

Solución Business Intelligence para la empresa Nutribox

Nicolás Miguel Ayut González
Grado de Ingeniería Informática
75.650 - TFG - Business Intelligence

Xavier Martinez Fontes
Atanasi Daradoumis Haralabus

06/2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Solución Business Intelligence para la empresa Nutribox</i>
Nombre del autor:	<i>Nicolás Miguel Ayut González</i>
Nombre del consultor/a:	<i>Xavier Martinez Fontes</i>
Nombre del PRA:	<i>Atanasi Daradoumis Haralabus</i>
Fecha de entrega (mm/aaaa):	06/2019
Titulación:	<i>Grado de Ingeniería Informática</i>
Área del Trabajo Final:	<i>75.650 - TFG - Business Intelligence</i>
Idioma del trabajo:	<i>Castellano.</i>
Palabras clave	<i>Cupones, Healthy discounts, Pentaho, ETL, Cuadro de mando, OLAP, Business Intelligence</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>Este trabajo de fin de grado trata sobre el diseño e implementación de un prototipo inicial de BI para la empresa Nutribox, una empresa creada para gestionar la actividad de una plataforma (aplicación móvil y web service) basada en “healthy discounts”.</p> <p>Esta organización surge a partir del programa piloto “TFG emprendedor interestudios”, por medio del cual, se propone crear una iniciativa emprendedora, a partir, de los trabajos de fin de grado de los estudiantes de diferentes áreas.</p> <p>Así este prototipo pretende dotar a la dirección ejecutiva de esta entidad de información relativa a la consecución de objetivos establecidos inicialmente, por medio de cuadros de mando e informes que permitirán valorar la evaluación del negocio y una toma de decisiones en base a la información obtenida.</p> <p>En este trabajo se detallan las fases de análisis, diseño e implantación de la solución. El prototipo creado consta de procesos de ETL, sistema de almacenamiento y gestión de datos (data warehouse), cubos OLAP, así como cuadros de mando e informes que permiten valorar los indicadores definidos en la fase de análisis.</p> <p>Finalmente, en el capítulo “Líneas de futuro” se exponen una serie de mejoras o ampliaciones a instaurar en la solución que permitirán estudiar nuevos indicadores y realizar nuevos análisis que dotarán a la empresa de un mayor conocimiento de su actividad.</p>	

En definitiva, se puede considerar que se ha construido un prototipo de solución BI para la empresa Nutribox que cumple los objetivos propuestos.

Abstract (in English, 250 words or less):

This final project deals with the design and implementation of an initial BI prototype for Nutribox, a company created to manage the activity of a platform (mobile application and web service) based on "healthy discounts".

This organization arises from the program "TFG emprendedor interestudios", by means of which it is proposed to create an enterprising initiative, from the final projects of students from different areas.

Therefore, this prototype aims to provide the executive management of this company the information related to the achievement of initially established objectives, through dashboards and reports, which will allow evaluating the business and making decisions based on the information obtained.

In this work the phases of analysis, design and implementation of the solution are detailed. The created prototype consists of ETL processes, storage system and data management (data warehouse), OLAP cubes, as well as dashboards and reports that allow to assess the indicators defined in the analysis phase.

Finally, in the chapter "Líneas de futuro" a series of improvements or extensions are exposed to be introduced in the solution that will allow studying new indicators and carry out new analyzes that will give the company a greater knowledge of its activity.

In conclusion, it can be considered that a prototype of BI solution has been built for the Nutribox company that agrees with the proposed objectives.

Índice

1.	Introducción	1
1.1.	Contexto y justificación del Trabajo	1
1.2.	Objetivos del Trabajo.....	3
1.3.	Planificación del Trabajo.....	4
1.4.	Breve resumen de productos obtenidos	7
1.5.	Breve descripción de los siguientes capítulos de la memoria.....	7
2.	Análisis	7
2.1.	Definición de fuentes y tipos de datos	8
2.2.	Definición de KPI	11
3.	Diseño.....	21
3.1.	Base de datos operativa Nutribox y datos de prueba.	22
3.2.	Componentes de la solución.	22
3.2.1.	Data warehouse	23
3.2.2.	Procesos ETL.....	25
3.2.3.	Cubo OLAP	25
3.2.4.	Diseño de dashboard o cuadro de mandos.....	27
3.2.4.1.	CM1: Cuadro de mando Administrativo Nutribox.....	27
3.2.4.2.	CM2: Cuadro de mando para empresa colaboradora.....	31
3.2.5.	Reports.....	34
3.2.5.1.	informe de canjes por categoría y año	35
3.2.5.2.	Informe top colaboradores del año	36
3.2.5.3.	Informe Productos canjeados del año.....	37
3.3.	Selección del software.....	38
3.3.1.	Selección del Software de virtualización	38
3.3.2.	Selección del sistema operativo utilizado en el prototipo	38
3.3.3.	Selección del sistema gestor de bases de datos.....	38
3.3.3.1.	MySQL.....	39
3.3.3.2.	PostgreSQL	39
3.3.3.3.	Elección de sistema gestor de base de datos.....	40
3.3.4.	Selección de Suite Business Intelligence	40
4.	Implantación.....	41
4.1.	Descarga e instalación de componentes necesarios para desarrollar la solución.	41
4.1.1.	Software de virtualización y sistema operativo.....	41
4.1.2.	Sistema gestor de bases de datos	42
4.1.3.	Instalación de Java.....	42
4.1.4.	Descarga e instalación de los componentes necesarios de la suite BI Pentaho.	42
4.2.	Creación de las bases de datos Nutribox y Datawarehouse	45
4.3.	Creación de la data set de pruebas.....	45
4.4.	Construcción de ETL	45
4.4.1.	Transformación HechosCanje.ktr	47
4.4.2.	Transformación DimTiempo.ktr	47
4.4.3.	Transformación DimProductos.ktr	48
4.4.4.	Transformación DimColaboradores.ktr.....	49

4.4.5.	Transformación DimUsuarios.ktr	50
4.5.	Cubos OLAP.....	51
4.6.	Creación de cuadros de mando.....	53
4.6.1.	CM1: Cuadro de mando para la dirección ejecutiva de Nutribox	53
4.6.2.	CM2: Cuadro de mando empresa colaboradora.	62
4.7.	Creación de informes dinámicos en Pentaho Report Designer.	69
4.7.1.	Informe de canjes por categoría y año	70
4.7.2.	Informe Top 10 colaboradores del año.....	71
4.7.3.	Informe Productos canjeados del año	73
5.	Líneas de futuro	74
6.	Conclusiones	76
7.	Bibliografía	77
8.	Anexos.....	78
8.1.	Descarga y uso del prototipo.	78
8.2.	Uso de generatedata.com para crear los datos de prueba	83
8.3.	Procesos ETL con Pentaho Data Integration.....	86
8.4.	Definición del esquema OLAP	93
8.5.	Análisis de los datos del cubo OLAP con jpivot.....	100
8.6.	Análisis de los datos del cubo OLAP con saiku.....	103
8.7.	Creación de cuadros de mando con Pentaho CDE Dashboard.....	105
8.8.	Creación de informes con Pentaho Report Designer.....	108

Lista de Imágenes

Imagen 1: Diagrama de Gantt del proyecto.....	6
Imagen 2: Estructura de la base de datos "Nutribox"	8
Imagen 3: Diseño de la infraestructura BI Nutribox.	22
Imagen 4: Diagrama del Data warehouse de Nutribox.....	24
Imagen 5: Ejemplo de Cubo OLAP	26
Imagen 6: Diseño cuadro de mando para la dirección ejecutiva de Nutribox... ..	27
Imagen 7: Diseño cuadro de mando para empresas colaboradoras.	31
Imagen 8: Iniciar pentaho al encender la máquina virtual.	44
Imagen 9: Transformación HechosCanje.ktr	47
Imagen 10: Transformación DimTiempo.ktr	47
Imagen 11: Transformación DimProductos.ktr	48
Imagen 12: Transformación DimColaboradores.ktr	49
Imagen 13: Transformación DimUsuarios.ktr	50
Imagen 14: Schema OLAP Nutribox	51
Imagen 15: Cuadro de mandos Nutribox para el año 2019.....	53
Imagen 16: Imagen cuadro de mandos Nutribox 2020.....	54
Imagen 17: Cuadro de mando para la empresa Papa Arrugada S.L.....	62
Imagen 18: Cuadro de mando para la empresa Cancajos S.L. del año 2020. .	63
Imagen 19: Importar servicio virtualizado.	78
Imagen 20: Indicar la ruta del archivo.	79
Imagen 21: Sistema virtualizado.	79
Imagen 22: Proceso de importación.....	80
Imagen 23: Iniciar máquina virtual.....	80
Imagen 24: Acceder al servidor Pentaho Business Analytics.....	81
Imagen 25: Detalle Pentaho Business Analytics.	81
Imagen 26: Browse files en Pentaho Business Analytics.	82
Imagen 27: Generar datos de usuarios.	83
Imagen 28: Generar datos de cupones.	84
Imagen 29: Generar transacciones.	85
Imagen 30: Nueva Transformación PDI.	86
Imagen 31: Otra Nueva transformación.	86
Imagen 32: Área de trabajo PDI.	87
Imagen 33: Detalle del objeto table input para la tabla couponUse.....	87
Imagen 34: Detalle del objeto table input para la tabla coupons.	88
Imagen 35: Detalle del objeto Sort rows.....	89
Imagen 36: Detalle del objeto Merge join.	89
Imagen 37: Detalle objeto Calculator	90
Imagen 38: Detalle objeto Split fields.	90
Imagen 39: Detalle objeto String operations.	91
Imagen 40: Detalle objeto Calculator2.	91
Imagen 41: Detalle objeto Combination lookup/update.	92
Imagen 42: Trabajo HechosCanjes.	93
Imagen 43: Área de trabajo Pentaho Schema Workbench.	93
Imagen 44: Detalle conexión con la base de datos.	94
Imagen 45: Área de trabajo Pentaho Schema Workbench 2.	94
Imagen 46: Dimensiones del cubo.	95
Imagen 47: Detalle nivel colaborador.	96

Imagen 48: Detalle nivel Año.....	96
Imagen 49: Detalle nivel Mes.	97
Imagen 50: Añadir uso de dimensión al cubo.....	97
Imagen 51: Añadir medida al cubo.....	98
Imagen 52: Detalle del esquema final.	98
Imagen 53: Publicar el esquema.....	99
Imagen 54: Crear nuevo.....	100
Imagen 55: Crear nueva vista JPivot.....	100
Imagen 56: Medidas y jerarquías.	101
Imagen 57: Ejemplo Jpivot tabla.	101
Imagen 58: Gráfico generado con Jpivot.....	102
Imagen 59: Crear nuevo.....	103
Imagen 60: Seleccionar cubo.	103
Imagen 61: Área de trabajo Saiku.....	104
Imagen 62: Vista en tabla de consulta con Saiku.....	104
Imagen 63: Gráfica generada con Saiku.	105
Imagen 64: Crear nuevo CDE Dashboard.....	105
Imagen 65: CDE Dashboard, detalle "Layout Panel"	106
Imagen 66: CDE Dashboard, detalle " Datasources Panel "	106
Imagen 67: CDE Dashboard, detalle "Component Panel"	107
Imagen 68: CDE Dashboard, guardar cuadro de mando.	107
Imagen 69: Nuevo reporte.....	108
Imagen 70: Crear nuevo reporte con el asistente.....	109
Imagen 71: Seleccionar apariencia.	109
Imagen 72: Fuentes de datos y consultas.....	110
Imagen 73: Tipo de conexión.	110
Imagen 74: Pentaho Analysis Data Source.....	111
Imagen 75: Consultas.	111
Imagen 76: Detalle final de la conexión.....	112
Imagen 77: Previsualización de la consulta	112
Imagen 78: Conexión creada en el asistente.	113
Imagen 79: Selector de campos.....	113
Imagen 80: Previsualización del informe.....	114
Imagen 81: Ajuste de formato del informe.....	115
Imagen 82: Segunda previsualización del informe.....	115
Imagen 83: Edición desde el área del trabajo.	116
Imagen 84: Tercera Previsualización.	116
Imagen 85: Insertar gráfico.....	117
Imagen 86: Otra previsualización.	117
Imagen 87: Añadir consulta.....	118
Imagen 88: Añadir parámetro.....	118
Imagen 89: Editar parámetro.....	119
Imagen 90: Modificar consulta anterior.	119
Imagen 91: Se ha modificado la cabecera.	120
Imagen 92: Informe final Vista del año 2018.	120
Imagen 93: Informe final Vista del año 2019.	121
Imagen 94: Publicar el informe.....	121
Imagen 95: Detalles de la conexión.	122
Imagen 96: Datos de la publicación.	122
Imagen 97: Confirmación reporte publicado.....	122

Lista de tablas

Tabla 1: Planificación inicial.....	5
Tabla 2: Tabla users (base de datos operativa Nutribox)	9
Tabla 3: Tabla collaborators (base de datos operativa Nutribox)	9
Tabla 4: Tabla products (base de datos operativa Nutribox)	10
Tabla 5: Tabla coupons (base de datos operativa Nutribox)	10
Tabla 6: Tabla couponUse (base de datos operativa Nutribox).....	10
Tabla 7: KPI 1 Popularidad del producto.....	11
Tabla 8: KPI 2 Tendencia Producto	12
Tabla 9: KPI 3 Porcentaje de canjes por categoría de productos.....	13
Tabla 10: KP I4 Popularidad del Colaborador	13
Tabla 11:KPI 5 Tendencia Colaborador	14
Tabla 12: KPI 6 Tendencia Canjes anuales.	14
Tabla 13: KPI 7 Total canjes.	15
Tabla 14: KPI 8 Usuarios registrados.....	15
Tabla 15: KPI 9 Usuarios activos.	16
Tabla 16: KPI 10 Tendencia Usuarios activos.....	17
Tabla 17:KP 11 Tendencia de clientes del colaborador.	17
Tabla 18: KPI 12 Actividad del cliente.....	18
Tabla 19: KPI 13 Tendencia en la actividad del cliente.	18
Tabla 20: KPI 14 Usuarios activos por colaborador.	19
Tabla 21: KPI 15 Popularidad del producto de colaborador	20
Tabla 22: KPI 16 Tendencia de canjes de producto por colaborador	20
Tabla 23: KPI 17 Tendencia Canjes anuales por colaborador.....	21
Tabla 24: KPI 18 Total canjes anuales por colaborador.....	21
Tabla 25: Diseño CM1 Componente1 “Número de usuarios registrados”.....	28
Tabla 26:Diseño CM1 Componente 2 “Número de usuarios activos”	28
Tabla 27: Diseño CM1 Componente 3 “Actividad mensual de los usuarios”	28
Tabla 28: Diseño CM1 Componente 4: “Top Colaboradores”.....	29
Tabla 29: Diseño CM1 Componente 5 “Por categoría”.....	29
Tabla 30: Diseño CM1 Componente "Top Productos".....	30
Tabla 31: Diseño CM1 Componente 7 "Canjes anuales"	30
Tabla 32:Diseño CM1 Componente 8 "Número de canjes realizados por mes"	31
Tabla 33: Diseño CM2 Componente 1 “Clientes en el ejercicio”	32
Tabla 34: Diseño CM2 Componente 2 “Actividad mensual de nuestros clientes”.....	32
Tabla 35: Diseño CM2 Componente 3: “Top Clientes”.....	33
Tabla 36: Diseño CM2 Componente "Top Productos".....	33
Tabla 37: Diseño CM2 Componente 5 "Canjes anuales"	34
Tabla 38:Diseño CM2 Componente 6 "Número de canjes realizados por mes"	34
Tabla 39: Descripción Informe Canjes por categoría.....	35
Tabla 40: Informe Top colaboradores del año.....	36
Tabla 41: Informe Productos canjeados del año.	37
Tabla 42: Procesos ETL generados con Pentaho Data Integration	46
Tabla 43: CM1 Componente 1 “Selector de año”	55
Tabla 44: Implantación CM1 Componente 2 “Actividad mensual de usuarios”	56
Tabla 45: Implantación CM1 Componente 3 “Usuarios totales”	56
Tabla 46: Implantación CM1 Componente 4 “Usuarios activos”	57
Tabla 47: Implantación CM1 Componente 5 “Tabla empresas colaboradoras”	58
Tabla 48: Implantación CM1 Componente 6 “Canjes por categoría”	59
Tabla 49: Implantación CM1 Componente 7 “Tabla de productos”.....	60
Tabla 50: Implantación CM1 Componente 8 “Canjeados al año”	61
Tabla 51: Implantación CM1 Componente 9 “Canjes por mes”	61
Tabla 52: Implantación CM2 Componente 1 “Selector de año”	64
Tabla 53: Implantación CM2 Componente 2 “Selector de empresa colaboradora”	64

Tabla 54: Implantación CM2 Componente 3 “Clientes en el ejercicio”	65
Tabla 55: Implantación CM2 Componente 4 “Actividad mensual de los clientes”	66
Tabla 56: Implantación CM2 Componente 5 “Top Clientes”	67
Tabla 57: Implantación CM2 Componente 6 “Top productos”	68
Tabla 58: Implantación CM2 Componente 7 “Canjes anuales”	68
Tabla 59: Implantación CM2 Componente 8 “Número de canjes realizados al mes.”..	69
Tabla 60: Informe de canjes por categoría y año	71
Tabla 61: Informe Top 10 colaboradores del año.....	72
Tabla 62: Informe Productos canjeados del año	74

1. Introducción

1.1. Contexto y justificación del Trabajo

Nutribox es una empresa creada para gestionar la actividad de una plataforma (aplicación móvil y web service) basada en “healthy discounts”. Esta empresa, trata de conseguir acuerdos con mercados locales (para promover la compra en estos pequeños comercios), centrándose en el descuento en comida saludable.

De esta forma las empresas colaboradoras se aprovecharán de este servicio para ofrecer sus productos a un mayor número de clientes y obtener una mayor visibilidad en internet, y por otro lado los usuarios tendrán a su disposición de una gran cantidad de ofertas en un único portal.

Esta organización surge a partir del programa piloto “TFG emprendedor interestudios”, por medio del cual, se propone de crear una iniciativa emprendedora, en base a los trabajos de fin de grado de los estudiantes de diferentes áreas, en este caso dos estudiantes del grado en administración y dirección de empresas y un estudiante del grado de ingeniería informática.

La participación en este programa “TFG emprendedor interestudios” supone una mayor complejidad a este proyecto debido a que no sólo se deberá tratar el área del trabajo de fin de grado, sino que será necesario un proceso de definición empresarial y un trabajo conjunto del equipo de emprendedor para desarrollar esta iniciativa.

Del mismo modo, hay que tener en cuenta que cada integrante del grupo deberá llevar a cabo un trabajo individual sobre su área. Para el caso de este trabajo sobre Business Intelligence al tratarse de una iniciativa a desarrollar no se tienen referencias sobre posibles necesidades, oportunidades u objetivos a valorar en este trabajo.

Así pues, con la definición de servicio o plataforma, se ha planteado la necesidad de integrar soluciones Business Intelligence (BI), para tratar la información generada por la propia plataforma a través de su funcionamiento y de otros datos externos (como podrían ser los extraídos de redes sociales, IoT, Open Data...), con la finalidad de extraer y compartir el conocimiento que pueda aportar valor a la solución.

Este TFG pretende cubrir esta necesidad, diseñando e implantando un prototipo inicial de BI para esta empresa; que permita dotar a la dirección ejecutiva de información relativa a la consecución de objetivos establecidos inicialmente por medio de cuadros de mando, informes y/o

reports que permitirán valorar la evaluación del negocio y la toma de decisiones en base a la información obtenida.

Con el objetivo de obtener cuadros de mando e informes que representen esta información, este prototipo inicial necesita la definición y producción de procesos de ETL, sistema de almacenamiento y gestión de datos (data warehouse), y un cubo OLAP que permita realizar valoraciones a los indicadores inicialmente establecidos.

Estos indicadores o KPI (del inglés, key performance indicator) se definen durante la realización de este trabajo, ya que actualmente, la plataforma Nutribox se encuentra en una fase idea, sin ningún prototipo desarrollado.

Por esta razón, en este TFG, se deberá considerar qué datos dispondrá la futura plataforma y cómo se tratarán, estudiando posibles técnicas y herramientas, con la finalidad de obtener una solución adecuada; y, por lo tanto; la definición de los KPI dependerá del modelado de datos inicial de la plataforma.

En consecuencia, el conjunto de indicadores se incrementará en posteriores ampliaciones del sistema, en base al análisis de las nuevas dimensiones del negocio, fuentes de datos, objetivos corporativos, etc; ya que este prototipo se pretende que sirva, como base para la solución BI de la empresa y su plataforma; que irá evolucionando, ampliando sus características y funcionalidades, en base a las necesidades del negocio.

Teniendo en cuenta que la plataforma de la empresa Nutribox se encuentra en un estado inicial o de definición empresarial, para la realización del prototipo se deberá generar un dataset ficticio con los que mostrar la funcionalidad de este prototipo.

Debido a la falta de financiación inicial del proyecto, se establece, el uso de soluciones basadas en Open Source necesarias para la implantación de un modelo conceptual, con la finalidad de ahorrar en el coste de licencias. Para futuras actualizaciones se podrá valorar el uso de otras soluciones, en base a nuevas necesidades o carencias que detectadas en el sistema.

1.2. Objetivos del Trabajo

Se han establecido los principales objetivos de este TFG:

- Valorar las posibles fuentes de datos del servicio/plataforma. En este trabajo se describe la base de datos operacional de Nutribox sobre la que se construye el prototipo inicial y posteriormente en el apartado de líneas de futuro se propondrán nuevas fuentes de datos, que podrá utilizar la solución BI de la organización.
- Definir KPI iniciales, que servirán para evaluar los resultados de la empresa. En esta fase inicial del servicio estos indicadores estarán destinados principalmente a la medición de aspectos como podrían ser el número de usuarios, popularidad de los diferentes productos ofrecidos desde la plataforma o diferentes tendencias.
- Diseñar e implantar un prototipo inicial de BI para esta plataforma, cuya finalidad será obtener reports/informes y cuadro de mando; con los que valorar los resultados de la empresa a partir de los KPI definidos en esta fase inicial del prototipo.
- Exponer las posibles mejoras, ampliaciones y nuevas funcionalidades a implementar en el prototipo. En el apartado de Líneas de futuro se expondrán posibles opciones como el uso de nuevas fuentes de datos o herramientas analíticas que mejoren y amplíen las capacidades del sistema.

Para cumplir los objetivos establecidos se definen los siguientes aspectos en el alcance del proyecto:

- Modelado de datos: Puesto que, como se ha explicado en la descripción del proyecto, la plataforma se encuentra en la fase de idea, sin siquiera, un prototipo, se hace necesario realizar un modelado de datos.

Para el prototipo inicial, que se tratará en este TFG, sólo se considerará la base de datos del servicio o plataforma; no se estudiarán otras fuentes de datos que se podrán plantear para futuras evoluciones de la solución BI de la organización.

- El proyecto incluye la definición y construcción de los sistemas de extracción, transformación, carga (ETL), construcción de un sistema de almacenamiento y gestión de datos (data warehouse), diseño de cubos OLAP, así como el diseño y construcción de cuadros de mando e informes que presenten los indicadores (KPI) que se definen en fases anteriores.

1.3. Planificación del Trabajo

En este proyecto se pueden describir los siguientes hitos:

- Definición y planificación del proyecto.
- Análisis y diseño conceptual.
 - Selección de técnicas y herramientas a utilizar para la creación de la solución.
 - Definición de fuentes y tipos de datos a tratar.
 - Definición de KPI.
 - Diseño de visualizaciones.
- Construcción del prototipo.
 - Creación de dataset de pruebas.
 - Construcción de los sistemas de extracción, transformación, carga (ETL).
 - Construcción de un sistema de almacenamiento y gestión de datos (data warehouse).
 - Diseño y creación de cubos OLAP (online analytical processing).
 - Creación de prototipo de cuadros de mando e informes.
- Cierre del proyecto
 - Propuesta de posibles evoluciones mejoras, ampliaciones y nuevas funcionalidades a implementar en el prototipo.
 - Redactar de informe de cierre.
 - Finalizar entregables (memoria del proyecto, producto, video presentación, informe de autoevaluación...).

A partir de los hitos que se han definido anteriormente y teniendo en cuenta las fechas de las entregas de las diferentes PEC se establece la siguiente planificación para desarrollar este trabajo:

Hito	Duración	Comienzo	Fin
PEC1	16 días	lun 25/02/19	lun 18/03/19
Definición y planificación del proyecto.	16 días	lun 25/02/19	lun 18/03/19
PEC2	20 días	mar 19/03/19	lun 15/04/19
Selección de técnicas y herramientas a utilizar	2 días	mar 19/03/19	mié 20/03/19
Definición de fuentes y tipos de datos	5 días	jue 21/03/19	mié 27/03/19
Definición de KPI	3 días	jue 28/03/19	lun 01/04/19
Creación de dataset de pruebas	5 días	mar 02/04/19	lun 08/04/19
Construcción de ETL	5 días	mar 09/04/19	lun 15/04/19
PEC3	25 días	mar 16/04/19	lun 20/05/19
Construcción de datawarehouse	4 días	mar 16/04/19	vie 19/04/19
Diseño y creación cubos OLAP	6 días	sáb 20/04/19	vie 26/04/19
Diseño de visualizaciones	5 días	sáb 27/04/19	jue 02/05/19
Creación de prototipo cuadro de mando	8 días	vie 03/05/19	mar 14/05/19
Creación de reports	4 días	mié 15/05/19	lun 20/05/19
Entrega Final	27 días	mar 21/05/19	lun 17/06/19
Propuesta de mejoras, ampliaciones y nuevas funcionalidades a implementar	9 días	mar 21/05/19	jue 30/05/19
Finalizar entregables	17 días	jue 30/05/19	lun 17/06/19

Tabla 1: Planificación inicial

Dicha planificación se puede representar en el siguiente diagrama de Gantt:

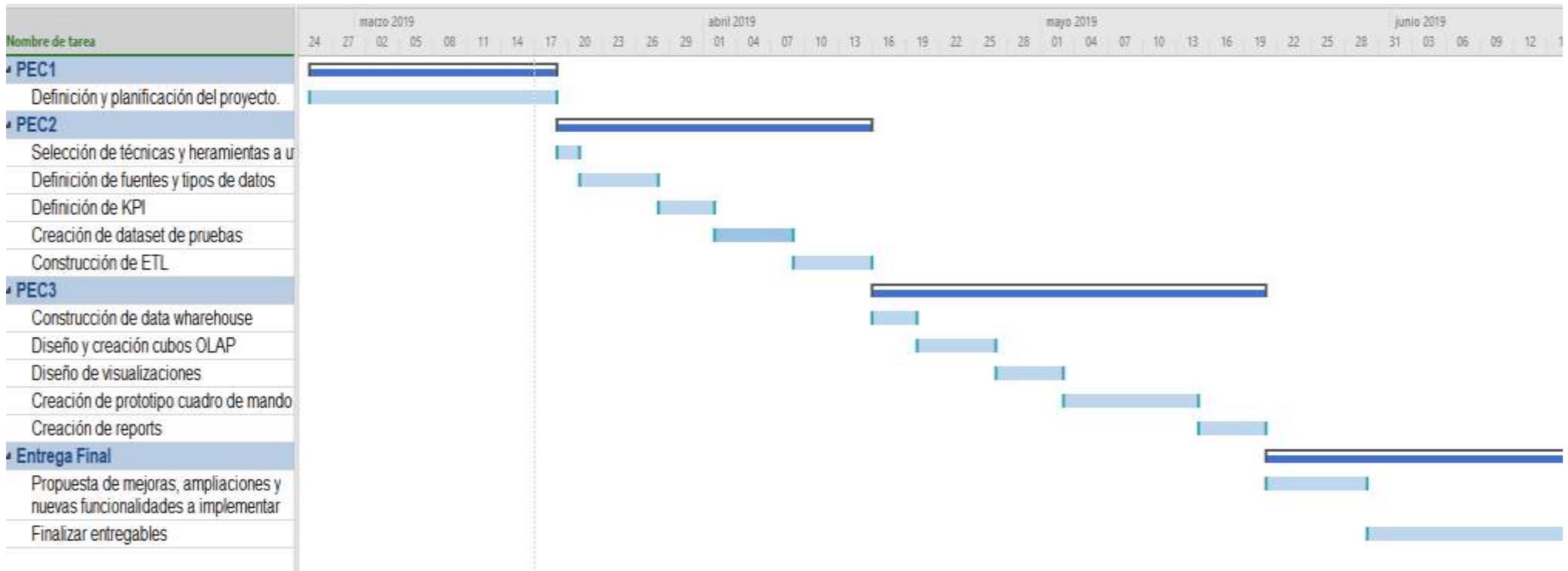


Imagen 1: Diagrama de Gantt del proyecto

1.4. Breve resumen de productos obtenidos

Se han obtenido los siguientes productos:

- Archivo VirtualBox
 - Prototipo.ova
- Archivos con los script sql:
 - Nutribox.sql
 - insertarDatos.sql
 - creaDatawarehouse.sql
- Archivos de Pentaho Data Integration:
 - DimColaboradores.ktr
 - DimProductos.ktr
 - DimTiempo.ktr
 - DimUsuarios.ktr
 - HechosCanje.ktr
- Archivos de Pentaho Report Designer:
 - PorCat.prpt
 - PorProducto.prpt
 - TopColaborador.prpt

1.5. Breve descripción de los siguientes capítulos de la memoria

En esta memoria se detalla todo el proyecto y se estructura en las fases de análisis, diseño e implantación, así como los capítulos de líneas de futuro y conclusión.

2. Análisis

En esta fase de análisis, como la empresa Nutribox se encuentra en una fase inicial o de idea, se realiza un trabajo de definición de las fuentes de datos de la plataforma, así como los KPI o principales indicadores que se valorarán en este prototipo de Business Intelligence.

De esta forma, es necesario que la empresa cuestione cómo el tratamiento de los datos que dispone puede ser usado para alcanzar los objetivos y metas de la organización, es decir, como los datos existentes se convierten en información de utilidad para la toma de decisiones.

Del mismo modo la definición de los indicadores debe ir alineada con las necesidades y cumplimiento de objetivos de la empresa, de forma que estos puedan medir las diferentes acciones definidas y la evolución de la entidad.

2.1. Definición de fuentes y tipos de datos

Como se ha descrito en el alcance del proyecto, para este prototipo inicial sólo se considerará la base de datos del servicio o plataforma como fuente de datos. Esta base de datos la podemos representar con el siguiente diagrama:

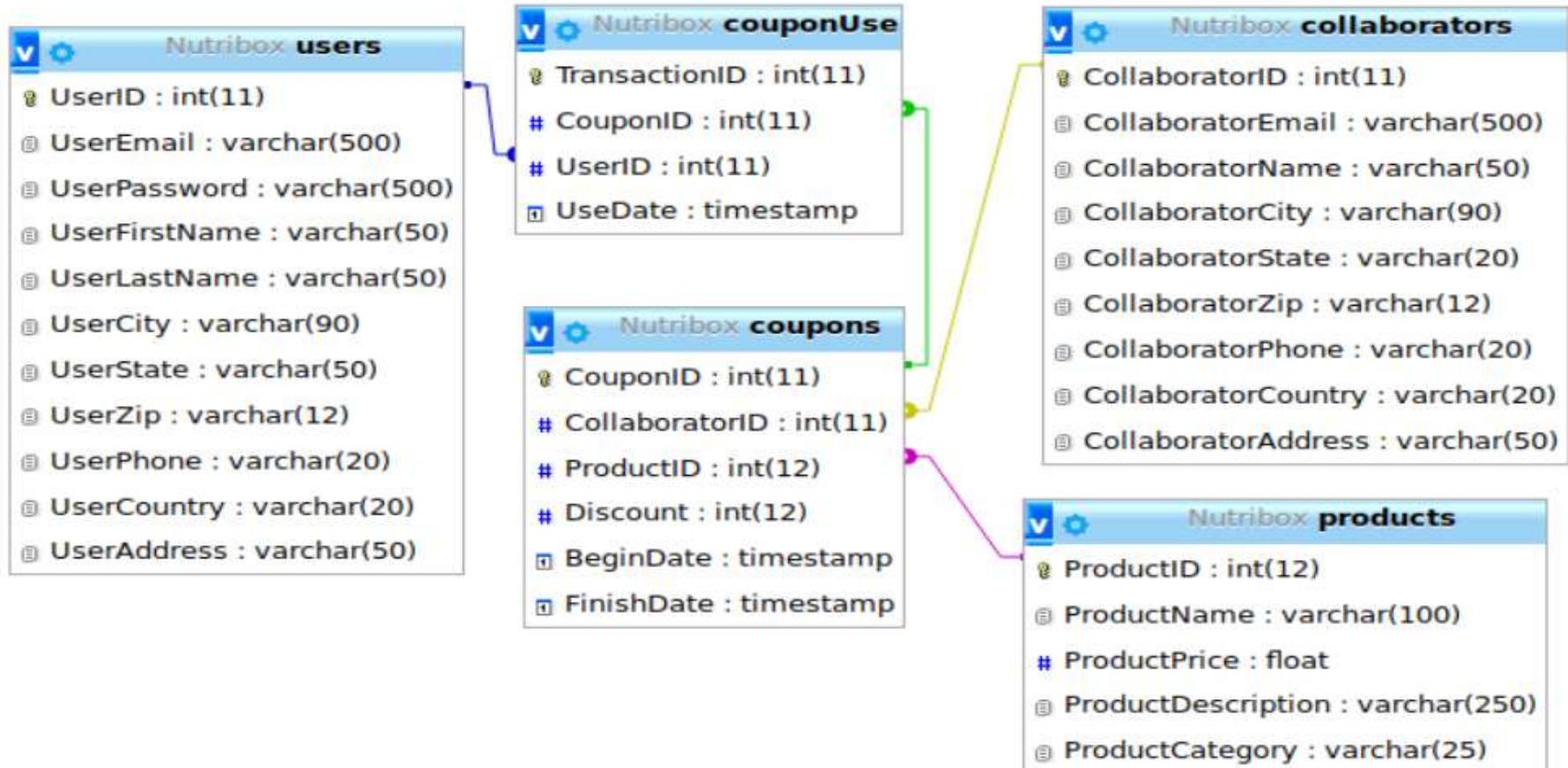


Imagen 2: Estructura de la base de datos "Nutribox"

La tabla users almacena la información de los usuarios de la plataforma y se compone de los siguientes campos:

Tabla users	
UserID	Se trata del identificador de usuario que se generará de forma automática.
UserEmail	En este campo se registra la dirección de correo electrónico del usuario.
UserPassword	En este campo se registra la contraseña del usuario.
UserFirstName	En este campo se registra el nombre del usuario.
UserLastName	En este campo se registra el/los apellidos del usuario.
UserCity	En este campo se registra la ciudad del usuario.
UserState	En este campo se registra la provincia del usuario.
UserZip	En este campo se registra el código postal del usuario.
UserPhone	En este campo se registra el número de teléfono del usuario.
UserCountry	En este campo se registra el País del usuario.
UserAddress	En este campo se registra la dirección del usuario.

Tabla 2: Tabla users (base de datos operativa Nutribox)

La tabla collaborators almacena la información de las empresas colaboradoras de la plataforma y consta de los siguientes campos:

Tabla collaborators	
CollaboratorID	Se trata del identificador de colaborador que se generará de forma automática.
CollaboratorEmail	En este campo se registra el correo electrónico de la empresa colaboradora.
CollaboratorPassword	En este campo se registra la contraseña de la empresa colaboradora en la plataforma.
CollaboratorName	En este campo se registra el nombre de la empresa colaboradora.
CollaboratorCity	En este campo se registra la ciudad de la empresa colaboradora.
CollaboratorState	En este campo se registra la provincia de la empresa colaboradora.
CollaboratorZip	En este campo se registra el código postal de la empresa colaboradora.
CollaboratorPhone	En este campo se registra el número de teléfono de la empresa colaboradora.
CollaboratorCountry	En este campo se registra el país de la empresa colaboradora.
CollaboratorAddress	En este campo se registra la dirección de la empresa colaboradora.

Tabla 3: Tabla collaborators (base de datos operativa Nutribox)

La tabla products almacena la información de los productos que las empresas colaboradoras promocionan a través de la plataforma y consta de los siguientes campos:

Tabla products	
ProductID	Identificador de producto que se generará de forma automática.
ProductName	En este campo se almacena el nombre del producto.
ProductPrice	En este campo se almacena el precio unitario del producto.
ProductDescription	En este campo se registra una breve descripción del producto.
ProductCategory	En este campo se registra la categoría del producto.

Tabla 4: Tabla products (base de datos operativa Nutribox)

La tabla coupons almacena la información acerca de los cupones/ofertas que se ofrecen en la plataforma y consta de los siguientes campos:

Tabla coupons	
CouponID	Identificador del cupón que se generará de forma automática.
CollaboratorID	Identificador del colaborador que registra el cupón.
ProductID	Identificador del producto que se oferta.
Discount	Descuento que se ofrece en este cupón.
BeginDate	Se registra la fecha en la que empieza la oferta.
FinishDate	Se registra la fecha en la que finaliza la oferta.

Tabla 5: Tabla coupons (base de datos operativa Nutribox)

La tabla couponUse almacena los datos de las diferentes transacciones o canjes de cupones y consta de los siguientes campos:

Tabla couponUse	
TransactionID	En este campo se almacena el identificador de la transacción.
CouponID	En este campo se almacenará el id del cupón a canjear.
UserID	En este campo se registra el identificador de usuario que utiliza el cupón.
UseDate	En este campo se registra la fecha en la que se ha utilizado el cupón.

Tabla 6: Tabla couponUse (base de datos operativa Nutribox)

2.2. Definición de KPI

Un KPI (key performance indicator), conocido también como indicador clave, medidor de desempeño o indicador clave de rendimiento, es una medida del nivel del rendimiento de un proceso. El valor del indicador está directamente relacionado con un objetivo fijado previamente.¹

Para la consecución de este proyecto, se definen KPI que permitan valorar el rendimiento de la plataforma Nutribox, relativos a diferentes aspectos como usuarios, empresas colaboradoras, productos o canjes de cupones. Estos indicadores deberán ser actualizados de acuerdo a las necesidades y estrategia de la empresa, así como al desarrollo de la plataforma BI. Inicialmente se definen los siguientes indicadores:

Nombre	KPI 1 Popularidad del producto
Finalidad	Mide la popularidad de un producto de acuerdo al número de canjes de cupones de un producto determinado.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	Nº de canjes de un producto por año.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	20/100. Se ha establecido que, en la fase inicial de la plataforma, cada producto ofertado en la plataforma debería tener entre 20 y 100 canjes anuales.
Iniciativas	-Crear campañas específicas para publicitar este producto. -Búsqueda de colaboradores que ofrezcan este tipo de producto.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 7: KPI 1 Popularidad del producto

¹ Indicador clave de rendimiento (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Indicador_clave_de_rendimiento>

Nombre	KPI 2 Tendencia Producto
Finalidad	Conocer la tendencia de los canjes de cupones de un producto determinado con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	(Canjes del producto/año)-(Canjes del producto/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas específicas para publicitar este producto. -Búsqueda de colaboradores que ofrezcan este tipo de producto.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 8: KPI 2 Tendencia Producto

Nombre	KPI 3 Porcentaje de canjes por Categoría de productos
Finalidad	Conocer el porcentaje canjes de cupones de una categoría determinada de productos con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	(Canjes categoría/canjes totales) *100
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual / Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0

Iniciativas	-Crear campañas específicas para publicitar esta categoría. -Búsqueda de colaboradores que ofrezcan este tipo de productos.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 9: KPI 3 Porcentaje de canjes por categoría de productos.

