

UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS

ESCUELA DE COMPUTACIÓN



**SOLUCIÓN DE INTELIGENCIA DE NEGOCIO PARA LA GENERACIÓN Y  
PRESENTACIÓN DE INDICADORES DEL PROCESO DE INTERMEDIACIÓN  
O SUBASTA DE PRODUCTOS RELACIONADOS CON EL GAS Y  
PETRÓLEO**

Trabajo Especial de Grado presentado ante la ilustre  
Universidad Central de Venezuela por  
Br. Ariam Salazar.

Para optar al título de Licenciado en Computación

**Tutor:**

Prof. Franky Uzcátegui

Mayo, 2018

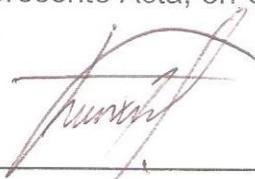
UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE CIENCIAS  
ESCUELA DE COMPUTACIÓN

### ACTA

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado titulado "Solución de Inteligencia de Negocio para la generación y presentación de indicadores del proceso de intermediación o subasta de productos relacionados con el gas y petróleo" y presentado por la bachiller: Br. Ariam Andreina Salazar Vivas titular de la Cédula de Identidad V-22.030.829, a los fines de optar al título de **Licenciada en Computación**, dejamos constancia de lo siguiente:

Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 24 de Mayo de 2018, a las 9,30 AM horas, para que la autora lo defendieran en forma pública, la que esta hizo en la Sala PB3 de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondió a las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobar con la nota de 19 puntos.

En fe de lo cual se levanta la presente Acta, en Caracas el día 24 de Mayo de 2018.

  
\_\_\_\_\_  
**Prof. Franky Uzcátegui**

Tutor

  
\_\_\_\_\_  
**Profa. Concettina Di Vasta**

Jurado

  
\_\_\_\_\_  
**Prof. José Mirabal**

Jurado

## **AGRADECIMIENTOS**

Cada persona tiene su meta personal y sus propios sueños para cumplir, puedo decir este es uno de los logros de mi vida, sin decir que es el más importante. Todos estos 6 años de estudios, de nuevos aprendizajes, de vivir nuevas experiencias, de grandes esfuerzos, de momentos buenos, de momentos malos, todo este tiempo se resume en este trabajo especial de grado. Un logro como este no se puede alcanzar sin personas que te ayuden y te apoyen a lo largo del camino, por esto, quiero darles las gracias a cada una de las personas que me apoyaron y ayudaron directamente o indirectamente en el desarrollo de mi trabajo especial de grado.

Para empezar quisiera darle las gracias a mi mamá María Vivas por ser una de la persona más importante de mi vida, todo lo que soy hoy en día te lo debo a ti, a la educación y a todas las cosas que me has enseñado como madre, cada día que pasa llego a ver en tu cara el orgullo que sientes por mí.

Quisiera de igual manera darles las gracias a mi papá Ricardo Salazar, por ser una persona incondicional en mi día a día, por tenerme paciencia con todo este proceso y por siempre apoyarme en todo lo que hago, y quisiera que estuvieras presente en todos los logros que tenga.

En esta carrera conocí a personas que hoy en día son esenciales para mí y que quiero conservar toda mi vida, quisiera agradecer a mis amigos que hacen que mi vida sea cada día más feliz, gracias por estar en cada momento esencial en mi vida, ustedes son como mi familia y aunque no estén físicamente conmigo en este momento sé que están de corazón y que comparten y sienten este logro tanto como yo, y tengo la fe en que cada uno cumplirá sus sueños y sus metas porque se lo merecen.

En especial quisiera agradecer a mi mejor amiga Fabiola por ser aquella persona que me impulsa a ser mejor persona y mejor profesional día a día, por ser la única constante en mi vida, por estar presente en cada momento todos estos años de amistad, y por último gracias por ser esa hermana que Dios me dio.

Quisiera agradecer a todas las personas involucradas en mi trabajo, sin ellas este proyecto no fuera sido desarrollado, gracias por su apoyo, por sus aprobaciones, por sus correcciones y por su participación en este maravilloso proyecto, que podamos seguir construyendo cosas nuevas para la evolución de la organización.

En último lugar pero no menos importante, quisiera agradecer a mi tutor Franky Uzcátegui por ser parte de mi trabajo especial de grado, por toda la paciencia que me ofreció, por los nuevos conocimientos que me transmitió, por su profesionalismo, por la dedicación de su tiempo para lograr que esto se desarrollara de manera correcta.

Finalmente quisiera agradecer a Dios y a la Virgen Rosa Mística, sin ellos nada de esto existiera ni fuera posible.

## DEDICATORIA

A mi madre que ha sido el ejemplo a seguir en todos mis logros, por enseñarme los valores de la familia, la importancia de ser constante en la vida, el nunca rendirse cuando las cosas se ponen difíciles, por darme fuerzas cuando más la necesito, por siempre apoyarme en todas mis ideas, por creer en mí, por cuidarme, por estar conmigo en los buenos y malos momentos, gracias por su amor.

A mi padre por estar ahí cuando lo más necesito y por apoyarme en mis decisiones, por cuidarme.

A mis amigos, aquellas personas que yo considero mi familia, mi lugar seguro, mi zona confort, porque nunca dudaron de mí, siempre me recordaron que cada trabajo tiene su recompensa y que las metas se cumplen con mucho trabajo y dedicación para obtenerlas.

Finalmente quiero terminar con una frase que tuve presente en todo este proceso:

**“Lo que es para ti, es para ti...sólo tienes que trabajar para conseguirlo”**

Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Computación

**SOLUCIÓN DE INTELIGENCIA DE NEGOCIO PARA LA GENERACIÓN Y  
PRESENTACIÓN DE INDICADORES DEL PROCESO DE INTERMEDIACIÓN  
O SUBASTA DE PRODUCTOS RELACIONADOS CON EL GAS Y  
PETRÓLEO**

**Autor:** Br. Ariam Salazar.

**Tutor:** Prof. Franky Uzcátegui.

**Fecha:** Mayo, 2018

**RESUMEN**

Los procesos de intermediación, compra y subasta de productos relacionados con el gas y petróleo requieren de un conjunto de indicadores que permiten su seguimiento y control. El siguiente Trabajo Especial de Grado tiene como finalidad diseñar, definir una solución de Inteligencia de Negocio, con el objetivo de medir el desempeño de estos procesos en un portal web específico. Para ellos se presentan conceptos, metodologías, esquemas y arquitecturas, que permiten la generación de indicadores de gestión, que ayuden a tomar mejores decisiones en la organización y manejar de manera más eficiente y eficaz la información. El desarrollo de la solución de Inteligencia de Negocio, se basa en la metodología ascendente de Ralph Kimball. Y de igual manera se usará la metodología de Scrum para poder tener un mayor control y manejo a lo largo del desarrollo del proyecto. Obteniendo como beneficio un mayor control y gestión de los procesos ejecutados en una organización aumentando de esta manera la operatividad de la misma.

**Palabras Clave:** Inteligencia de Negocio, Almacén de datos, Indicadores de gestión, Tableros de control.

## ÍNDICE DE CONTENIDO

<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO 1</b> .....	4
<b>PROBLEMA DE INVESTIGACIÓN</b> .....	4
1.1 Planteamiento del Problema .....	4
1.2 Objetivos .....	6
1.2.1 Objetivo General .....	6
1.2.2 Objetivos Específicos .....	6
1.3 Solución Propuesta .....	6
1.4 Justificación .....	8
1.5 Alcance .....	9
<b>CAPÍTULO 2</b> .....	10
2.1 Almacén de datos (Datawarehouse) .....	10
2.1.1 Antecedentes .....	10
2.1.2 Definición.....	11
2.1.3 Características de un almacén de datos .....	12
2.1.4 Elementos para una arquitectura de un almacén de datos .....	14
2.2 Bodega de datos (Datamart).....	16
2.2.1 Tipos de datamart. ....	16
2.2.2 Datawarehouse vs Datamart.....	17
2.3 Modelo Dimensional .....	18
2.3.1 Dimensión .....	19
2.3.2 Tabla dimensión .....	20
2.3.3 Hechos .....	20
2.3.4 Tabla de hechos.....	21
2.3.5 Jerarquía .....	21
2.3.6 Granularidad.....	22

2.3.7	Agregación .....	22
2.3.8	Cubo.....	22
2.3.9	Esquema de estrellas .....	22
2.3.10	Esquema copo de nieve .....	23
2.3.11	Esquema constelación .....	24
2.3.12	Beneficios del Modelo Dimensional.....	25
2.4	Inteligencia de Negocio.....	26
2.4.1	Antecedentes .....	26
2.4.2	Definición.....	27
2.4.3	Inteligencia de Negocio a nivel operativo, táctico y estratégico. 28	
2.4.4	Beneficios de la inteligencia de negocio.....	29
2.4.5	Arquitectura de una solución de Inteligencia de Negocio .....	30
2.5	Importancia de la Inteligencia de Negocio .....	33
2.6	Indicadores de gestión .....	34
2.6.1	Categorías de indicadores de gestión .....	34
2.6.2	Propósito y beneficios de los indicadores de gestión .....	35
2.6.3	¿Por qué medir y para qué? .....	36
2.7	Herramientas de Inteligencia de Negocio .....	37
2.8	Tipos de productos BI .....	37
2.9	Conceptos orientados a los procesos del portal web .....	45
2.9.1	Compra.....	45
2.9.2	Subastas electrónicas .....	45
2.9.3	Intermediación.....	46
2.9.4	Surplus .....	47
<b>CAPÍTULO 3</b>	.....	<b>49</b>
MARCO METODOLÓGICO	.....	49
3.1	Metodología según Ralph Kimball (Ascendente). .....	49

3.1.1	Planificación del proyecto.....	52
3.1.2	Definición de requerimientos del Negocio .....	53
3.1.3	Modelo Dimensional.....	54
3.1.3.1	Elegir el proceso de negocio .....	54
3.1.3.2	Establecer el nivel de granularidad .....	54
3.1.3.3	Elegir las dimensiones .....	55
3.1.3.4	Identificar las tablas de hechos y medidas.....	55
3.1.3.5	Modelo gráfico de alto nivel.....	55
3.1.3.6	Identificación de atributos de dimensiones y tablas de hechos	56
3.1.4	Modelo Físico .....	57
3.1.5	Diseño e implementación del subsistema de ETL.....	57
3.1.6	Diseño Técnico de la arquitectura .....	57
3.1.7	Selección e instalación de productos .....	58
3.1.8	Especificación y desarrollo de aplicaciones de BI .....	59
3.1.8.1	Informes estándar Los informes estándar son la base del espectro de aplicaciones de BI .....	59
3.1.8.2	Aplicaciones analíticas Las aplicaciones analíticas son más complejas que los informes estándar .....	60
3.1.9	Implementación .....	60
3.1.10	Mantenimiento y crecimiento.....	61
3.1.11	Administración de Proyecto.....	62
3.2	Metodología Scrum.....	62
3.2.1	Características de Scrum .....	63
<b>CAPÍTULO 4.....</b>		<b>65</b>
4.1	Fases del Proyecto. ....	66
4.1.1	Planificación del Proyecto. ....	66
4.1.2	Definición de los Requerimientos del Negocio .....	69

4.1.3	Diseño Técnico de la Arquitectura.....	74
4.1.4	Selección de Productos e Instalación.....	76
4.1.5	Modelado Dimensional.....	78
4.1.5.1	Definir el Proceso de Negocio.....	78
4.1.5.2	Identificar el Nivel de Granularidad.....	78
4.1.5.3	Definir las Dimensiones.....	80
4.1.5.4	Identificación de los Hechos y las Tablas de Hechos.....	81
4.1.5.5	Modelo gráfico de alto nivel.....	83
4.1.6	Diseño Físico.....	87
4.1.7	Diseño y Desarrollo de Procesos ETL.....	88
4.1.8	Especificaciones de las Aplicaciones Analíticas.....	93
4.1.9	Desarrollo de las Aplicaciones Analíticas.....	95
4.1.10	Implementación.....	107
4.1.10.1	Pruebas de Calidad.....	108
4.1.10.2	Manual de usuarios.....	127
<b>CONCLUSIONES Y RECOMENDACIONES.....</b>		<b>152</b>
<b>GLOSARIO DE TÉRMINOS.....</b>		<b>154</b>
<b>REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES.....</b>		<b>156</b>

## ÍNDICE DE ILUSTRACIONES

Ilustración 1. Arquitectura de la solución de Inteligencia de Negocio propuesta	7
Ilustración 2. Modelo Dimensional.....	18
Ilustración 3. Esquema de estrella con una sola tabla de hechos con enlaces a varias tablas de dimensiones. Estrella. ....	23
Ilustración 4. Esquema de copo de nieve con dos dimensiones y tres niveles cada una.....	24
Ilustración 5. Esquema de constelación con un hecho y dos dimensiones que comparten un outrigger.. ....	25
Ilustración 6. Niveles de Inteligencia de Negocio .....	28
Ilustración 7. Arquitectura de una solución de Inteligencia de Negocio ....	<b>¡Error! Marcador no definido.</b>
Ilustración 8. Producto publicado por BON (Buy option now).....	45
Ilustración 9. Producto publicado por subasta.....	46
Ilustración 10. Producto publicado por Contacto Vendedor. ....	47
Ilustración 11. Sección de Surplus del portal web. ....	48
Ilustración 12. Metodología Bottom-Up .....	50
Ilustración 13. Ciclo de vida de Kimball .....	<b>¡Error! Marcador no definido.</b>
Ilustración 14. Ejemplo de Modelo final de alto nivel de la sesión inicial de diseño .....	56
Ilustración 15. Ciclo de vida de la metodología de SCRUM .....	64
Ilustración 16. Arquitectura de la solución de Inteligencia de Negocio.....	74
Ilustración 17. Relaciones Jerárquicas del modelo dimensional de la solución propuesta .....	81
Ilustración 18. Modelo gráfico de alto nivel de Visitas .....	84
Ilustración 19. Modelo gráfico de alto nivel de Contacto Vendedor.....	84
Ilustración 20. Modelo gráfico de alto nivel de Publicaciones .....	85
Ilustración 21. Modelo gráfico de alto nivel de Surplus publicación.....	85
Ilustración 22. Modelo gráfico de alto nivel de Compra Inmediata .....	86
Ilustración 23. Modelo gráfico de alto nivel de Subastas.....	86
Ilustración 24. Modelo físico del Almacén de Datos implementado en la base de datos .....	87
Ilustración 25. Repositorio de Archivos de la primera fase de ETL .....	89

Ilustración 26. Repositorio de Archivos de la segunda fase de ETL.....	89
Ilustración 27. ETL para cargar la dimensión Usuario.....	90
Ilustración 28. ETL para cargar la Tabla Hecho “fact_visita”.....	91
Ilustración 29. Trabajo (Job) que ejecuta todas las transformaciones.....	92
Ilustración 30. ETL para cargar la dimensión categoría. <b>¡Error! Marcador no definido.</b>	
Ilustración 31. ETL para cargar Tabla Hecho “fact_visita” del almacén de datos final..... <b>¡Error! Marcador no definido.</b>	
Ilustración 32. Trabajo (Job) que ejecuta todas las transformaciones..... <b>¡Error! Marcador no definido.</b>	
Ilustración 33. Reportes iniciales del portal web desde el administrador. ....	93
Ilustración 34. Vista de la base de datos desarrollada con Navicat.....	95
Ilustración 35. Conexión a una base de datos desde Tableau.....	96
Ilustración 36. Selección de la base de datos con que se trabajará en la creación de reportes y tableros de control (dashboard).....	97
Ilustración 37. Elementos de la interfaz de Tableau Desktop para la creación de las hojas de trabajo. ....	98
Ilustración 38. Elementos de la interfaz de Tableau Desktop para la creación de los tableros de control (dashboard). ....	99
Ilustración 39. Tablero de Control de visitas del portal web. ....	100
Ilustración 40. Tablero de Control de Surplus del portal web. ....	101
Ilustración 41. Tablero de Control de Subastas del portal web .....	101
Ilustración 42. Tablero de Control de Publicaciones del portal web. ....	102
Ilustración 43. Tablero de Control de Intermediación de Ventas del portal web. ....	102
Ilustración 44. Tablero de Control de Intermediación de Ventas del portal web. ....	103
Ilustración 45. Tablero de Control de Resumen general de los reportes.....	103
Ilustración 46. Tableau Public. ....	104
Ilustración 47. Página Inicio del portal web administrativo para acceder a los tableros de control, y a los reportes. ....	105
Ilustración 48. Sección del portal web administrativo que muestra el resumen de los reportes más destacados del portal web principal. ....	105

Ilustración 49. Sección del portal web administrativo para acceder a los tableros de control por procesos del portal web principal. ....	106
Ilustración 50. Vista interna de cada tablero de control (dashboard). En este caso del tablero de control de Rendimiento de las publicaciones del portal web. ..	106
Ilustración 51. Vista interna de cada tablero de control (por libro) desde Tableau Public. ....	107
Ilustración 52. Consulta de Base de datos fuente de la dimensión Usuario...	108
Ilustración 53. Consulta del Almacén de datos de la dimensión usuario .....	109
Ilustración 54. Consulta de Base de datos fuente de las publicaciones .....	110
Ilustración 55. Consulta del almacén de datos de la tabla hecho publicaciones .....	110
Ilustración 56. Cantidad de visitas activas en el portal web (desde Almacén de datos). ....	111
Ilustración 57. Cantidad de visitas activas en el portal web (desde Fuente de datos). ....	112
Ilustración 58. Cantidad de visitas por Sección en el portal web (desde Almacén de datos), en este caso son la cantidad de visitas en las secciones de Marcas, Categorías, Información. ....	112
Ilustración 59. Cantidad de visitas por Sección en el portal web (desde fuente de datos), en este caso son la cantidad de visitas en las secciones de Marcas, Categorías, Información. ....	113
Ilustración 60. Cantidad de visitas por Sección en el portal web (desde fuente de datos), en este caso son la cantidad de visitas en las secciones de Home y Surplus. ....	113
Ilustración 61. Cantidad de visitas por Sección en el portal web (desde el almacén de datos), en este caso son la cantidad de visitas en las secciones de Home y Surplus. ....	114
Ilustración 62. Cantidad de publicaciones (productos activos) activas en el portal web (desde fuente de datos). ....	114
Ilustración 63. Cantidad de publicaciones (productos activos) activas en el portal web (desde almacén de datos). ....	115
Ilustración 64. Cantidad de publicaciones (productos inactivos) inactivos en el portal web (desde fuente de datos. ....	115

Ilustración 65. Cantidad de publicaciones (productos inactivos) inactivos en el portal web (desde almacén de datos). .....	116
Ilustración 66. Cantidad de productos publicados por marca (desde la fuente de datos). .....	117
Ilustración 67. Cantidad de productos publicados por marca (desde el almacén de datos). .....	117
Ilustración 68. Cantidad de surplus publicados (desde el almacén de datos). .....	118
Ilustración 69. Cantidad de surplus publicados (desde la fuente de datos). ....	118
Ilustración 70. Promedio de costo de Surplus (desde el almacén de datos) ..	119
Ilustración 71. Promedio de costo de Surplus (desde la fuente de datos). ....	119
Ilustración 72. Cantidad de veces que han sido contactados un producto (desde la fuente de datos). .....	120
Ilustración 73. Cantidad de veces que han sido contactado un producto (desde el almacén de datos). .....	121
Ilustración 74. Cantidad de productos contactados del portal web (desde la fuente de datos). .....	121
Ilustración 75. Cantidad de productos contactados del portal web (desde el almacén de datos). .....	122
Ilustración 76. Productos contactados en el portal web por categoría (desde la fuente de datos). .....	122
Ilustración 77. Productos contactados en el portal web por categoría (desde el almacén de datos). .....	123
Ilustración 78. Cantidad de subastas (desde la fuente de datos). .....	123
Ilustración 79. Cantidad de subastas (desde el almacén de datos). .....	124
Ilustración 80. Ganancias de compras inmediatas en el portal web (desde la fuente de datos). .....	124
Ilustración 81. Ganancias de compras inmediatas en el portal web (desde el almacén de datos). .....	125
Ilustración 82. Cantidad de compras inmediatas en el portal web (desde la fuente de datos). .....	125
Ilustración 83. Cantidad de compras inmediatas en el portal web (desde el almacén de datos). .....	126
Ilustración 84. Ganancia de las publicaciones del portal web (desde la fuente de datos). .....	126

Ilustración 85. Ganancia de las publicaciones del portal web (desde el almacén de datos). .....	127
Ilustración 86. URL de acceso del portal web administrativo. ....	127
Ilustración 87. Interfaz para ingresar en el portal web administrativo.....	128
Ilustración 88. Página principal del portal web administrativo. ....	128
Ilustración 89. Menú principal del portal web administrativo.....	129
Ilustración 90. Menú principal del portal web administrativo (con segundo nivel de despliegue).....	130
Ilustración 91. Tablero de control del resumen general de los reportes del portal web.....	131
Ilustración 92. Tablero de control del resumen general de los reportes del portal web.....	131
Ilustración 93. Opciones de compartir, descargar, full pantalla del dashboard. ....	133
Ilustración 94. Interfaz de indicadores del portal web administrativo.....	133
Ilustración 95. Tablero de control de “Control de Visitas del portal web”.....	134
Ilustración 96. Tablero de control de “Control de Visitas del portal web.....	135
Ilustración 97. Tablero de control de “Control de Visitas del portal web.....	136
Ilustración 98. Tablero de control de “Rendimiento en las ventas de productos”. ....	137
Ilustración 99. Tablero de control de “Rendimiento en las ventas de productos”. ....	137
Ilustración 100. Tablero de control de “Control de Subastas del portal web”. ....	139
Ilustración 101. Tablero de control de “Control de Subastas del portal web”. ....	139
Ilustración 102. Tablero de control de “Rendimiento de las publicaciones del portal web”.....	140
Ilustración 103. Tablero de control de “Rendimiento de las publicaciones del portal web”.....	141
Ilustración 104. Tablero de control de “Rendimiento de las publicaciones del portal web”.....	142
Ilustración 105. Tablero de control de “Rendimiento de las publicaciones del portal web”.....	143
Ilustración 106. Botón para volver a la interfaz de todos los indicadores. ....	144
Ilustración 107. Pregunta #1 de la encuesta .....	147

Ilustración 108. Pregunta #2 de la encuesta .....	147
Ilustración 109. Pregunta #3 de la encuesta .....	148
Ilustración 110. Pregunta #4 de la encuesta .....	148
Ilustración 111. Pregunta #5 de la encuesta .....	149
Ilustración 112. Pregunta #6 de la encuesta .....	149
Ilustración 113. Pregunta #7 de la encuesta .....	150
Ilustración 114. Pregunta #8 de la encuesta .....	150

## ÍNDICE DE TABLAS

Tabla 1. Herramientas de tecnologías en Inteligencia de Negocio y su descripción. ....	39
Tabla 2. Productos a entregables de la planificación de SCRUM. ....	65
Tabla 3. Actividades a realizar para la implementación de la solución de Inteligencia de Negocio. ....	67
Tabla 4. Indicadores a desarrollar en la Solución de Inteligencia de Negocio.	70
Tabla 5. Herramientas para desarrollar la solución de Inteligencia de Negocio. ....	77
Tabla 6. Dimensiones del Modelo Dimensional de la solución. ....	80
Tabla 7. Hechos de la tabla "Fact_visita" .....	81
Tabla 8. Hechos de la tabla "Fact_contacto_vendedor" .....	82
Tabla 9. Hechos de la tabla "Fact_publicaciones" .....	82
Tabla 10. Hechos de la tabla "Fact_compra_inmediata" .....	82
Tabla 11. Hechos de la tabla "Fact_subasta" .....	83
Tabla 12. Hechos de la tabla "Surplus_publicacion" .....	83
Tabla 13. Pautas y estándares de la aplicación del portal web administrativo. ....	<b>¡Error! Marcador no definido.</b>

## INTRODUCCIÓN

En la actualidad, existen diversos portales web que permiten realizar procesos de compras, ventas, subastas, entre otras operaciones de algún producto en específico en un área en particular. Sin embargo, en el área del gas y petróleo son muy escasos los portales web que cumplan con todas estas funcionalidades en uno sólo. Normalmente el proceso de buscar, obtener o adquirir un producto relacionado con el gas y petróleo puede ser un poco engorroso, y puede tomar mucho tiempo.

El portal web que se estudia en el siguiente trabajo es una plataforma global de Marketplace diseñada para facilitar la compra de maquinarias, equipos, herramientas, partes, y todos aquellos elementos relacionados con la industria del gas, energía y petróleo. Teniendo una serie de ventajas como la gran variedad de productos que se puedan encontrar en el portal, poder crear presupuestos directamente desde el portal, la comunicación directa con las personas que vendan estos productos, entre otros.

Este portal permite realizar procesos de intermediación, subasta, ventas de productos, a su vez constantemente se hacen registros de empresas o usuarios para poder hacer esas operaciones en el portal. Uno de los procesos más importantes del portal es el proceso de publicación de un producto, ya que es un proceso que pasa por varias etapas, finalizando con el pago de la publicación, existe otro proceso de solicitud para publicar en una sección denominada "Surplus", que es un listado de productos sobrantes de un proyecto relacionado con el petróleo para posteriormente venderlos.

Todos los procesos nombrados anteriormente generan una gran cantidad de datos de los cuales se requieren obtener información útil a través de reportes e indicadores, que ayuden a facilitar el control y seguimiento de ellos y de esta manera poder tomar mejores decisiones que permitan el avance y la escalabilidad operativa de la organización.

Todo esto tiene como resultado un almacenamiento de gran cantidad de datos que necesitan ser procesados y transformados en información que pueda ser útil, sin embargo, no todos los portales en esta área del petróleo y gas poseen un sistema analítico encargado de realizar el análisis y monitoreo en función a los datos recolectados, y eso afecta a la toma de acciones.

Es por ello la necesidad de implementar una solución de Inteligencia de Negocio, como una alternativa tecnológica que permita el manejo de los datos, empezando por la extracción, depuración y transformación de los mismos, hasta la visualización de la información a través de herramientas tecnológicas de fácil uso que ayuden a organizaciones relacionadas con el gas y petróleo a tomar mejores decisiones para incrementar sus procesos operativos dentro de la organización.

Este Trabajo Especial de Grado (TEG) se basa fundamental en el desarrollo de una solución de Inteligencia de Negocio debido a la necesidad de generar y presentar indicadores de los procesos de intermediación, compras o subastas de productos relacionados con el gas y petróleo. De acuerdo con lo señalado, el trabajo de investigación se estructuró en cuatro capítulos, a saber:

**Capítulo 1. Planteamiento del problema:** En este capítulo se define el planteamiento del problema, el objetivo general y los objetivos específicos a cumplir en este trabajo de investigación. Se argumenta la solución propuesta y justificación de la solución que se presenta.

**Capítulo 2. Marco conceptual:** En él se describen los conceptos básicos requeridos para tener un mayor conocimiento de los procesos en el análisis, diseño y construcción de una solución de Inteligencia de Negocio. De igual manera, se definen y comparan herramientas tecnológicas que apoyan a la solución de Inteligencia de Negocio, herramientas de base de datos y herramientas de extracción, transformación y carga de los datos, con el fin de conocer mejor sus aspectos y características.

**Capítulo 3. Marco metodológico:** Se presenta y se explica de manera ampliada la metodología usada para desarrollar la solución de Inteligencia de Negocio. Se procede a definir y describir cada fase que involucra los procesos de construcción de una solución de Inteligencia de Negocio.

**Capítulo 4. Marco Aplicativo:** Se especifican las actividades que se realizaron en cada una de las fases de la metodología definida para la solución planteada.

Por último, se presentan las conclusiones, referencias bibliográficas y digitales consultadas y utilizadas para la elaboración de este trabajo de investigación.

# CAPÍTULO 1

## PROBLEMA DE INVESTIGACIÓN

### 1.1 Planteamiento del Problema

La capacidad para tomar decisiones de negocio precisas y de forma rápida se ha convertido en una de las claves para que una empresa llegue al éxito. Sin embargo, los sistemas de información tradicionales, suelen presentar una estructura poco flexible para este fin.

Aunque su diseño se adapta con mayor o menor medida para manejar los datos de la organización, no permite obtener la información de los mismos, y mucho menos extrapolar el conocimiento almacenado en el día a día de las bases de datos de la misma.

Los portales web que se encargan de la intermediación, compra, subasta, de productos relacionados con el gas y petróleo se ven directamente afectados por este problema, ya que para estos portales es de vital importancia tener conocimiento y seguimiento de los procesos que realizan, y así mismo poder tener un mayor control de ellos. Para esto, se utilizan diferentes mecanismos, técnicas y herramientas para realizar el control y seguimiento de los procesos previamente mencionados.

La planificación, ejecución y control de los procesos de intermediación, compra, venta, subastas en un portal web requieren de un constante control y seguimiento para que la organización tenga una mejor gestión, si no se realizan estas acciones la organización podría tener pérdidas de recursos o graves consecuencias de producción.

Este tipo de portales web manejan normalmente un gran volumen de datos almacenados y están en un constante crecimiento en el manejo de datos. Sin embargo ese crecimiento no significa que se tenga una información real de ella, debido a que puede ser difícil procesar todos esos datos con los métodos clásicos y extraer información que sería de gran utilidad para la toma de decisiones.

Todo esto es con el objetivo que a través del análisis de reportes gerenciales se pueda obtener información útil y poder encontrar dificultades y plantear mejores estrategias en la organización.

La mayoría de las organizaciones cada vez más manejan mayor cantidad de datos pero no tienen el tiempo para analizarlos y conseguir el máximo provecho de ellos. El problema radica en poder obtener y generar información que pueda ser manejada y conocida a nivel gerencial en la organización, para poder responder todas las preguntas asociadas a la situación de los procesos ejecutados en ella y tomar de una manera más acertada las decisiones.

## **1.2 Objetivos**

### **1.2.1 Objetivo General**

Desarrollar una solución de Inteligencia de Negocio para la generación y presentación de indicadores del proceso de intermediación o subasta de productos relacionados con el gas y petróleo.

### **1.2.2 Objetivos Específicos**

1. Definir los requerimientos de negocios obtenidos por parte de la organización acerca de los procesos realizados en el portal web de intermediación, ventas y subastas de productos relacionados con el gas y petróleo.
2. Diseñar el almacén de datos y los procesos ETLs relacionados con los requerimientos obtenidos por parte de la organización.
3. Desarrollar un almacén de datos y los procesos ETIs que sean adaptables a la Solución de Inteligencia de Negocio a partir de los requerimientos previamente identificados.
4. Desarrollar los indicadores de gestión en base a los requerimientos definidos por parte de la organización.
5. Realizar las pruebas funcionales, pruebas de usabilidad de la aplicación de BI, y pruebas de calidad de los datos almacenados en el almacén de datos.

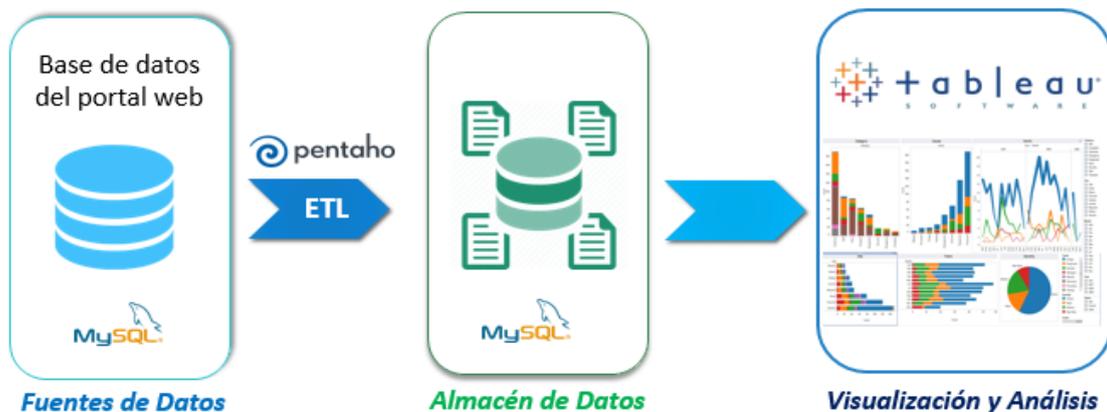
## **1.3 Solución Propuesta**

Se propone una solución de Inteligencia de Negocio que permita obtener y presentar indicadores de gestión que sean necesarios y tengan relevancia relacionados con los procesos de ventas, compras y subastas de productos en el área petrolera de un portal web en específico.

Como se observa en la ilustración 1, la solución está conformada por los siguientes componentes:

- Las fuentes de datos provenientes de la base de datos del portal web, las cuales son administradas por el Sistema Manejador de Base de Datos MySQL.

- Un conjunto de procesos de Extracción, Transformación y Carga (ETL o ETC) que permiten extraer los datos desde las fuentes de datos (base de datos), realizar procesos de depuración, transformación y operaciones sobre los datos y por último cargarlos en el almacén de datos. Estos procesos se realizarán con la herramienta de Pentaho denominada “Pentaho Data Integration”, debido a que estos son de gran importancia para la creación del almacén de datos que se utiliza por la solución de Inteligencia de Negocio.
- Un almacén de datos, basado en un esquema constelación, a través de la implementación del modelo dimensional, utilizando SMBD MySQL para su administración. Este almacén de datos está acorde a los indicadores que pudieron ser obtenidos de los datos previamente almacenados, y a su vez que puedan ser útiles de acuerdo con las necesidades que vayan surgiendo en la organización.
- Tableros de control (Dashboard) con la herramienta Tableau que nos ayude a visualizar los indicadores de gestión obtenidos.



**Ilustración 1. Arquitectura de la solución de Inteligencia de Negocio propuesta (Cano, 2007)**

## **1.4 Justificación**

Actualmente, vivimos en una época donde la información es clave para alcanzar la misión, la visión y los objetivos de una organización, en donde los ejecutivos requieren un acceso rápido de esta información para ayudar a tomar las mejores decisiones, y de esta manera obtener mejores beneficios para mantener y mejorar el desempeño funcional de la organización.

El escaso tiempo que tienen las organizaciones para el análisis de los datos hace más complicado la toma de decisiones en un ambiente sujeto a constantes cambios.

Debido a que en el portal web se puede manejar una gran cantidad de datos de todos los procesos de negocios de la organización, las personas responsables de llevar el control y seguimiento deben constantemente monitorear la evolución de los mismos, y aquí nace la necesidad de una herramienta que ayude a encontrar prioridades sobre los procesos de negocios que requieran importancia, así como también tener un acceso a información importante y tomar decisiones efectivas para mejorar los procesos.

Por otro lado, al tener indicadores de gestión se logra una mejor administración y control de los procesos de negocios del portal web, ya que los responsables del control y seguimiento de estos procesos, no invertirán más tiempo del necesario analizando cada uno por separado, sino que podrán ver en los indicadores aquellos procesos del negocio que necesiten mayor atención.

Además con la generación de reportes se analiza de manera más detallada esta información para buscar una solución a los problemas que se presenten, estos reportes se generan de manera inmediata. Debido a esto, la implementación de una Solución de Inteligencia de Negocio, es una alternativa tecnológica, permitiendo tomar mejores decisiones en todos los niveles de una organización a través del manejo de la información que se obtenga, desde la extracción, depuración y transformación de los datos, hasta al exploración y análisis de la información mediante herramientas de fácil uso para los usuarios, con el fin de ayudar a las organizaciones encargadas de intermediación, compras, ventas, subastas de productos relacionados con el gas y petróleo.

## **1.5 Alcance**

El desarrollo de la solución de Inteligencia de Negocio se enfoca en el diseño y despliegue de los indicadores asociados a una organización petrolera que se enfoca en los servicios de intermediación de compra, y subastas de productos relacionados a gas y petróleo en un portal web, para poder obtener indicadores de los procesos mencionados.

A su vez, se les proporciona un conjunto de métricas para cada usuario definido por la función que realizan en la organización, ayudando a estos a tomar mejores decisiones en la organización, apoyándose en la información generada por la aplicación de BI, esta información ser mostrada en el tablero de control (dashboard), un dashboard es una herramienta de Inteligencia de Negocio, que es una exposición gráfica de los KPI's que afectan a la consecución de los objetivos de organización, utilizando los datos de la base de datos del portal web como fuente para la carga del almacén de datos.

## **CAPÍTULO 2**

### **MARCOCONCEPTUAL**

A continuación se describen ciertos conceptos básicos requeridos para tener un mayor conocimiento de los procesos en el análisis, diseño y construcción de una solución de Inteligencia de Negocio.

#### **2.1 Almacén de datos (Datawarehouse)**

##### **2.1.1 Antecedentes**

###### **1960's**

- Nacen las tarjetas perforadas y cintas magnéticas en donde el acceso se hacía de forma secuencial.

###### **1970 's**

- Nacen los discos de acceso directo (dasd).
- Bases de datos jerárquicas y reticulares.
- Primeras arquitecturas de oltp trabajando en línea.
- Sistemas cliente, nacen las terminales tontas.

###### **1980's.**

- Surgen las pc y emulaciones.
- Nace 4gl que es sql con oltp y olap.
- Toman importancia los administradores de bases de datos, esto debido a que se empiezan a separar los datos de las aplicaciones.

###### **1990's**

- Surgen problemas por hacer consulta en la misma base de datos en la cual se hacen las transacciones, por lo que surge la necesidad de separar las bases de datos en OLAP y OLTP.

### **2.1.2 Definición.**

Inmon (2000), define los Almacenes de Datos (Data Warehouse) como una colección de datos orientados a temas, integrados, no volátiles y variantes en el tiempo que apoyan a las decisiones de la gerencia.

Los almacenes de datos dan acceso a datos para análisis complejos dando respuesta a las demandas de alto rendimiento de datos e información de una organización.

A su vez Ralph Kimball, define un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un data warehouse no era más que: "la unión de todos los Data marts de una entidad". Defiende por tanto una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos.

#### *Función de un almacén de datos*

En un almacén de datos lo que se quiere es contener datos que son necesarios o útiles para una organización, es decir, que se utiliza como un repositorio de datos para posteriormente transformarlos en información útil para el usuario. Un almacén de datos debe entregar la información correcta a la gente indicada en el momento óptimo y en el formato adecuado.

El almacén de datos da respuesta a las necesidades de usuarios expertos, utilizando Sistemas de Soporte a Decisiones (DSS), Sistemas de información ejecutiva (EIS) o herramientas para hacer consultas o informes. Los usuarios finales pueden hacer fácilmente consultas sobre sus almacenes de datos sin tocar o afectar la operación del sistema.

En el funcionamiento de un almacén de datos son muy importantes las siguientes ideas:

1. Integración de los datos provenientes de bases de datos distribuidas por las diferentes unidades de la organización y que con frecuencia tendrán diferentes estructuras (fuentes heterogéneas). Se debe facilitar una descripción global y un análisis comprensivo de toda la organización en el almacén de datos.
2. Separación de los datos usados en operaciones diarias de los datos usados en el almacén de datos para los propósitos de divulgación, de ayuda en la toma de decisiones, para el análisis y para operaciones de control. Ambos tipos de datos no deben coincidir en la misma base de datos, ya que obedecen a objetivos muy distintos y podrían interponerse entre sí.

### 2.1.3 Características de un almacén de datos

El término Datawarehouse fue acuñado por primera vez por Bill Inmon, y se traduce literalmente como *almacén de datos*. No obstante, y como cabe suponer, es mucho más que eso. Según definió el propio Bill Inmon, un datawarehouse se caracteriza por ser:

- **Integrado:** Los datos almacenados en el datawarehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático:** Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.

Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del datawarehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:** El tiempo es parte implícita de la información contenida en un datawarehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el datawarehouse sirve, entre otras cosas, para realizar análisis de tendencias.
  
- **No volátil:** El almacén de información de un datawarehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del datawarehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

#### Ventajas de un almacén de datos.

1. Proporciona información clave para la toma de decisiones empresariales.
2. Mejora la calidad de las decisiones tomadas.
3. Especialmente útil para el medio y largo plazo.
4. Son sistemas relativamente sencillos de instalar si las fuentes de datos y los objetivos están claros.
5. Muy útiles para el almacenamiento de análisis y consultas de históricos.
6. Proporciona un gran poder de procesamiento de información.
7. Permite una mayor flexibilidad y rapidez en el acceso a la información.
8. Facilita la toma de decisiones en los negocios.
9. Las empresas obtienen un aumento de la productividad.
10. Proporciona una comunicación fiable entre todos los departamentos de la empresa.
11. Mejora las relaciones con los proveedores y los clientes.
12. Permite conocer qué está pasando en el negocio, es decir, estar siempre enterado de los buenos y malos resultados.
13. Transforma los datos en información y la información en conocimiento.

### Desventajas de un almacén de datos.

1. No es muy útil para la toma de decisiones en tiempo real debido al largo tiempo de procesamiento que puede requerir. En cualquier caso la tendencia de los productos actuales es la de solventar este problema convirtiendo la desventaja en una ventaja.
2. Requiere de continua limpieza, transformación e integración de datos.
3. Mantenimiento.
4. En un proceso de implantación puede encontrarse dificultades ante los diferentes objetivos que pretende una organización.
5. Una vez implementado puede ser complicado añadir nuevas fuentes de datos.
6. Requieren una revisión del modelo de datos, objetos, transacciones y además del almacenamiento.
7. Tienen un diseño complejo y multidisciplinar.
8. Tienen un alto coste.
9. Requieren sistemas, aplicaciones y almacenamiento específico.

#### **2.1.4 Elementos para una arquitectura de un almacén de datos**

Una Arquitectura Data Warehouse (Data Warehouse Architecture – DWA) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.

- **Base de datos operacional / Nivel de base de datos externo:** Los sistemas operacionales procesan datos para apoyar las necesidades operacionales críticas. Para hacer eso, se han creado las bases de datos operacionales históricas que proveen una estructura de procesamiento eficiente, para un número relativamente pequeño de transacciones comerciales bien definidas.

- **Nivel de acceso a la información:** El nivel de acceso a la información de la arquitectura datawarehouse, es el nivel del que el usuario final se encarga directamente.
- **Nivel de acceso a los datos:** El nivel de acceso a los datos de la arquitectura datawarehouse está involucrado con el nivel de acceso a la información para conversar en el nivel operacional. El nivel de acceso a los datos no solamente conecta DBMS diferentes y sistemas de archivos sobre el mismo hardware, sino también a los fabricantes y protocolos de red. Una de las claves de una estrategia data warehousing es proveer a los usuarios finales con “acceso a datos universales”.
- **Nivel de Directorio de Datos (Metadata):** A fin de proveer el acceso a los datos universales, es absolutamente necesario mantener alguna forma de directorio de datos o repositorio de la información metadata. La metadata es la información alrededor de los datos dentro de la empresa.
- **Nivel de Gestión de Procesos:** El nivel de gestión de procesos tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener la datawarehouse y la información del directorio de datos.
- **Nivel Datawarehouse (Físico):** En la datawarehouse (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos. En algunos casos, uno puede pensar del datawarehouse simplemente como una vista lógica o virtual de datos. En muchos ejemplos, el datawarehouse puede no involucrar almacenamiento de datos.
- **Nivel de Organización de Datos:** El componente final de la arquitectura datawarehouse es la organización de los datos. Se llama también gestión de copia o réplica, pero de hecho, incluye todos los procesos necesarios como seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas.

## **2.2 Bodega de datos (Datamart)**

Es un pequeño Datawarehouse, para un determinado número de usuarios, para un área funcional, específica de la compañía. También podemos definir que un DataMart es un subconjunto de una bodega de datos para un propósito específico.

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un Datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

### **2.2.1 Tipos de datamart.**

#### *Datamart OLAP*

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

#### *Datamart OLTP*

Pueden basarse en un simple extracto del datawarehouse, no obstante, lo común es introducir mejoras en su rendimiento aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser fact-tables reducidas y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries.

Los datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

1. Poco volumen de datos
2. Mayor rapidez de consulta
3. Consultas SQL y/o MDX sencillas
4. Validación directa de la información
5. Facilidad para la historización de los datos

### **2.2.2 Datawarehouse vs Datamart**

Según define Meta Group, "un DataMart es una aplicación de Data Warehouse, construida rápidamente para soportar una línea de negocio simple". Los Data Marts, tienen las mismas características de integración, no volatilidad, orientación temática y no volatilidad que el Data Warehouse. Representan una estrategia de "divide y vencerás" para ámbitos muy genéricos de un Data Warehouse.

Esta estrategia es particularmente apropiada cuando el Data Warehouse central crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él. La creación de estos Data Marts requiere algo más que una simple réplica de los datos: se necesitarán tanto la segmentación como algunos métodos adicionales de consolidación.

