

ANÁLISIS DE LAS PRINCIPALES HERRAMIENTAS DE BUSINESS INTELLIGENCE EN ENTORNOS DE BIG DATA

por

David Moisés Blanco Quevedo

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un
i Universidad
Internacional
de Andalucía
A

November 2017

ANALYSIS OF THE MAIN TOOLS OF BUSINESS INTELLIGENCE AND ITS APPLICATION IN BIG DATA ENVIRONMENTS

David Moisés Blanco Quevedo

Máster en Economía, Finanzas y Computación

Antonio Peregrín Rubio

Universidad de Huelva y Universidad Internacional de Andalucía

2017

Abstract

The purpose of this study is to analyze the main tools existing in the market for the realization of Business Intelligence processes. We start from scratch, reflecting on the current situation in the world of data and information, on how they have arrived here and where they are heading. In this process, we will check the organizations' needs in terms of analysis, and how these needs can be met with current technology. We will demonstrate how comprehensive the existing solutions are, by reviewing their capabilities and shortcomings, emphasizing on a key aspect, their capabilities for managing large volumes of data, with which they can operate in a context of increasing and frenetic growth of the data available in the world.

Key words: Business Intelligence, data, information, technology.

Resumen

El presente estudio tiene por objeto analizar las principales herramientas existentes en el mercado para la realización de procesos de Inteligencia de Negocio. Partimos desde cero, reflexionando sobre la situación actual del mundo de los datos y la información, sobre cómo se ha llegado hasta aquí y hacia dónde se dirigen. En este proceso, comprobaremos las necesidades de las organizaciones en cuanto a análisis se refiere, y la forma en que estas necesidades pueden ser

cubiertas con la tecnología actual. Estudiaremos cuán completas son las soluciones existentes, mediante una revisión de sus capacidades y deficiencias, enfatizando en un aspecto clave, sus capacidades para el manejo de grandes volúmenes de datos, con la que puedan desenvolverse en un contexto de creciente y frenético crecimiento de los datos disponibles.

Agradecimientos

Quiero expresar mi agradecimiento a mi tutor, el profesor Antonio Peregrín, pues sin su ayuda y consejo no habría sido posible realizar este estudio. A la ciudad de Huelva, que me acogió con cariño y ahora considero mi segundo hogar, y a mis compañeros Pepe e Israel.

Tabla de Contenidos

1	Introducción	1
2	Aproximación al Business Intelligence	5
2.1	Marco teórico	5
2.2	Elección de las soluciones estudiadas	7
3	Comparativa de herramientas	11
3.1	Tableau	11
3.1.1	Introducción y versiones	11
3.1.2	Arquitectura	17
3.1.3	Aspectos técnicos	18
3.1.4	Funcionalidades de usuario	20
3.1.5	Seguridad, volumetría y Big Data	31
3.2	Qlik	33
3.2.1	Introducción y versiones	33
3.2.2	Arquitectura	37
3.2.3	Aspectos técnicos	39
3.2.4	Funcionalidades de usuario	40
3.2.5	Seguridad, volumetría y Big Data	52
3.3	PowerBI	53
3.3.1	Introducción y versiones	54
3.3.2	Arquitectura	57
3.3.3	Aspectos técnicos	59
3.3.4	Funcionalidades de usuario	60
3.3.5	Seguridad, volumetría y Big Data	69
3.4	Pentaho	71
3.4.1	Introducción y versiones	71
3.4.2	Arquitectura	77
3.4.3	Aspectos técnicos	78
3.4.4	Funcionalidades de usuario	79
3.4.5	Seguridad, volumetría y Big Data	88

4	Conclusiones	89
	Referencias Bibliográficas	95
	Bibliografía General	96

Lista de Figuras

Figura 1. Internet of Things, ejemplos. Source: Elaboración propia.	2
Figura 2. Principales aplicaciones y servicios web en cifras. Fuente: Elaboración propia	3
Figura 3. Proceso ETL. Fuente: www.vperacto.com	4
Figura 4. Esquema Cuadrante Mágico Gartner. Fuente: www.gartner.com	8
Figura 5. Herramientas de BI en Cuadrante Gartner. Fuente: www.gartner.com	10
Figura 6. Logo Tableau. Fuente: www.tableau.com	11
Figura 7. Versiones Tableau. Fuente: www.tableau.com	11
Figura 8. Interfaz Tableau Desktop. Fuente: Elaboración propia	12
Figura 9. Interfaz Tableau Online. Fuente: elaboración propia	13
Figura 10. Control accesos Tableau Online. Fuente: elaboración propia	14
Figura 11. Logo Tableau Server. Fuente: elaboración propia	14
Figura 12. Fuentes datos Tableau Server. Fuente: elaboración propia	15
Figura 13. App Tableau en Play Store (Android). Fuente: elaboración propia	15
Figura 14. Interfaz App Tableau Android. Fuente: elaboración propia	16
Figura 15. Arquitectura cliente-servidor en Tableau. Fuente: www.intellipaat.com	17
Figura 16. Arquitectura Tableau. Fuente: www.tableau.com	17
Figura 17. Arrastrar y soltar en Tableau. Fuente: elaboración propia	19
Figura 18. Menú conexión a datos. Fuente: elaboración propia	21
Figura 19. Menú integración datos. Fuente: elaboración propia	21
Figura 20. Previsualización de datos. Fuente: elaboración propia	22

Figura 21. Interfaz principal. Fuente: elaboración propia	22
Figura 22. Hojas en Tableau. Fuente: elaboración propia	23
Figura 23. Hoja de trabajo. Fuente: elaboración propia	23
Figura 24. Dashboard. Fuente: elaboración propia	24
Figura 25. Historias. Fuente: elaboración propia	25
Figura 26. Cuadros y tarjetas. Fuente: elaboración propia	26
Figura 27. Columnas y filas. Fuente: elaboración propia	26
Figura 28. Pestaña Mostrarme. Fuente: elaboración propia	27
Figura 29. Línea de tendencia y P-value. Fuente: elaboración propia	28
Figura 30. Gráfico líneas de tendencia por regiones. Fuente: elaboración propia	28
Figura 31. Creación nuevas variables. Fuente: elaboración propia	29
Figura 32. Clúster con tendencias. Fuente: elaboración propia	30
Figura 33. Esquema integración Tableau y R. Fuente: www.thorogood.com	31
Figura 34. Logo Qlik. Fuente: www.qlik.com	33
Figura 35. Interfaz Qlik Sense Desktop. Fuente: elaboración propia	34
Figura 36. Logo Qlik Sense Cloud. Fuente: www.qlik.com	34
Figura 37. Interfaz Qlik View. Fuente: www.qlikblog.com	36
Figura 38. Logo Qlik NPrinting. Fuente: www.qlik.com	36
Figura 39. Logo Qlik GeoAnalytics. Fuente: www.qlik.com	37
Figura 40. Arquitectura Qlik. Fuente: www.globustrainings.com	38

Figura 41. Interfaz Qlik Desktop. Fuente: elaboración propia	39
Figura 42. Fuentes de datos en Qlik. Fuente: www.qlik.com	41
Figura 43. Archivos admitidos Qlik. Fuente: elaboración propia	41
Figura 44. Centro de control. Fuente: elaboración propia	42
Figura 45. Hojas en Qlik. Fuente: elaboración propia	43
Figura 46. Pestañas. Fuente: elaboración propia	43
Figura 47. Botonera interactiva Qlik. Fuente: elaboración propia	44
Figura 48. Información de hoja. Fuente: elaboración propia	44
Figura 49. Botonera. Fuente: elaboración propia	45
Figura 50. Menú dimensión y medida sobre gráfico. Fuente: elaboración propia	45
Figura 51. Menú lateral Qlik. Fuente: elaboración propia	46
Figura 52. Menú visualizaciones Qlik. Fuente: elaboración propia	47
Figura 53. Muestra gráficos Qlik 1. Fuente: elaboración propia	48
Figura 54. Muestra gráficos Qlik 2. Fuente: elaboración propia	48
Figura 55. Menú creación variables. Fuente: elaboración propia	49
Figura 56. Añadir campo calculado. Fuente: elaboración propia	49
Figura 57. Logos ezdatamunch y sistel. Fuente: www.ezdatamunch.com y www.sistel.com	50
Figura 58. Qlik DataMarket. Fuente: www.qlik.com	51
Figura 59. Captura Qlik DataMarket. Fuente: elaboración propia	51
Figura 60. Captura tipos de cambio. Fuente: elaboración propia	52

Figura 61. Logo PowerBI. Fuente: www.powerbi.microsoft.com	53
Figura 62. Estilo de interfaz PowerBI. Fuente: elaboración propia	54
Figura 63. Limitaciones versión free y premium. Fuente: www.powerbi.microsoft.com	55
Figura 64. PowerBI Mobile Play Store. Fuente: elaboración propia	56
Figura 65. PowerBI Embedded. Fuente: www.powerbi.microsoft.com	57
Figura 66. PowerBI front-end. Fuente: www.powerbi.microsoft.com	58
Figura 67. PowerBI back-end. Fuente: www.powerbi.microsoft.com	59
Figura 68. Ejemplo estructura sentencias DAX. Fuente: elaboración propia	60
Figura 69. Gráfico DAX ejemplo. Fuente: elaboración propia	60
Figura 70. Menú ingestión datos. Fuente: elaboración propia	61
Figura 71. Botón Get data en menú. Fuente: elaboración propia	61
Figura 72. Interfaz PowerBI. Fuente: elaboración propia	62
Figura 73. Hojas y menú interactivos. Fuente: elaboración propia	63
Figura 74. Pestaña de Data en PowerBI. Fuente: elaboración propia	63
Figura 75. Pestaña de conexiones existentes en PowerBI. Fuente: elaboración propia	64
Figura 76. Menú visualizaciones. Fuente: elaboración propia	64
Figura 77. Menú filtros. Fuente: elaboración propia	65
Figura 78. Ejemplo visualizaciones 1. Fuente: elaboración propia	65
Figura 79. Ejemplo visualizaciones 2. Fuente: elaboración propia	66
Figura 80. Ejemplo visualizaciones 3. Fuente: elaboración propia	66

Figura 81. RScript Studio. Fuente: elaboración propia	68
Figura 82. Visualizacion creada a partir de script R en PowerBI. Fuente: elaboración propia	68
Figura 83. Azure Active Directory Logo. Fuente: elaboración propia	69
Figura 84. Kafka, Apache Storm y Spark en Streaming con Azure HDInsight. Fuente: www.azure.microsoft.com	70
Figura 85. Componentes y utilidades en clústeres HDInsight. Fuente: www.docs.microsoft.es	71
Figura 86. Logo Pentaho. Fuente: www.pentaho.com	71
Figura 87. Opciones de instalación Pentaho. Fuente: elaboración propia	72
Figura 88. Interfaz Pentaho Aggregation Designer. Fuente: elaboración propia	73
Figura 89. Interfaz PDI. Fuente: elaboración propia	74
Figura 90. Interfaz Pentaho Report Designer. Fuente: elaboración propia	74
Figura 91. Start/stop Pentaho Server. Fuente: elaboración propia	76
Figura 92. Uso de memoria de Pentaho Server. Fuente: elaboración propia	76
Figura 93. Interfaz principal consola web Pentaho. Fuente: elaboración propia	76
Figura 94. Consola de usuario, opciones. Fuente: www.pentaho.com	77
Figura 95. Servidor analítico. Fuente: www.pentaho.com	77
Figura 96. Fuentes de datos para el servidor. Fuente: www.pentaho.com	78
Figura 97. Logo Pentaho Open Source. Fuente: www.bigdataanalyticsnews.com	78
Figura 98. Data Source Wizard. Fuente: elaboración propia	81
Figura 99. Menú interactivo principal. Fuente: elaboración propia	82
Figura 100. Menú rápido Opened. Fuente: elaboración propia	82

Figura 101. Menú rápido, Browse Files. Fuente: elaboración propia	83
Figura 102. Menú rápido Schedules. Fuente: elaboración propia	84
Figura 103. Menú rápido Administration. Fuente: elaboración propia	84
Figura 104. Vista general espacio de trabajo. Fuente: elaboración propia	85
Figura 105. Vistas de gráfico disponibles en Pentaho. Fuente: elaboración propia	86
Figura 106. Ejemplos gráficos con Pentaho. Fuente: elaboración propia	86
Figura 107. Opciones de transformación de datos en Pentaho PDI. Fuente: elaboración propia	87
Figura 108. Orden Arquitectura. Fuente: elaboración propia	90
Figura 109. Orden Facilidad uso. Fuente: elaboración propia	90
Figura 110. Orden Facilidad con dimensiones. Fuente: elaboración propia	91
Figura 111. Orden Atractivo interfaz. Fuente: elaboración propia	92
Figura 112. Orden Atractivo visualizaciones. Fuente: elaboración propia	93
Figura 113. PowerBI y Pentaho, soluciones de BI con funcinoalidad BigData. Fuente: elaboración propia.	94

1 Introducción

La información es poder. Nos permite tomar buenas decisiones, reducir riesgos y optimizar recursos. Desde nuestros orígenes hemos buscado datos con los que generar esta información útil y durante mucho tiempo esta búsqueda ha sido el principal problema. Sin embargo, los avances tecnológicos han hecho mutar esta situación, siendo ahora el problema no la búsqueda, sino el almacenamiento y el correcto tratamiento de la enorme cantidad de datos disponibles.