Nombre	KPI 4 Popularidad del Colaborador
Finalidad	Mide la popularidad de un colaborador de acuerdo al número de canjes de cupones ofertados.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	Nº de canjes de cupones publicados por el colaborador en el año.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	200/20. Se ha establecido que, en la fase inicial de la plataforma, cada colaborador consiga al menos 20 canjes de sus productos para garantizar su continuidad en la plataforma. Se plantea un objetivo inicial de 200 canjes anuales.
Iniciativas	-Crear campañas específicas para publicitar colaboradores que no han cumplido el objetivo mínimo de 20 canjes.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 10: KP I4 Popularidad del Colaborador

Nombre	KPI 5 Tendencia Colaborador
Finalidad	Conocer la tendencia de los canjes de cupones por colaborador con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	(Canjes categoría/año)-(Canjes categoría/año anterior)

Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas específicas para publicitar a cada colaborador. -Ofrecer beneficios específicos a los colaboradores que cumplen objetivos. -Establecimiento de unos requisitos mínimos para los colaboradores de la plataforma.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 11:KPI 5 Tendencia Colaborador

Nombre	KPI 6 Tendencia Canjes anuales
Finalidad	Conocer la tendencia del total de canjes efectuados en la plataforma con respecto al ejercicio anterior (si se han producido más, menos o las mismas transacciones) y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	(Canjes categoría/año)-(Canjes categoría/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Promover los productos menos vendidos por medio de promociones específicas.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 12: KPI 6 Tendencia Canjes anuales.

Nombre	KPI 7 Total canjes
Finalidad	Conocer el total de canjes efectuados en la plataforma en un ejercicio para valorar el volumen de negocio en ese periodo.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	Contar todos los canjes realizados en el periodo a valorar.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual / Mensual / Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Promover los productos menos vendidos por medio de promociones específicas.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 13: KPI 7 Total canjes.

Nombre	KPI 8 Usuarios registrados
Finalidad	El número de usuarios registrados es un indicador que mide el volumen de negocio de la compañía.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de operaciones
Definición/fórmula	Count-distict(UserID)
Proceso relacionado	Registro de clientes, Satisfacción y fidelización de clientes, Atención al cliente.
Frecuencia	Anual
Quien lo mide	Dirección de operaciones
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	2000/ 1500 / 1000
Iniciativas	-Crear campañas de marketing para publicitar la plataforma. -Programa de afiliados.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 14: KPI 8 Usuarios registrados.

Nombre	KPI 9 Usuarios activos
Finalidad	Conocer el número de usuarios únicos activos en el año, este un indicador ayuda a medir el volumen de negocio de la compañía.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox, Departamento de marketing de Nutribox y departamento de operaciones.
Definición/fórmula	Cuenta usuarios que hayan canjeado algún cupón
Proceso relacionado	Registro de clientes, Satisfacción y fidelización de clientes, Atención al cliente.
Frecuencia	Anual / Mensual / Semanal
Quien lo mide	Departamento de operaciones
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	2000/ 1500 / 1000
Iniciativas	-Crear campañas de marketing para publicitar la plataforma. -Programa de afiliados -Ofrecer beneficios específicos a los usuarios con mayor actividad.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 15: KPI 9 Usuarios activos.

Nombre	KPI 10 Tendencia Usuarios activos
Finalidad	Conocer la tendencia de actividad de los usuarios de la plataforma y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva Nutribox / Departamento de marketing de Nutribox
Definición/fórmula	$(\text{usuariosActivos/año}) - (\text{usuariosActivos/año anterior})$
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas de marketing para publicitar la plataforma.

	-Programa de afiliados -Ofrecer beneficios específicos a los usuarios más activos.
Quien lo recibe	Dirección de Marketing, Dirección de finanzas, Dirección ejecutiva de Nutribox.

Tabla 16: KPI 10 Tendencia Usuarios activos

Nombre	KPI 11 Tendencia de clientes del Colaborador
Finalidad	Conocer la tendencia del número de clientes por colaborador con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Operativo
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	(Clientes/año)-(Clientes/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas específicas para publicitar a cada colaborador. -Ofrecer beneficios específicos a los colaboradores que cumplen objetivos. -Establecimiento de unos requisitos mínimos para los colaboradores de la plataforma.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 17:KP 11 Tendencia de clientes del colaborador.

Nombre	KPI 12 Actividad del cliente.
Finalidad	Mide la actividad del cliente de acuerdo al número de canjes de cupones ofertados por un colaborador determinado.
Categoría	Operativo
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	Nº de canjes de productos ofertado por un colaborador.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	2/0.
Iniciativas	-Crear campañas para perfiles específicos de clientes.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 18: KPI 12 Actividad del cliente.

Nombre	KPI 13 Tendencia en la actividad del cliente.
Finalidad	Conocer la tendencia de los canjes de cupones ofrecidos por un colaborador, realizados de un cliente determinado con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	(Canjes del cliente/año)-(Canjes del cliente/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas para perfiles específicos de clientes.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 19: KPI 13 Tendencia en la actividad del cliente.

Nombre	KPI 14 Usuarios activos por colaborador
Finalidad	Conocer el número de usuarios únicos activos que realizan canjes de los productos del colaborador en un mes concreto, este un indicador ayuda a medir el volumen de negocio del colaborador.
Categoría	Operativo
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	Cuenta usuarios que hayan canjeado algún cupón ofertado por la empresa colaboradora.
Proceso relacionado	Registro de clientes, Satisfacción y fidelización de clientes, Atención al cliente.
Frecuencia	Anual / Mensual/ Semanal
Quien lo mide	Departamento de operaciones
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	15/ 3
Iniciativas	-Crear campañas de marketing para publicitar la plataforma. -Programa de afiliados -Ofrecer beneficios específicos a los usuarios con mayor actividad.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 20: KPI 14 Usuarios activos por colaborador.

Nombre	KPI 15 Popularidad del producto de colaborador X
Finalidad	Mide la popularidad de un producto ofertado por un colaborador de acuerdo al número de canjes de cupones de un producto determinado ofrecidos por este colaborador.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	Nº de canjes de un producto ofrecido por un colaborador determinado al año.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual /Mensual /Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.

Meta/Umbral	20/100. Se ha establecido que, en la fase inicial de la plataforma, cada producto ofertado en la plataforma debería tener entre 20 y 100 canjes anuales.
Iniciativas	-Crear campañas específicas para publicitar este producto.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 21: KPI 15 Popularidad del producto de colaborador

Nombre	KPI 16 Tendencia de canjes de producto por colaborador
Finalidad	Conocer la tendencia de los canjes de cupones de un producto determinado ofrecidos por un colaborador con respecto al ejercicio anterior y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	(Canjes del producto/año)-(Canjes del producto/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes.
Frecuencia	Anual/ Mensual/ Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Crear campañas específicas para publicitar este producto.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 22: KPI 16 Tendencia de canjes de producto por colaborador

Nombre	KPI 17 Tendencia Canjes anuales por colaborador
Finalidad	Conocer la tendencia del total de canjes efectuados en la plataforma sobre productos ofrecidos por el colaborador con respecto al ejercicio anterior (si se han producido más, menos o las mismas transacciones) y actuar en consecuencia.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	(Canjes categoría/año)-(Canjes categoría/año anterior)
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores

Frecuencia	Anual
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Promover los productos menos vendidos por medio de promociones específicas.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 23: KPI 17 Tendencia Canjes anuales por colaborador

Nombre	KPI 18 Total canjes por colaborador
Finalidad	Conocer el total de canjes efectuados en la plataforma sobre las ofertas de un colaborador en un ejercicio para valorar el volumen de negocio del colaborador en ese periodo.
Categoría	Estratégico
Responsabilidad	Dirección ejecutiva empresa colaboradora.
Definición/fórmula	Contar todos los canjes realizados en el año de productos de este colaborador.
Proceso relacionado	Promoción de productos, Satisfacción y fidelización de clientes, Atención a colaboradores/proveedores
Frecuencia	Anual/Mensual/Semanal
Quien lo mide	Dirección de Marketing
Fuente de los datos	Base de datos operacional de Nutribox.
Meta/Umbral	<0/0
Iniciativas	-Promover los productos menos vendidos por medio de promociones específicas.
Quien lo recibe	Dirección ejecutiva empresa colaboradora.

Tabla 24: KPI 18 Total canjes anuales por colaborador.

3. Diseño

En esta fase del proyecto se definen las características de la solución a realizar. Para ello es necesario determinar los datos que usará el prototipo en base al análisis realizado en la fase anterior, describir los diferentes componentes de la solución y seleccionar el software necesario para construir el producto final.

3.1. Base de datos operativa Nutribox y datos de prueba.

Como ya se ha indicado en la fase de análisis la empresa Nutribox no está operativa actualmente por lo que es necesario generar unos datos iniciales o de prueba, que permitan mostrar la utilidad del prototipo.

Con este fin en el prototipo se crea una base de datos llamada Nutribox que simulará la base de datos operativa de la plataforma y teniendo en cuenta la estructura de la base de datos descrita en la fase de análisis, se generarán:

- 20 registros en la tabla de empresas colaboradoras.
- 20 registros en la tabla de productos
- 200 registros de cupones
- 1000 registros en la tabla usuarios
- 5000 registros de transacciones o canjes de cupones en la base de datos Nutribox.
-

3.2. Componentes de la solución.

Para obtener los resultados previstos es necesario diseñar e implantar una infraestructura capaz de extraer los datos de la base de datos operacional de Nutribox y que tras su procesamiento permita obtener unos reportes y cuadros de mando o dashboards, capaces de realizar un seguimiento de los principales KPI. Esta infraestructura se describe en las siguientes secciones de esta memoria y la podemos representar en el siguiente diagrama:

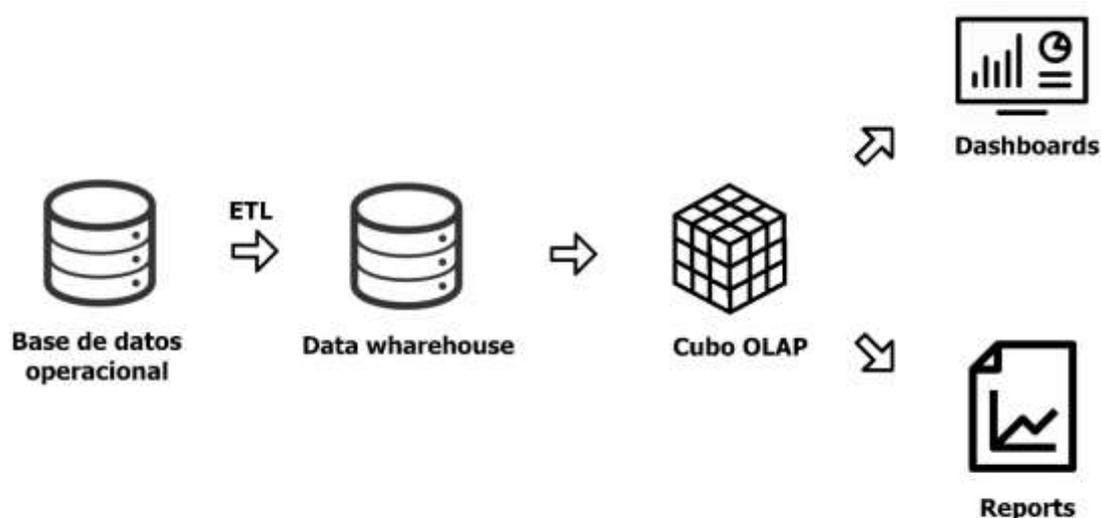


Imagen 3: Diseño de la infraestructura BI Nutribox.

3.2.1. Data warehouse

Un data warehouse, es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se usa para realizar informes (reports) y análisis de datos, siendo un componente fundamental de la inteligencia empresarial.² Los data warehouse se estructuran principalmente siguiendo los esquemas en estrella o en copo de nieve.

Para esta solución se ha escogido el esquema en estrella³. Este esquema separa los datos del proceso de negocios en: hechos y dimensiones. Los hechos contienen datos medibles, cuantitativos, relacionados a la transacción del negocio, y las dimensiones son atributos que describen los datos indicados en los hechos.

Una razón para utilizar los esquemas en estrella es su simplicidad desde el punto de vista del usuario final. Las consultas no son complicadas, ya que las condiciones y las uniones necesarias sólo involucran a la tabla de hechos y a las de dimensiones, no haciendo falta que se encadenen uniones y condiciones a dos o más niveles como ocurriría en un esquema en copo de nieve.

Además, es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

Este datawarehouse se estructurará en una tabla de hechos HechosCanje y las tablas de dimensiones DimProductos, DimUsuarios, DimColaboradores y DimFecha. Como se puede ver representada en el siguiente diagrama:

² Almacén de datos (2019) [En línea] de Wikipedia
[Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos>

³ Esquema en estrella (2019) [En línea] de Wikipedia
[Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Esquema_en_estrella>

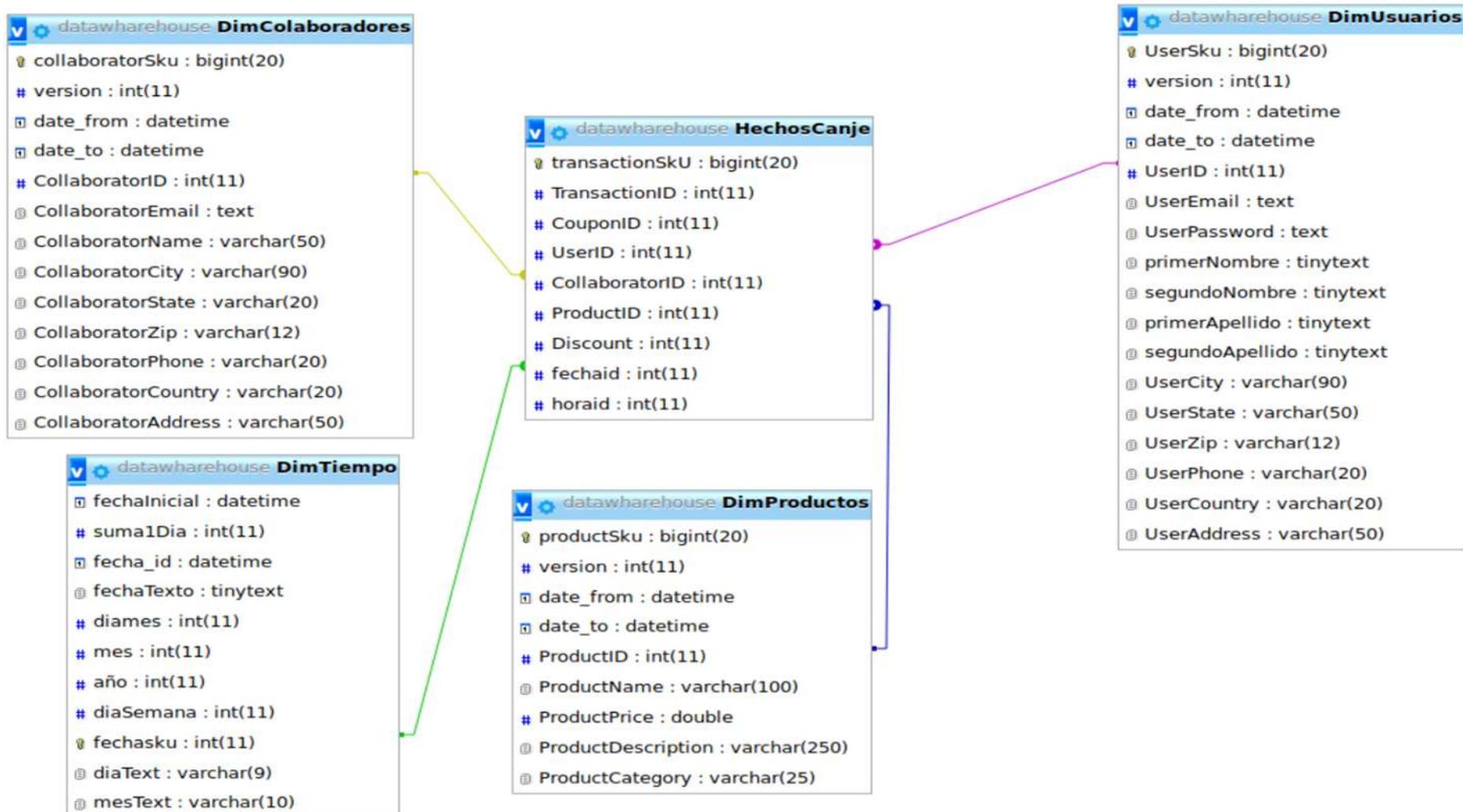


Imagen 4: Diagrama del Data warehouse de Nutribox

3.2.2. Procesos ETL

Una vez definida la estructura del data warehouse, serán necesarios definir los procesos ETL que permitan extraer la información a partir de la base de datos operativa, volcándolos en el data warehouse de la Compañía, tras realizar las operaciones de formato y limpieza que sean necesarios. Para ello, será necesario implementar los siguientes trabajos:

- A partir de los datos incluidos en las tablas CouponUse y coupons de la base de datos operativa Nutribox poblar la tabla HechosCange, tras haber realizado el tratamiento necesario de los datos.
- A partir de los datos incluidos en la tabla products de la base de datos operativa Nutribox poblar la tabla DimProducts, tras haber realizado el tratamiento necesario de los datos.
- A partir de los datos incluidos en la tabla users de la base de datos operativa Nutribox poblar la tabla DimUsuarios, tras haber realizado el tratamiento necesario de los datos.
- A partir de los datos incluidos en la tabla collaborators de la base de datos operativa Nutribox poblar la tabla DimColaboradores, tras haber realizado el tratamiento necesario de los datos.
- Genere los datos necesarios para crear la dimensión tiempo y su carga en la tabla DimTiempo del data warehouse.

3.2.3. Cubo OLAP

OLAP es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Su objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras de datos diversas, normalmente multidimensionales (o Cubos OLAP), que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP).⁴

En este caso se podrán definir diferentes cubos incluyendo las dimensiones colaboradores, usuarios, tiempo y productos, y diferentes métricas que ayuden a valorar los diferentes indicadores que se han establecido.

⁴ OLAP (2019) [En línea] de Wikipedia
[Consulta: 1 de Junio de 2019] <<https://es.wikipedia.org/wiki/OLAP>>

Por ejemplo, se podría generar un cubo con las dimensiones tiempo, usuarios y producto de forma que cada celda contendrá los canjes de ofertas del productoX realizados por el UsuarioX en un mes indicado. Este cubo se representa en el siguiente diagrama:

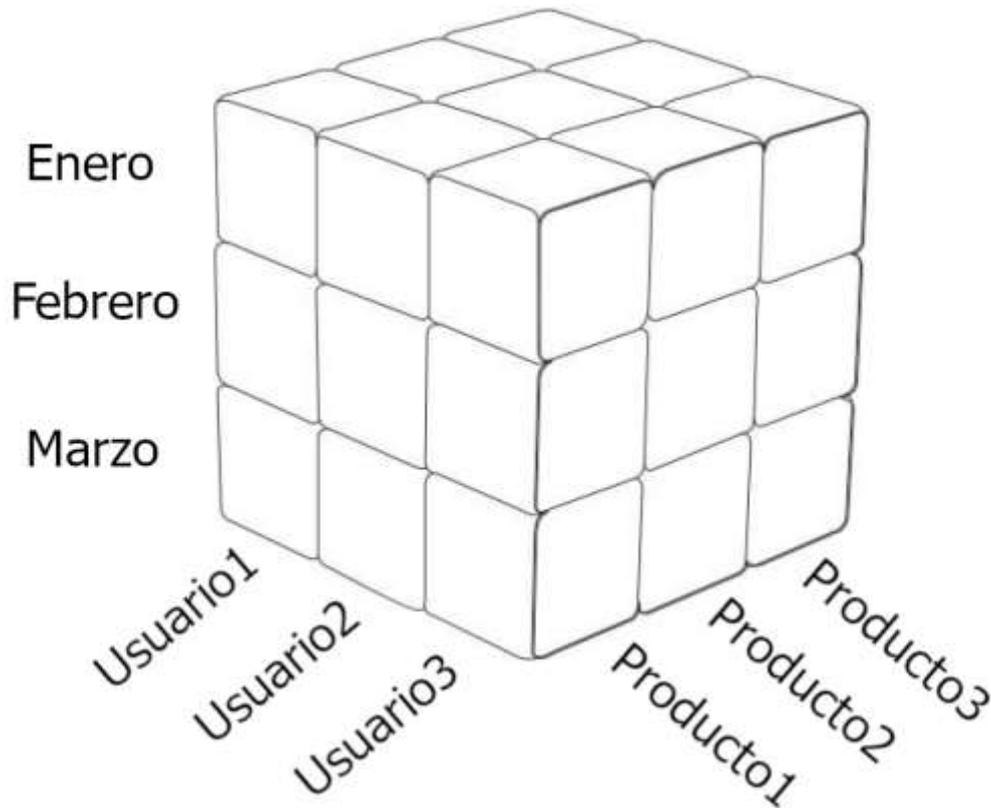


Imagen 5: Ejemplo de Cubo OLAP

Para el diseño del prototipo inicial se plantea la definición de un único cubo OLAP con las dimensiones colaboradores, usuarios, tiempo y productos, para valorar las métricas:

- cuponesCanjeados: Cuenta todos los identificadores de canje (transactionSKU) distintos de una selección.
- Número de colaboradores: Cuenta todos los identificadores de colaborador (CollaboratorID) distintos de una selección.
- Número de usuarios: Cuenta todos los identificadores de usuario (UserID) distintos de una selección.

- Número de productos: Cuenta todos los identificadores de producto (ProductID) distintos de una selección.
- cuponesUnicos: Cuenta todos los identificadores de cupón (CouponID) distintos de una selección.

3.2.4. Diseño de dashboard o cuadro de mandos.

Con la finalidad de representar gráficamente en valor de los diferentes indicadores que se han establecido en la fase de análisis, se diseñan diferentes cuadros de mando que permiten su visualización de una forma rápida, sencilla y accesible.

Para la realización de los bocetos se ha usado la herramienta online Balsamiq Cloud disponible a través de la URL: <https://balsamiq.cloud/>.

3.2.4.1. CM1: Cuadro de mando Administrativo Nutribox

Este cuadro de mando se ha diseñado para que la gerencia de la entidad pueda obtener una representación accesible y simplificada de los principales indicadores que afectan a la evolución anual del proceso de canjes de cupones en la plataforma:

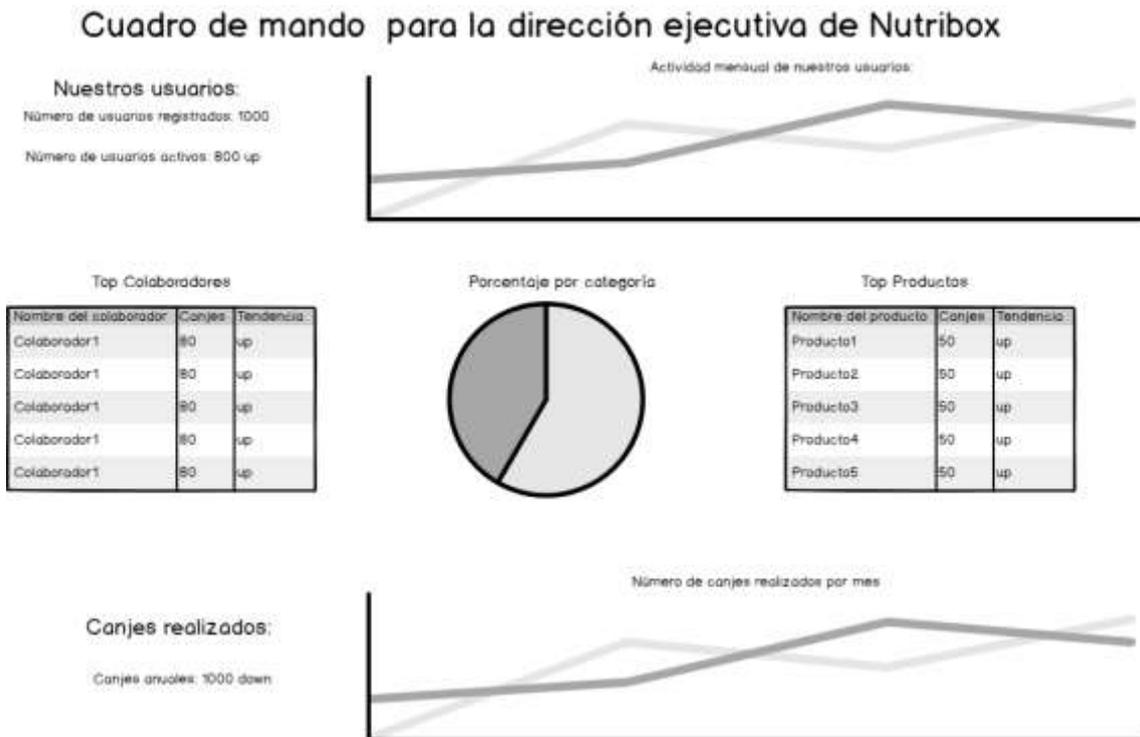


Imagen 6: Diseño cuadro de mando para la dirección ejecutiva de Nutribox

A continuación, se detalla cada elemento:

Diseño CM1 Componente 1: Número de usuarios registrados	
Número de usuarios registrados: 1000	
Descripción:	Muestra el número total de usuarios registrados en la plataforma.
Indicadores relacionados:	KPI 9 Usuarios registrados

Tabla 25: Diseño CM1 Componente1 “Número de usuarios registrados”.

Diseño CM1 Componente 2: Número de usuarios activos	
Número de usuarios activos: 800 up	
Descripción:	Muestra el número total de usuarios activos (que han realizado algún canje de cupón en la plataforma) en el año indicado, así como su tendencia con respecto al ejercicio anterior.
Indicadores relacionados:	KPI 10 Usuarios activos al año. KPI 12 Tendencia Usuarios activos.

Tabla 26: Diseño CM1 Componente 2 “Número de usuarios activos”.

Diseño CM1 Componente 3: Actividad mensual de los usuarios	
	
Descripción:	Muestra el número total de usuarios activos (que han realizado algún canje de cupón en la plataforma) por mes.
Indicadores relacionados:	KPI 11 Usuarios activos por mes.

Tabla 27: Diseño CM1 Componente 3 “Actividad mensual de los usuarios”.

Diseño CM1 Componente 4: Top Colaboradores

Top Colaboradores		
Nombre del colaborador	Canjes	Tendencia
Colaborador1	80	up

Descripción:	Muestra en formato de tabla los 5 colaboradores que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.
Indicadores relacionados:	KPI 4 Popularidad del Colaborador KPI 5 Tendencia Colaborador.

Tabla 28: Diseño CM1 Componente 4: “Top Colaboradores”.

Diseño CM1 Componente 5: Por categoría

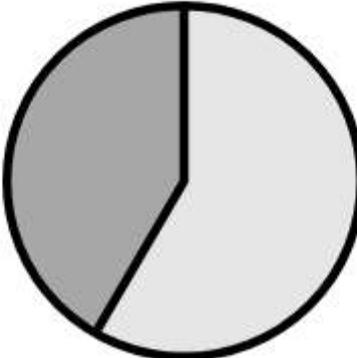
Porcentaje por categoría	
	
Descripción:	Muestra a partir de un gráfico tipo pastel que representa el porcentaje de canjes de cupones por categoría de productos.
Indicadores relacionados:	KPI 3 Porcentaje de canjes por Categoría de productos.

Tabla 29: Diseño CM1 Componente 5 “Por categoría”.

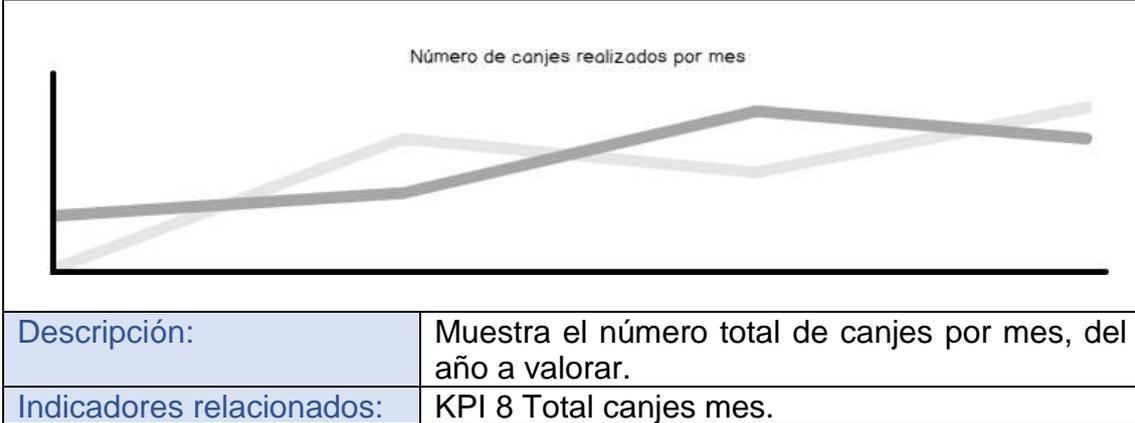
Diseño CM1 Componente 6: Top Productos.																			
<p>Top Productos</p> <table border="1"> <thead> <tr> <th>Nombre del producto</th> <th>Canjes</th> <th>Tendencia</th> </tr> </thead> <tbody> <tr> <td>Producto1</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto2</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto3</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto4</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto5</td> <td>50</td> <td>up</td> </tr> </tbody> </table>		Nombre del producto	Canjes	Tendencia	Producto1	50	up	Producto2	50	up	Producto3	50	up	Producto4	50	up	Producto5	50	up
Nombre del producto	Canjes	Tendencia																	
Producto1	50	up																	
Producto2	50	up																	
Producto3	50	up																	
Producto4	50	up																	
Producto5	50	up																	
Descripción:	Muestra en formato de tabla los 5 productos que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.																		
Indicadores relacionados:	Popularidad del producto, Tendencia Producto.																		

Tabla 30: Diseño CM1 Componente "Top Productos".

Diseño CM1 Componente 7: Canjes anuales	
<p>Canjes anuales: 1000 down</p>	
Descripción:	Muestra el número de canjes de cupones total a través de la plataforma en el año a analizar, así como su tendencia con respecto al ejercicio anterior.
Indicadores relacionados:	KPI 7 Total canjes anuales KPI 6 Tendencia Canjes anuales

Tabla 31: Diseño CM1 Componente 7 "Canjes anuales"

Diseño CM1 Componente 8: Número de canjes realizados por mes



Descripción:	Muestra el número total de canjes por mes, del año a valorar.
Indicadores relacionados:	KPI 8 Total canjes mes.

Tabla 32: Diseño CM1 Componente 8 "Número de canjes realizados por mes"

3.2.4.2. CM2: Cuadro de mando para empresa colaboradora

Este cuadro de mando se ha diseñado para que cada empresa colaboradora de la compañía pueda obtener una representación accesible y simplificada de los principales indicadores que afectan a la evolución anual del proceso de canjes de cupones de los productos que han ofertado en la plataforma:

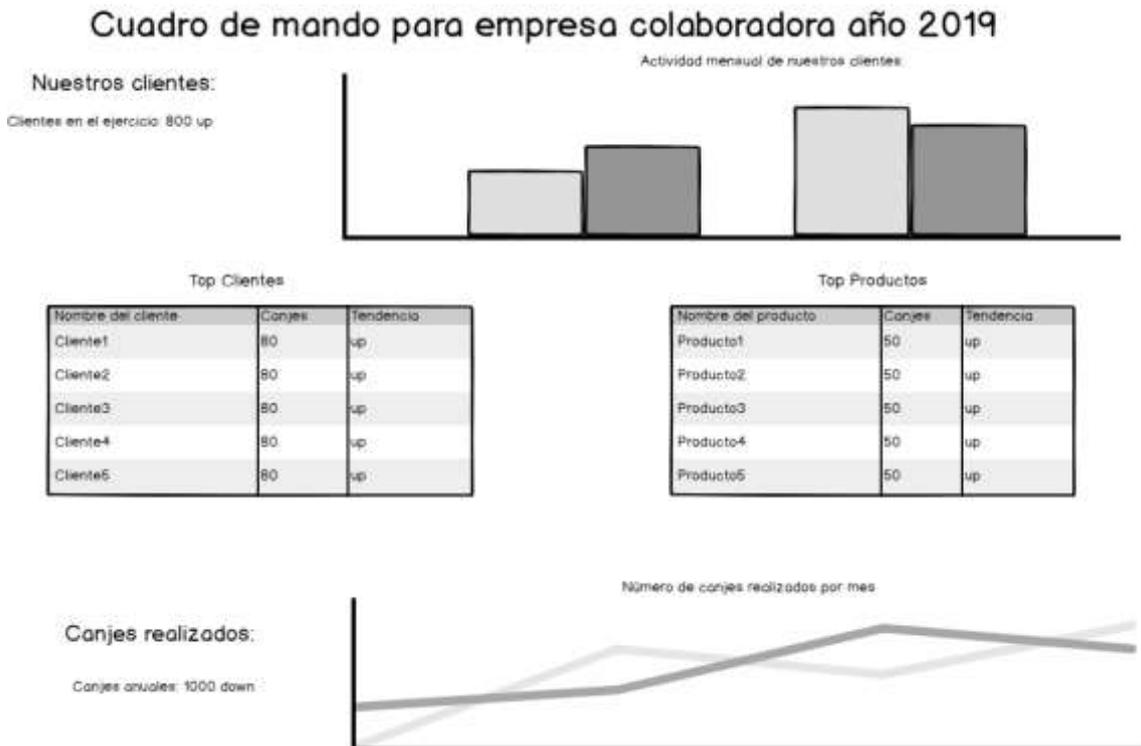


Imagen 7: Diseño cuadro de mando para empresas colaboradoras.

A continuación, se detalla cada elemento:

Diseño CM2 Componente 1: Clientes en el ejercicio	
Clientes en el ejercicio: 800 up	
Descripción:	Muestra el número total de usuarios que han realizado algún canje de cupones del colaborador, en el año indicado, así como su tendencia con respecto al ejercicio anterior.
Indicadores relacionados:	KPI 13 Tendencia clientes del colaborador.

Tabla 33: Diseño CM2 Componente 1 “Clientes en el ejercicio”.

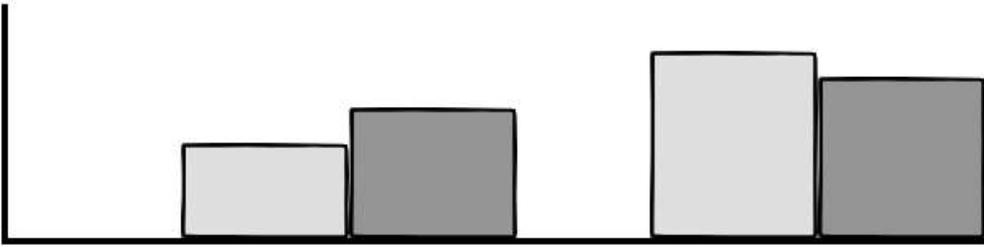
Diseño CM2 Componente 2: Actividad mensual de nuestros clientes.	
Actividad mensual de nuestros clientes: 	
Descripción:	Muestra el número total de usuarios activos que han realizado algún canje de cupón ofertado por este colaborador en la plataforma, por mes.
Indicadores relacionados:	KPI 16 Usuarios activos por colaborador y mes

Tabla 34: Diseño CM2 Componente 2 “Actividad mensual de nuestros clientes”.

Diseño CM2 Componente 3: Top Clientes.		
Top Clientes		
Nombre del cliente	Canjes	Tendencia
Cliente1	80	up
Cliente2	80	up
Cliente3	80	up
Cliente4	80	up
Cliente5	80	up

Descripción:	Muestra en formato de tabla los 5 clientes que han realizado más canjes de cupones ofertados por este colaborador, que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.
Indicadores relacionados:	KPI 14: Actividad del cliente. KPI 15: Tendencia en la actividad del cliente.

Tabla 35: Diseño CM2 Componente 3: "Top Clientes".

Diseño CM2 Componente 4: Top Productos																				
<p>Top Productos</p> <table border="1"> <thead> <tr> <th>Nombre del producto</th> <th>Canjes</th> <th>Tendencia</th> </tr> </thead> <tbody> <tr> <td>Producto1</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto2</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto3</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto4</td> <td>50</td> <td>up</td> </tr> <tr> <td>Producto5</td> <td>50</td> <td>up</td> </tr> </tbody> </table>			Nombre del producto	Canjes	Tendencia	Producto1	50	up	Producto2	50	up	Producto3	50	up	Producto4	50	up	Producto5	50	up
Nombre del producto	Canjes	Tendencia																		
Producto1	50	up																		
Producto2	50	up																		
Producto3	50	up																		
Producto4	50	up																		
Producto5	50	up																		
Descripción:	Muestra en formato de tabla los 5 productos ofertados por este colaborador, que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.																			
Indicadores relacionados:	KPI 17 Popularidad del producto de colaborador X. KPI 18 Tendencia de canjes de producto por colaborador.																			

Tabla 36: Diseño CM2 Componente "Top Productos".

Diseño CM2 Componente 5: Canjes anuales	
Canjes anuales: 1000 down	
Descripción:	Muestra el número de canjes de cupones total que ha obtenido este colaborador en el año a analizar, así como su tendencia con respecto al ejercicio anterior.
Indicadores relacionados:	KPI 19 Tendencia Canjes anuales por colaborador. KPI 20 Total canjes anuales por colaborador.

Tabla 37: Diseño CM2 Componente 5 "Canjes anuales"

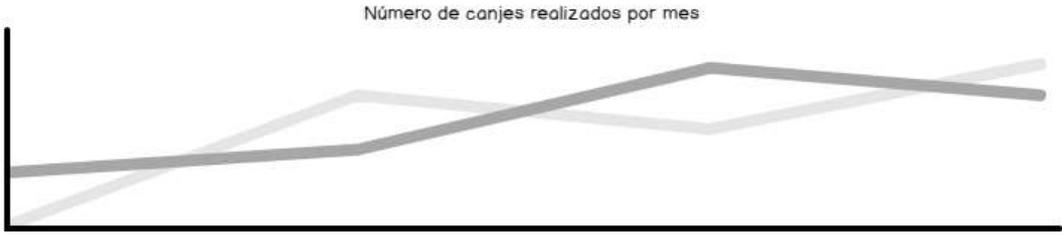
Diseño CM2 Componente 6: Número de canjes realizados por mes	
	
Descripción:	Muestra el número total de canjes de ofertas de este colaborador por mes, del año a valorar.
Indicadores relacionados:	KPI 21 Total canjes del colaborador mes

Tabla 38: Diseño CM2 Componente 6 "Número de canjes realizados por mes"

3.2.5. Reports

A partir de Pentaho Report Designer se crearán diferentes informes dinámicos que servirán para representar la información obtenida a partir de los datos de la organización. Estos informes podrán ser impresos o exportarse en diversos formatos.