En efecto, un enfoque más adecuado sería la coordinación de la gestión de información de todos los Data Marts en un Data Warehouse centralizado. En esta situación los Data Marts obtendrían la información necesaria, ya previamente cargada y depurada en el Data Warehouse corporativo, simplificando el crecimiento de una base de conocimientos a nivel de toda la empresa.

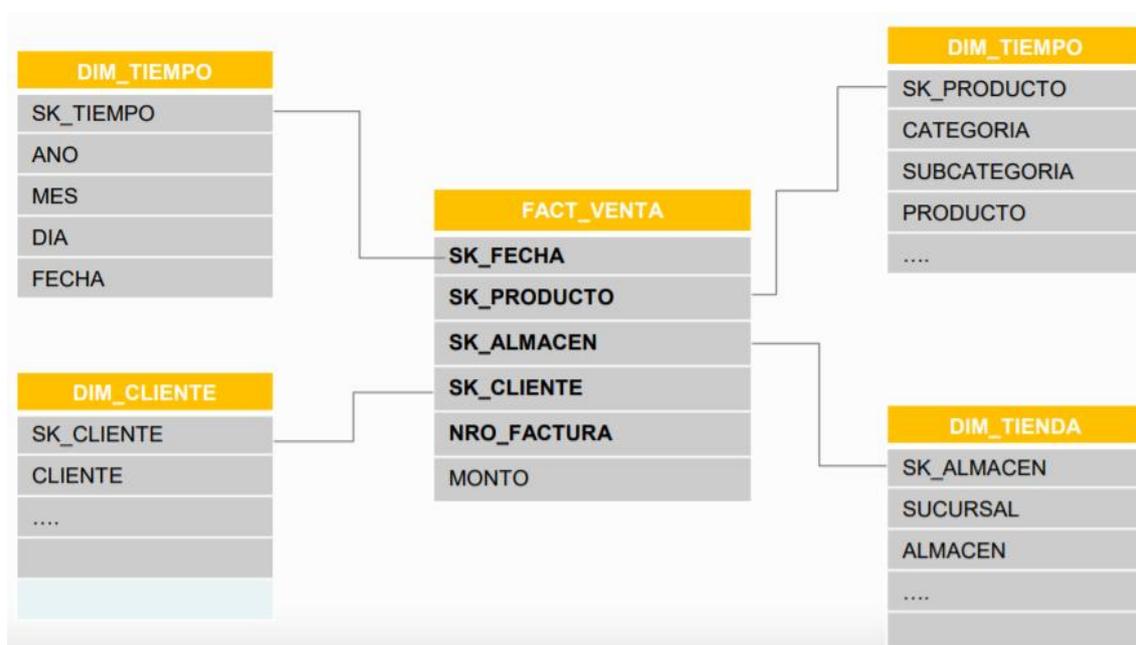
Esta simplificación provendría de la centralización de las labores de gestión de los Data Marts, en el Data Warehouse corporativo, generando economías de escala en la gestión de los Data Marts implicados.

## 2.3 Modelo Dimensional

Modelado dimensional nombra a un conjunto de técnicas y conceptos utilizados en el diseño de almacenes de datos. El modelado de dimensiones no implica necesariamente una base de datos relacional, el mismo enfoque de modelado, a nivel lógico, se puede utilizar para cualquier forma física, tal como archivos de bases de datos multidimensionales o planas.

Según el consultor de almacenamiento de datos Ralph Kimball, el modelado dimensional es una técnica de diseño de bases de datos destinadas a apoyar a las consultas de los usuarios finales en un almacén de datos. Se orienta en torno a la comprensibilidad y rendimiento.

Según él, aunque el modelo entidad relacional orientado a transacciones es muy útil para la captura de transacción, se debe evitar en la entrega al usuario final. En la ilustración 2 se puede observar cuales son los elementos principales de un modelo dimensional:



**Ilustración 2. Modelo Dimensional (Kimball & Caserta, 2008).**

### 2.3.1 Dimensión

“La Dimensión es una entidad independiente en el modelo dimensional que sirve como un punto de entrada o como un mecanismo de reordenamiento y fraccionamiento de las medidas sumariadas, localizadas en la tabla de hechos del modelo” (Kimball & Caserta, 2008).

Podemos distinguir varios tipos de dimensiones:

- **Dimensiones Normales:** aquellas que agrupan diferentes atributos que están relacionados por el ámbito al que se refieren (todas las características de un cliente, los diferentes componentes de la dimensión tiempo, etc.).
- **Dimensiones Heterogéneas:** dimensiones que agrupan conjuntos heterogéneos de atributos, que no están relacionados entre sí.
- **Dimensiones Roll-Up:** es una dimensión que es un subconjunto de otra, necesarias para el caso en que tenemos tablas de hechos con diferente granularidad.
- **Dimensiones Junk:** dimensión que agrupa indicadores de baja cardinalidad como pueden ser flags o indicadores.
- **Dimensiones Role-playing:** Cuando una misma dimensión interviene en una tabla de hechos varias veces, es necesario reutilizar la misma dimensión, pues no tiene sentido crear tantas dimensiones como usos se hagan de ella. Para ello se definen las dimensiones Role-playing. Podemos crear vistas sobre la tabla de la dimensión completa que nos permiten utilizarla varias veces o jugar con los alias de tabla. La misma dimensión juega un rol diferente según el sitio donde se utiliza.
- **Dimensiones Degeneradas:** son dimensiones que no tienen ningún atributo y por tanto, no tienen una tabla específica de dimensión. Solo se incluye para ellas un identificador en la tabla de hechos, que identifica completamente a la dimensión (por ejemplo, un pedido de ventas).

Nos interesa tener identificada la transacción (para realizar data mining, por ejemplo), pero los datos interesantes de este elemento los tenemos repartidos en las diferentes dimensiones (cliente, producto, etc.).

- **Mini dimensiones o Dimensiones Outrigger:** conjunto de atributos de una dimensión que se extraen la tabla de dimensión principal pues se suelen analizar de forma diferente. El típico ejemplo son los datos socio demográficos asociados a un cliente.

### 2.3.2 Tabla dimensión

“La tabla de dimensión es una tabla en el modelo dimensional con una clave primaria simple y columnas de atributos descriptivos” (Kimball & Caserta, 2008). En la construcción de cubos OLAP, las tablas de dimensiones son elementos que contienen atributos (o campos) que se utilizan para restringir y agrupar los datos almacenados en una tabla de hechos cuando se realizan consultas sobre dicho datos en un entorno de almacén de datos o datamart.

Estos datos sobre dimensiones son parámetros de los que dependen otros datos que serán objeto de estudio y análisis y que están contenidos en la tabla de hechos. Las tablas de dimensiones ayudan a realizar ese estudio/análisis aportando información sobre los datos de la tabla de hechos, por lo que puede decirse que en un cubo OLAP, la tabla de hechos contiene los datos de interés y las tablas de dimensiones contienen metadatos sobre dichos hechos.

### 2.3.3 Hechos

“Un hecho es una medida del desempeño empresarial, usualmente son valores numéricos y aditivos que son almacenados en una tabla de hechos” (Kimball & Caserta, 2008).

Por lo general la mayoría de la veces los hechos que se almacenan son números enteros (ejemplo: cantidad vendida de un producto, cantidad solicitada de un producto, entre otros) ó números reales (ejemplo: costo de producción de un producto, monto obtenido por venta, entre otros).

#### 2.3.4 Tabla de hechos

“En el modelo dimensional, la tabla de hechos es la tabla principal con las mediciones de rendimiento numéricas que son caracterizadas por una clave compuesta, donde cada elemento de la misma es una clave foránea que corresponde a una tabla de dimensiones” (Kimball & Caserta, 2008).

Las tablas de hechos incluyen los indicadores asociados a un proceso de negocio en concreto, además de las claves de las dimensiones que intervienen en dicho proceso, en el mínimo nivel de granularidad o detalle. Podemos tener varios tipos de tablas de hechos:

- Tabla de hechos transaccionales: Representan eventos que suceden en un determinado espacio-tiempo. Se caracterizan por permitir analizar los datos con el máximo detalle. Reflejan las transacciones relacionadas con nuestros procesos de negocio (ventas, compras, inventario, contabilidad, etc).
- Tabla de hechos sin hechos: Son tablas que no tienen medidas y representan la ocurrencia de un evento determinado. Por ejemplo, la asistencia a un curso puede ser una tabla de hechos sin métricas asociadas.
- Tabla de hechos periódicos: Son tablas de hecho usadas para recoger información de forma periódica a intervalos de tiempo regulares sobre un hecho. Nos permiten tomar una foto de la situación en un momento determinado (por ejemplo al final del día, de una semana o de un mes). Un ejemplo puede ser la foto del stock de materiales al final de cada día.
- Tabla de hecho acumulativa (Snapshots): Representan el ciclo de vida completo de una actividad o proceso, que tiene un principio y final. Suelen representar valores acumulados.

#### 2.3.5 Jerarquía

“La jerarquía es una serie de relaciones en cascada de uno a muchos” (Kimball & Caserta, 2008).

Una jerarquía define la posición relativa de un atributo con respecto a otros pertenecientes a la misma dimensión. Representados bajo una relación de tipo jerárquica o bien conocida como forma de árbol, donde los atributos serán progresivamente más detallados si se recorre de manera descendente el árbol hasta llegar a las hojas, quiénes son los que tienen mayor nivel de detalle.

### **2.3.6 Granularidad**

“La granularidad es el nivel de detalle que posee cada registro de una tabla de hechos” (Kimball & Caserta, 2008).

Se entiende como el nivel de detalle (mínimo) con el que se almacenarán los datos en la tabla de hechos de acuerdo con la actividad a modelar. La granularidad determina el significado de las tuplas de la tabla de hechos.

### **2.3.7 Agregación**

“La agregación es una medida, es decir un dato contable” (William I., 2007). Por ejemplo, si se quiere tener el total del salario de una empresa, lo que se hace es hacerle un proceso de agregación, definiendo como operación la suma de los sueldos de los empleados, logrando que cada vez que entre más información en el atributo sueldo, se va a ir mostrando sumariado.

### **2.3.8 Cubo**

Es una estructura multidimensional generada por la intersección de las dimensiones enfocadas en el hecho a medir a través de procesamiento analítico en línea (OLAP).

### **2.3.9 Esquema de estrellas**

“El esquema de estrella es la representación genérica de un modelo dimensional en una base de datos relacional en donde una tabla de hechos con una clave compuesta es unida a un número de tablas de dimensiones, cada una con una clave primaria simple” (Kimball & Caserta, 2008).

Un esquema de estrella es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

En la ilustración 3 se presenta un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones.



***Ilustración 3. Esquema de estrella con una sola tabla de hechos con enlaces a varias tablas de dimensiones. Estrella (Kimball & Caserta, 2008).***

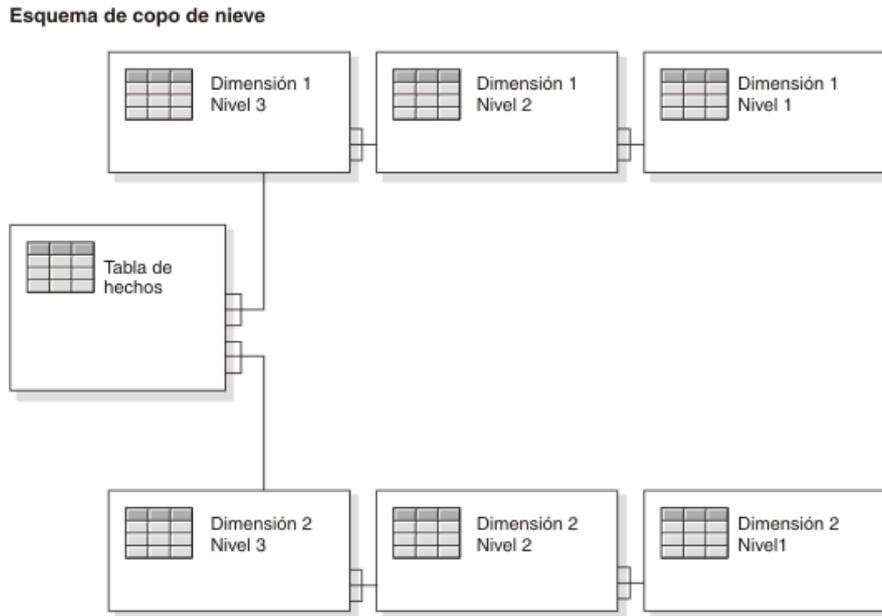
### **2.3.10 Esquema copo de nieve**

“El esquema de copo de nieve es una dimensión normalizada donde una tabla de dimensión simple es descompuesta en una estructura de árbol con potencialmente muchos niveles” (Kimball & Margy, 2002).

El esquema de copo de nieve consta de una tabla de hechos que está conectada a muchas tablas de dimensiones, que pueden estar conectadas a otras tablas de dimensiones a través de una relación de muchos a uno.

Las tablas de un esquema de copo de nieve generalmente se normalizan en el tercer formulario de normalización. Cada tabla de dimensiones representa exactamente un nivel en una jerarquía.

En la ilustración 4 se presenta un esquema de copo de nieve con dos dimensiones, cada una con tres niveles. Un esquema de copo de nieve puede tener varias dimensiones y cada dimensión puede tener varios niveles.



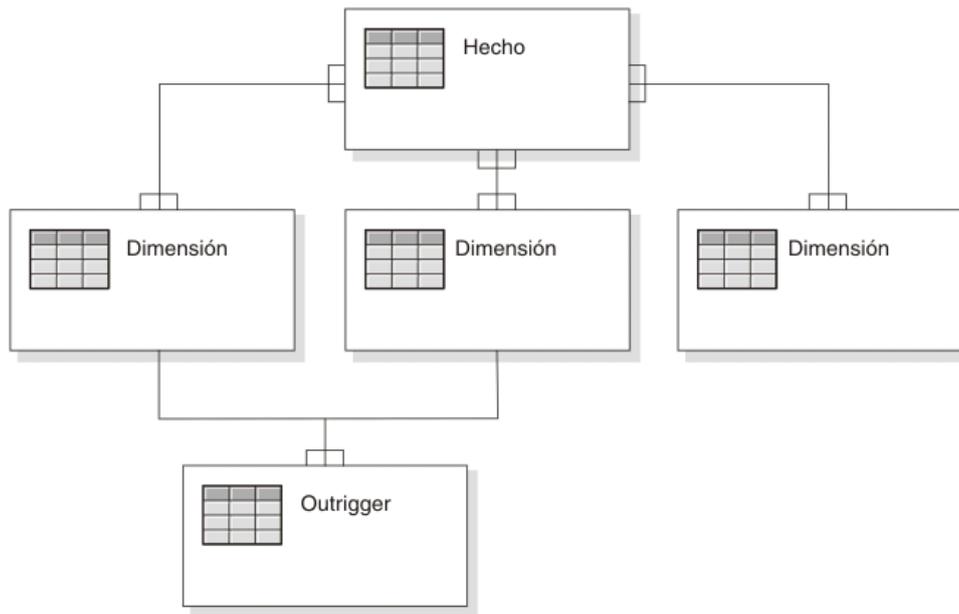
***Ilustración 4. Esquema de copo de nieve con dos dimensiones y tres niveles cada una. (Kimball & Caserta, 2008).***

### **2.3.11 Esquema constelación**

Un esquema de constelación es una combinación de un esquema de estrella y un esquema de copo de nieve. Los esquemas de constelación son esquemas de copo de nieve en los que sólo algunas de las tablas de dimensiones se han desnormalizado. El objetivo de los esquemas de constelación es aprovechar las ventajas de los esquemas de estrella y de copo de nieve. Las jerarquías de los esquemas de estrella están desnormalizadas, mientras que las jerarquías de los esquemas de copo de nieve están normalizadas.

En la ilustración 5 se presenta un esquema de constelación de ejemplo:

### Esquema de constelación



***Ilustración 5. Esquema de constelación con un hecho y dos dimensiones que comparten un outrigger. (Kimball & Caserta, 2008).***

Los esquemas de constelación están normalizados para eliminar las redundancias de las dimensiones. Para normalizar el esquema, las jerarquías dimensionales compartidas se colocan en outriggers.

### 2.3.12 Beneficios del Modelo Dimensional

**Compresibilidad** - En comparación con el modelo normalizado, el modelo dimensional es más fácil de entender y más intuitivo. En los modelos dimensionales, la información se agrupa en dimensiones coherentes, por lo que es más fácil de leer e interpretar.

La simplicidad también permite al software navegar las bases de datos de manera eficiente. En los modelos normalizados, los datos se divide en muchas entidades discretas e incluso un proceso de negocio simple podría resultar en docenas de tablas unidas entre sí de una manera compleja.

**El rendimiento de consultas** - Los modelos dimensionales están más optimizados para las consultas de datos, mientras que los modelos normalizados buscan eliminar redundancias de datos y están optimizados para la carga de transacciones y actualización.

El marco predecible de un modelo dimensional permite a la base de datos hacer fuertes supuestos sobre los datos, por lo que puede tener un impacto positivo en el rendimiento.

Cada dimensión es el equivalente a un punto de entrada en la tabla de hechos, y esta estructura simétrica permite un manejo eficaz de consultas complejas. La optimización de consulta se hace simple, predecible y controlable.

**Extensibilidad** - Los modelos dimensionales son escalables y fácilmente acomodados a nuevos datos inesperados. Las tablas existentes pueden ser cambiados, ya sea por la simple adición de nuevas filas de datos en la tabla o ejecutar en SQL algún comando de tipo "Alter Table". Las consultas o aplicaciones montadas sobre el almacén de datos no necesitan ser reprogramadas para acomodarse a los nuevos cambios.

Las consultas y aplicaciones antiguas continúan funcionando sin producir resultados diferentes. Pero en los modelos normalizados cada modificación se debe considerar cuidadosamente, debido a las complejas dependencias entre las tablas de bases de datos.

## **2.4 Inteligencia de Negocio**

### **2.4.1 Antecedentes**

La Inteligencia de Negocio (Business Intelligence o B.I.) no es un concepto nuevo, en octubre de 1958 Hans Peter Luhn (investigador de IBM), acuñó el término en el artículo "A Business Intelligence System", en el cual hace mención de la siguiente definición: "la habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada".

Pero es hasta el año 1989 que Howard Dresden (analista de Gartner), propone una definición más formal del BI, la cual es: “conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”.

Para el año 2000, la Democracia de la Información emergerá en las empresas de vanguardia, con las aplicaciones de Inteligencia de Negocio ampliamente disponibles a nivel de empleados, consultores, clientes, proveedores y el público en general. La clave para surgir en un mercado competitivo es mantenerse delante de sus competidores. Se requiere más que intuición para tomar decisiones correctas basadas en información exacta y actualizada.

Las herramientas de reporte, consulta y análisis de datos pueden ayudar a los usuarios de negocios a navegar a través de un mar de información para sintetizar la información valiosa que en él se encuentra – hoy en día esta categoría de herramientas se les llama “Inteligencia de Negocio”.

#### **2.4.2 Definición**

La Inteligencia de Negocio (Business Intelligence) es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. (Caralt, 2010).

Según Luis Méndez Del Río, la Inteligencia de Negocio (Business Intelligence) es un conjunto de herramientas y aplicaciones para la ayuda a la toma de decisiones que posibilitan acceso interactivo, análisis y manipulación de información corporativa de misión crítica.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc.

## La Inteligencia de Negocio sirve para:

- Ampliar la visión estratégica, reducir el riesgo y la incertidumbre en la toma de decisiones empresariales y construir ventajas competitivas de largo plazo en base a su información inteligente
- Tener una mejora continua de la organización, gracias a la información oportuna que genera el conocimiento que enriquece la toma de decisiones.
- Que las organizaciones sean proactivas y ágiles en la gestión de la información que utilizan.

### 2.4.3 Inteligencia de Negocio a nivel operativo, táctico y estratégico

Se le conoce como Inteligencia de Negocio o Business Intelligence, al conjunto de productos y servicios que permite que los usuarios puedan acceder y analizar de forma rápida y sencilla, a la información para su adecuada toma de decisiones a nivel operativo, táctico y estratégico. En la ilustración 6 podemos observar los niveles de Inteligencia de Negocio según Cono.



**Ilustración 6. Niveles de Inteligencia de Negocio (Cono, 2007)**

- Nivel operativo: La Inteligencia de Negocio permite a los colaboradores de éste nivel que en su rutina diaria manejan información, la reciben de una manera oportuna, exacta y adecuada.

El nivel de operativo maneja principalmente herramientas de reportes u hojas de cálculo con formatos fijos que está en constante actualización.

- Nivel táctico: Aporta información para los mandos medios en análisis y decisiones mensuales que son de utilidad para revisiones de seguimiento y toma de acciones.
- Nivel estratégico: A este nivel las decisiones son de mayor impacto en la compañía siendo utilizada la información por la alta dirección.

#### **2.4.4 Beneficios de la inteligencia de negocio**

Algunas de las ventajas que puedes tener en tu empresa al utilizar la Inteligencia de Negocio son las siguientes:

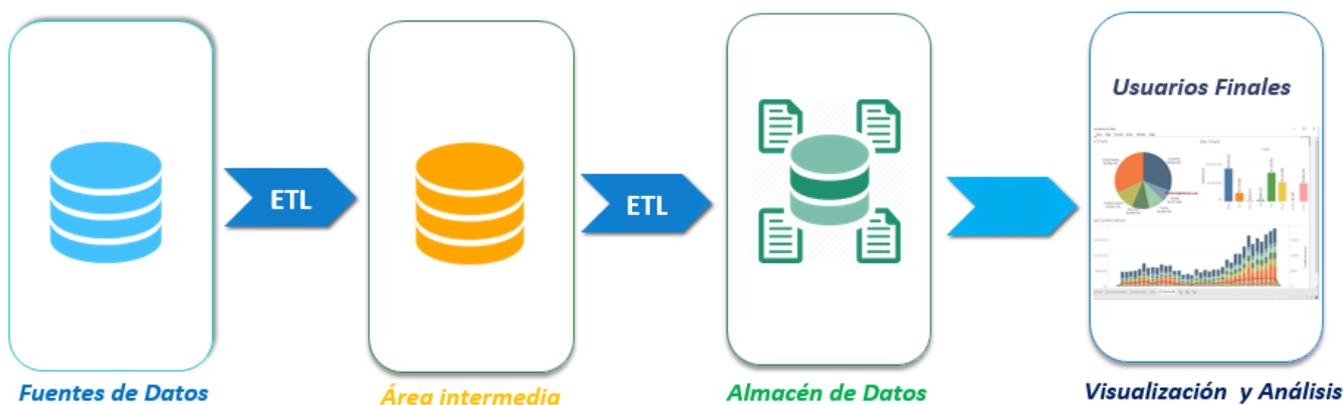
- 1. Incremento de la eficiencia:** Al contar con los datos de manera accesible y ágil puedes generar información de valor centralizada la cual podrás visualizar en una única plataforma para aprovecharla de manera óptima para realizar análisis y tomar decisiones informadas y en tiempo.
- 2. Respuestas rápidas a situaciones de negocio:** Para poder tomar decisiones en el momento indicado es importante contar con la información a la mano de manera sencilla y no perder tiempo en buscar y consolidar datos. Gracias al BI puedes tener las respuestas en minutos de manera clara y concisa por medio de reportes de indicadores y tableros de datos.
- 3. Control de las áreas funcionales de la empresa:** En todas las áreas de tu empresa se genera información de valor día a día, puedes aprovecharla de la mejor manera para conocer tendencias, proyectar datos, analizar escenarios, etc.
- 4. Mejora tu servicio al cliente:** Al contar con la información más importante y en tiempo real puedes ofrecer a tus clientes un servicio de mayor calidad desde el pedido hasta el servicio post venta al conocer más acerca de ellos y sus necesidades. Analiza hábitos de compra, reconoce los productos más vendidos, etc.
- 5. Presenta información por medio de tableros de indicadores:** para una comunicación más simple y directa de la situación de la empresa. Al tener la posibilidad de crear distintos tableros para control puedes enfocarte en los datos más relevantes que mostrar sin necesidad de revisar grandes cantidades de información.

### 2.4.5 Arquitectura de una solución de Inteligencia de Negocio

Una solución de Business Intelligence parte de los sistemas de origen de una organización, sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico. Para ello se realiza una fase de extracción, transformación y carga (ETL) de datos. Esta etapa suele apoyarse en un almacén intermedio, llamado ODS, que actúa como pasarela entre los sistemas fuente y los sistemas destino (generalmente un datawarehouse), y cuyo principal objetivo consiste en evitar la saturación de los servidores funcionales de la organización.

La información resultante, ya unificada, depurada y consolidada, se almacena en un datawarehouse corporativo, que puede servir como base para la construcción de distintos datamarts departamentales. Estos datamarts se caracterizan por poseer la estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante bases de datos transaccionales (OLTP) o mediante bases de datos analíticas (OLAP).

Los datos albergados en el datawarehouse o en cada datamart se explotan utilizando herramientas comerciales de análisis.



**Ilustración 7. Arquitectura de una solución de Inteligencia de Negocio (Cano, 2007)**

A continuación se definen los elementos que integran una arquitectura de una solución de Inteligencia.

### **2.4.5.1 Fuente de datos**

La primera capa considera todas las fuentes de datos relevantes para el BI a implementar. Las fuentes de datos pueden tener variado origen:

- Datos operacionales procedentes de bases de datos corporativas
- Generalmente procedentes de ambientes transaccionales OLTP.
- Datos generados por sistemas de información
- Datos privados
- Datos externos (Internet, BD Comerciales, de clientes, información económica, etc.)

### **2.4.5.2 Capa de integración de los datos**

Esta capa es la encargada de la integración de los datos desde las múltiples fuentes existentes para la construcción del Datawarehouse y los Datamarts de la base analítica.

Se utilizan rutinas para Extracción, Transformación y Carga (ETL).

- Las rutinas ETL tienen la función de manipular los datos de las distintas fuentes de datos a fin de integrarlos todos al Datawarehouse según formatos y estructuras que requieren.
- Integración de fuentes heterogéneas.
- Limpia los datos removiendo problemas transaccionales y asegurando consistencia en los datos.

#### **Proceso Extracción, Transformación y Carga (ETL's)**

Los procesos de ETL consideran:

- 1. Extracción:** Se considera la consolidación de los datos desde las diversas fuentes que usan diferentes estructuras y formatos. Establece el acceso a las fuentes de datos.

Extrae eficientemente cada una de las fuentes de datos, con el mejor procesamiento posible, buscando solo los datos requeridos. Se pueden almacenar datos temporalmente en un área especializada para su transformación e integración con otras fuentes.

**2. Transformación:** La fase de transformación aplica una serie de reglas o funciones a los datos extraídos desde la fuente para generar los datos a cargar al destino final (el Data Warehouse o Data Marts).

Algunos datos requerirán muy poca o ninguna manipulación. En otros casos, se requerirá una serie de pasos de transformación para satisfacer las necesidades técnicas y del negocio de la base analítica.

**3. Carga:** Carga de los datos a la base analítica (DataMarts / Datawarehouse/Staging Area (ODS))

#### **2.4.5.3 Capa analítica**

Es un área intermedia de almacenamiento de datos utilizada para el procesamiento de los mismos durante procesos de extracción, transformación y carga (ETL). Esta área se encuentra entre la fuente de los datos y su destino, que a menudo son almacenes de datos, datamarts u otros repositorios de datos.

- Generalmente de naturaleza transitoria.
- Contenido se borrará antes de ejecutar un proceso de ETL o inmediatamente después de haberlo finalizado con éxito.
- Existen arquitecturas con áreas intermedias de datos permanentes con la finalidad de mantener un archivo de los mismos o para poder resolver problemas detectados a posteriori.

#### **2.4.5.4 Capa de visualización y explotación de datos**

Esta capa es la encargada de desplegar los análisis sobre los cubos mediante sus diversos cortes de información, cruces y filtros. También despliega los reportes estáticos, ad hoc, tableros de control (dashboards), etc.

Esta capa tiene diversos medios por los que desplegar información: Pantallas, Portales web, Archivos (Excel, pdf, powerpoint), correo electrónico. Consultas e informes simples: Corresponden a los reportes operacionales y listados que generalmente son estáticos y muestran datos detallados y actualizados. Pueden incluir gráficos y cuadros. Pueden ser generados de manera automática o pueden ser elaborados cada vez que se requieran Apoyan en los procesos de negocio y excepciones.

## **2.5 Importancia de la Inteligencia de Negocio**

- Tenemos datos pero carecemos de información – Es importante almacenar los datos de clientes, empleados, departamentos, compras, ventas, entre otros en aplicaciones, sistemas financieros o fuentes de datos. Si queremos que nuestra empresa tenga mayor ventaja sobre la competencia esta gestión no es suficiente.
- Fragmentación – Poseen aplicaciones independientes a través de todos los departamentos pero se carece de una visión global de la empresa. Tal vez por la incapacidad de las herramientas de BI de integrar fuentes de datos heterogéneas.

Esto limita a la empresa a tomar decisiones importantes sin tener todos los elementos imprescindibles a la mano. Esta fragmentación conduce a lo que se llama diferentes versiones de la verdad. Los gerenciales solicitan informes a los distintos departamentos obteniendo diferentes resultados del mismo informe.

- Manipulación manual – La necesidad de generar análisis de negocios e informes nos ha llevado a utilizar herramientas de BI y/o de reportes que no son las más confiables. Esta práctica conlleva la exportación de datos a distintas herramientas que resultan en un proceso lento, costoso, duplicación de trabajo, poca confiabilidad en los informes, propenso a errores y sujetos a la interpretación individual.
- Poca agilidad – Debido a la carencia de información, la fragmentación y la manipulación manual me mantiene en un nivel de rendimiento bajo.

## 2.6 Indicadores de gestión

“Los indicadores de gestión son elementos del sistema de control de gestión que proporcionan información significativa sobre aspectos críticos o claves de una organización mediante la relación de dos o más datos” (Pacheco, 2002).

Otras definiciones:

- Medios, instrumentos o mecanismos para evaluar hasta qué punto o en qué medida se están logrando los objetivos estratégicos. Representan una unidad de medida gerencial que permite evaluar el desempeño de una organización frente a sus metas, objetivos y responsabilidades con los grupos de referencia.
- Producen información para analizar el desempeño de cualquier área de la organización y verificar el cumplimiento de los objetivos en términos de resultados.
- Detectan y prevén desviaciones en el logro de los objetivos.
- El análisis de los indicadores conlleva a generar Alertas Sobre La Acción, no perder la dirección, bajo el supuesto de que la organización está perfectamente alineada con el plan.

### 2.6.1 Categorías de indicadores de gestión

Se debe saber discernir entre indicadores de cumplimiento, de evaluación, de eficiencia, de eficacia e indicadores de gestión. Como un ejemplo vale más que mil palabras este se realizará teniendo en cuenta los indicadores que se pueden encontrar en la gestión de un pedido.

1. *Indicadores de resultado*: Son aquellos que miden la consecuencia del objetivo estratégico. También se les llama indicadores de efecto.
2. *Indicadores de desempeño*: Los indicadores de desempeño son instrumentos de medición de las principales variables asociadas al cumplimiento de los objetivos, que a su vez constituyen.

3. *Indicadores de eficiencia:* Teniendo en cuenta que eficiencia tiene que ver con la actitud y la capacidad para llevar a cabo un trabajo o una tarea con el mínimo de recursos. Los indicadores de eficiencia están relacionados con las razones que indican los recursos invertidos en la consecución de tareas y/o trabajos. Ejemplo: Tiempo fabricación de un producto, razón de piezas / hora, rotación de inventarios.
4. *Indicadores de eficacia:* Los indicadores de eficacia están relacionados con las razones que indican capacidad o acierto en la consecución de tareas y/o trabajos. Ejemplo: grado de satisfacción de los clientes con relación a los pedidos.
5. *Indicadores de calidad:* Son medidas estadísticas basadas en cifras o ratios que se utilizan como criterio para juzgar y evaluar el desempeño de una organización, un sistema o un proceso

#### **2.6.2 Propósito y beneficios de los indicadores de gestión**

Podría decirse que el objetivo de los sistemas de medición es aportar a la empresa un camino correcto para que ésta logre cumplir con las metas establecidas.

Todo sistema de medición debe satisfacer los siguientes objetivos:

- Comunicar la estrategia.
- Comunicar las metas.
- Identificar problemas y oportunidades.
- Diagnosticar problemas.
- Entender procesos.
- Definir responsabilidades.
- Mejorar el control de la empresa.
- Identificar iniciativas y acciones necesarias.
- Medir comportamientos.
- Facilitar la delegación en las personas.
- Integrar la compensación con la actuación.

### **2.6.3 ¿Por qué medir y para qué?**

Si no se mide lo que se hace, no se puede controlar y si no se puede controlar, no se puede dirigir y si no se puede dirigir no se puede mejorar. A partir de las últimas décadas del siglo XX, las empresas están experimentando un proceso de cambios revolucionarios, pasando de una situación de protección regulada a entornos abiertos altamente competitivos.

Esta situación, de transformaciones constantes del ambiente de negocio hace necesario que las empresas, para mantener e incrementar su participación de mercado en estas condiciones, deben tener claro la forma de cómo analizar y evaluar los procesos de su negocio, es decir deben tener claro su sistema de medición de desempeño.

La medición del desempeño puede ser definida generalmente, como una serie de acciones orientadas a medir, evaluar, ajustar y regular las actividades de una empresa.

## **Entonces, ¿Por qué medir?**

- Por qué la empresa debe tomar decisiones.
- Por qué se necesita conocer la eficiencia de las empresas (caso contrario, se marcha “a ciegas”, tomando decisiones sobre suposiciones o intuiciones).
- Por qué se requiere saber si se está en el camino correcto o no en cada área.
- Por qué se necesita mejorar en cada área de la empresa, principalmente en aquellos puntos donde se está más débil. Por qué se requiere saber, en lo posible, en tiempo real, que pasa en la empresa (eficiencia o ineficiencia).

## **2.7 Herramientas de Inteligencia de Negocio**

Las herramientas de software de Inteligencia de Negocio son usadas para acceder a los datos de los negocios y proporcionar reportes, análisis, visualizaciones y alertas a los usuarios. La gran mayoría de las herramientas de Inteligencia de Negocio son usadas por usuarios finales para acceder, analizar y reportar contra los datos que más frecuentemente residen en datawarehouse, datamarts y almacenes de datos operacionales

## **2.8 Tipos de productos BI**

Actualmente el mercado de herramientas de Inteligencia de Negocio se encuentra constituido por dos subsegmentos: suites de Inteligencia de Negocio empresarial (EBIS, por sus siglas en inglés) y plataformas de BI.

- Las plataformas de Inteligencia de Negocio se usan cuando hay una necesidad de analizar aplicaciones complejas con muchos cálculos (por ejemplo, rentabilidad de un producto) o para crear aplicaciones amigables para usuarios ocasionales.
- Las plataformas de Inteligencia de Negocio se usan cuando hay una necesidad de analizar aplicaciones complejas con muchos cálculos (por ejemplo, rentabilidad de un producto) o para crear aplicaciones amigables para usuarios ocasionales.

- Minería de datos: Extracción de información de las bases de datos acerca del consumidor, mediante la utilización de aplicaciones que pueden aislar e identificar patrones o tendencias del consumidor en un alto volumen de datos.
- Digital Dashboards o paneles de Control Digital También conocidos como Business Intelligence Dashboards, o Dashboards Ejecutivos.: Son resúmenes visuales de información del negocio, que muestran de una mirada la comprensión del global de las condiciones del negocio mediante métricas e Indicadores Clave de Desempeño (KPIs). Esta es una Herramienta de Inteligencia de Negocio muy popular desde hace unos pocos años.
- OLAP(incluido HOLAP, ROLAP and MOLAP): Es la capacidad de algunos sistemas de soporte de decisiones gerenciales que permiten examinar de manera interactiva grandes volúmenes de información desde varias perspectivas.

**Tabla 1. Herramientas de tecnologías en Inteligencia de Negocio y su descripción.**

<b>Herramienta</b>	<b>Descripción</b>
<b><u>Pentaho</u></b>	<p>Pentaho ofrece una versión gratuita de su suite que incluye análisis de negocios, así como la integración de datos, un diseñador de informes y un mercado de plugins, desarrollados por la comunidad para ampliar las capacidades de la plataforma Pentaho.</p> <p>La empresa promueve esta versión de prueba y evaluación antes de actualizar el producto comercial al completo, que incluye el análisis visual, el modelado predictivo, cuadros de mando y un elemento móvil</p>
<b><u>Power BI</u></b>	<p>Herramienta con servicio en la nube que permite subir, compartir y tener acceso a informes de excel realizados en una determinada compañía.</p> <p>Además, puedes acceder desde cualquier dispositivo, ya sea un ordenador, una tablet o un smartphone.</p> <p>La integración de Power BI también permite el acceso a orígenes de datos locales, orígenes de bases de datos y servicios en la nube.</p> <p>Asimismo, podrás interactuar con gráficos y filtros, generar alertas de determinados datos, consultar datos de informes de Power View y compartir información actualizada con toda tu plantilla en cualquier momento.</p>

<b>Herramienta</b>	<b>Descripción</b>
<u>Qlik</u>	<p>Es una plataforma enfocada al análisis visual de datos y aplicaciones interactivas que tiene por objetivo mejorar el proceso de acceso a los datos de cara al usuario. Como, por ejemplo, acceder a ciertas visualizaciones ‘limpias’ y fáciles de comprender, diseños de gráficos llamativos, entre otros.</p> <p>Siguiendo el informe de Gartner, esta herramienta cuenta, en porcentaje, con el mayor número de usuarios que la utilizan. Además, Qlik funciona con el motor Qlik Indexing Engine (QIX), “motor de indexación asociativa de datos más potente del mundo”, según explican en la web principal de la herramienta.</p>
<u>Tableau</u>	<p>Esta otra herramienta BI sirve para la visualización interactiva de los datos, con los que los usuarios pueden interactuar de varias maneras: comparando datos, filtrándolos o creando una conexión entre unas variables y otras.</p> <p>Con esta herramienta uno puede tener acceso a grandes bases de datos en común como MySQL, Greenplum, Oracle o Microsoft. Es la herramienta más visual en cuanto a la representación de datos, gráficos y tablas de cara al usuario.</p>

<b>Herramienta</b>	<b>Descripción</b>
<u>Spago BI</u>	<p>SpagoBI se describe a sí mismo como la única plataforma abierta 100% y completa de Business Intelligence. Su arquitectura modular, basada en estándares abiertos, facilita la personalización y la integración en entornos.</p> <p>SpagoBI tiene un modelo de conducta que regula la visibilidad de los datos y el comportamiento de los documentos simples con relación a los roles de los usuarios finales.</p> <p>SpagoBI es una plataforma de integración (y no una plataforma de producto), ya que no se construye y se cierra en torno a un conjunto predefinido de herramientas.</p> <p>SpagoBI es enteramente de software libre (FOSS), solamente posee una versión la cual es gratuita (a diferencias de otras similares como Pentaho que tienen una versión comunitaria (gratuita) y otra empresarial (de pago)); lleva el BI Open Source al nivel empresarial, de hecho, el modelo de comportamiento regula la visibilidad en los datos de acuerdo a las responsabilidades empresariales y cuenta con una arquitectura escalable que continúa en la lógica del software libre.</p>

<b>Herramienta</b>	<b>Descripción</b>
<u>Microstrategy</u>	<p>MicroStrategy provee software de reporte, análisis y monitoreo integrados que permite a las empresas analizar datos almacenados de la empresa y de todas las áreas para tomar mejores decisiones. Las empresas escogen a MicroStrategy como herramienta de business intelligence por su fácil uso, análisis sofisticado y escalabilidad de datos y usuarios a grandes volúmenes.</p> <p>Una herramienta de escritorio rápida y fácil de usar que le permite acceder de forma intuitiva y explorar sus datos para crear visualizaciones en minutos en su PC o MAC.</p> <p>Uno de sus beneficios es que tiene una licencia de usuario. Establezca, diseñe y modifique sus pantallas de estadísticas. Descubra sus datos de una manera visual intuitiva y amigable.</p> <p>Permite transformar grandes cantidades de datos en paneles de control intuitivos, que facilitan una mayor comprensión analítica para realizar informes y tomar decisiones de negocio basadas en los datos. Con la herramienta de business intelligence, mejore la productividad de su empresa, de seguimiento de las tendencias, descubra anomalías y detecte oportunidades de negocio.</p>

<b>Herramienta</b>	<b>Descripción</b>
<u>LiteBI</u>	<p><b>Liteb</b>ies un software completo de Business Intelligence ofrecido en modalidad SaaS, para poder contrarrestar el elevado coste y complejidad del software BI tradicionales.</p> <p>La arquitectura SaaS es cada vez más popular, y tiene muchas fortalezas que lo convierten en una opción interesante. En otros ámbitos, el SaaS ya es una realidad, sin embargo en entornos Business Intelligence se trata de un enfoque novedoso y Litebi es uno de los primeros proveedores en ofrecer esta posibilidad. La principal dificultad del SaaS Business Intelligence reside en la cantidad de información que debe transmitirse diariamente por la red, por este motivo, puede resultar inasumible en organizaciones con grandes cantidades de información.</p> <p>Además, se diferencian de otros proveedores de BI en que su producto es menos complejo, y está dirigido especialmente a empresas de tamaño "mediano", por lo que las implantaciones pueden realizarse en un menor tiempo y a un menor coste.</p>

<b>Herramienta</b>	<b>Descripción</b>
<b>Oracle Business Intelligence</b>	<p>Oracle proporciona a la industria la <b>Suite Oracle Business Intelligence Foundation</b>, plataforma única de inteligencia de negocio que permite a los clientes descubrir nuevas ideas y tomar decisiones de negocio más informadas y de una forma más rápida, ofreciendo análisis visuales ágiles junto con aplicaciones analíticas.</p> <p>Oracle BI es una solución integral que reduce el costo total de propiedad y aumenta la rentabilidad de la inversión para toda la organización.</p> <p>Entregar los mejores productos de su clase para consultas y análisis, OLAP, informes y cuadros de mando. Así como proporcionar a los clientes una visión superior en cuanto alineamiento empresarial, visibilidad y un mejor más reducido para generar valor.</p>
<b>Cognos</b>	<p><b>IBM Cognos BI</b> es una de las suites de Business Intelligence más utilizadas. Es un software de BI bastante completo, y a la vez manejable, y uno de los líderes del mercado de BI. Las aplicaciones principales se utilizan desde un portal web que controla el servidor de Business Intelligence, que es el corazón de la herramienta.</p>

## 2.9 Conceptos orientados a los procesos del portal web de Tangerix

### 2.9.1 Compra

La compra hace referencia a la acción de obtener o adquirir, a cambio de un precio determinado, un producto o un servicio. Pero también se considera “compra” el objeto adquirido, una vez consumado el acto de adquisición.

En este caso del portal web se le denomina “Compra inmediata”, se le denomina a la acción de adquirir un producto publicado que sea de tipo BON (Buy Option Now) en el portal web pero de forma inmediata, se le da este nombre debido a que en el momento en que el usuario seleccione esta opción adquiere el producto, de esta manera el producto sale del aire en el portal.