Según un estudio realizado por IBM, en la actualidad se generan **2,5 cuatrillones de bytes de datos al día**¹, llegando a la situación en que nos encontramos, donde más del 90% de los datos existentes en el mundo se han creado en los últimos dos años.

Las fuentes de datos son innumerables, abarcando todos los ámbitos de la sociedad:

- **Ciencia:** astronomía, medicina, genómica, medioambiente, transporte, etc.
- **Ciencias sociales y humanidades:** digitalización de documentos, datos económicos, sociales, etc.
- **Negocio y comercio:** cifras de compra venta, cotización de acciones, transacción de mercados, transacciones bancarias, etc.
- **Entretenimiento y ocio:** imágenes digitales, fotografía, películas, audio, etc.
- **Medicina:** historiales médicos,
- **Industria:** multitud de sensores, software, etc.

Y estas cifras van en aumento, pues cada vez son más los dispositivos conectados de nuestro entorno que capturan datos conformando el denominado *Internet Of Things*², **Figura 1**. Nos encontramos en un momento donde todo lo que nos rodea se vuelve inteligente. Los instrumentos

¹ “Bringing big data to the enterprise.” (IBM, 2017)

² Término usado para referirse al conjunto de elementos de nuestro entorno que ahora cuentan con la tecnología suficiente para captar y compartir datos

y dispositivos más cotidianos se transforman en potenciales captadores de datos, desde nuestros dispositivos móviles, a los *wearables*³, así como nuestro vehículo, TV o nevera.



Figura 1. Internet of Things, ejemplos. Source: Elaboración propia.

Somos generadores de información, con las compras que realizamos, tanto de forma física como online, las consultas que lanzamos a la web, el uso de aplicaciones, mensajes de correo electrónico, y cómo no, con la enorme cantidad de datos que volcamos en las redes sociales. Estas últimas han cobrado especial relevancia en los últimos años, siendo espacios donde incorporamos información de todo tipo, más o menos sensible. Mandamos en ellas mensajes, opiniones, imágenes, fotografías personales, nuestra documentación, opiniones, gustos, etc. Se muestran a continuación algunas cifras que nos pueden ayudar a entender la situación que nos encontramos, **Figura 2**, en cuanto al volumen de información que generamos a través de las mismas.

³ Término utilizado para referirse a la *tecnología vestible*. Incluye a todos los dispositivos que llevamos encima y que interactúan con nosotros, de forma consciente e/o inconsciente, como por ejemplo relojes, pulseras, gafas, prendas de ropa con fibras inteligentes, etc.



Figura 2. Principales aplicaciones y servicios web en cifras. Fuente: Elaboración propia

La previsión es que estas cifras no hagan más que aumentar, así como el ritmo en que lo hacen. Empresas y organizaciones se han visto en muy poco tiempo desbordadas ante este cambio de paradigma, pasando desde la escasez a la avalancha de datos. Hasta hace poco tiempo, la forma en la que se abordaban los problemas de procesamiento de volúmenes de datos cada vez mayores era la inversión en equipos más potentes, con mayor cantidad de CPU, memoria RAM, GPU, etc. Esta solución sirvió durante un tiempo, pero poco a poco resultó insuficiente y los problemas empezaron a hacerse insostenibles, siendo muy frecuente encontrar problemas como los “*cuellos de botella*” en el acceso a disco, donde los sistemas de almacenamiento resultaron ser incapaces de evolucionar al ritmo que lo hacían los medios con los que se procesaban los datos.

El bajo rendimiento en los sistemas de recogida y almacenamiento masivo de datos ha venido derivado por dos problemas principales. El primero, la **falta de preparación** por parte de muchas empresas que no han previsto estos cambios con suficiente antelación y ahora luchan por recuperar el ritmo. Por otro lado, la **falta de medios y conocimiento** suficiente para abordar un problema, este bajo rendimiento, que requiere de un cambio radical en la forma de pensar y actuar.

Para que las organizaciones puedan hacer uso de estos datos de forma eficaz deben seguir un proceso compuesto por tres pasos, denominado como **ETL (Extraer, transformar y cargar)**. Este proceso básico permite a las organizaciones trasladar la información captada desde múltiples fuentes, transformarla de forma que sea manejable y cargarla en una base de datos propia para posteriormente estudiarla, **Figura 3**.

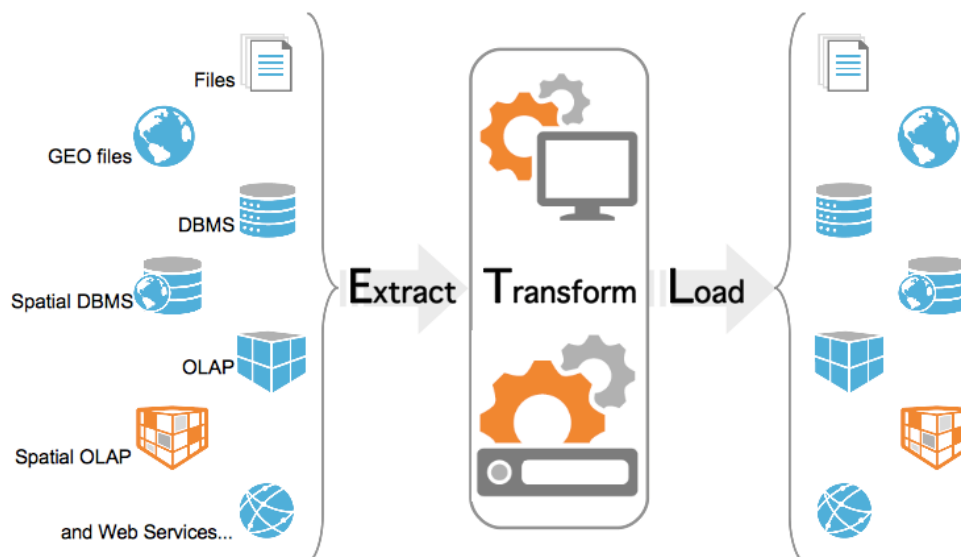


Figura 3. Proceso ETL. Fuente: www.vperacto.com

En el primer paso, la **extracción**, el sistema debe obtener los datos desde la fuente en que se encuentran. Esta fuente puede ser una base de datos externa, una **API**⁴ de otra aplicación o red social, una herramienta de *web crawling*⁵, etc. La **transformación** será el proceso de limpieza de los datos, eliminando toda aquello que no nos resulte útil o cuente con algún tipo de error o problema. Por último, la **carga**, será el momento en el que estos datos útiles serán volcados a nuestro sistema para su almacenamiento y explotación.

Será aquí donde entrarán en juego las herramientas de *Business Intelligence*, que supondrán el punto de partida del analista para sacar todo el rendimiento posible de los datos con los que se cuentan. Será su labor por tanto utilizar estas herramientas para extraer información elemental,

⁴ Siglas de *Application Programming Interface*. Término utilizado para referirse al conjunto de funciones y procedimientos de una aplicación que son accesibles para ser utilizadas por otro software externo.

⁵ Programa informático dedicado a la inspección sistemática de la web.

como por ejemplo cifras de venta, coste y beneficio, así como realizar lo que será la auténtica *mina de oro*, el descubrimiento de todo aquello que no es fácilmente observable. Nos referimos a patrones, conductas, relaciones ocultas, predicciones de comportamiento, previsiones de consumo, cambios del entorno que podrían beneficiar o afectar a nuestra actividad, etc. Es por tanto un punto clave del proceso, en el que los conocimientos del analista y la herramienta utilizada serán claves para garantizar el éxito.

Contar con la mayor cantidad y calidad de datos para realizar este análisis es sumamente valioso. Como veremos a continuación, no todas las herramientas de *Business Intelligence* existentes en el mercado cuentan con los mismos recursos, y no todas nos permiten trabajar verdaderamente grandes conjuntos de datos. Serán tan solo unas pocas las que nos permitan realizar procesos de computación paralela y distribuida, trabajando con los denominados *clústeres*. Se trata de un elemento diferencial que puede darnos la nota diferencial respecto de nuestros competidores.

Centraremos nuestro estudio en el estudio de las mejores herramientas de *BI*⁶ del mercado, viendo sus capacidades, puntos fuertes y débiles, pero enfatizando sobre todo en esta capacidad para trabajar con *Big Data*.

2 Aproximación al Business Intelligence

2.1 Marco teórico

En el año 1958 el profesor *Hans Peter Luhn*⁷, investigador de *IBM*, dio la que puede ser una de las primeras definiciones de *Business Intelligence*, en un artículo titulado *A Business Intelligence System*⁸. En este lo define como “la habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada” (Luhn, 1958). Vemos en esta definición dos primeras claves, aprender las relaciones y guiar las acciones. Cazando con esta idea, la

⁶ *Business Intelligence*.

⁷ *Hans Peter Luhn*, <https://www.ithistory.org/honor-roll/mr-hans-peter-luhn>

⁸ *A Business Intelligence System*, <https://goo.gl/xEiJaj>

prestigiosa consultora *Gartner Inc.*⁹ amplía esta definición años después adaptándolo aún más al entorno empresarial, definiendo esto mismo como “*las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento*”(Gartner Inc., 2017). Será primordial para las empresas contar con instrumentos para substraer datos que les permitan, bajo estas técnicas, obtener información con la que apoyar su toma de decisiones.

Se presume que en la actualidad existen cuatro grandes fuentes de datos, que a su vez se pueden clasificar en dos categorías: la información *humana* y la *digital*.

Dentro del primer grupo encontramos a su vez dos apartados, siendo el primero de ellos la “información social”. Nos referimos a las fotos, videos, correos electrónicos, mensajes de texto personales, publicaciones en redes sociales. Se trata de información no estructurada que acostumbra a ser almacenada utilizando *bases de datos no relacionales*¹⁰. Por otro lado, encontramos la información que se genera, de manera más o menos intencional, cuando realizamos nuestras transacciones comerciales (incluimos aquí también a la información generada por las empresas). Hablamos de la información logística, stocks, sistemas CRM y demás. Esta información es almacenada, en contrapunto de la anterior, en *bases de datos relacionales*¹¹.

Dentro del segundo grupo se sitúan otros dos apartados. El primero, el muy popular estos días ***IOT*** (*Internet of Things*), que supone toda la información generada de manera masiva, constante y automática, por parte de los objetos conectados que nos rodean. Se incluyen aquí sensores, vehículos, dispositivos móviles, electrodomésticos etc. Se trata de información estructurada. El segundo corresponde a los ***metadatos***, toda la información que se genera a partir de la propia información, por ejemplo, los registros de lugares, fechas y horas donde se produce la información y que supone a su vez otra fuente de interés. Se trata también de una fuente de datos estructurada.

⁹ Gartner Inc., <https://www.gartner.com/technology/home.jsp>

¹⁰ “Una base de datos no relacional es una base de datos que se trata como un conjunto de tablas y se manipula de acuerdo con el modelo de datos relacional” (Carlos Paramio, 2011)

¹¹ “Una base de datos relacional es una recopilación de elementos de datos con relaciones predefinidas entre ellos. Estos elementos se organizan como un conjunto de tablas con columnas y filas” (Amazon Web Services, 2017)

Todos estos datos esperan a ser capturados, desmontados, organizados, almacenados y analizados. Para ello se reconoce la necesidad de que los procesos sean **veloces** y **certeros**, amén de que los datos seleccionados de entre todos los existentes sean **valiosos**, **variados** y **veraces**. ¿Cómo conseguir todo esto?

Resulta primordial detenernos precisamente en las mencionadas características que debe tener la información (valor, variedad y veracidad), pues el objetivo fundamental buscado con el uso de cualquier herramienta de análisis de información, en este caso herramientas de *Business Intelligence*, es la transformación de los **datos** en **información útil y fácilmente accesible** para los usuarios que la necesiten. Será por tanto un elemento clave la correcta elección de la herramienta utilizada, teniendo en cuenta nuestras necesidades, así como de los recursos con los que contamos para ponerla en funcionamiento.