Para la realización de los bocetos se ha usado la herramienta online Balsamiq Cloud disponible a través de la URL: <https://balsamiq.cloud/>. A continuación, se define el diseño de varios posibles informes generados en este trabajo:

3.2.5.1. informe de canjes por categoría y año

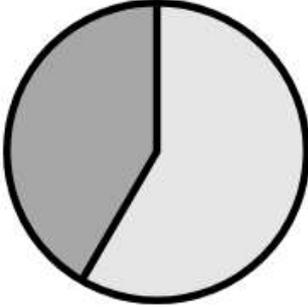
Informe de canjes por categoría y año													
<p>Canjes por categoría realizados el año 2019</p> 													
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%;">Categoría</th> <th style="width: 50%;">Cantidad</th> </tr> </thead> <tbody> <tr> <td>Categoría</td> <td>10</td> </tr> </tbody> </table>		Categoría	Cantidad	Categoría	10								
Categoría	Cantidad												
Categoría	10												
Categoría	10												
Categoría	10												
Categoría	10												
Categoría	10												
Descripción:	En este informe se representa en un gráfico tipo pastel, el porcentaje de canjes realizados sobre una categoría determinada y en una tabla se presenta un listado con las diferentes categorías y el número de canjes realizados.												
Indicadores relacionados:	KPI 3 Porcentaje de canjes por Categoría de productos.												

Tabla 39: Descripción Informe Canjes por categoría.

3.2.5.2. Informe top colaboradores del año

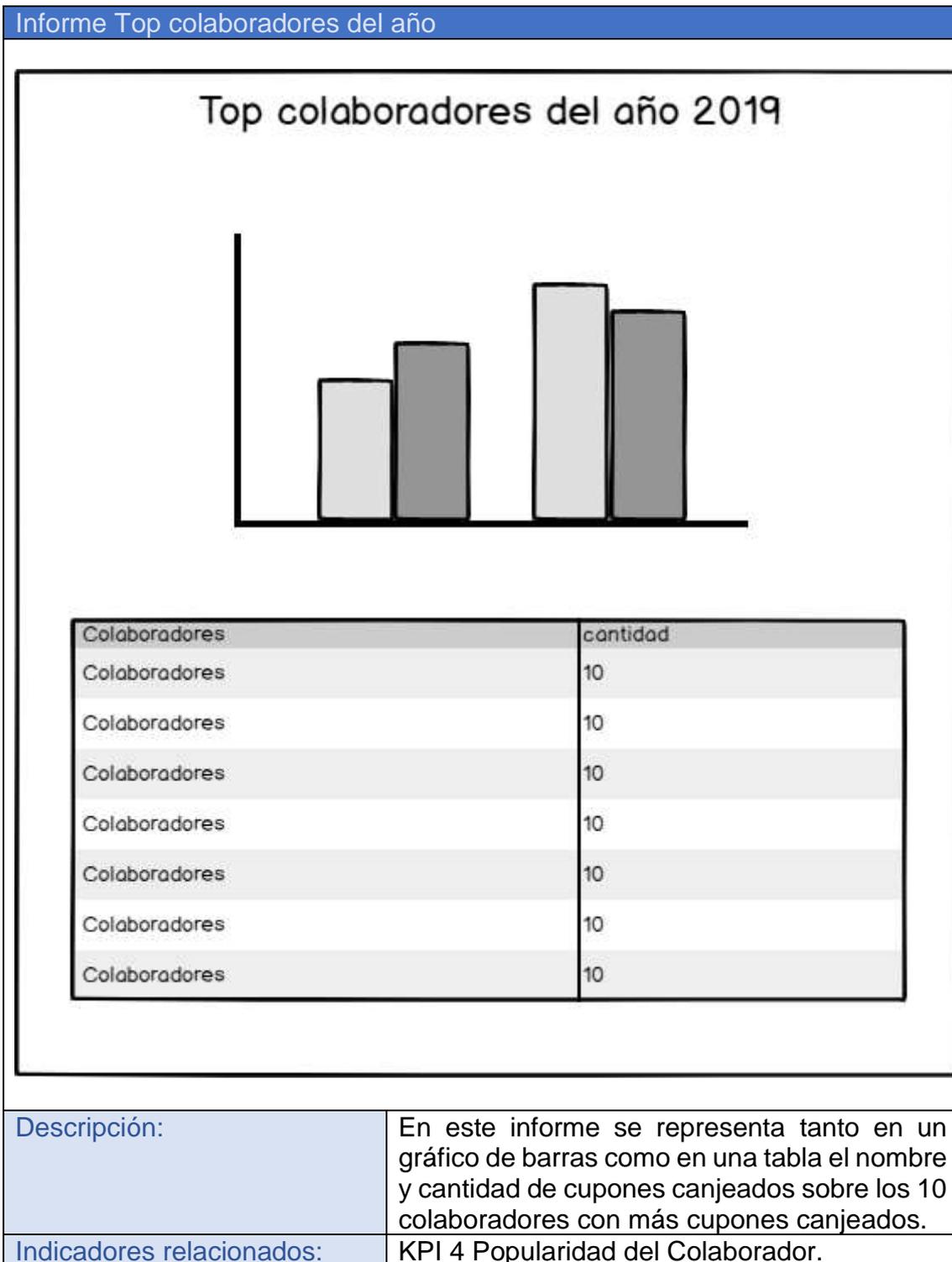


Tabla 40: Informe Top colaboradores del año.

3.3. Selección del software

En este apartado se justificará la elección de las diferentes herramientas y soluciones software que componen la arquitectura del prototipo en base al diseño planteado en el apartado anterior.

3.3.1. Selección del Software de virtualización

Para cumplir los objetivos propuestos, se creará una máquina virtual sobre la que se instalará todo el software necesario. Para este proyecto se ha seleccionado el software Virtualbox, debido principalmente a su facilidad de uso y experiencia previa con esta herramienta.

Aunque es propiedad de Oracle Corporation, es desarrollado bajo un modelo de código abierto, licenciado bajo GNU General Public License (GPL) Version 2 y su uso es totalmente gratuito, tanto para uso personal como profesional⁵, por lo que cumple con los requisitos de este proyecto.

3.3.2. Selección del sistema operativo utilizado en el prototipo

Ubuntu es una distribución de Linux basada en la arquitectura de Debian y está orientada al usuario promedio, con un fuerte enfoque en la facilidad de uso y en mejorar la experiencia del usuario. Estadísticas web sugieren que la cuota de mercado de Ubuntu dentro de las distribuciones Linux es, aproximadamente, del 52% y con una tendencia a aumentar como servidor web.

Se ha seleccionado el sistema operativo Ubuntu 18.04 ya que es una de las distribuciones Linux más utilizadas en la actualidad, por lo que cuenta con una amplia comunidad, así como infinidad de documentación, manuales y otros recursos gratuitos que facilitan su uso.

Además, este sistema operativo es gratuito y de código abierto; cumpliendo los requisitos impuestos para este prototipo. Por otro lado, esta versión (la 18.04 LTS) cuenta con soporte extendido de 5 años, un aspecto a tener en cuenta a la hora de implantar una solución definitiva.

3.3.3. Selección del sistema gestor de bases de datos.

Para la creación de este prototipo es necesaria la instalación de un sistema gestor de bases de datos que servirá, por un lado, para simular la base de datos operacional de Nutribox que será tomada como fuente

⁵ Oracle VM VirtualBox (2019) [En línea] de ORACLE
[Consulta: 1 de Junio de 2019] <<https://www.oracle.com/es/virtualization/virtualbox/>>

de datos en este proyecto y por otro lado como almacén para el datawarehouse de la compañía.

Entre los sistemas gestores de bases de datos Open Source destacan MySQL y PostgreSQL. Ambos sistemas se consideran válidos para este proyecto, por lo que se realiza una comparativa de ambas soluciones, para poder realizar una correcta elección.

3.3.3.1. MySQL

MySQL es la base de datos de código abierto más popular del mercado. Gracias a su rendimiento probado, a su fiabilidad y a su facilidad de uso, MySQL se ha convertido en la base de datos líder elegida para las aplicaciones basadas en web, así como una elección muy popular como base de datos integrada.⁶

Ventajas:

- Velocidad a la hora de realizar las operaciones, lo que le hace uno de los gestores que ofrecen mayor rendimiento.
- Su bajo consumo lo hace apto para ser ejecutado en una máquina con escasos recursos sin ningún problema.
- Gran facilidad de configuración e instalación.
- Tiene una probabilidad muy reducida de corromper los datos, incluso en los casos en los que los errores no se produzcan en el propio gestor, sino en el sistema en el que está.
- El conjunto de aplicaciones Apache-PHP-MySQL es uno de los más utilizados en Internet.

Inconvenientes:

- Carece de soporte para transacciones, rollback's y subconsultas.
- No maneja la integridad referencial.
- Peor rendimiento con grandes bases de datos.⁷

3.3.3.2. PostgreSQL

PostgreSQL se ha enfocado más que nada a la fiabilidad e integridad de los datos, así como también a mejorar como es visto del lado del desarrollador. Actualmente cuenta con planificador de consultas extremadamente sofisticado, que nos permite unir tablas sin importar lo grande que sean de una forma sencilla y eficiente; algo muy importante para aquellos sistemas donde se trabaja con gran cantidad de datos.⁸

⁶ Oracle MySQL (2019) [En línea] de ORACLE
[Consulta: 1 de Junio de 2019] <<https://www.oracle.com/es/mysql/>>

⁷ PostGreSQL vs. MySQL (2016) [En línea] de Daniel Pecos Martínez
[Consulta: 1 de Junio de 2019] <<https://danielpecos.com/documents/postgresql-vs-mysql/>>

⁸ Infografía MySQL vs PostgreSQL (2016) [En línea] de infranetworking
[Consulta: 1 de Junio de 2019] <<https://blog.infranetworking.com/infografia-mysql-vs-postgresql/>>

Ventajas:

- Posee una gran escalabilidad. Es capaz de ajustarse al número de CPUs y a la cantidad de memoria que posee el sistema de forma óptima, haciéndole capaz de soportar una mayor cantidad de peticiones simultáneas de manera correcta.
- Implementa el uso de rollback's, subconsultas y transacciones, haciendo su funcionamiento mucho más eficaz, y ofreciendo soluciones en campos en las que MySQL no podría.
- Tiene la capacidad de comprobar la integridad referencial, así como también la de almacenar procedimientos en la propia base de datos.

Inconvenientes:

- Consume gran cantidad de recursos.
- Tiene un límite de 8K por fila, aunque se puede aumentar a 32K, con una disminución considerable del rendimiento.
- Más lento que MySQL.⁹

3.3.3.3. Elección de sistema gestor de base de datos

Teniendo en cuenta las ventajas e inconvenientes de ambas alternativas, se escoge para este proyecto el uso de MySQL, debido principalmente a la menor cantidad de recursos que requiere y su mayor rendimiento; ya que la construcción de este prototipo sobre una máquina virtual con recursos limitados podría ser un factor limitante.

Además, su mayor popularidad y uso favorecen la obtención de manuales, tutoriales y otros recursos didácticos necesarios para poder obtener unos conocimientos adecuados para desarrollar el proyecto adecuadamente.

3.3.4. Selección de Suite Business Intelligence

Para cumplir con los objetivos previstos es necesario disponer de una solución que permita diseñar y ejecutar procesos ETL que, con la finalidad de extraer los datos de la base de datos operativa, tratarlos y almacenarlos en el data warehouse de la compañía; definir de cubos OLAP para su posterior tratamiento, así como, la creación de cuadros de mandos y reportes.

Para este proyecto la solución de BI escogida es Pentaho Community Edition. Es considerada líder en el mercado actualmente, por lo que existe una gran cantidad de recursos como manuales, tutoriales, videos y otros materiales, que facilitan el aprendizaje.

La suite Pentaho se compone de las siguientes herramientas necesarias para cumplir los objetivos del proyecto:

- Pentaho Business Analytics Platform: desde este servidor de la plataforma podremos gestionar las conexiones con las distintas fuentes

⁹ PostGreSQL vs. MySQL (2016) [En línea] de Daniel Pecos Martínez
[Consulta: 1 de Junio de 2019] <<https://danielpecos.com/documents/postgresql-vs-mysql/>>

de datos, integración con otras herramientas como cde para crear los cuadros de mando, jpivot o saiku para hacer análisis OLAP o acceder a un Marketplace que permite añadir otras funcionalidades

- Pentaho Data Integration: Es la herramienta de ETL de la Suite.
- Pentaho Report Designer (PRD): Herramienta para crear informes.
- Pentaho Schema Workbench (PSW): en este caso se trata de una interfaz gráfica de diseño para crear y probar esquemas para los cubos OLAP de Mondrian.

4. Implantación

En este apartado se describirá la implantación de la solución, de acuerdo a las valoraciones realizadas en fases anteriores y su uso para obtener los distintos entregables y objetivos planteados.

4.1. Descarga e instalación de componentes necesarios para desarrollar la solución.

Con la finalidad de cumplir los objetivos descritos en apartados anteriores es necesario descargar e instalar los diferentes componentes de la solución.

4.1.1. Software de virtualización y sistema operativo.

El primer paso para la creación de la solución escogida será la instalación del software de virtualización y el sistema operativo sobre el que se construye la solución.

Se ha descargado Virtualbox desde el enlace:

<https://www.virtualbox.org/wiki/Downloads>

Tras instalar este software en el equipo anfitrión, se crea una nueva máquina con la configuración por defecto para Ubuntu 18.04. y se instala este sistema operativo que previamente ha sido descargado desde el siguiente enlace:

<https://www.ubuntu.com/download/desktop>

Se ha generado el usuario “nico” con contraseña “123456”

4.1.2. Sistema gestor de bases de datos

Para instalar el gestor de bases de datos mysql, se ejecuta el siguiente comando desde terminal:

```
sudo apt-get install mysql-server
```

Posteriormente se configura la seguridad del servidor por medio del comando:

```
sudo mysql_secure_installation
```

Tras seguir el asistente ya se ha configurado mysql creando un usuario administrador "admin" con la contraseña "12345678". Este será el usuario usado en el resto del desarrollo del proyecto.

4.1.3. Instalación de Java.

La Suite Pentaho se ejecuta sobre java por lo que será un requisito indispensable para la solución, la instalación este software. Con este fin se ejecuta desde terminal los siguientes comandos:

```
sudo add-apt-repository ppa:webupd8team/java  
sudo apt update  
sudo apt-get install oracle-java8-set-default
```

Una vez instalado se debe configurar la variable del sistema JAVA_HOME, para lo cual se edita el archivo del sistema /etc/environment Introduciendo la siguiente línea:

```
JAVA_HOME="/usr/lib/jvm/java-8-oracle/"
```

Tras estos cambios se actualiza esta configuración del sistema mediante el comando:

```
sudo source /etc/environment
```

4.1.4. Descarga e instalación de los componentes necesarios de la suite BI Pentaho.

Será necesaria la descarga de los siguientes componentes:

Servidor Pentaho BI server:

<https://sourceforge.net/projects/pentaho/files/Pentaho%208.2/server/pentaho-server-ce-8.2.0.0-342.zip/download>

Pentaho Data Integration:

<https://sourceforge.net/projects/pentaho/files/Pentaho%208.2/client-tools/>

Pentaho schema workbench:

<https://sourceforge.net/projects/mondrian/files/schema%20workbench/3.14.0/>

Pentaho Report Designer:

<https://sourceforge.net/projects/pentaho/files/Report%20Designer/7.1/prd-ce-7.1.0.0-12.zip/download>

Se descarga el plugin Saiku ce desde el siguiente enlace:

<https://github.com/OSBI/saiku>

Tras descargar los archivos se han descomprimido en la carpeta /home/nico/Downloads/ para poder ser usados. El plugin de saiku tras descomprimirse se mueve a la carpeta /home/nico/Downloads/pentaho-server-ce-8.2.0.0-342/pentaho-server/pentaho-solutions/system/saiku/

Para realizar las conexiones con las bases de datos es necesario descargar el archivo mysql-connector-java-5.1.23.jar desde el siguiente enlace:

<https://dev.mysql.com/downloads/connector/j/5.1.html>

Para copiarlo en las siguientes carpetas:

- /home/nico/Downloads/pentaho-server-ce-8.2.0.0-342/pentaho-server/tomcat/lib/
- /home/nico/Downloads/psw-ce-3.14.0.0-12/schema-workbench/lib/
- /home/nico/Downloads/pdi-ce-8.2.0.0-342/data-integration/lib/
- /home/nico/Downloads/prd-ce-7.0.1.0-12/report-designer/lib/

Además, es necesario modificar los permisos del contenido de la carpeta /home/nico/Downloads/pentaho-server-ce-8.2.0.0-342/ con el siguiente comando:

```
sudo chmod 777 -R /home/nico/Downloads/pentaho-server-ce-8.2.0.0-342/
```

Para iniciar el servidor cuando encienda la máquina virtual se añade un inicio automático a través del “Startup Applications Preferences” de Ubuntu indicando el siguiente comando:

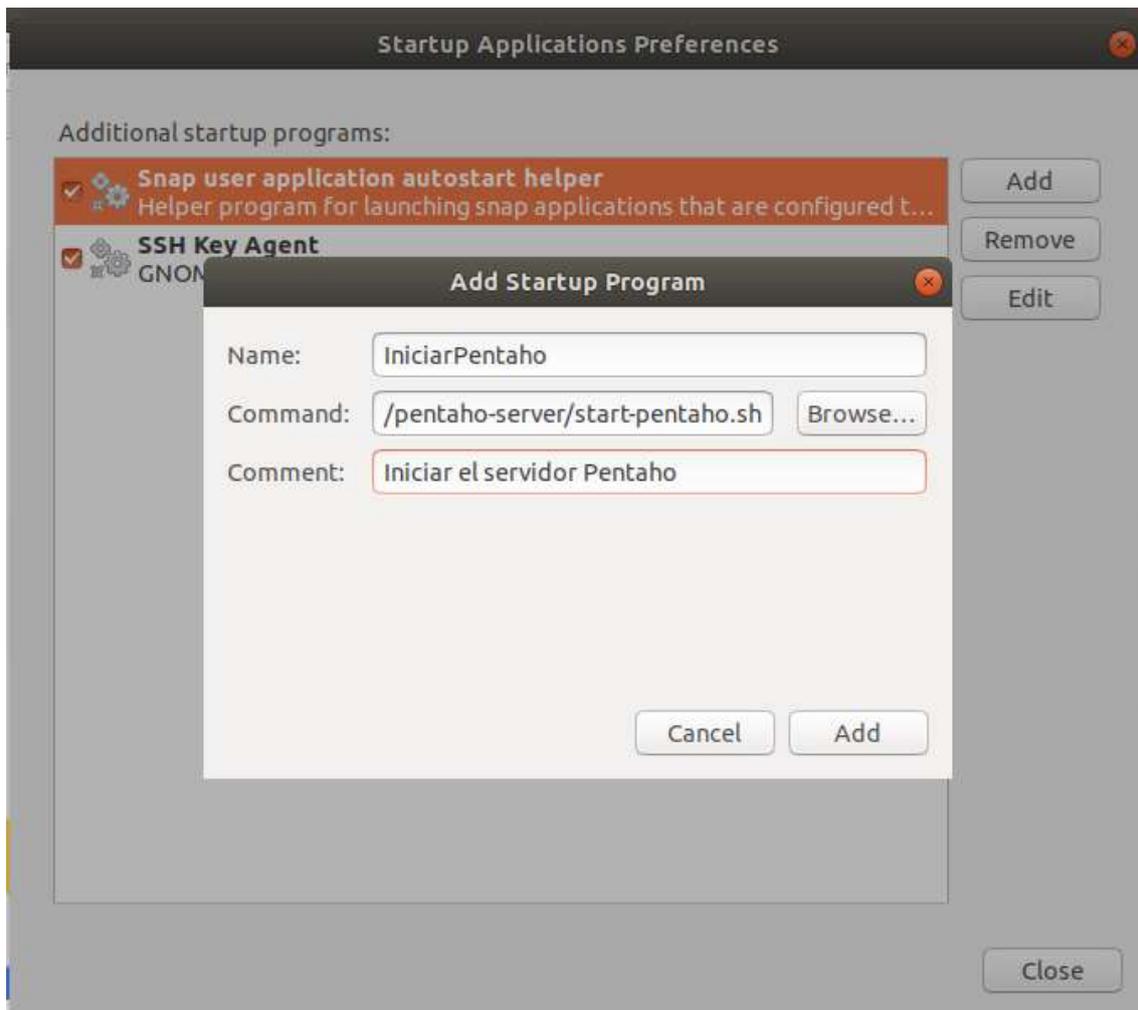


Imagen 8: Iniciar pentaho al encender la máquina virtual.

```
sh /home/nico/Downloads/pentaho-server-ce-8.2.0.0-342/pentaho-server/  
start-pentaho.sh
```

Con esto ya se encuentra preparado en el equipo todo el software necesario para este prototipo.

Para acceder al servidor de Pentaho hay que introducir la siguiente url <http://localhost:8080> desde el navegador y validarse con el usuario administrador por defecto:

usuario: Admin

Contraseña: password

4.2. Creación de las bases de datos Nutribox y Datawarehouse

Se crea el script “crearDatawarehouse.sql” y “Nutribox.sql” que a través de código sql genera las bases de datos, las diferentes tablas, así como sus relaciones.

Estos script se ejecutarán desde terminal del servidor utilizando los comandos:

```
mysql -u admin -p < Nutribox.sql  
  
mysql -u admin -p Nutribox < crearDatawarehouse.sql
```

4.3. Creación de la data set de pruebas

Como ya se ha indicado anteriormente, la empresa Nutribox no está operativa actualmente por lo que se generan unos datos iniciales o de prueba, que servirán para mostrar la utilidad del prototipo.

Para introducir los datos de prueba en la base de datos se ha generado un script SQL que inserta 20 registros en la tabla de empresas colaboradoras, 20 registros en la tabla de productos, 200 registros de cupones, 1000 registros en la tabla usuarios y 5000 registros de transacciones o canjes de cupones en la base de datos Nutribox.

Para generar los 20 registros de empresas y los 20 registros de productos se ha escrito el código sql correspondiente, mientras que para el resto de los registros se utiliza la web <http://generatedata.com/> para simplificar el proceso. Se ha descrito en el anexo “Uso de generatedata.com para crear los datos de prueba” que valores se han introducido en la web para generar estas tablas.

Tanto el código generado como el obtenido de generatedata.com se copia a un archivo nombrado insertarDatos.sql. Este script se ejecutará desde terminal para introducir los datos utilizando el siguiente comando:

```
mysql -u admin -p Nutribox < insertarDatos.sql
```

4.4. Construcción de ETL

Extract, Transform and Load («extraer, transformar y cargar», frecuentemente abreviado ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse

para analizar, o en otro sistema operacional para apoyar un proceso de negocio.¹⁰

Para la creación de estos procesos ETL en este proyecto se utilizan transformaciones de Pentaho Data Integration, que extraen los datos de la base de datos operativa de la plataforma, realiza las transformaciones y cambios necesarios para posteriormente volcarlos en el datawarehouse.

En la siguiente tabla se detallan los procesos generados con Pentaho Data Integration:

Proceso	Descripción
HechosCanje.ktr	A partir de los datos incluidos en las tablas CouponUse y coupons de la base de datos operativa Nutribox poblar la tabla HechosCange, tras haber realizado el tratamiento necesario de los datos.
DimTiempo.ktr	Genera los datos necesarios para crear la dimensión tiempo y su carga en la tabla DimTiempo del data warehouse.
DimProductos.ktr	A partir de los datos incluidos en la tabla products de la base de datos operativa Nutribox poblar la tabla DimProducts, tras haber realizado el tratamiento necesario de los datos.
DimColaboradores.ktr	A partir de los datos incluidos en la tabla collaborators de la base de datos operativa Nutribox poblar la tabla DimColaboradores, tras haber realizado el tratamiento necesario de los datos.
DimUsuarios.ktr	A partir de los datos incluidos en la tabla users de la base de datos operativa Nutribox poblar la tabla DimUsuarios, tras haber realizado el tratamiento necesario de los datos.

Tabla 42: Procesos ETL generados con Pentaho Data Integration

En el anexo “Procesos ETL con Pentaho Data Integration” se detalla cómo construir uno de estos trabajos. Además, los archivos .ktr generados serán entregados con el prototipo.

¹⁰ Extract, transform and load (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] < https://es.wikipedia.org/wiki/Extract,_transform_and_load >

4.4.1. Transformación HechosCanje.ktr



Imagen 9: Transformación HechosCanje.ktr

Esta transformación realiza las siguientes acciones:

- Conecta con la base de datos operativa de Nutibox para extraer los datos de las tablas couponUse y coupons.
- Crea dos nuevos campos tipo string que identifican a la fecha (fechaid) y la hora (horaid) del canje.
- Realiza transformaciones para normalizar los datos (eliminando puntos y otros caracteres especiales, dejando sólo los números de estos dos campos fechaid y horaid)
- Finalmente vuelca incrementalmente los datos, en la tabla HechosCanje del datawarehouse por medio del operador Combination lookup/update.

4.4.2. Transformación DimTiempo.ktr

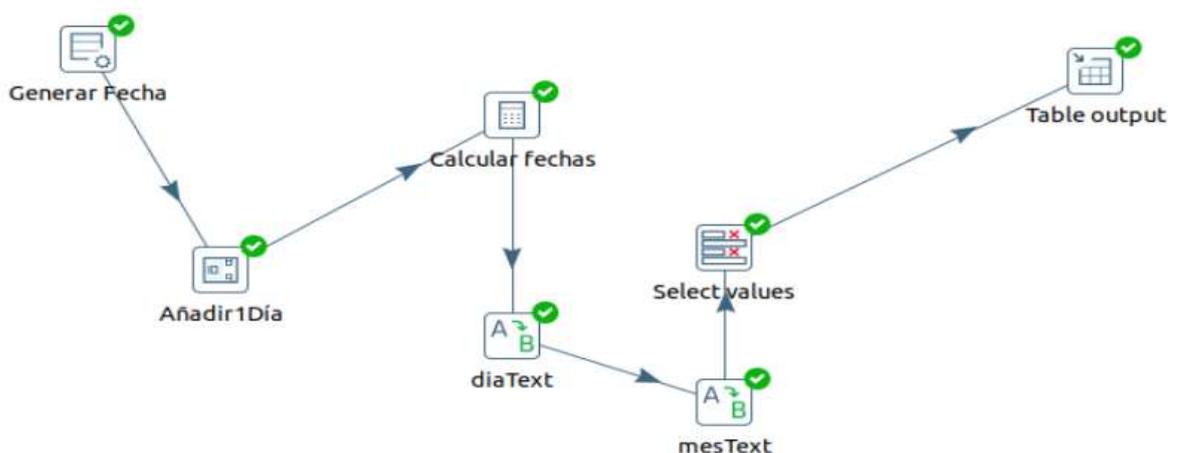


Imagen 10: Transformación DimTiempo.ktr

Este trabajo sólo se ejecutará una vez para cargar los datos de esta dimensión puesto que no dependerá de las operaciones de la plataforma. Al ejecutarlo realiza las siguientes acciones:

- Genera un dataset con 10000 registros tipo fecha con el valor 20180101
- Genera una variable para incrementar el valor de los registros de fecha en un día.
- A partir de los campos generados anteriormente se generan otros campos como la fecha en tipo texto; día del mes, mes, año y día de la semana en formato numérico.
- A partir del campo diaSemana (valor numérico que representa los días de la semana) se genera otro campo indicando el día de la semana en formato texto (Lunes, Martes...)
- A partir del campo mes (valor numérico que representa los días el mes) se genera otro campo indicando el mes en formato texto (Enero, Febrero,...)
- Se seleccionan los valores, descartando los valores auxiliares generados para crear nuevos atributos.
- Finalmente vuelcan los datos en la tabla DimTiempo los datos en el datawarehouse.

4.4.3. Transformación DimProductos.ktr

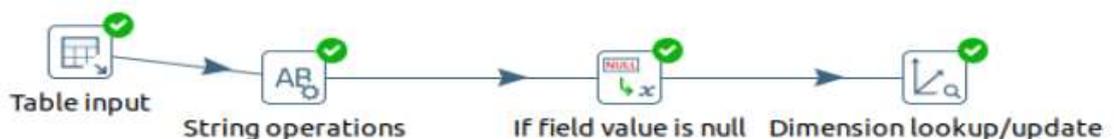


Imagen 11: Transformación DimProductos.ktr

Esta transformación realiza las siguientes acciones:

- Conecta con la base de datos operativa de Nutibox para extraer los datos de la tabla products.
- Crea modificaciones en algunos campos para estandarizar su formato.

- Comprueba si hay registros con el valor NULL, y en su caso completándolo con la cadena “aaaaaaa”.
- Finalmente vuelca incrementalmente los datos en la tabla DimProductos del datawarehouse por medio del operador Dimension lookup/update, el cual sirve además para actualizar las dimensiones lentamente cambiantes, ya que crea claves subrogadas que permiten mantener una historia tanto de los hechos como de la propia dimensión.

4.4.4. Transformación DimColaboradores.ktr

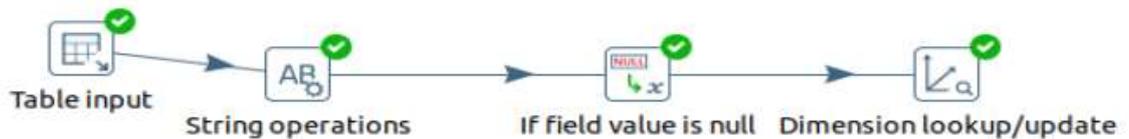


Imagen 12: Transformación DimColaboradores.ktr

Esta transformación realiza las siguientes acciones:

- Conecta con la base de datos operativa de Nutibox para extraer los datos de la tabla collaborators.
- Crea modificaciones en algunos campos para estandarizar su formato.
- Comprueba si hay registros con el valor NULL, y en su caso completándolo con la cadena vacía.
- Finalmente vuelca incrementalmente los datos en la tabla DimColaboradores del datawarehouse por medio del operador Dimension lookup/update, el cual sirve además para actualizar las dimensiones lentamente cambiantes, ya que crea claves subrogadas que permiten mantener una historia tanto de los hechos como de la propia dimensión.

4.4.5. Transformación DimUsuarios.ktr

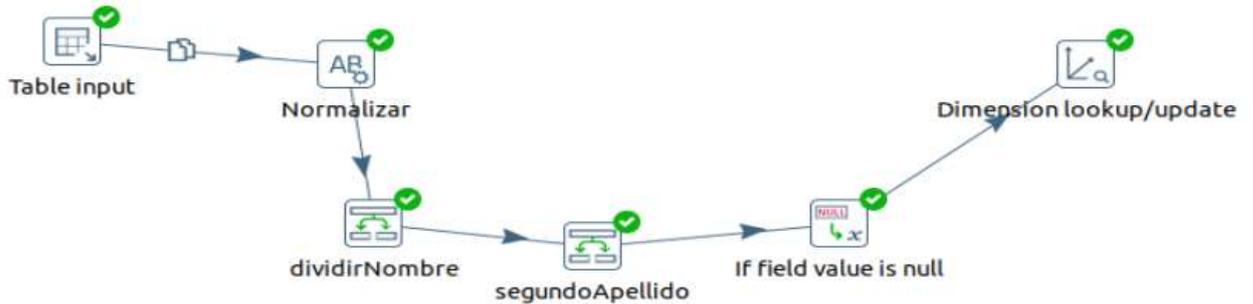


Imagen 13: Transformación DimUsuarios.ktr

Esta transformación realiza las siguientes acciones:

- Conecta con la base de datos operativa de Nutibox para extraer los datos de la tabla users.
- Crea modificaciones en algunos campos para estandarizar su formato.
- Divide el campo que contiene el nombre en dos campos (para nombres compuestos).
- Divide el campo que contiene los apellidos en dos campos.
- Comprueba si hay registros con el valor NULL, y en su caso completándolo con la cadena vacía.
- Finalmente vuelca incrementalmente los datos en la tabla DimUsuarios del datawarehouse por medio del operador Dimension lookup/update, el cual sirve además para actualizar las dimensiones lentamente cambiantes, ya que crea claves subrogadas que permiten mantener una historia tanto de los hechos como de la propia dimensión.

4.5. Cubos OLAP.

Para la definición de los cubos es necesario crear un esquema xml a través de la aplicación pentaho schema workbench. En el anexo definición del esquema OLAP se detallan los pasos realizados para crear y publicar el esquema NutriboxSchema.xml utilizado en este proyecto.

En base a lo descrito en este anexo se ha creado el cubo Olap Nutribox con la estructura que se muestra en la siguiente imagen:

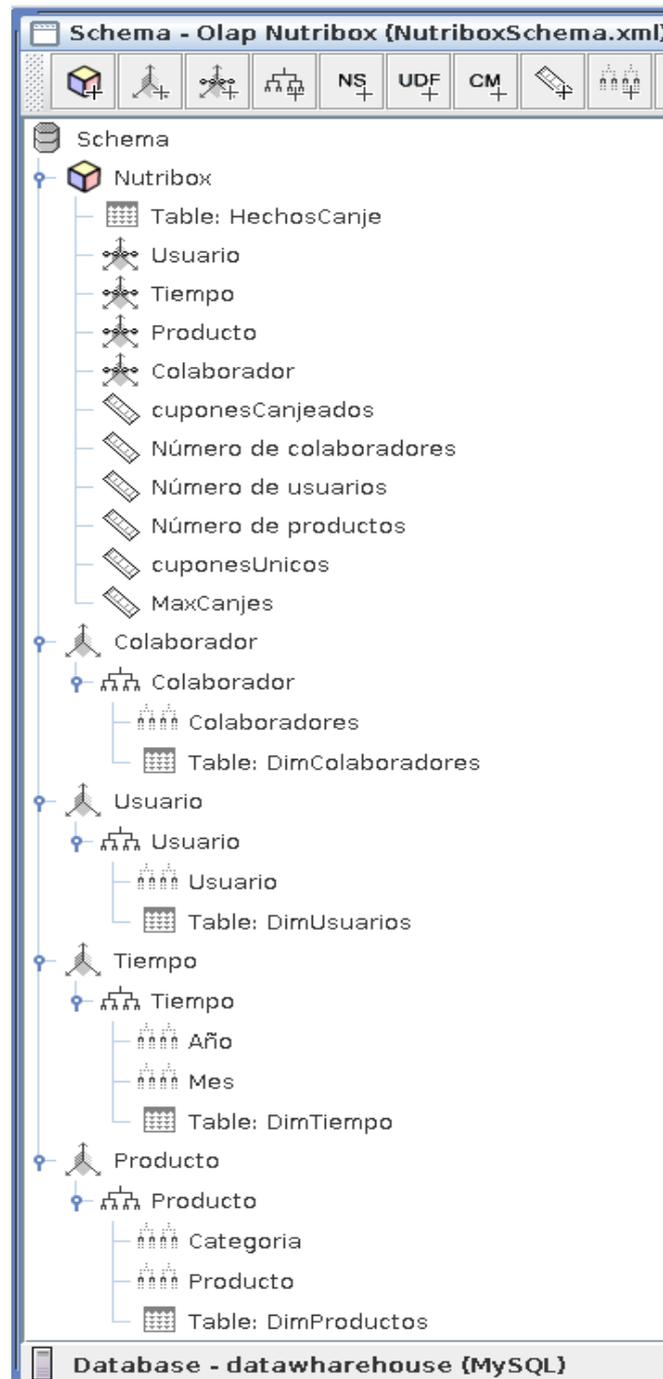


Imagen 14: Schema OLAP Nutribox

Este esquema define las dimensiones:

- Colaborador, que usa como fuente de datos la tabla Dimcolaboradores de la base de datos datawarehouse y tiene el nivel colaboradores.
- Usuario, que usa como fuente de datos la tabla DimUsuarios de la base de datos datawarehouse y tiene el nivel Usuario.
- Tiempo, que usa como fuente de datos la tabla DimTiempo de la base de datos datawarehouse y tiene los niveles Año y Mes.
- Producto, que usa como fuente de datos la tabla DimProductos de la base de datos datawarehouse y tiene el nivel Producto.

Se ha definido el cubo Nutribox, que usa las dimensiones descritas anteriormente, la tabla de hechos HechosCanje y se han definido las métricas:

- cuponesCanjeados: Cuenta todos los identificadores de canje (transactionSKU) distintos de una selección.
- Número de colaboradores: Cuenta todos los identificadores de colaborador (CollaboratorID) distintos de una selección.
- Número de usuarios: Cuenta todos los identificadores de usuario (UserID) distintos de una selección.
- Número de productos: Cuenta todos los identificadores de producto (ProductID) distintos de una selección.
- cuponesUnicos: Cuenta todos los identificadores de cupón (CouponID) distintos de una selección.

Este cubo servirá como fuente de datos estructurada para crear diferentes componentes de cuadros de mando, o reports y permitirá mostrar diferentes análisis a través de herramientas como jpivot o saiku (instaladas en el prototipo del proyecto.)

En los anexos “Análisis de los datos del cubo con jpivot” y “Análisis de los datos del cubo con saiku” se muestran ejemplos del uso de estas herramientas.

4.6. Creación de cuadros de mando.

En base a los diseños planteados en la fase de diseño de la solución, se han definidos un conjunto de cuadros de mando que representan la información de la organización y permitiendo valorar los indicadores de una forma rápida, sencilla y accesible.

4.6.1. CM1: Cuadro de mando para la dirección ejecutiva de Nutribox

Se ha generado el siguiente cuadro de mando para la dirección de la empresa Nutribox en base a los diseños realizados:



Imagen 15: Cuadro de mandos Nutribox para el año 2019

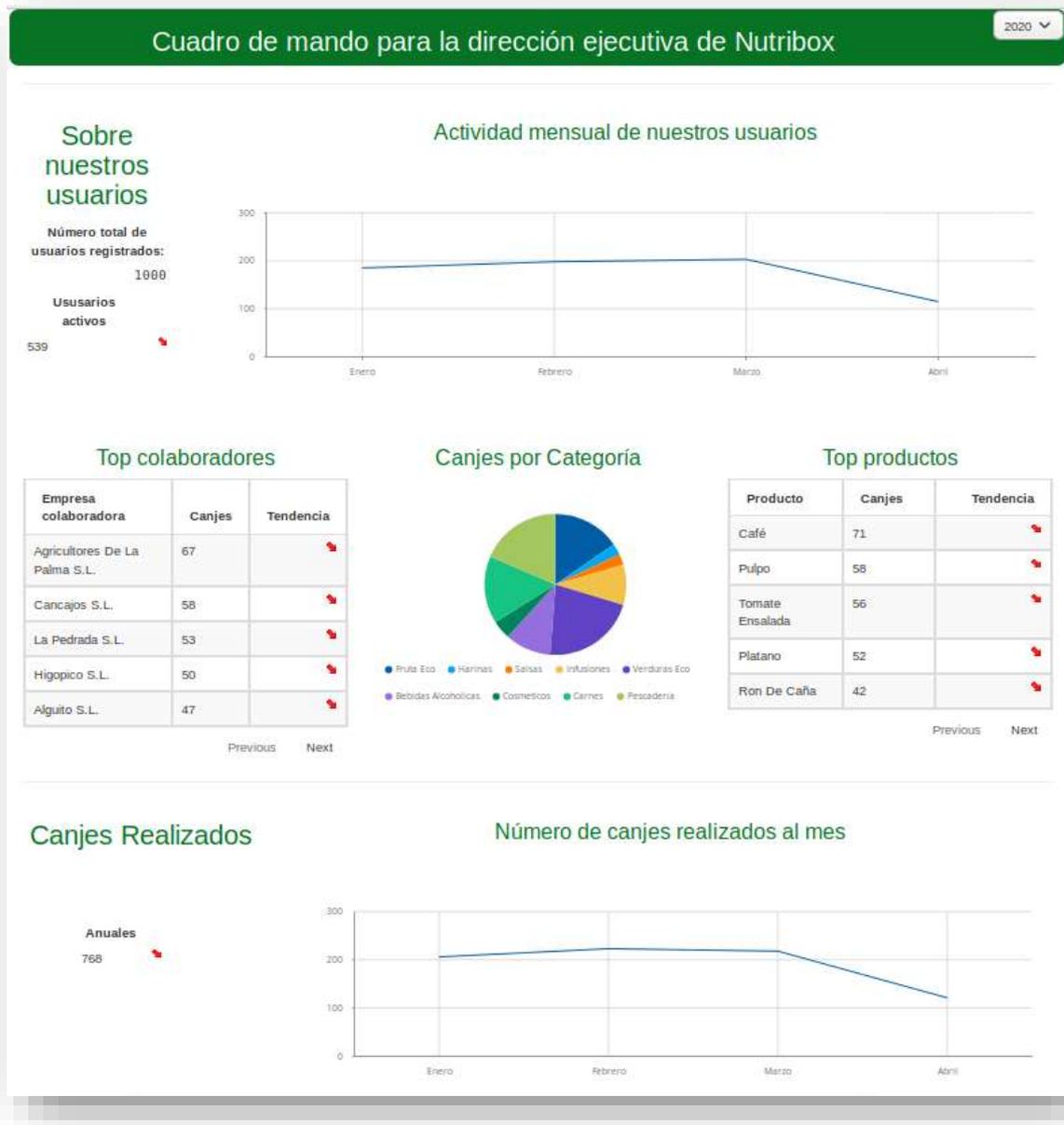
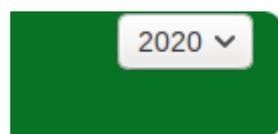


Imagen 16: Imagen cuadro de mandos Nutribox 2020

Para su creación se ha utilizado pentaho CDE que se encuentra integrado en el servidor de pentaho business Analytics del prototipo. A continuación, se describe cada componente.