HALCO ENERGY EQUIPMENT COMPANY

### Gear Box 1750 H.P.

Item Cod. # 092

QUANTITY AVAILABLE	YEAR	USAGE
2 Units	2012	8500 Hours

So. LA  
New

USD 456000.00

ADVERTISING END  
04-27-2018  
Millquip, LLC

[Add to Budget](#) [Buy now](#)

DOWNLOAD: [Datasheet](#)

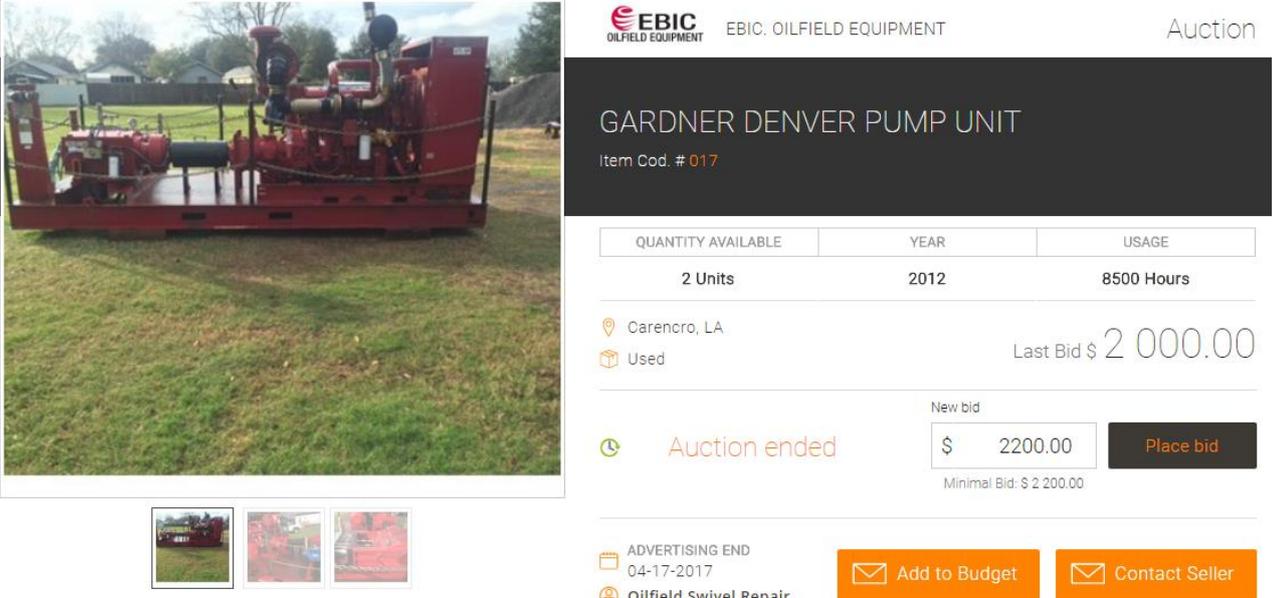
**Ilustración 8. Producto publicado por BON (Buy option now)**

### 2.9.2 Subastas electrónicas

La subasta electrónica / subasta inversa es un proceso dinámico de negociación de precios online entre proveedores preseleccionados para conseguir una parte del negocio. Es una manifestación del B2B (Business to Business).

Se trata de un nuevo sistema de negociación en el que, en cierto modo, desaparece la relación personal con el cliente para convertirla en una negociación a través de internet.

En este caso, las subastas ocurren en un período de tiempo en el cual pueden participar cualquier usuario registrado en el portal web, a su vez cada usuario puede pujar una cantidad de dinero por el producto publicado. Una vez que la subasta finalice se le proporciona la información del vendedor del producto al usuario ganador de la subasta y el producto sale del aire en el portal web.



The screenshot displays an auction listing for a Gardner Denver Pump Unit. The main image shows a large red industrial pump unit mounted on a trailer, situated outdoors on a grassy area. The page header includes the EBIC Oilfield Equipment logo and the word 'Auction'. The item title is 'GARDNER DENVER PUMP UNIT' with an item code of #017.

QUANTITY AVAILABLE	YEAR	USAGE
2 Units	2012	8500 Hours

Location: Carencro, LA  
Condition: Used  
Last Bid: \$ 2 000.00

Auction ended

New bid: \$ 2200.00 (Place bid)  
Minimal Bid: \$ 2 200.00

ADVERTISING END: 04-17-2017  
Oilfield Swivel Repair

Buttons: Add to Budget, Contact Seller

**Ilustración 9. Producto publicado por subasta**

### 2.9.3 Intermediación

Es el proceso de mediación de una persona o institución en un conflicto en específico, en este caso sería la mediación entre los usuarios compradores y vendedores para la adquisición de productos en un portal web. El objetivo de la intermediación del portal es que si un usuario quiere adquirir un producto publicado de tipo "Contact seller" es decir, contacto del vendedor, con tan solo el usuario darle click a la opción de "Contact Seller" se le proporciona la información del vendedor del producto para proceder a finalizar la compra.



### Spools & Flowback 4-1/16 & 3-1/16 10k & 15k

Item Cod. # 0508

QUANTITY AVAILABLE	YEAR	USAGE
2 Units	2012	8500 Hours

Crowley, La.

Used

USD 1000.00

ADVERTISING END  
10-03-2018

Millquip, LLC

Add to Budget

Contact Seller

DOWNLOAD:

Datasheet



**Ilustración 10. Producto publicado por Contacto Vendedor.**

#### 2.9.4 Surplus

Se le denomina “Surplus” en el ámbito petrolero a aquellos productos sobrantes de un proyecto de petróleo y gas que nunca se usaron para posteriormente vender todos estos productos a otra empresa petrolera que los necesite para otros fines.

En este caso, existe una sección en el portal web donde si existen usuarios que tengan Surplus para vender puedan ser publicados en esta sección para que otras empresas u organizaciones se contacten con ellos y poder terminar la compra de los mismos.

field Equipment

Equipments Machinery Parts & Supplies Materials Refineries Petrochemicals Surplus

Add Surplus (+)



USER/COMPANY	DOWNLOAD LIST
Millquip, LLC	<a href="#">↓</a>
Millquip, LLC	<a href="#">↓</a>
Millquip, LLC	<a href="#">↓</a>
Ritchie Bros	<a href="#">↓</a>
Equipment & Supply, LLC	<a href="#">↓</a>

**Ilustración 11. Sección de Surplus del portal web.**

## CAPÍTULO 3

### MARCO METODOLÓGICO

En el desarrollo de una solución de Inteligencia de Negocio es de gran importancia tomar en cuenta ciertos aspectos, empezando por un estudio exhaustivo de los procesos de la organización, hasta la selección de las herramientas tecnológicas que se adapten a las necesidades de la misma, teniendo mejores resultados confiables y satisfactorios.

Es un requerimiento fundamental estudiar una o más metodologías que se ajusten adecuadamente al análisis, diseño y construcción de la solución de Inteligencia de Negocio. Es importante tener una relación con lo mencionado anteriormente y la elaboración de un almacén de datos, ya que es un factor importante, debido a todo abarca desde la obtención de datos desde las fuentes, hasta la elaboración de un modelo que integre los datos para poder procesarlos posteriormente.

En el desarrollo de una estructura de datos que ayuda a la toma de decisiones (almacenes de datos), se destacan varias metodologías de desarrollo entre ellas Ralph Kimball y Bill Inmon, sin embargo se seleccionó como base la metodología propuesta por Ralph Kimball, ya que conduce a una solución que se puede desarrollar de manera más rápida y se adapta a los requerimientos propuestos por la organización.

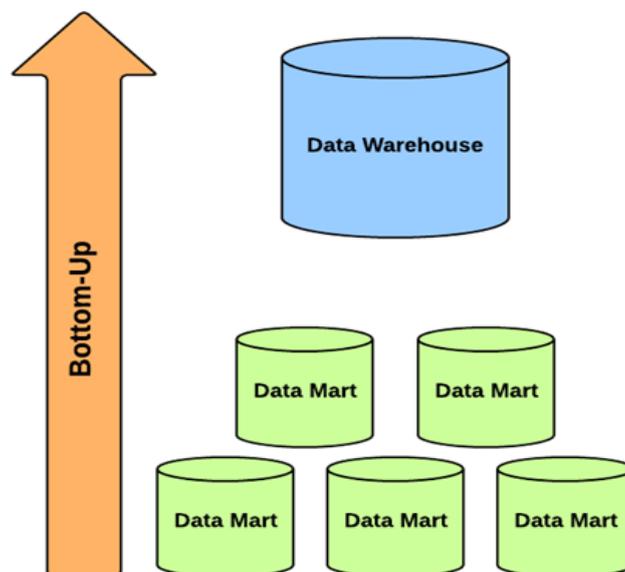
#### **3.1 Metodología según Ralph Kimball (Ascendente).**

La metodología de Ralph Kimball define un Almacén de Datos como “una copia de los datos transaccionales específicamente estructurados para consultas y análisis” (Kimball & Caserta, The Data Warehouse ETL Toolkit (2nd edition), 2008), también da una segunda definición como la unión de todas las Bodegas de Datos de una entidad.

La metodología de Kimball es ampliamente utilizada en las corporaciones por su capacidad de ofrecer resultados medibles en un corto periodo de tiempo (en comparación a la metodología de Inmon).

Ralph Kimball es un autor considerado como el "gurú" en el tema de almacenes de datos. Se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que este concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de proyectos de inteligencia de negocio. Por esta razón, esta metodología se ha convertido en el estándar de facto en el área de apoyo a las decisiones empresariales.

Kimball, plantea un enfoque ascendente (bottom-up), en donde se definen previamente las bodegas de datos para luego integrarlas cuando se diseñe un almacén de datos centralizado (Kimball R.1998), como se observa en la ilustración 12. Esto permite que la metodología sea flexible y sencilla de implementar, se construye como una bodega de datos como primer elemento del sistema de análisis, y luego ir añadiendo otras bodegas que compartan las dimensiones ya definidas o incluyan otras nuevas, hasta integrarlas y formar un almacén de datos corporativo.



**Ilustración 12. Metodología Bottom-Up**

La metodología de Kimball está descrita en su libro publicado en 1998 "The Data Warehouse Lifecycle toolkit", en este libro Kimball establece una serie de pasos que deben seguirse para la realización de un almacén de datos. Este conjunto de pasos no es lineal, puesto que en el ciclo de desarrollo ciertas tareas pueden ser desarrolladas de forma paralela.

Con esta metodología se establecen los estándares para el diseño, construcción, e implementación de una solución de Inteligencia de Negocio, también denominada como “Ciclo de Vida Dimensional del Negocio y está basada en cuatro principios que son:

Con esta metodología Kimball establece los estándares necesarios para el diseño, construcción e implementación de una solución de inteligencia de Negocio, la cual también es conocida como el Ciclo de Vida Dimensional del Negocio y está basada en cuatro principios básicos que son:

1. **Centrarse en el negocio:** Se basa en la identificación de los requerimientos del negocio.
2. **Construcción de una infraestructura de información adecuada:** Diseñar una base que contenga información única, integrada, fácil de usar, de alto rendimiento donde se reflejan requerimientos de negocio identificados en la organización.
3. **Realizar entregas en incrementos significativos:** Crear el almacén de datos de forma incremental a través de entregables en plazos de 6 a 12 meses aproximadamente.
4. **Ofrecer la solución completa:** Significa la entrega de los elementos necesarios para entregar a los usuarios, empezando por un almacén de datos bien implementado y, herramientas de consulta, aplicaciones para informes y análisis avanzado, capacitación, soporte y documentación.

La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente complejo, y Kimball nos propone una metodología que nos ayuda a simplificar esta complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la ilustración 13.

Cabe destacar que no todas las fases de las tareas del ciclo de vida son ejecutadas en todos los proyectos, todo va a depender de la planificación y el alcance que tengan estos.



**Ilustración 13. Adaptación del ciclo de vida de Kimball (Kimball & Ross, 2013)**

El conjunto de pasos que plantea Kimball son descritos a continuación:

### 3.1.1 Planificación del proyecto

En este proceso se determina el propósito del proyecto de Inteligencia de Negocio, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información. En la visión de programas y proyectos de Kimball, Proyecto, se refiere a una iteración simple del KLC (Kimball Life Cycle).

Esta tarea incluye las siguientes acciones típicas de un plan de proyecto: Definir el alcance (entender los requerimientos del negocio).

- Identificar las tareas
- Programar las tareas
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos

- Elaboración de un documento final que representa un plan del proyecto. Además en esta parte definimos cómo realizar la administración o gestión de esta subfase que es todo un proyecto en sí mismo, con las siguientes actividades:
  - Monitoreo del estado de los procesos y actividades.
  - Rastreo de problemas
  - Desarrollo de un plan de comunicación comprensiva que dirija la empresa y las áreas de TI

### **3.1.2 Definición de requerimientos del Negocio**

La definición de los requerimientos es en gran medida un proceso de entrevistar al personal de negocio y técnico. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo.

Hay que leer todos los informes posibles de la organización; rastrear los documentos de estrategia interna; entrevistar a los empleados, analizar lo que se dice en la prensa acerca de la organización, la competencia y la industria. Se deben conocer los términos y la terminología del negocio.

Parte del proceso de preparación es averiguar a quién se debe realmente entrevistar. Esto normalmente implica examinar cuidadosamente el organigrama de la organización. Hay básicamente cuatro grupos de personas con las que hablar desde el principio: el directivo responsable de tomar las decisiones estratégicas; los administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones; personal de sistemas, si existen, la gente que realmente sabe qué tipos de problemas informáticos y de datos existen; y por último, la gente que se necesita entrevistar por razones políticas.

### **3.1.3 Modelo Dimensional**

El levantamiento de los requerimientos define el modelo dimensional requerido para resolver las necesidades de los usuarios. Las fuentes de datos y los requerimientos van a definir la granularidad del modelo, los indicadores, dimensiones, jerarquías y atributos.

Básicamente se comienza con una matriz donde se determina la dimensión de cada indicador y luego se especifican los diferentes grados de detalle (atributos), dentro de cada concepto del negocio (dimensión), así como la granularidad de cada indicador (variable o métrica) y las jerarquías que dan forma al modelo dimensional del negocio (BDM) o mapa dimensional.

El proceso consiste en cuatro pasos:

1. Elegir el proceso de negocio.
2. Establecer el nivel de granularidad.
3. Elegir las dimensiones. 4. Identificar medidas y las tablas de hechos.
4. Identificar medidas y las tablas de hechos.

#### **3.1.3.1 Elegir el proceso de negocio**

El primer paso es elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.

#### **3.1.3.2 Establecer el nivel de granularidad**

La granularidad significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. La sugerencia general es comenzar a diseñar el datawarehouse al mayor nivel de detalle posible, ya que se podría luego realizar agrupamientos al nivel deseado. En caso contrario no sería posible abrir (drill-down) las sumalizaciones en caso de que el nivel de detalle no lo permita.

### **3.1.3.3 Elegir las dimensiones**

Las dimensiones surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos.

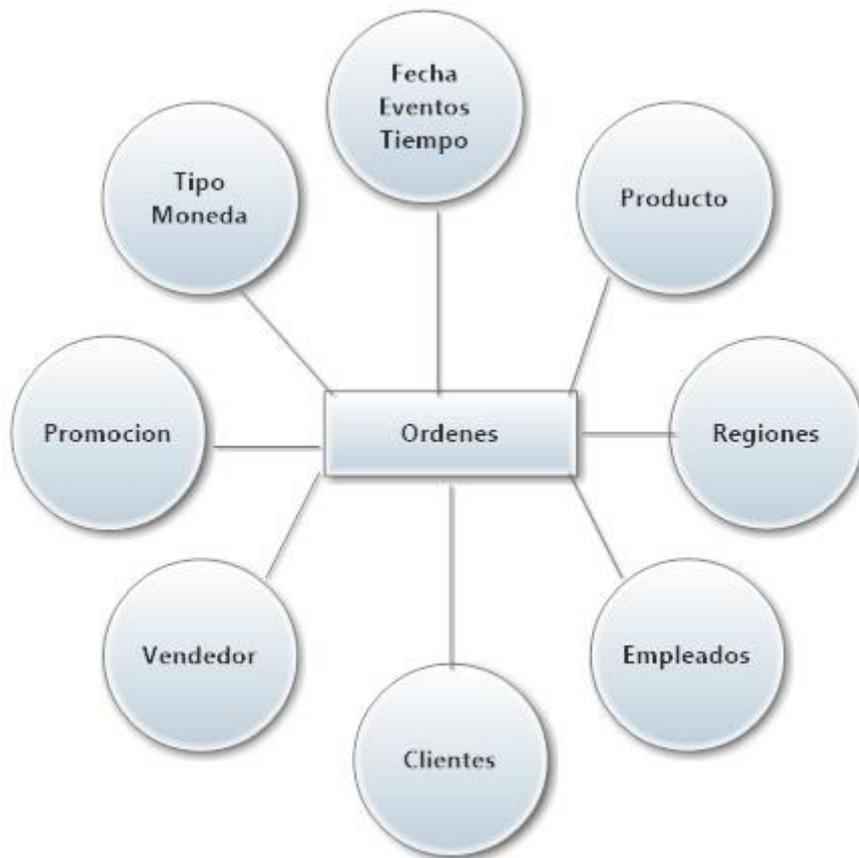
### **3.1.3.4 Identificar las tablas de hechos y medidas**

El último paso consiste en identificar las medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla que se desea analizar, sumando o agrupando sus datos, usando los criterios de corte conocidos como dimensiones. Las medidas habitualmente se vinculan con el nivel de granularidad del punto 2.3.2., y se encuentran en tablas que denominamos tablas de hechos.

Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos. Un registro contiene una medida expresada en números, como ser cantidad, tiempo, dinero, etc., sobre la cual se desea realizar una operación de agregación (promedio, conteo, suma, etc.). La granularidad es el nivel de detalle que posee cada registro de una tabla de hechos.

### **3.1.3.5 Modelo gráfico de alto nivel**

Para concluir con el proceso dimensional inicial se realiza un gráfico denominado modelo dimensional de alto nivel (o gráfico de burbujas, Bubble chart, en el léxico de Kimball), como se presenta en la Ilustración 14.



***Ilustración 14. Ejemplo de Modelo final de alto nivel de la sesión inicial de diseño (Mundy & Thornthwaite 06)***

### **3.1.3.6 Identificación de atributos de dimensiones y tablas de hechos**

La segunda parte de la sesión inicial de diseño consiste en completar cada tabla con una lista de atributos bien formada. Esta lista o grilla se forma colocando en las filas los atributos de la tabla, y en las columnas la siguiente información:

- Características relacionadas con la futura tabla dimensional del almacén de datos (target), por ejemplo tipo de datos.
- El origen de los datos (source, por lo general atributos de las tablas transaccionales).
- Reglas de conversión, transformación y carga (ETL rules), que nos dicen cómo transformar los datos de las tablas de origen a las del almacén de datos.

### **3.1.4 Modelo Físico**

En esa fase consiste en la construcción de las estructuras físicas dentro de la base de datos que soportaran el modelo dimensional. Además, se definen estándares para los nombres de los atributos o elementos, se configura la base de datos, se crean índices o demás estructuras auxiliares requeridas.

### **3.1.5 Diseño e implementación del subsistema de ETL**

El sistema de Extracción, Transformación y Carga (ETL) es la base sobre la cual se alimenta el Datawarehouse. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el DW en un formato acorde para la utilización por parte de las herramientas de análisis.

### **3.1.6 Diseño técnico de la arquitectura**

La infraestructura será la encargada de soportar o permitir la realización del almacén de datos. La infraestructura incluye el hardware, las redes, elementos de seguridad, funciones de bajo nivel requeridas para la existencia del almacén de datos.

Dentro de la infraestructura también se puede señalar la metada, porque es una estructura lógica da soporte a los datos y la cual puede ser utilizada por distintas herramientas para extraer información de los modelos de datos definidos dentro del sistema (metada de las fuentes de datos y metada del modelo dimensional).

Dentro de la infraestructura es importante establecer las variables que van a afectar la elección de la infraestructura y documentar la arquitectura que se posee.

La elección de la arquitectura dependerá del nivel o aspecto que sea visualizado. En este sentido a nivel del servidor, que es lugar donde se tendrá el almacén de datos, se tiene que evaluar la concurrencia de usuarios, el tamaño de los datos, la capacidad de realizar procesamiento paralelo, el sistema operativo, la volatilidad del sistema entre otros. Dentro del servidor tiene que establecerse el tipo de base de datos a utilizar, relacional o multidimensional.

A nivel de red debe evaluarse el ancho de banda que se posee desde el servidor hasta las distintas terminales de los usuarios, así como los distintos conectores que se posean (conectores a bases de datos o alguna aplicación específica).

Por último, se tiene que evaluar el terminal donde el usuario va a acceder al almacén de datos. Los distintos sistemas operativos o aplicaciones que se pueden tener, evaluando que el desempeño de las aplicaciones de usuario no se vea afectado, es decir, existe una independencia de la arquitectura. También es importante evaluar las capacidades de procesamiento y memoria del terminal del usuario final con el fin de mantener tiempos óptimos de respuesta.

### **3.1.7 Selección e instalación de productos**

Una vez pasada la parte teórica de diseño de la arquitectura, se pasa a la selección de los productos requeridos para soportar la infraestructura. Muchas veces estos dos procesos se ejecutan de forma paralela en el sentido de que en la evaluación de los componentes dentro del diseño de la arquitectura, ya se selecciona la mejor opción para el almacén de datos a desarrollar.

Para cada componente se debe realizar una evaluación de los distintos productos existentes en el mercado, esto atado a escoger herramientas que puedan soportar los requerimientos de los usuarios. Dentro de la selección de productos Kimball (1998) establece que existen 4 áreas importantes donde se tiene que adquirir componentes y que son los que harán posibles la implementación del almacén de datos, las cuales son:

- 1. Plataforma de hardware:** Los equipos redes que soportaran el almacén de datos.

2. **Sistema manejador de base datos:** Encargado de gestionar la base de datos.
3. **Herramientas de ETL y área intermedia:** Los elementos que tomaran las fuentes de datos y pasaran estos al almacén de datos, haciendo las transformaciones pertinentes para que estos datos puedan ser utilizados.
4. **Herramienta de acceso a los datos:** Todo el conjunto de herramientas que servirán a los usuarios finales para realizar análisis.

Se debe estudiar los productos de cada uno de estos elementos, quienes son los fabricantes, referencias sobre ellos, comparación de implementación, costos, y ver cuales productos se adaptan más a los requerimientos.

### **3.1.8 Especificación y desarrollo de aplicaciones de BI**

Una parte fundamental de todo proyecto de DW/BI está en proporcionarles a una gran comunidad de usuarios una forma más estructurada y por lo tanto, más fácil, de acceder al almacén de datos. Proporcionamos este acceso estructurado a través de lo que llamamos aplicaciones de inteligencia de negocio. Las aplicaciones de BI son la cara visible de la inteligencia de negocio: los informes y aplicaciones de análisis proporcionan información útil a los usuarios.

Las aplicaciones de BI incluye un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio. Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama informes estándar y aplicaciones analíticas.

#### **3.1.8.1 Informes estándar Los informes estándar son la base del espectro de aplicaciones de BI**

Por lo general son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos. En el caso más simple, son informes estáticos pre almacenado.

Los informes estándar proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa. Este tipo de aplicaciones son el caballo de batalla de la BI de la empresa. Son informes que los usuarios usan día a día. La mayor parte de lo que piden las personas durante el proceso de definición de requisitos se clasificaría como informes estándar.

### **3.1.8.2 Aplicaciones analíticas Las aplicaciones analíticas son más complejas que los informes estándar**

Estas aplicaciones pueden ser muy avanzadas e incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los datos. Otra característica avanzada en algunas aplicaciones analíticas es que el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de BI. En el otro extremo del espectro, algunas aplicaciones analíticas se venden como soluciones cerradas o enlatados, y son independientes de las aplicaciones particulares de la empresa. Algunas aplicaciones analíticas comunes incluyen:

- Análisis de la eficacia de la promociones.
- Análisis de rutas de acceso en un sitio Web.
- Planificación del espacio en espacios comerciales.
- Detección de fraudes.

### **3.1.9 Implementación**

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Existen varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de feedback.

La implementación puede darse a partir de un lanzamiento de prueba (beta), donde solo accedan un grupo de usuarios definidos, esto permite que la puesta en producción no abarque toda la organización, entendiendo que una primera versión de la misma puede contener múltiples errores o bugs. Los usuarios beta sirven entonces para probar la herramienta y asegurar que el producto que será puesto en producción posee ciertos criterios de calidad previamente definidos (eficiencia, efectividad, usabilidad, entre otros).

### **3.1.10 Mantenimiento y crecimiento**

Para administrar el entorno del Datawarehouse existente es importante enfocarse en los usuarios de negocio, los cuales son el motivo de su existencia, además de gestionar adecuadamente las operaciones del Datawarehouse, medir y proyectar su éxito y comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación.

En esto consiste el Mantenimiento. Finalmente, es importante sentar las bases para el crecimiento y evolución del Datawarehouse en donde el aspecto clave es manejar el crecimiento y evolución de forma iterativa utilizando el Ciclo de Vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel prioridad.

### **3.1.11 Administración de Proyecto**

Es una tarea que se ejecuta durante todo el desarrollo del proyecto. Esta tarea implica la verificación de que las actividades del proyecto se están ejecutando según lo planificado. Por tanto, la gestión del proyecto se basa en monitorizar todas las fases del proyecto en busca de que la ejecución del proyecto no se salga de los parámetros establecidos (tiempo, costo, producto final) y en dado caso de que exista algún elemento que pueda afectar el desarrollo del proyecto, dicho factor pueda ser detectado y corregido a tiempo.

Si un elemento no puede ser corregido, y está afectando la planificación del proyecto, se tienen que reajustar las demás actividades y sus tiempos de ejecución.

En caso de que un elemento no pueda ser corregido y afecte la planificación del proyecto, se tienen que reajustar las demás actividades y sus tiempos de ejecución. Todo esto para tener como última opción la modificación de la planificación establecida, la cual debe modificarse si y solo si no existe forma de realizar el proyecto con los parámetros preestablecidos.

### **3.2 Metodología Scrum**

Es una metodología en la que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto. Estas prácticas se apoyan unas a otras y su selección tiene origen en un estudio de la manera de trabajar de equipos altamente productivos.

Es una opción de gestión ideal para acometer proyectos desarrollados en entornos complejos que exigen rapidez en los resultados y en los que la flexibilidad es un requisito imprescindible. Scrum ofrece agilidad y el, resultado, siempre, valor.

### 3.2.1 Características de Scrum

SCRUM es un modelo de referencia que define un conjunto de prácticas y roles, y que puede tomarse como punto de partida para definir el proceso de desarrollo que se ejecutará durante un proyecto.

Los roles principales en Scrum son el 'Scrum Master, que procura facilitar la aplicación de Scrum y gestionar cambios, el Product Owner, que representa a los stakeholders (interesados externos o internos), y el Team (equipo) que ejecuta el desarrollo y demás elementos relacionados con él. Durante cada sprint, un periodo entre una y cuatro semanas (la magnitud es definida por el equipo y debe ser lo más corta posible), el equipo crea un incremento de software potencialmente entregable (utilizable).

El conjunto de características que forma parte de cada sprint viene del Product Backlog, que es un conjunto de requisitos de alto nivel priorizados que definen el trabajo a realizar (PBI, Product Backlog Item). Los elementos del Product Backlog que forman parte del sprint se determinan durante la reunión de Sprint Planning.

Durante esta reunión, el Product Owner identifica los elementos del Product Backlog que quiere ver completados y los hace del conocimiento del equipo. Entonces, el equipo conversa con el Product Owner buscando la claridad y magnitud adecuadas (Cumpliendo el INVEST) para luego determinar la cantidad de ese trabajo que puede comprometerse a completar durante el siguiente sprint. Durante el sprint, nadie puede cambiar el Sprint Backlog, lo que significa que los requisitos están congelados durante el sprint.

Scrum permite la creación de equipos auto organizados impulsando la co-localización de todos los miembros del equipo, y la comunicación verbal entre todos los miembros y disciplinas involucrados en el proyecto.

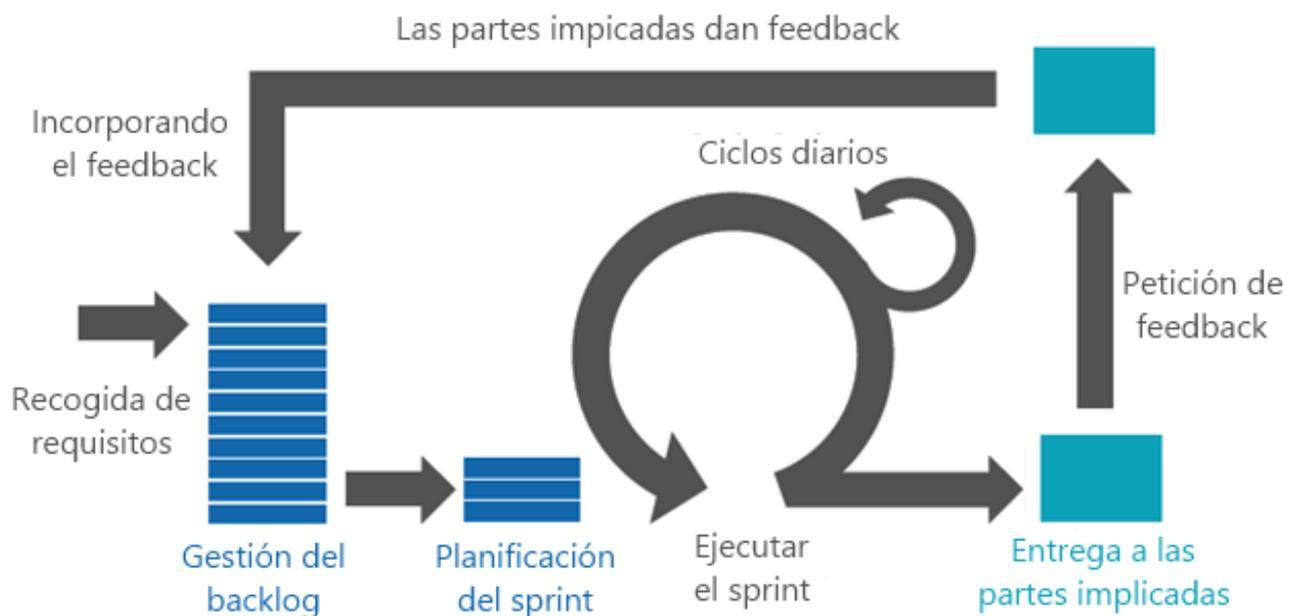
## Beneficios de la metodología Scrum

Los beneficios son amplios y repercuten en el equipo, en los Stakeholders y en la organización en su conjunto.

Se fomenta el trabajo en equipo, focalizando todos los esfuerzos en alcanzar un objetivo común. Se trata de un modelo basado en la auto-disciplina y la auto-gestión, lo que repercute positivamente en la responsabilidad. Respecto al aspecto comunicativo, esta metodología fomenta la comunicación entre los distintos miembros del equipo.

Los Stakeholders tienen un mayor control y transparencia sobre el proyecto, permitiendo una mejor organización. El cliente puede hacer seguimiento más cercano de lo que pasa, sin tener que esperar a un resultado final que no le convenza. Con las metas intermedias se minimizan riesgos.

En definitiva, la adopción de estas buenas prácticas permite reducir el tiempo de desarrollo de productos, más capacidad de adaptación y flexibilidad frente a un entorno y unos requisitos cambiantes aumentando el valor que se aporta a los clientes. En la ilustración 15, se puede observar el ciclo de vida de SCRUM:



**Ilustración 15. Ciclo de vida de la metodología de SCRUM**

## CAPÍTULO 4

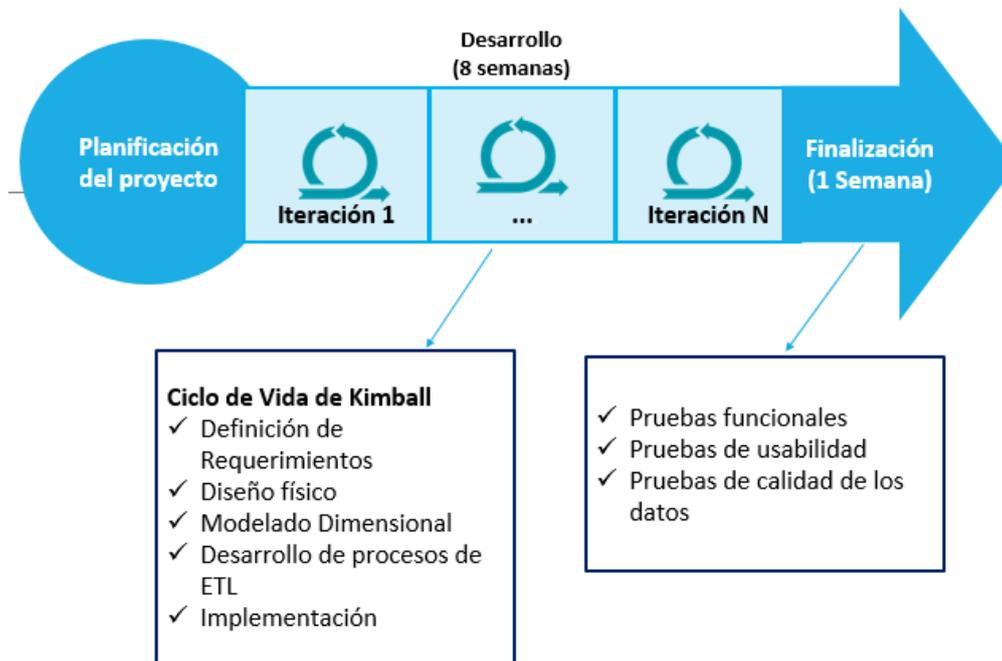
### MARCO APLICATIVO

La implementación de este TEG, se desarrolló basada en la metodología definida por Ralph Kimball, que permite la construcción de una solución de Inteligencia de Negocio de manera eficiente a través de una serie de fases. Y a su vez, para tener un mayor control de proyecto en períodos de trabajo se usa la metodología Scrum.

Primero se va a definir las actividades a realizar con sus respectivos períodos de tiempo (Sprint) en este proyecto. Cabe destacar que en cada Sprint se involucran las actividades del ciclo de vida de Kimball.

**Tabla 2. Productos a entregables de la planificación de SCRUM.**

<b>Productos a entregables</b>	<b>Sprint</b>	<b>Duración</b>
1. Definir los requerimientos de negocios	Sprint #1	1 semana
2. Diseñar el almacén de datos y ETL	Sprint #2	3 semanas
3. Desarrollar el almacén de datos y ETLs	Sprint #3	4 semanas
4. Desarrollar indicadores de gestión	Sprint #4	4 semanas
5. Pruebas funcionales, pruebas de usabilidad, de calidad de datos	Sprint #5	1 semanas



A continuación, se describen las actividades realizadas en todo el proceso de desarrollo de la solución propuesta, distribuidas en cada fase que comprende la metodología aplicada de Ralph Kimball, desde la planificación del proyecto, el diseño del almacén de datos y la elaboración de los procesos de extracción, transformación y carga de los datos del almacén, hasta la realización de las consultas analíticas para el desarrollo y visualización de los indicadores, terminando con la implementación un tablero de control(Dashboard) mediante una herramienta de Inteligencia de Negocio seleccionada, para poder cumplir con los requerimientos analíticos de los procesos de intermediación, compra y subasta de productos relacionados con el gas y petróleo.

#### **4.1 Fases del Ciclo de vida de Kimball.**

##### **4.1.1 Planificación del Proyecto.**

En esta primera fase se establecen unos lineamientos y actividades planteadas en períodos de tiempo específicos, que permiten la implementación de la solución de Inteligencia de Negocio planteada.

Kimball en su metodología define que en la primera fase de la implementación de una solución de inteligencia de negocio se tiene que realizar un levantamiento y análisis de los requerimientos, definir la arquitectura de la solución, instalar y configurar los productos tecnológicos seleccionados, diseñar e implementar el almacén de datos, diseñar y elaborar los procesos de extracción, transformación y carga de los datos ETLs, y para finalizar el desarrollo del tablero de control, de los reportes e informes de análisis con los indicadores propuestos.

*Las actividades serán presentadas en la siguiente tabla de planificación:*

**Tabla 3. Actividades a realizar para la implementación de la solución de Inteligencia de Negocio.**

<b>Actividades</b>		<b>Duración estimada</b>	<b>Fecha</b>
1. Definir los requerimientos de negocio	1. Definir y analizar los requerimientos del negocio 2. Identificar temas analíticos del negocio 3. Identificar los procesos de negocios de la organización.	2 días	12/03/18 al – 14/03/18
2. Diseñar el almacén de datos y ETL	1 Diseñar modelo entidad/relación de la base de datos a utilizar 2 Diseñar el modelo dimensional a utilizar 3 Seleccionar el modelo físico a utilizar a implementar en la aplicación 4 Diseñar los procesos de ETL.	1 semana	14/03/18 al 21/03/18

<b>Fase</b>	<b>Actividades</b>	<b>Duración estimada</b>	<b>Fecha</b>
3. Desarrollar el almacén de datos y ETLs	<ol style="list-style-type: none"> <li>1. Seleccionar las herramientas tecnológicas que se usarán para el desarrollo de la aplicación de BI</li> <li>2. Descargar, instalar y configurar las herramientas tecnológicas que se utilizarán para el desarrollo de la aplicación de BI</li> <li>3. Implementar el modelo dimensional.</li> <li>4. Implementar el almacén de datos</li> <li>5. Desarrollar los procesos ETL asociados</li> </ol>	4 semanas	21/03/18 al 10/04/18
4. Desarrollar indicadores de gestión	<ol style="list-style-type: none"> <li>1. Desarrollar la aplicación de BI</li> <li>2. Desarrollar los indicadores de gestión</li> <li>3. Realizar reportes y análisis de los indicadores de gestión.</li> </ol>	4 semanas	10/04/18 10/05/18
5. Pruebas funcionales, pruebas de usabilidad, de calidad de datos	<ol style="list-style-type: none"> <li>1. Aplicar pruebas funcionales y pruebas de usabilidad de la aplicación de BI</li> <li>2. Aplicar pruebas de calidad de datos de la aplicación</li> </ol>	1 semana	10/05/18 17/05/18

#### **4.1.2 Definición de los Requerimientos del Negocio**

El objetivo de implementar una solución de Inteligencia de Negocio es que sea de gran utilidad para las empresas relacionadas con el gas y petróleo, para que exista un mayor control y manejo de la información que se genera en ellas, y su vez que exista una herramienta que ayude a tomar mejores decisiones. Debido a esto, se realizó una investigación con la finalidad de conocer los procesos y administración de la empresa a través de entrevistas y reuniones con los directivos y empleados de varios departamentos de la empresa para obtener cual era la información que requerían saber de sus operaciones.

Obteniendo como resultado, una lista de requerimientos:

- ✓ Cantidad de visitas activas e inactivas en el portal web
- ✓ Porcentaje las publicaciones activas e inactivas que se encuentran en el portal web
- ✓ Secciones más visitadas del portal web.
- ✓ Distribución porcentual de los productos publicados por categorías en el portal web.
- ✓ Cantidad Surplus (ventas de productos sobrantes de un proyecto de petróleo) publicados mensualmente en el portal web
- ✓ Promedio del costo de un Surplus en el portal web por año.
- ✓ Productos más contactados en el portal web.
- ✓ Cantidad de productos contactados en el portal web al mes.
- ✓ Distribución de productos publicados en el portal web por marca.
- ✓ Categorías más buscadas en el portal web.
- ✓ Cantidad de productos contactados a través del portal web en años diferentes
- ✓ Cantidad de subastas en el portal web anual.
- ✓ Ganancia de compra inmediata (BON) Vs. Ganancia esperada por mes
- ✓ Cantidad de productos vendidos por compra inmediata (por marca)
- ✓ Ganancia de publicaciones en el portal web (por trimestre)

En la siguiente tabla se presentará la información necesaria para calcular los indicadores de la solución de Inteligencia de Negocio:

Fórmula (Base\_de\_datos.nombre\_tabla.campo)

**Tabla 4. Indicadores a desarrollar en la Solución de Inteligencia de Negocio.**

<b>Nombre del indicador</b>	<b>Variables a usar</b>	<b>Fórmula para el cálculo</b>	<b>Tipo de indicador</b>
<b>Cantidad de visitas activas e inactivas</b>	CVA - Cantidad de visitas activas VP - Visitas en el portalweb	$CVA = \sum VP$ CV (Tabla de visitas) VP = Tangerix_mvp.visit.id_visit	Eficacia
<b>Porcentaje de publicaciones activas e inactivas</b>	PPA – Porcentaje de publicaciones activas CPA – Cantidad de publicaciones activas en el portal web PP – Publicaciones en el portal web	$PPA = \frac{CPA}{PP} * 100$ CPA (Tabla de productos, campo estatus = “Activo” e “Inactivo”) CPA= Tangerix_mvp.productos.estatus PP (Suma total de los productos publicados)	Eficacia
<b>Secciones más visitadas en el portal web</b>	CSP= Cantidad de visitas en las secciones de la página Max= máximo de visitas por sección	$CSP = Max(\sum CPSi)$ CPSi (Tabla de visitas con el campo, sección) CPS= Tangerix_mvp.visit.section	Eficiencia

<b>Nombre del indicador</b>	<b>VARIABLES A USAR</b>	<b>Fórmula para el cálculo</b>	<b>Tipo de indicador</b>
<b>Distribución porcentual de productos publicados (por categoría)</b>	DPC= Distribución porcentual de las publicaciones por categorías CC= Cantidad de productos por categoría CP= Cantidad de Publicaciones	$DPC = \frac{\sum CC}{(CP)} * 100$ CC (Tabla productos, campo cate CC=Tangerix_mvp.productos.id_categoria CP (Tabla de productos, campo estatus activo) CP=Tangerix_mvp.productos.es_tatus	Proporcional
<b>Cantidad de Surplus publicados</b>	TS= Total surplus S= Surplus publicados	$TS = \sum S$ S (Tabla de surplus) S=Tangerix_mvp.surplus.id_surplus	Eficacia
<b>Promedio del costo de Surplus</b>	PS= Promedio de ítems de surplus CPS= Costo por surplus TS= Total de surplus	$PS = \frac{\sum CPS}{TS}$ TS (Total de surplus) TS=Tangerix_vp.surplus.id_surplus CPS (Tabla de surplus, campo costo) CPS=Tangerix_mvp.surplus.costo	Eficacia
<b>Productos más contactados del portal web</b>	PM= Producto más contactado del portal web PC= Producto contacto	$PM = \text{Max}(\sum PC)$ PC(Tabla de contacto, campo id producto) PC=Tangerix_mvp.contact_sellers.id_producto	

<b>Nombre del indicador</b>	<b>VARIABLES A USAR</b>	<b>Fórmula para el cálculo</b>	<b>Tipo de indicador</b>
<b>Cantidad de productos contactados</b>	<p>CPr= Cantidad de productos</p> <p>PC= Productos contactados</p>	$CPR = \sum PC$ <p>PC (Tabla de contacto, campo id del producto, campo fecha que sea del mes que se quiere buscar)</p> <p>PC= Tangerix_mvp.contact_sellers.contact_date</p>	Eficacia
<b>Distribución de productos publicados (por marca)</b>	<p>DPM= Distribución porcentual de productos por marca</p> <p>CPM= Cantidad de publicación por marca</p> <p>CP= Cantidad de publicaciones en total</p>	$DPM = \frac{\sum CPM}{CP} * 100$ <p>CPM (Tabla de productos, campo marca)</p> <p>CPM=Tangreix_mvp.productos.id_marca</p> <p>CP (Tabla de productos, campo estatus activo)</p> <p>CP=Tangerix_mvp.productos.estatus</p>	Proporcional
<b>Categorías más buscadas(contactadas) en el portal web</b>	<p>CMB= Categoría más buscada</p> <p>CC= Categoría contactada</p>	$CMB = MAX (\sum CC)$ <p>CC (Tabla de visitas, campo sección categorías)</p> <p>CC=Tangerix_mvp.visit.id_identifier</p> <p>Section='categorias'</p>	Eficiencia

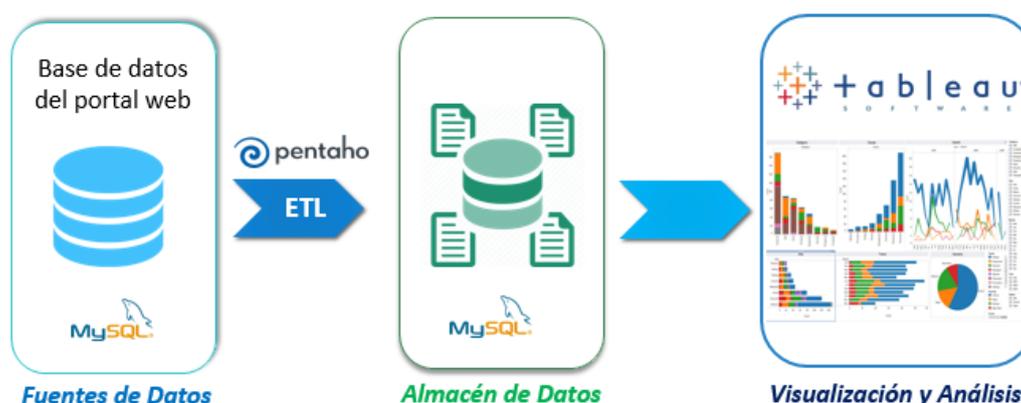
<b>Nombre del indicador</b>	<b>VARIABLES A USAR</b>	<b>Fórmula para el cálculo</b>	<b>Tipo de indicador</b>
<b>Variación de productos contactados en el portal web (entre dos años diferentes)</b>	VP= Variación de Productos contactados PC1=Productos contactados año 1 PC2= Productos contactados año 2	$VP = \frac{(\sum PC1 - \sum PC2)}{PC1} * 100$ PC1 y PC2 (Tabla productos, campo año que se requiera comparar) PC1=Tangerix_mvp.contact_sellers.contact_date PC2=Tangerix_mvp.contact_sellers.contact_date	Variación proporcional
<b>Cantidad de subastas anuales</b>	CSA= Cantidad de subastas anuales S= Subastas	$CSA = (\sum S)$ S (Tabla subastas, campo año) S= Tangerix_mvp.bids.fecha	Eficacia
<b>Ganancia de compra inmediata (BON) Vs. Ganancia Esperada por mes</b>	MBON= Monto total por BON M= Monto por cada BON GE= Ganancia esperada	$\frac{MBON}{GE} * 100$ M(Tabla de compra inmediata, campo monto) M=Tangerix_mvp.purchases.price	Eficacia
<b>Cantidad de productos vendidos por compra inmediata</b>	CBON=Cantidad de BON PBON= Productos vendidos por BON	$CBON = \sum PBON$ PBON(Tabla de compra inmediata, campo id de producto)	Eficacia

<b>Nombre del indicador</b>	<b>VARIABLES A USAR</b>	<b>Fórmula para el cálculo</b>	<b>Tipo de indicador</b>
<b>Ganancia de publicaciones (por trimestre)</b>	MT= Monto trimestral ganado MG = Monto ganado ME= Monto esperado	$MT = \frac{\sum MG}{ME} * 100$ MG(Tabla de compras, campo total compras) MG=Tangerix_mvp.compras.total_compra	Eficiencia

### 4.1.3 Diseño de la Arquitectura Técnica

Previamente estudiada la situación del negocio y encontrado los problemas previamente mencionados, se propone un diseño de la arquitectura de la solución de Inteligencia de Negocio con el fin de obtener indicadores de gestión que sean necesarios y tengan relevancia relacionados con los procesos de ventas, compras y subastas de productos en el área petrolera de un portal web en específico.