Cabe preguntarnos, ¿todos estos esfuerzos tienen una repercusión real en el beneficio o mejora de las organizaciones? Por un lado, los sistemas de información facilitan a las empresas algo que resulta primordial para ellas, reducir la **incertidumbre en la toma de decisiones**, se trata en este caso de un beneficio **intangible**. Pero no todo es tan etéreo, y es que también puede provocar una mejora directa de resultados. Analizar la información puede revertir en nuestra organización, por ejemplo, una reducción en los costes de determinadas áreas, reducción de tiempos de actividad, detección de ociosidades, entre otros. Todo ello supondrá un beneficio **tangible** que la organización obtendrá a corto y medio plazo.

2.2 Elección de las soluciones estudiadas

Existen en la actualidad multitud de herramientas dedicadas a ofrecer soluciones de *Business Intelligence*. Se trata de un mercado con una demanda creciente, pues cada vez son más las empresas interesadas en hacer uso de estas herramientas, pero en el cual ya existen muchas firmas con ofertas maduras.

La ya mencionada agencia *Gartner Inc.*, agencia de consultoría tecnológica, elabora cada año un informe en el cuál presenta sus conclusiones tras analizar las distintas soluciones ofrecidas por las empresas existentes en el mercado para un determinado producto. Para hacer estos informes se valen de los mencionados *Cuadrantes Mágicos de Gartner*, **Figura 4**, la cual es una representación gráfica de la situación del mercado para un determinado tipo de producto tecnológico.

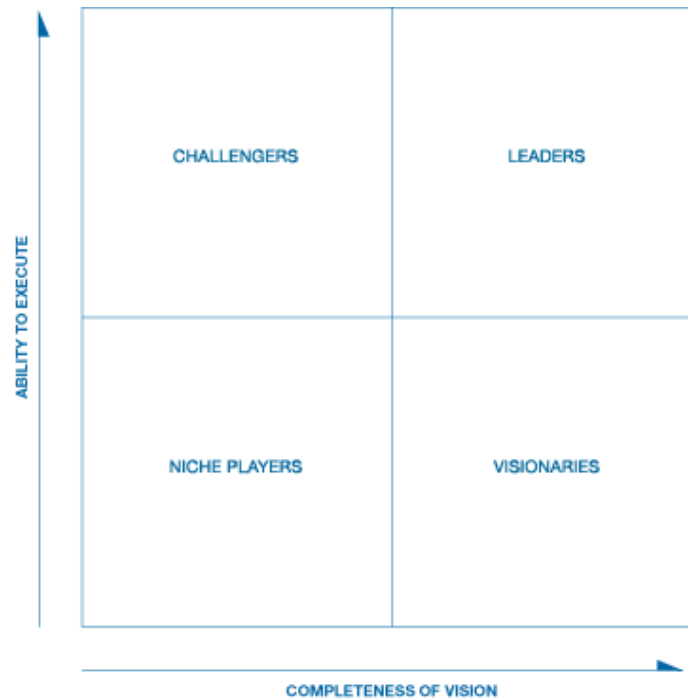


Figura 4. Esquema Cuadrante Mágico Gartner. Fuente: www.gartner.com

El cuadrante nos permite identificar el tipo de posicionamiento adoptado o conseguido por las empresas desarrolladoras, encontrando cuatro cuadrantes con los que identificamos estos posicionamientos, basándose en el grado en que cumplen sus soluciones dos características. La primera, correspondiente con el eje de las X, se define como la **integridad de la visión**, y expresa el grado en el que la solución puede generar valor para el usuario. En el eje opuesto encontramos la **capacidad de ejecutar**, que mide la capacidad para alcanzar con éxito su visión del mercado.

Para posicionar cada una de las soluciones en estos puntos revisan los siguientes puntos:

1. Infraestructura:

- Administración de la plataforma
- Seguridad
- Nube
- Conectividad e ingestión de las fuentes de datos

2. Administración de datos

- Administración de metadatos
- Extracción, transformación, carga y almacenamiento de datos

- Preparación de datos
- 3. Análisis y creación de contenido**
 - Incrustación de analíticas avanzadas
 - Dashboards
 - Exploración visual interactiva
 - Análisis inteligente de datos
 - Exploración móvil
- 4. Capacidad para compartir resultados**
 - Incrustación de análisis de contenido
 - Publicación, compartición y colaboración

En base a los resultados obtenidos en los puntos anteriores y atendiendo a los ejes del gráfico, las distintas soluciones se posicionan en un punto u otro del gráfico, ocupando uno de los cuatro roles posibles:

1. **Líderes:** son las herramientas y desarrolladores que obtienen la mayor puntuación, alcanzada al contar con una gran capacidad para visionar el mercado y una gran habilidad para ejecutar esa visión.
2. **Retadores/aspirantes:** sus herramientas ofrecen buenas funciones, pero se centran en unas pocas características por lo que su abanico de opciones es menor que la de los líderes. Sin embargo, buscan mejorar el mercado, innovando y a través de la diferencia.
3. **Visionarios:** cuentan con una elevada capacidad de visión del futuro del mercado, pero no disponen de los medios para implantar las mejoras antes que los líderes.
4. **Jugadores de nicho:** no destacan en nada y simplemente tratan de cubrir aquellos segmentos que dejan menos cubiertos sus competidores.

El estudio realizado para el año 2017 es el siguiente, **Figura 5**, el cuál muestra algunas curiosidades.



Figura 5. Herramientas de BI en Cuadrante Gartner. Fuente: www.gartner.com

Vemos como la mayoría de los desarrolladores se encuentran en los cuadrantes tres y cuatro, correspondientes a los **Jugadores de nicho** y los **Visionarios**. Se observa la existencia de homogeneidad en las soluciones, algo que queda aún más puntualizado con la inexistencia de ninguna solución en el cuadro de **Retadores**.

Para el presente estudio se han seleccionado las herramientas existentes en el cuadro de **Líderes**, por ser las que a priori, están dominando el mercado y por ende deben contar con las funcionalidades y características punteras. Se trata de *Qlik*, *Tableau* y *PowerBI*.

Así mismo, hemos seleccionado otra herramienta que si bien no destaca en el cuadrante si nos atrae por dos aspectos. El primero, es la única herramienta que nació como un proyecto *Open Source* de todas las analizadas por *Gartner*. El segundo, es una herramienta que está siendo muy comentada en los círculos y comunidades de usuarios/expertos, aparentemente por realizar una apuesta de producto muy completa, algo que analizaremos en profundidad, *Pentaho*.

3 Comparativa de herramientas

3.1 Tableau

Tableau es una herramienta de análisis, especializada en la temática del *Business Intelligence*, y que tiene como principal estandarte facilitar la realización de análisis complicados y técnicos a usuarios sin conocimientos específicos en la materia de una forma gráfica e intuitiva, **Figura 6**.



Figura 6. Logo Tableau. Fuente: www.tableau.com

3.1.1 Introducción y versiones

Cuenta con tres versiones principales del programa: *Desktop*, *Online* y *Server*, **Figura 7**. Estas tres versiones se encuentran totalmente integradas, pudiendo consultar y modificar la información elaborada en cualquiera de ellas en las demás. Esta dinámica obedece a la máxima de conseguir un entorno de trabajo colaborativo. Además de esto cuenta con otras versiones menores del programa, siendo las versiones: *Mobile*, *Public* y *Reader*.



Figura 7. Versiones Tableau. Fuente: www.tableau.com

3.1.1.1 Tableau Desktop

Se trata de la versión de escritorio del programa, **Figura 8**, la más potente y que incluye la totalidad de las funcionalidades analíticas. Está disponible para sistemas operativos **Microsoft Windows (versión 7 o posterior)** y **macOS (versión 10.10 o posterior)**. Los entornos virtuales disponibles son **VMWare, Citrix, Hyper-V y Parallels**, advirtiéndose además que con el hardware y sistema operativo apropiado pueden llegar a funcionar en otros entornos virtualizados, sin mayor precisión. Se encuentra habilitado para **Unicode**, siendo por tanto compatible con cualquier idioma. Su documentación se encuentra disponible en inglés, francés, alemán, español, portugués, japonés, coreano y chino simplificado.

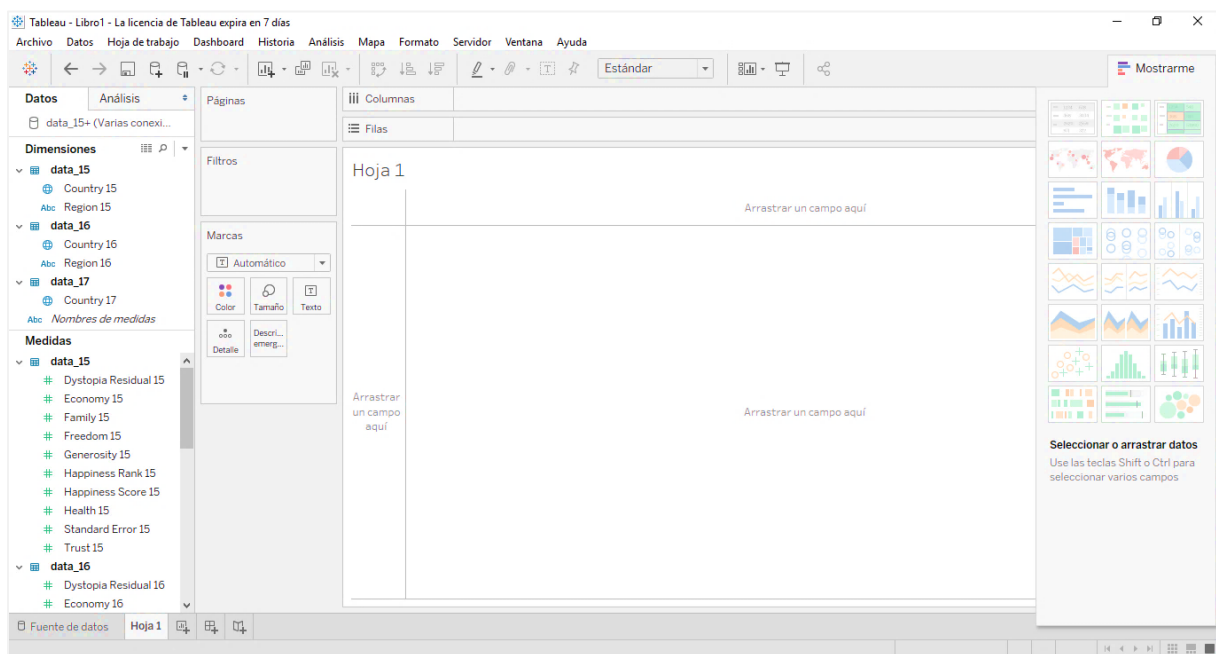


Figura 8. Interfaz Tableau Desktop. Fuente: elaboración propia

Se trata de una herramienta de análisis totalmente integrado en la nube, **Figura 9**, por lo que accediendo a esta a través de un navegador web tendremos acceso a una interfaz idéntica a la de escritorio, realizar las mismas acciones y funcionalidades que con esta última.