CM1 Componente 1: Selector de año.



Descripción: Este componente del tipo selector permite seleccionar el año a estudiar en el cuadro de mando y en él lista todos los años con transacciones registradas. El año seleccionado se almacena en un parámetro “Tiempo” que se utilizará para actualizar el resto de los elementos.

Consultas necesarias para la creación del componente:

Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma:

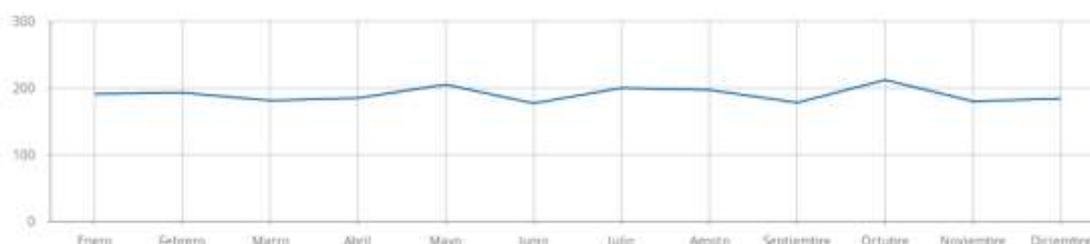
```
with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName'
select TopCount( filter({Descendants([Tiempo].[All Tiempos]
,[Tiempo].[Año])}, not isempty(([Tiempo].CurrentMember)) ) , 50) on ROWS,
{[Measures].[Name]} on Columns
from [Nutribox]
```

A partir de este selector se crea el parámetro tiempo.

Tabla 43: CM1 Componente 1 “Selector de año”

Implantación CM1 Componente 2: Actividad mensual de usuarios.

Actividad mensual de nuestros usuarios



Descripción: Este componente de tipo gráfico de líneas muestra la tendencia de usuarios activos (que han canjeado algún cupón) a lo largo del año según el valor mensual de usuarios activos.

Consultas necesarias para la creación del componente:

Para obtener estos datos se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:

```
select NON EMPTY({Descendants([Tiempo].[${Tiempo}]
,[Tiempo].[Mes])}) on ROWS,
NON EMPTY({[Measures].[Número de usuarios]}) on Columns
from [Nutribox]
```

Indicadores
relacionados

KPI 9 Usuarios activos.

Tabla 44: Implantación CM1 Componente 2 “Actividad mensual de usuarios”

Implantación CM1 Componente 3: Usuarios totales.

**Número total de
usuarios registrados:**

1000

Descripción:

Este componente muestra el número total de usuarios registrados en el sistema.

Consultas necesarias para la creación del componente:

Para obtener este dato se utiliza la siguiente consulta SQL a la base de datos datawarehouse:

```
SELECT Count(DISTINCT UserID) As "Usuarios totales"
From DimUsuarios
```

Indicadores
relacionados

KPI 8 Usuarios registrados

Tabla 45: Implantación CM1 Componente 3 “Usuarios totales”

Implantación CM1 Componente 4: Usuarios activos.

**Usuarios
activos**

539



Descripción:

Este componente muestra el número de usuarios que han canjeado algún cupón el año a evaluar y su tendencia con respecto al año anterior. Para obtener estos años se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:

Consultas necesarias para la creación del componente:

Para obtener estos datos se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:

```
WITH
MEMBER [Tiempo].[Tendencia] as [Tiempo].[${Tiempo}] -
[Tiempo].[${Tiempo}].PREVMEMBER
SET [~C] AS
    {[Tiempo].[${Tiempo}], [Tiempo].[${Tiempo}].PREVMEMBER,
[Tiempo].[Tendencia]}
SELECT
NON EMPTY [~C] ON COLUMNS,
NON EMPTY {[Measures].[cuponesCanjeados]} ON ROWS
FROM [Nutribox]
```

Indicadores relacionados	KPI 9 Usuarios activos. KPI 10 Tendencia Usuarios activos.
--------------------------	---

Tabla 46: Implantación CM1 Componente 4 “Usuarios activos”

Implantación CM1 Componente 5: Tabla empresas colaboradoras

Top colaboradores

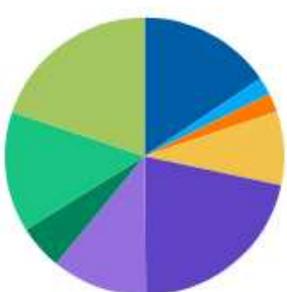
Empresa colaboradora	Canjes	Tendencia
Agricultores De La Palma S.L.	201	
Cooperativa Palmera	172	
Higopico S.L.	163	
Bimbache S.L.	162	
Alguito S.L.	156	

Previous Next

Descripción: Este componente muestra una tabla con los valores de 5 Empresas colaboradoras ordenados por el número de cupones canjeados para adquirir sus productos. En esta tabla se indica el número de canjes totales en el año indicado, así como su tendencia en comparación con el ejercicio anterior.

	Aunque inicialmente sólo muestra el top 5 en orden de canjes realizados pueden verse el resto de los registros utilizando las etiquetas de Previous y Next.
Consultas necesarias para la creación del componente:	
<p>Para obtener estos años se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:</p> <pre> WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~COLUMNS] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER,[Tiempo].[Tendencia]} SET [~ROWS] AS {[Colaborador].[Colaboradores].Members} SELECT NON EMPTY CrossJoin([~COLUMNS], {[Measures].[cuponesCanjeados]}) ON COLUMNS, NON EMPTY [~ROWS] ON ROWS FROM [Nutribox] </pre>	
Indicadores relacionados	KPI 4 Popularidad del Colaborador KPI 5 Tendencia Colaborador.

Tabla 47: Implantación CM1 Componente 5 “Tabla empresas colaboradoras”

Implantación CM1 Componente 6: Gráfico por categorías	
<p style="text-align: center;">Canjes por Categoría</p>  <p style="text-align: center;"> ● Fruta Eco ● Harinas ● Salsas ● Infusiones ● Verduras Eco ● Bebidas Alcoholicas ● Cosmeticos ● Carnes ● Pescadería </p>	
Descripción:	Se trata de un gráfico que representa el porcentaje de canjes por categorías en el año indicado.

Consultas necesarias para la creación del componente:

Para obtener estos valores se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:

```
select NON EMPTY({Descendants([Producto].[All Productos],
[Producto].[Categoria])}) on ROWS,
NON EMPTY({[Measures].[cuponesCanjeados]}) on Columns
from [Nutribox]
WHERE {[Tiempo].[${Tiempo}]}
```

Indicadores relacionados

KPI 3 Porcentaje de canjes por Categoría de productos.

Tabla 48: Implantación CM1 Componente 6 “Canjes por categoría”

Implantación CM1 Componente /: Tabla de productos

Top productos

Producto	Canjes	Tendencia
Café	218	
Pulpo	183	
Tomate Ensalada	156	
Ron De Caña	155	
Platano	154	

Previous Next

Descripción:

Este componente muestra una tabla con los valores de 5 productos ordenados por el número de cupones canjeados para adquirir ese producto en el año a estudiar. En esta tabla se indica el número de canjes totales en el año indicado, así como su tendencia en comparación con el ejercicio anterior.

Aunque inicialmente sólo muestra el top 5 en orden de canjes realizados pueden verse el resto de los registros utilizando las etiquetas de Previous y Next.

Consultas necesarias para la creación del componente:

<p>Para obtener estos valores se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:</p> <p>Para obtener estos años se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:</p> <pre> WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~COLUMNS] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER, [Tiempo].[Tendencia]} SET [~ROWS] AS {[Producto].[Producto].Members} SELECT NON EMPTY CrossJoin([~COLUMNS], {[Measures].[cuponesCanjeados]}) ON COLUMNS, NON EMPTY [~ROWS] ON ROWS FROM [Nutribox] </pre>	
Indicadores relacionados	<p>KPI 1 Popularidad del producto.</p> <p>KPI 2 Tendencia Producto.</p>

Tabla 49: Implantación CM1 Componente 7 “Tabla de productos”

Implantación CM1 Componente 8: Canjeados al año	
<p>Anuales</p> <p>2548 </p>	
Descripción:	Indica el número de cupones canjeados el año indicado.
Consultas necesarias para la creación del componente:	
<p>Para obtener estos valores se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:</p> <pre> WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~C] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER, [Tiempo].[Tendencia]} SELECT NON EMPTY [~C] ON COLUMNS, </pre>	

NON EMPTY {[Measures].[cuponesCanjeados]} ON ROWS FROM [Nutribox]	
Indicadores relacionados	KPI 7 Total canjes KPI 6 Tendencia Canjes anuales

Tabla 50: Implantación CM1 Componente 8 “Canjeados al año”

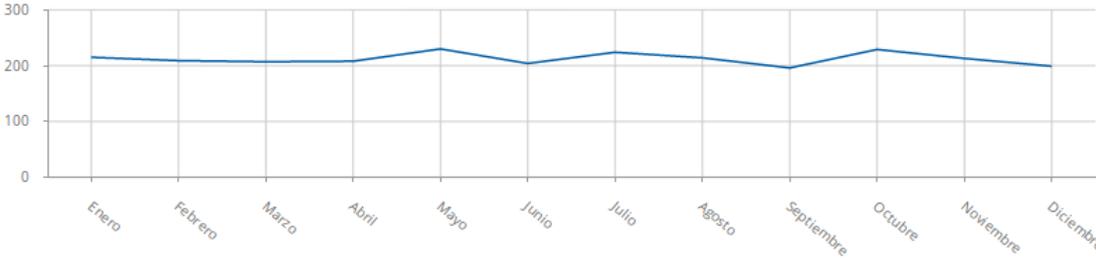
Implantación CM1 Componente 9: Canjes por mes																											
<p>Número de canjes realizados al mes</p>  <table border="1"> <caption>Data for 'Número de canjes realizados al mes'</caption> <thead> <tr> <th>Mes</th> <th>Número de canjes</th> </tr> </thead> <tbody> <tr><td>Enero</td><td>220</td></tr> <tr><td>Febrero</td><td>210</td></tr> <tr><td>Marzo</td><td>210</td></tr> <tr><td>Abril</td><td>210</td></tr> <tr><td>Mayo</td><td>230</td></tr> <tr><td>Junio</td><td>210</td></tr> <tr><td>Julio</td><td>230</td></tr> <tr><td>Agosto</td><td>220</td></tr> <tr><td>Septiembre</td><td>200</td></tr> <tr><td>Octubre</td><td>230</td></tr> <tr><td>Noviembre</td><td>210</td></tr> <tr><td>Diciembre</td><td>200</td></tr> </tbody> </table>		Mes	Número de canjes	Enero	220	Febrero	210	Marzo	210	Abril	210	Mayo	230	Junio	210	Julio	230	Agosto	220	Septiembre	200	Octubre	230	Noviembre	210	Diciembre	200
Mes	Número de canjes																										
Enero	220																										
Febrero	210																										
Marzo	210																										
Abril	210																										
Mayo	230																										
Junio	210																										
Julio	230																										
Agosto	220																										
Septiembre	200																										
Octubre	230																										
Noviembre	210																										
Diciembre	200																										
Descripción:	Grafica que representa la tendencia de cupones canjeados al mes en el año indicado de acuerdo al número de canjes realizados en la plataforma por mes.																										
Consultas necesarias para la creación del componente:																											
<p>Para obtener estos años se utiliza la siguiente consulta MDX al cubo OLAP Nutribox:</p> <pre> select NON EMPTY({Descendants([Tiempo].[\${Tiempo}] ,[Tiempo].[Mes])}) on ROWS, NON EMPTY({[Measures].[cuponesCanjeados]}) on Columns from [Nutribox] </pre>																											
Indicadores relacionados	KPI 7 Total canjes.																										

Tabla 51: Implantación CM1 Componente 9 “Canjes por mes”

4.6.2. CM2: Cuadro de mando empresa colaboradora.

Se ha generado el siguiente cuadro de mando para la dirección ejecutiva de las diferentes empresas colaboradoras, en base a los diseños realizados:



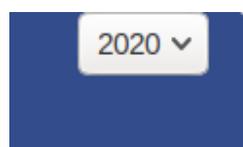
Imagen 17: Cuadro de mando para la empresa Papa Arrugada S.L.



Imagen 18: Cuadro de mando para la empresa Cancajos S.L. del año 2020.

Para su creación se ha utilizado pentaho CDE que se encuentra integrado en el servidor de pentaho business Analytics del prototipo. A continuación, se describe cada componente.

CM2 Componente 1: Selector de año.



Descripción: Este componente del tipo selector permite seleccionar el año a estudiar en el cuadro de mando y en él lista todos los años

	con transacciones registradas. El año seleccionado se almacena en un parámetro “Tiempo” que se utilizará para actualizar el resto de los elementos.
Consultas necesarias para la creación del componente:	
<p>Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma:</p> <pre>with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName' select TopCount(filter({Descendants([Tiempo].[All Tiempos] ,[Tiempo].[Año])}, not isempty(([Tiempo].CurrentMember))) , 50) on ROWS, {[Measures].[Name]} on Columns from [Nutribox]</pre> <p>A partir de este selector se crea el parámetro Tiempo.</p>	

Tabla 52: Implantación CM2 Componente 1 “Selector de año”.

Implantación CM2 Componente 2: Selector de empresa colaboradora.	
	
Descripción:	Este componente del tipo selector permite seleccionar la empresa colaboradora que se mostrará en el cuadro de mando. La empresa seleccionada se almacena en un parámetro que se utilizará para actualizar el resto de los elementos.
Consultas necesarias para la creación del componente:	
<p>Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma:</p> <pre>with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName' select TopCount(filter({Descendants([Tiempo].[All Tiempos] ,[Tiempo].[Año])}, not isempty(([Tiempo].CurrentMember))) , 50) on ROWS, {[Measures].[Name]} on Columns from [Nutribox]</pre> <p>A partir de este selector se crea el parámetro Col.</p>	

Tabla 53: Implantación CM2 Componente 2 “Selector de empresa colaboradora”.

Implantación CM2 Componente 3: Clientes en el ejercicio.	
<p>Clientes en el ejercicio:</p> <p>57 </p>	
Descripción:	Muestra el número total de usuarios que han realizado algún canje de cupones del colaborador, en el año indicado, así como su tendencia con respecto al ejercicio anterior.
Consultas necesarias para la creación del componente:	
<pre>WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~C] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER, [Tiempo].[Tendencia]} SELECT NON EMPTY [~C] ON COLUMNS, NON EMPTY {[Measures].[Número de usuarios]} ON ROWS FROM [Nutribox] where [Colaborador].[\${Col}]</pre>	
Indicadores relacionados	KPI 11 Tendencia clientes del colaborador.

Tabla 54: Implantación CM2 Componente 3 “Clientes en el ejercicio”.

Implantación CM2 Componente 4: Actividad mensual de nuestros clientes.																					
<p>Actividad mensual de nuestros clientes</p>  <table border="1" style="display: none;"> <caption>Data for 'Actividad mensual de nuestros clientes'</caption> <thead> <tr> <th>Mes</th> <th>Número de usuarios activos</th> </tr> </thead> <tbody> <tr><td>Abril</td><td>2</td></tr> <tr><td>Mayo</td><td>15</td></tr> <tr><td>Junio</td><td>3</td></tr> <tr><td>Julio</td><td>10</td></tr> <tr><td>Agosto</td><td>9</td></tr> <tr><td>Septiembre</td><td>12</td></tr> <tr><td>Octubre</td><td>9</td></tr> <tr><td>Noviembre</td><td>11</td></tr> <tr><td>Diciembre</td><td>12</td></tr> </tbody> </table>		Mes	Número de usuarios activos	Abril	2	Mayo	15	Junio	3	Julio	10	Agosto	9	Septiembre	12	Octubre	9	Noviembre	11	Diciembre	12
Mes	Número de usuarios activos																				
Abril	2																				
Mayo	15																				
Junio	3																				
Julio	10																				
Agosto	9																				
Septiembre	12																				
Octubre	9																				
Noviembre	11																				
Diciembre	12																				
Descripción:	Muestra el número total de usuarios activos que han realizado algún canje de cupón ofertado por este colaborador en la plataforma, por mes.																				

Consultas necesarias para la creación del componente:	
<pre>select NON EMPTY({Descendants([Tiempo].[\${Tiempo}] ,[Tiempo].[Mes])}) on ROWS, NON EMPTY({[Measures].[Número de usuarios]}) on Columns from [Nutribox] where [Colaborador].[\${Col}]</pre>	
Indicadores relacionados	KPI 14 Usuarios activos por colaborador y mes

Tabla 55: Implantación CM2 Componente 4 “Actividad mensual de los clientes”

Implantación CM2 Componente 5: Top Clientes.																			
<h3>Top Clientes</h3> <table border="1" style="margin: auto;"> <thead> <tr> <th>Usuario</th> <th>Canjes</th> <th>Tendencia</th> </tr> </thead> <tbody> <tr> <td>Jonah Sweet</td> <td>2</td> <td style="text-align: right;">↑</td> </tr> <tr> <td>Reese Hooper</td> <td>1</td> <td style="text-align: right;">↑</td> </tr> <tr> <td>Adria Reese</td> <td>1</td> <td style="text-align: right;">↑</td> </tr> <tr> <td>Colton Andrews</td> <td>1</td> <td style="text-align: right;">↓</td> </tr> <tr> <td>Jelani Gilmore</td> <td>1</td> <td style="text-align: right;">↑</td> </tr> </tbody> </table> <p style="text-align: right; margin-right: 50px;">Previous Next</p>		Usuario	Canjes	Tendencia	Jonah Sweet	2	↑	Reese Hooper	1	↑	Adria Reese	1	↑	Colton Andrews	1	↓	Jelani Gilmore	1	↑
Usuario	Canjes	Tendencia																	
Jonah Sweet	2	↑																	
Reese Hooper	1	↑																	
Adria Reese	1	↑																	
Colton Andrews	1	↓																	
Jelani Gilmore	1	↑																	
Descripción:	Muestra en formato de tabla los 5 clientes que han realizado más canjes de cupones ofertados por este colaborador, que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.																		
Consultas necesarias para la creación del componente:																			
<pre>WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~COLUMNS] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER, [Tiempo].[Tendencia]} SET [~ROWS] AS {[Usuario].[Usuario].Members} SELECT</pre>																			

```

NON EMPTY CrossJoin([~COLUMNS], {[Measures].[cuponesCanjeados]}) ON
COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Nutribox]
where [Colaborador].[${Col}]

```

Indicadores relacionados	KPI 12: Actividad del cliente. KPI 13: Tendencia en la actividad del cliente.
--------------------------	--

Tabla 56: Implantación CM2 Componente 5 “Top Clientes”.

Implantación CM2 Componente 6: Top Productos.

Top productos

Producto	Canjes	Tendencia
Tomate Ensalada	9	
Pepino Eco	9	
Piña Tropical	8	
Aguacate	6	
Pulpo	6	

Previous
Next

Descripción:	Muestra en formato de tabla los 5 productos ofertados por este colaborador, que más éxito han tenido en la plataforma en base al número de cupones canjeados de sus ofertas. Muestra además del valor numérico del año actual, su tendencia respecto al año anterior.
---------------------	---

Consultas necesarias para la creación del componente:

```

WITH
MEMBER [Tiempo].[Tendencia] as [Tiempo].[${Tiempo}] -
[Tiempo].[${Tiempo}].PREVMEMBER

SET [~COLUMNS] AS
    {[Tiempo].[${Tiempo}], [Tiempo].[${Tiempo}].PREVMEMBER,
[Tiempo].[Tendencia]}
SET [~ROWS] AS
    {[Producto].[Producto].Members}

```

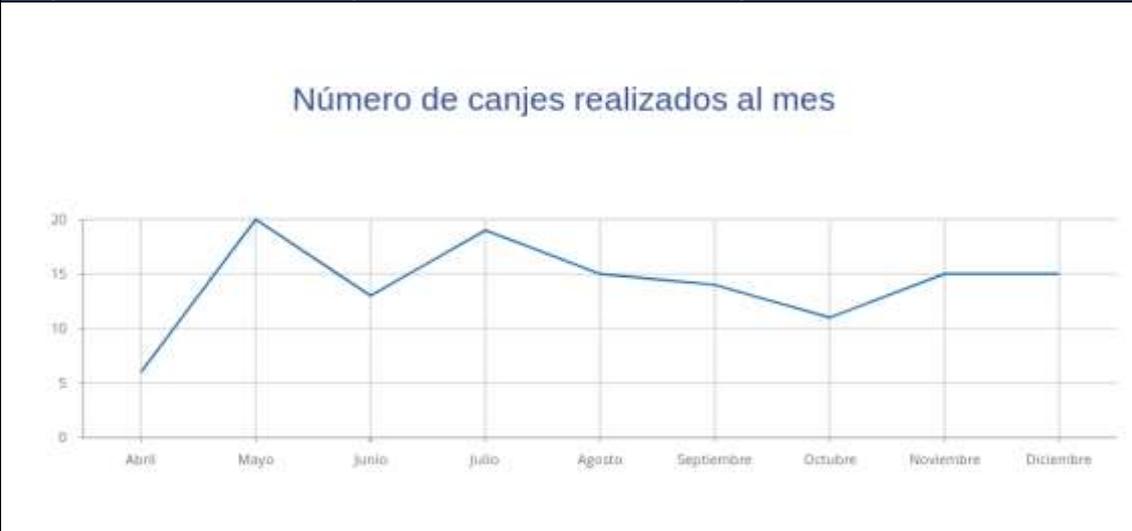
SELECT NON EMPTY CrossJoin([~COLUMNS], {[Measures].[cuponesCanjeados]}) ON COLUMNS, NON EMPTY [~ROWS] ON ROWS FROM [Nutribox] where [Colaborador].[\${Col}]	
Indicadores relacionados	KPI 15 Popularidad del producto de colaborador X. KPI 16 Tendencia de canjes de producto por colaborador.

Tabla 57: Implantación CM2 Componente 6 “Top productos”.

Implantación CM2 Componente 7: Canjes anuales.	
<p>Anuales</p> <p>201 </p>	
Descripción:	Muestra el número de canjes de cupones total que ha obtenido este colaborador en el año a analizar, así como su tendencia con respecto al ejercicio anterior.
Consultas necesarias para la creación del componente:	
WITH MEMBER [Tiempo].[Tendencia] as [Tiempo].[\${Tiempo}] - [Tiempo].[\${Tiempo}].PREVMEMBER SET [~C] AS {[Tiempo].[\${Tiempo}], [Tiempo].[\${Tiempo}].PREVMEMBER, [Tiempo].[Tendencia]} SELECT NON EMPTY [~C] ON COLUMNS, NON EMPTY {[Measures].[cuponesCanjeados]} ON ROWS FROM [Nutribox] where [Colaborador].[\${Col}]	
Indicadores relacionados	KPI 17 Tendencia Canjes anuales por colaborador. KPI 18 Total canjes anuales por colaborador.

Tabla 58: Implantación CM2 Componente 7 “Canjes anuales”.

Implantación CM2 Componente 8: Número de canjes realizados al mes.



Descripción: Muestra el número total de canjes de ofertas de este colaborador por mes, del año a valorar.

Consultas necesarias para la creación del componente:

```
select NON EMPTY({Descendants([Tiempo].[${Tiempo}] ,[Tiempo].[Mes])}) on
ROWS,
NON EMPTY({[Measures].[cuponesCanjeados]}) on Columns
from [Nutribox]
where [Colaborador].[${Col}]
```

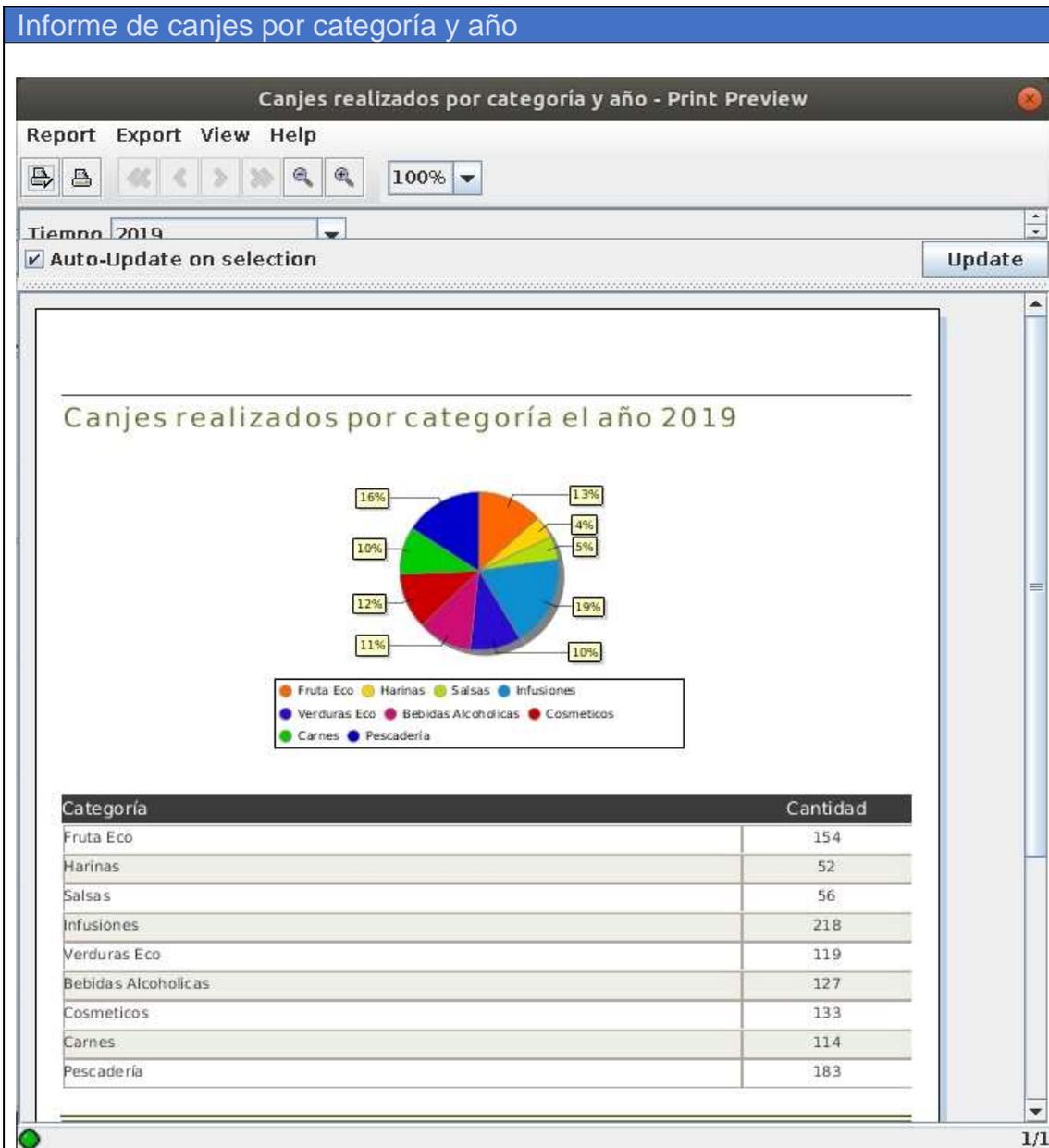
Indicadores relacionados KPI 18 Total canjes del colaborador

Tabla 59: Implantación CM2 Componente 8 “Número de canjes realizados al mes.”.

4.7. Creación de informes dinámicos en Pentaho Report Designer.

En base a los diseños planteados en la fase de diseño de la solución, se han definen un conjunto de informes dinámicos que representan información de la organización y permiten valorar los indicadores establecidos en la fase de análisis. Se puede ver un ejemplo de su construcción en el anexo Creación de informes con Pentaho Report Designer. A continuación, se detallan algunos de estos informes:

4.7.1. Informe de canjes por categoría y año



Descripción:

En este informe se representa en un gráfico tipo pastel, el porcentaje de canjes realizados sobre una categoría determinada y en una tabla se presenta un listado con las diferentes categorías y el número de canjes realizados. En la parte superior se ha incluido un selector para seleccionar el año a representar en el informe.

Consultas necesarias para la creación del informe:

Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma.

```
with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName' select
TopCount( filter({Descendants([Tiempo].[All Tiempos] ,[Tiempo].[Año])}, not
isempty([Tiempo].CurrentMember)) , 50) on ROWS,
{[Measures].[Name]} on Columns
from [Nutribox]
```

A partir de este selector se crea el parámetro tiempo.

Para obtener los datos que se representan en el gráfico y la tabla del informe se crea la siguiente consulta MDX al cubo OLAP Nutribox:

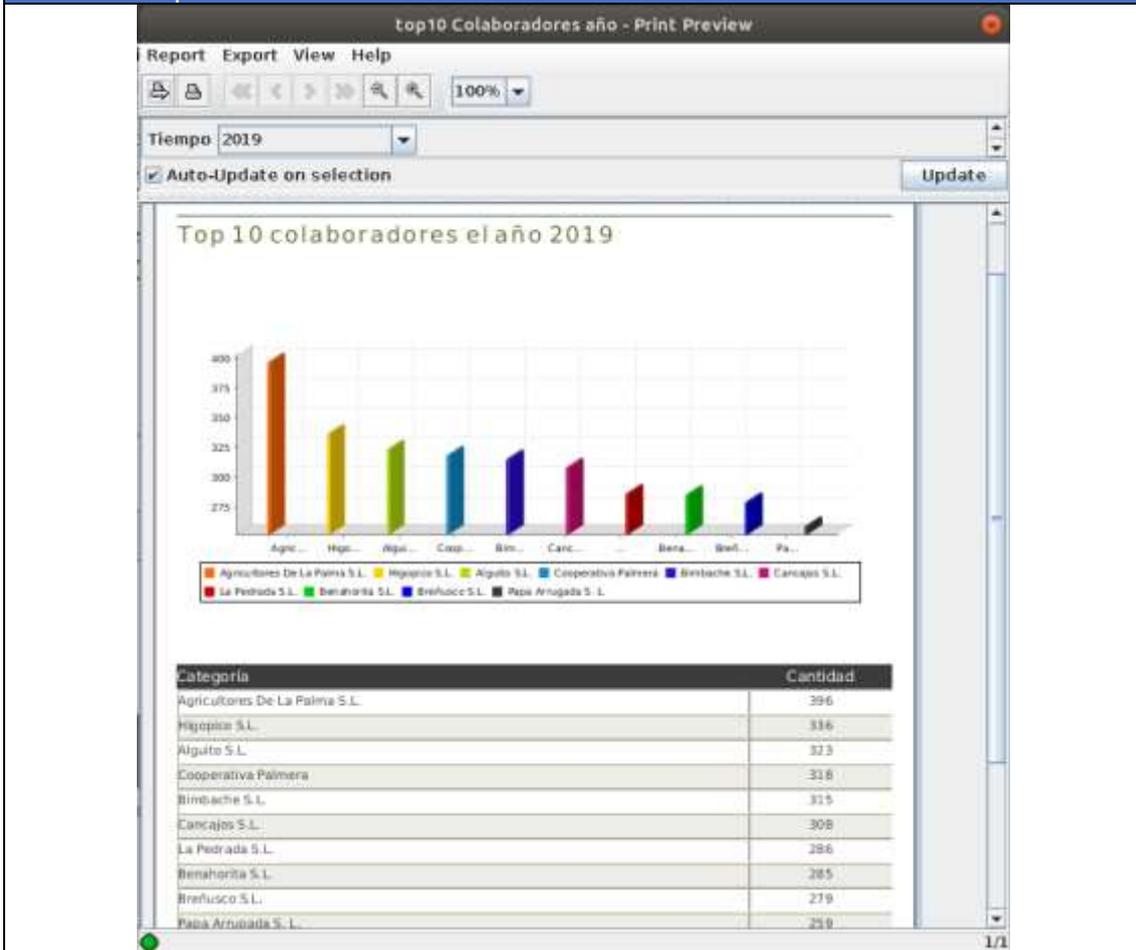
```
select NON EMPTY({Descendants([Producto].[All Productos]
,[Producto].[Categoria])}) on ROWS,
NON EMPTY({[Measures].[cuponesCanjeados]}) on Columns
from [Nutribox]
WHERE {[Tiempo].[${Tiempo}]}
```

Indicadores relacionados:	KPI 3 Porcentaje de canjes por Categoría de productos.
---------------------------	--

Tabla 60: Informe de canjes por categoría y año

4.7.2. Informe Top 10 colaboradores del año

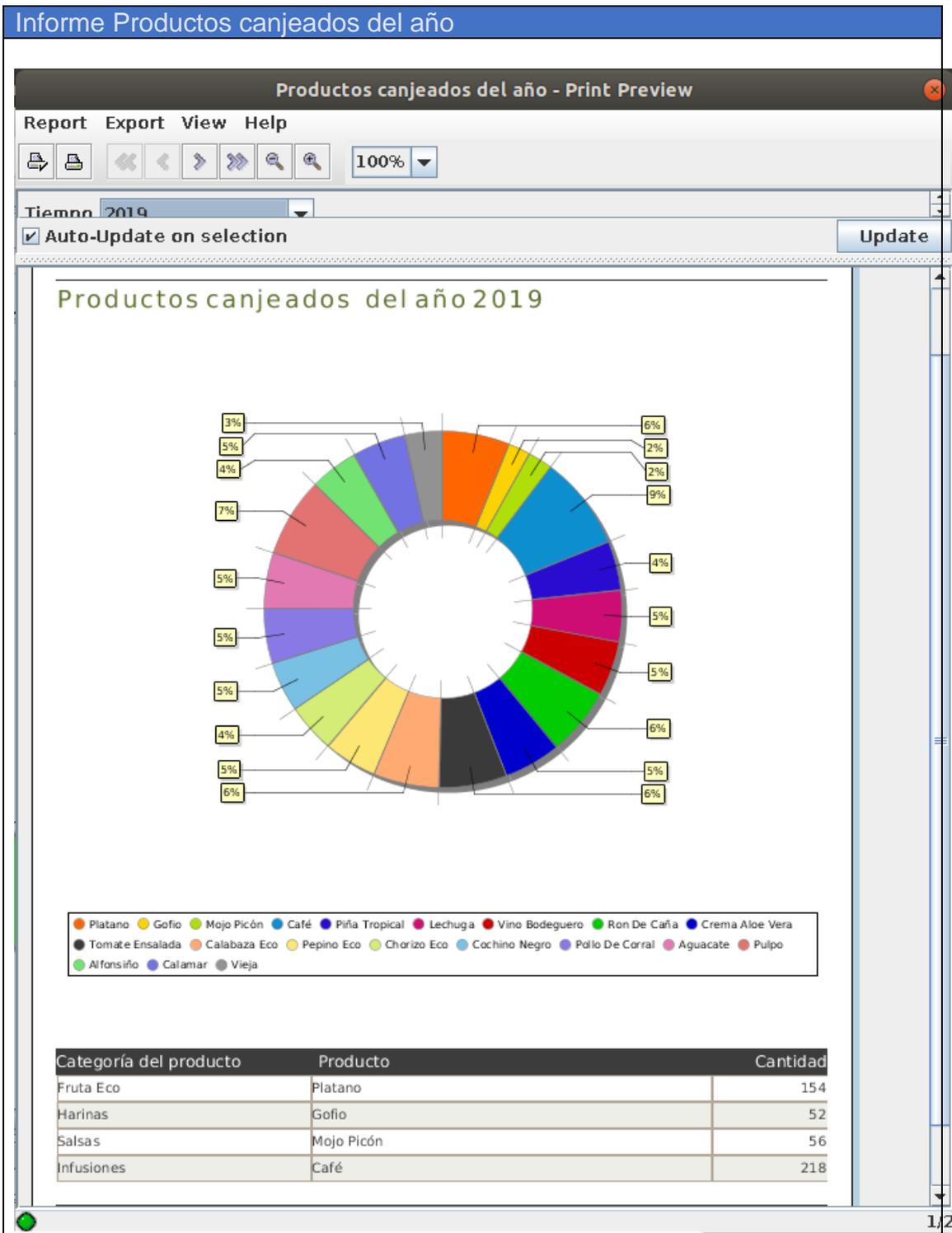
Informe Top 10 de colaboradores del año



<p>Descripción:</p>	<p>En este informe se representa en un gráfico de barras y en una tabla el número de canjes realizados a los productos ofertados por los 10 colaboradores más demandados de un año concreto. En la parte superior se ha incluido un selector para seleccionar el año a representar en el informe.</p>
<p>Consultas necesarias para la creación del informe:</p>	
<p>Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma.</p> <pre>with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName' select TopCount(filter({Descendants([Tiempo].[All Tiempos] ,[Tiempo].[Año])}, not isempty([Tiempo].CurrentMember))) , 50) on ROWS, {[Measures].[Name]} on Columns from [Nutribox]</pre> <p>A partir de este selector se crea el parámetro tiempo.</p> <p>Para obtener los datos que se representan en el gráfico y la tabla del informe se crea la siguiente consulta MDX al cubo OLAP Nutribox:</p> <pre>WITH SET [~ROWS] AS {[Colaborador].[Colaboradores].Members} SELECT NON EMPTY {[Measures].[cuponesCanjeados]} ON COLUMNS, NON EMPTY TopCount([~ROWS], 10, [Measures].[cuponesCanjeados]) ON ROWS FROM [Nutribox]</pre>	
<p>Indicadores relacionados:</p>	<p>KPI 4 Popularidad del colaborador.</p>

Tabla 61: Informe Top 10 colaboradores del año

4.7.3. Informe Productos canjeados del año



Descripción:

En este informe se representa en un gráfico de pastel el porcentaje de canjes por producto y en una tabla el nombre, su categoría y la cantidad de canjes de cada producto.

Consultas necesarias para la creación del informe:

Para el selector del año se ha usado una consulta MDX sobre el cubo OLAP Nutribox que recoge los años en los que ha habido actividad en la plataforma.

```
with member [Measures].[Name] as '[Tiempo].CurrentMember.UniqueName' select
TopCount( filter({Descendants([Tiempo].[All Tiempos] ,[Tiempo].[Año])}, not
isempty(([Tiempo].CurrentMember)) ) , 50) on ROWS,
{[Measures].[Name]} on Columns
from [Nutribox]
```

A partir de este selector se crea el parámetro tiempo.