Como se observa en la ilustración 16, la arquitectura está formada por los siguientes componentes: datos fuentes que los tenemos con una base de datos proporcionada por el portal web que almacena los datos; un almacén de datos, los procesos de extracción, transformación y carga, cuadros de mando y reportes que soportan los indicadores de gestión.



**Ilustración 16. Arquitectura de la solución de Inteligencia de Negocio (Cano, 2007).**

Partimos por la base de datos fuentes del portal web del negocio, los cuales tienen que pasar por un proceso de ETL, que tiene como consecuencia la creación de un almacén de datos la cual será una base de datos relacional cuyo diseño lógico sigue un modelo E/R genérico de las ventas y compras del portal web.

Se procede al diseño y construcción de un almacén de datos en MySQL, debido a que la base de datos del portal web está relacionado con este sistema manejador de base de datos, este almacén de datos está acorde a los indicadores que pudieron ser obtenidos de los datos previamente almacenados, y a su vez que puedan ser útiles de acuerdo con las necesidades que vayan surgiendo en el sistema.

Posteriormente, se comienza a extraer y transformar los datos con el proceso de ETL para ser cargados en el almacén de datos, este proceso se realiza con la herramienta de Pentaho denominada "Pentaho Data Integration", estos son de gran importancia para la creación del almacén de datos que se utiliza por la solución de Inteligencia de Negocio.

Finalmente al terminar el proceso de creación del almacén de datos, se desarrolla la capa de presentación, es decir, las herramientas de acceso a los datos dispuestos en el almacén para permitir que los usuarios finales puedan visualizar y analizar los distintos indicadores y reportes contemplados para esta solución de inteligencia de negocio.

En este componente se recurre al conjunto de componentes que provee la herramienta llamada "Tableau Software" para el desarrollo de los tableros de control (DashBoard), Tableau Server que permite conectar cualquier fuente de datos de manera segura, ya sea en las instalaciones físicas o en la nube. Y su vez, Tableau Online la cual es una plataforma de análisis completamente hospedada en la nube que permite compartir los tableros de control con otros usuarios interesados, haciendo todo posible a través de un navegador, o un dispositivo móvil con la aplicación móvil de Tableau.

#### **4.1.4 Selección de Productos e Instalación**

El proceso de selección de una herramienta tecnológica para el desarrollo e implementación de una solución de inteligencia de negocio, resulta un poco complejo, ya que existen diversas herramientas en el mercado y a la variada gama de funcionalidades que cada una de ellas ofrece, estas herramientas son: Pentaho Business Intelligence, Tableau Software, Oracle Business Intelligence (OBI), IBM Cognos, SpagoBI y Microsoft Power Business Intelligence, entre otra más.

Luego de estudiar las distintas herramientas presentes en el mercado, se optó utilizar para el desarrollo de la solución de Inteligencia de Negocio planteada, las componentes de la herramienta de Tableau Software, la herramienta de Pentaho denominada "Pentaho Data Integration", y la base de datos MySQL.

**Tabla 5. Herramientas para desarrollar la solución de Inteligencia de Negocio.**

<b>Herramienta</b>	<b>Uso</b>	<b>Requisitos de Instalación</b>
Pentaho Data Integration 7.1	Creación y Administración de Procesos ETL	Sistema Operativo: Windows o Linux Memoria RAM: min 1 GB, Recomendable > 2 GB Memoria en disco duro: 200 MB
Tableau Desktop	Creación y Administración de tableros de control, así como servidor para alojar los reportes	Memoria RAM recomendable: 2GB Espacio Libre en disco: 15GB Sistema Operativo: Windows o Linux.
MySQL 5.0	Creación y administración de bases de datos	512 Mb de memoria Ram. 1 GB de espacio de disco duro. Sistema operativo: Windows, Linux o Unix. Arquitectura del sistema 32/64 bit.
Navicat 11.0.8 Premium	Creación, administración y conexión con las bases de datos.	Microsoft Windows Vista, Windows 7, Windows 8, Windows 8.1, Windows 10, Server 2008, Server 2012, Server 2016

#### **4.1.5 Modelado Dimensional**

Teniendo en cuenta los requerimientos obtenidos en el punto 4.1.2 se validaron las relaciones entre los criterios y la lógica del negocio de acuerdo con la estructura de la base de datos del portal web de la organización, y así poder determinar cuáles elementos podrían considerarse como hechos medibles o dimensiones que puedan formar parte de la solución de Inteligencia de Negocio y con esto poder dar inicio al diseño del modelo dimensional del almacén de datos, el cual comprende los procesos de intermediación, subastas y compra inmediata de productos relacionados con el gas y petróleo.

Para el desarrollo del modelo dimensional hay que seguir un conjunto de actividades propuestas por Ralph Kimball, las cuales son las siguientes:

##### **4.1.5.1 Definir el Proceso de Negocio**

Se modela el proceso de intermediación, compra o subastas de productos relacionados con el gas y petróleo de un portal web a través de indicadores.

##### **4.1.5.2 Identificar el Nivel de Granularidad**

Es necesario especificar cómo se representaría una fila en la tabla de hechos para determinar el nivel de granularidad, porque gracias a esto, se define el nivel de detalle de la información que se desee almacenar del proceso a modelar, permitiendo identificar de manera más sencilla lo que se espera media con cada indicador propuesto.

Se procede a establecer el siguiente nivel de granularidad:

- ✓ Visitas de una sección en particular en el portal web, por un usuario específico, en una fecha específica.
- ✓ Contactos realizados hacia un producto publicado en el portal web con su respectivo contacto (contact\_sellers), en una fecha específica, por un usuario específico.

- ✓ Publicaciones de un producto en específico en el portal web, publicado por un usuario, en una fecha específica y con un monto en específico de publicación.
- ✓ Listados de Surplus publicados en el portal web, por un usuario en una fecha específica con un costo específico de publicación.
- ✓ Compras de manera inmediata (BON = Buy Option Now) de un producto en específico en el portal web, comprado por un usuario, en una fecha específica y con un monto en específico de compra.
- ✓ Subastas de un producto en específico en el portal web, con la participación de un usuario, en una fecha específica y con un monto específico de puja en la subasta.

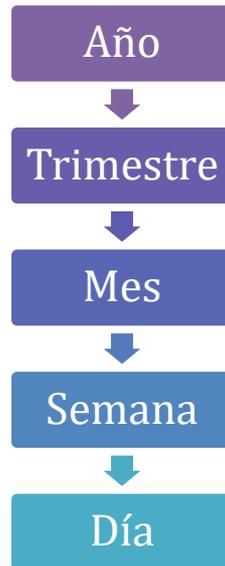
### 4.1.5.3 Definir las Dimensiones

De acuerdo con los requerimientos de información obtenidos previamente y el análisis de la base de datos del portal web, en la fase de diseño, se identificaron las siguientes dimensiones:

**Tabla 6. Dimensiones del Modelo Dimensional de la solución.**

<b>Dimensión</b>	<b>Nombre en Almacén de datos</b>	<b>Descripción</b>
Tiempo	Dim_tiempo	En esta tabla se guardan todos los datos referentes al tiempo (año, semestre, mes, día).
Usuario	Dim_usuario	Esta tabla contiene toda la información (nombre, apellido, etc.) de los usuarios registrados en el portal web.
Producto	Dim_producto	Esta tabla almacena la información (nombre, tipo, estatus, etc.) de los productos publicados en el portal web.
Categoría	Dim_categoria	Esta tabla contiene la información de las diferentes categorías que poseen los productos del portal web.
Marca	Dim_marca	Esta tabla contiene la información de las diferentes marcas que poseen los productos del portal web.
Contacto	Dim_contacto	Esta tabla almacena la información (nombre, teléfono, email, etc.) de la persona contacto del producto publicado en el portal web.
Sección	Dim_seccion	En esta tabla tenemos la información (home, categorías, productos, etc.) de las diferentes secciones del portal web.
Surplus	Dim_surplus	Almacena la información (archivo, mensaje, costo) de los Surplus publicados en el portal web

A su vez, se identificaron las relaciones jerárquicas que surgen entre los atributos de las diferentes dimensiones según la lógica de negocio, como se observa en la ilustración 17:



**Ilustración 17. Relaciones Jerárquicas del modelo dimensional de la solución propuesta**

#### 4.1.5.4 Identificación de los Hechos y las Tablas de Hechos

Una vez definidas todas las dimensiones, se generó la tabla de hechos:

- ✓ **FACT\_VISITA:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes con respecto a las visitas generadas en el portal web.

**Tabla 7. Hechos de la tabla "Fact\_visita"**

<i>Hecho</i>	<i>Nombre en Almacén de datos</i>
Cantidad de visitas	Cantidad

- ✓ **FACT\_CONTACTO\_VENDEDOR:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes con respecto a los productos contactados a través del portal web para la intermediación de la compra de los mismos.

**Tabla 8. Hechos de la tabla "Fact\_contacto\_vendedor"**

<b>Hecho</b>	<b>Nombre en Almacén de datos</b>
Cantidad de contacto	Cantidad

- ✓ **FACT\_PUBLICACIONES:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes con respecto a los productos publicados en el portal web.

**Tabla 9. Hechos de la tabla "Fact\_publicaciones"**

<b>Hecho</b>	<b>Nombre en Almacén de datos</b>
Cantidad de publicaciones	Cantidad
Estatus de publicación	Estatus
Ganancia por publicación	Ganancia

- ✓ **FACT\_COMPRA\_INMEDIATA:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes asociados a los productos comprados de manera inmediata en el portal web

**Tabla 10. Hechos de la tabla "Fact\_compra\_inmediata"**

<b>Hecho</b>	<b>Nombre en Almacén de datos</b>
Cantidad de compra inmediata	Cantidad
Estatus de publicación	Estatus

- ✓ **FACT\_SUBASTA:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes con respecto a las subastas de los productos realizadas en el portal web.

**Tabla 11. Hechos de la tabla "Fact\_subasta"**

<b>Hecho</b>	<b>Nombre en Almacén de datos</b>
Cantidad de subastas	Cantidad
Monto de la puja	Monto_puja

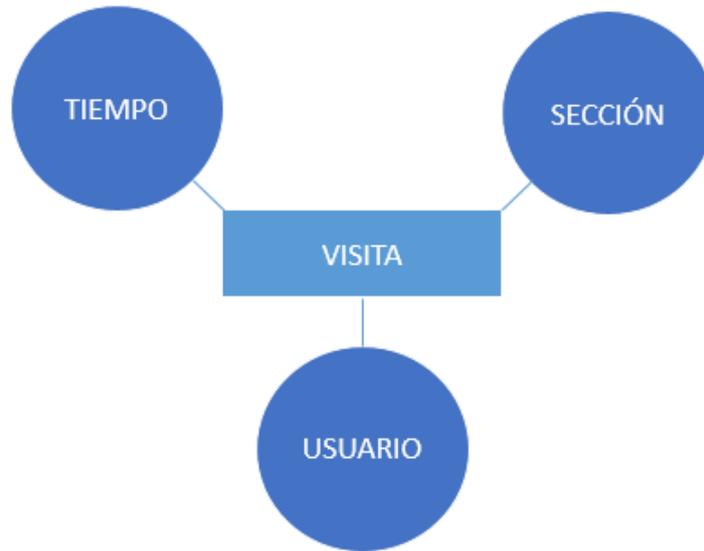
- ✓ **FACT\_SURPLUS\_PUBLICACIÓN:** Almacena toda la información necesaria para implementar el tablero de control (Dashboard) y generar reportes asociados a los Surplus publicados en el portal web.

**Tabla 12. Hechos de la tabla "Surplus\_publicacion"**

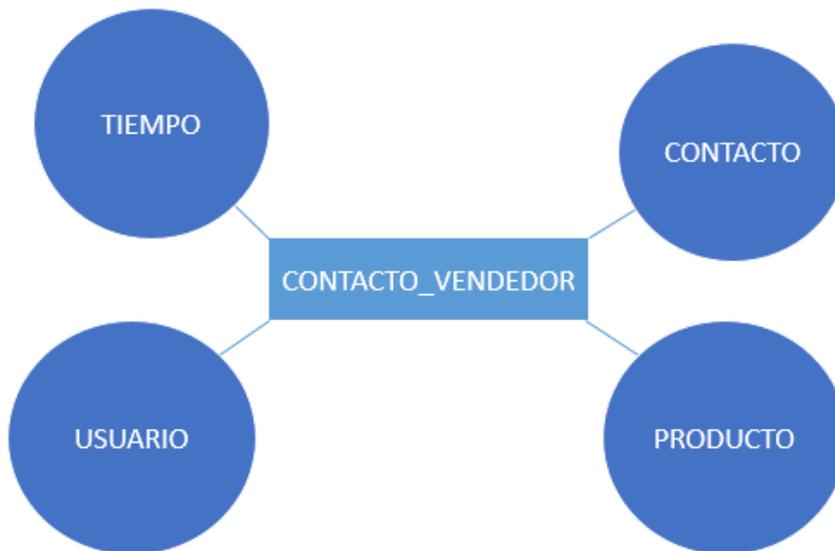
<b>Hecho</b>	<b>Nombre en Almacén de datos</b>
Cantidad de surplus	Cantidad
Costo del Surplus	Costo

#### **4.1.5.5 Modelo gráfico de alto nivel**

Una vez identificadas las dimensiones, los hechos y las tablas hechos se realizan varios gráficos que tienen el nombre de "modelo dimensional de alto nivel (o gráfico de burbujas, Bubble chart, en el léxico de Kimball). Como se presenta en las ilustraciones 18 hasta 23.



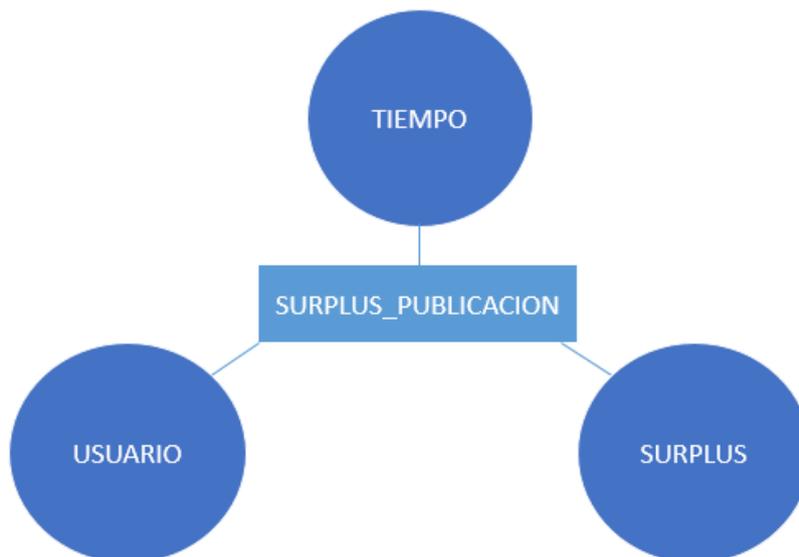
**Ilustración 18. Modelo gráfico de alto nivel de Visitas**



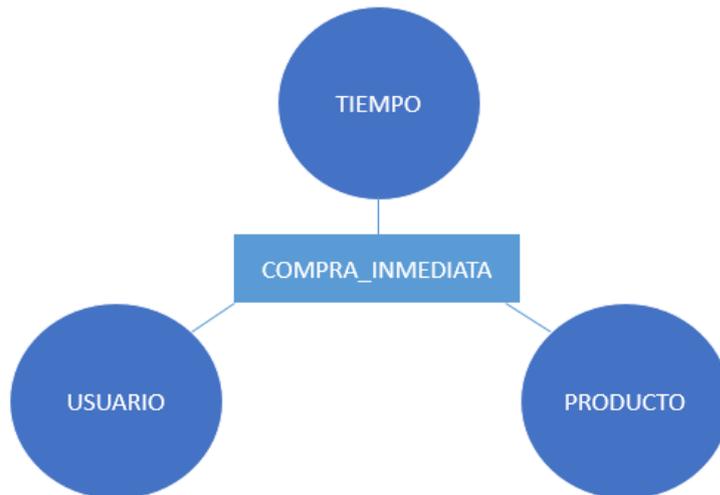
**Ilustración 19. Modelo gráfico de alto nivel de Contacto Vendedor**



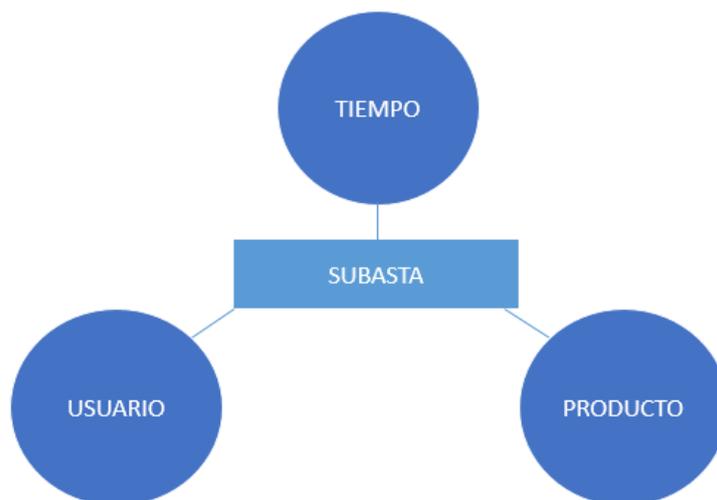
***Ilustración 20. Modelo gráfico de alto nivel de Publicaciones***



***Ilustración 21. Modelo gráfico de alto nivel de Surplus publicación***



***Ilustración 22. Modelo gráfico de alto nivel de Compra Inmediata***

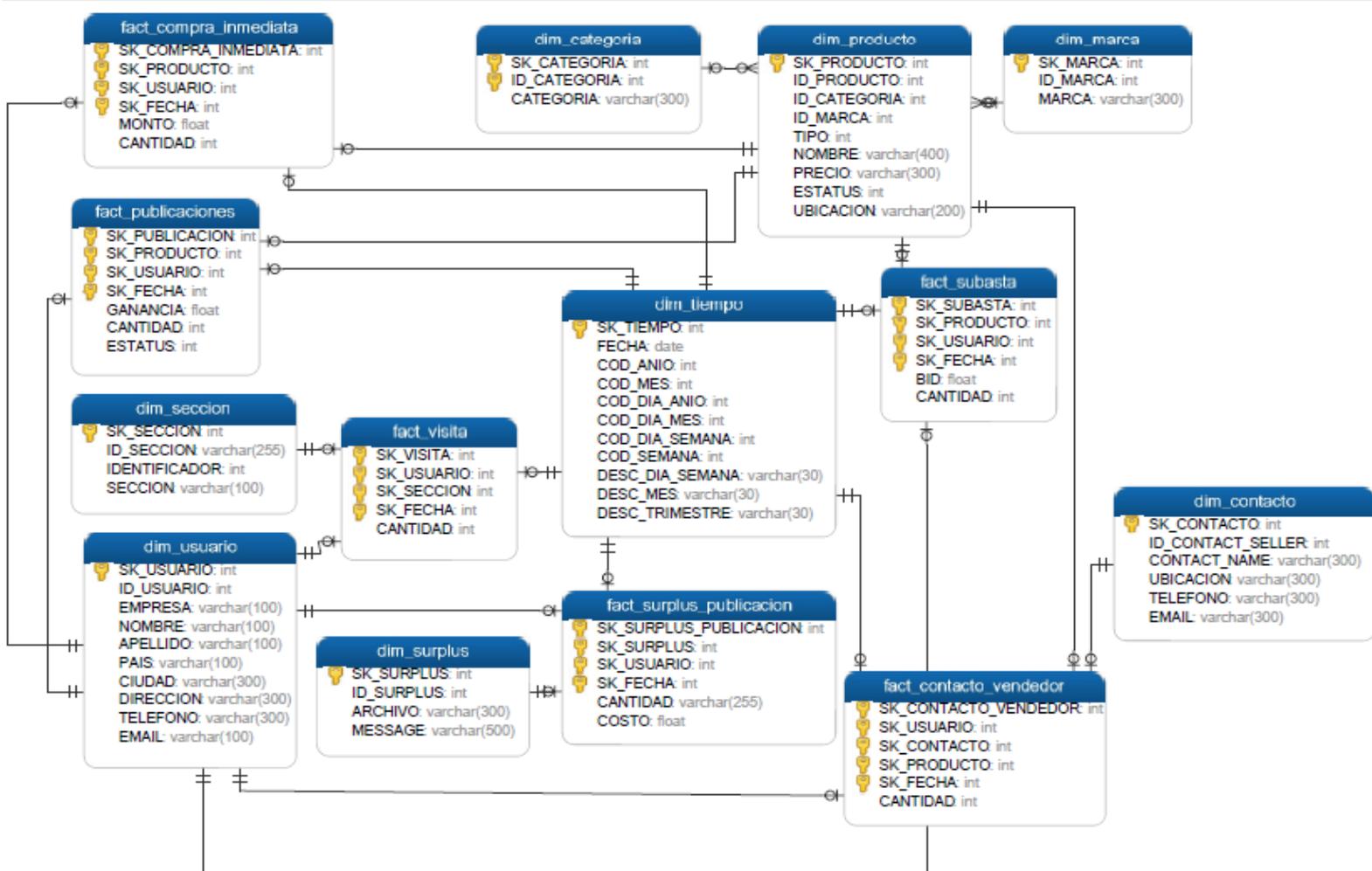


***Ilustración 23. Modelo gráfico de alto nivel de Subastas***

#### 4.1.6 Diseño Físico

Una vez definido el modelo dimensional, utilizando Navicat, el cual es un administrador gráfico de base de datos provista por dicha base de datos, se procede a la construcción de las estructuras físicas (Dimensiones y Tablas de Hechos) del almacén de datos dentro de la base de datos por medio de sentencias SQL. Para ello se definen las siguientes propiedades de la base de datos:

Servidor: localhost	Puerto: 3306
Nombre de la base de datos: Dw_tangerix_mvp	Usuario: Root Password: “ “



**Ilustración 24. Modelo físico del Almacén de Datos implementado en la base de datos**

En la ilustración 24, se observa el modelo físico del almacén de datos ya implementado dentro de la base de datos.

#### **4.1.7 Diseño y Desarrollo de Procesos ETL**

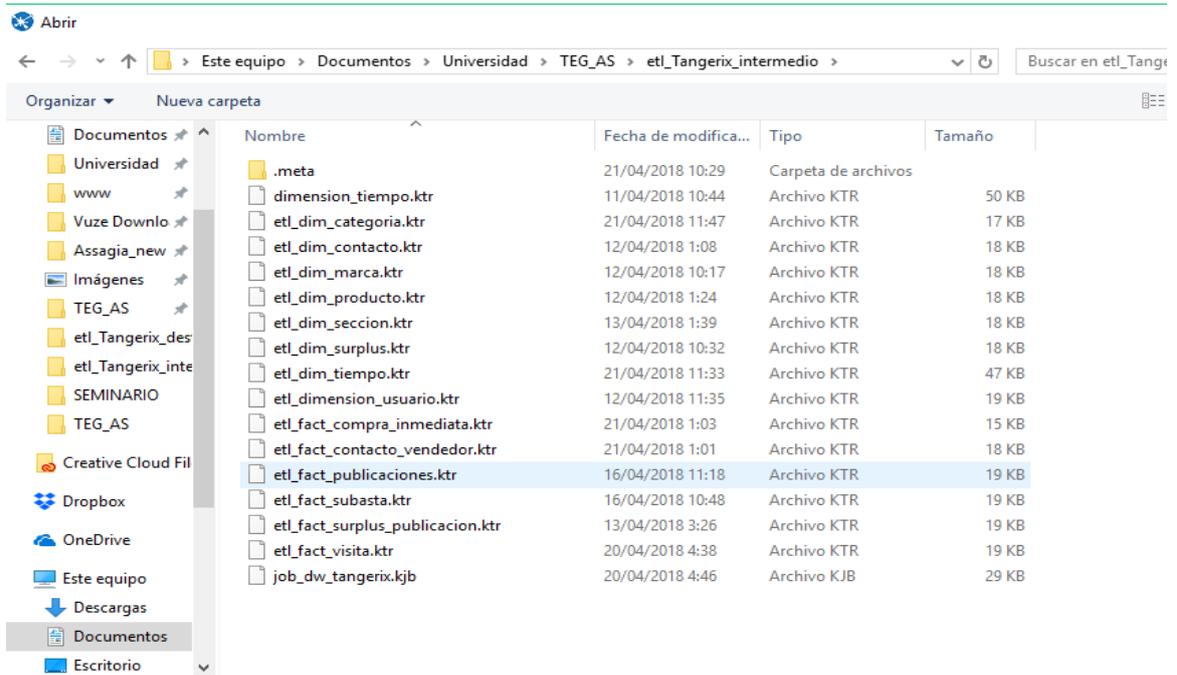
Una vez creado el almacén de datos en físico dentro de la base de datos de MySQL, se comienza a diseñar e implementar los procesos ETL para extraer, transformar y cargar los datos procedentes de las fuentes de datos.

Cabe señalar que los datos usados como fuente de datos son aquellos provenientes de base de datos del portal web que aún está en un ambiente de prueba (ambiente beta), entonces quiere decir que los resultados arrojados por la aplicación son simulados.

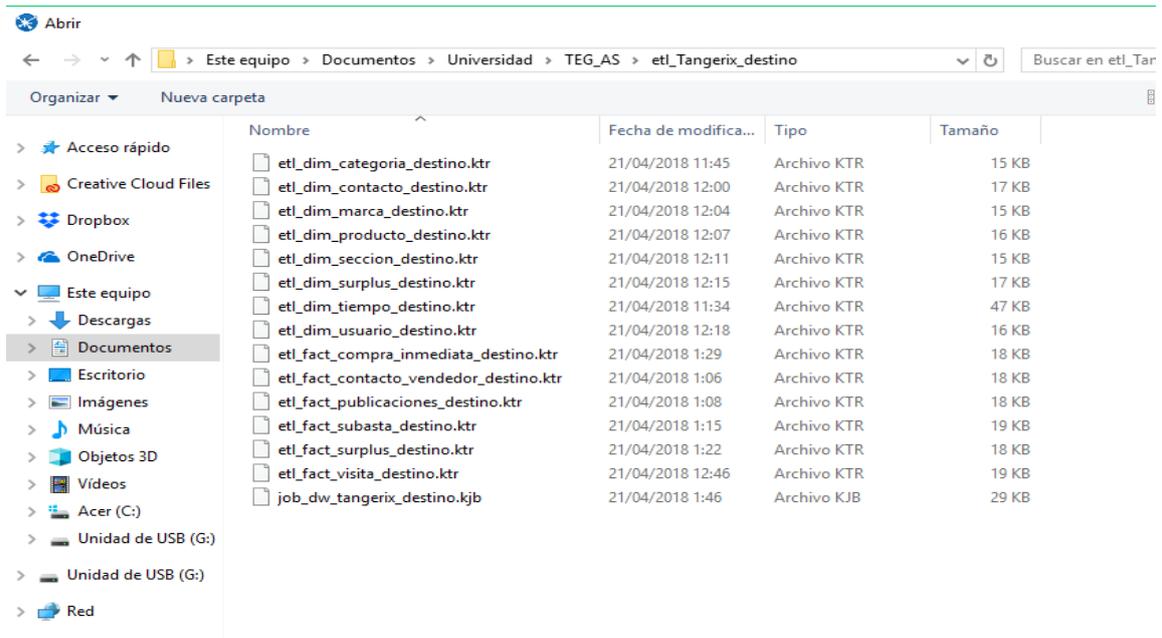
Para el desarrollo de los procesos de ETL, se utilizó la herramienta Pentaho Data Integration (PDI), mencionada previamente en la fase de selección de productos. Esta herramienta permite la creación de las transformaciones (Transformations), conectadas desde las distintas fuentes de datos, donde se realizan consultas para obtener datos y procesarlos, para posteriormente tomar esos resultados e insertarlos o actualizarlos en las correspondientes fuentes de destino (Tablas que constituyen el almacén de datos).

A su vez se crean procesos denominados Jobs, que tienen como función ejecutar un conjunto de transformaciones automáticamente, evitando la ejecución de una por una.

Las transformaciones y trabajos (Jobs) se crean dentro de un repositorio de archivos, tal como se muestra en la Ilustración 25 y 26. Este repositorio es un mismo lugar a través de carpetas, todas las transformaciones y los trabajos (Jobs) desarrollados.

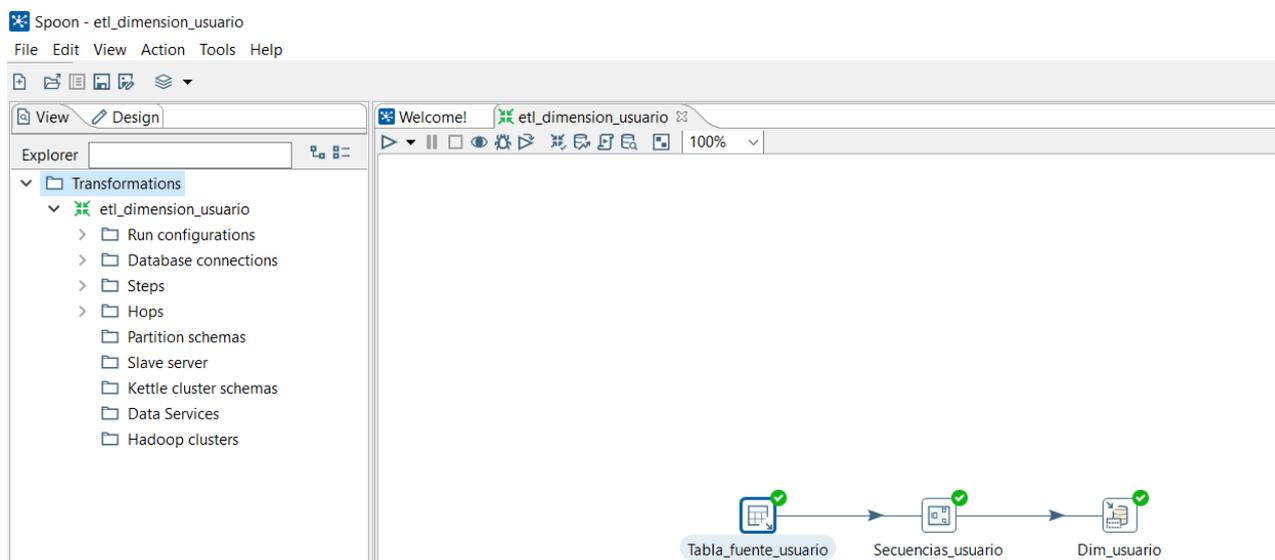


**Ilustración 25. Repositorio de Archivos de la primera fase de ETL**



**Ilustración 26. Repositorio de Archivos de la segunda fase de ETL**

En la ilustración 27 se observa una transformación a una de las tablas que conforman el almacén de datos, para las demás tablas del almacén se desarrollan transformaciones similares a esta.



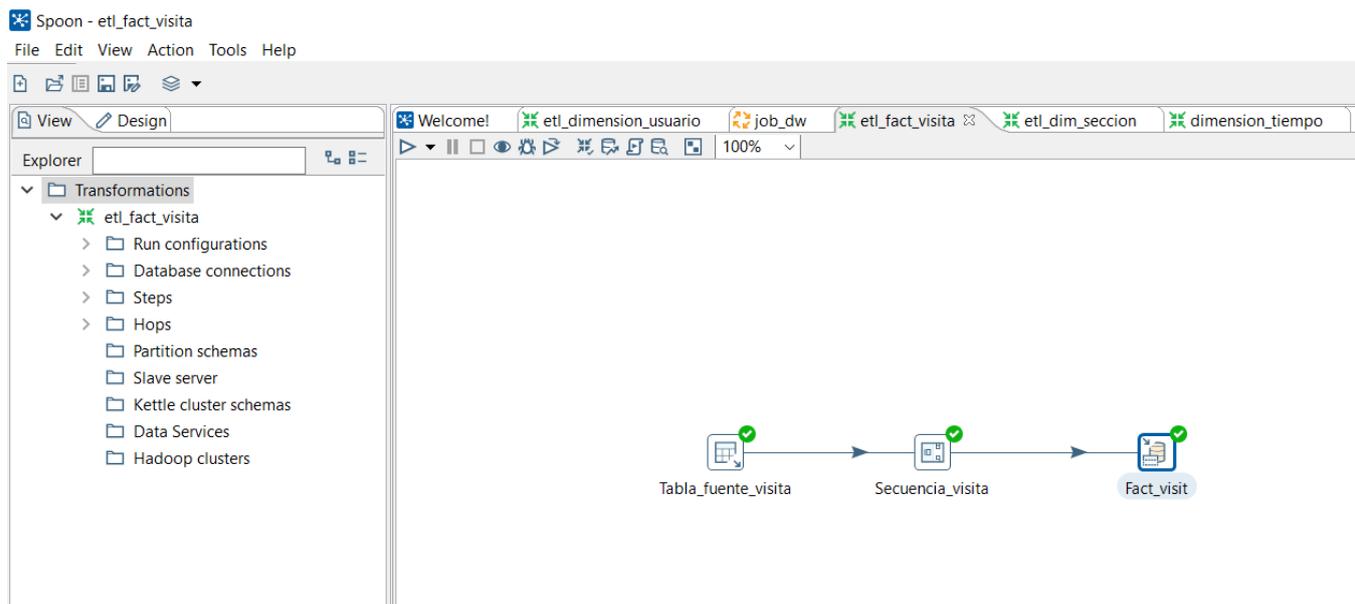
**Ilustración 27. ETL para cargar la dimensión Usuario**

En la ilustración 27, se puede observar que para cargar información en una dimensión del almacén de datos, se empieza realizando una consulta a una tabla fuente, en este caso es la tabla fuente llamada “usuarios” (tabla proveniente de la base de datos del portal web). Luego se procede a crear un paso llamado “secuencias\_usuario” la cual nos arroja como resultados un número incrementable en cada inserción en la dimensión destino, todo eso con la finalidad de que sea el valor de las SK (Clave subrogada) de las dimensiones.

Finalmente se hace una inserción/actualización en la dimensión destino de los datos obtenidos previamente en la consulta del primer paso, en este caso en la dimensión llamada “dim\_usuario” (tabla hecho del almacén de datos).

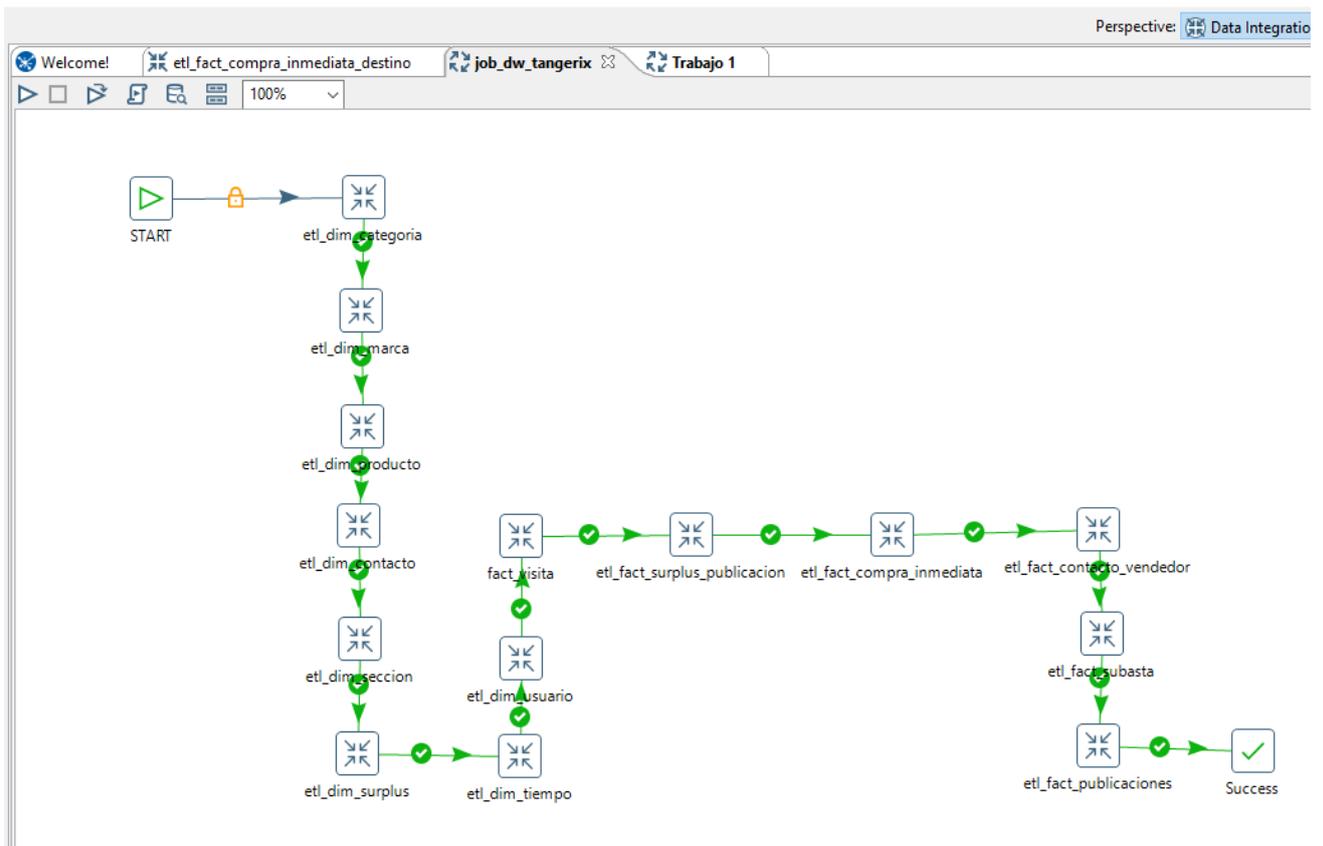
Asimismo, todas las transformaciones que permiten cargar con datos el resto de las tablas que integran al almacén de datos poseen una estructura similar a la de esta transformación.

Aunado a lo mencionado previamente, como se muestra en la ilustración 28, se crea una transformación para poblar una de las tablas de hechos del almacén de datos, en este caso se cargará la información en la tabla llamada “fact\_visita” (tabla hecho del almacén de datos). Puede parecer un proceso similar al anterior, sin embargo puede ser un poco más complejo debido a que se realizan consultas a la fuente de datos (base de datos) y a las dimensiones previamente cargadas, con el objetivo de obtener todos los registros en base a las claves primarias de cada dimensión.



**Ilustración 28. ETL para cargar la Tabla Hecho “fact\_visita”.**

Para evitar la ejecución una a una de las transformaciones previamente desarrolladas, se crea un trabajo (job) para poblar el almacén de datos. Cada trabajo (job) posee una estructura parecida a la reflejada en la ilustración 29:

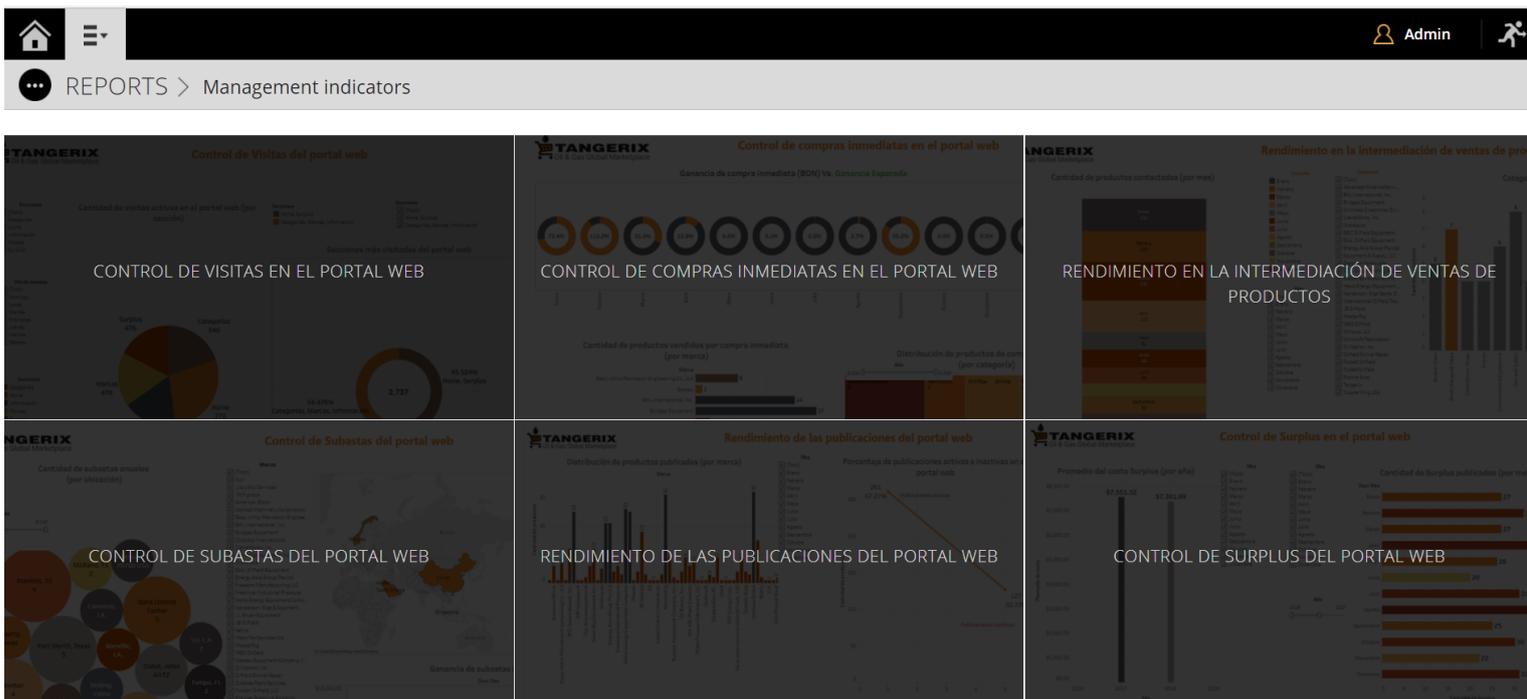


**Ilustración 29. Trabajo (Job) que ejecuta todas las transformaciones.**

En la anterior ilustración, se observa que el job comienza desde un punto de partida (Start) y luego, se ejecutan todas las transformaciones que representan cada uno de los procesos ETL que cargan con datos las tablas del modelo dimensional previamente mencionado. Terminando la ejecución del trabajo (job) con un "Success", que indica que todo el trabajo se ejecutó correctamente.

#### 4.1.8 Especificaciones de las Aplicaciones Analíticas

Los requerimientos del usuario dependen de la necesidad de utilizar herramientas que permitan de manera más fácil obtener información, por ende es necesario crear aplicaciones fáciles de usar, eficientes, eficaces y flexibles para este fin. En la ilustración 33 se puede observar cómo se planteó el conjunto inicial de reportes.