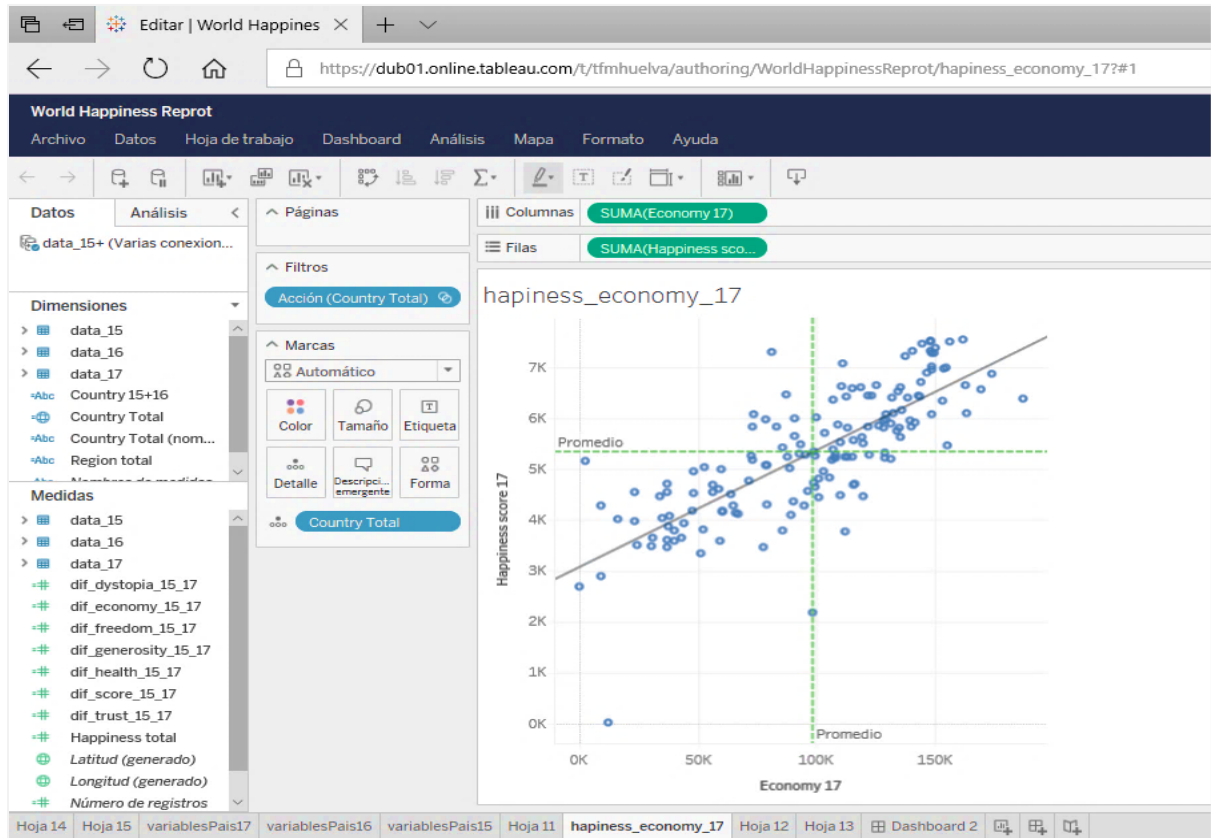


Figura 9. Interfaz Tableau Online. Fuente: elaboración propia

Se puede hacer uso del mismo desde cualquier navegador o dispositivo móvil. Cuenta además con un interfaz reducido cuando no se pretende modificar o crear contenido y lo único que deseamos es visualizar e interactuar con los elementos, por ejemplo, *dashboards* (más adelante veremos qué son) creados con la versión de escritorio.

Cuenta con un eficiente y sencillo sistema para controlar el acceso de los usuarios a la información,

Figura 10.

hido	Usuarios	Grupos	Programas	Tareas	Estado	Configuración	🔍
Usuarios							
Nombre de visualización	Nombre de usuario	Función de sitio	Grupos	Autenticación	Último inicio de sesión		
 david blanco	*** davidmbq@gmail.com	Administrador de sitio	1	Tableau	29 de oct. de 2017 1:38		

Figura 10. Control accesos Tableau Online. Fuente: elaboración propia

3.1.1.2 Tableau Server

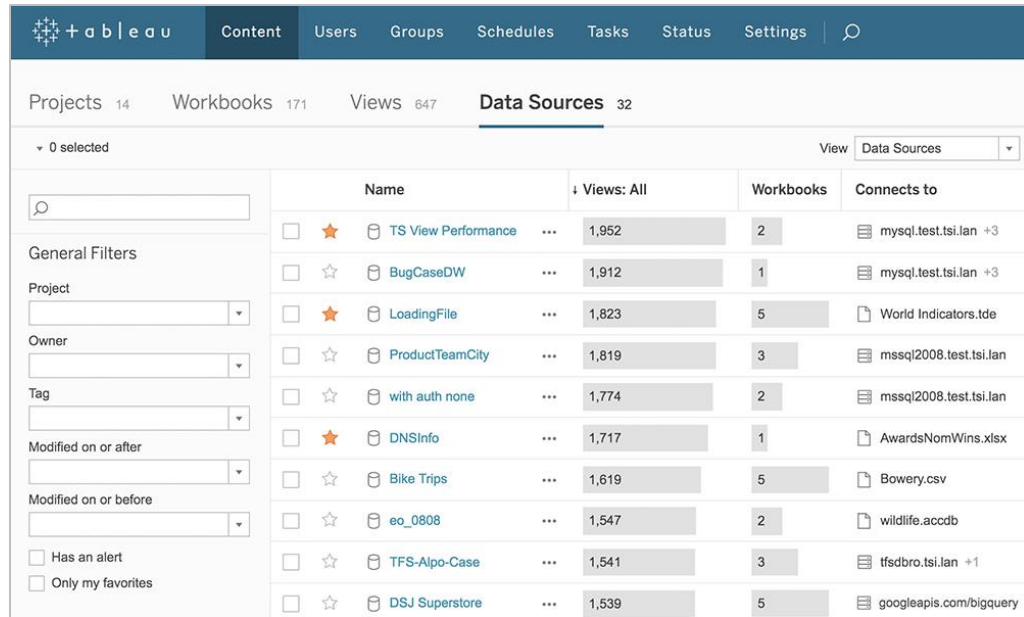
La versión para servidor de *Tableau*, **Figura 11**, permite integrar la aplicación dentro de los sistemas de la empresa. Su capacidad de adaptación y administración a fuentes de datos es muy extensa, abarcando casi la totalidad de las opciones más usadas y existentes en el mercado en la actualidad.



Figura 11. Logo Tableau Server. Fuente: elaboración propia

Este acceso a los datos por parte del programa no se limita a los datos almacenados dentro de nuestros sistemas, sino que también puede integrar los que se encuentren en la nube (este punto será tratado en profundidad más adelante).

Desde la propia aplicación, **Figura 12**, podemos controlar el acceso por parte de los usuarios a según qué información, controlar la conectividad de fuentes de datos y el estado del servidor por parte de los mismos. Cuenta además con una elevada escalabilidad, pudiendo adaptarse a empresas de reducidas dimensiones tanto como a grandes multinacionales. Esta versión del programa solo está disponible para servidores con sistema operativo *Windows*.



Name	Views: All	Workbooks	Connects to
TS View Performance	1,952	2	mysql.test.tsi.lan +3
BugCaseDW	1,912	1	mysql.test.tsi.lan +3
LoadingFile	1,823	5	World Indicators.tde
ProductTeamCity	1,819	3	mssql2008.test.tsi.lan
with auth none	1,774	2	mssql2008.test.tsi.lan
DNSInfo	1,717	1	AwardsNomWins.xlsx
Bike Trips	1,619	5	Bowery.csv
eo_0808	1,547	2	wildlife.acddb
TFS-Alpo-Case	1,541	3	tfsdbro.tsi.lan +1
DSJ Superstore	1,539	5	googleapis.com/bigquery

Figura 12. Fuentes datos Tableau Server. Fuente: elaboración propia

3.1.1.3 Tableau Mobile

Se trata de la versión móvil de Tableau y se encuentra disponible para los sistemas operativos iOS y Android en las respectivas plataformas de aplicaciones, App Store y Google Play. En la Play Store, **Figura 13**, encontramos que la aplicación cuenta con más de cincuenta mil descargas y una ratio de calificaciones medio, aunque especialmente bajo para las últimas versiones del programa.

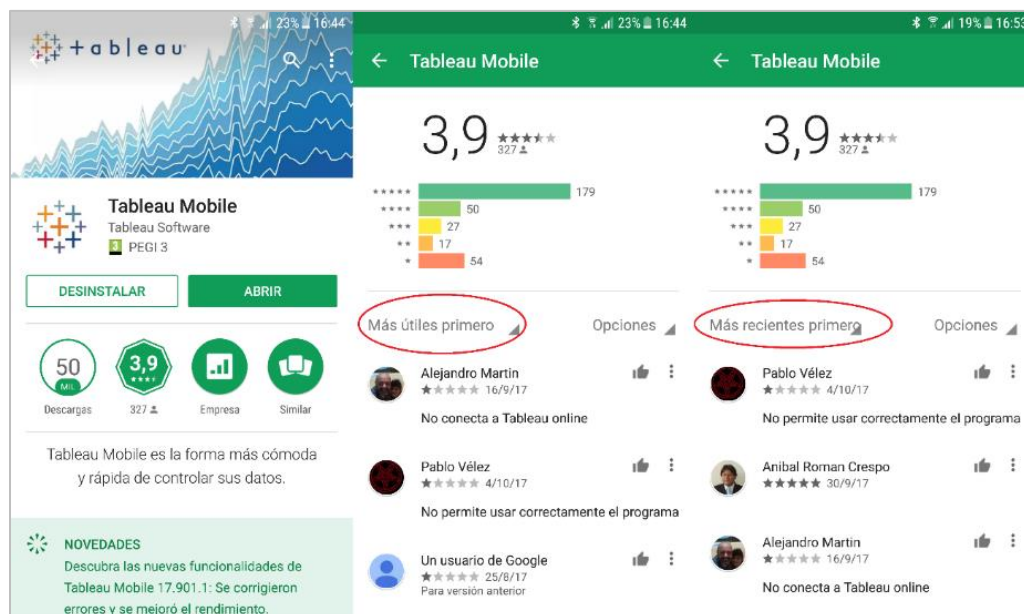


Figura 13. App Tableau en Play Store (Android). Fuente: elaboración propia

Se trata de una aplicación destinada a hacer más portable la visualización de contenidos previamente desarrollados, **Figura 14**, en las versiones de escritorio y online. Esta aplicación también permite, según versiones y plataformas, crear y editar informes.

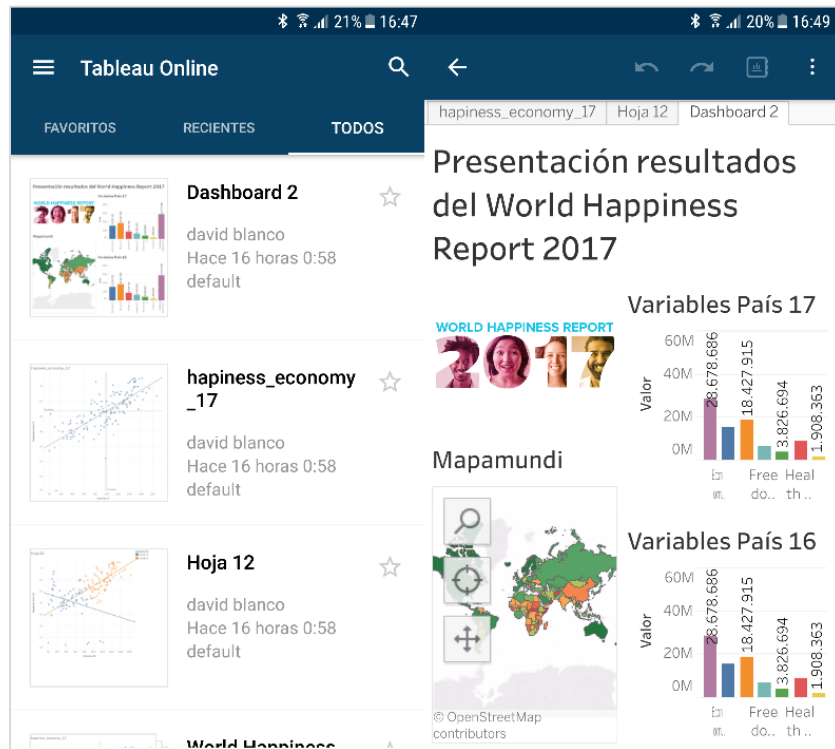


Figura 14. Interfaz App Tableau Android. Fuente: elaboración propia

3.1.1.4 Tableau Public

Se trata de una versión gratuita del programa disponible para descarga desde la web y que cuenta, como es lógico, con numerosas limitaciones tanto por el tipo de fuente de datos utilizable como de funcionalidades de cara a análisis. Como fuente de datos puede utilizarse Excel, Access, ficheros de texto o un servidor de tipo *OData*. Otra limitación importante es que no permite la descarga del trabajo en local, siendo de uso exclusivo en el perfil online. Disponible para Windows y Mac.

3.1.1.5 Tableau Reader

Versión de escritorio exclusivamente para visualización de datos, disponible para Windows y Mac.

3.1.2 Arquitectura

Cuenta con una arquitectura tipo *cliente-servidor*¹², que se encargará de suministrar (el servidor) la información a los demandantes, **Figura 15**.



Figura 15. Arquitectura cliente-servidor en Tableau. Fuente: www.intellipaat.com

Esta arquitectura a nivel está compuesta por una serie de niveles, siendo estos la capa de datos, conectores de datos, componentes, puerta de enlace/equilibrador de carga y clientes, **Figura 16**.