Para obtener los datos que se representan en el gráfico y la tabla del informe se crea la siguiente consulta MDX al cubo OLAP Nutribox:

```
WITH
SET [~ROWS] AS
    {[Producto].[Producto].Members}
SELECT
NON EMPTY {[Measures].[cuponesCanjeados]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Nutribox]
WHERE ([Tiempo].[${Tiempo}])
```

Indicadores relacionados:	KPI1 Popularidad del producto.
---------------------------	--------------------------------

Tabla 62: Informe Productos canjeados del año

5. Líneas de futuro

La solución BI de la empresa Nutribox deberá adaptarse a las necesidades de la organización de forma que sea útil para obtener conocimiento a través de los datos disponibles. Para ello se seguirá un modelo de desarrollo iterativo de la solución donde se continuarán las fases de análisis, diseño e implantación de los cambios de las diferentes versiones de esta solución.

Así una vez implantada una versión se deberá iniciar una fase de análisis donde, como se ha hecho en este proyecto, se deberán plantear los indicadores y fuentes de datos con el objetivo de valorar el cumplimiento de los nuevos objetivos establecidos.

Ampliación del cubo OLAP existente definiendo nuevas dimensiones, ampliar las existentes, o la creación de nuevos cubos. Así, por ejemplo, se propone la creación de los niveles Semana y Hora, que permitan realizar nuevos análisis a partir de los datos disponibles.

Para siguientes iteraciones o versiones de la solución se propone el registro y tratamiento de otros datos obtenidos a partir de la actividad de

la plataforma que no han sido valorados en este prototipo inicial, como pueden ser la fecha de registro de los usuarios, que permitirá estudiar la evolución de registros y nuevos usuarios de la plataforma.

Además, con la finalidad de ofrecer un mejor servicio de los usuarios, se hace necesario un conocimiento mayor de los mismos por lo que se contempla que en el registro se obtengan de forma voluntaria otros datos personales y que estos puedan ser tratados en la solución de business intelligence de la compañía.

Del mismo modo, en posteriores evoluciones de la solución se plantea el uso de técnicas de web scraping a través de la herramienta Pentaho Data Integration usada en este proyecto, para obtener datos que puedan dar lugar a información relevante para la compañía.

Por ejemplo, el uso de estas técnicas sobre diferentes redes sociales puede suministrar un mayor conocimiento sobre los posibles usuarios de la compañía. También pueden servir para obtener datos disponibles en diferentes repositorios de Open Data que pueden aportar información acerca del entorno o necesidades de estos usuarios.

Por otro lado, el uso de datos geográficos o de localización pueden servir para proveer soluciones a determinados problemas. Así el tratamiento de estos datos y representación de estos a través de la solución propuesta, podrá ser otra posibilidad a tener en cuenta en evoluciones futuras.

Así, por ejemplo, a través de datos de localización de usuarios se podrá medir la distancia desde estos a los diferentes establecimientos de las empresas colaboradoras de Nutribox y representarlos por medio de componentes de mapa (Map component) en los cuadros de mando realizados por la solución BI a través de Pentaho, de forma que esta información resulte accesible y fácilmente interpretada.

Con el objeto de dotar a la plataforma de una mayor capacidad analítica se podrán integrar otras herramientas como Weka, que se encuentra disponible como plugin de Pentaho y permite aplicar algoritmos de machine learning (aprendizaje automático) y minería de datos.

Así, por ejemplo, se podrían utilizar algoritmos de agrupamiento con la finalidad de clasificar los diferentes perfiles de usuarios o el uso de árboles de decisión que permitirán observar patrones de compra de nuestros usuarios y poder realizar sugerencias de compras.

Finalmente, si el volumen de datos a tratar por la solución excediera a las capacidades de la base de datos actual, se podrá plantear el uso de Hadoop, que es actualmente soportado por Pentaho y permite el tratamiento escalable y distribuido de los datos aumentando la capacidad del sistema.

6. Conclusiones

En primer lugar, cabe destacar la dificultad que ha supuesto basar este trabajo en una idea de negocio propuesta a partir de la iniciativa de la UOC “Emprendedores interestudios” a la hora de definir, planificar y ejecutar este proyecto.

Como integrantes de esta iniciativa hemos trabajado sobre temáticas muy diferenciadas, las compañeras de Administración de empresas en el ámbito empresarial y yo sobre una solución de BI para esta empresa, por lo que finalmente la mayoría del trabajo realizado lo hemos llevado a cabo de forma autónoma.

Desde mi punto de vista, para que surgieran las sinergias esperadas entre los diferentes perfiles, era necesario un perfil de desarrollo o ingeniería de software que tratara el caso de la plataforma sobre la que se basa el servicio de Nutribox y sirviera de nexo entre los demás trabajos.

Por otro lado, puesto que esta empresa se encuentra en fase de idea, no se dispone de una fuente de datos real sobre la que trabajar por lo que ha sido necesario el diseño de una base de datos inicial sobre la que se basa esta solución BI y la creación de unos datos ficticios sobre los que poder mostrar los resultados.

Además, como esta empresa se encuentra en proceso de definición, no se tienen referencias sobre los posibles problemas o necesidades de esta organización, lo que ha dificultado en gran medida la definición de objetivos, alcance e indicadores a tratar por esta solución.

Tenido estos hechos en cuenta, se ha definido y simulado la base de datos operativa de la plataforma introduciendo un conjunto de datos o dataset de prueba. Posteriormente se han definido los principales indicadores o KPI a valorar por él prototipo.

Utilizando herramientas Open Source se ha procedido a la definición y construcción de los procesos de extracción, transformación, carga (ETL), construcción de un sistema de almacenamiento y gestión de datos (data warehouse), diseño de cubos OLAP, así como el diseño y construcción de cuadros de mando e informes que presenten los indicadores (KPI) definidos en la fase de análisis.

Hay que destacar que el uso de diferentes herramientas, como Pentaho Data Integration, Pentaho Schema Workbench, Pentaho report designer o pentaho cde, ha supuesto una gran dedicación para comprender su funcionamiento y configuración; y así poder obtener los resultados esperados.

Finalmente, se han descrito una serie de líneas de futuro que pueden servir como base para las futuras versiones de la solución implementada,

por lo que se puede considerar que se han alcanzado los objetivos planteados en la definición del trabajo.

7. Bibliografía

- Indicador clave de rendimiento (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Indicador_clave_de_rendimiento>
- Almacén de datos (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos>
- Almacén de datos (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos>
- Esquema en estrella (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Esquema_en_estrella>
- OLAP (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <<https://es.wikipedia.org/wiki/OLAP>>
- Oracle VM VirtualBox (2019) [En línea] de ORACLE [Consulta: 1 de Junio de 2019] <<https://www.oracle.com/es/virtualization/virtualbox/>>
- Oracle MySQL (2019) [En línea] de ORACLE [Consulta: 1 de Junio de 2019] <<https://www.oracle.com/es/mysql/>>
- PostGreSQL vs. MySQL (2016) [En línea] de Daniel Pecos Martínez [Consulta: 1 de Junio de 2019] <<https://danielpecos.com/documents/postgresql-vs-mysql/>>
- Infografía MySQL vs PostgreSQL (2016) [En línea] de infranetworking [Consulta: 1 de Junio de 2019] <<https://blog.infranetworking.com/infografia-mysql-vs-postgresql/>>
- Extract, transform and load (2019) [En línea] de Wikipedia [Consulta: 1 de Junio de 2019] <https://es.wikipedia.org/wiki/Extract,_transform_and_load>

8. Anexos

8.1. Descarga y uso del prototipo.

Se ha alojado el archivo Prototipo.ova en OneDrive para que pueda ser descargado a través del siguiente enlace:

https://uoc0-my.sharepoint.com/:u:/g/personal/nayut_uoc_edu/EapS0JNmHsJLsM8nPfDeJzMBBGjO6mM2fg03fBFECdnrdw?e=TmS6b2

Para poder utilizar este archivo que contiene el servicio virtualizado, es necesario tener instalado el software VirtualBox que puede ser descargado desde:

<https://www.virtualbox.org/wiki/Downloads>

Desde esta aplicación se deberá importar el servicio virtualizado a través del menú Archivo, seleccionar la opción “Importar el servicio virtualizado” como se muestra en la siguiente imagen:

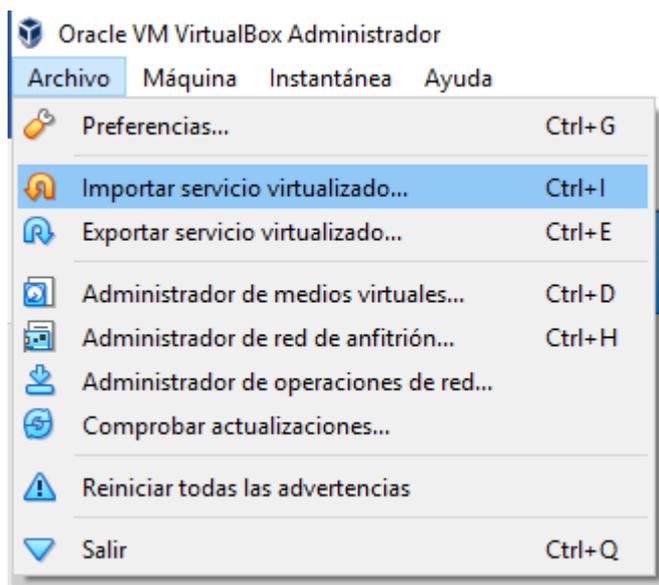


Imagen 19: Importar servicio virtualizado.

Se mostrará una ventana donde se deberá indicar la ruta del archivo Prototipo.ova:

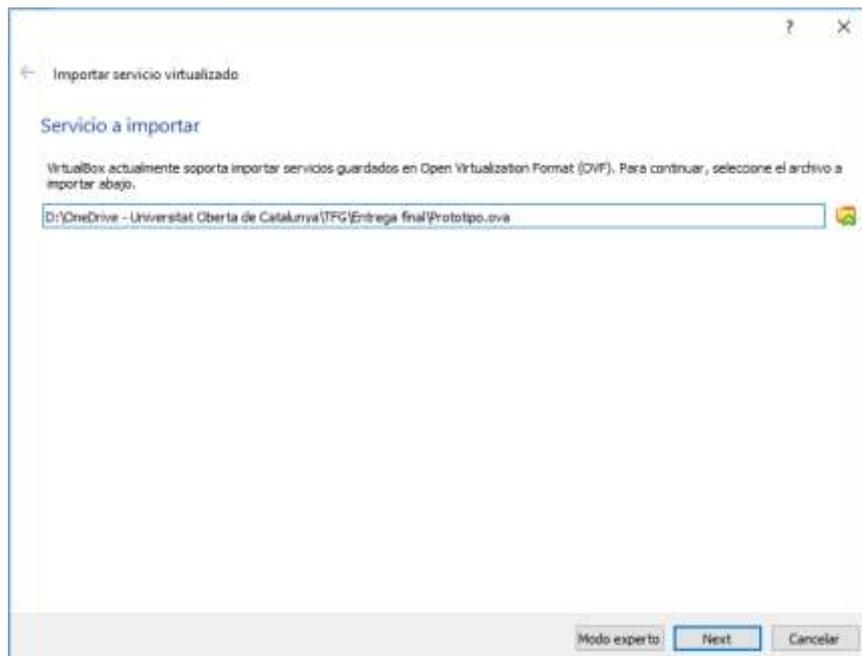


Imagen 20: Indicar la ruta del archivo.

Tras pulsar en siguiente o “Next” se mostrará la información del archivo y se podrá indicar la carpeta que contendrá las máquinas virtuales:

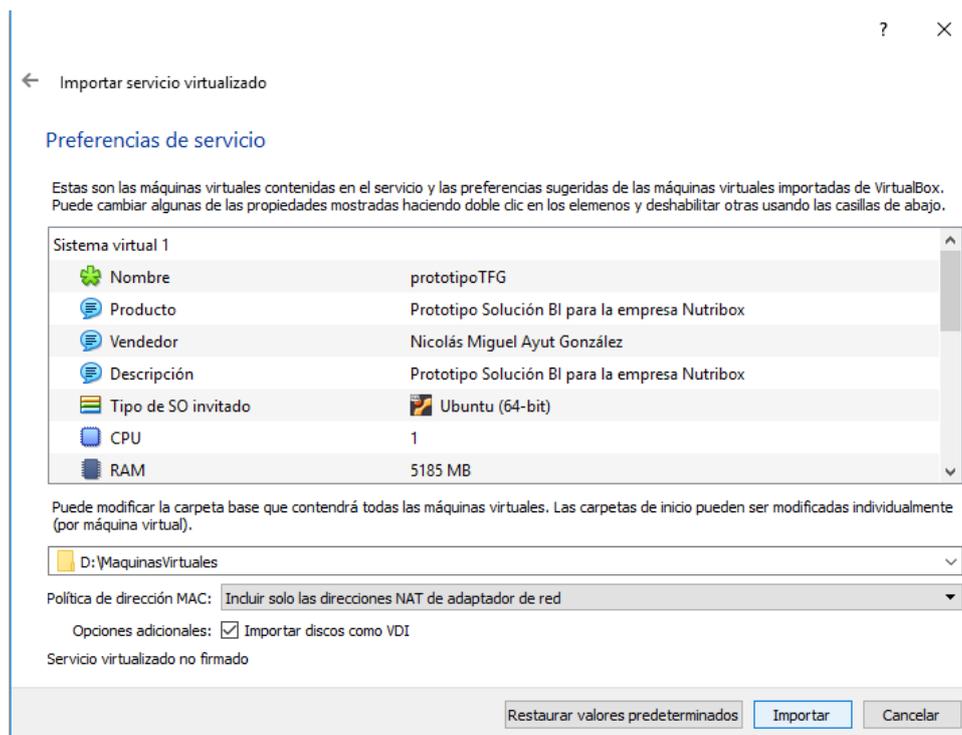


Imagen 21: Sistema virtualizado.

Tras pulsar sobre la opción Importar en la ventana anterior comenzará el proceso de importación:

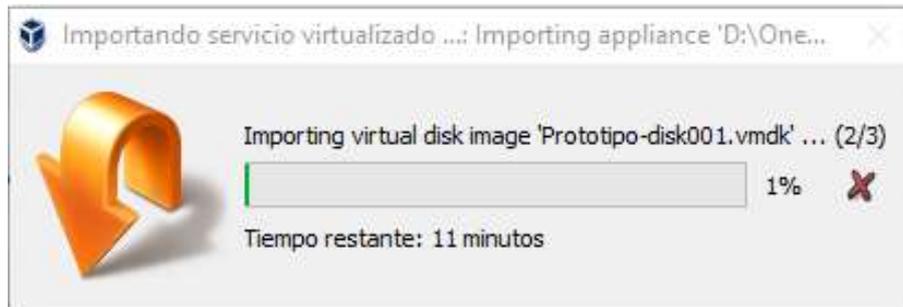


Imagen 22: Proceso de importación

Una vez finalizado el proceso de importación y tras seleccionar la máquina virtual prototipoTFG como se ve en la imagen, se podrá iniciar la máquina usando el botón de Iniciar:



Imagen 23: Iniciar máquina virtual.

Se iniciará la máquina virtual y automáticamente cargará la sesión del usuario por defecto (usuario: nico, contraseña: 123456). Se ha configurado este servicio para que tras la carga del sistema operativo inicie el servidor Pentaho Business Analytics.

Se recomienda esperar unos minutos antes de empezar a utilizarlo ya que los servicios del servidor Pentaho tardan en estar listos para su uso. Una vez el servidor esté operativo se podrá acceder a él desde el navegador introduciendo la url: localhost:8080/:

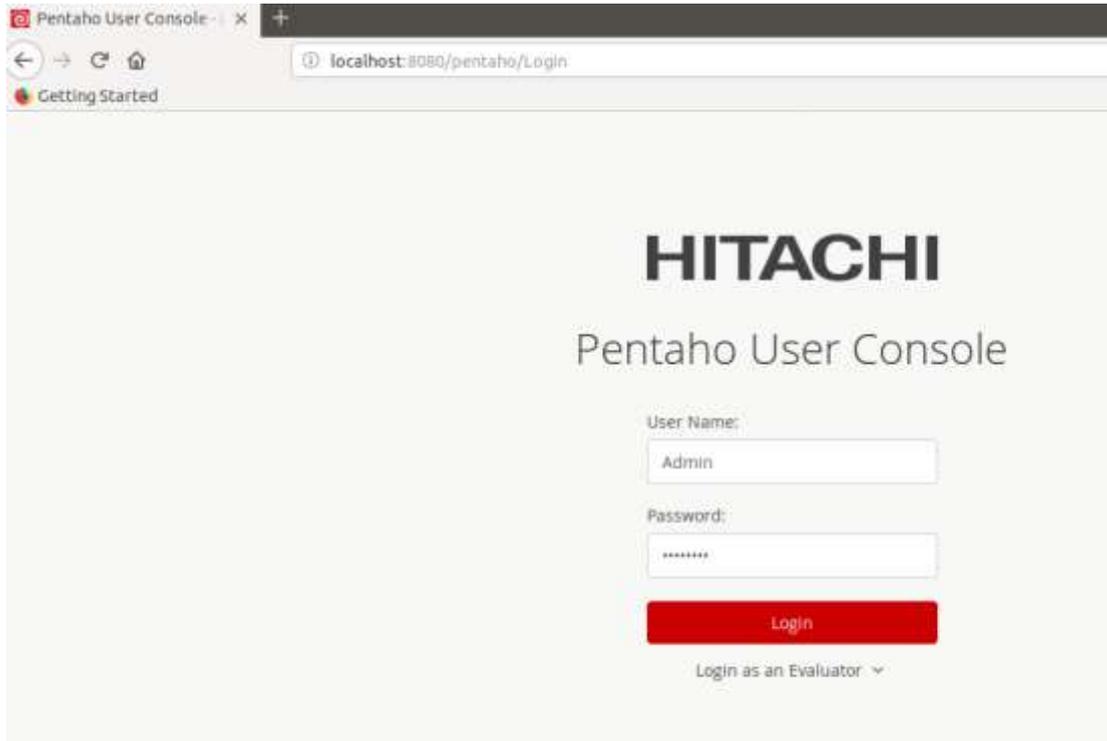


Imagen 24: Acceder al servidor Pentaho Business Analytics.

Tras validar con el usuario administrador por defecto (Usuario: Admin, contraseña: password). Será posible acceder a todos los módulos utilizados en este prototipo, así como a los cuadros de mando e informes publicados en este prototipo. Así para poder acceder a estos archivos, se debe pulsar sobre la opción Browse Files del servidor Pentaho Business Analytics:

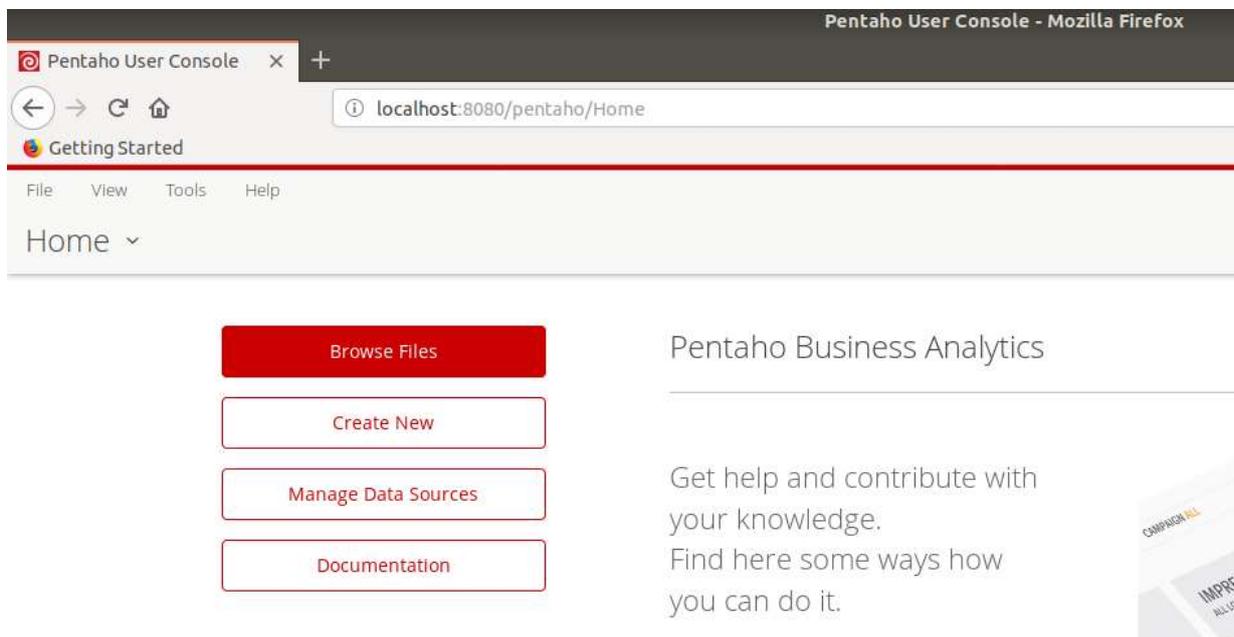


Imagen 25: Detalle Pentaho Business Analytics.

Al abrir, esta opción se muestra una sección de carpetas donde se puede observar todo el árbol de directorios del servidor y una sección de archivos, donde se muestra el contenido de la carpeta seleccionada en la sección anterior. Los archivos publicados en este trabajo se han ubicado en la carpeta Public, tras crear los directorios CuadrosDeMando e Informes como puede verse en la siguiente imagen:

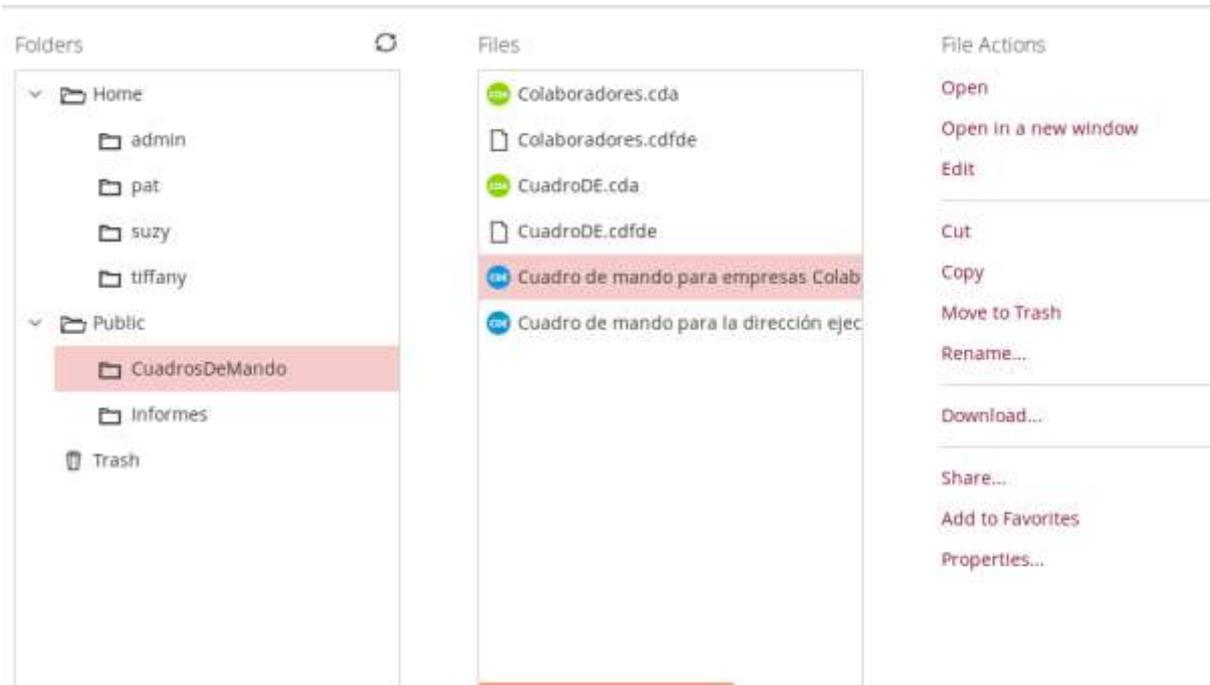


Imagen 26: Browse files en Pentaho Business Analytics.

Así se podrá realizar diferentes acciones sobre los archivos como abrirlo, abrirlo en una nueva ventana o editarlo, usando las acciones de la columna derecha.

Para utilizar las herramientas pentaho report designer, pentaho data integration y pentaho schema workbench, se deberán ejecutar desde terminal los siguientes comandos:

Para Pentaho data integration:

```
sudo sh /home/nico/Downloads/pdi-ce-8.2.0.0-342/data-integration/spoon.sh
```

Para Pentaho report designer:

```
sudo sh /home/nico/Downloads/prd-ce-7.1.0.0-12/report-designer/report-designer.sh
```

Para Pentaho schema workbench:

```
sudo sh /home/nico/Downloads/psw-ce-3.14.0.0-12/schema-workbench/workbench.sh
```

8.2. Uso de generatedata.com para crear los datos de prueba

En este anexo se detallan los pasos seguidos para obtener los distintos datos de pruebas a partir de esta web. Así, para obtener el código que generará los usuarios se introduce los siguientes valores en la web:

generatedata.com Conexión | Español

Generar Acerca de Noticias Donar

Users SALVAR

DATOS ESPECÍFICOS DE PAÍS ?

Spain X

CONJUNTO DE DATOS ?

Orden	Columna de la tabla	Tipo de dato	Ejemplos	Opciones	Ayuda	Elim
1	UserEmail	Correo electrónico	No hay ejemplos disponibles.	No hay opciones disponibles.	?	
2	UserPassword	Alfanumérico	eZg29gdF5K1 (Contraseña)	LLLxxLLxLL	?	
3	UserFirstName	Nombres	Alex (cualquier sexo)	Name	?	
4	UserLastName	Nombres	Smith (apellido)	Surname	?	
5	UserCity	Ciudad	No hay ejemplos disponibles.	No hay opciones disponibles.	?	
6	UserState	Región	No hay ejemplos disponibles.	<input checked="" type="checkbox"/> Spanish Autonomies <input checked="" type="checkbox"/> Complete Porto	?	
7	UserZip	Código postal	No hay ejemplos disponibles.	<input checked="" type="checkbox"/> Spanish Autonomies	?	
8	UserPhone	Teléfono / Fax	Formatos diferentes	6xxxxxxx	?	
9	UserCountry	País	No hay ejemplos disponibles.	<input type="checkbox"/> Limitar los países a los elegidos más arriba	?	
10	UserAddress	Dirección	No hay ejemplos disponibles.	No hay opciones disponibles.	?	

Orden Columna de la tabla Tipo de dato Ejemplos Opciones Ayuda Elim

Añadir

TIPOS DE EXPORTACIÓN ?

CSV Excel HTML JSON LDIF Lenguaje de programación: SQL XML - ocultar opciones de formato de datos

Nombre de tabla Tipo de sentencia INSERT INSERT IGNORE UPDATE

Tipo de base de datos INSERT tamaño de lote

Opciones varias Incluir sentencia CREATE TABLE Incluir sentencia DROP TABLE Entrecomillar nombres (tablas, columnas) Ninguno Añadir columna auto-incrementada por defecto

Imagen 27: Generar datos de usuarios.

Para obtener el código que generará los cupones se introduce los siguientes valores:

generatedata.com Login | English

Generate About News Donate

coupons SAVE

COUNTRY-SPECIFIC DATA

Spain ×

DATA SET

Order	Table Column	Data Type	Examples	Options	Help	Del
1	CollaboratorID	Number Range	No examples available.	Between <input type="text" value="1"/> and <input type="text" value="20"/>		<input type="checkbox"/>
2	ProductID	Number Range	No examples available.	Between <input type="text" value="1"/> and <input type="text" value="20"/>		<input type="checkbox"/>
3	Discount	Number Range	No examples available.	Between <input type="text" value="1"/> and <input type="text" value="100"/>		<input type="checkbox"/>

Order Table Column Data Type Examples Options Help Del

Add Row(s)

EXPORT TYPES

CSV Excel HTML JSON LDIF Programming Language **SQL** XML - hide data format options

Database table name Statement Type INSERT INSERT IGNORE UPDATE

Database Type INSERT batch size

Misc Options: Include CREATE TABLE query Primary Key None

Include DROP TABLE query Add default auto-increment column

Enclose table and field names with backquotes

Generate rows Generate in-page New window/tab Prompt to download Zip? Generate

3.2.8 | Documentation | Report a bug

Imagen 28: Generar datos de cupones.

Para obtener el código que generará las transacciones se introduce los siguientes valores:

generatedata.com Login | English

Generate About News Donate

couponUse SAVE

COUNTRY-SPECIFIC DATA
Spain

DATA SET

Order	Table Column	Data Type	Examples	Options	Help	Del
1	CouponID	Number Range	No examples available.	Between 1 and 200	?	
2	UserID	Number Range	No examples available.	Between 1 and 1000	?	
3	UseDate	Date	MySQL datetime	From: 04/19/2018 To: 04/19/2020 Format code: Y-m-d H:i:s	?	
4		Select Data Type				

Add 1 Row(s)

EXPORT TYPES
CSV Excel HTML JSON LDIF Programming Language SQL XML - hide data format options

Database table name: couponUse
Database Type: MySQL
Statement Type: INSERT INSERT IGNORE UPDATE
INSERT batch size: 100
Primary Key: None Add default auto-increment column
Misc Options:
 Include CREATE TABLE query
 Include DROP TABLE query
 Enclose table and field names with backquotes

Generate 100 rows Generate in-page New window/tab Prompt to download Zip? **Generate**

Imagen 29: Generar transacciones.

Tanto el código generado como el obtenido de generatedata.com se copia a un archivo nombrado insertarDatos.sql.

8.3. Procesos ETL con Pentaho Data Integration

En este anexo se indica a modo de ejemplo cómo se ha creado el archivo HechosCanje.ktr en la aplicación Pentaho Data Integration. El primer paso será crear la nueva transformación, esto lo podemos hacer desde el menú file, new, transformation:

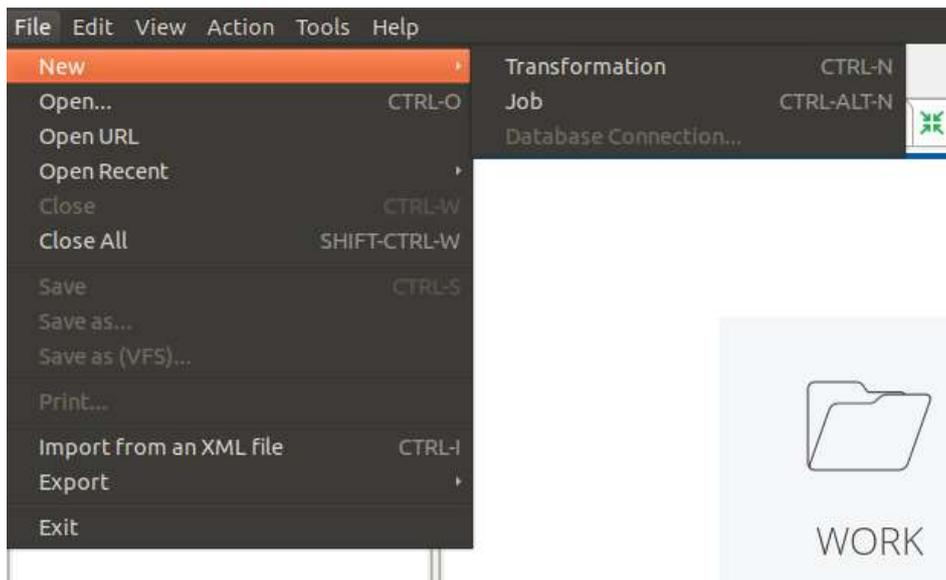


Imagen 30: Nueva Transformación PDI.

O bien desde la pantalla de bienvenida, pulsando sobre la opción “New Transformation”:

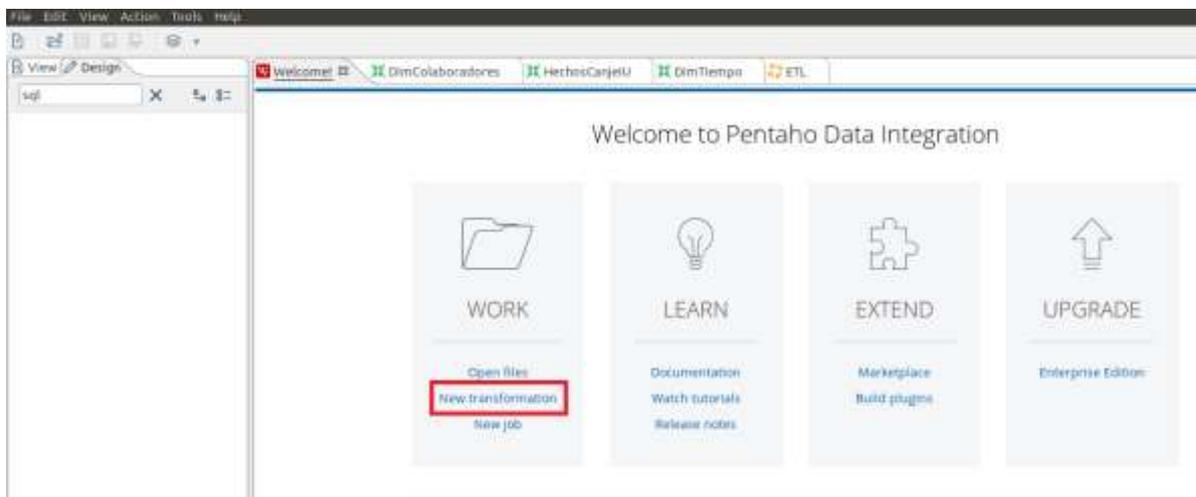


Imagen 31: Otra Nueva transformación.

En el área de trabajo iremos arrastrando los diferentes elementos desde la columna de elementos o design:

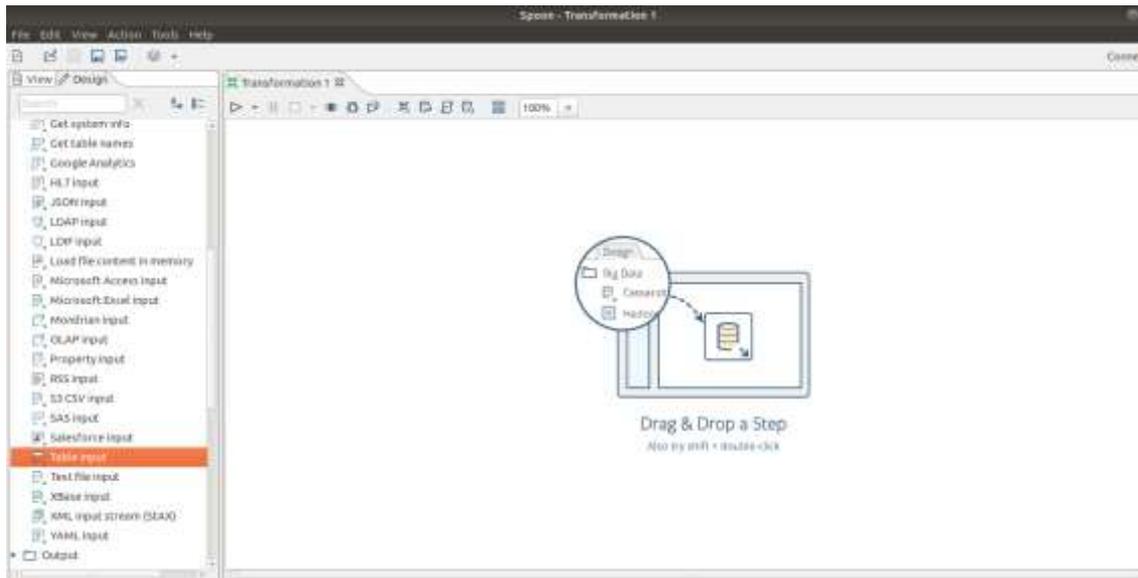


Imagen 32: Área de trabajo PDI.

Como vamos a obtener como origen de datos para este trabajo dos tablas de la base de datos Nutribox (coupons y couponUse) se arrastran dos objetos table input que configuraremos haciendo doble clic sobre ellos:

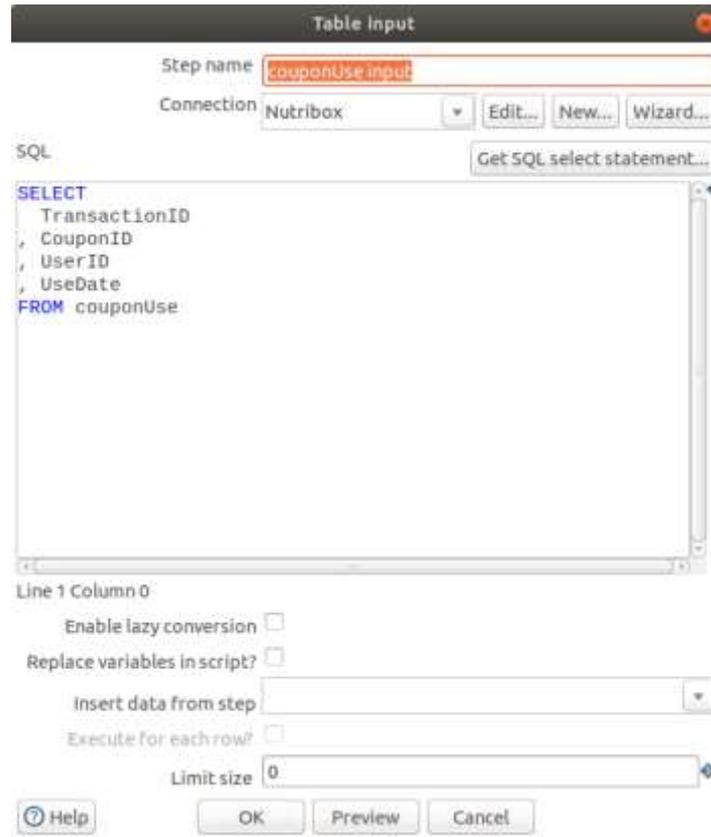


Imagen 33: Detalle del objeto table input para la tabla couponUse.