**Ilustración 30. Reportes iniciales del portal web desde el administrador.**

Tomando en consideración lo mencionado previamente, se planteó el desarrollo de un administrador del portal web (aplicación que se encarga de la administración y control del contenido del portal web principal), con el objetivo de que sea el medio por el cual el usuario de la aplicación pueda interactuar con los tableros de control (dashboard) que contiene los indicadores desarrollados en esta solución de Inteligencia de Negocio. Por motivos de seguridad, la aplicación del administrador permitirá que los usuarios puedan acceder a través de un usuario y una clave para poder interactuar con el tablero de control. A continuación se definen un conjunto de pautas y estándares para el fácil uso y entendimiento de la aplicación del portal web administrativo:

**Tabla 13. Pautas y estándares de la aplicación del portal web administrativo.**

<p><b>Autenticación</b></p>	<p>Los usuarios que hacen uso de la aplicación del administrador tienen que autenticarse al momento de acceder a la sección de reportes para interactuar con los tableros de control (dashboard) a través del uso de un usuario y una clave asignada por la organización.</p>
<p><b>Diseño minimalista</b></p>	<p>El portal web administrativo, los reportes, los tableros de control se cargan de manera rápida y sin consumir muchos recursos, agrupando los indicadores por los procesos que se ejecutan en la empresa (compras, intermediación, surplus, etc.)</p>
<p><b>Vista de reportes y indicadores</b></p>	<p>La representación de los reportes e indicadores consta de un conjunto de tablas combinadas con elementos gráficos, como por ejemplo: gráficos barras, gráficos de donas, gráficos de torta, gráficos de línea, etc.</p>
<p><b>Consultas</b></p>	<p>Cada reporte y tablero de control tiene un conjunto de filtros, que cumplen con la función de ayudar al usuario a obtener la información que necesite analizar a través de estos reportes creados.</p>
<p><b>Formato estándar</b></p>	<p>Cada reporte consta con su título, el logotipo de la organización, y el título de cada indicador en sus gráficos.</p>

### 4.1.9 Desarrollo de las Aplicaciones Analíticas

A través de la herramienta de gestión y diseño de bases de datos conocida como “Navicat”, se inicia el desarrollo de la aplicación analítica de la Inteligencia de Negocio propuesta, en la ilustración 34 se puede observar cómo se genera una base de datos con toda la estructura y los datos del almacén de datos implementados anteriormente, sirviendo como base para la obtención de los datos para la implementación de los reportes y tableros de control (dashboard).

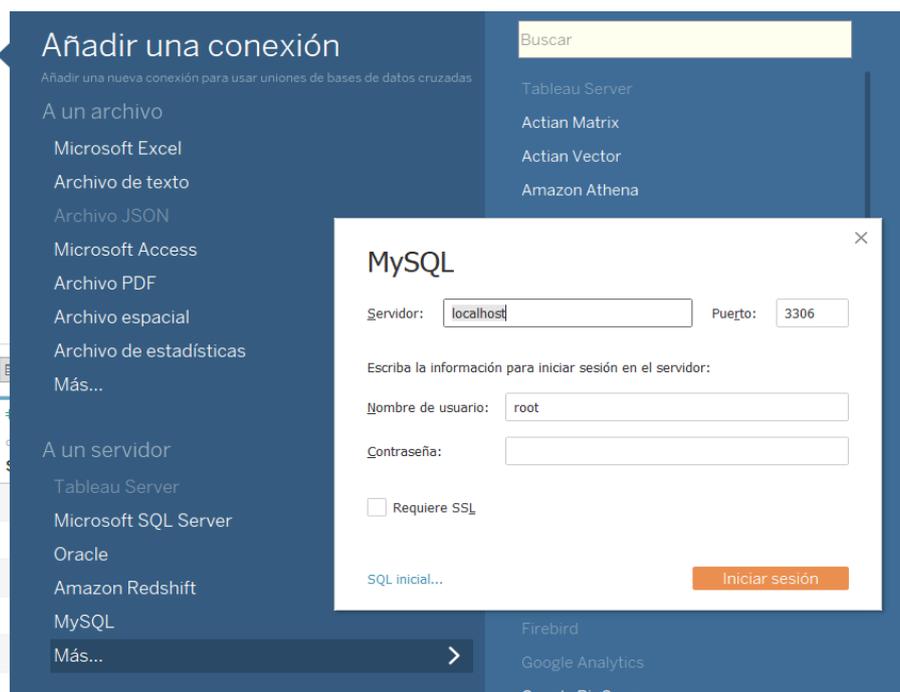
SK_COMPRA_INMEDIATA	SK_PRODUCTO	SK_USUARIO	SK_FECHA	MONTO	CANTIDAD
1	136	39	370	850	1
2	192	28	370	6000	1
3	287	47	371	695	1
4	279	35	373	380	1
5	211	39	377	20000	1
6	241	44	377	8000	1
7	455	33	377	55000	1
8	70	9	380	870000	1
9	192	7	380	6000	1
10	422	48	380	25000	1
11	264	29	383	800	1
12	454	8	384	1000	1
13	265	34	385	7000	1
14	352	39	386	1200	1
15	457	21	398	3200	1
16	27	21	399	2000000	1
17	147	8	399	2000	1
18	356	45	399	1000000	1
19	366	31	399	25000	1
20	256	32	401	150000	1
21	342	44	401	12	1
22	90	18	402	15000	1

**Ilustración 31. Vista de la base de datos desarrollada con Navicat.**

Como se observa en la ilustración 34, una base de datos está conformada por un conjunto de tablas, donde se dividen entre tablas que conforman las dimensiones del modelo dimensional, y otras tablas que conforman las tablas hechos del mismo modelo.

Para la creación de los tableros de control y reportes de indicadores en la aplicación de Tableau conocida como “Tableau Desktop” se deben realizar varios procesos, iniciando por la creación de elementos denominados “Hojas de trabajo”, en donde se realizarán los reportes por cada indicador con sus respectivos filtros, leyendas, etc. y al finalizar cada indicador se crea un dashboard (tablero de control) donde se observa la unificación de estos reportes. Cada uno de estos reportes y tableros de control se almacenan en un elemento denominado “Libro” (donde se almacenan las hojas de trabajos, dashboard, historiales, etc.).

En la siguiente ilustración 35, se muestra que para empezar la creación de estos reportes y tableros de control se necesita crear una conexión a un sistema manejador de base de datos para seleccionar la base de datos a usar, en este caso el la conexión es con MySQL, con el nombre de Usuario “root” y sin contraseña en el servidor de localhost por el puerto 3306.



**Ilustración 32. Conexión a una base de datos desde Tableau.**

Posteriormente a esto, se procede a seleccionar con cual base de datos se trabajará para la creación de los reportes y tableros de control, arrastrando hacia el lado izquierdo aquellas tablas que serán utilizadas en el libro.

Tableau - Libro\_tangerix\_visitas

Archivo Datos Servidor Ventana Ayuda

Conexiones Añadir

localhost  
MySQL

Base de datos

dw\_tangerix\_mvp

Escriba el nombre de la base de datos

- agora
- arlintong
- dw\_intermedia\_tg
- dw\_tangerix\_mvp
- gestion\_compras
- information\_schema
- mysql
- performance\_schema
- prevenauto\_corp
- reservas
- salvage
- tangerix\_mvp
- tangerixbd
- test
- xplore\_venezuela

fact\_surplus\_publicacion

fact\_visita

Nueva personalización de SQL

Nueva unión

fact\_visita+ (dw\_tangerix\_mvp)

fact\_visita

dim\_seccion

dim\_usuario

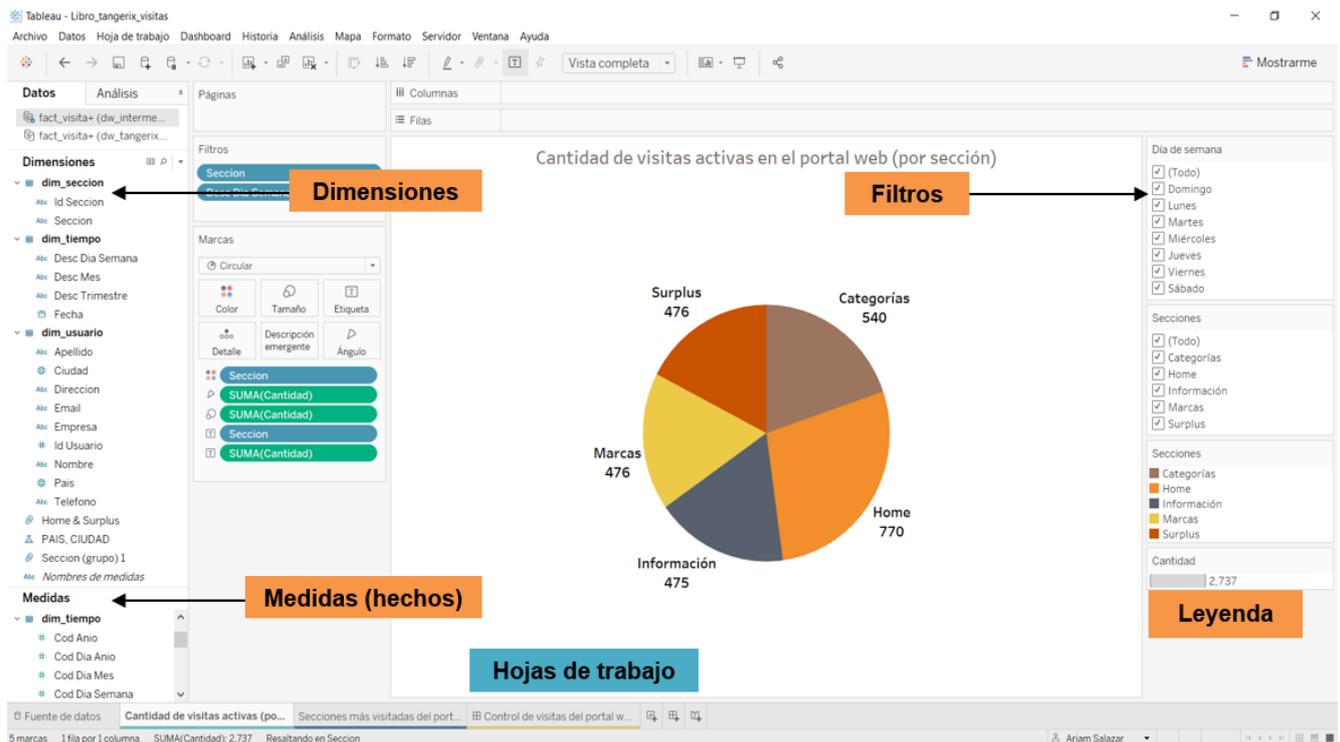
Para combinar tablas de bases de datos distintas, cree una conexión nueva. Sin una conexión nueva, solo podrá unir tablas que se encuentren en la misma base de datos.

**Seleccionar base de datos**

Ordenar campos Orden de fuente de datr

#	Abc	#	Abc	#	
dim_seccion	dim_seccion	dim_seccion	dim_seccion	dim_tiempo	dim_t
SK SECCION (...)	Id Seccion	Identificador	Seccion	Sk Tiempo	Fect
	17	17	0 surplus	384	1/18
	20	20	0 surplus	385	1/19
	21	21	0 surplus	370	1/4/
	32	32	0 surplus	412	2/15
	33	33	0 surplus	368	1/2/
	38	38	0 surplus	390	1/24
	51	51	0 surplus	403	2/6/
	54	54	0 surplus	371	1/5/
	55	55	0 surplus	418	2/21

**Ilustración 33. Selección de la base de datos con que se trabajará en la creación de reportes y tableros de control (dashboard).**

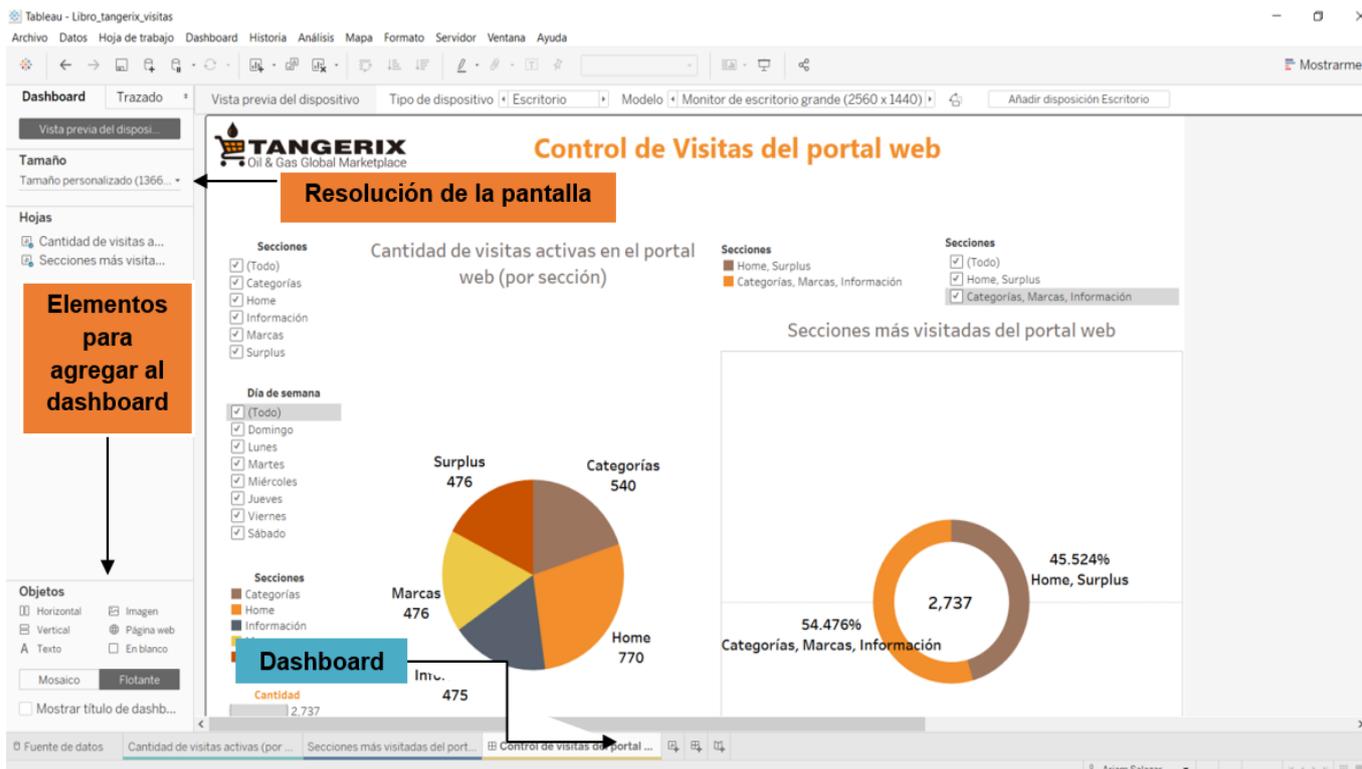


**Ilustración 34. Elementos de la interfaz de Tableau Desktop para la creación de las hojas de trabajo.**

Como se observa en la ilustración 37, para la creación de una hoja de trabajo, se tienen una serie de elementos en la interfaz de la aplicación de Tableau Desktop. En el lado izquierdo se tienen lo que son las dimensiones (que vienen previamente de las dimensiones cargadas de la fuente de datos) y las medidas (que vienen previamente de las tablas hecho cargadas previamente), los elementos del lado derecho son aquellos que se arrastran hacia las filas o a las columnas para la creación de los gráficos.

Una vez cargadas las dimensiones y medidas por las cuales se va a desarrollar el gráfico, se procede a seleccionar cual es el mejor gráfico para representar el indicador a desarrollar (gráficos de barra, gráfico de torta, etc.). De igual manera en esta interfaz se procede a crear los filtros que usarán los usuarios al final en los tableros de control (dashboard), así como se observa de igual manera las leyendas de las gráficas.

Una vez creado cada gráfico por separado en las hojas de trabajo, se crea los tableros de control (dashboard), como se observa en la ilustración 38.



**Ilustración 35. Elementos de la interfaz de Tableau Desktop para la creación de los tableros de control (dashboard).**

Por último, el tablero de control permite la visualización de la información a través de gráficos de barras, gráficos de línea, tablas y tacómetro. En la interfaz de la creación de los tableros de control (dashboard) se tiene una serie de elementos del lado izquierdo que nos permiten la personalización del dashboard, como por ejemplo seleccionar el tamaño de la pantalla para los tableros de control, seleccionar los elementos que serán insertados en el tablero de control (imagen, páginas web, textos, etc.), y a su vez arrastrar aquellos libros previamente creados que quieran ser mostrados en el tablero de control (dashboard).

En las siguientes ilustraciones se presentarán los dashboard creados en cada libro en la aplicación de Tableau Desktop, por cada proceso del portal web.

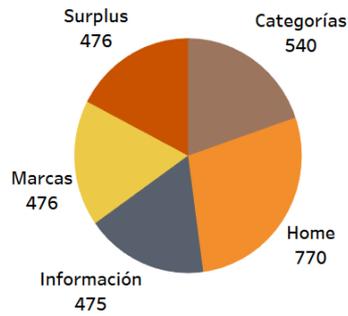
- Secciones**
- (Todo)
  - Categorías
  - Home
  - Información
  - Marcas
  - Surplus

- Día de semana**
- (Todo)
  - Domingo
  - Lunes
  - Martes
  - Miércoles
  - Jueves
  - Viernes
  - Sábado

- Secciones**
- Categorías
  - Home
  - Información
  - Marcas
  - Surplus

**Cantidad**  
2,737

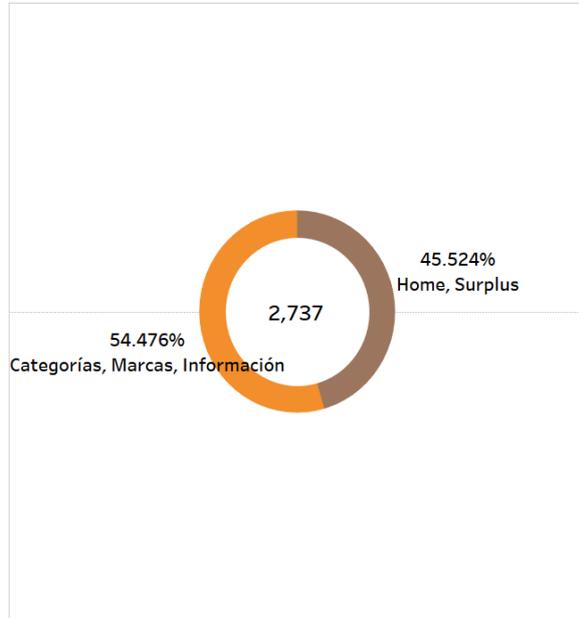
Cantidad de visitas activas en el portal web (por sección)



- Secciones**
- Home, Surplus
  - Categorías, Marcas, Información

- Secciones**
- (Todo)
  - Home, Surplus
  - Categorías, Marcas, Información

Secciones más visitadas del portal web



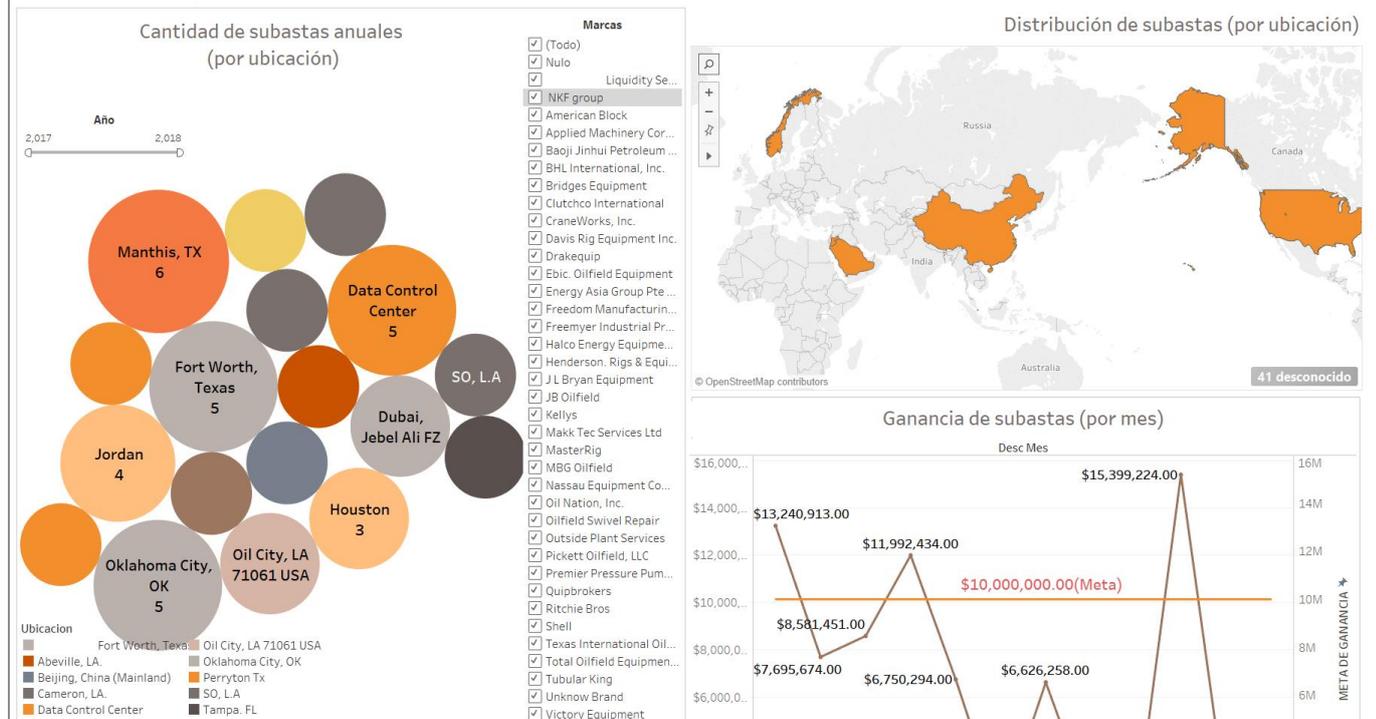
**Ilustración 36. Tablero de Control de visitas del portal web.**

## Control de Surplus en el portal web



**Ilustración 37. Tablero de Control de Surplus del portal web.**

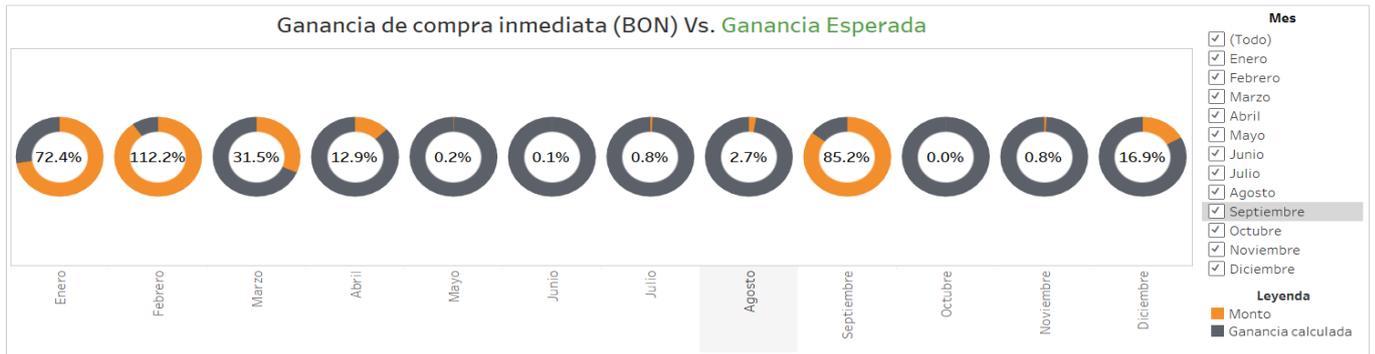
## Control de Subastas del portal web



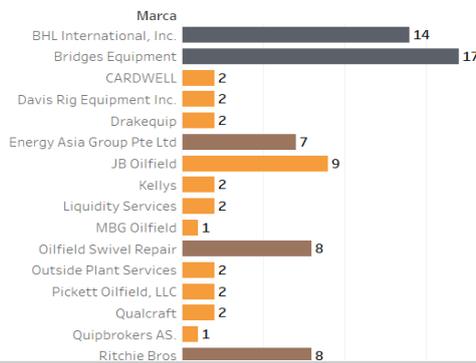
**Ilustración 38. Tablero de Control de Subastas del portal web**



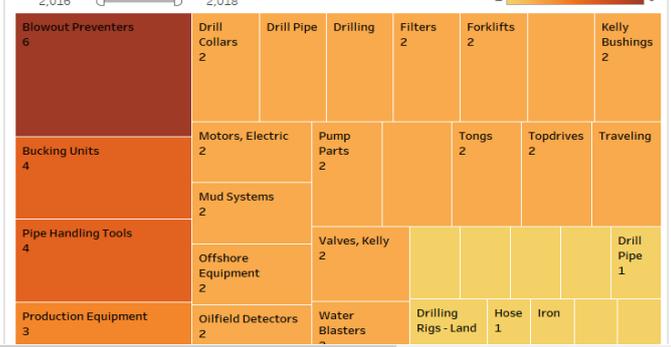
## Control de compras inmediatas en el portal web



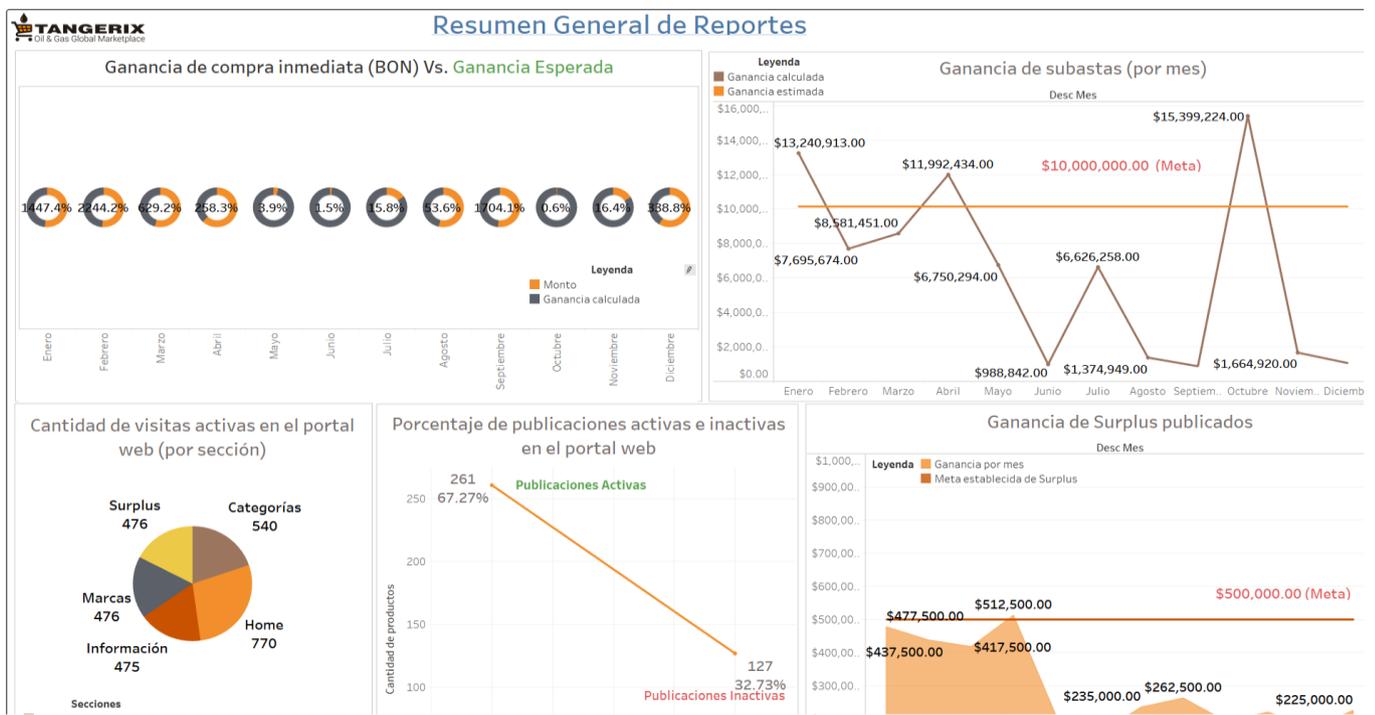
### Cantidad de productos vendidos por compra inmediata (por marca)



### Distribución de productos de compra inmediata (por categoría)

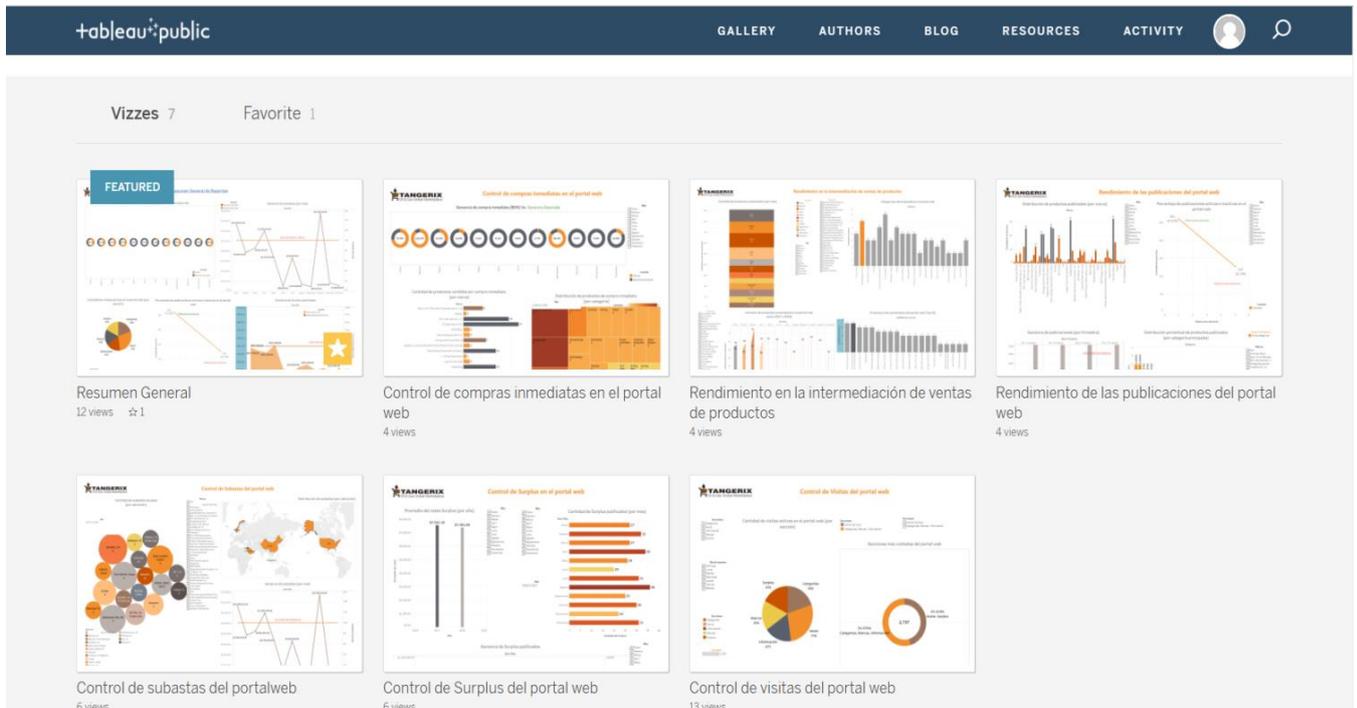


**Ilustración 41. Tablero de Control de Intermediación de Ventas del portal web.**



**Ilustración 42. Tablero de Control de Resumen general de los reportes.**

Al finalizar la creación de los tableros de control (dashboard) Tableau Desktop permite publicar estos libros en una nube dentro de la aplicación denominada “Tableau Public”, lo cual se define como un software gratuito en donde los usuarios puedan ver de manera online sus libros creados previamente en la aplicación de Tableau Desktop, a su vez nos permite compartir cada libro con otros usuarios interesados en ver estos reportes y tableros de control.



**Ilustración 43. Tableau Public.**

Como se puede observar en la ilustración 46, se tiene una serie de libros creados desde la aplicación de Tableau Desktop, cabe destacar que existe la posibilidad de marcar como favoritos o destacados los libros que el usuario desee.

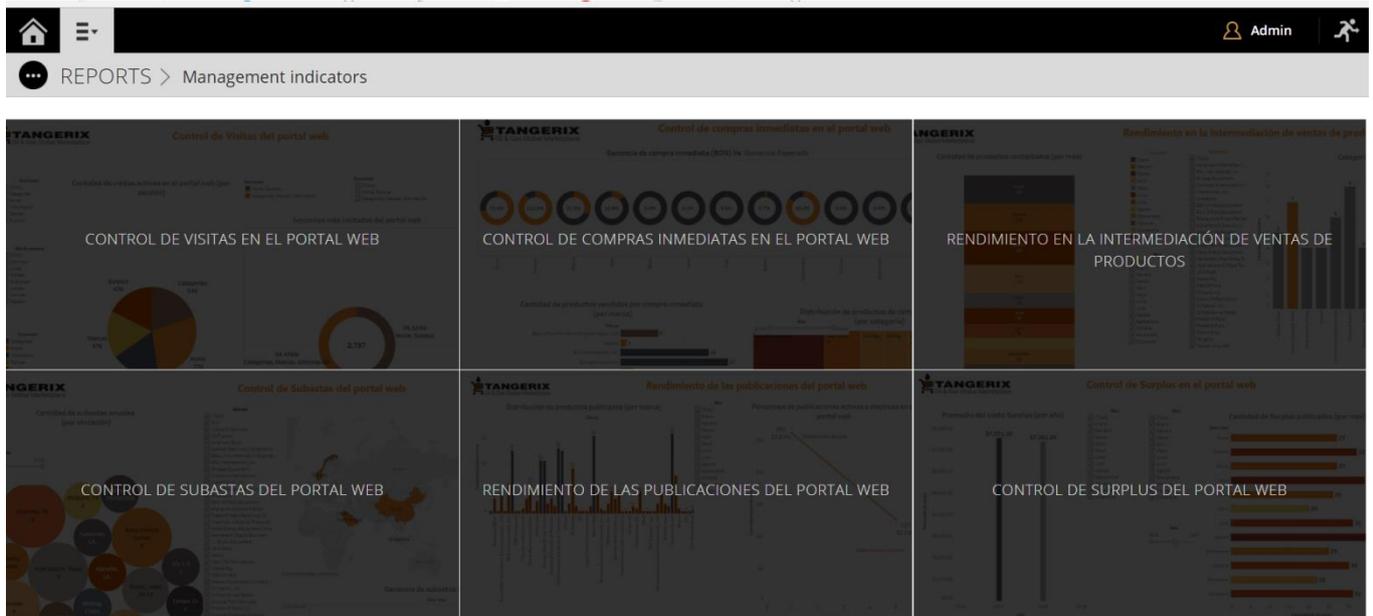
Finalmente, el acceso al conjunto de reportes, al tablero de control se puede realizar como se muestra en la ilustración 47 a través de un portal web administrativo.



**Ilustración 44. Página Inicio del portal web administrativo para acceder a los tableros de control, y a los reportes.**



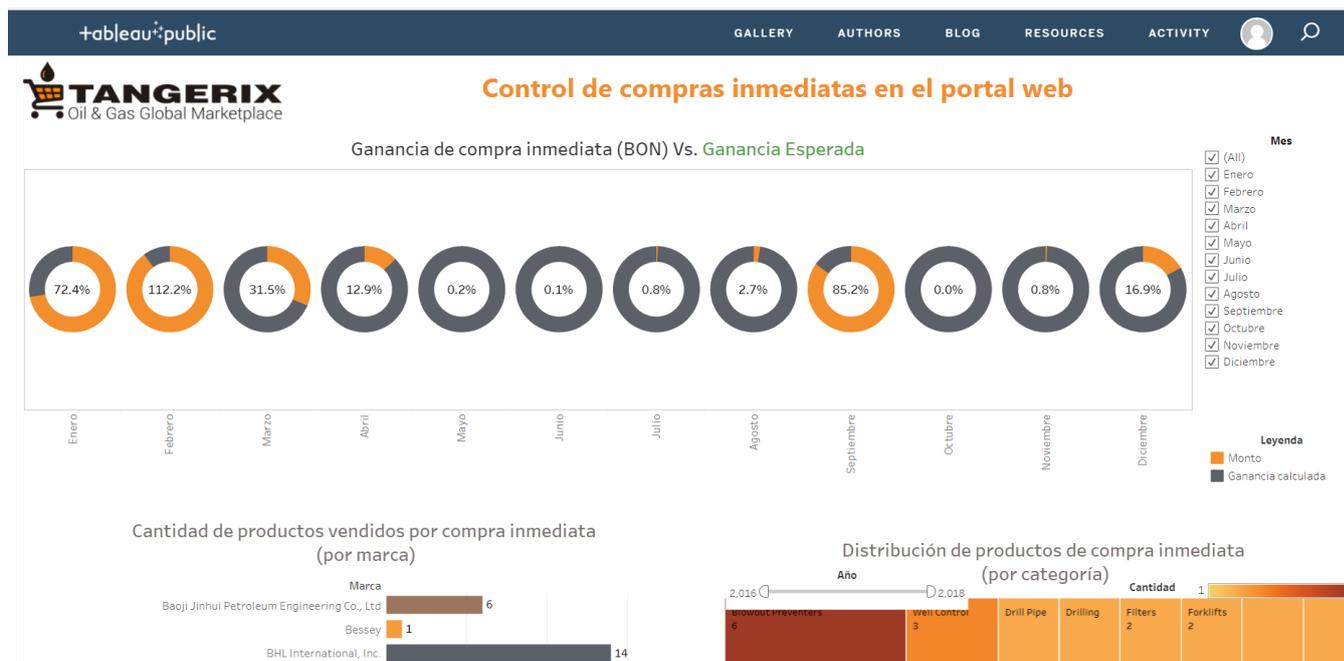
**Ilustración 45. Sección del portal web administrativo que muestra el resumen de los reportes más destacados del portal web principal.**



**Ilustración 46. Sección del portal web administrativo para acceder a los tableros de control por procesos del portal web principal.**



**Ilustración 47. Vista interna de cada tablero de control (dashboard). En este caso del tablero de control de Rendimiento de las publicaciones del portal web.**



**Ilustración 48. Vista interna de cada tablero de control (por libro) desde Tableau Public.**

#### 4.1.10 Implementación

El proceso de implementación lleva varios procesos, el primero se refiere a toda la instalación y configuración de los elementos en el ambiente de producción. Normalmente por motivos de seguridad y de calidad, el desarrollo de las aplicaciones, el almacén de datos y los procesos de ETL no se realizan directamente en la organización, sino que se ejecutan a partir de un conjunto de datos dados por la organización en un ambiente de ejecución distinto, en este caso sería en un ambiente de prueba.

En el caso de esta investigación, en el ambiente de producción de instalaron todas las herramientas mencionadas previamente en el punto 4.1.4 y posteriormente se ejecutaron los procesos ETL para la carga de los datos en el almacén de datos, luego de eso, se extrajeron los datos de la base de datos fuente para desarrollar los reportes y tableros de control, obteniendo de esta manera la solución de Inteligencia de Negocio ya implementada en el ambiente de producción, finalmente se realizaron pruebas para verificar que la implementación se ejecute de manera correcta y eficiente.

#### 4.1.10.1 Pruebas de Calidad

Después de haber implementado el almacén de datos, se deben realizar un conjunto de pruebas de calidad con el objetivo de evaluar el rendimiento del sistema, es decir, que los datos están siendo cargados de forma correcta y así verificar que los datos contenidos en el almacén de datos corresponden con los que se encuentran en el modelo dimensional.

El plan de calidad consiste en verificar que los datos que fueron migrados correctamente desde la fuente de datos al almacén de datos de final por los procesos ETL's comprobando que se obtuvo una información de buena calidad. El primer paso de este plan de calidad es la verificación de los datos de la fuente de datos realizando una consulta en la fuente de datos origen para poder observar los registros que nos retorna la misma, como se puede observar con un ejemplo de ETL para cargar la dimensión usuario en la ilustración 52:

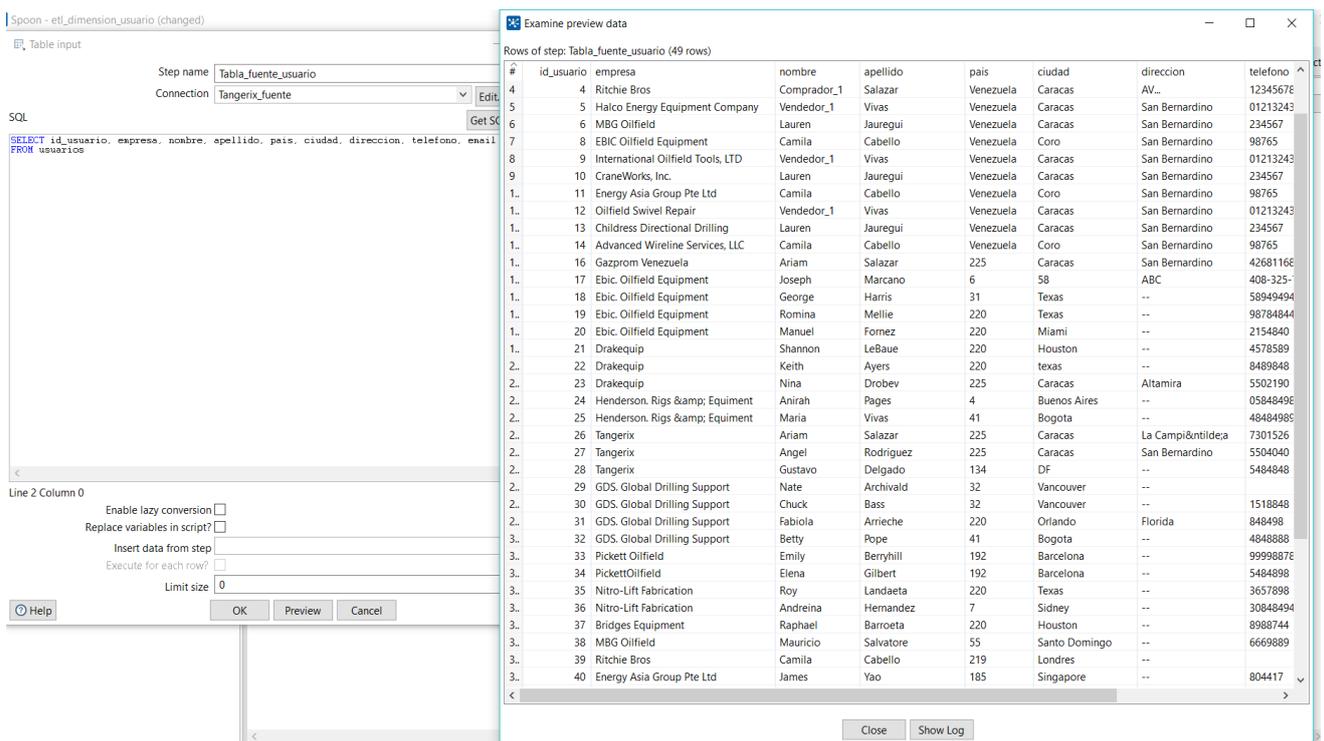


Ilustración 49. Consulta de Base de datos fuente de la dimensión Usuario

Una vez ejecutado la consulta en la fuente de datos, se procede a realizar la misma consulta en el almacén de datos destino del proceso ETL, como se observa en la ilustración 53:

The screenshot shows a SQL query editor with the following query: `SELECT id_usuario, empresa, nombre, apellido, pais, ciudad, direccion, telefono, email FROM dim_usuario;`

The results are displayed in a table with the following columns: id\_usuario, empresa, nombre, apellido, pais, ciudad, direccion, telefono, email.

id_usuario	empresa	nombre	apellido	pais	ciudad	direccion	telefono	email
3	Equipment & Supply, LLC	Matias	Rodriguez	Venezuela	Caracas	San Bernardino	02125504040	cimadelmundo@hotmail.co
4	Ritchie Bros	Comprador_1	Salazar	Venezuela	Caracas	AV...	123456789	comprador@gmail.com
5	Halco Energy Equipment Co	Vendedor_1	Vivas	Venezuela	Caracas	San Bernardino	012132435	vendedor@gmail.com
6	MBG Oilfield	Lauren	Jauregui	Venezuela	Caracas	San Bernardino	234567	lauren@gmail.com
8	EBIC Oilfield Equipment	Camila	Cabello	Venezuela	Coro	San Bernardino	98765	cam@gmail.com
9	International Oilfield Tools,	Vendedor_1	Vivas	Venezuela	Caracas	San Bernardino	012132435	vendedor@gmail.com
10	CraneWorks, Inc.	Lauren	Jauregui	Venezuela	Caracas	San Bernardino	234567	lauren@gmail.com
11	Energy Asia Group Pte Ltd	Camila	Cabello	Venezuela	Coro	San Bernardino	98765	cam@gmail.com
12	Oilfield Swivel Repair	Vendedor_1	Vivas	Venezuela	Caracas	San Bernardino	012132435	vendedor@gmail.com
13	Childress Directional Drilling	Lauren	Jauregui	Venezuela	Caracas	San Bernardino	234567	lauren@gmail.com
14	Advanced Wireline Services,	Camila	Cabello	Venezuela	Coro	San Bernardino	98765	cam@gmail.com
16	Gazprom Venezuela	Ariam	Salazar	225	Caracas	San Bernardino	4268116869	cimadelmundo3000@gmail.
17	Ebic. Oilfield Equipment	Joseph	Marcano	6	58	ABC	408-325-7669	joseph@edicoilfield.com
18	Ebic. Oilfield Equipment	George	Harris	31	Texas	--	5894949484	george@ebicoilfield.com
19	Ebic. Oilfield Equipment	Romina	Mellie	220	Texas	--	987848441	romina@gmail.com
20	Ebic. Oilfield Equipment	Manuel	Fornez	220	Miami	--	2154840	manuel@hotmail.com
21	Drakequip	Shannon	LeBaue	220	Houston	--	4578589	shannon@drakequip.com
22	Drakequip	Keith	Ayers	220	texas	--	8489848	keith@gmail.com
23	Drakequip	Nina	Drobev	225	Caracas	Altamira	5502190	nina@yahoo.com
24	Henderson. Rigs & Equ	Anirah	Pages	4	Buenos Aires	--	058484984	apares@marketing.com
25	Henderson. Rigs & Equ	Maria	Vivas	41	Bogota	--	484849898	mariavivas@gmail.com
26	Tangerix	Ariam	Salazar	225	Caracas	La Campi&ntilde;a	7301526	a.salazar@tangerix.com

**Ilustración 50. Consulta del Almacén de datos de la dimensión usuario**

Se procede a la confirmación de carga de la fuente de datos a las tablas hechos en el almacén de datos, el proceso es similar al anterior, se realiza una consulta tanto en la fuente de datos como en el almacén de datos para verificar que se cargo la información de manera exitosa y correcta.