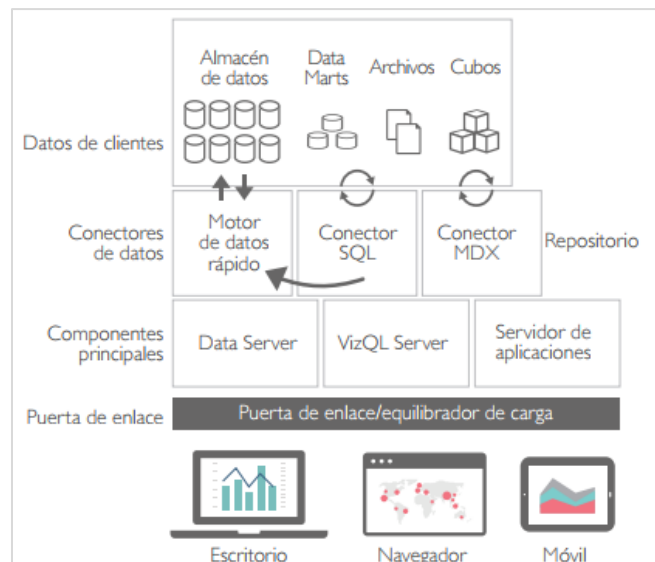


Figura 16. Arquitectura Tableau. Fuente: www.tableau.com

¹² “Esta arquitectura consiste básicamente en un cliente que realiza peticiones a otro programa (el servidor) que le da respuesta” (Ecured.cu, 2016)

La **capa de datos** se caracteriza por ser muy flexible, no requiriendo que todos los datos se encuentren alojados en una misma fuente. En lo que respecta a la **gestión de memoria**, *Tableau* permite simultanear la carga de los datos tanto en memoria como en disco. Por su parte, los conectores de datos son múltiples, encontrando más de cuarenta distintos, así como un conector genérico de tipo **ODBC**, que permiten tanto la conexión en vivo como en memoria. La aplicación requiere el funcionamiento conjunto de cuatro procesos de servidor para funcionar:

1. **Servidor de aplicaciones:** controlan la exploración, administración del servidor y ejecución de las interfaces.
2. **VizQL Server:** funciona como el punto intermedio entre la fuente de datos y el usuario, sirviendo como puente entre estos para las consultas realizadas.
3. **Data Server:** Este proceso se encarga de administrar y almacenar las fuentes de datos utilizadas por la aplicación.
4. **Componentes de segundo plano:** se encarga de las demás funciones que los tres procesos anteriores no ejecutan, como las extracciones programadas, el servicio de notificaciones y otras tareas.

La **puerta de enlace** direcciona las solicitudes desde los clientes hacia los componentes, funcionando además como un balanceador natural, distribuyendo los procesos de forma equitativa.

3.1.3 Aspectos técnicos

3.1.3.1 Arrastrar y soltar

El sistema de trabajo de *Tableau* se basa en la intuición y familiaridad, valiéndose de una tecnología en forma de lenguaje denominado *VizQL*, un proyecto surgido en el año 2003 de la universidad de *Stanford* y se basa en el concepto de “arrastra-soltar”, **Figura 17**, para la creación de visualizaciones sofisticadas.

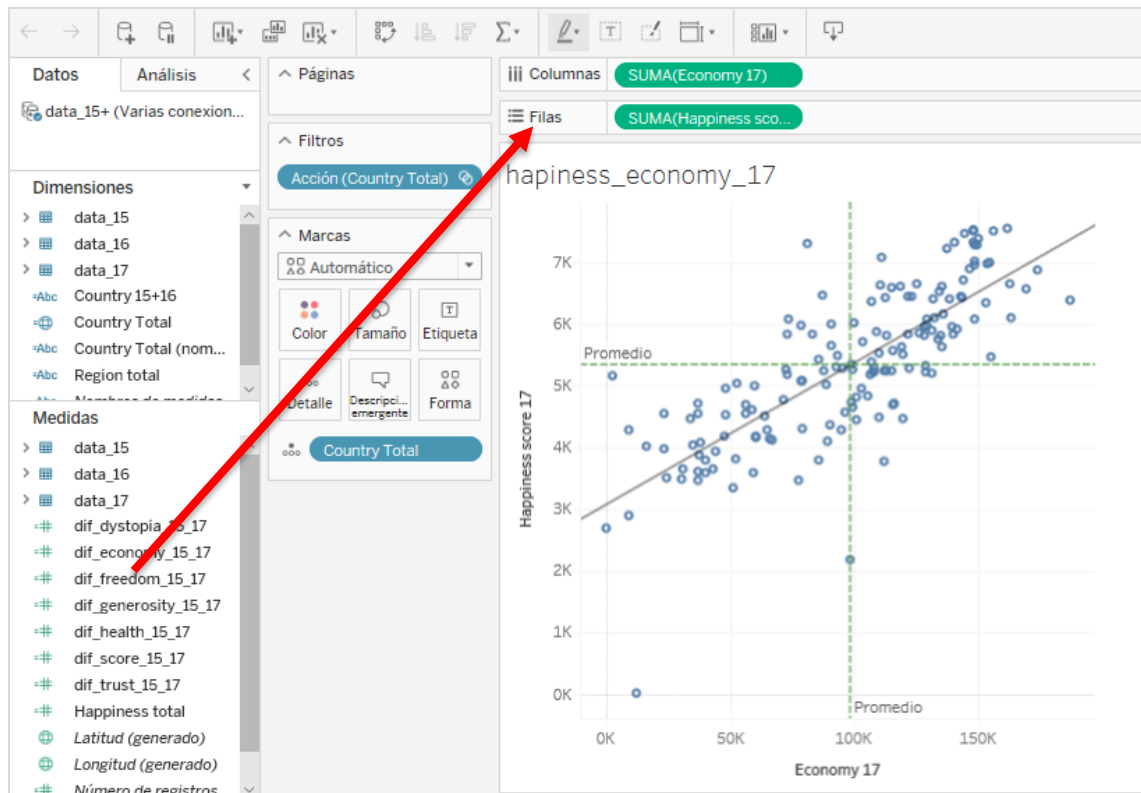


Figura 17. Arrastrar y soltar en Tableau. Fuente: elaboración propia

Bajo esta idea el usuario, con los datos cargados, tendrá una interfaz desde la cuál simplemente arrastrará aquellas variables que quiere estudiar y las colocará sobre el gráfico que pretende construir, generando automáticamente los mismos, contando además con un menú rápido para seleccionar el tipo de gráfico que desea (barras, líneas, mapa, etc.)

3.1.3.2 Motor de consultas interactivo y en vivo

El motor de consultas incluido en la herramienta *Tableau* permite de manera muy sencilla que usuarios *legos* en la materia puedan acceder a fuentes de datos muy diversos (bases de datos, nube, Hadoop...) sin que para ello tengan que tener conocimientos de programación. Siguiendo las instrucciones obtenidas en pantalla y con no más complicación de la que supone instalar un programa o acceder a una ruta con el explorador de archivos de *Windows* se puede hacer la ingestión de datos.

3.1.4 Funcionalidades de usuario

3.1.4.1 Fuentes y conexión a datos

Tableau tiene disponible multitud de fuentes de datos distintas, pudiendo combinarse de la forma que el usuario determine, siendo las siguientes:

- **Archivo:** Excel, archivos de texto, Access, archivos JSON, archivos PDF, archivos especiales, archivos de estadística...
- **Bases de datos:** *Tableau server, Microsoft SQL server, MySQL Oracle, Amazon Redshift, IBM DB2, Odata, Teradata, PosgtgreSQL*. Cuenta además con un conector de tipo *ODCB* genérico para cualquier sistema sin conexión nativa.
- **Ficheros de tipo SAS**
- **Cloud/Big data:** permite la conexión a entornos de Big Data, como por ejemplo *Hadoop, Spark, Microsoft, Azure, Cloudera*.
- **Cubos OLAP**¹³
- **Otros:** *Tableau* permite además la conexión a fuentes de datos distintas de las anteriores, como páginas webs, RRSS, repositorios, *Google Analytics*, etc.
- **Modo offline:** el programa cuenta con la opción de almacenar los datos en memoria para mejorar el tiempo de respuesta de las consultas y también a almacenamiento en disco.

La ingestión de los datos es muy simple, pudiendo realizarse directamente a través de la interfaz principal que aparece al iniciar el programa. Concretamente se realiza mediante el menú *Conectar*, que nos indicará los tipos de conexiones disponibles, **Figura 18**.

¹³ *Online Analytical Processing*, es el término dado para hacer referencias a una base de datos multidimensional caracterizada por el uso de vectores multidimensionales como medio para el almacenamiento físico.



Figura 18. Menú conexión a datos. Fuente: elaboración propia

Tableau nos permite hacer la conexión entre las distintas fuentes de datos y gestionar de qué forma queremos realizar la integración de la información de una forma muy intuitiva, mediante un diagrama de unión configurable. Se trata del *Menú de integración*, **Figura 19**, que nos permite seleccionar hasta 4 tipos de uniones de fuentes de datos.

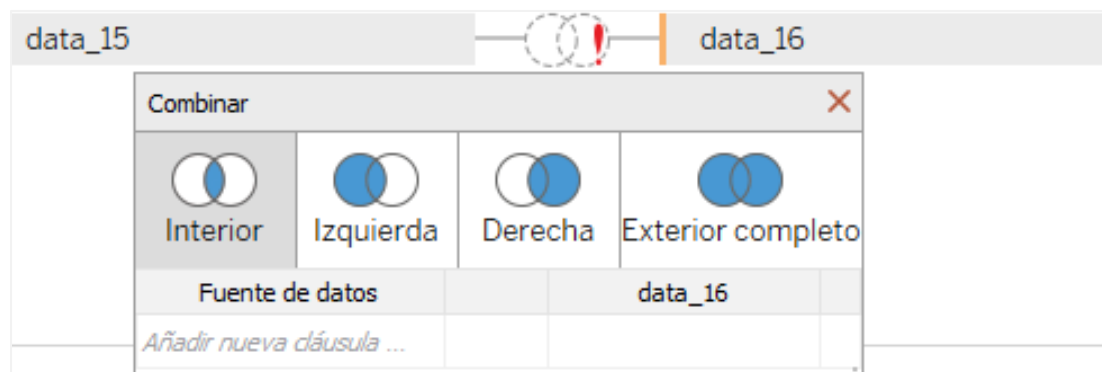


Figura 19. Menú integración datos. Fuente: elaboración propia

Una vez cargados los datos la aplicación nos ofrece una previsualización de los datos, **Figura 20**, donde además podemos si queremos ejecutar modificaciones en los mismos, cambiando por ejemplo nombres de dimensiones ya existentes, o generando otras nuevas.

data15 Country 15	data16 Country 16	data17 Country 17	# data15 Dystopia Residual ...	# data16 Dystopia Residual ...	# data17 Dystopia Residual ...
Taiwan	Taiwan	nulo	232.323	261.523	nulo
Kuwait	Kuwait	Kuwait	187.634	228.085	1,64043
Suriname	Suriname	nulo	279.094	306.852	nulo

Figura 20. Previsualización de datos. Fuente: elaboración propia

3.1.4.2 Interfaz

La interfaz de la aplicación resulta muy sencilla, contando con distintos menús y botones que nos facilitarán las tareas a realizar, **Figura 21**.