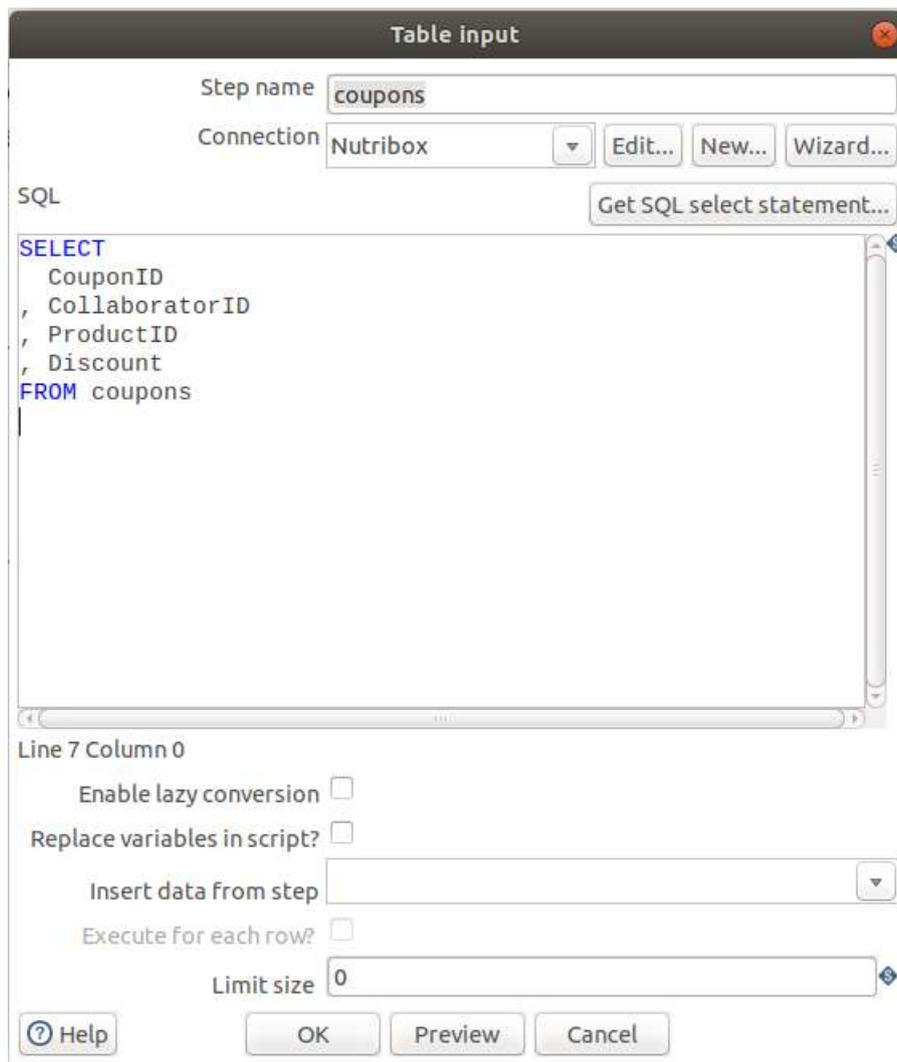


Imagen 34: Detalle del objeto table input para la tabla coupons.

Para poder unir estas tablas deberemos previamente ordenar sus registros por el atributo CouponID. Para ello cada uno de los dos elementos Table Input creados anteriormente se conectarán a elementos Sort rows que se configurarán de la siguiente forma:

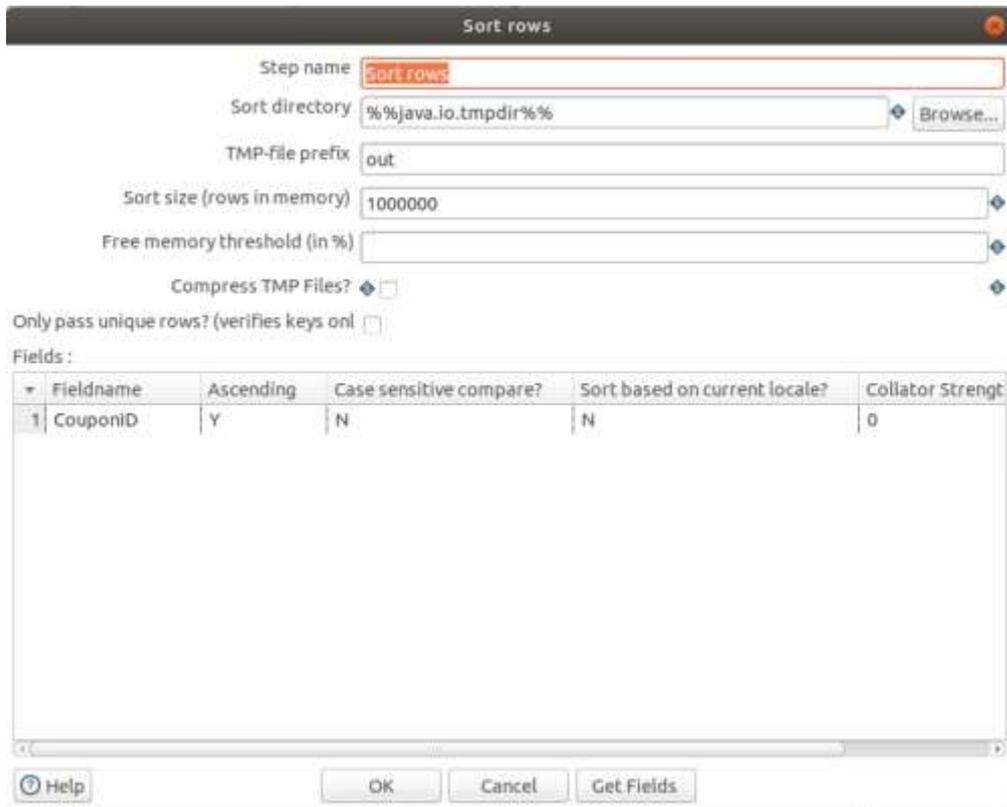


Imagen 35: Detalle del objeto Sort rows.

Cada uno de los elementos Sort rows (que han ordenado las tablas por CouponID se conectarán a un elemento Merge join que unirá las dos tablas (coupons y couponUse), para lo cual habrá que configurarlo de la siguiente forma:

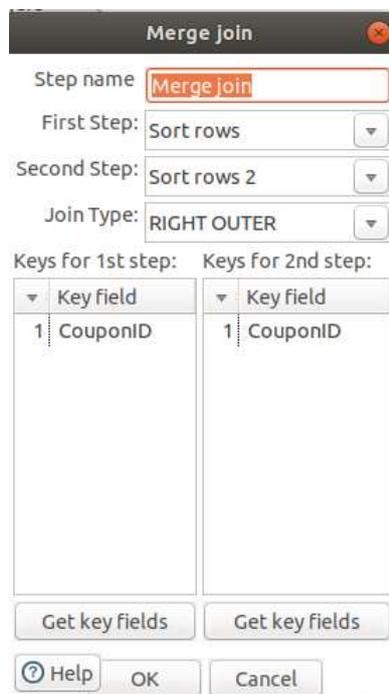


Imagen 36: Detalle del objeto Merge join.

Con la finalidad de crear dos campos, fechasku y horasku a partir del atributo Usedate, se usa un elemento Calculator que crea un campo UseDateString que será una copia del atributo Usedate:

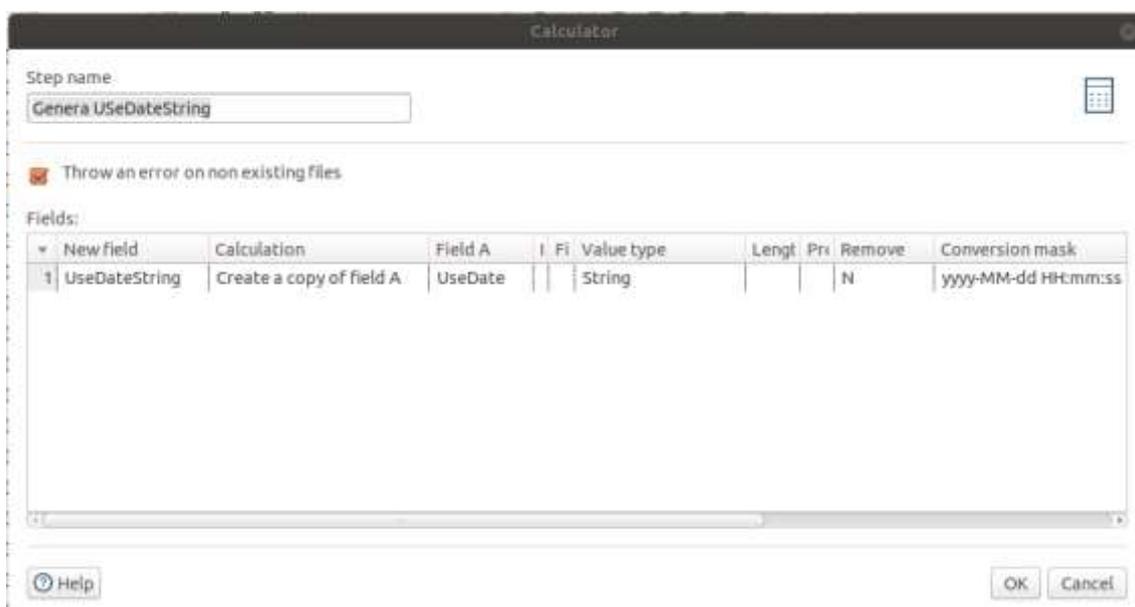


Imagen 37: Detalle objeto Calculator

Y posteriormente un Split fields que genera fechasku y horasku a partir de UseDateString:

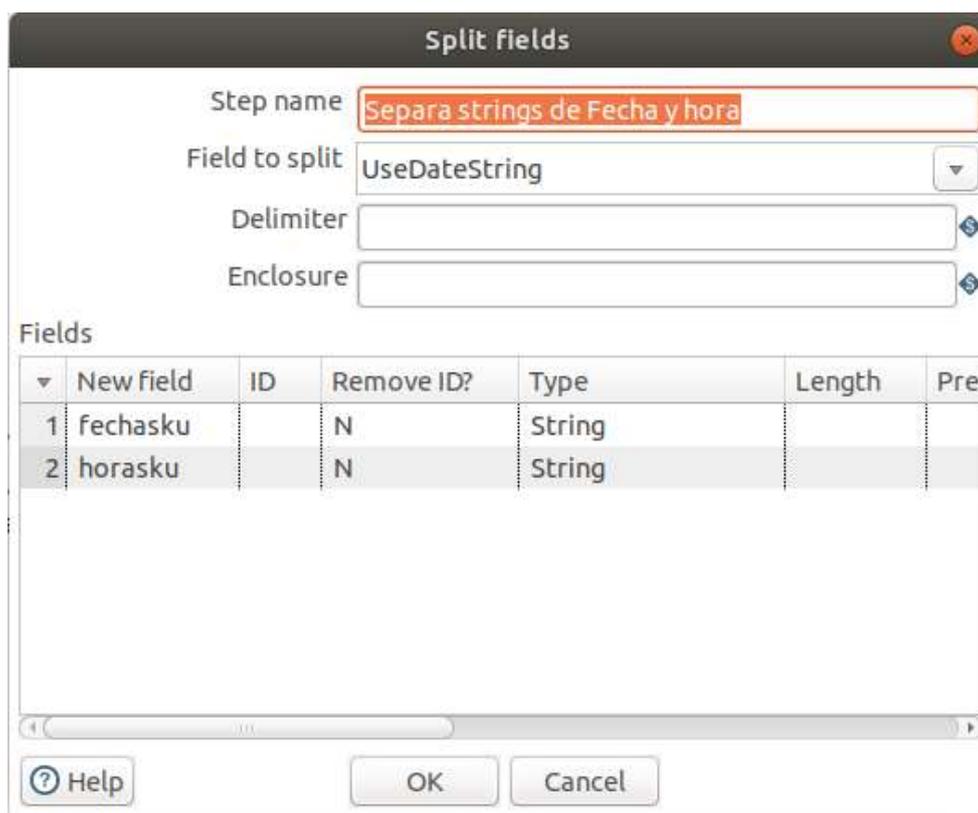


Imagen 38: Detalle objeto Split fields.

Posteriormente con un elemento String operations, se eliminan los caracteres especiales como “.” y espacios que originalmente tienen los atributos fechasku y horasku:

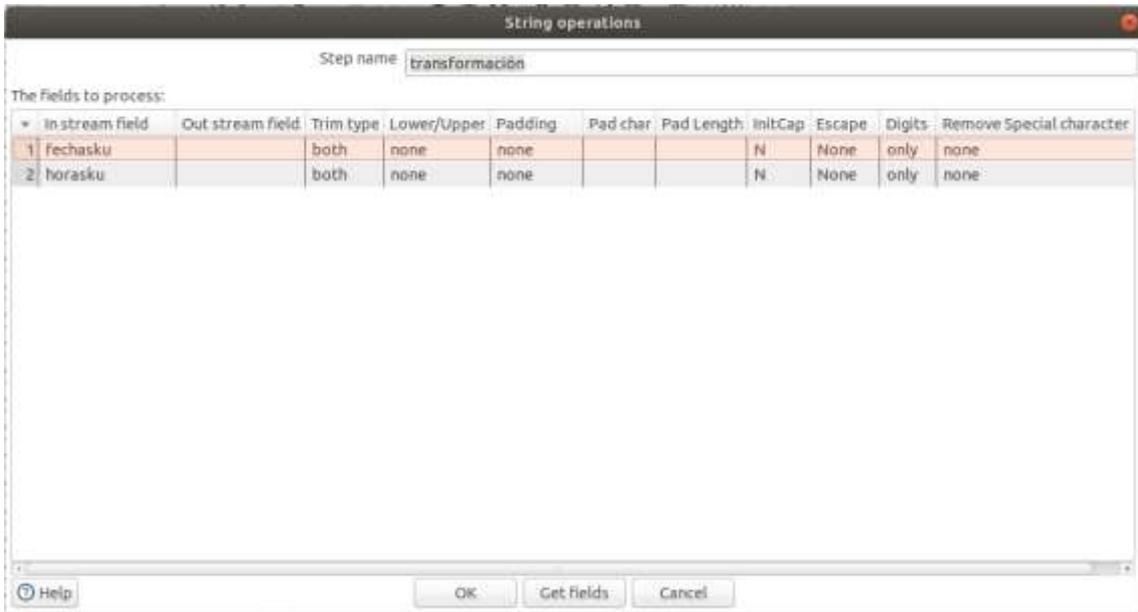


Imagen 39: Detalle objeto String operations.

Se conecta a otro elemento Calculator que a partir de fechasku y horasku, que son de tipo string, los nuevos campos fechaid y horasid del tipo Integer:

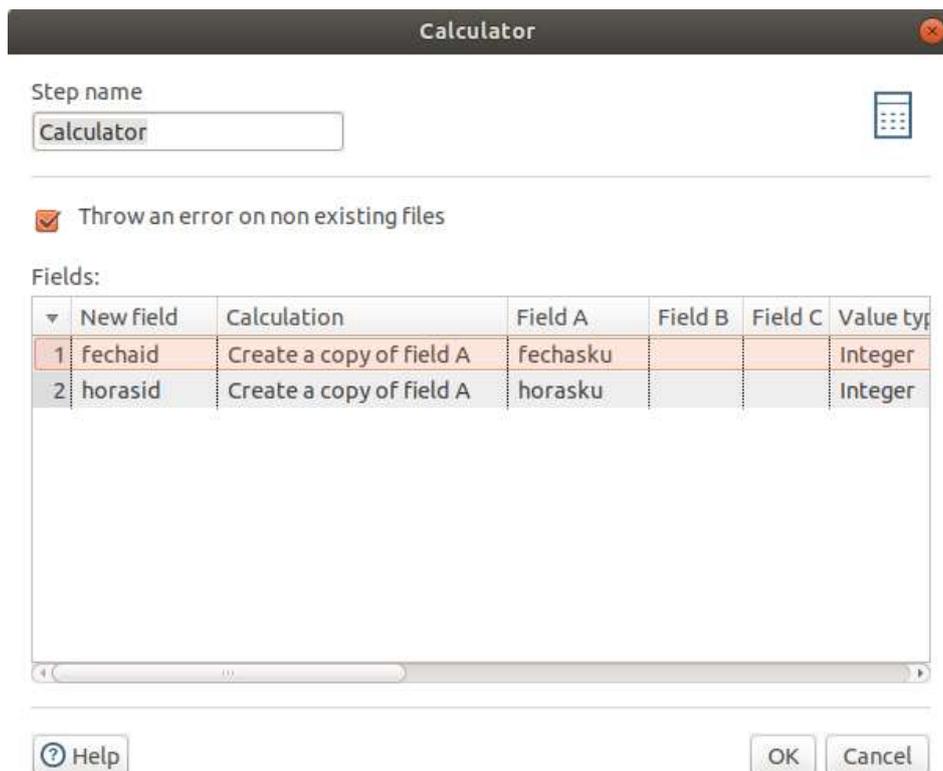


Imagen 40: Detalle objeto Calculator2.

Finalmente, con un elemento Combinatio lookup/update almacenará los atributos TransactionID, CouponID, UserID, Discount, fechaid, CollaboratorID, ProductID en la base de datos datawarehouse generando los atributos transactionSKU y versión, de forma que si estos datos se actualizan en la base de datos original en la base de datos datawarehouse se mantendrán los datos originales y se copiará la nueva versión:

The screenshot shows the 'Combination lookup/update' dialog box with the following configuration:

- Step name: Combination lookup/update
- Connection: datawarehouse
- Target schema: (empty)
- Target table: HechosCanje
- Commit size: 100
- Cache size: 9999
- Pre-load the cache?:

Key fields (to look up row in table):

Dimension field	Field in stream
1 TransactionID	TransactionID
2 CouponID	CouponID
3 UserID	UserID
4 Discount	Discount
5 fechaid	fechaid
6 CollaboratorID	CollaboratorID
7 ProductID	ProductID

Technical key field: transactionSKU

Creation of technical key:

- Use table maximum + 1
- Use sequence
- Use auto increment field

Remove lookup fields?

Use hashcode?

Hashcode field in table: (empty)

Date of last update field (optional): (empty)

Buttons: Help, OK, Cancel, Get Fields, SQL

Imagen 41: Detalle objeto Combination lookup/update.

Con estos elementos conectados como se ha indicado se obtiene el trabajo completo HechosCanje.ktr:



Imagen 42: Trabajo HechosCanjes.

8.4. Definición del esquema OLAP

En este anexo se detalla el proceso de creación del esquema OLAP que se ha usado en el prototipo.

Para la creación del esquema el primer paso después de abrir la aplicación es definir la conexión a los datos. Para ello desde la aplicación vamos al menú “Options” y se pulsa en el apartado “Connection...”

Como se muestra en la siguiente ventana:

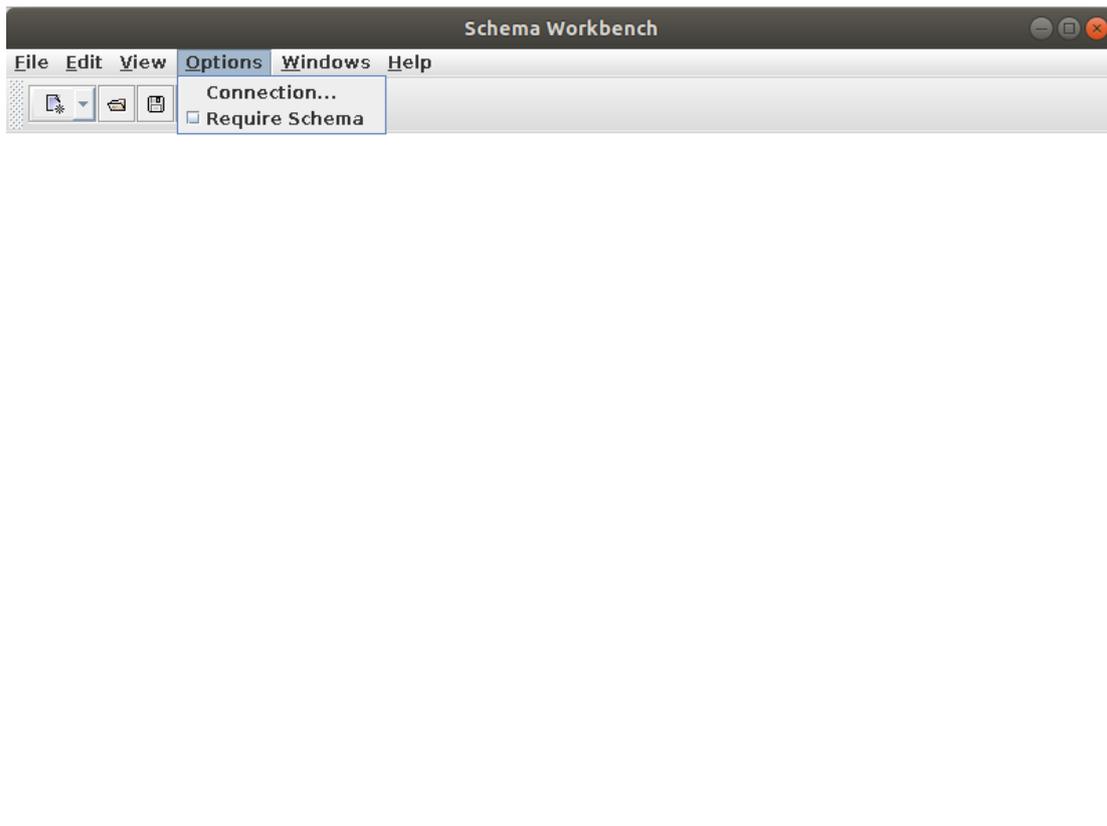


Imagen 43: Área de trabajo Pentaho Schema Workbench.

Para este caso será necesario establecer como tipo de conexión MySQL con la base de datos datawarehouse, alojada en localhost, usando el usuario admin, contraseña 12345678 y puerto número 3306; como se puede ver en la siguiente imagen:

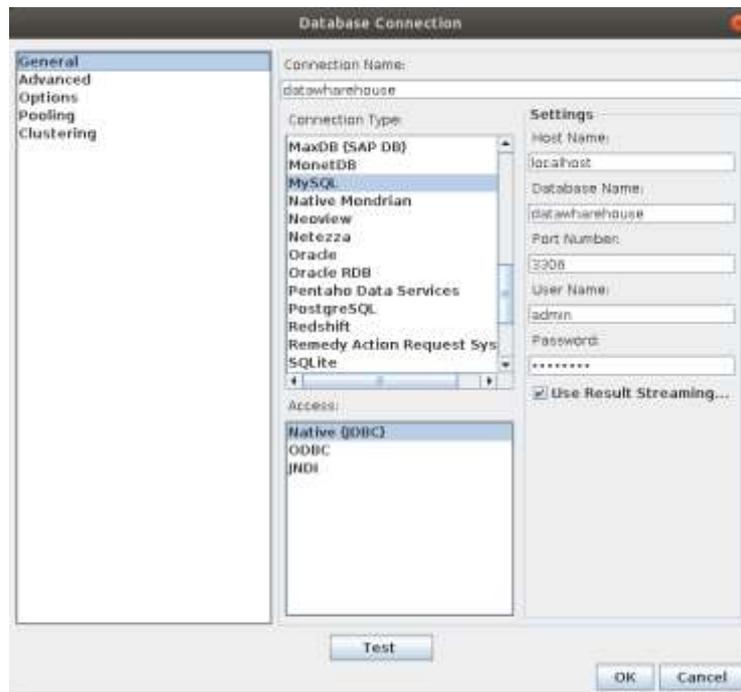


Imagen 44: Detalle conexión con la base de datos.

Tras pulsar ok, volveremos a la ventana inicial y seleccionaremos crear un nuevo esquema a través de “file”, “new”, “schema” en el menú superior. Si pulsamos sobre el icono del schema podremos añadir cubos y dimensiones en el siguiente menú:

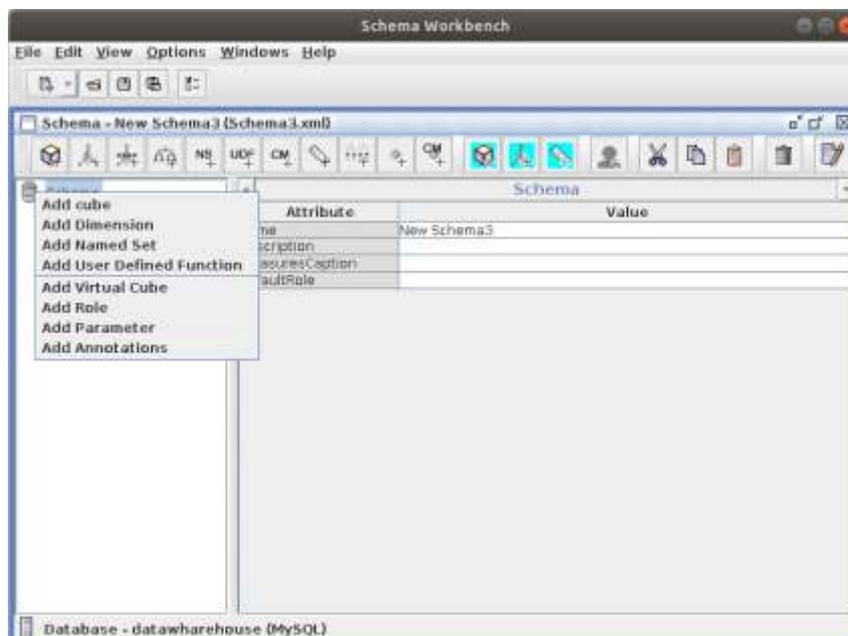


Imagen 45: Área de trabajo Pentaho Schema Workbench 2.

Las dimensiones se pueden generar internamente en los cubos o como se ha hecho en este caso en la raíz del esquema de forma que puedan ser reutilizadas por diferentes cubos. En el esquema creado para este proyecto se crean las dimensiones Colaborador, Usuario, Tiempo y Producto, como se puede ver en la siguiente imagen:

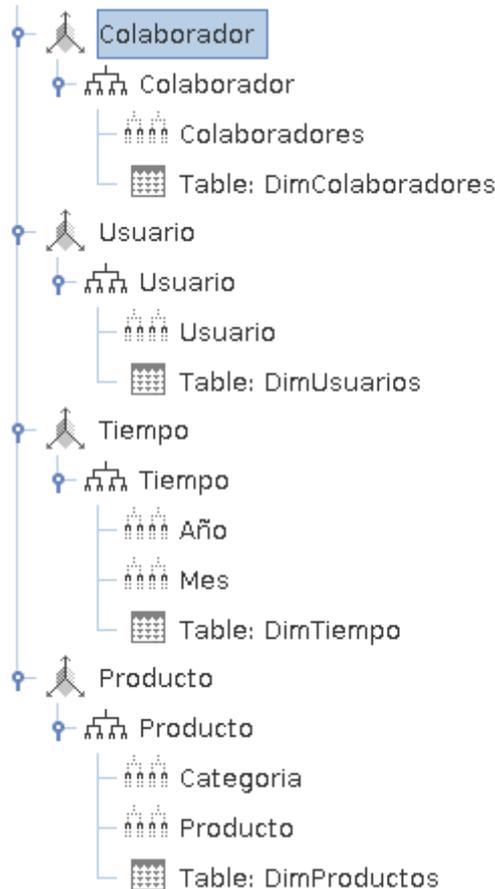


Imagen 46: Dimensiones del cubo.

Para cada dimensión se debe definir las jerarquías, que se componen de los niveles y las tablas de datos. Así para la dimensión Colaborador, se crean el nivel colaborador y se define la tabla DimColaboradores. Como se puede observar en el diagrama anterior se pueden definir varios niveles de jerarquía; en este caso se han definido para tiempo los niveles Año y Mes, o en la dimensión producto los niveles Categoría y Producto. Esto permitirá filtrar la información por año y mes o agrupar por categorías y productos.

Para definir los niveles se debe introducir los valores en su tabla de atributos indicando el nombre del nivel, la columna donde se encuentra el índice que relaciona con la tabla de hechos el nombre de la columna, el tipo de datos devueltos, si son únicos los datos, el tipo de nivel o si es o no un nivel visible. La siguiente imagen es un ejemplo de como se ha definido el nivel Colaboradores:

Level for 'Colaborador' Hierarchy	
Attribute	Value
name	Colaboradores
description	
table	
column	CollaboratorID
nameColumn	CollaboratorName
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	Regular
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Imagen 47: Detalle nivel colaborador.

Los niveles año y mes de la dimensión Tiempo se crean con un tipo TimeYears y TimeMonths respectivamente; mientras que el resto de los niveles definidos en este esquema, serán del tipo Regular.

Attribute	Value
name	Año
description	
table	
column	año
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	Integer
internalType	
uniqueMembers	<input checked="" type="checkbox"/>
levelType	TimeYears
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

Imagen 48: Detalle nivel Año.

Level for 'Tiempo' Hierarchy	
Attribute	Value
name	Mes
description	
table	
column	mes
nameColumn	mesText
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input type="checkbox"/>
levelType	TimeMonths
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	mesText
formatter	
visible	<input checked="" type="checkbox"/>

Imagen 49: Detalle nivel Mes.

Tras definir las dimensiones, se define el cubo, que en este caso se denomina Nutribox para lo cual, sobre el schema se pulse el botón derecho del ratón y se selecciona la opción add cube. Tras nombrarlo añadiendo el valor en su tabla de atributos, hay que indicar su tabla de hechos, en este caso HechosCanje para lo cual hay que pulsar sobre el cubo con el botón derecho y pulsar sobre “ad table”.

Además, se deberán indicar sus dimensiones para lo cual usaremos la opción “add dimensión usage” e introduciremos los atributos correspondientes:

Dimension Usage for 'Nutribox' Cube	
Attribute	Value
name	Usuario
foreignKey	UserID
source	Usuario
level	
usagePrefix	
caption	
visible	<input checked="" type="checkbox"/>

Imagen 50: Añadir uso de dimensión al cubo.

Por ejemplo, en el caso de la dimensión usuario (como se ve en la imagen anterior) indicaremos el nombre usuario, la foreignKey UserID, y source el nombre de la dimensión definida anteriormente, en este caso también “usuario”.

Una vez definidas las dimensiones del cubo, habrá que definir las medidas que se incluirán en el cubo, para ello sobre el cubo, tras hacer clic con el botón derecho del ratón, pulsaremos sobre la opción “Add measure” e introducir sus valores en su tabla de definición. Se puede ver en la siguiente imagen como se ha definido la medida “Número de colaboradores”:

Measure for 'Nutribox' Cube	
Attribute	Value
name	Número de colaboradores
description	
aggregator	distinct-count
column	CollaboratorID
formatString	
datatype	
formatter	
caption	
visible	<input checked="" type="checkbox"/>

Imagen 51: Añadir medida al cubo.

En esta tabla de definición se ha asignado un nombre en el atributo name, se indica el tipo de operación a realizar en el campo “aggregator”, que en este caso es distinct-count y se indica la columna sobre la que se aplica la operación, en este caso Collaborator ID.

Para nombrar el esquema, sobre el icono del schema se puede asignar un nombre, en este caso Olap Nutribox a través de su tabla de atributos:

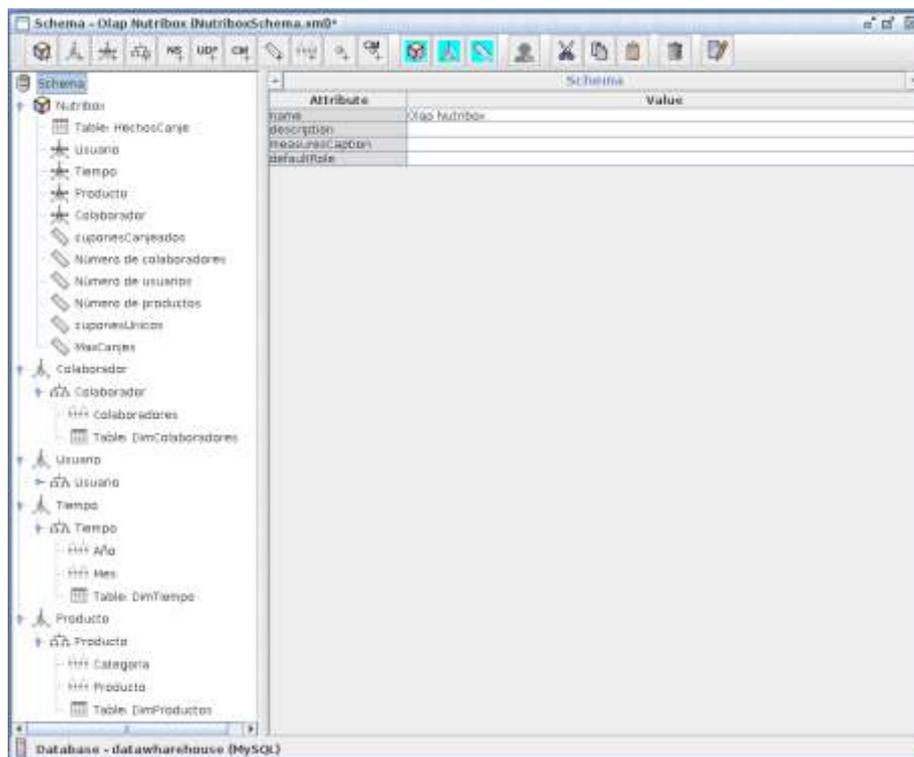
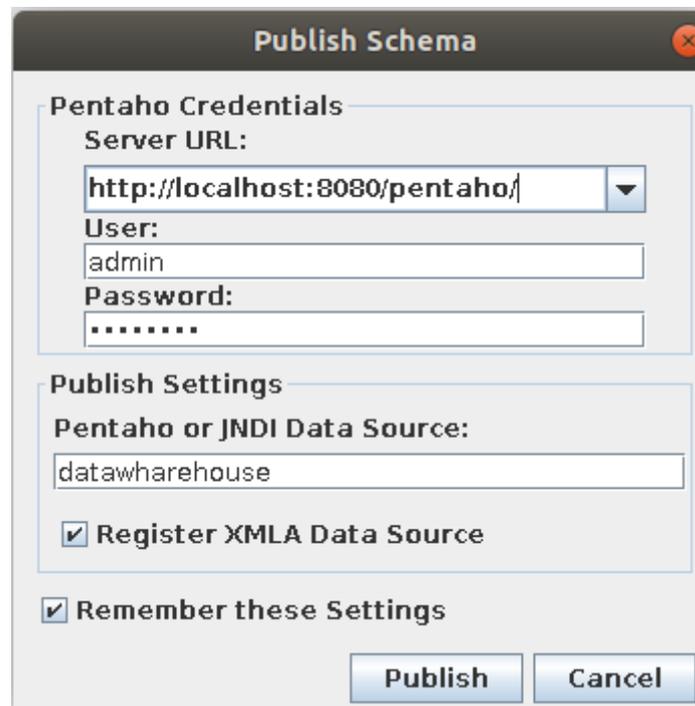


Imagen 52: Detalle del esquema final.

Una vez creado el esquema es necesario guardar el archivo en la opción “save as...”el menú file.

Para poder ser utilizado por el resto de las herramientas de la suite pentaho, este esquema hay que publicarlo, a través de la opción “Publish” del menú “file” e introduciendo la información del servidor pentaho en la ventana que aparece:



Publish Schema

Pentaho Credentials

Server URL:
http://localhost:8080/pentaho/

User:
admin

Password:

Publish Settings

Pentaho or JNDI Data Source:
datawarehouse

Register XMLA Data Source

Remember these Settings

Publish **Cancel**

Imagen 53: Publicar el esquema.

Una vez introducidos los datos del servidor hay que pulsar en el botón “Publish” y una vez realizado el proceso se desplegará una ventana indicando que se ha publicado con éxito el esquema.

8.5. Análisis de los datos del cubo OLAP con jpivot

En este anexo se ejemplifica un análisis OLAP con jpivot a partir del cubo Olap Nutribox.

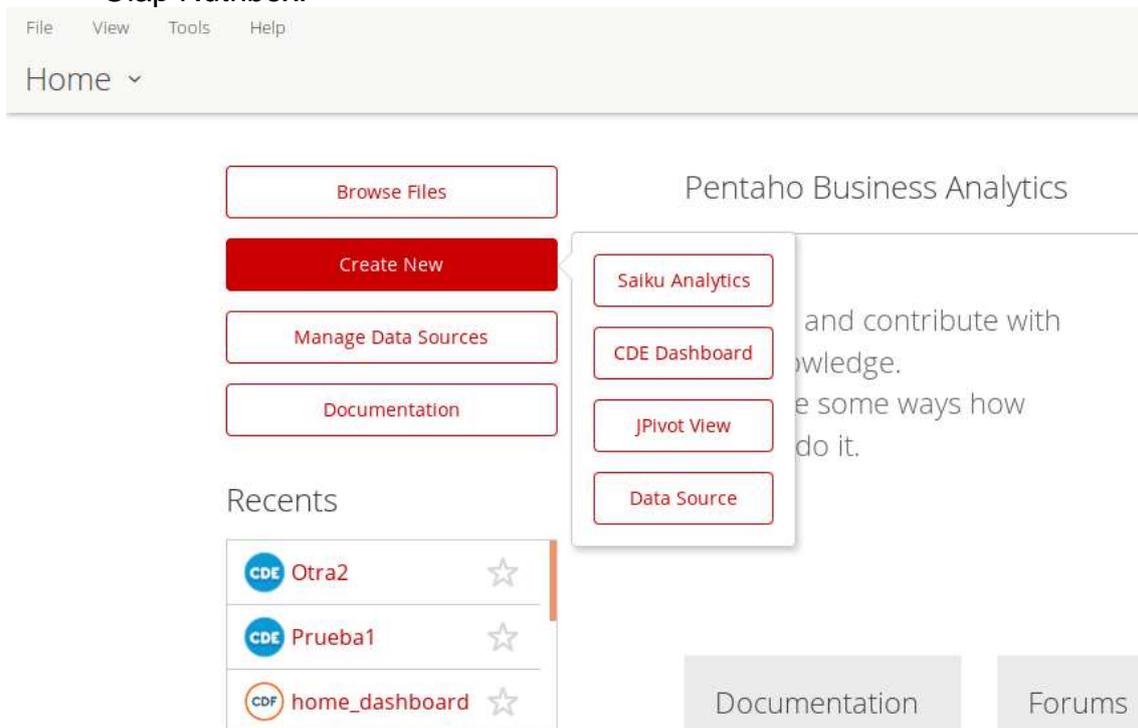


Imagen 54: Crear nuevo.

Tras validar en el servidor de pentaho business Analytics, entrando en <http://localhost:8080/pentaho> desde el navegador en el prototipo, al pulsar sobre el botón “Create New” y seleccionando JPivot View permitirá crear un análisis de los datos del cubo, para lo cual será necesario seleccionar el esquema y cubo entre los disponibles en el servidor:

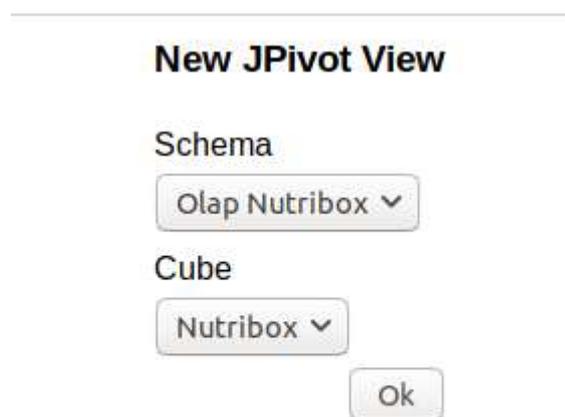


Imagen 55: Crear nueva vista JPivot.

Pulsando sobre el botón del cubo  se pueden definir las medidas y las jerarquías o agrupaciones que se quieren analizar:

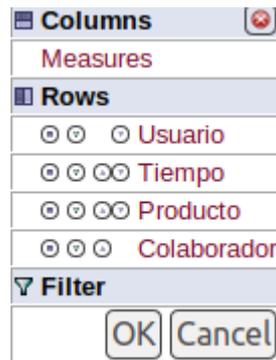


Imagen 56: Medidas y jerarquías.