Spoon - etl\_fact\_publicaciones

Table input

Step name: Tabla\_fuente\_publicacion  
 Connection: tangerix\_fuente

SQL

```

SELECT dw_intermedia_tg.dim_producto.sk_producto as SK_PRODUCTO,
dw_intermedia_tg.dim_usuario.sk_usuario as SK_USUARIO,
dw_intermedia_tg.dim_tiempo.sk_tiempo as SK_FECHA,
tangerix_mvp.compras.total_compra as GANANCIA
FROM tangerix_mvp.compras
JOIN dw_intermedia_tg.dim_producto
ON tangerix_mvp.compras.id_producto=dw_intermedia_tg.dim_producto.id_producto
JOIN dw_intermedia_tg.dim_usuario
ON tangerix_mvp.compras.id_usuario=dw_intermedia_tg.dim_usuario.id_usuario
JOIN dw_intermedia_tg.dim_tiempo
ON tangerix_mvp.compras.fecha=dw_intermedia_tg.dim_tiempo.fecha
  
```

Line 1 Column 0

Enable lazy conversion   
 Replace variables in script?   
 Insert data from step   
 Execute for each row?   
 Limit size: 0

OK Preview Cancel

Examine preview data

Rows of step: Tabla\_fuente\_publicacion (464 rows)

#	SK_PRODUCTO	SK_USUARIO	SK_FECHA	GANANCIA
1	34	3	377	550.0
2	257	21	377	550.0
3	299	6	377	550.0
4	356	13	377	1750.0
5	276	6	378	550.0
6	252	5	379	1350.0
7	314	42	380	1350.0
8	247	48	382	1350.0
9	97	35	385	1250.0
1.	378	3	385	1950.0
1.	158	43	386	1850.0
1.	384	39	387	1750.0
1.	219	14	388	1350.0
1.	113	4	390	1850.0
1.	244	9	390	650.0
1.	29	22	391	350.0
1.	442	11	391	1450.0
1.	46	11	392	350.0
1.	210	3	393	550.0
2.	70	3	394	1780.0
2.	365	19	395	1350.0
2.	143	27	398	550.0
2.	343	41	398	550.0
2.	462	6	398	1750.0
2.	53	22	399	550.0
2.	290	38	400	1850.0
2.	292	18	401	550.0
2.	341	27	402	550.0
2.	437	13	403	480.0
3.	416	3	405	1350.0
3.	189	24	413	350.0

**Ilustración 51. Consulta de Base de datos fuente de las publicaciones**

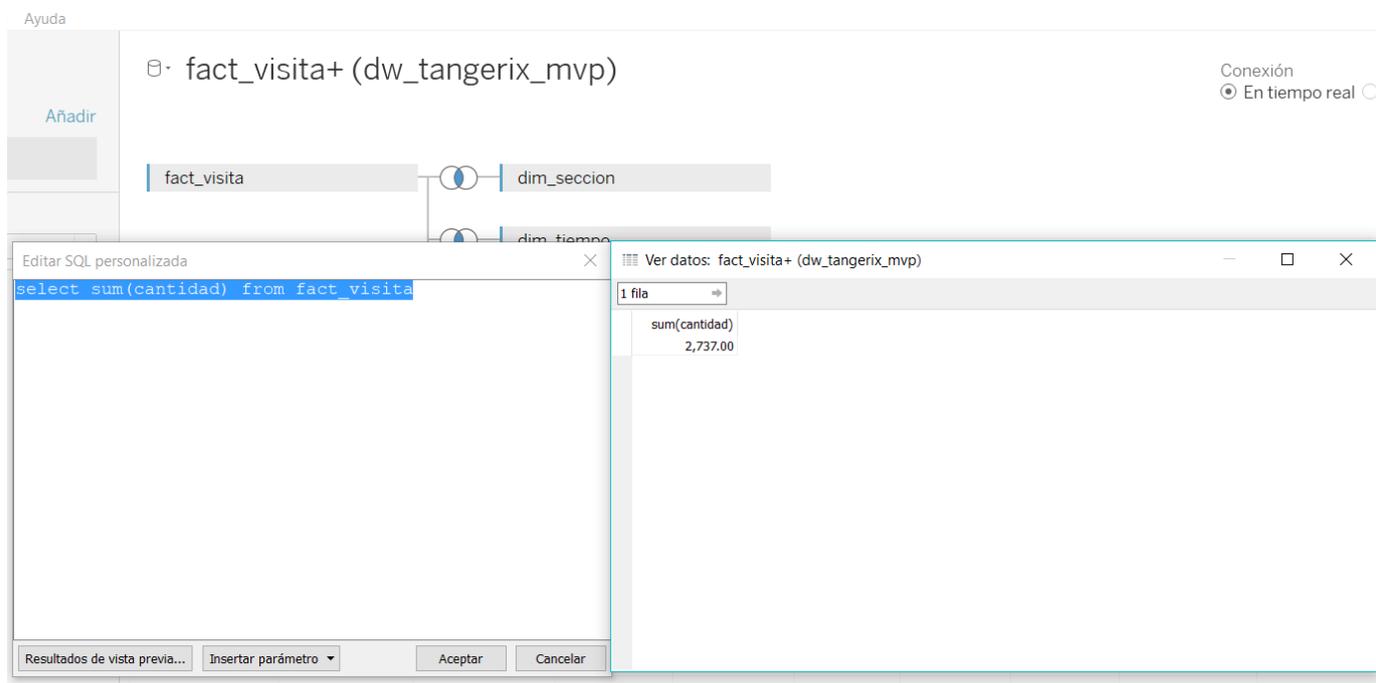
mysql> SELECT dw\_intermedia\_tg.dim\_producto.sk\_producto as SK\_PRODUCTO,  
 dw\_intermedia\_tg.dim\_usuario.sk\_usuario as SK\_USUARIO,  
 dw\_intermedia\_tg.dim\_tiempo.sk\_tiempo as SK\_FECHA,  
 tangerix\_mvp.compras.total\_compra as GANANCIA  
 FROM tangerix\_mvp.compras  
 JOIN dw\_intermedia\_tg.dim\_producto  
 ON tangerix\_mvp.compras.id\_producto=dw\_intermedia\_tg.dim\_producto.id\_producto  
 JOIN dw\_intermedia\_tg.dim\_usuario  
 ON tangerix\_mvp.compras.id\_usuario=dw\_intermedia\_tg.dim\_usuario.id\_usuario  
 JOIN dw\_intermedia\_tg.dim\_tiempo  
 ON tangerix\_mvp.compras.fecha=dw\_intermedia\_tg.dim\_tiempo.fecha;

SK_PRODUCTO	SK_USUARIO	SK_FECHA	GANANCIA
34	3	377	550
257	21	377	550
299	6	377	550
356	13	377	1750
276	6	378	550
252	5	379	1350
314	42	380	1350
247	48	382	1350
97	35	385	1250
378	3	385	1950
158	43	386	1850
384	39	387	1750
219	14	388	1350
113	4	390	1850
244	9	390	650
29	22	391	350
442	11	391	1450
46	11	392	350
210	3	393	550
70	3	394	1780
365	19	395	1350
143	27	398	550
343	41	398	550
462	6	398	1750
53	22	399	550

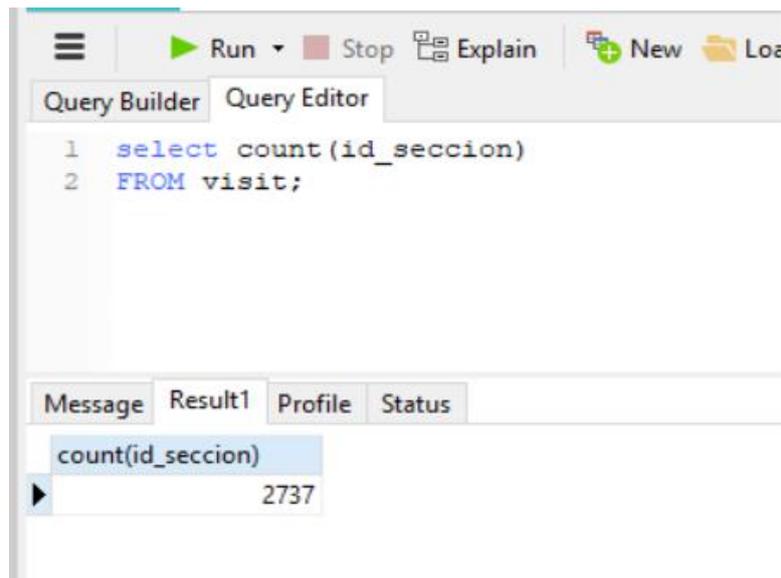
**Ilustración 52. Consulta del almacén de datos de la tabla hecho publicaciones**

Las demás comprobaciones se ejecutan de manera similar a las anteriores, con el objetivo de verificar que todas las dimensiones del almacén de datos se carguen con la información correcta obtenida de la fuente de datos del portal web. Posteriormente se realizan consultas para confirmar las coincidencias entre las cantidades de registros entre la fuente de datos y el almacén de datos final, obteniendo como resultado:

1. Aplicar una consulta para verificar la coincidencia en la cantidad de visitas activas en el portal web entre la fuente de datos y la tabla hecho fact\_visita del almacén de datos. (Ilustración 56 y 57).

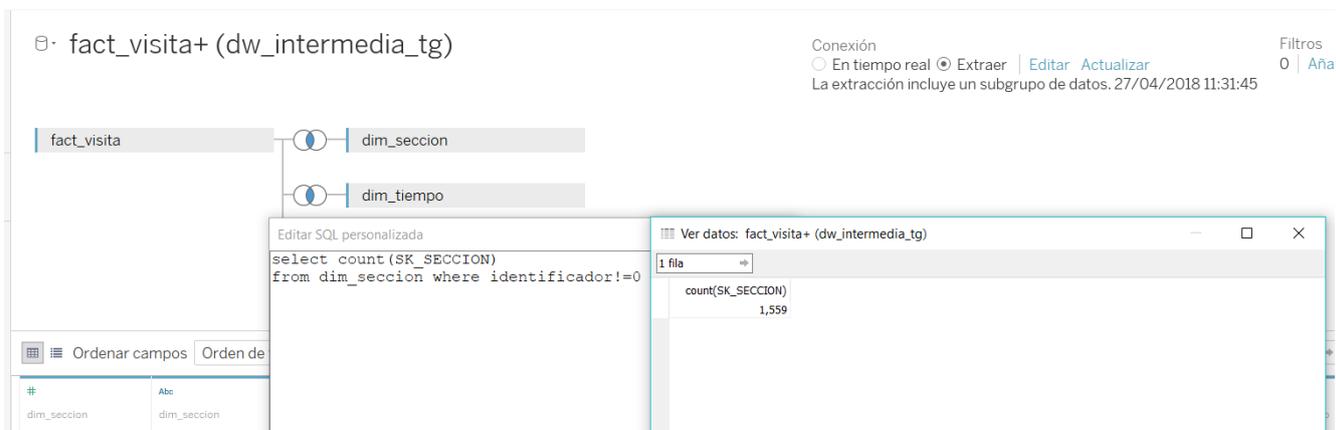


**Ilustración 53. Cantidad de visitas activas en el portal web (desde Almacén de datos).**

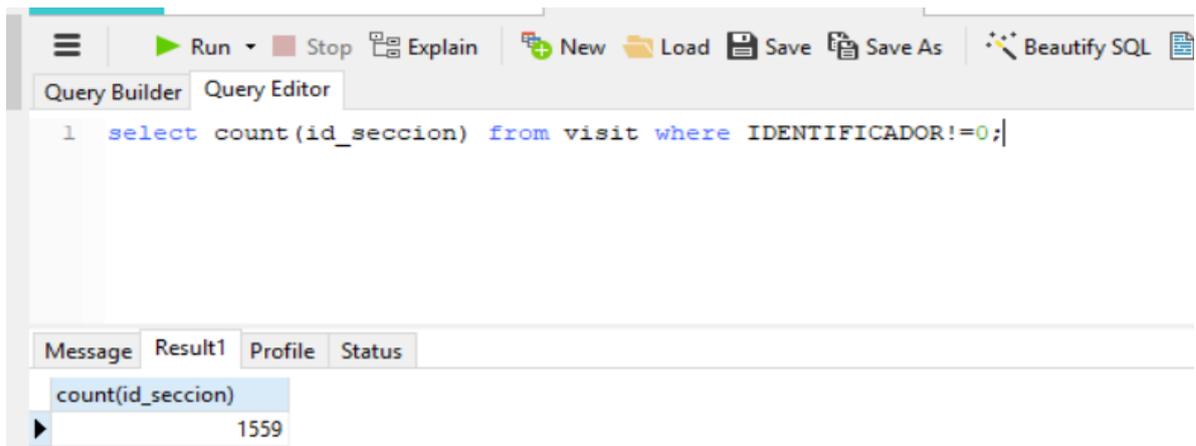


**Ilustración 54. Cantidad de visitas activas en el portal web (desde Fuente de datos).**

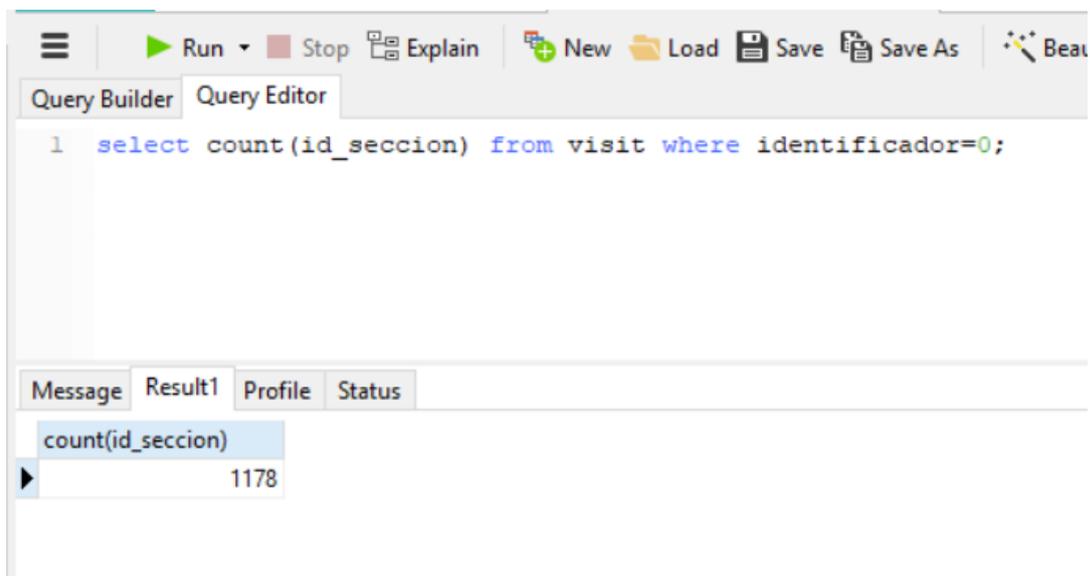
2. Aplicar una consulta para verificar la coincidencia en la cantidad de las secciones más visitadas del portal web entre la fuente de datos y la tabla hecho fact\_visita del almacén de datos. (Ilustración 58, 59, 60, 61).



**Ilustración 55. Cantidad de visitas por Sección en el portal web (desde Almacén de datos), en este caso son la cantidad de visitas en las secciones de Marcas, Categorías, Información.**



**Ilustración 56. Cantidad de visitas por Sección en el portal web (desde fuente de datos), en este caso son la cantidad de visitas en las secciones de Marcas, Categorías, Información.**



**Ilustración 57. Cantidad de visitas por Sección en el portal web (desde fuente de datos), en este caso son la cantidad de visitas en las secciones de Home y Surplus.**

fact\_visita+ (dw\_tangerix\_mv)

Conexión  
 En tiempo real  Extraer

Filtros  
0 | Añadir

The screenshot shows a data tool interface. On the left, there is a diagram with three tables: 'fact\_visita', 'dim\_seccion', and 'dim\_tiempo'. The 'fact\_visita' table is connected to both 'dim\_seccion' and 'dim\_tiempo'. In the center, a window titled 'Editar SQL personalizada' contains the following SQL query: 

```
select count(SK_SECCION)
from dim_seccion where identificador=0
```

 On the right, a window titled 'Ver datos: fact\_visita+ (dw\_tangerix\_mv)' displays the results of the query in a table with one row: 

count(SK_SECCION)
1,178

**Ilustración 58. Cantidad de visitas por Sección en el portal web (desde el almacén de datos), en este caso son la cantidad de visitas en las secciones de Home y Surplus.**

3. Aplicar una consulta para verificar la coincidencia entre cantidad de las publicaciones activas e inactivas (productos publicados) del portal web entre la fuente de datos y la tabla hecho fact\_publicaciones del almacén de datos. (Ilustración 61, 62, 63,64).

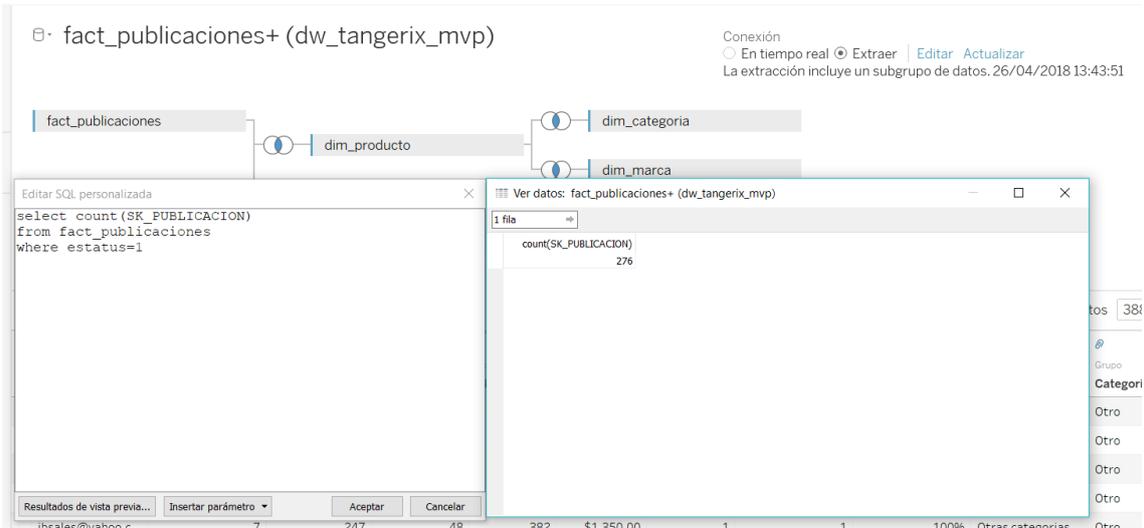
The screenshot shows a query editor interface. The top toolbar includes buttons for 'Run', 'Stop', 'Explain', 'New', 'Load', 'Save', 'Save As', and 'Beautify SQL'. Below the toolbar, the 'Query Editor' tab is active, showing the following SQL query: 

```
1 select count(id_producto) from productos where estatus=1;
```

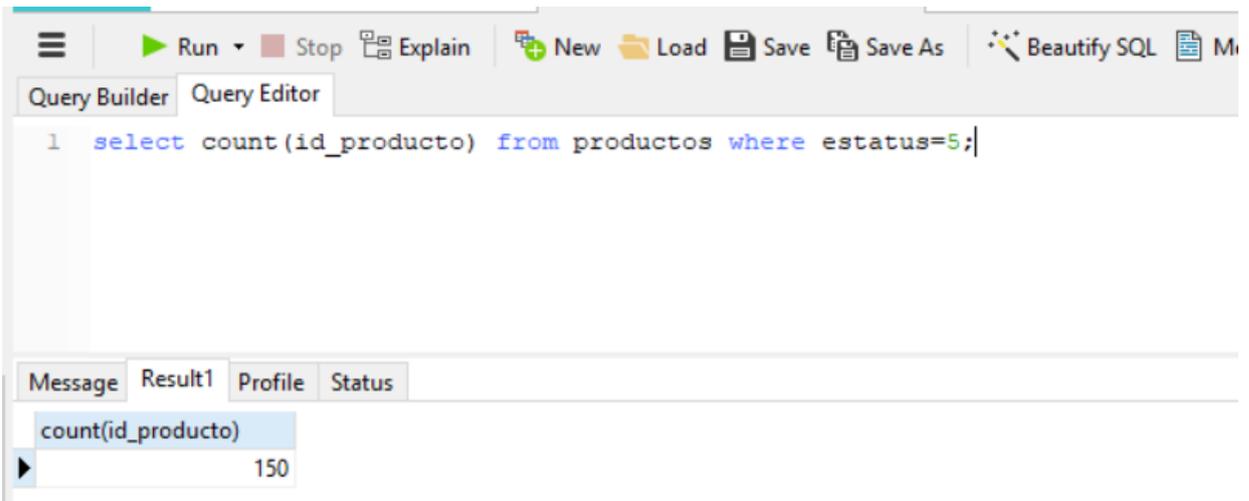
 At the bottom, there is a 'Result1' tab showing the query result in a table: 

count(id_producto)
276

**Ilustración 59. Cantidad de publicaciones (productos activos) activas en el portal web (desde fuente de datos).**



**Ilustración 60. Cantidad de publicaciones (productos activos) activas en el portal web (desde almacén de datos).**



**Ilustración 61. Cantidad de publicaciones (productos inactivos) inactivos en el portal web (desde fuente de datos).**

fact\_publicaciones+ (dw\_tangerix\_mvp)

Conexión  
 En tiempo real  Extraer | [Editar](#) [Actualizar](#)  
La extracción incluye un subgrupo de datos. 26/04/2018 13:43:51

The screenshot shows a data tool interface. At the top, a fact table 'fact\_publicaciones' is connected to three dimension tables: 'dim\_producto', 'dim\_categoria', and 'dim\_marca'. Below this, there is a 'Editar SQL personalizada' (Edit custom SQL) window with the following query:

```
select sum(cantidad)
from fact_publicaciones
where ESTATUS=5
```

To the right of the SQL editor is a 'Ver datos: fact\_publicaciones+ (dw\_tangerix\_mvp)' (View data) window showing the results of the query:

sum(cantidad)
150.000

At the bottom left, there is a table with columns 'Email' and 'Sk Publicacion'.

Email	Sk Publicacion
nina@yahoo.com	
cam@gmail.com	
vendedor@gmail...	
ricardo@yahoo.c...	
jbsales@yahoo.c...	

**Ilustración 62. Cantidad de publicaciones (productos inactivos) inactivos en el portal web (desde almacén de datos).**

4. Aplicar una consulta para verificar la coincidencia entre la distribución de productos publicados por marca en portal web entre la fuente de datos y la tabla dimensión dim\_producto del almacén de datos. (Ilustración 66,67).

Query Editor

```

1 SELECT count(*), p.id_marca
2 FROM productos as p
3 GROUP BY(p.id_marca)
4 ORDER BY p.id_marca ASC;

```

Message Result1 Profile Status

count(*)	id_marca
8	0
43	1
8	3
27	4
5	5
1	6
27	7
3	8
12	9
6	10
16	11
32	13
24	14

**Ilustración 63. Cantidad de productos publicados por marca (desde la fuente de datos).**

Editar SQL personalizada

```

SELECT count(*), p.id_marca
FROM dim_producto as p
GROUP BY(p.id_marca)
ORDER BY p.id_marca ASC

```

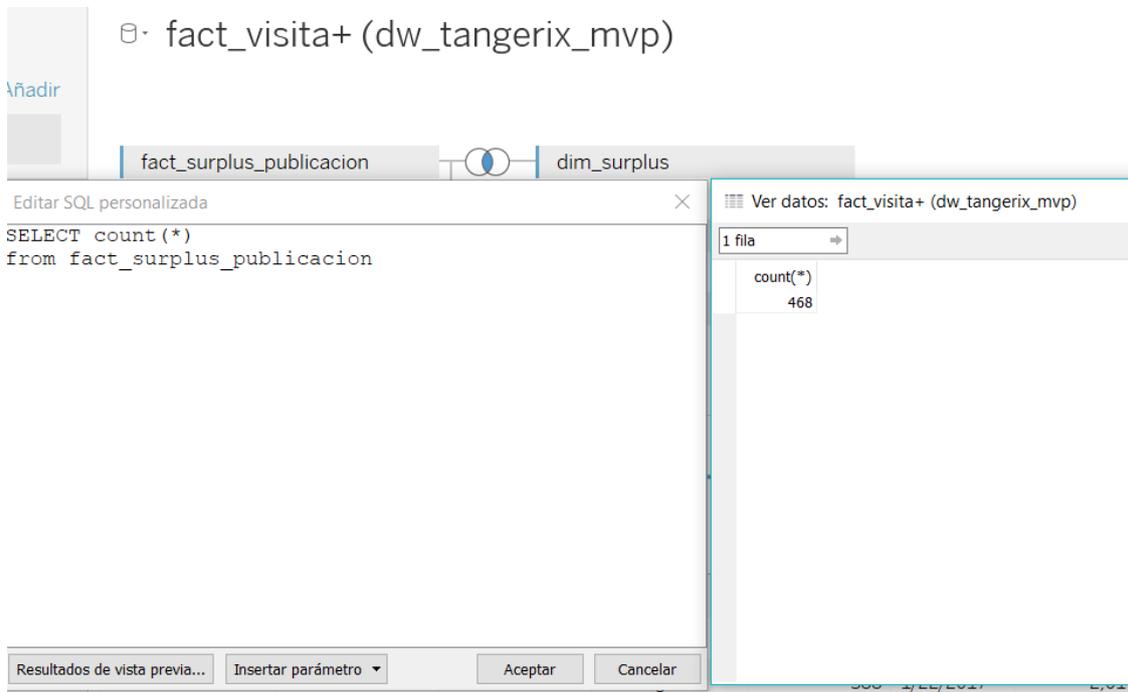
Ver datos: fact\_publicaciones+ (dw\_tangerix\_mvp)

78 filas

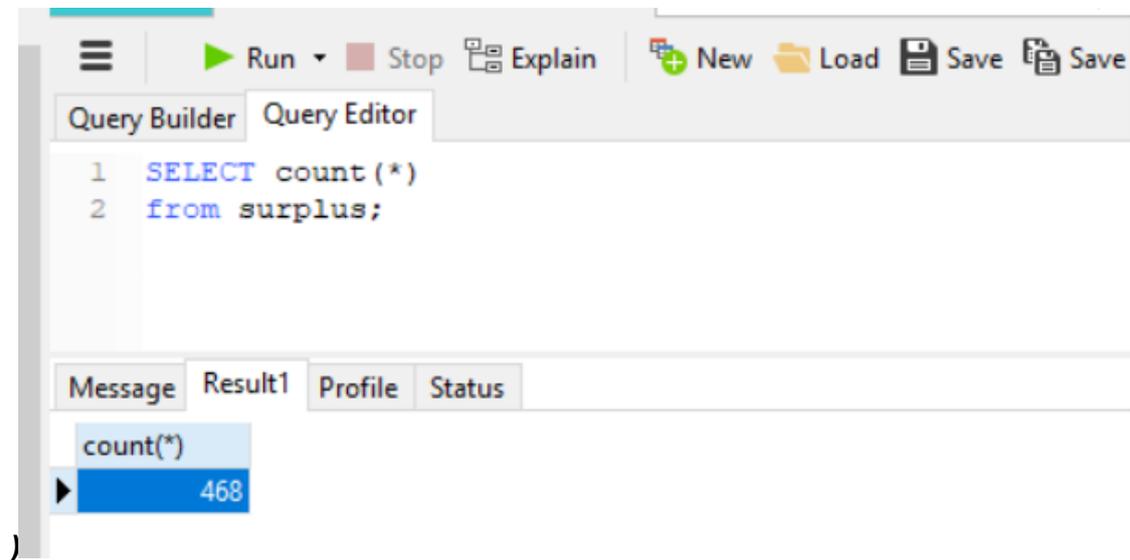
count(*)	id_marca
8	0
43	1
8	3
27	4
5	5
1	6
27	7
3	8
12	9
6	10
16	11
32	13
24	14
11	15
8	16
29	17
10	18
6	19

**Ilustración 64. Cantidad de productos publicados por marca (desde el almacén de datos).**

6. Aplicar una consulta para verificar la coincidencia entre la cantidad de surplus publicados en portal web entre la fuente de datos y la tabla dimensión fact\_surplus\_publicacion del almacén de datos. (Ilustración 68,69).

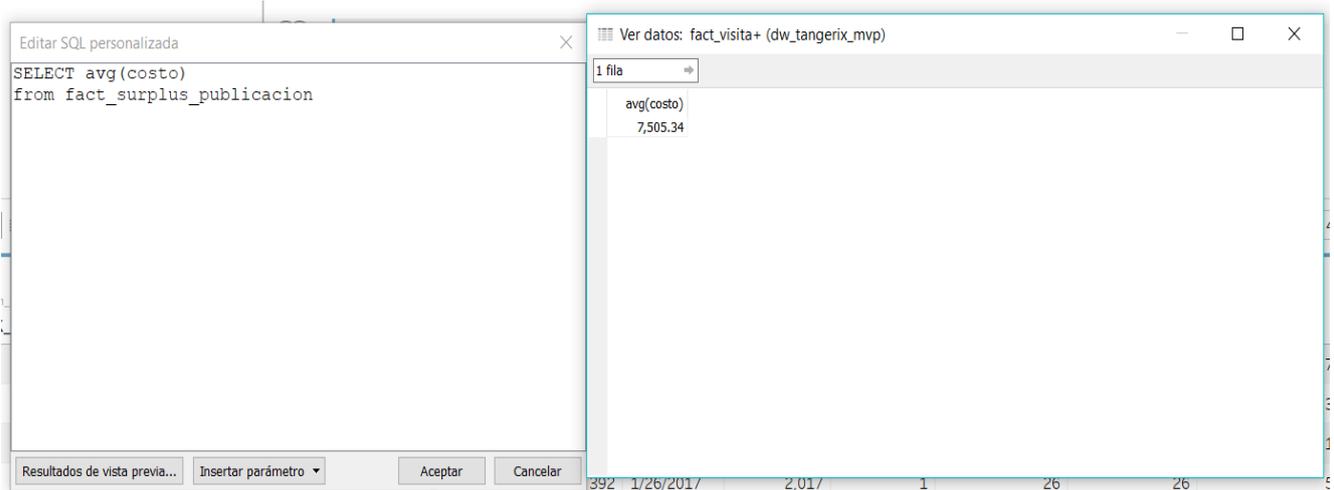


**Ilustración 65. Cantidad de surplus publicados (desde el almacén de datos).**

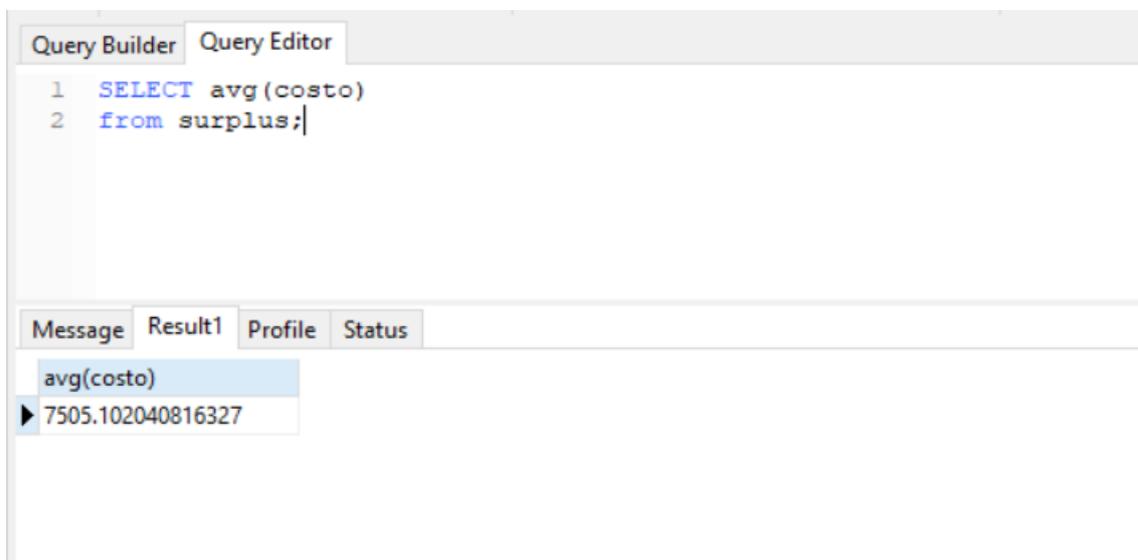


**Ilustración 66. Cantidad de surplus publicados (desde la fuente de datos).**

7. Aplicar una consulta para verificar la coincidencia entre el promedio de costo los surplus publicados en portal web entre la fuente de datos y la tabla dimensión fact\_surplus\_publicacion del almacén de datos. (Ilustración 70,71).



**Ilustración 67. Promedio de costo de Surplus (desde el almacén de datos)**



**Ilustración 68. Promedio de costo de Surplus (desde la fuente de datos).**

8. Aplicar una consulta para verificar la coincidencia entre los productos contactados a través portal web entre la fuente de datos y la tabla dimensión fact\_contacto\_vendedor del almacén de datos. (Ilustración 72,73).

Run Stop Explain New Load Save

Query Builder Query Editor

```

1 SELECT count(*), id_producto
2 from contact_sellers
3 GROUP BY id_producto
4

```

Message Result1 Profile Status

count(*)	id_producto
6	462
5	465
2	467
8	470
3	475
4	476
2	477
12	478
8	487
7	492
5	493
4	494
3	496
5	497
3	499
5	500
6	501
6	502
7	503
5	504

+ - ✓ ✕ ↺ ⌂

**Ilustración 69. Cantidad de veces que han sido contactados un producto (desde la fuente de datos).**

Editar SQL personalizada

```
SELECT count(*), SK_PRODUCTO
from fact_contacto_vendedor
GROUP BY SK_PRODUCTO
```

Ver datos: fact\_contacto\_vendedor+ (dw\_tangerix\_mvvp)

218 filas

count(*)	SK_PRODUCTO
7	397
3	402
4	403
2	404
12	405
7	414
7	419
5	420
4	421
3	423
5	424
3	426
5	427
6	428
6	429
7	430
4	431
2	432

Resultados de vista previa... Insertar parámetro Aceptar Cancelar

**Ilustración 70. Cantidad de veces que han sido contactado un producto (desde el almacén de datos).**

9. Aplicar una consulta para verificar la coincidencia entre la cantidad de productos contactados a través del portal web entre la fuente de datos y la tabla dimensión fact\_contacto\_vendedor del almacén de datos. (Ilustración 74,75).

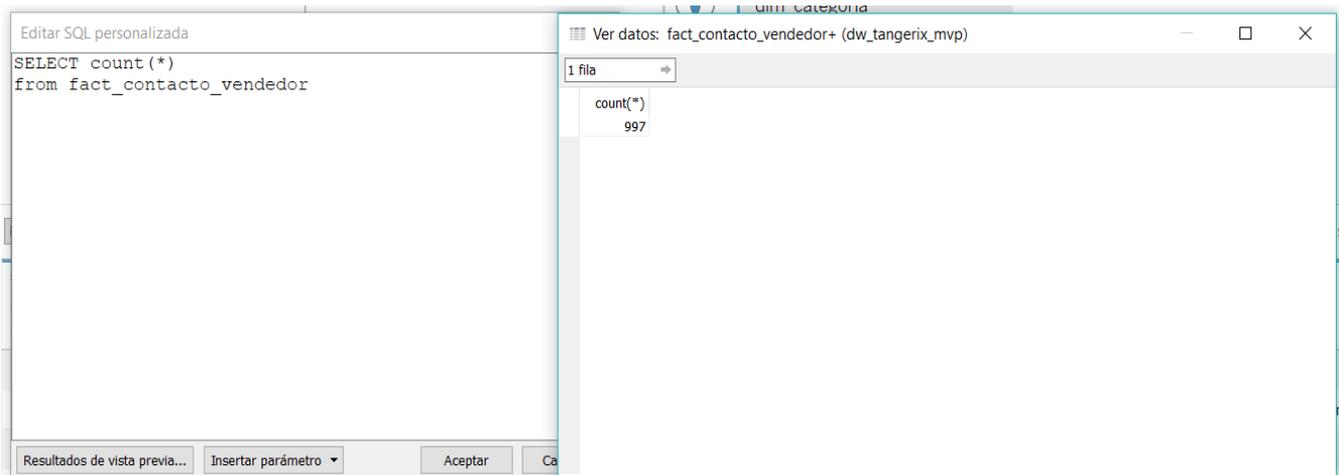
Query Builder Query Editor

```
1 SELECT count(*)
2 from contact_sellers;
```

Message Result1 Profile Status

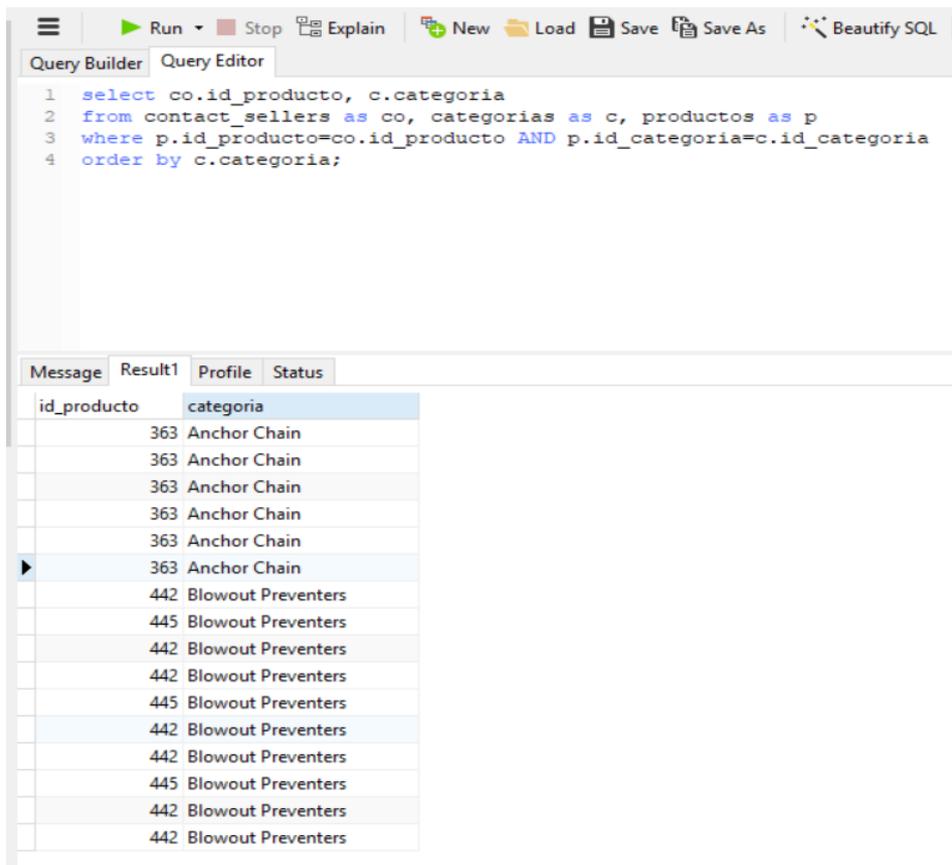
count(*)
997

**Ilustración 71. Cantidad de productos contactados del portal web (desde la fuente de datos).**



**Ilustración 72. Cantidad de productos contactados del portal web (desde el almacén de datos).**

10. Aplicar una consulta para verificar la coincidencia las categorías más buscadas (contactadas) del portal web entre la fuente de datos y la tabla dimensión fact\_contacto\_vendedor del almacén de datos. (Ilustración 76,77).



**Ilustración 73. Productos contactados en el portal web por categoría (desde la fuente de datos).**

Editar SQL personalizada

```
select co.sk_producto, c.categoria
from fact_contacto_vendedor as co, dim_categoria as c,
where p.SK_PRODUCTO=co.SK_PRODUCTO AND p.id_categoria=c
order by c.categoria
```

Ver datos: fact\_contacto\_vendedor+ (dw\_tangerix\_mvp)

909 filas

categoria	sk_producto
Accumulator Units	54
Accumulator Units	55
Accumulator Units	55
Accumulator Units	54
Accumulator Units	55
Accumulator Units	55
Accumulator Units	54
Anchor Chain	307
Bearings	60
Bearings	59
Bearings	59
Bearings	60
Bearings	59
Bearings	60

**Ilustración 74. Productos contactados en el portal web por categoría (desde el almacén de datos).**

11. Aplicar una consulta para verificar la coincidencia la cantidad de subastas publicadas en el portal web entre la fuente de datos y la tabla dimensión fact\_subasta del almacén de datos. (Ilustración 78,79).