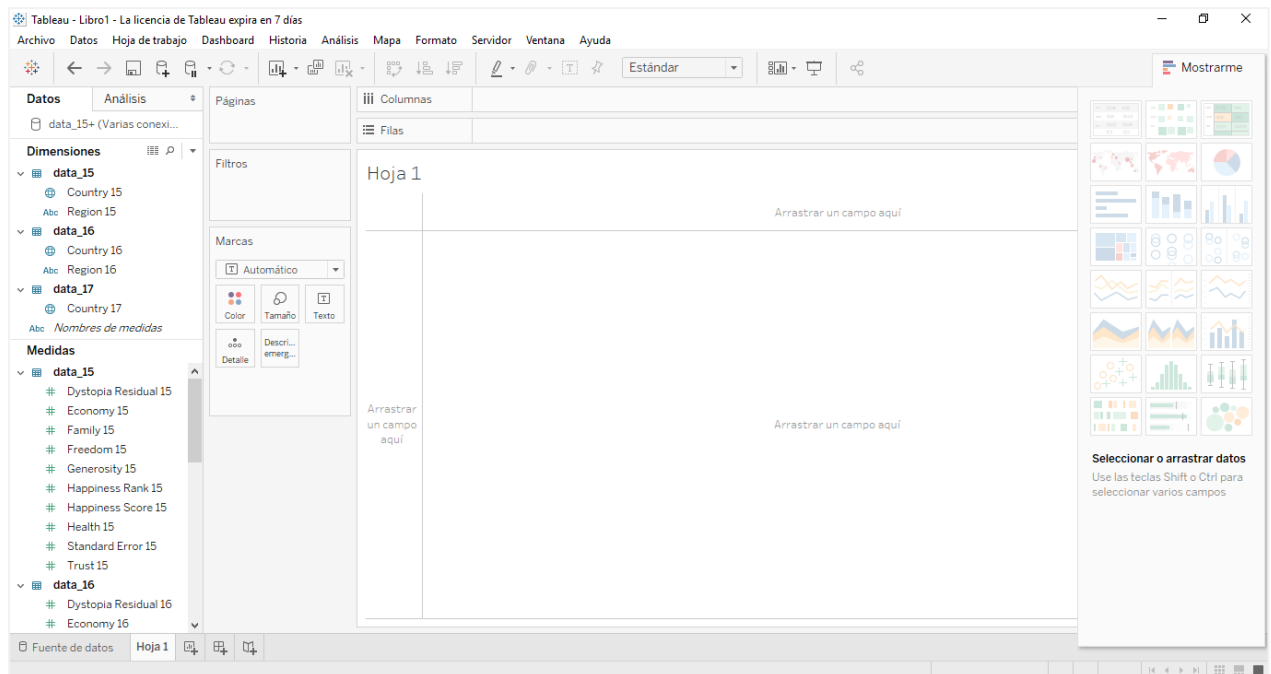


Figura 21. Interfaz principal. Fuente: elaboración propia

A continuación, se analizan los distintos apartados que componen la estructura de trabajo en *Tableau*.

En primer lugar, los **Libros de Trabajo**. Estos elementos son los proyectos, compuestos por *hojas*, en los que trabajaremos. Por ejemplo, podríamos contar con un libro de trabajo para análisis de

recursos humanos y otro para realizar un informe de resultados financieros del ejercicio. Posteriormente, las **Hojas**, que son el pilar fundamental de la estructura de trabajo en *Tableau*, encontrando tres tipos de hoja, **Figura 22**



Figura 22. Hojas en Tableau. Fuente: elaboración propia

El primer tipo es la **hoja de trabajo**, **Figura 23**, que consistirá en un espacio central en blanco sobre el que se construirán los gráficos a partir de los elementos que decidamos estudiar y que situaremos sobre los ejes (o en su defecto en los cuadros de filas y columnas que comentaremos posteriormente)



Figura 23. Hoja de trabajo. Fuente: elaboración propia

El segundo tipo de hoja se denomina **dashboard** y son un tipo especial de hoja donde podemos situar varios gráficos juntos que interactuarán entre sí. Si realizamos sobre alguno de ellos algún tipo de selección. Estos dashboards resultarán un elemento muy potente con el que estudiar distintos aspectos/variables sobre un mismo elemento de forma conjunta. Por ejemplo, en la **Figura 24**, encontramos una de estas hojas donde hay un mapa del mundo.

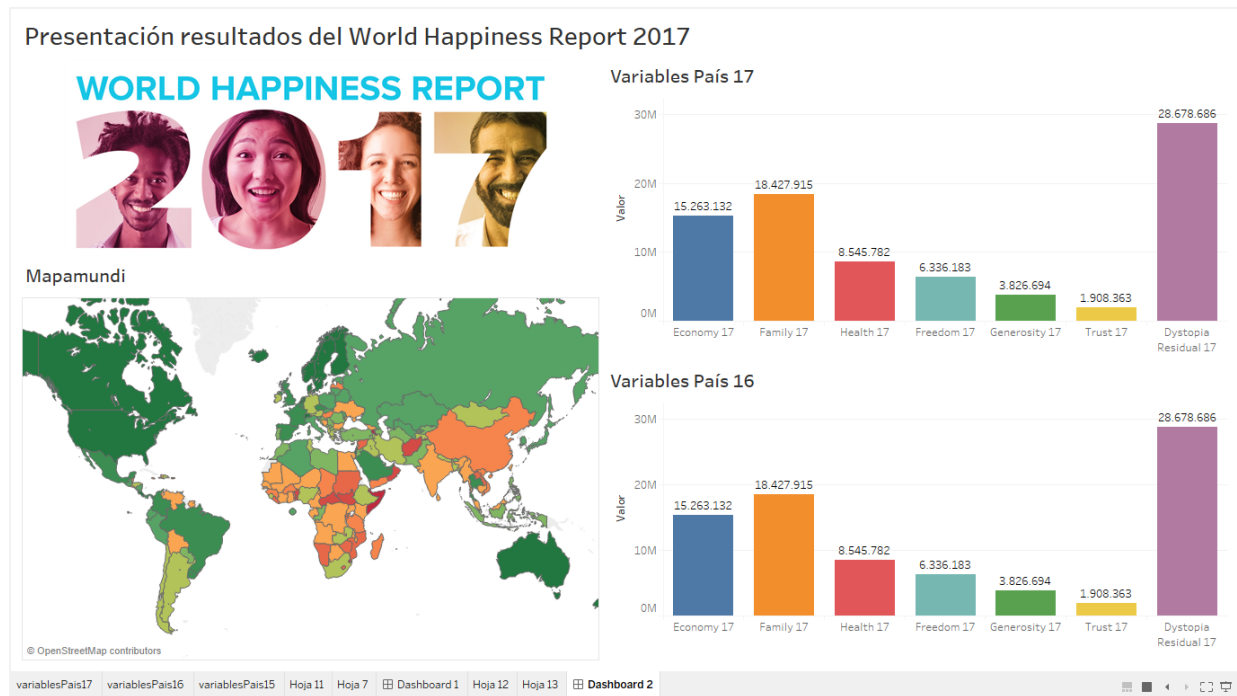


Figura 24. Dashboard. Fuente: elaboración propia

Al seleccionar alguno de los países el resto de los gráficos seleccionarán y mostrarán exclusivamente la información relativa a dicho país. Se trata de algo muy interesante también de cara a exposiciones y presentación de información ante público.

El tercer tipo de hoja se denomina **Historias** y se trata de un sistema muy eficaz para mostrar la evolución de un mismo gráfico o variable con el paso del tiempo. Este tipo de gráfico nos permitirá realizar una exposición temporal de manera clara y dinámica. En la **Figura 25**, se muestra una secuencia de cuatro gráficos que compondrían la historia, en este caso la evolución del par yuan/dólar y su evolución, remarcando fechas clave.

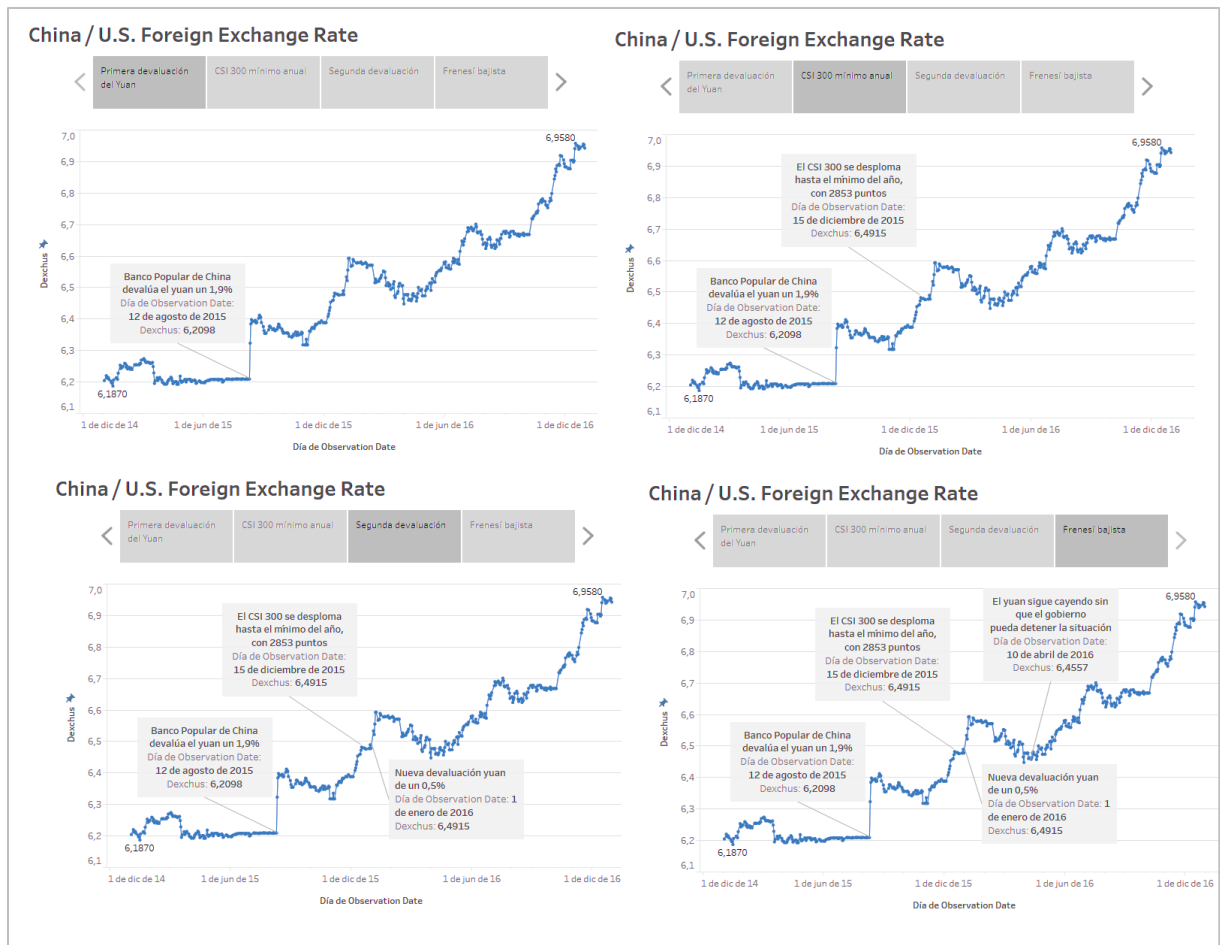


Figura 25. Historias. Fuente: elaboración propia

Posteriormente, los **cuadros/tarjetas**. Se trata de los espacios definidos para realizar las configuraciones que den lugar a nuestros gráficos. Encontramos el **cuadro de páginas**, el de **filtros** para aplicar, marcas (para delimitar colores, tamaños, etiquetas...) y el cuadro central para mostrar el gráfico resultante, como se muestra en la **Figura 26**.

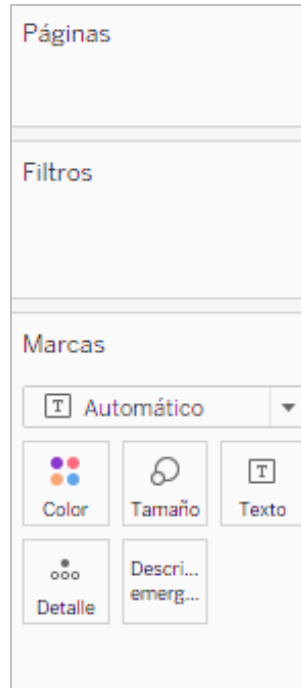


Figura 26. Cuadros y tarjetas. Fuente: elaboración propia

Por último, las **columnas y filas**. Estos espacios, **Figura 27**, están reservados para introducamos en ellos, arrastrando desde el espacio de variables y dimensiones, aquellos elementos con los que desee realizar el gráfico.



Figura 27. Columnas y filas. Fuente: elaboración propia

3.1.4.3 Visualizaciones

Como se ha comentado, el componente visual juega un papel clave en la estructura de funcionamiento de *Tableau* y es que la mayoría de las acciones se realizan no mediante escritura, sino mediante interacción modular con la interfaz mediante la acción arrastrar-soltar anteriormente descrita. Para generar los gráficos y **visualizaciones** basta con arrastrar hasta el área de dibujo las variables objeto de estudio, determinar mediante los cuadros los filtros colores y etiquetas, y por último hacer uso de una pestaña situada en la parte derecha de la interfaz denominada **Mostrar** que funciona como menú de acceso rápido a los tipos de visualización disponibles, **Figura 28**.



Figura 28. Pestaña Mostrarme. Fuente: elaboración propia

Existen un total de trece tipos de gráficos disponibles: barras, líneas, circular, mapa, diagrama de dispersión, gráfico de Gantt, gráfico de burbuja, gráfico de histograma, gráfico de bala, mapa de calor, tabla de resalto, diagrama de árbol, diagrama de caja y valores.