Pulsando sobre el botón  se desplegará un editor de MDX donde podremos observar o editar el código MDX de la consulta. Por ejemplo, en nuestro prototipo, para obtener el número de cupones canjeados por empresa colaboradora bastará con introducir la siguiente consulta:

```
SELECT
NON EMPTY {[Measures].[cuponesCanjeados]} ON COLUMNS,
NON EMPTY {[Colaborador].[Colaboradores].Members} ON ROWS
FROM [Nutribox]
```

Mostrará los siguientes resultados:

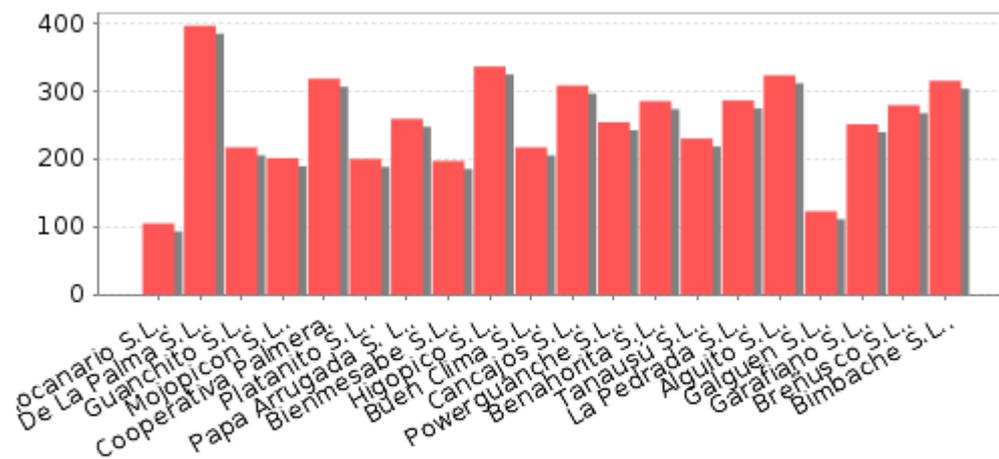
	Measures
Colaborador	☉ cuponesCanjeados
Platanocanario S.L.	105
Agricultores De La Palma S.L.	396
Guanchito S.L.	217
Mojopicon S.L.	201
Cooperativa Palmera	318
Platanito S.L.	200
Papa Arrugada S. L.	259
Bienmesabe S.L.	197
Higopico S.L.	336
Buen Clima S.L.	217
Cancajos S.L.	308
Powerguanche S.L.	254
Benahorita S.L.	285
Tanausú S.L.	230
La Pedrada S.L.	286
Alguito S.L.	323
Galguen S.L.	123
Garafiano S.L.	251
Breñusco S.L.	279

Imagen 57: Ejemplo Jpivot tabla.

Además, será posible obtener su gráfico al pulsar el botón :

Empresa	Cupones
Tanausu S.L.	230
La Pedrada S.L.	286
Alguito S.L.	323
Galguen S.L.	123
Garafiano S.L.	251
Breñusco S.L.	279
Bimbache S.L.	315

Slicer:



Slicer:

cuponesCanjeados.

JPIVOT IS A COMMUNITY

Imagen 58: Gráfico generado con Jpivot.

8.6. Análisis de los datos del cubo OLAP con saiku

En este anexo se ejemplifica un análisis OLAP con Saiku a partir del cubo Olap Nutribox.

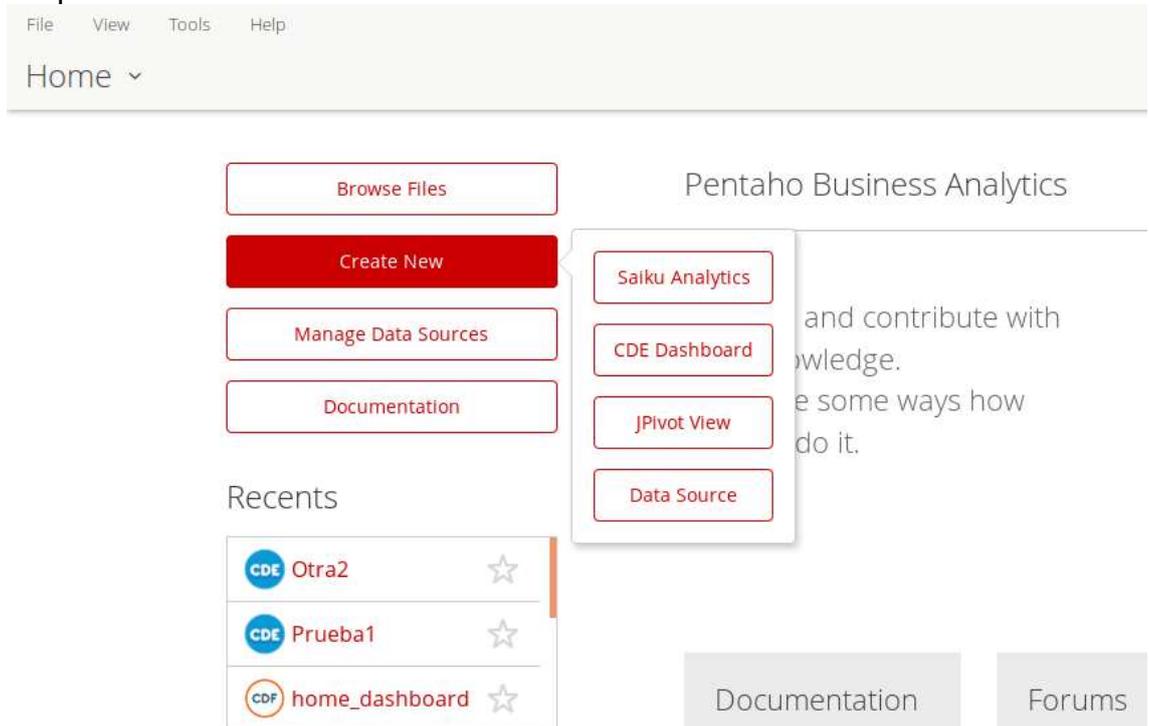


Imagen 59: Crear nuevo.

Tras validar en el servidor de pentaho business Analytics, entrando en <http://localhost:8080/pentaho> desde el navegador en el prototipo, al pulsar sobre el botón "Create New" y seleccionando Saiku Analytics permitirá crear un análisis de los datos del cubo, para lo cual será necesario seleccionar el esquema y cubo entre los disponibles en el servidor:

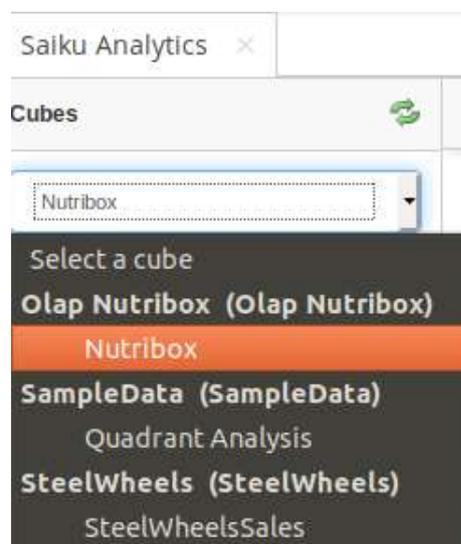


Imagen 60: Seleccionar cubo.

Una vez seleccionado el cubo sobre el que trabajar en la columna de la derecha aparecerán ordenadas todas las medidas y dimensiones del cubo.

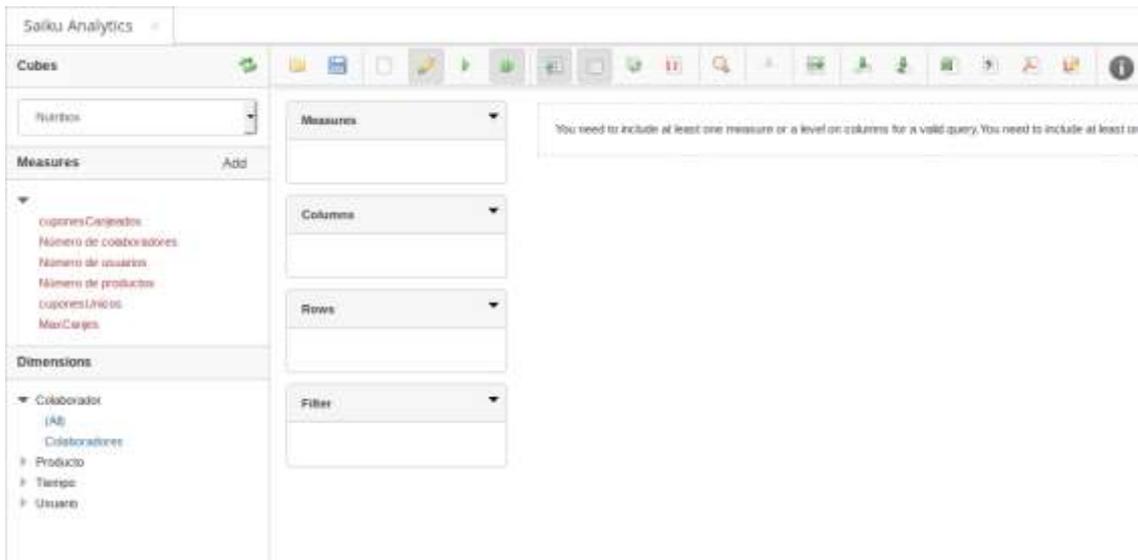


Imagen 61: Área de trabajo Saiku.

Arrastrando estos parámetros a las cajas Measures, Columns, Rows y Filter del área de trabajo, se generará una consulta que se mostrará los datos indicados. Por ejemplo, si queremos obtener el número de cupones canjeados por empresa colaboradora (ejemplo realizado anteriormente con jpivot) bastará con arrastrar cuponesCanjeados a la caja de Measures y Colaboradores a Rows:

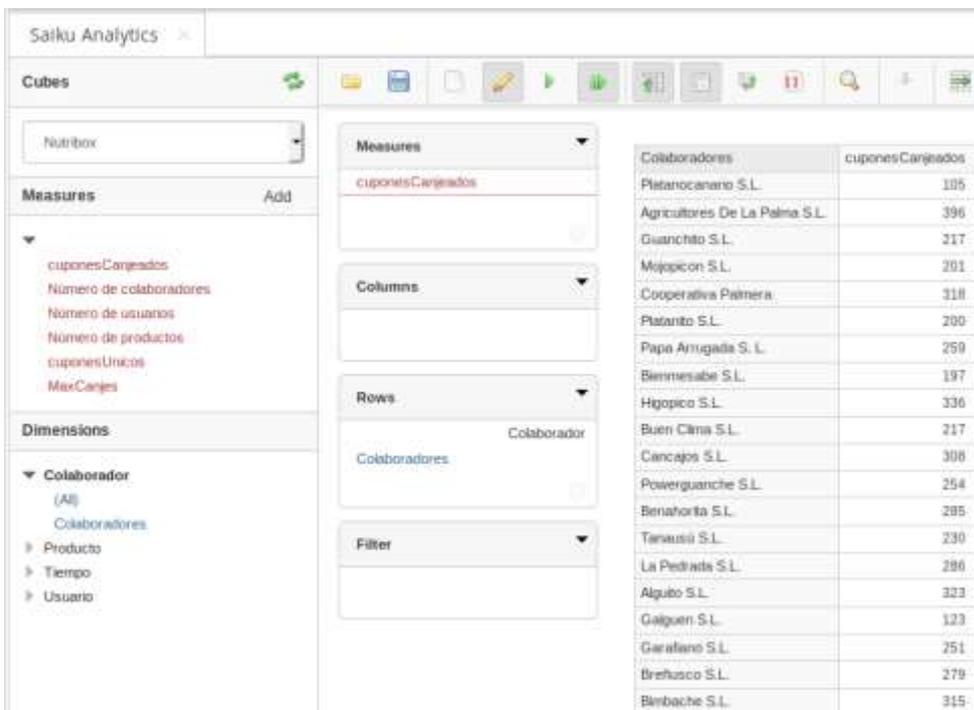


Imagen 62: Vista en tabla de consulta con Saiku.

Además, en la esquina superior izquierda, podemos alternar entre la



vista de tabla o la gráfica en los botones :

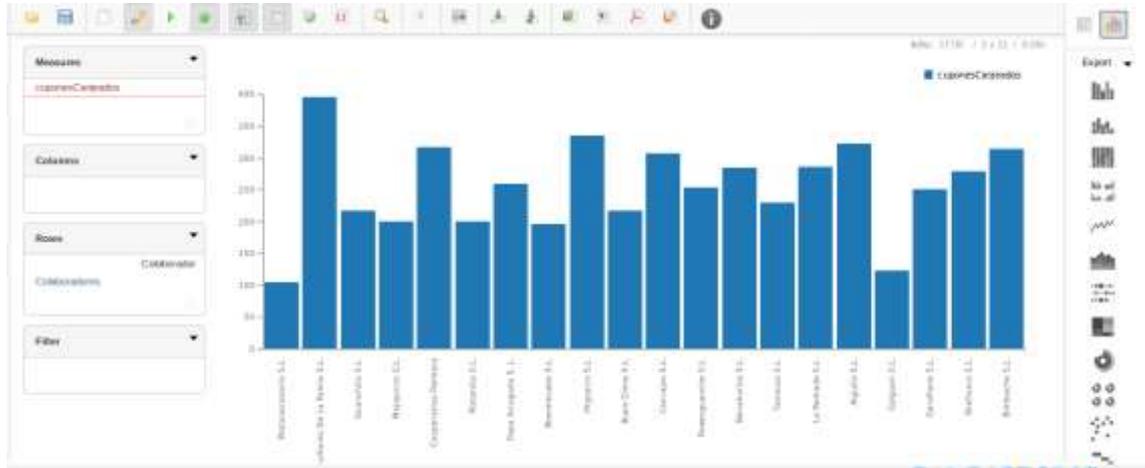


Imagen 63: Gráfica generada con Saiku.

8.7. Creación de cuadros de mando con Pentaho CDE Dashboard.

Tras acceder a Pentaho BI server se deberá pulsar sobre la opción Create New, CDE Dashboard:

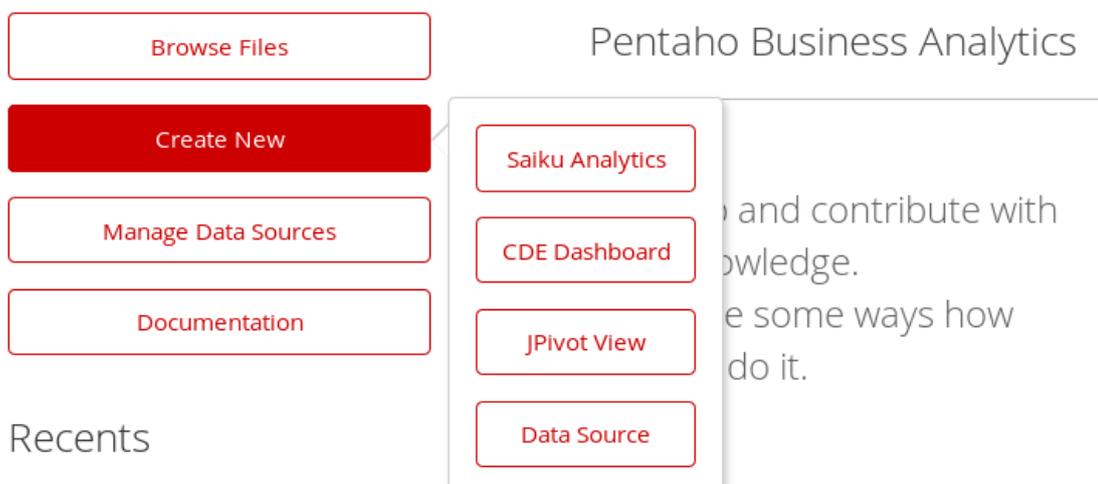


Imagen 64: Crear nuevo CDE Dashboard.

Una vez dentro del área de trabajo hay que configurar la distribución del cuadro de mando. Esta distribución se debe definir en forma de tabla, añadiendo filas y columnas desde el "Layout Panel".

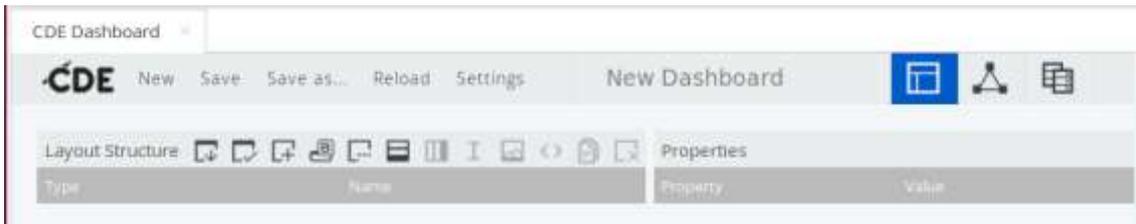


Imagen 65: CDE Dashboard, detalle "Layout Panel"

Desde este panel es posible añadir código html o javascript y se define el nombre del área que se usará como referencia para ubicar los componentes disponibles.

Desde el panel "Datasources Panel" se deben indicar las distintas fuentes de datos, que se usarán en los distintos componentes del cuadro de mando:

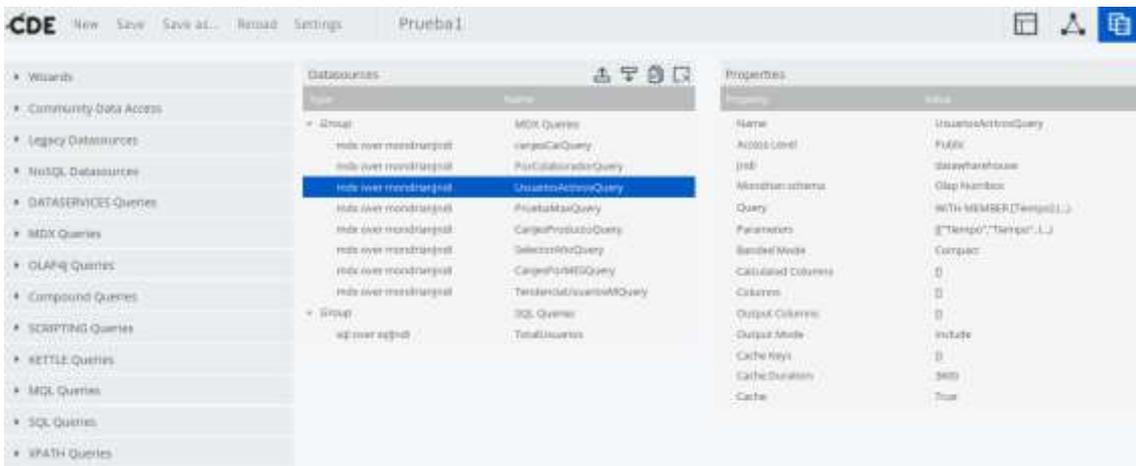


Imagen 66: CDE Dashboard, detalle " Datasources Panel "

Una vez definidos los orígenes de los datos a representar en el cuadro de mandos (en el caso de este trabajo, diferentes consultas MDX al cubo OLAP Nutribox y SQL a la base de datos datawarehouse), habrá que definir los componentes desde el "Component Panel":

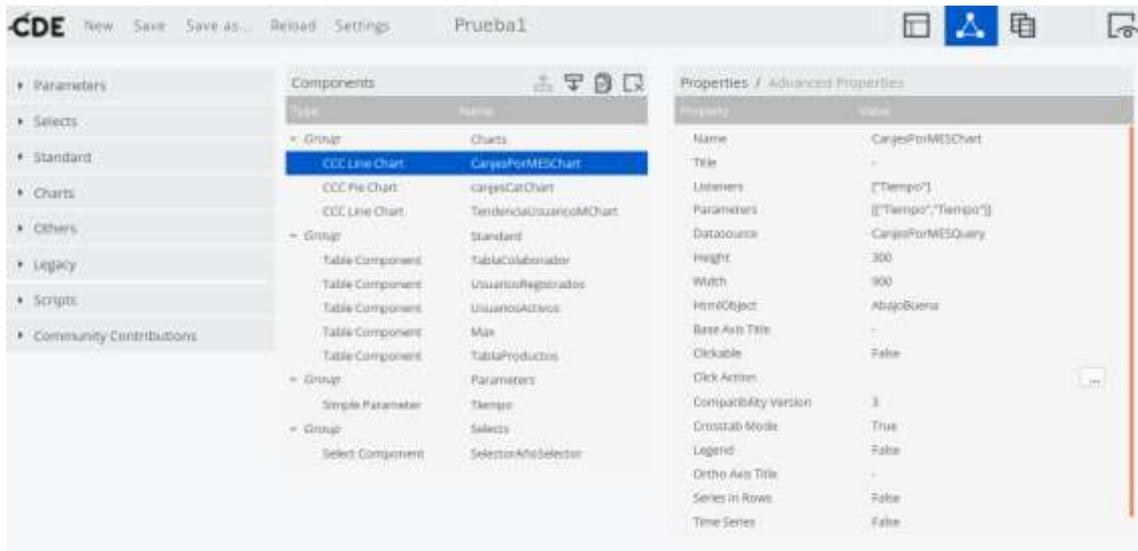


Imagen 67: CDE Dashboard, detalle "Component Panel"

Una vez definidos los diferentes componentes, se ubican en las distintas áreas generadas en el "Layout Panel" con la propiedad HtmlObject.

Para finalizar es necesario guardar el trabajo en el servidor desde la opción Save as:

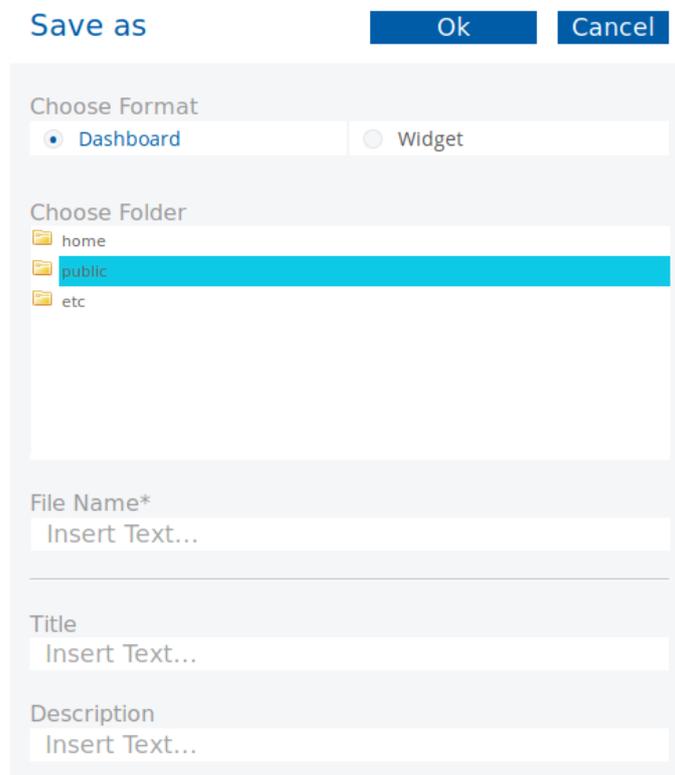


Imagen 68: CDE Dashboard, guardar cuadro de mando.

En esta nueva ventana indicaremos la ruta donde se guardará el archivo, su nombre, el título y una descripción.

8.8. Creación de informes con Pentaho Report Designer

En este anexo se describe la creación de un reporte usando la herramienta Pentaho Report Designer.

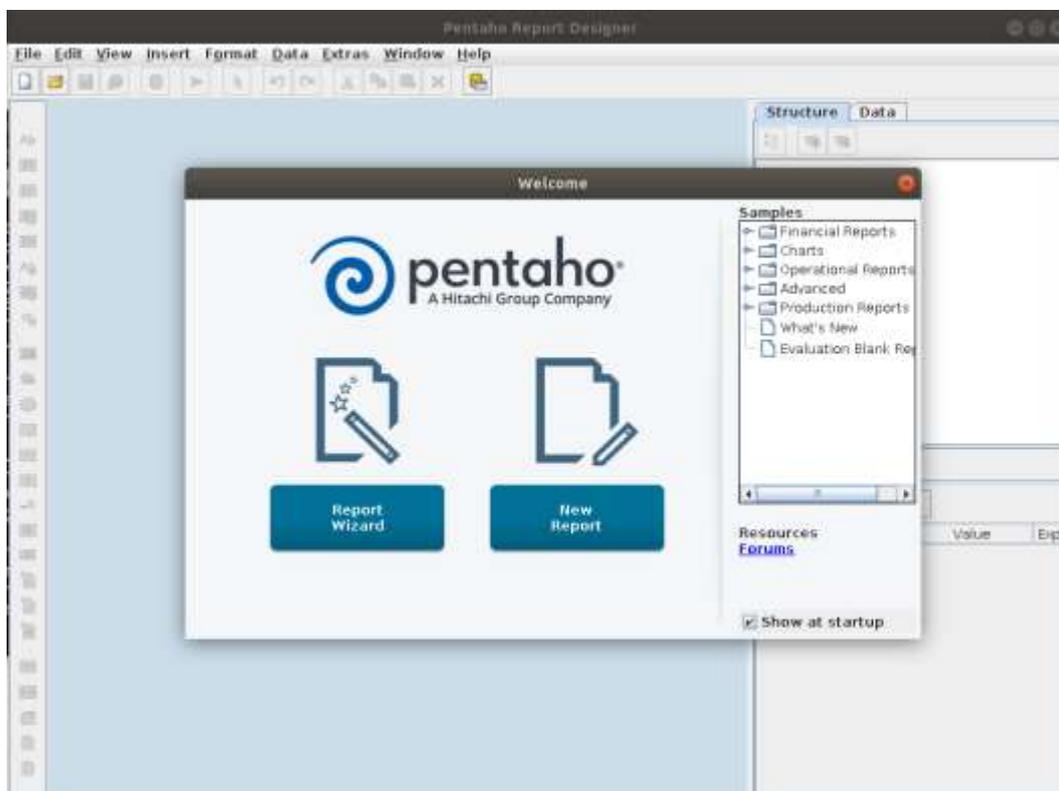


Imagen 69: Nuevo reporte.

Al ejecutar el programa, se muestra una pantalla de bienvenida que nos permitirá crear un nuevo archivo de reporting utilizando un asistente (Report Wizard) o desde cero (New Report), así como ver los archivos de ejemplo que contiene la aplicación y que se encuentran en la columna “Samples de la derecha”:

En este caso se partirá del asistente, por lo que será necesario pulsar la opción “Report Wizard” o bien desde la barra de menú ir a file, Report Design Wizard ...:

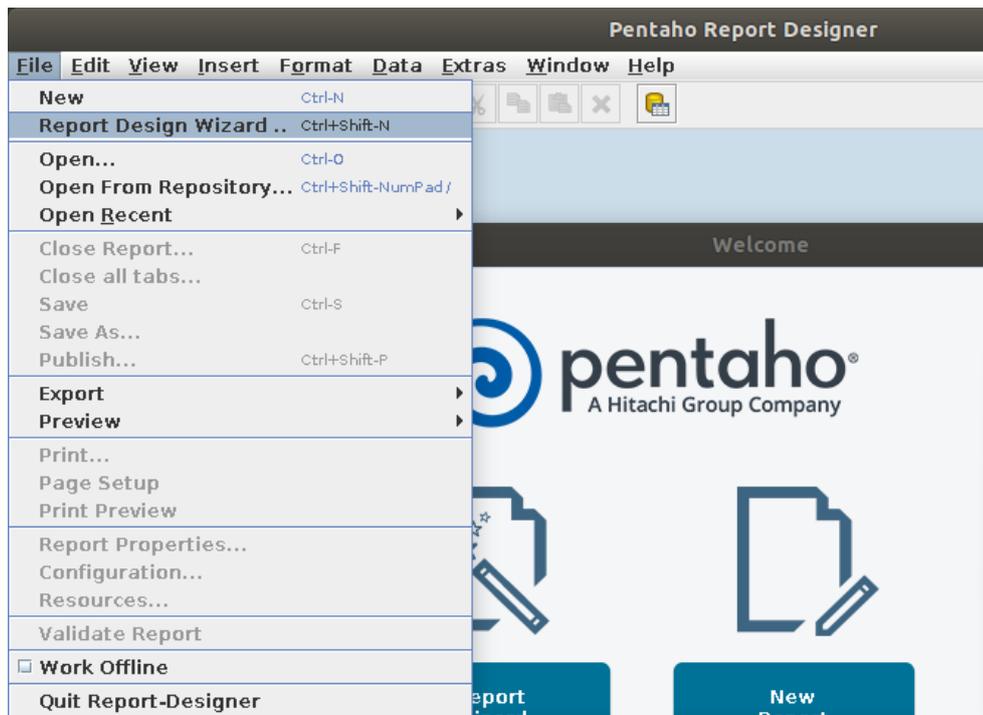


Imagen 70: Crear nuevo reporte con el asistente.

El primer paso de este asistente será el denominado Look and Feel donde asignaremos una apariencia, a partir de una plantilla predefinida al seleccionar la opción Template o bien de un documento existente al elegir la opción Report Document. En este ejemplo se ha utilizado la apariencia predefinida Jade:

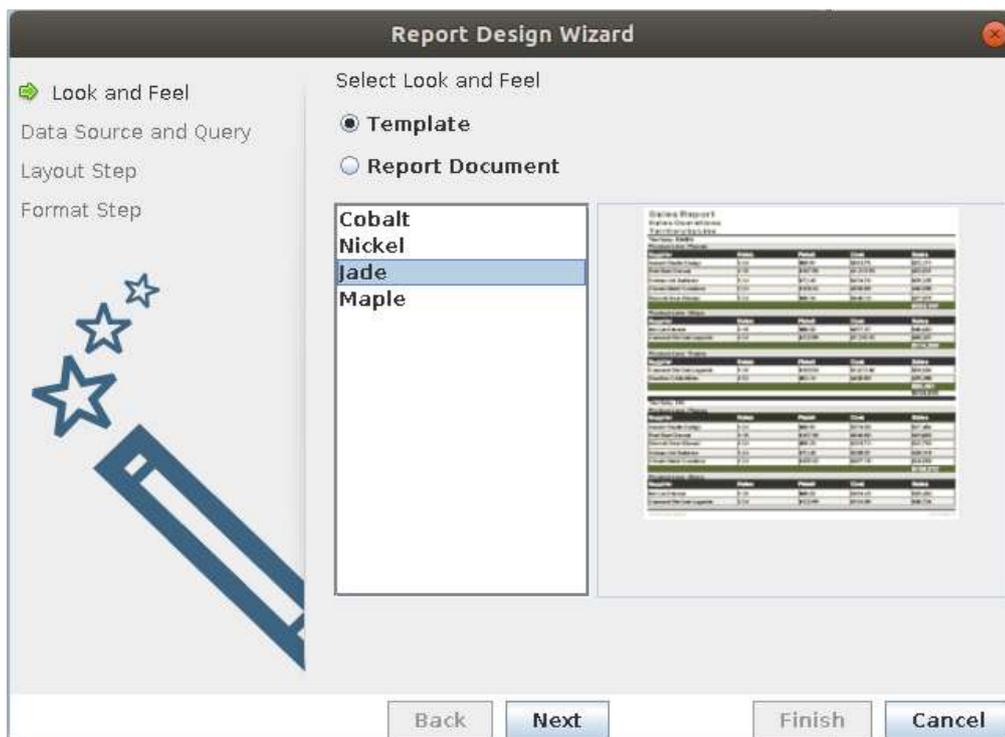


Imagen 71: Seleccionar apariencia.

Tras pulsar en el botón Next se mostrará el paso siguiente donde se deberá indicar las fuentes de datos a partir de los cuales se generará el informe. Por defecto, se encuentran asignadas unas fuentes de datos de pruebas como se ve en la siguiente imagen:

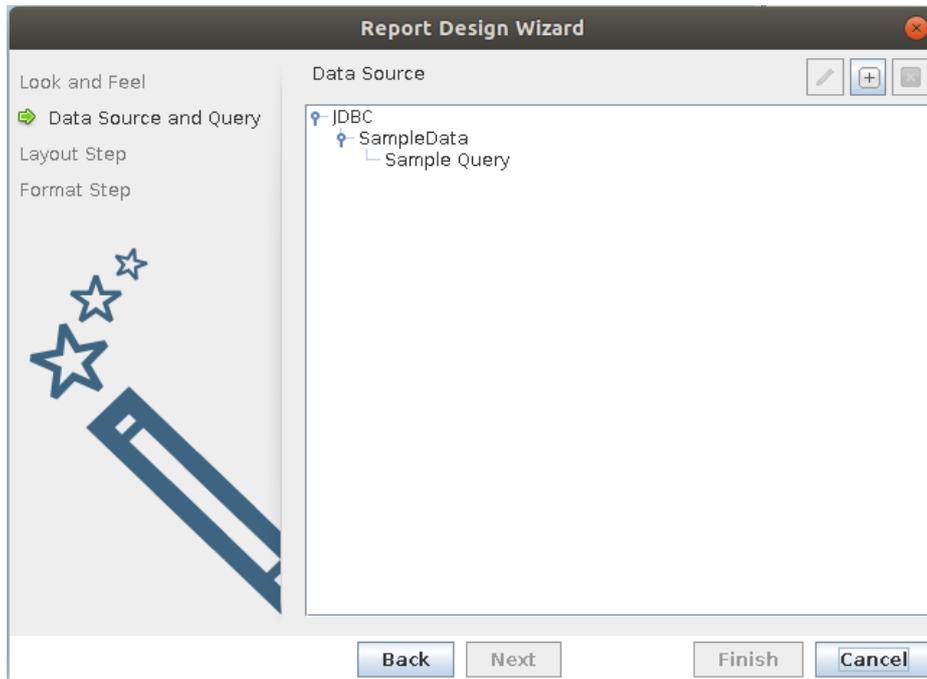


Imagen 72: Fuentes de datos y consultas.

Se deberán eliminar estas fuentes de datos para lo cual hay que pulsar sobre su nombre y posteriormente sobre el botón  en la esquina superior derecha. Tras este paso, será necesario añadir las nuevas fuentes de datos, pulsando sobre el botón .

En la siguiente ventana será necesario escoger el tipo de origen de datos. En este caso se utilizará el cubo OLAP que ha sido previamente definido a través de Pentaho schema workbench y publicado en el servidor Pentaho Business Analytics, seleccionando la opción "Pentaho Analysis":

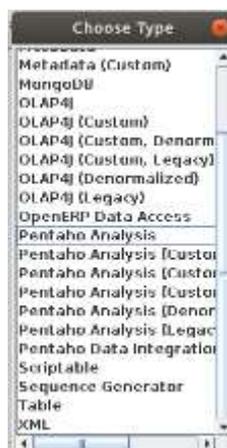


Imagen 73: Tipo de conexión.

En la siguiente imagen se deberá indicar la ruta del esquema NutriboxSchema.xml a través de la opción “Pentaho Analysis Schema File:” y seleccionar la conexión a la base de datos “datawarehouse” en el apartado “Connections:” cómo se puede ver en la siguiente imagen:

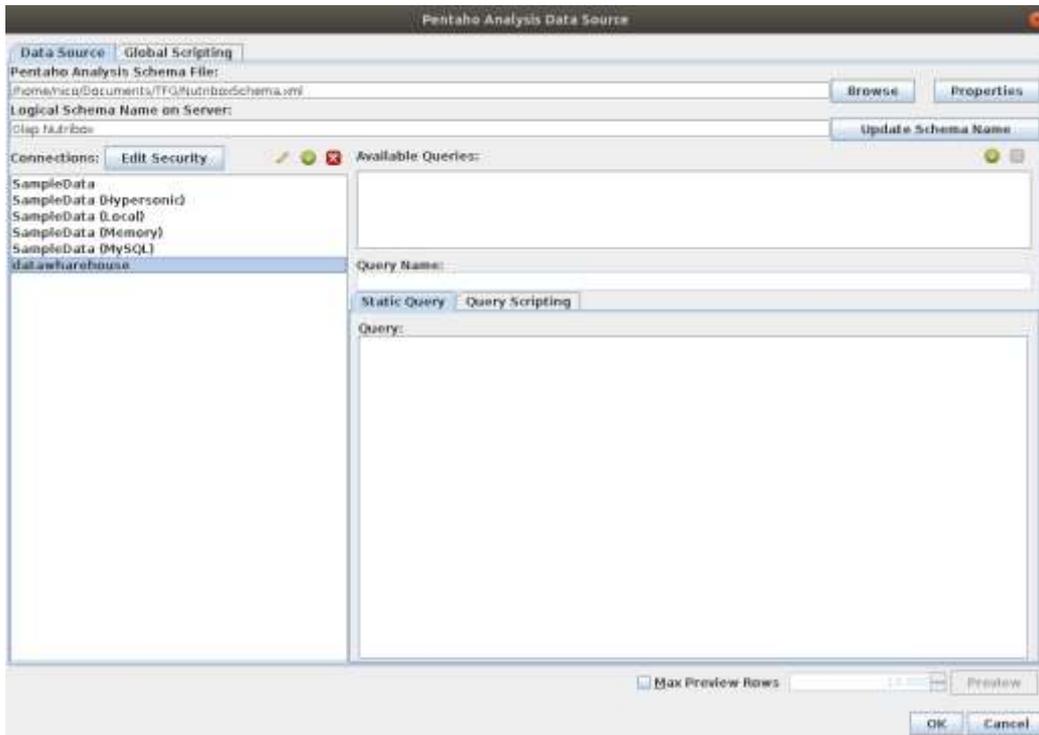


Imagen 74: Pentaho Analysis Data Source.

Sobre esta ventana, es necesario indicar las consultas para obtener los datos para lo cual será necesario pulsar sobre el botón  y modificar su nombre en “Query Name:”. En este caso se le ha llamado porCat.

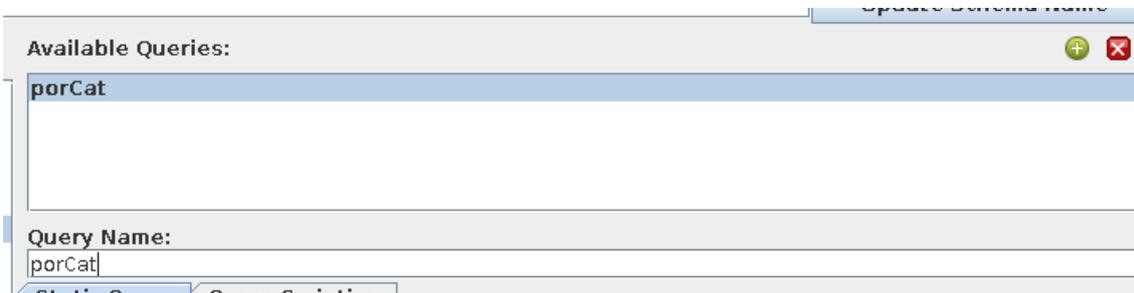


Imagen 75: Consultas.

En el espacio asignado introduciremos la consulta necesaria para obtener los datos, como puede verse en la siguiente ventana:

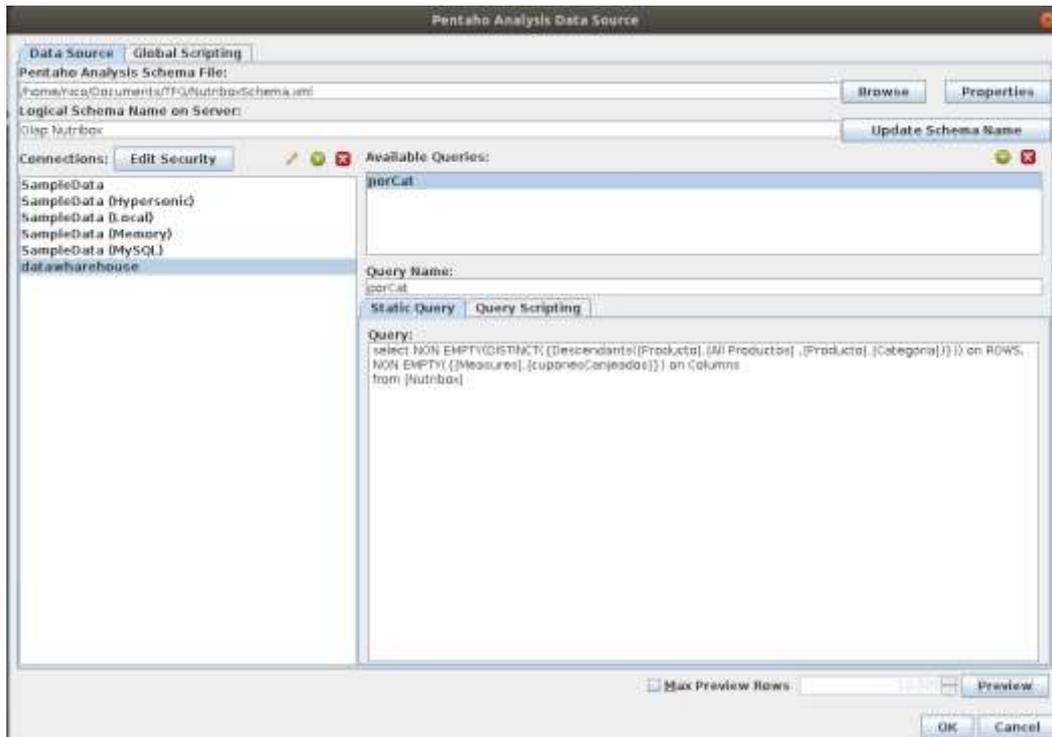


Imagen 76: Detalle final de la conexión.