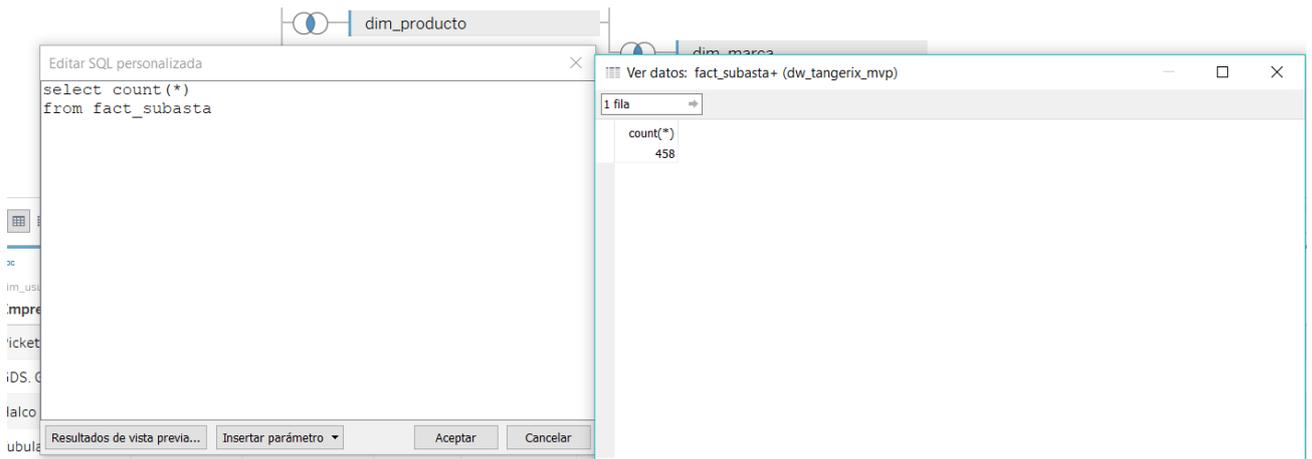
Query Builder Query Editor

```
1 select count(*)
2 from bids;
```

Message Result1 Profile Status

count(*)
458

**Ilustración 75. Cantidad de subastas (desde la fuente de datos).**

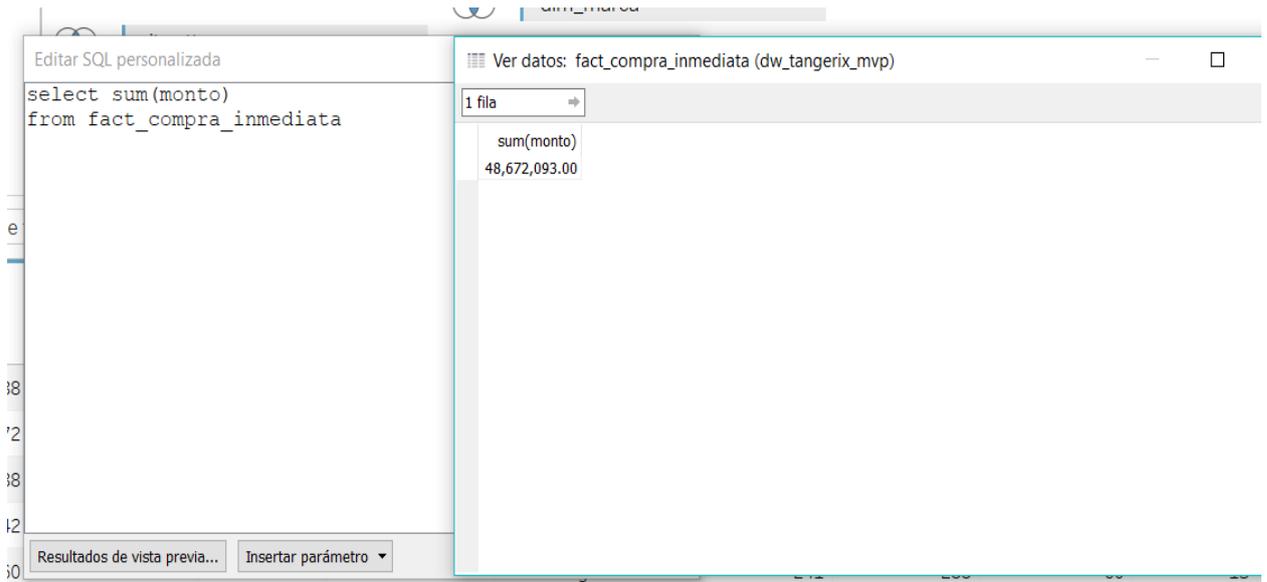


**Ilustración 76. Cantidad de subastas (desde el almacén de datos).**

12. Aplicar una consulta para verificar la coincidencia la ganancia de compra inmediata (BON) en el portal web entre la fuente de datos y la tabla dimensión fact\_compra\_inmediata del almacén de datos. (Ilustración 80,81).

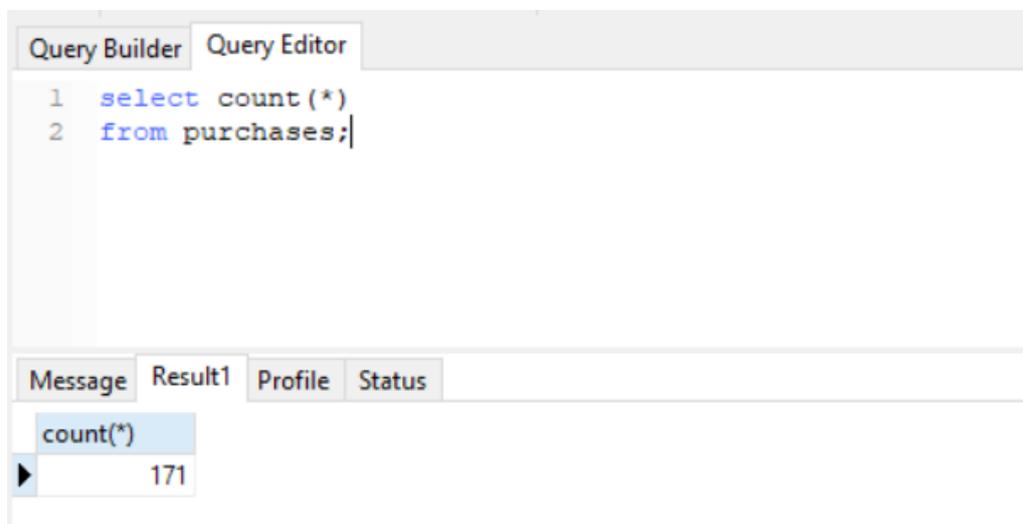


**Ilustración 77. Ganancias de compras inmediatas en el portal web (desde la fuente de datos).**

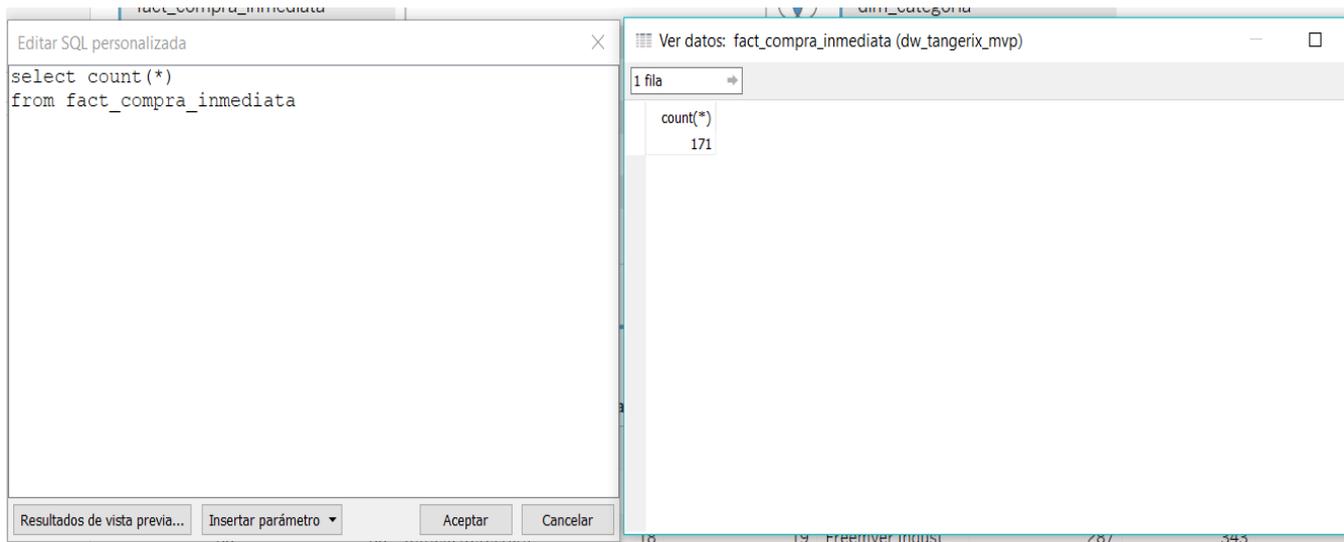


**Ilustración 78. Ganancias de compras inmediatas en el portal web (desde el almacén de datos).**

13. Aplicar una consulta para verificar la coincidencia la cantidad de productos vendidos por compra inmediata (Buy option now) en el portal web entre la fuentes de datos y la tabla dimensión fact\_compra\_inmediata del almacén de datos. (Ilustración 82,83).

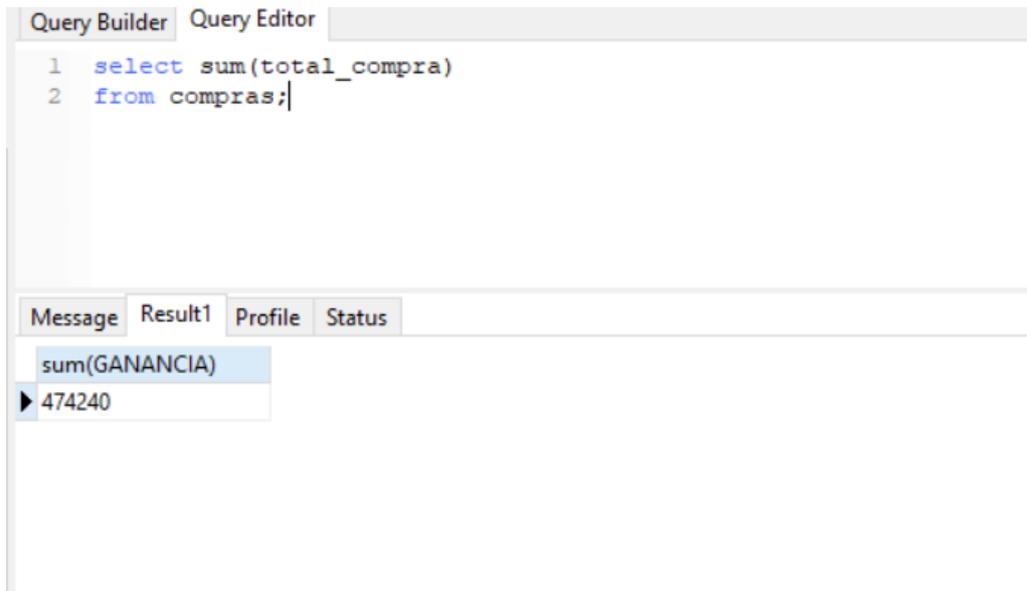


**Ilustración 79. Cantidad de compras inmediatas en el portal web (desde la fuente de datos).**

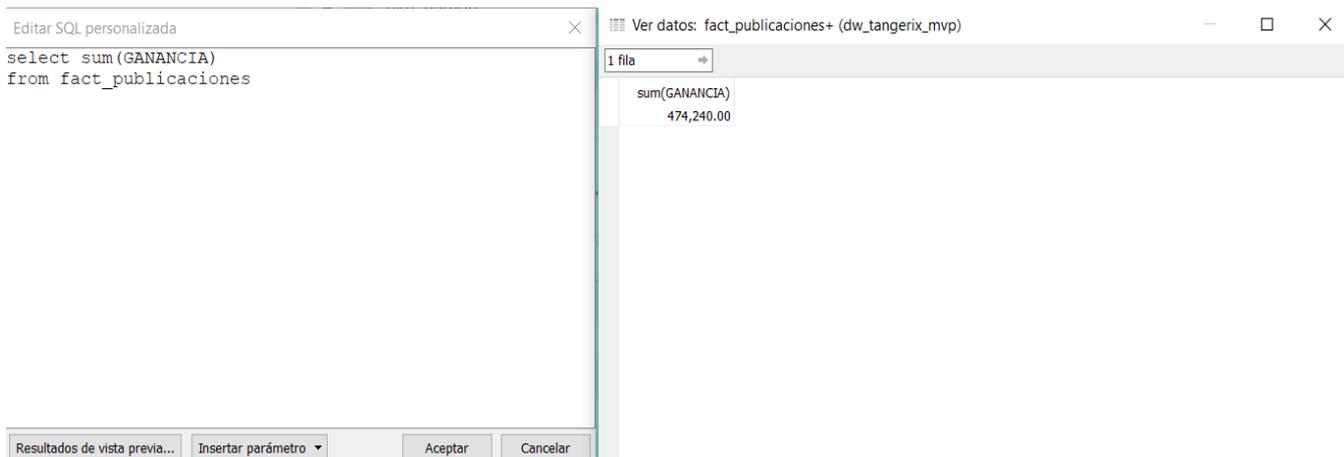


**Ilustración 80. Cantidad de compras inmediatas en el portal web (desde el almacén de datos).**

14. Aplicar una consulta para verificar la coincidencia la ganancia de las publicaciones en el portal web entre la fuente de datos y la tabla dimensión fact\_publicaciones del almacén de datos. (Ilustración 84,85).



**Ilustración 81. Ganancia de las publicaciones del portal web (desde la fuente de datos).**



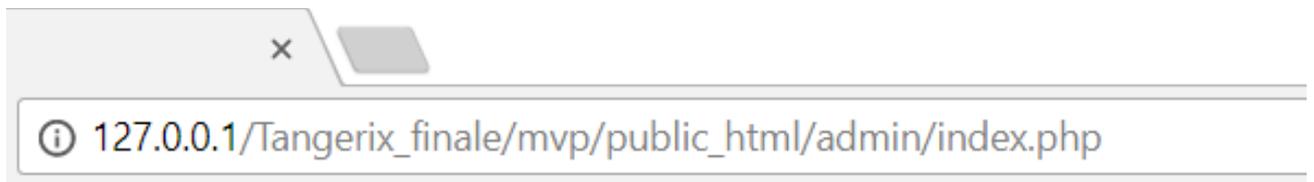
**Ilustración 82. Ganancia de las publicaciones del portal web (desde el almacén de datos).**

Al finalizar el desarrollo de la solución de Inteligencia de Negocio, se procede a realizar un manual de usuarios para la educación del uso de la herramienta, lo que requiere una serie de instrucciones que permita a estos usuarios comprender el almacén de datos, así como el conjunto de herramientas que giran en torno al almacén de datos.

#### 4.1.10.2 Manual de usuarios

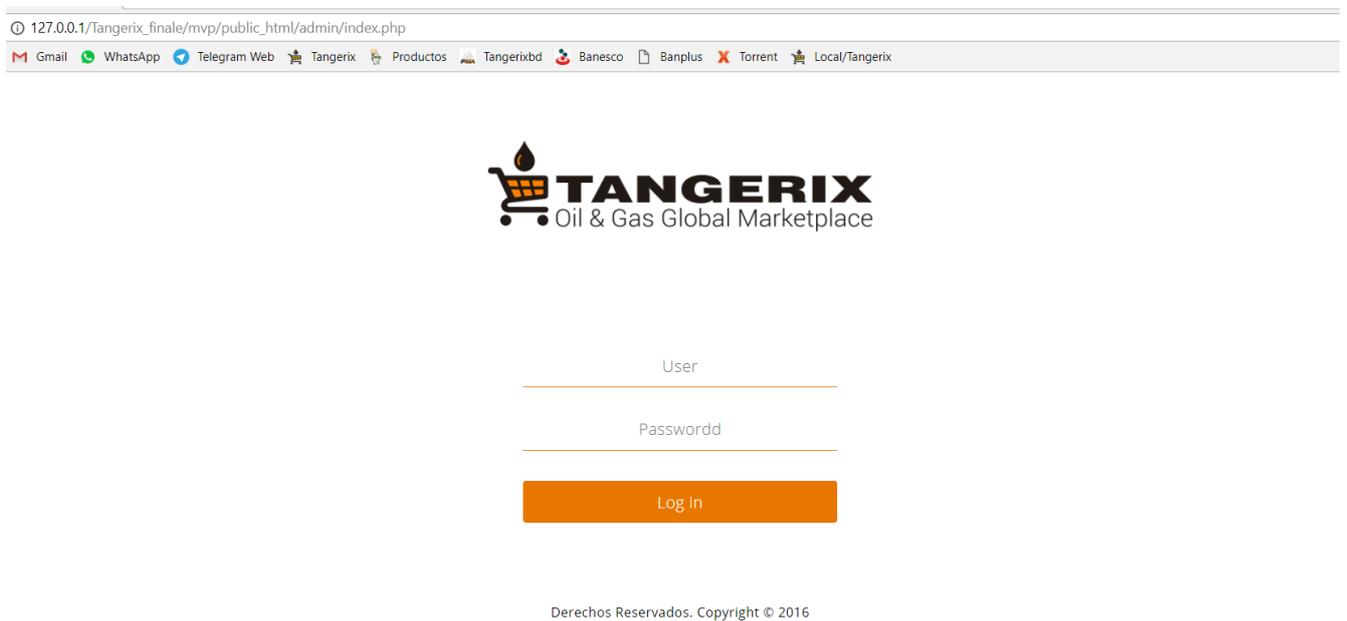
Para entender el uso de la herramienta de visualización de la información, como lo son los tableros de control y reportes, el tablero tiene que seguir las siguientes instrucciones:

1. Ingresar en la barra de búsqueda del explorador de internet el siguiente enlace: [http://localhost/Tangerix\\_finale/mvp/public\\_html/admin/](http://localhost/Tangerix_finale/mvp/public_html/admin/)



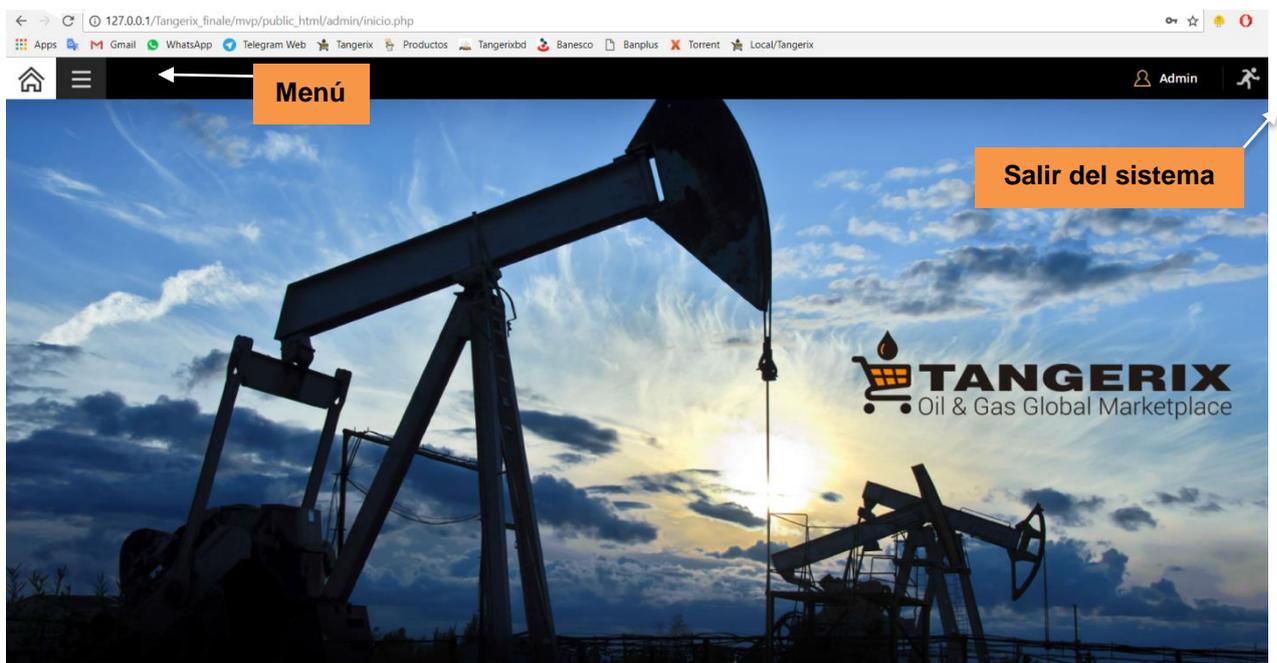
**Ilustración 83. URL de acceso del portal web administrativo.**

2. Para ingresar en el portal web administrativo el usuario debe proporcionar un usuario y una clave, los cuales son asignados por la organización:



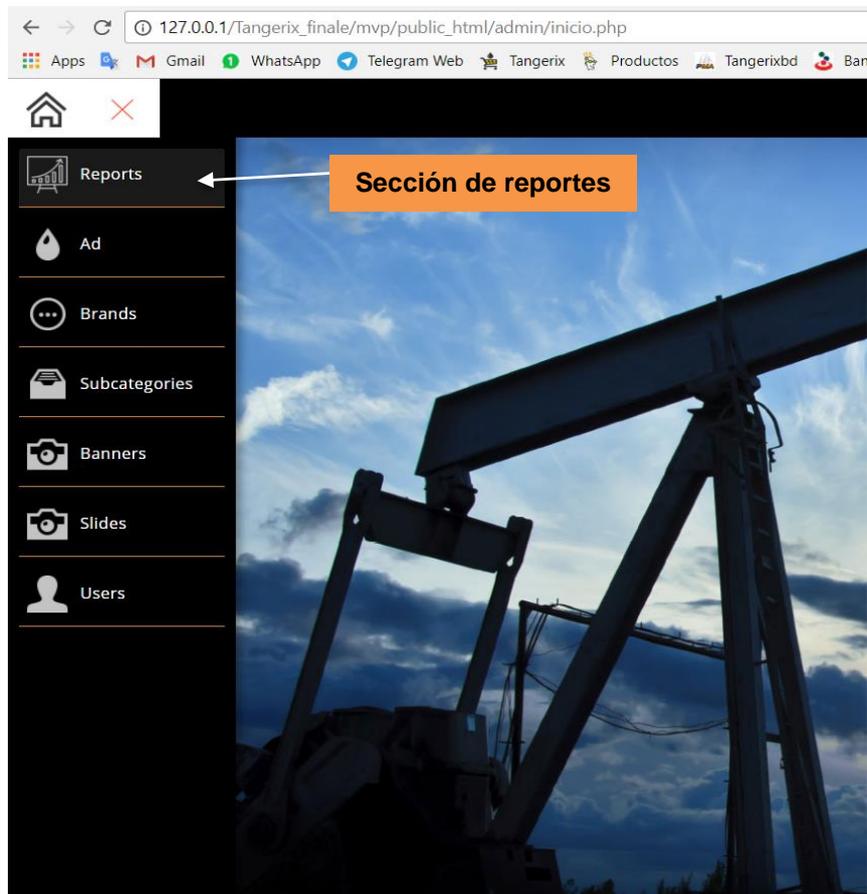
**Ilustración 84. Interfaz para ingresar en el portal web administrativo.**

3. Una vez ingresado el usuario, accede a la página principal del portal administrativo:



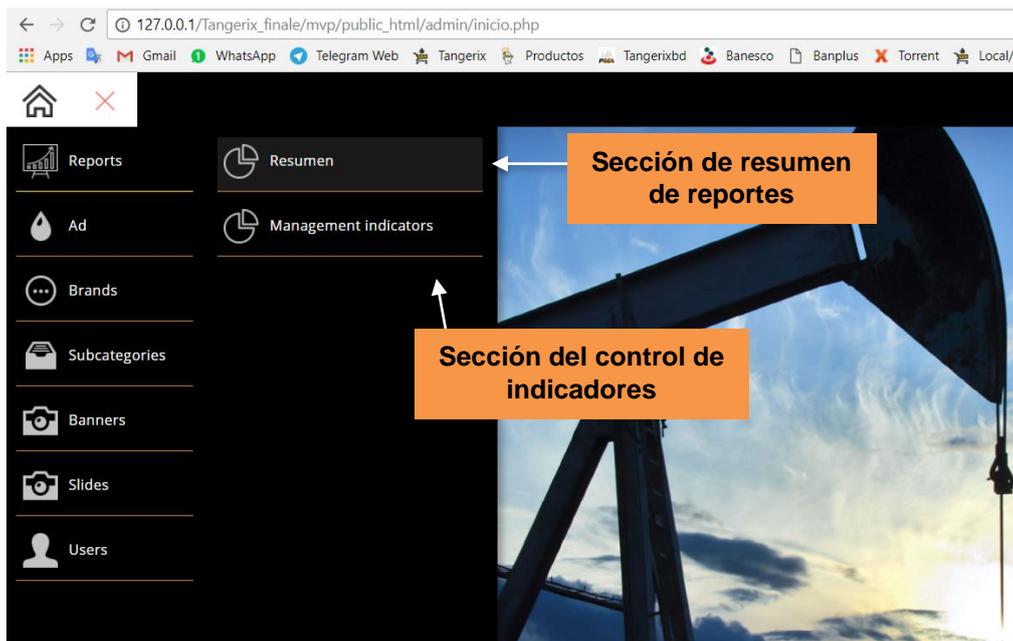
**Ilustración 85. Página principal del portal web administrativo.**

4. Si el usuario desea acceder a las secciones de reportes y tableros de control, tiene que abrir el menú ubicado en la derecha superior de la página principal.



**Ilustración 86. Menú principal del portal web administrativo.**

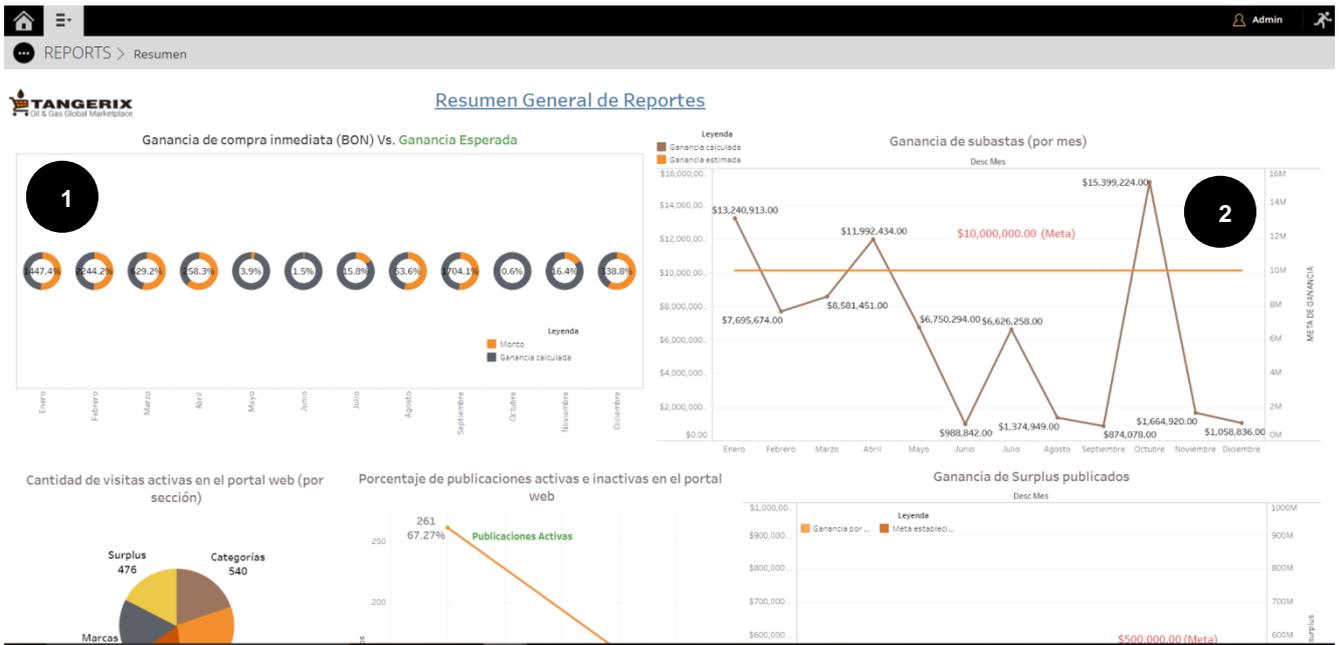
5. El menú principal tiene varias opciones de navegación, el usuario tendría que darle click a la opción de “Reports” para que se despliegue el siguiente nivel de ese menú que muestra dos opciones más (Resumen, Management Indicators)



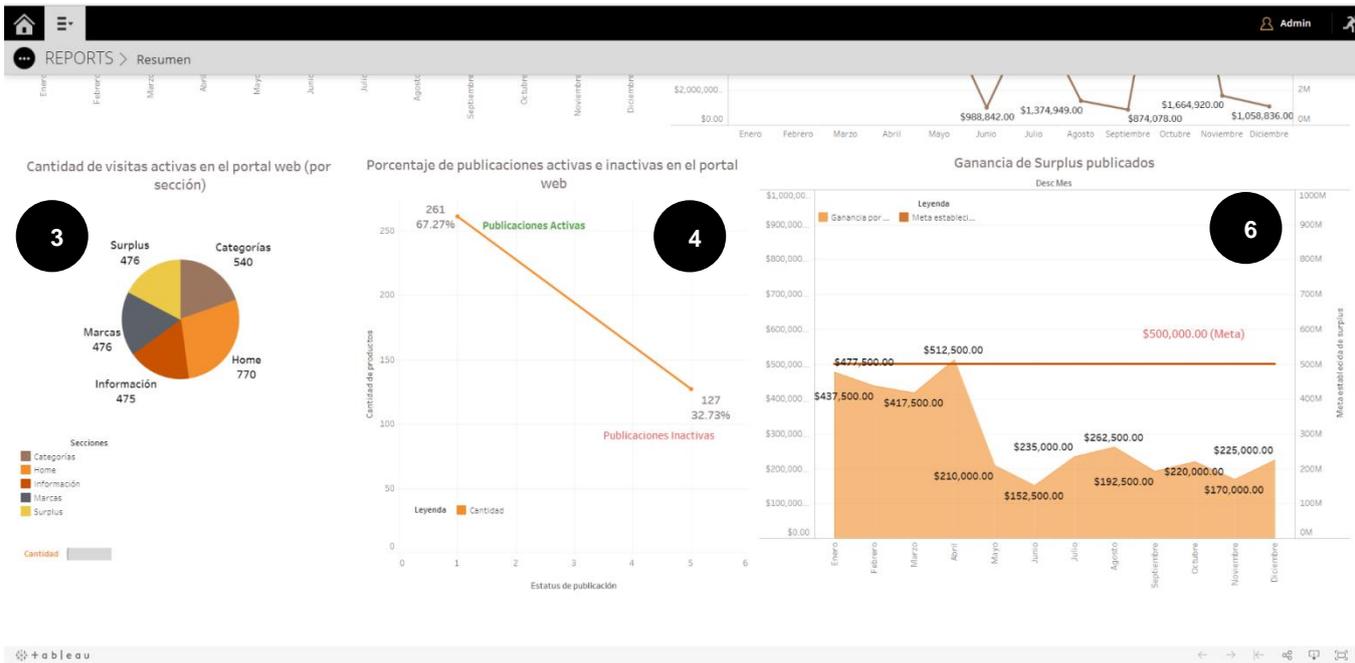
***Ilustración 87. Menú principal del portal web administrativo (con segundo nivel de despliegue).***

6. Existen dos opciones en el portal web administrativo para que el usuario acceda a los indicadores de gestión desarrollados en el proyecto:

- a. **Opción #1 - Resumen:** Sección donde se presenta un tablero de control, el cual es un resumen general de los reportes más destacados del portal web principal.



**Ilustración 88. Tablero de control del resumen general de los reportes del portal web.**



**Ilustración 89. Tablero de control del resumen general de los reportes del portal web.**

Como se observa en la ilustración 91 y 92, se presentan los cinco (5) reportes más destacados de los procesos del portal web principal:

Los 5 reportes más destacados son los siguientes:

\*(1): Gráfico de donas donde se compara la Ganancia de compra inmediata Vs la Ganancia Calculada por mes, si el usuario desea saber con detalle el monto de ganancia y el monto calculado puede pasar el mouse por el gráfico de donas y se le mostrarán los detalles de esta información. A su vez, tenemos una leyenda de colores para un mayor entendimiento de la gráfica.

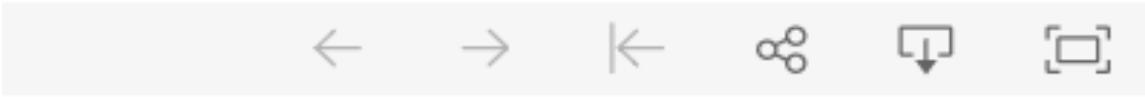
\*(2): Gráfico de líneas donde se observa el monto de la ganancia por cada mes y a su vez se va comparando con una meta que se planteó por la organización mensual, para poder captar cuanto se supera la meta o no. De igual manera, se presenta una leyenda para el mayor entendimiento de la gráfica.

\*(3): Gráfico de torta (pie-chart) donde presenta la distribución de la cantidad de visitas activas en el portal web por sección (home, información, categorías, marcas, surplus), de igual manera se observa la cantidad total de visitas activas en el portal web.

\*(4): Gráfico de líneas representando el porcentaje y la cantidad de publicaciones activas e inactivas del portal web, mostrando una leyenda para tener un mayor conocimiento de la gráfica

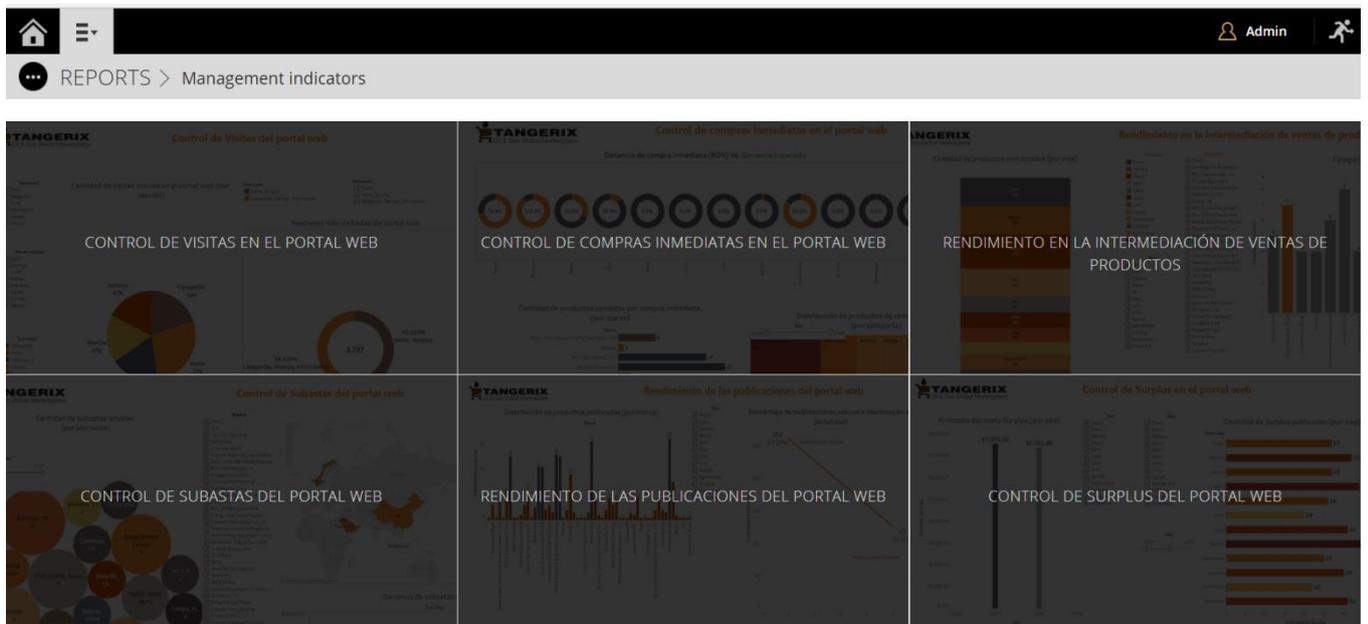
\*(5): Gráfica de líneas que representa el monto ganado por los surplus publicados en el portal web para poder compararlo con la meta mensual que se tiene planteada para este proceso, si se desea conocer detalles el usuario tiene que arrastrar el mouse por encima y se presentarán los detalles. Aunado a esto, se tiene su leyenda para el mayor entendimiento de la gráfica.

Cabe destacar que en la parte inferior derecha tenemos varios elementos que nos permiten interactuar con el tablero de control (dashboard), como por ejemplo se puede con compartir(a través de una URL), descargar, o colocar pantalla completa.



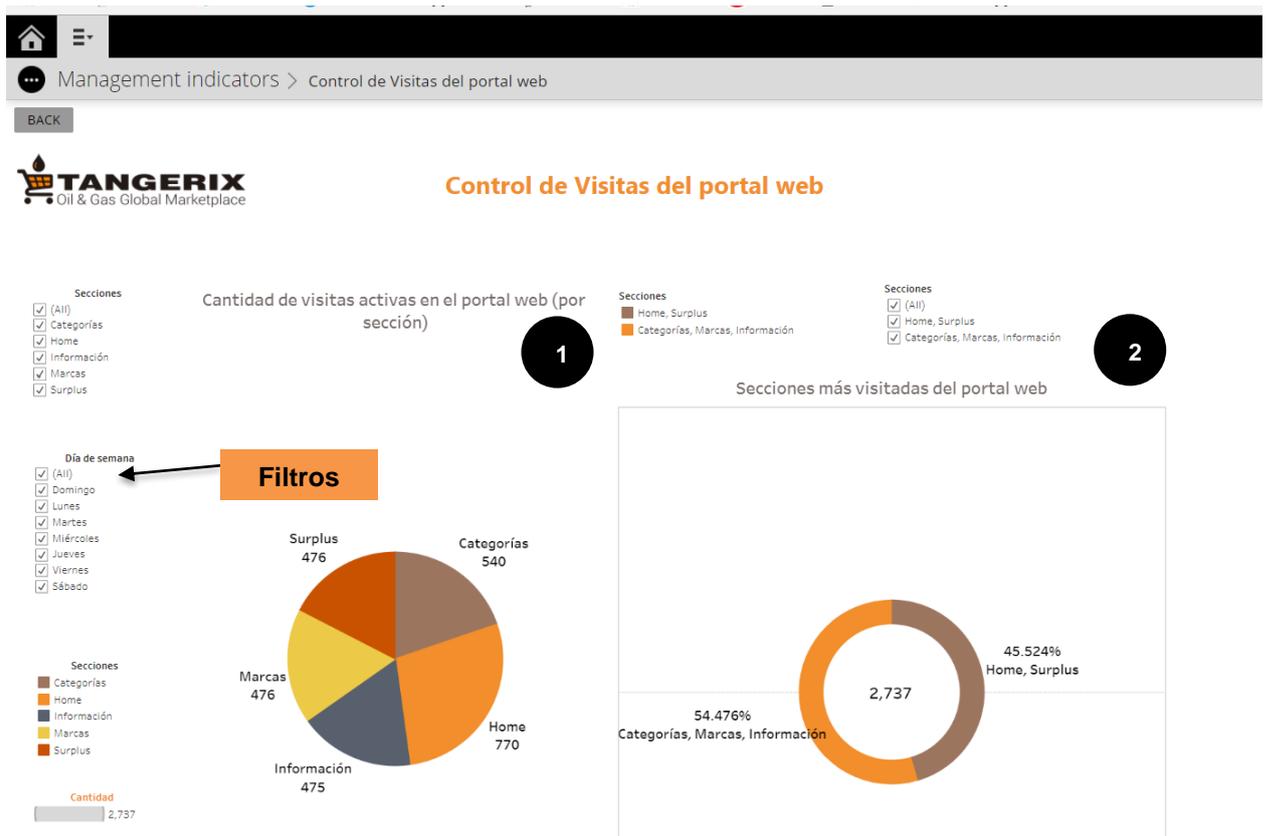
**Ilustración 90. Opciones de compartir, descargar, full pantalla del dashboard.**

**b. Opción #2 - Management Indicators:** Sección donde se presentan el conjunto de indicadores planteados al principio del proyecto, dándole la opción al usuario para que seleccione(con un click) cuál indicador desee ver.



**Ilustración 91. Interfaz de indicadores del portal web administrativo.**

**b.1. Control de visitas en el portal web:** Tablero de control (dashboard) donde se presentan todos los reportes con respecto al control de visitas del portal web.



**Ilustración 92. Tablero de control de “Control de Visitas del portal web”.**

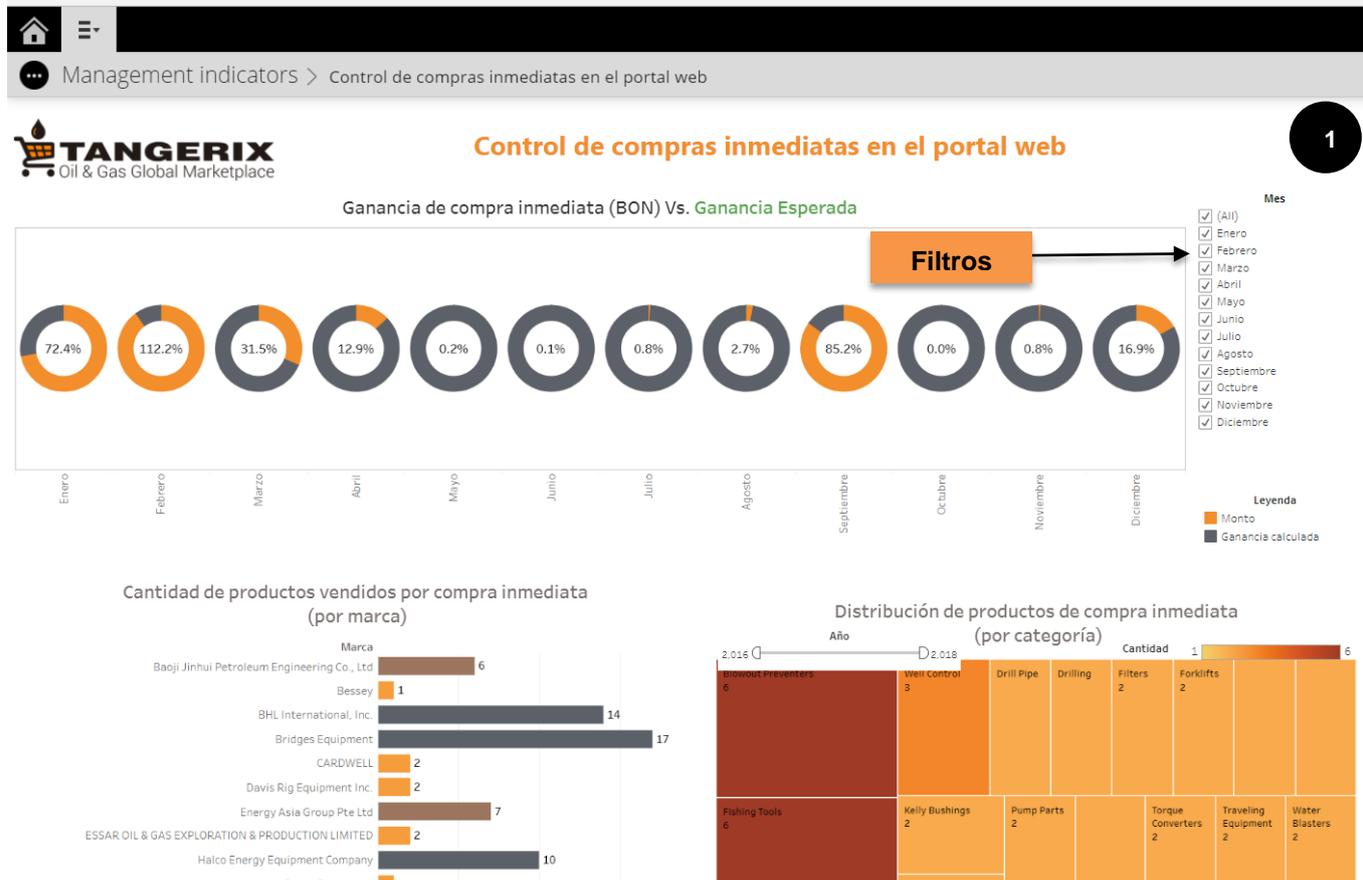
Los indicadores que se muestran en el tablero de control son los siguientes:

\*(1): Gráfico de torta que representa la cantidad de visitas activas en el portal web distribuidos por las secciones del portal (home, información, categorías, marcas, surplus), si el usuario quiere detalles de la información puede arrastrar el punto a cualquier parte de la gráfica.