Además, estos gráficos pueden enriquecerse añadiendo líneas de tendencia/referencia, contando con: líneas de constante, promedio, mediana con cuartiles, diagrama de cajas, totales, líneas de referencias, bandas de referencias y bandas de distribución, **Figura 29 y Figura 30.**

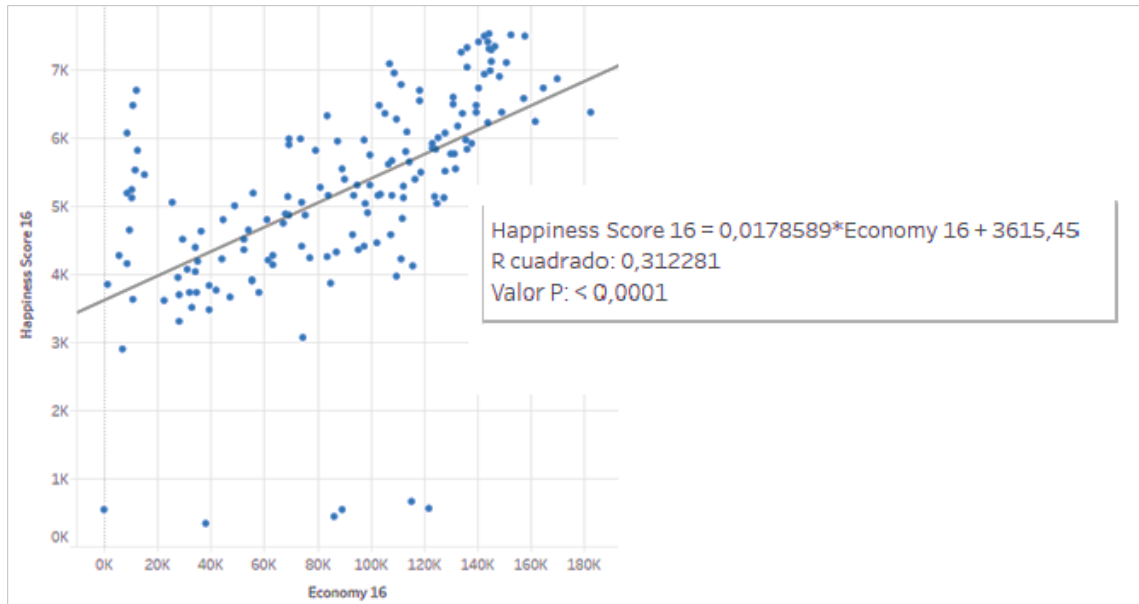


Figura 29. Línea de tendencia y P-value. Fuente: elaboración propia



Figura 30. Gráfico líneas de tendencia por regiones. Fuente: elaboración propia

Es posible la creación de **mapas**, mediante la introducción de variables del tipo: país, ciudad, código postal, latitud o longitud. Estos mapas nos servirán a su vez, mediante colores, para generar poderosos mapas de calor con los que analizar variables a nivel geográfico con gran facilidad.

Para poder realizar estas funciones el programa cuenta con una base de datos integrada que le permite reconocer los nombres o campos geográficos que se le arrojan mediante la fuente de datos suministrada. Si se diera el caso de que alguno de los parámetros introducidos le resultara desconocido, siendo posibles en estos casos su introducción manualmente. También permite la incorporación de mapas desde *Mapbox*¹⁴.

3.1.4.4 Creación de funciones y variables

Creación de nuevas variables: Una vez cargados los datos desde la fuente de datos seleccionado *Tableau* dispondrá para nosotros las variables, dimensiones que conforman dicho *dataset*. Sin embargo, es muy probable que a efectos de realizar nuestros análisis necesitemos crear nuevas variables. Esta es una posibilidad contemplada por la aplicación y que se resuelve a partir de un sistema para creación de variables a partir de cálculos y fórmulas personalizadas, mediante el uso del **crear campo calculado**. Este menú nos dará acceso a una serie de funciones que nos permitirá crear nuestra nueva variable. Este punto parece el menos intuitivo de todos, y es que según el tipo de variable que deseemos crear, y aún más dependiendo de la disposición en que se encuentre nuestra fuente de datos de origen, puede ser necesario contar con ciertos conocimientos (básicos) de programación y es aquí donde más flaquea el programa para un usuario no-técnico.

En la siguiente **Figura 31**, se muestra el menú para la creación de nuevas variables con un ejemplo real en el que resulta necesario saber cómo solventar el tratamiento de los parámetros nulos.

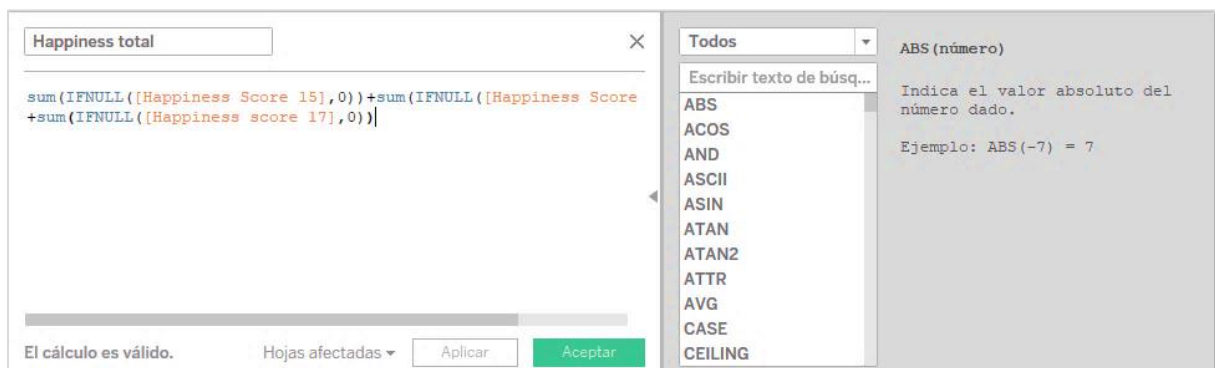


Figura 31. Creación de nuevas variables. Fuente: elaboración propia

¹⁴ Mapbox es un proveedor de mapas on-line realizados por peticiones personalizadas.

3.1.4.5 Predicciones, clúster y R

Tableau permite, a partir de una serie de datos anteriores, en el que debemos contar como mínimo con un campo de tipo fecha y una medida, realizar pronósticos o **predicciones** a futuro. Cuenta con distintas opciones para la realización de dichos pronósticos, como determinar el periodo temporal que se pretende pronosticar, así como los datos que se desean tener en cuenta para hacer dicha predicción.

Por su parte, es posible realizar clúster con mucha facilidad, tan solo utilizando el filtro correspondiente sobre nuestro gráfico con los datos dispuestos, obtendríamos algo como lo que figura en la **Figura 32**.



Figura 32. Clúster con tendencias. Fuente: elaboración propia

Para la realización de **análisis estadísticos** más avanzados *Tableau* permite integración con otras herramientas, como por ejemplo *R*, *Weka* y *Python*. Sin embargo, reflejar que lo único que hace la aplicación es ejecutar los scripts previamente realizados en el servicio de *R*, no permitiendo la edición en la propia aplicación.

La herramienta cuenta con un sistema de integración nativo para **R**. Tanto la versión desktop como la versión servidor de *Tableau* funcionarán como intermediadores entre las fuentes de datos y el

servidor de R, al que lanzarán los datos y recibirán los resultados, como se observa en el esquema de la **Figura 33**, para posteriormente suministrarlo a los usuarios de la herramienta. Con esta característica se amplía exponencialmente la capacidad analítica del usuario.

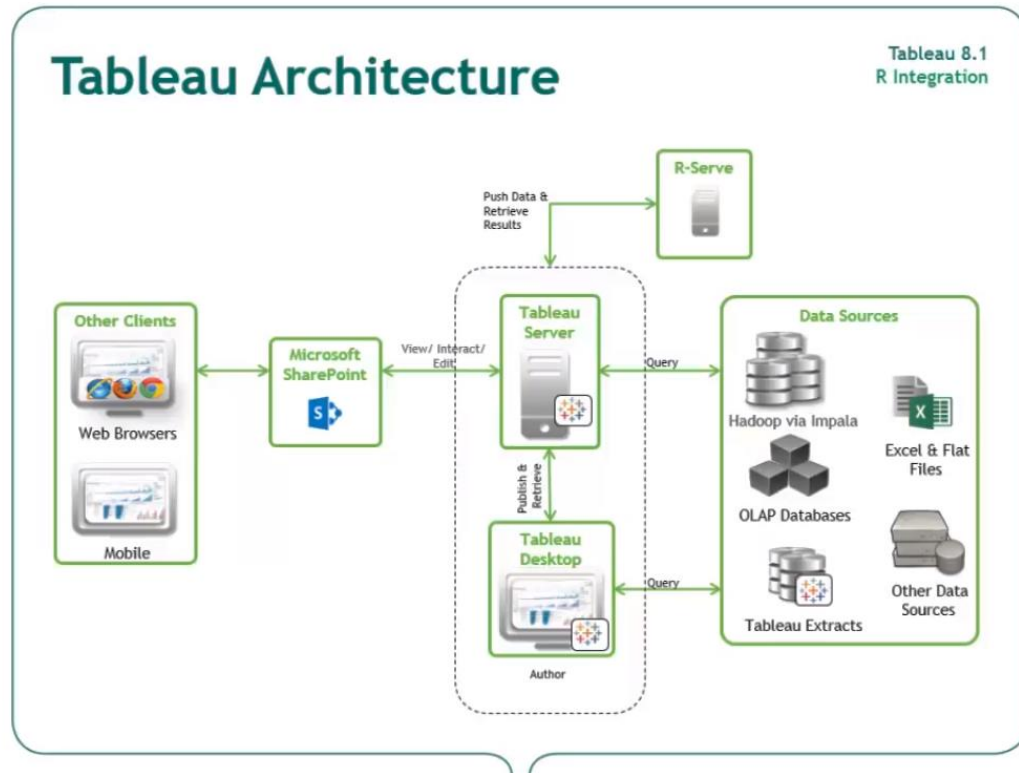


Figura 33. Esquema integración Tableau y R. Fuente: www.thorogood.com

3.1.4.6 Exportación de datos

La **exportación de datos** es también muy sencilla, pues en la mayoría de los casos (gráficos, tablas...) tan solo es necesario hacer clic derecho sobre el mismo, acceder al menú desplegable y seleccionar la opción de "exportar". La exportación de datos está permitida para gráficos, dashboards, así como tablas de valores en formatos csv, xls yxlsx. Se permite también la colaboración con otros usuarios para los dashboards.

3.1.5 Seguridad, volumetría y Big Data

Tableau dispone de sistemas para proporcionar la **seguridad** de las organizaciones que hagan uso de su herramienta. Cuenta con un sistema para administrar autenticaciones, autorizaciones, seguridad de datos y seguridad de red.

En primer lugar, cuenta con un sistema de **autenticación** para establecer la seguridad a nivel usuario, pudiendo generar también roles de usuario para facilitar la tarea. Esta autenticación se puede realizar mediante:

- Active Directory de Microsoft (SSPI/NTLM y Kerberos)
- SAML
- Autenticación confinable
- OAuth
- Autenticación nativa mediante *Tableau Server*

El sistema de **autorizaciones** permite delimitar la capacidad de acceso de cada usuario a la información. Estas autorizaciones controlan mediante dos sistemas:

- Funciones y permisos
- Licenciamiento y derechos de usuario

A nivel de datos la herramienta cuenta con sistemas, por un lado, para proteger los datos a nivel de servidor y, por otro, sistemas para garantizar la seguridad en la transmisión de red. Para esto pueden configurarse sistemas de HTTPS/SSL para controlar las solicitudes, utilización de controladores nativos para las bases de datos etc.

En el apartado de **volumetría**, la plataforma *Tableau* permite el acceso a plataformas de big data, con más de 40 fuentes de datos diferentes disponibles, muchas de ellas de este tipo. Encontramos conexiones disponibles a Hadoop (Cloudera IMplala, Hive, Amazon EMR), NoSQL, SPark, Amazon Redshift, Google BigQuery, Splunk. Para todas estas fuentes cuenta con un servicio de visualización con el que puede acceder a estas bases de datos de gran tamaño, apoyándose en su sistema de extracciones para agilizar los procesos. Sin embargo, no puede ir más allá de esto, no contando con procesos propios de programación distribuida en múltiples nodos para **Big Data**, no siendo posible por tanto ejecutar tareas de este estilo con esta herramienta.

3.2 Qlik

3.2.1 Introducción y versiones

Qlik, **Figura 34**, es el nombre de una de las soluciones de inteligencia de negocio que goza en la actualidad con mayor aceptación y reconocimiento. Su principal valor, dicen sus desarrolladores, la gran autonomía que otorga al usuario desde el primer momento en que comienza a trabajar.