Pulsando sobre el botón “Preview”, en la parte inferior derecha de la ventana podemos tener una vista previa de los datos que nos devuelve la consulta, que en este ejemplo se trata del número de canjes realizados por categoría de productos:

[Producto].[All]	[Producto].[Categoria]	[Measures].[cuponesCanjeados]
All Productos	Fruta Eco	348
All Productos	Harinas	116
All Productos	Salsas	118
All Productos	Infusiones	442
All Productos	Verduras Eco	221
All Productos	Bebidas Alcoholicas	244
All Productos	Cosméticos	259
All Productos	Carnes	244
All Productos	Pescadería	405

Imagen 77: Previsualización de la consulta

Una vez introducidas las consultas necesarias (en este caso inicialmente sólo es necesaria esta) serán visibles en el “Data Source” del asistente.

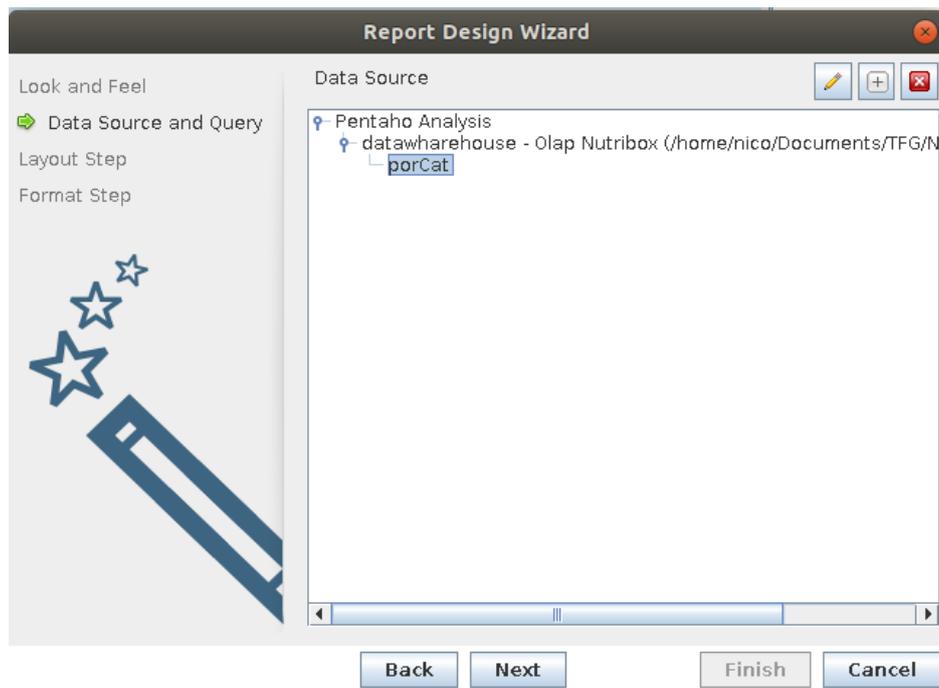


Imagen 78: Conexión creada en el asistente.

Tras pulsar en Next agregaremos los campos a mostrar en la tabla del informe. En este caso las categorías ([Producto].[Categoria]) y las cantidades ([Measures].[cuponesCanjeados]), sin ningún tipo de agrupación:

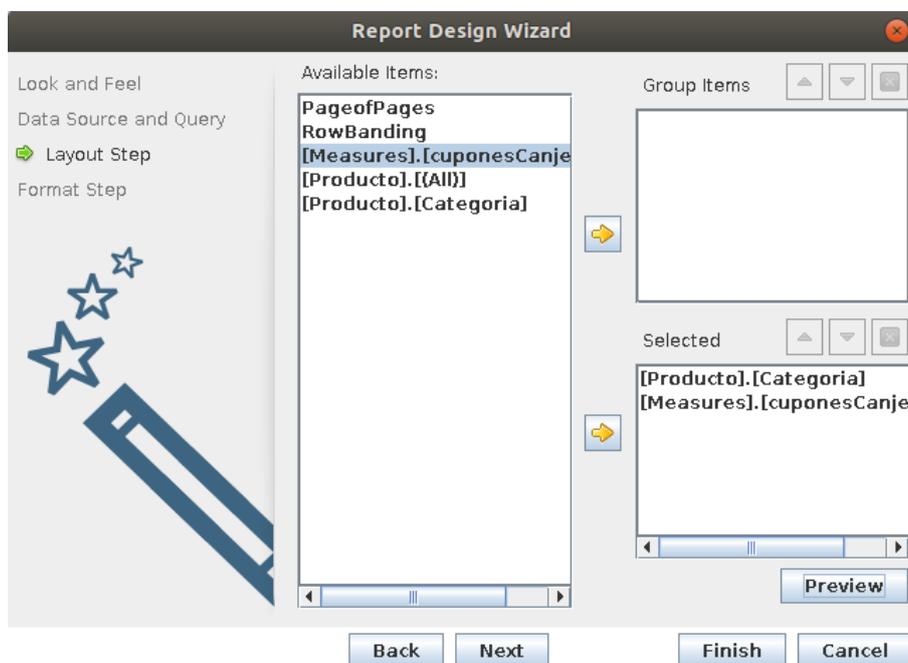


Imagen 79: Selector de campos.

Pulsando en “Preview” podemos obtener una vista previa del informe:

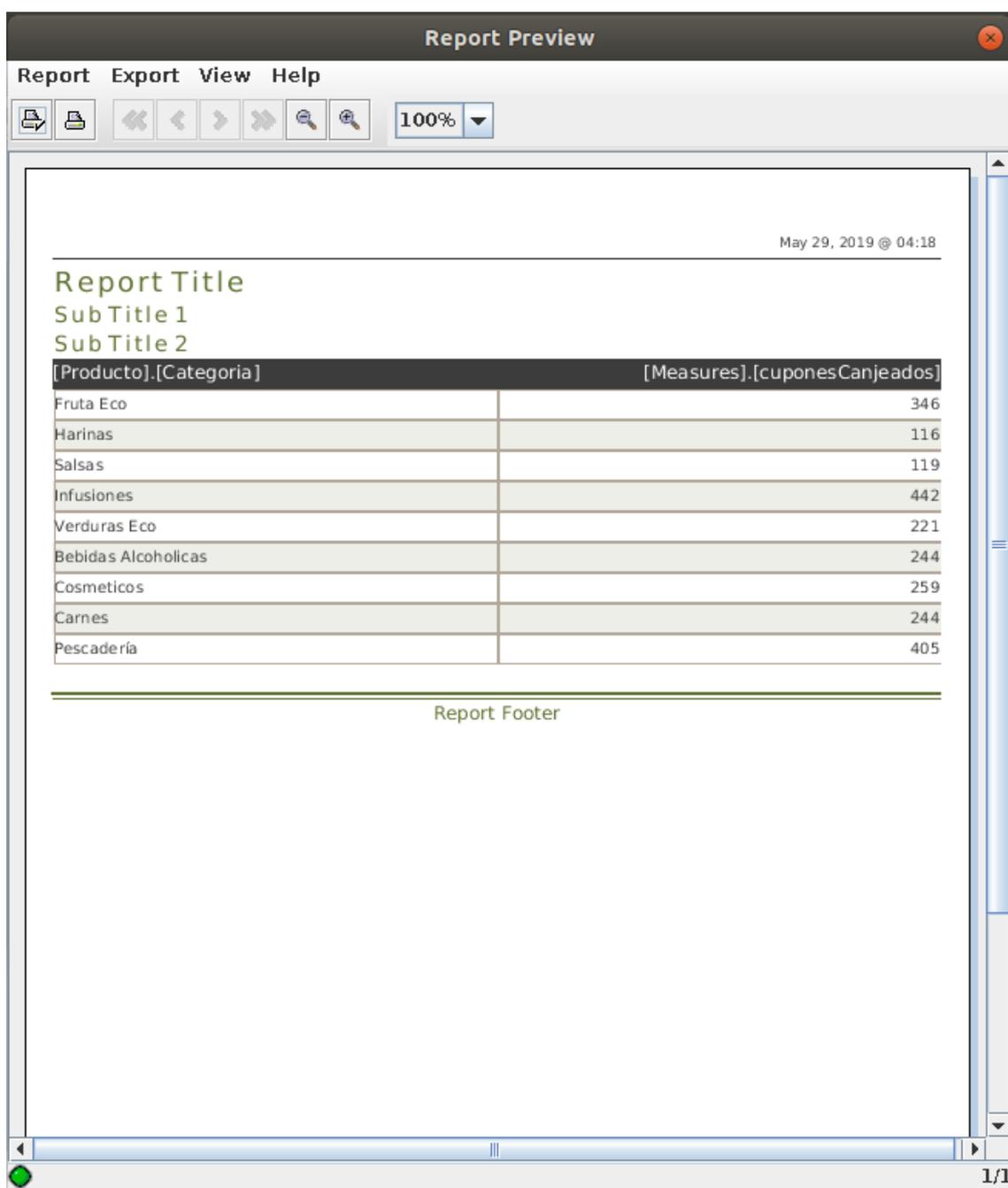


Imagen 80: Previsualización del informe.

En el siguiente paso podemos indicar el formato, en este caso damos al nombre de categoría un 80% del espacio y alineación centrada para las cantidades:

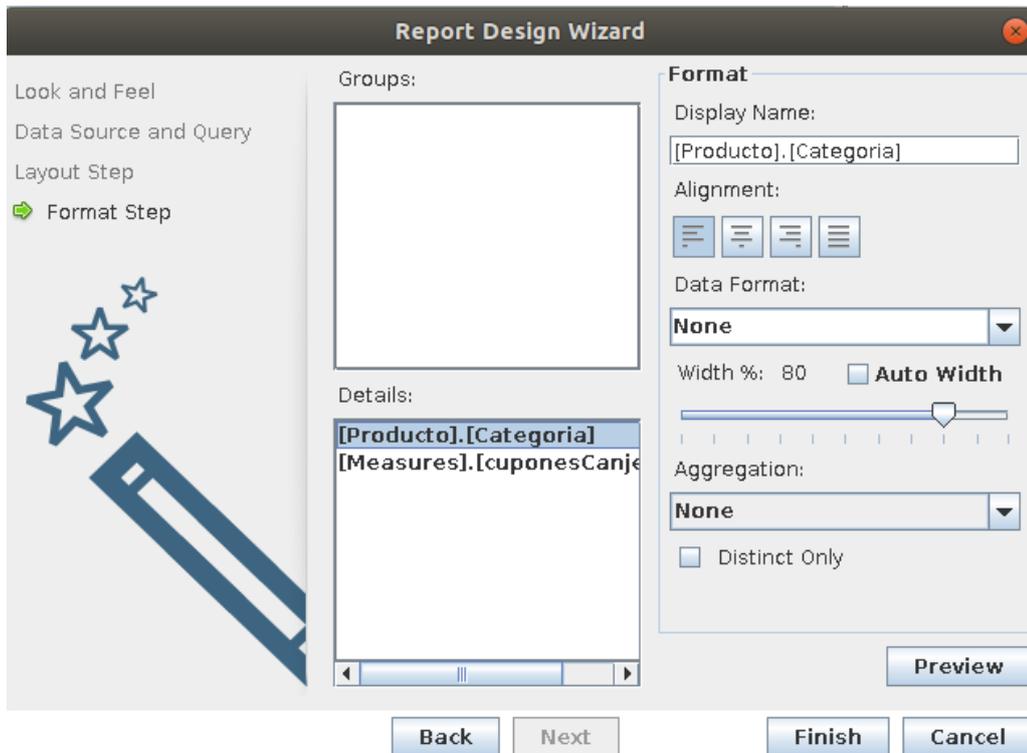


Imagen 81: Ajuste de formato del informe.

Tras pulsar en "Preview" se muestra el siguiente resultado:

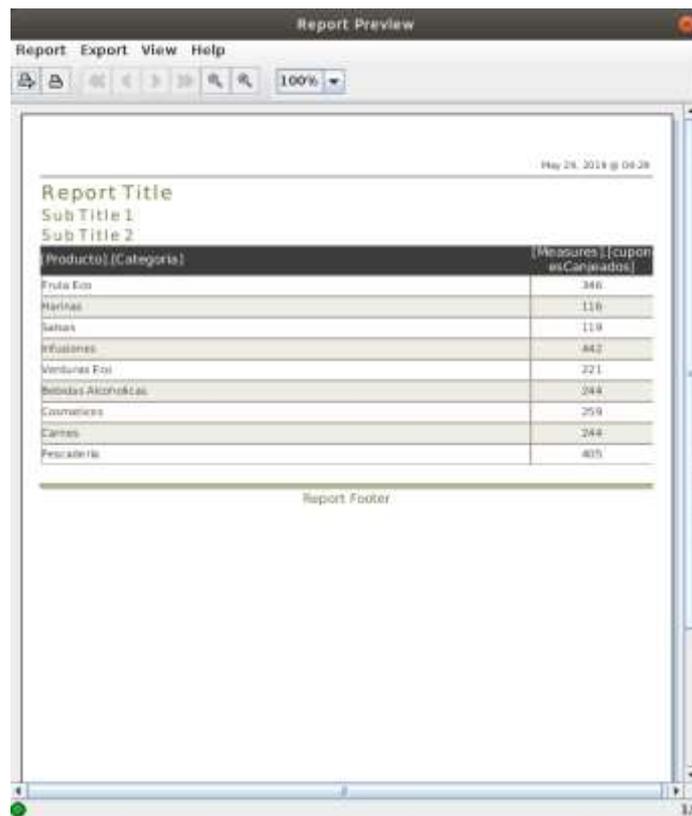


Imagen 82: Segunda previsualización del informe.

Tras cerrar la previsualización y pulsar en “Finish” en el asistente se cargarán estos datos en el área de trabajo de la aplicación:

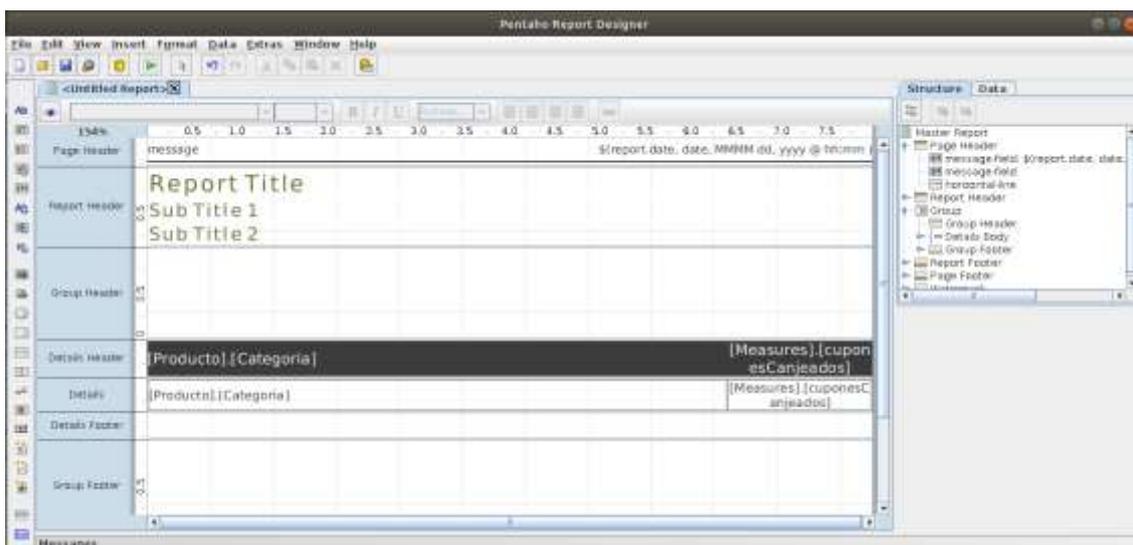


Imagen 83: Edición desde el área del trabajo.

Se edita el título del documento, así como las cabeceras de las tablas y podemos visualizar una previsualización pulsando sobre el botón  y luego la opción Print Preview:

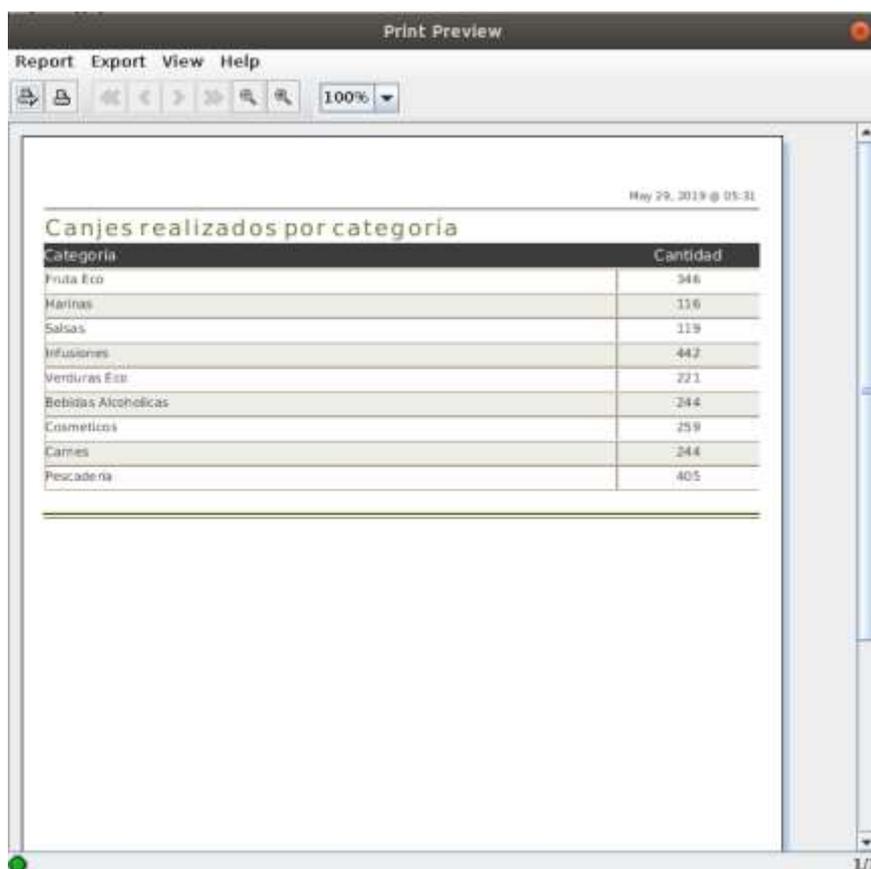


Imagen 84: Tercera Previsualización.

Para una mayor comprensión de los datos se añadirá un gráfico de tipo pastel, para lo cual se selecciona la posición donde irá ubicado (en este caso Group Header) y se pulsará sobre el botón . Se incluirá un gráfico en esta posición que se podrá configurar haciendo doble clic sobre él:

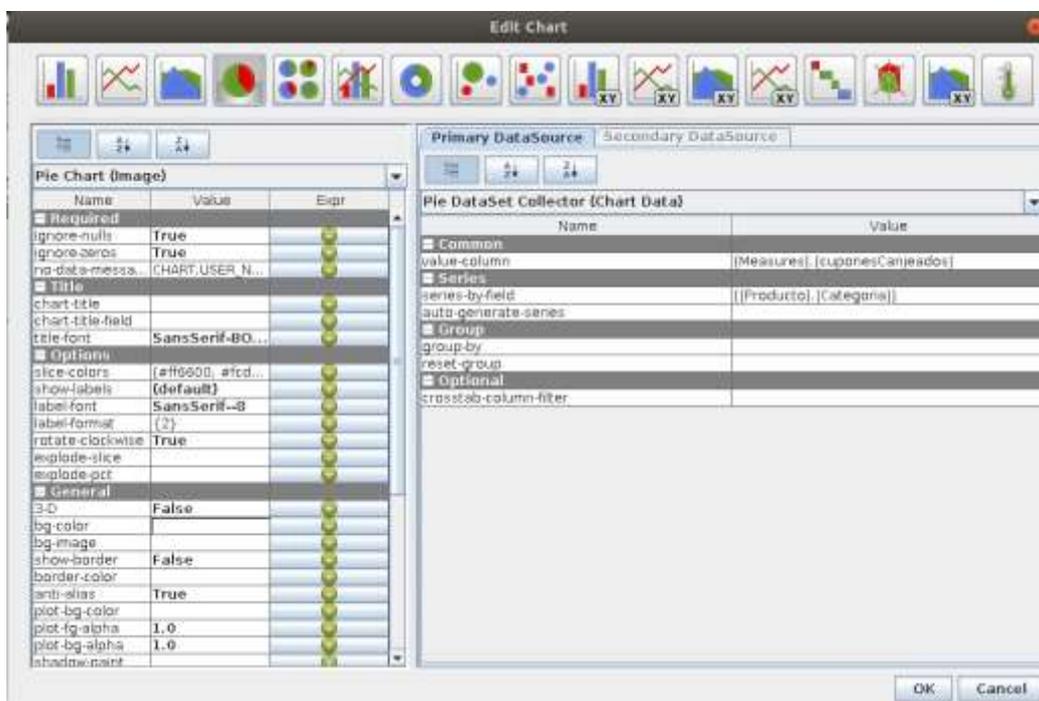


Imagen 85: Insertar gráfico.

Se obtiene el siguiente resultado:

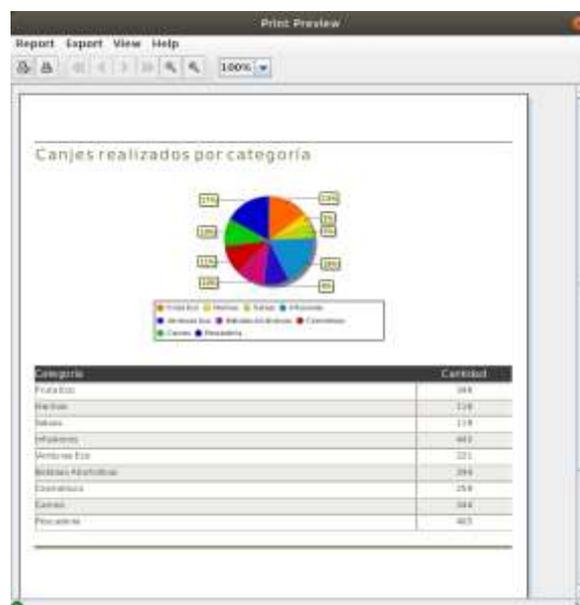


Imagen 86: Otra previsualización.

Para que este informe se genere para cada año y poder analizar cada ejercicio individualmente es necesario crear un parámetro, que se basará en los diferentes años con canjes registrados en nuestro cubo, por lo que será necesario añadir otra query a las fuentes de datos:

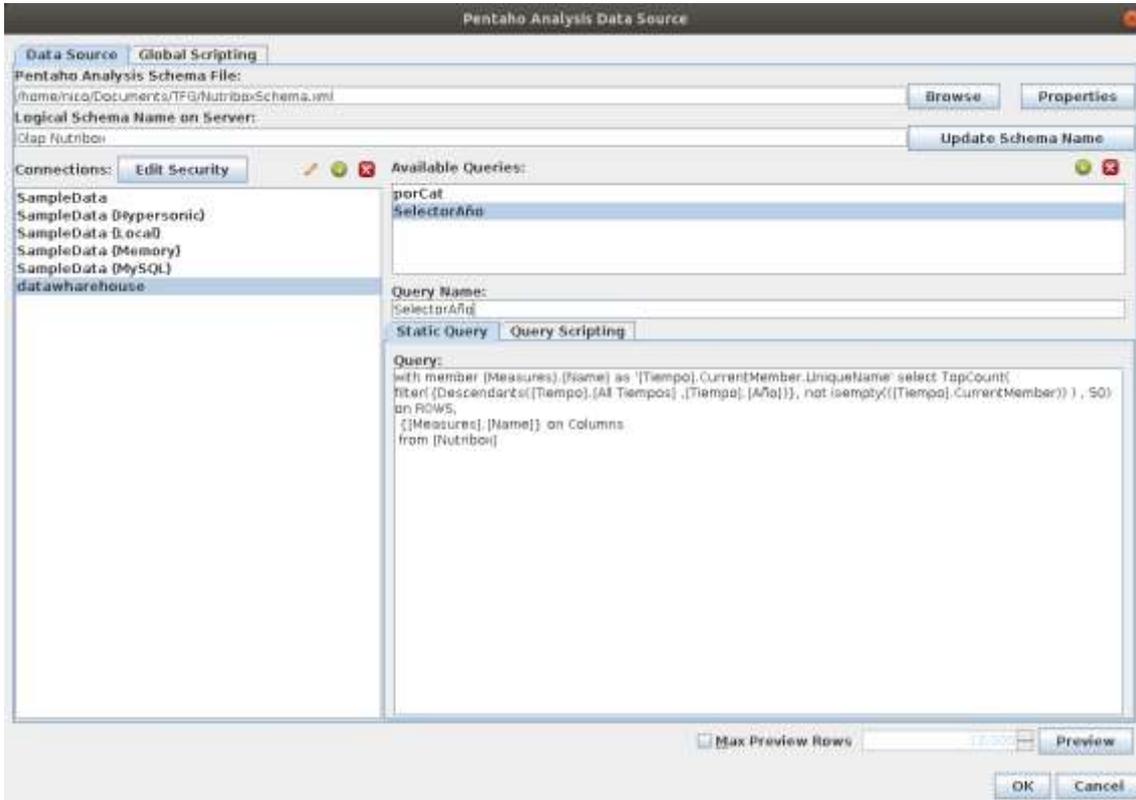


Imagen 87: Añadir consulta.

Posteriormente, para crear el parámetro vamos a la pestaña data al lado derecho del área de trabajo, pulsar en “Add parameter...”:

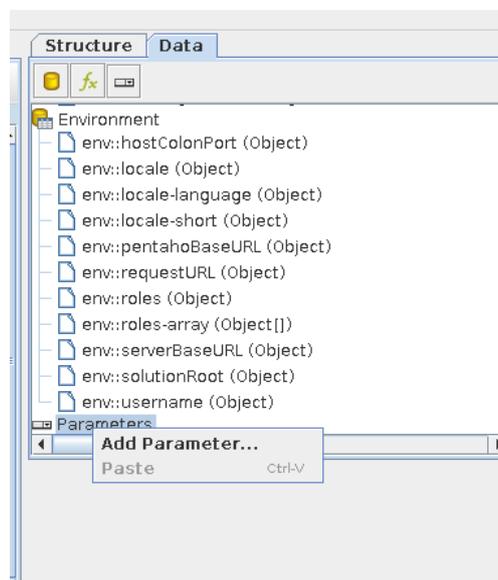


Imagen 88: Añadir parámetro.

Introduciendo los siguientes valores:

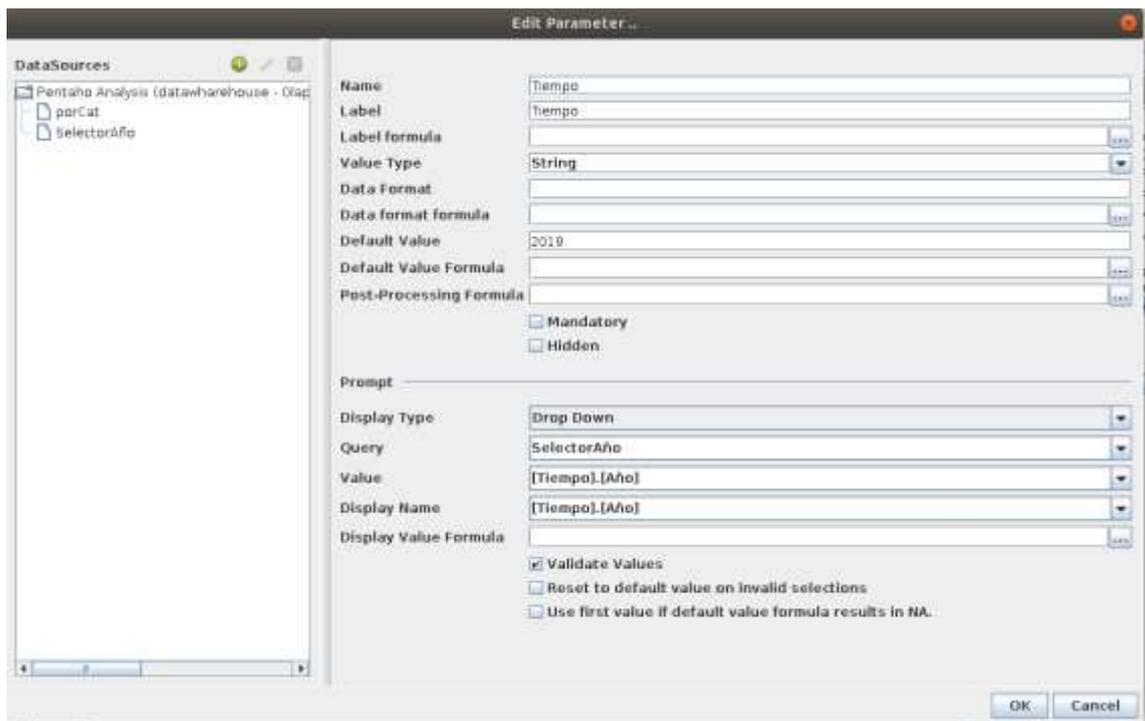


Imagen 89: Editar parámetro.

Finalmente hay que cambiar la consulta porCat que servía como fuente de datos para que sólo muestre los datos del año seleccionado:

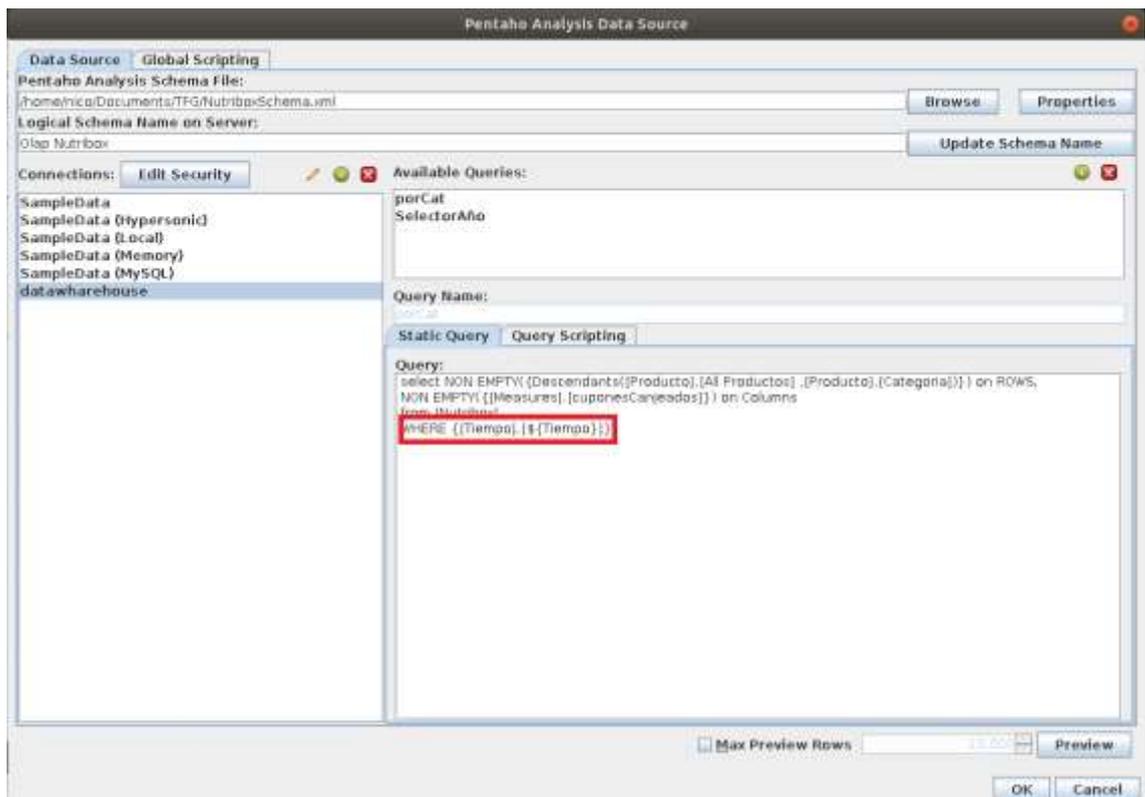


Imagen 90: Modificar consulta anterior.

Se modifica la cabecera para que indique el año a valorar a partir del parámetro creado:



Imagen 91: Se ha modificado la cabecera.

Así se incluirá un selector y mostrará los datos del año indicado:

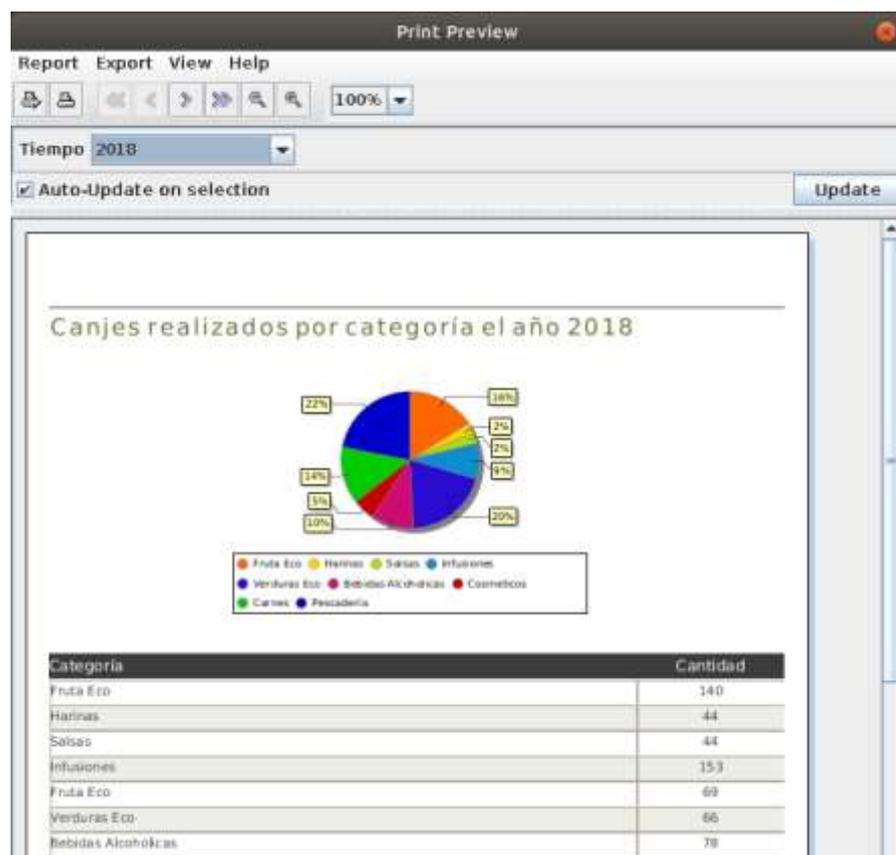


Imagen 92: Informe final Vista del año 2018.



Imagen 93: Informe final Vista del año 2019.

Finalmente, sólo queda guardarlo, desde el menú “file”, “save as...” o publicarlo en el servidor BI de pentaho desde la opción “Publish...”:

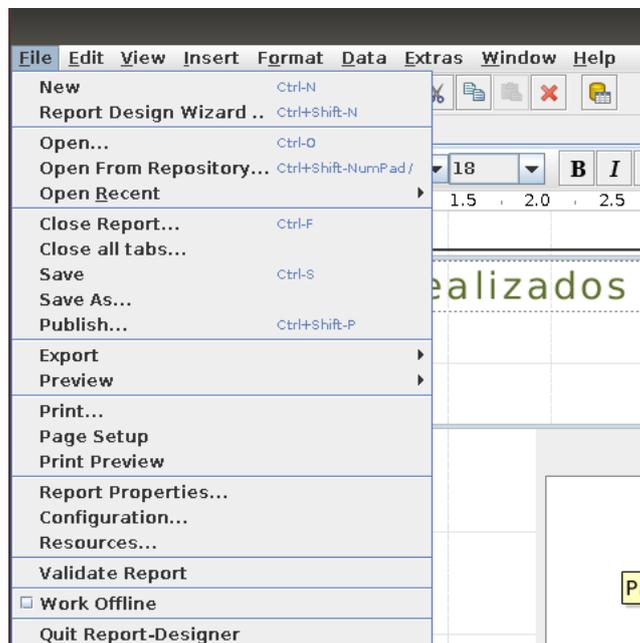


Imagen 94: Publicar el informe.

En este caso se publicará indicando los siguientes datos de conexión:

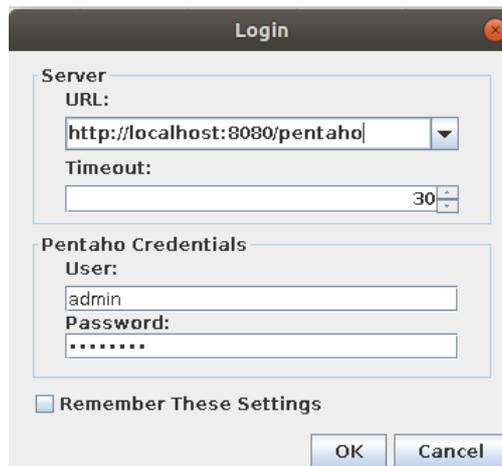


Imagen 95: Detalles de la conexión.

Tras esto el servidor solicitará los datos de publicación:

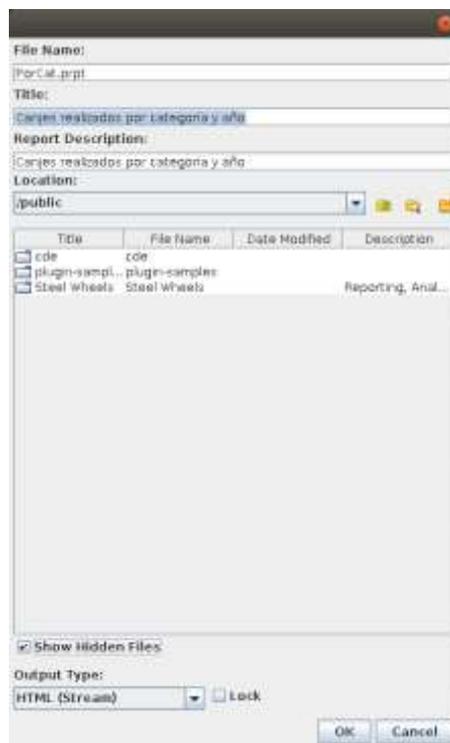


Imagen 96: Datos de la publicación.

Y una vez publicado nos mostrará su confirmación:



Imagen 97: Confirmación reporte publicado.