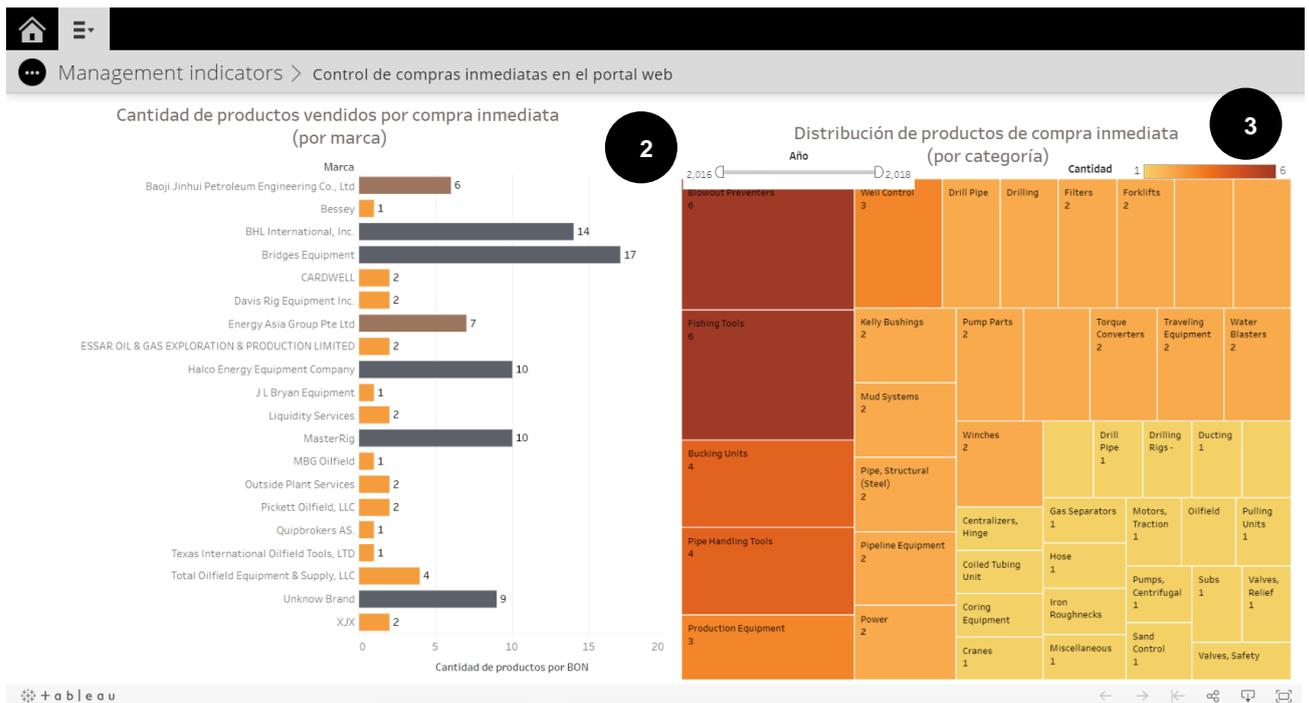
En la parte superior izquierda tenemos los filtros que el usuario puede usar, el primer filtro permite al usuario filtrar por sección del portal web, el segundo filtro permite filtrar por día de semana. Y finalizando la gráfica que nos indica el color de cada sección para entender la gráfica de manera más fácil.

\*(2): Gráfica de dona que permite representar las secciones más visitadas del portal web, comparando la cantidad de visitas de la secciones de (home y surplus) y categorías, marcas e información. La gráfica tiene un pequeño filtro para que el usuario pueda filtrar por los grupos de las secciones.

**b.2. Control de compras inmediatas en el portal web:** Tablero de control (dashboard) donde se presentan todos los reportes asociados al control de compras inmediatas (Buy Option Now) del portal web.



**Ilustración 93. Tablero de control de "Control de Visitas del portal web."**



**Ilustración 94. Tablero de control de “Control de Visitas del portal web.**

Los indicadores que se muestran en el tablero de control son los siguientes:

\*(1): Gráfica de donas que permite comparar las ganancias obtenidas por las compras inmediatas Vs. la ganancia esperada por cada mes, apreciándose con un porcentaje si se obtuvo una ganancia o no, del lado derecho en la parte superior se observa un filtro que le permite al usuario alterar las gráficas por mes, y finalmente tenemos una leyenda donde nos explica la información de la gráfica.

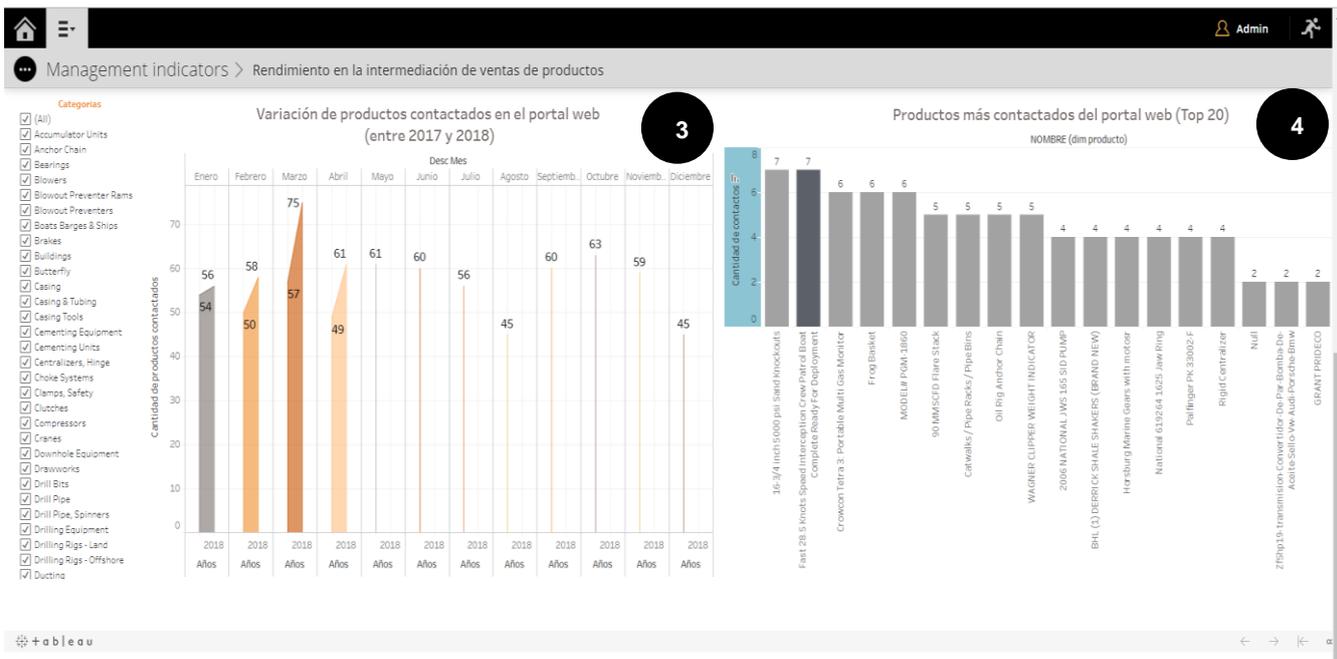
\*(2): Gráfico de barras que presenta la cantidad de productos que fueron vendidos en la modalidad de “Compra Inmediata”, se observa que el gráfico está distribuido por marcas

\*(3): Gráfico de cuadros que representa la distribución de productos publicados en el portal web con la modalidad de “Compra inmediata” por categoría, observamos que en la gráfica se muestra la cantidad de productos publicados por categoría. Aunado esto, tenemos un filtro que permite al usuario seleccionar el rango de tiempo que quiere observar este indicador (por año).

**b.3. Rendimiento en la intermediación de ventas de productos:** Tablero de control (dashboard) donde se presentan los reportes relacionados al rendimiento en la información de ventas de productos.



**Ilustración 95. Tablero de control de “Rendimiento en las ventas de productos”.**



**Ilustración 96. Tablero de control de “Rendimiento en las ventas de productos”.**

Los indicadores que se muestran en el tablero de control son los siguientes:

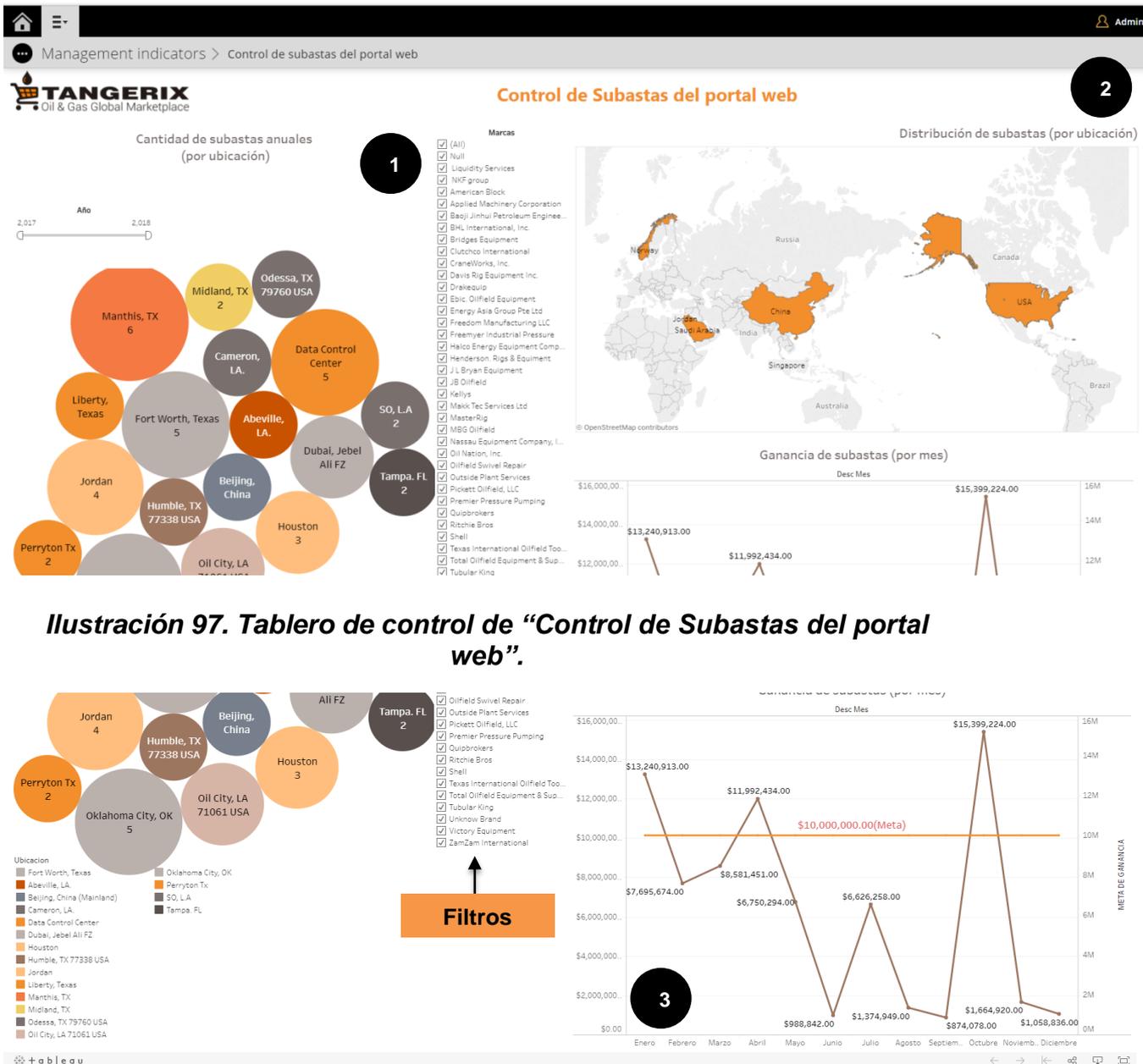
\*(1): Gráfica de barras que representa la cantidad de productos que fueron contactados a través del portal web por cada mes, en el lado derecho de la gráfica se observa un filtro por mes, que permite alterar al usuario la gráfica, y su vez tenemos una leyenda que muestra el color de cada mes para el mejor entendimiento de la gráfica.

\*(2): Gráfica de barras que indica las categorías más buscadas (más contactadas) en el portal web, a su lado izquierdo tenemos un filtro por empresa que puede utilizar el usuario para ver esta gráfica por una empresa específica.

\*(3): Gráfico de línea que muestra la variación de los productos que han sido contactados a través del portal web entre el 2017 y 2018, en la gráfica se observa la variación entre estos dos años y son distribuidos por cada mes, al lado izquierdo podemos observar un filtro por categoría para que el usuario pueda usarlo.

\*(4): Gráfica de barras que representa los productos más contactados del portal web (con un límite de 20 productos), lo cual significa cuántas veces ha sido contactado un producto a través del portal para concretar una venta.

**b.4. Control de subastas del portal web:** Tablero de control (dashboard) donde se presentan los reportes asociados a las subastas publicadas en el portal web.



**Ilustración 97. Tablero de control de “Control de Subastas del portal web”.**

**Ilustración 98. Tablero de control de “Control de Subastas del portal web”.**

Los indicadores que se muestran en el tablero de control son los siguientes:

\*(1): Gráfico circular que presenta la cantidad de subastas publicadas en el portal web por ubicación, en el lado derecho el usuario puede filtrar por marcas y finalmente tenemos la leyenda informativa para que el usuario pueda entender mejor la gráfica.

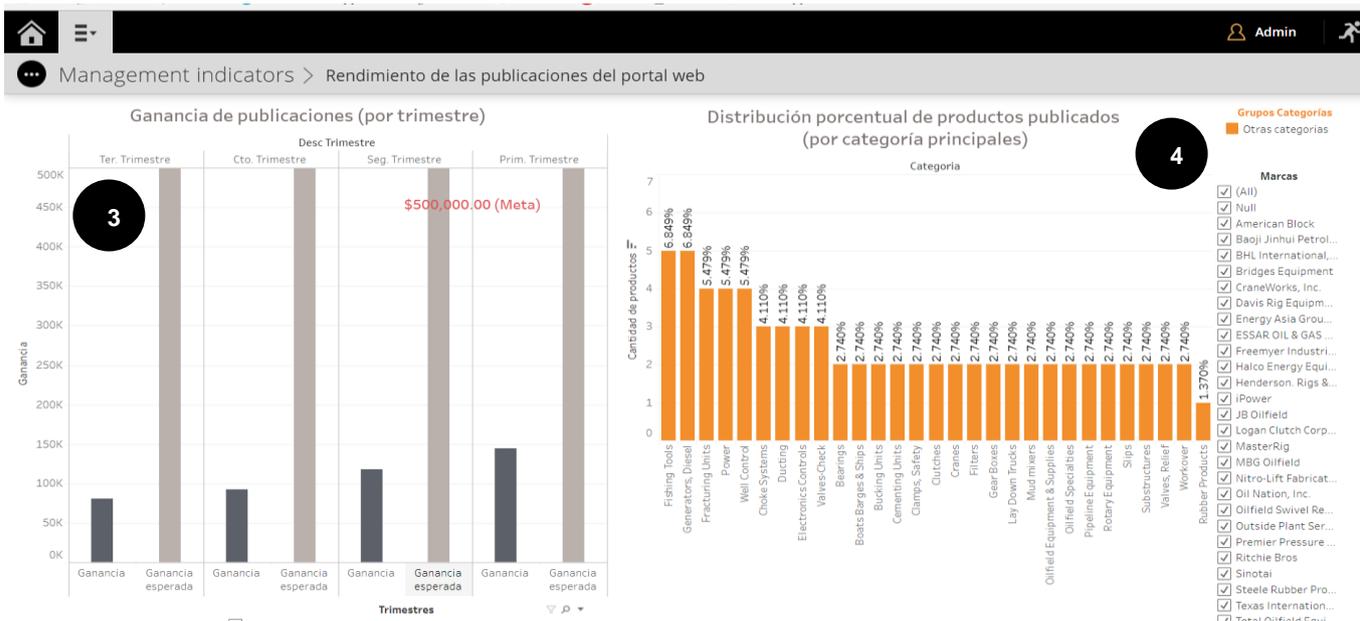
\*(2): Mapa geográfico que representa los países en donde se encuentran los productos publicados con la modalidad de “subastas”.

\*(3): Gráfica de líneas donde permite comparar la ganancia obtenida por las subastas publicadas en el portal web por mes y ser comparadas con una meta planteada por la organización.

b.5. Rendimiento de las publicaciones del portal web: Tablero de control (dashboard) donde se presentan los reportes asociados a las publicaciones del portal web.



**Ilustración 99. Tablero de control de “Rendimiento de las publicaciones del portal web”.**



**Ilustración 100. Tablero de control de “Rendimiento de las publicaciones del portal web”.**

Los indicadores que se muestran en el tablero de control son los siguientes:

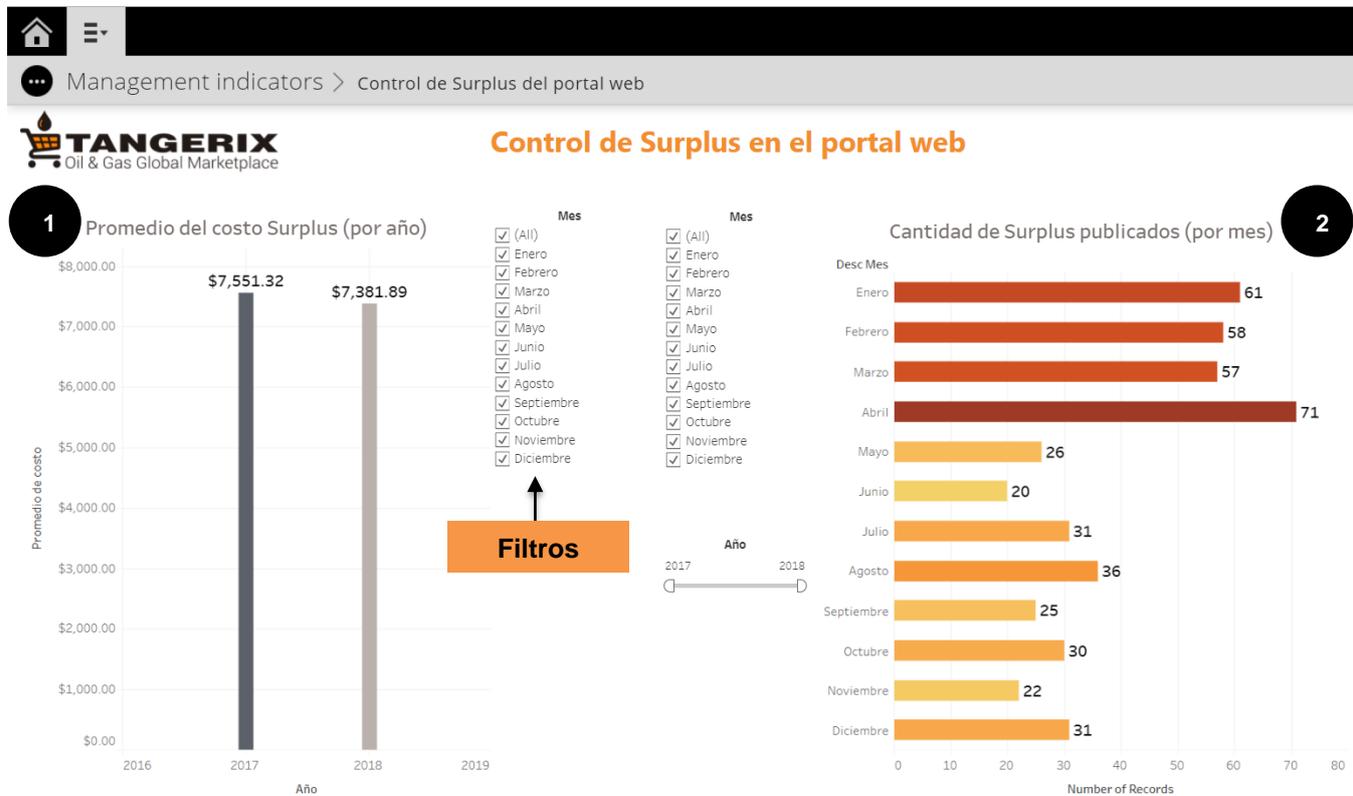
\* (1): Gráfica de barras que presenta la cantidad de productos publicados en el portal web por marca, el usuario al lado derecho tiene la posibilidad de poder filtrar por mes para obtener una información más detallada.

\* (2): Gráfica de líneas donde se puede observar la cantidad y el porcentaje de las publicaciones activas e inactivas del portal web, de la misma manera tiene en la parte superior derecha el filtro por mes para alterar la gráfica.

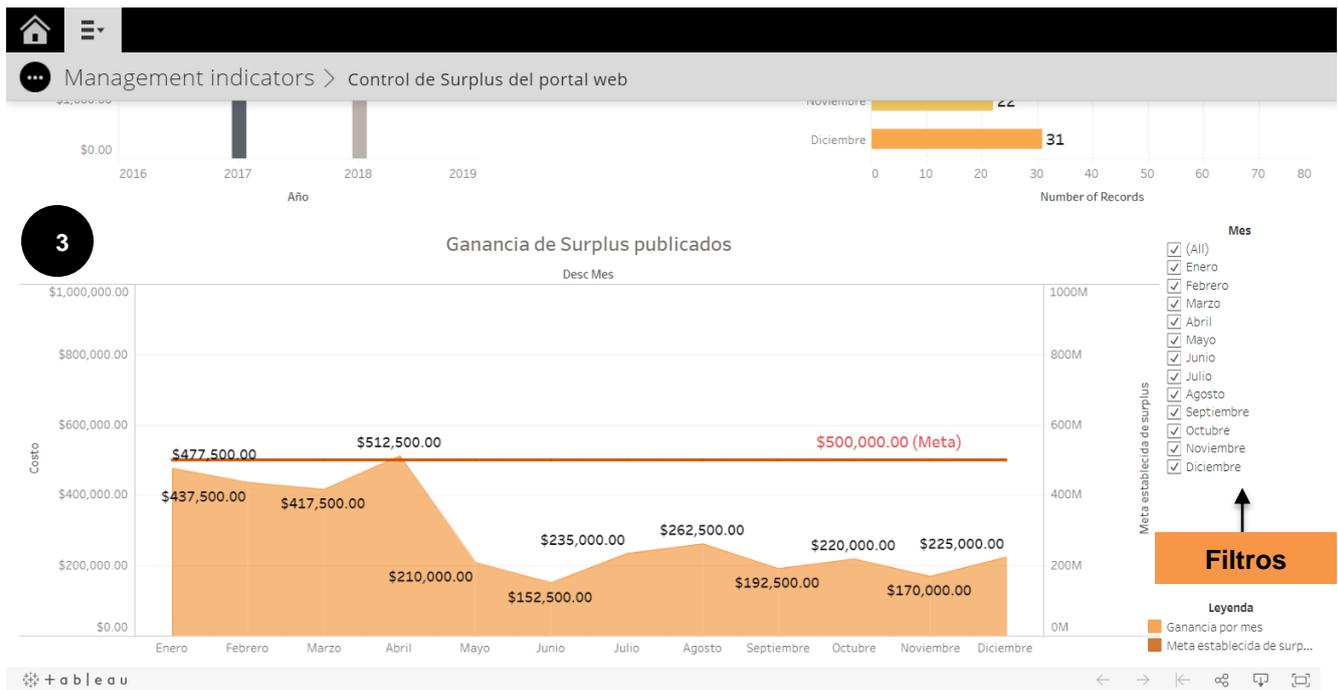
\* (3): Gráfica de barras que representa las ganancias obtenidas por las publicaciones en el portal web por trimestre, en cada trimestre se compara con una meta mensual planteada por la organización y el usuario en la parte inferior de la gráfica puede filtrar por trimestre.

\* (4): Gráfica de barras que nos muestra la distribución de publicaciones en el portal web pero esta vez por categoría (con un límite de categorías), y finalmente se tiene un filtro por marca de producto para que el usuario pueda seleccionar una marca en específica.

**b.6.Control Surplus del portal web:** Tablero de control (dashboard) donde se presentan los reportes relacionados a los Surplus publicados en el portal web.



**Ilustración 101. Tablero de control de “Rendimiento de las publicaciones del portal web”.**



**Ilustración 102. Tablero de control de “Rendimiento de las publicaciones del portal web”.**

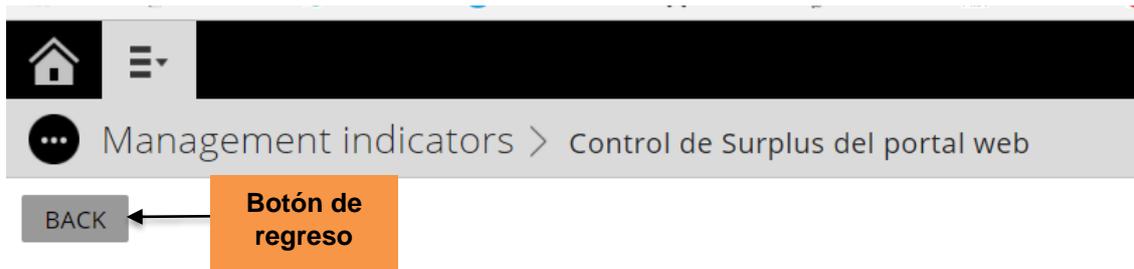
Los indicadores que se muestran en el tablero de control son los siguientes:

\* (1): Gráfica de barras donde nos indica el promedio de costo de un Surplus publicado en el portal web, se observa un promedio de costo en el 2017 y otro en el 2018. En la parte izquierda el usuario tiene la capacidad de filtrar por mes estos resultados.

\* (2): Gráfica de barras que nos presenta la cantidad de Surplus publicados mensualmente en el portal web, donde se tienen dos filtros en esta gráfica, se puede filtrar por mes y por año.

\* (3): Gráfica de líneas donde representa la ganancia mensual obtenida por los surplus publicados en el portal web, comparando esta con una meta mensual planteada por la organización, de igual manera se tiene un filtro por mes que puede ser usado por el usuario y una leyenda de colores para aumentar el entendimiento de la gráfica.

Finalmente si el usuario se encuentra dentro de un tablero de control (dashboard) en específico y quiere devolverse a la sección donde se encuentran los demás indicadores, sólo tiene que darle click al botón “Back”.



***Ilustración 103. Botón para volver a la interfaz de todos los indicadores.***

#### 4.1.10.3. Pruebas de Usabilidad

Al igual que con otros aspectos, es importante probar para validar el nivel de usabilidad que muestra un producto de software, en este caso del portal web administrativo previamente implementado.

Las pruebas de usabilidad se llevó a cabo observando a las personas mientras usan el producto a probar, realizando pruebas de las reglas heurísticas de Jakob Nielsen y realizando una encuesta al final para chequear la usabilidad del sistema, y a su vez este grupo de personas es seleccionado considerando ciertas características, según las condiciones que se desean valorar, categorizándolos por ejemplo como usuarios expertos, medios o inexpertos.

El portal web administrativo fue desarrollado en un tiempo atrás, donde fueron corregidas las fallas presentadas por los usuarios que lo usaban, finalmente siendo corregidas el sistema pasó la prueba de usabilidad, sin embargo al integrarse un nuevo módulo de “Reportes e Indicadores” a este portal, se volvieron a aplicarlas 10 reglas heurísticas del apodado “padre” o “gurú” de la usabilidad Jakob Nielsen para evaluar al mismo portal. Y de igual manera se volvió a realizar una encuesta a los usuarios de la organización que usan el portal web administrativo.

A continuación se aplican las reglas heurísticas de Jakob Nielsen:

1. **Visibilidad:** Siempre le muestra al usuario cuál es el estado del sistema en cada momento, y lo mantiene informado del estatus del mismo.
2. **Relación con la realidad:** Utiliza un lenguaje familiar y apropiado para los usuarios que usan el portal web.
3. **Control y libertad:** Ofrece funciones de rehacer y deshacer que permite al usuario tener el control de sus interacciones con una mayor libertad.
4. **Consistencia y estándares:** Establece convenciones lógicas y las mantiene de manera constante.

5. **Prevención de errores:** El portal web ayuda a los usuarios a evitar equivocarse antes de que cometan el error.
6. **Reconocimiento:** Hace visible todo lo que sea posible, no espera que los usuarios recuerden o memoricen información.
7. **Flexibilidad:** Permite que el sistema pueda adaptarse a los usuarios frecuentes, diseña la realización de tareas avanzadas de manera fluida y eficiente.
8. **Estética y minimalismo:** Muestra sólo lo necesario y relevante en cada situación, siendo lo más minimalista posible.
9. **Recuperarse de los errores:** Ayuda a los usuarios a reconocer y corregir sus errores, indica siempre el problema concreto que está ocurriendo.
10. **Ayuda y documentación:** La información de ayuda es breve, concisa, fácil de buscar y enfocada a las tareas del usuario.

Se realizó una encuesta con ocho (8) preguntas puntuales asociadas al portal web administrativo, obteniendo 15 respuestas de esta. A continuación los resultados de la encuesta obtenida por parte de los usuarios de la organización que usan el portal web administrativo:

15 respuestas



RESUMEN

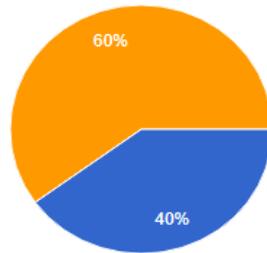
INDIVIDUAL

Se aceptan respuestas



Considera que el nuevo módulo es fácil de usar?

15 respuestas



- Si
- No
- Con el manual de usuarios se entiende sin problema.

**Ilustración 104. Pregunta #1 de la encuesta**

El nuevo módulo tiene relación con los otros módulos desarrollados

15 respuestas

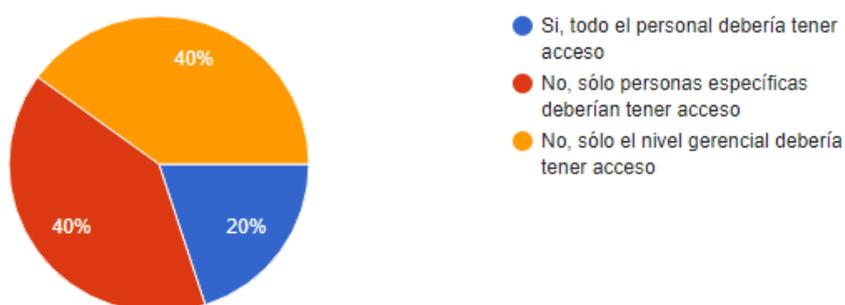


- Si, tiene relación con algunos
- No, no tiene relación con ninguno

**Ilustración 105. Pregunta #2 de la encuesta**

Cree que todo el personal de la organización debería de tener acceso a este nuevo módulo "Reportes e Indicadores"

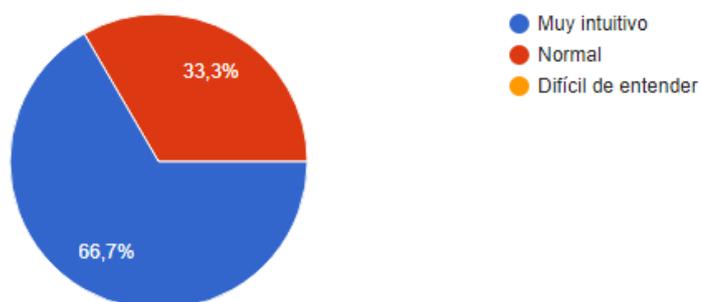
15 respuestas



**Ilustración 106. Pregunta #3 de la encuesta**

Que tan intuitivo le pareció el manual de usuario del portal web administrativo con el nuevo módulo?

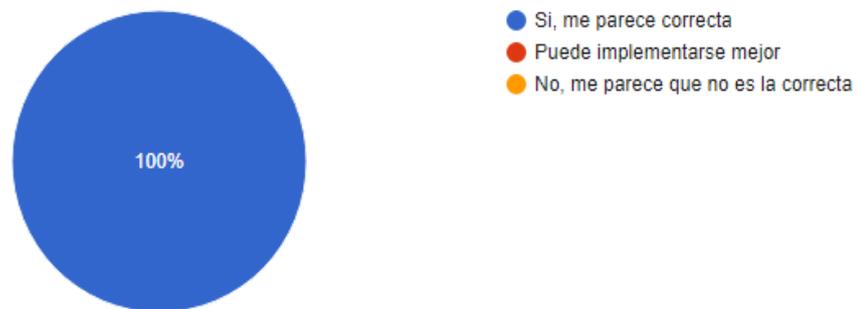
15 respuestas



**Ilustración 107. Pregunta #4 de la encuesta**

La distribución de los reportes le parece correcta e intuitiva?

15 respuestas



**Ilustración 108. Pregunta #5 de la encuesta**

El sistema muestra los errores y la manera de prevenirlos?

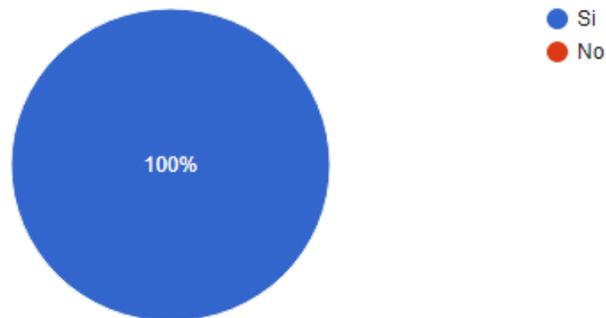
15 respuestas



**Ilustración 109. Pregunta #6 de la encuesta**

El sistema permite que usted como usuario pueda tener control de las acciones que realiza?

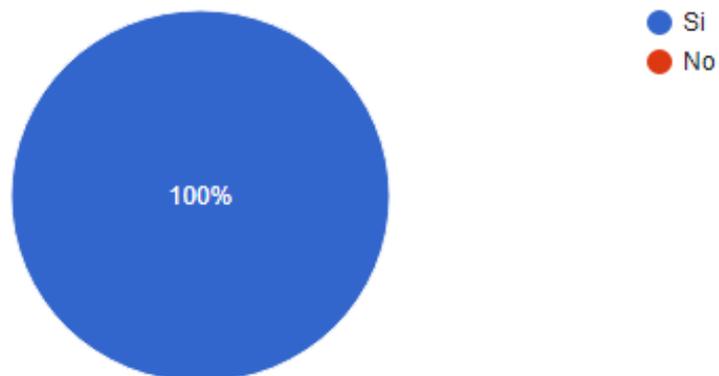
15 respuestas



**Ilustración 110. Pregunta #7 de la encuesta**

El nuevo módulo es de gran utilidad para la organización?

15 respuestas



**Ilustración 111. Pregunta #8 de la encuesta**

#### **4.1.11. Gestión del proyecto**

La gestión de proyecto es un proceso que se ejecuta durante todo el desarrollo del proyecto, lo cual implica la verificación constante que las actividades del proyecto se ejecuten de acuerdo a lo planificado. Se dice entonces, que durante el desarrollo de la solución de Inteligencia de Negocio las actividades se realizaron según lo planificado.

## CONCLUSIONES Y RECOMENDACIONES

Después de haber desarrollado una solución de Inteligencia de Negocio para la generación y presentación de indicadores del proceso de intermediación o subasta de productos relacionados con el gas y petróleo, se logró satisfactoriamente el objetivo del TEG con el fin de apoyar las decisiones de la organización, a través de la obtención de información de calidad basada en los procesos reales del portal web.

Debido a esto, se realizaron estudios sobre los procesos del portal web (subastas, compras, intermediación) planteados por la organización, así como también fueron analizadas sus necesidades con el fin de cubrir de manera satisfactoria la definición de los requerimientos, obteniendo de esta manera como resultado la construcción e implementación de los indicadores que faciliten el control de estos procesos.

El desarrollo de esta solución se planteó bajo la metodología de Ralph Kimball, desarrollando cada una de las etapas involucradas en ella, esta metodología se adapta perfectamente para aquellas organizaciones que están empezando a desarrollar una infraestructura de Inteligencia de Negocio, ya que se puede instaurar una solución de negocios completa sobre un área específica y permite que un proyecto pueda salir a producción más rápido.

Las herramientas usadas para la implementación de la solución de Inteligencia de Negocios fueron: Navicat (administrador de base de datos), el sistema manejador de base de datos MySQL, la herramienta de Pentaho denominada "Pentaho Data Integration", y para la visualización de los reportes y tableros de control se utilizó la herramienta analítica "Tableau".

Es importante mencionar que la mayoría de estas herramientas son de software libre (open source) por lo cual la solución de inteligencia de negocios está desarrollada en este tipo de software.

Sin embargo Tableau no es de software libre debido a que tiene un período gratuito sólo por 15 días, normalmente después de este período se tendría que pagar la mensualidad para seguir usando sus servicios en la organización, no obstante a esto en el portal web de Tableau se puede solicitar una licencia académica para poder descargar y usar la aplicación por un período más amplio de tiempo, en este caso la solicitud se otorgó de manera exitosa y se pudo usar la aplicación para fines académicos. Si la organización desea seguir utilizando los servicios de Tableau, se debería optar por pagar un paquete de los mismos.

Para finalizar, se obtuvo como resultado una solución de Inteligencia de Negocio donde se integran datos de una fuente de datos en un almacén de datos, con el objetivo de aprovechar un almacenamiento eficiente de los datos que ayuda a tener acceso a un gran volumen de información de forma rápida y fácil, optimizando así los procesos de análisis para obtener un conjunto de reportes y tableros de control (dashboard) sobre el rendimiento y desempeño de los procesos ejecutados en el portal web teniendo los indicadores de planificación, ejecución y gestión que nos aportan información para un mejor control y seguimiento hacia los mismos.

Aunado a esto, se tiene la capacidad de poder visualizar los procesos más débiles y apoyar a la toma de decisiones para ejecutar las mejores acciones al respecto, así como también corregir errores que se estén presentando en estos procesos, y aumentar el desempeño en general, ganando con todo esto un gran impulso para todas aquellas organizaciones que cuenten con portales web relacionados a gas y petróleo.

Se recomienda a las organizaciones del ámbito petrolero continuar desarrollando estos proyectos debido a que garantizan la escalabilidad operativa y el mejor rendimiento de sus procesos. De igual manera se recomienda mantener una correcta comunicación en las organizaciones, ya que es de vital importancia para lograr mejores objetivos y alcanzar las meta de manera eficiente.

## GLOSARIO DE TÉRMINOS

<b><i>Término</i></b>	<b><i>Significado</i></b>
Dato	Son los hechos que describen sucesos y entidades
Información	Consiste en datos seleccionados y ordenados con un propósito específico.
Información corporativa	Es aquella información una empresa lanza al exterior, de manera que le permita controlar sus canales y sus contenidos.
Sistema	Es la colección de componentes interrelacionados que trabajan conjuntamente para cumplir algún objetivo”.
Inteligencia	Es la capacidad general que posee el individuo para ajustar conscientemente su pensamiento a nuevas y cambiantes exigencias, aplicando y adaptando a las cosas y a los hechos concretos las nociones abstractas y generales.
Negocio	Aquella ocupación que detenta un individuo y que está encaminada a obtener un beneficio de tipo económico.
Indicador	Son herramientas para clarificar y definir, de forma más precisa, objetivos e impactos. Son medidas verificables de cambio o resultado diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas, facilitan el reparto de insumos, produciendo productos y alcanzando objetivos.

<b><i>Término</i></b>	<b><i>Significado</i></b>
Toma de decisiones	Una decisión puede ser descrita como la respuesta a un problema o la elección entre distintas alternativas para conseguir unos objetivos.
Sistema de información	Un sistema de información es un conjunto de elementos interrelacionados con el propósito de prestar atención a las demandas de información de una organización.
Base de Datos	Es una colección de datos persistentes que pueden compartirse e interrelacionarse: Esta visones muy general y enfatiza en la persistencia de los datos.
Sistema gestor de base de datos	Es un conjunto de programas que permiten crear y mantener una Base de datos, asegurando su integridad, confidencialidad y seguridad.
Modelo relacional	Se trata de un modelo teórico matemático que, además de proporcionarnos los elementos básicos de modelado (las relaciones), incluye un conjunto de operadores para su manipulación, sin ambigüedad posible.
Base de datos relacional	Una base de datos relacional es un conjunto de una o más tablas estructuradas en registros (líneas) y campos (columnas), que se vinculan entre sí por un campo en común, en ambos casos posee las mismas características como por ejemplo el nombre de campo, tipo y longitud; a este campo generalmente se le denomina ID, identificador o clave.

## REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES

Imhoff & Galletto, Mastering Data Warehouse Design: Relational and Dimensional Techniques, Wiley Publishing, 2003

Inmon, Building the Data Warehouse, (Third Edition). John Wiley & Sons, 2002  
Cuadernos de la Facultad n. 5, 2010 71

Kimball & Caserta, The Data Warehouse ETL Toolkit, Indianapolis, Wiley, 2004.

Kimball & Merz, The Data Warehouse Toolkit: Building the WebEnabled Data Warehouse, Wiley, 2000.

Kimball & Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition), New York, Wiley, 2002.

Kimball & Ross, The Kimball Group Reader; Relentlessly Practical Tools for Data Warehousing and Business Intelligence, Indianapolis, Wiley, 2010.

Kimball et al., The Data Warehouse Lifecycle Toolkit. 2nd Edition. New York, Wiley, 2008

Kimball et al., The Data Warehouse Lifecycle Toolkit. New York, Wiley, 1998.

Monografías – Datos Recuperado de:  
<http://www.monografias.com/trabajos14/datos/datos.shtml#ixzz4uach6kLO>

Promo Negocios – Definición de Información. Recuperado de:  
<https://www.promonegocios.net/mercadotecnia/definicion-informacion.html>

UPF- Sistemas de Información. Recuperado de:  
[https://www.upf.edu/hipertextnet/numero-1/sistem\\_infor.html](https://www.upf.edu/hipertextnet/numero-1/sistem_infor.html)

Slides Share- Definiciones de Sistemas de Información. Recuperado de:  
<https://es.slideshare.net/PaolaBorges/definiciones-de-sistemas-informaticos-11740753>

Cursos Gerenciales. Recuperado de: <http://www.cursosgerenciales.org/ineig>

EconLink – Sistemas de Información. Recuperado de:  
<https://www.econlink.com.ar/sistemas-informacion/definicion>

Principios de Sistemas de Información. Recuperado de:  
<https://richardunefa.wordpress.com/principios-de-sistemas-de-informacion/>

Pertutatis – Sistemas de apoyo a los ejecutivos. Recuperado de:  
<http://pertutatis.cat/que-son-los-sistemas-de-apoyo-a-los-ejecutivos-ess/>

Tipos de sistema de información. Recuperado de:  
<https://sanmiranda.weebly.com/sistema-de-informacioacuten>

EUMED LIBROS – Tipos de Sistemas de Información. Recuperado de:  
[http://www.eumed.net/libros-gratis/2012a/1169/tipos\\_de\\_sistema\\_de\\_informacion.html](http://www.eumed.net/libros-gratis/2012a/1169/tipos_de_sistema_de_informacion.html)

Cavsi – Sistema Gestor de Base de datos. Recuperado de:  
<http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-bases-de-datos-o-sgbd/>

Rodcasdatawarehouse -Antecedentes de Datawarehouse. Recuperado de:  
<http://rodcasdatawarehouse.blogspot.com/2013/01/antecedentes.html>

Wikipedia, La enciclopedia Libre – Almacén de Datos. Recuperado de:  
[https://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)

Sinnexus – Datawarehouse. Recuperado de:  
[http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx)

Datawarehouse – Ventajas y Desventajas. Recuperado de:  
<https://datawarehouse.es.tl/Ventajas-y-desventajas.htm>

Todo Technology – Datamar. Recuperado de:  
<http://todotecnology.blogspot.com/2009/09/datamart.html>

Conceptos del modelado Dimensional. Recuperado de:  
<https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>

Wikipedia, La enciclopedia Libre – Tabla dimensión. Recuperado de:  
[https://es.wikipedia.org/wiki/Tabla\\_de\\_dimensi%C3%B3n](https://es.wikipedia.org/wiki/Tabla_de_dimensi%C3%B3n)

Wikipedia, La enciclopedia Libre – Tabla hechos. Recuperado de:  
[https://es.wikipedia.org/wiki/Tabla\\_de\\_hechos](https://es.wikipedia.org/wiki/Tabla_de_hechos)

IBM – Recuperado de:  
[https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_8.1.0/com.ibm.data.tools dimensional.ui.doc/topics/c\\_dm\\_dimschemadesign.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_8.1.0/com.ibm.data.tools dimensional.ui.doc/topics/c_dm_dimschemadesign.html)

[https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_8.1.0/com.ibm.data.tools dimensional.ui.doc/topics/c\\_dm\\_schema\\_starflake.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_8.1.0/com.ibm.data.tools dimensional.ui.doc/topics/c_dm_schema_starflake.html)

Gestiopolis – Inteligencia de Negocios. Recuperado de:  
<https://www.gestiopolis.com/inteligencia-de-negocios-business-intelligence/>

Gestiopolis – Indicadores de Gestión. Recuperado de:  
<https://www.gestiopolis.com/indicadores-de-gestion-que-son-y-por-que-usarlos/>

Decisiones y Tecnologías – Metodología E Implantación de un sistema de BI.  
Recuperado de:  
<https://decisionesytecnologia.wordpress.com/2013/02/07/metodologia-para-el-diseno-e-implantacion-de-un-sistema-de-bi/>

IEBS School - Metodología Scrum. Recuperado de:  
<http://www.iebschool.com/blog/metodologia-scrum-agile-scrum/>

La importancia de las metodologías ágiles. Recuperado de:  
<https://sq.com.mx/buzz/la-importancia-las-metodologias-agiles#.WhdNF2hSxPY>