Figura 34. Logo Qlik. Fuente: www.qlik.com

3.2.1.1 Qlik Sense

Qlik cuenta con un amplio abanico de soluciones, **Figura 35**, destinado a satisfacer necesidades de distintos tipos de usuario. Estas soluciones se agrupan en tres familias, **Sense**, **Analytics** y **View**, además de dos complementos denominados **NPrinting** y **GeoAnalytics**.

Qlik Sense es la herramienta destinada a la exploración de datos y realización de procesos de minería de datos, para facilitar la toma de decisiones. Se caracteriza por su facilidad de uso, no siendo necesario contar con conocimientos técnicos o específicos para hacer uso de la misma. Cuenta a su vez con tres modalidades.

La primera, **Qlik Sense Desktop**, es la aplicación de escritorio, para realizar visualizaciones de datos, informes y cuadros de mando interactivos totalmente configurables, **Figura 35**. Permite múltiples fuentes de datos y se basa en el concepto de arrastrar-soltar para la elaboración de los mismos. Se encuentra disponible en exclusiva para sistemas operativos Windows. Se trata de un producto gratuito, para uso personal e interno en la empresa.

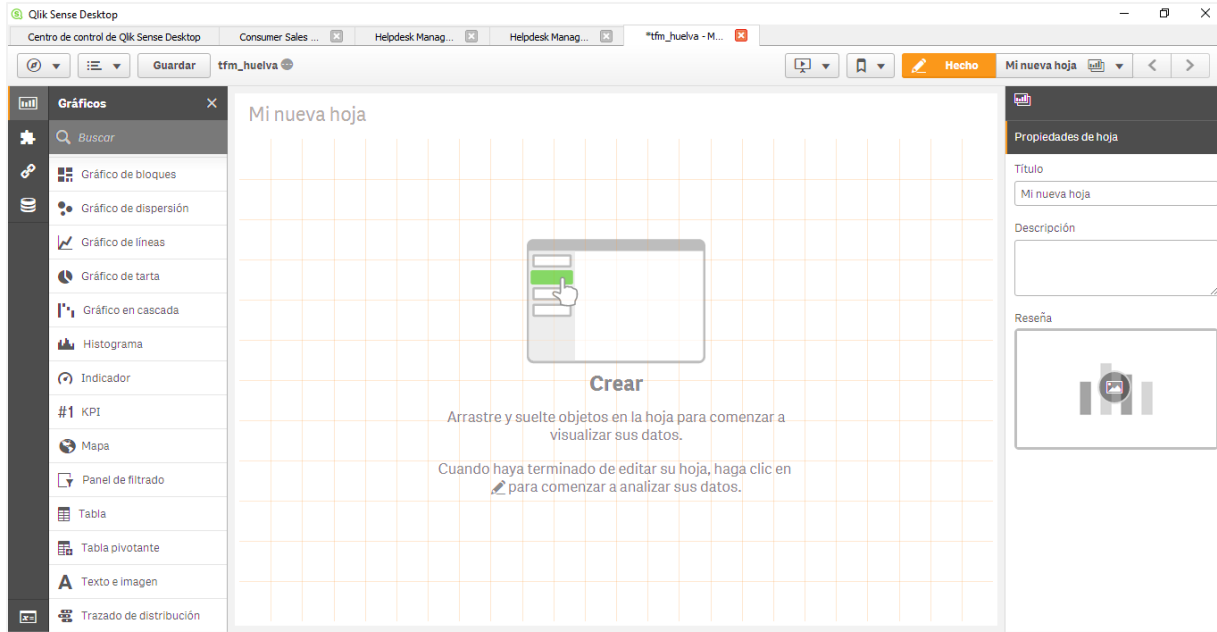


Figura 35. Interfaz Qlik Sense Desktop. Fuente: elaboración propia

Por contra las limitaciones que tiene esta versión, y es que las aplicaciones generadas no se pueden publicar (aunque sí compartir mediante la nube), no admiten *stream*¹⁵, no tienen sistema de seguridad o autenticación, ni tampoco autoguardado.

Con **Qlik Sense Cloud** encontramos otra aplicación para uso personal y que cuenta con todas las funcionalidades de la versión anterior, a la cuál añade la posibilidad de realizar visualizaciones incrustadas y de colaboración, haciendo para ello uso de la nube, amén de las ventajas inherentes al uso de una herramienta de estas características (mayor movilidad etc.), **Figura 36**.



Figura 36. Logo Qlik Sense Cloud. Fuente: www.qlik.com

¹⁵ Reproducción o uso justo mientras se produce la descarga.

La versión para empresas de la herramienta Qlik, **Qlik Sense Enterprise**, cuenta con todas las funcionalidades de las versiones de escritorio y nube, a las que suma otras tantas específicamente diseñadas para el entorno empresarial. Cuenta, por ejemplo, con un sistema para gestión de licencias flexibles, sistema para supervisión de uso, motor de reglas de seguridad y sistemas de auditoría y registro. Como elementos añadidos respecto de la anterior cuenta además con librerías de visualización compartidas, exploración y acceso desde dispositivos móviles y acceso a APIs y kits de herramientas.

3.2.1.2 Qlik Analytics

Por su parte, **Qlik Analytics Platform** es una versión que consiste en un paquete de herramientas que permite tanto a desarrolladores como socios *OEM*¹⁶ la inclusión de las capacidades analíticas y visuales de Qlik en sus propias herramientas y webs.

3.2.1.3 Qlik View

Se trata de la herramienta original desarrollada por la compañía Qlik. Comparte con la primera el ser una interfaz de usuario, **Figura 37**, que utilizan un mismo motor de análisis (compartiendo por tanto scripts y fórmulas). Esta aplicación tiene como principio dotar al usuario de un medio para utilizar aplicaciones empresariales previamente diseñadas, lo que Qlik denomina *Analítica Guiada*.

Se podría concluir que es una herramienta enfocada aún más a usuarios con escaso conocimiento técnico que busca una serie de plantillas preconfiguradas que le sirvan para encarar determinados problemas, pudiendo usar estas para interactuar, analizar y obtener conclusiones, pero que no busca en ningún caso desarrollar sus propias aplicaciones.

¹⁶ “Un fabricante de equipos originales (OEM) fabrica piezas o componentes que se utilizan en los productos de otra empresa.” (Hewlett Packard Enterprise, 2017)

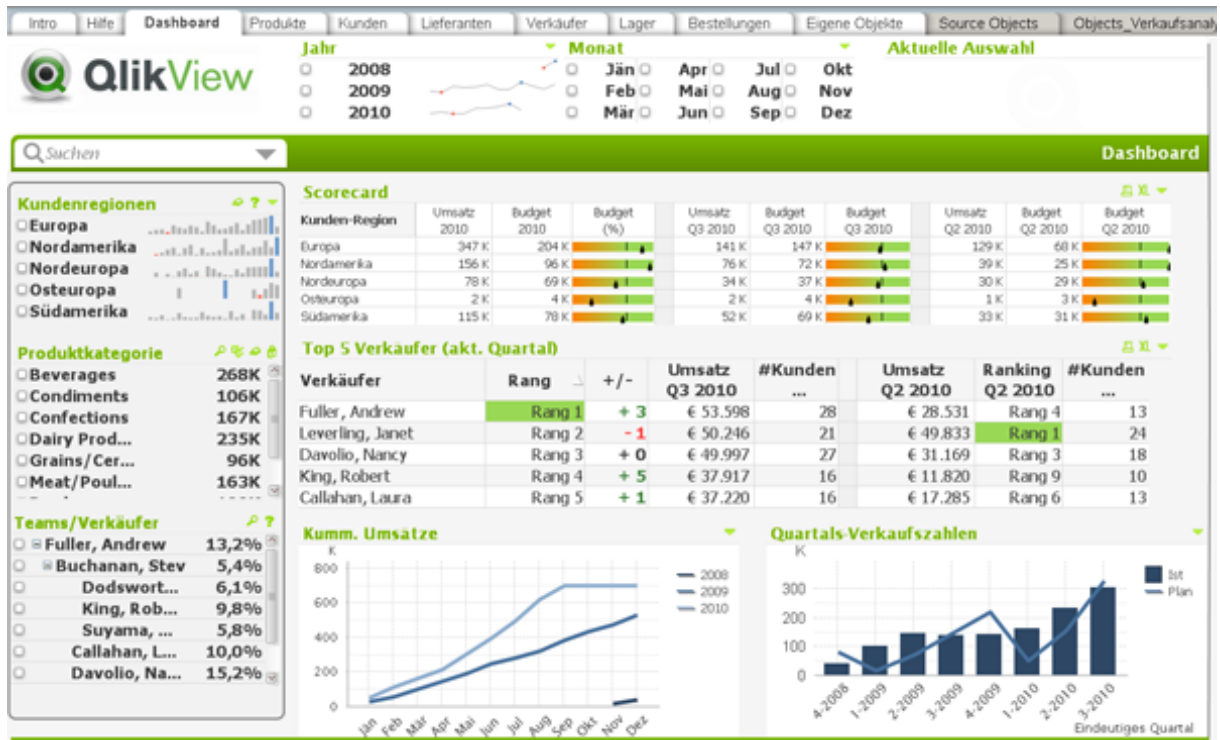


Figura 37. Interfaz Qlik View. Fuente: www.qlikblog.com

3.2.1.4 Qlik NPrinting

Este paquete supone un complemento de funciones extra para las versiones anteriores. *Qlik Nprinting* surge de la compra de Qlik a la compañía Vizubi, especializada en la generación de informes interactivos. Con esta herramienta se podrá, a partir de los datos trabajados en las versiones *View* y *Sense*, desarrollar informes interactivos en múltiples formatos del tipo PDF, Excel, Word, PowerPoint y HTML, **Figura 38**.

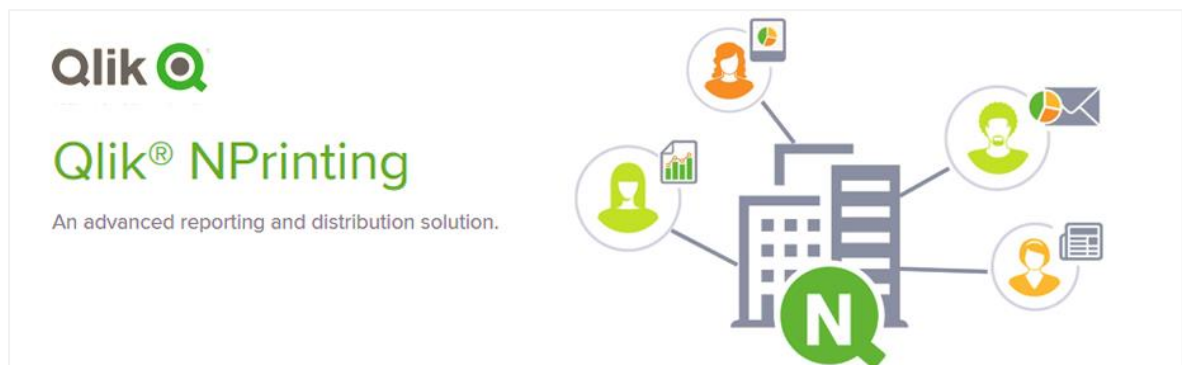


Figura 38. Logo Qlik NPrinting. Fuente: www.qlik.com

3.2.1.5 Qlik Geonalytics

GeoAnalytics es la solución propuesta por la compañía para integrar, en las versiones de *Sense* y *View*, soluciones de visualización en mapas muy potentes, **Figura 39**.



Figura 39. Logo Qlik GeoAnalytics. Fuente: www.qlik.com

Entre las principales funciones, destaca la posibilidad de realizar mapeo sin coordenadas, por lo que, tan solo introduciendo nombre del país, estado, ciudad o área postal, el sistema conectará automáticamente el punto en el lugar preciso del mapa.

El nivel de detalle alcanzado es muy profundo, pudiendo trabajar con mapas a nivel país, pero pudiendo profundizar si así se requiere a nivel de región o ciudad. Además, amplía el abanico de visualizaciones disponibles, con **burbuja**s, **líneas**, **áreas**, **gráficos circulares**, **gráficos de barras** y **mapas de calor**.

3.2.2 Arquitectura

La arquitectura de la herramienta sigue la estructura de cliente-servidor, pudiendo desgranarse en tres niveles, *QlickView Desktop*, *QlickView Server* y *QlickView Publisher*, siguiendo para su funcionamiento el esquema mostrado en la **Figura 40**.

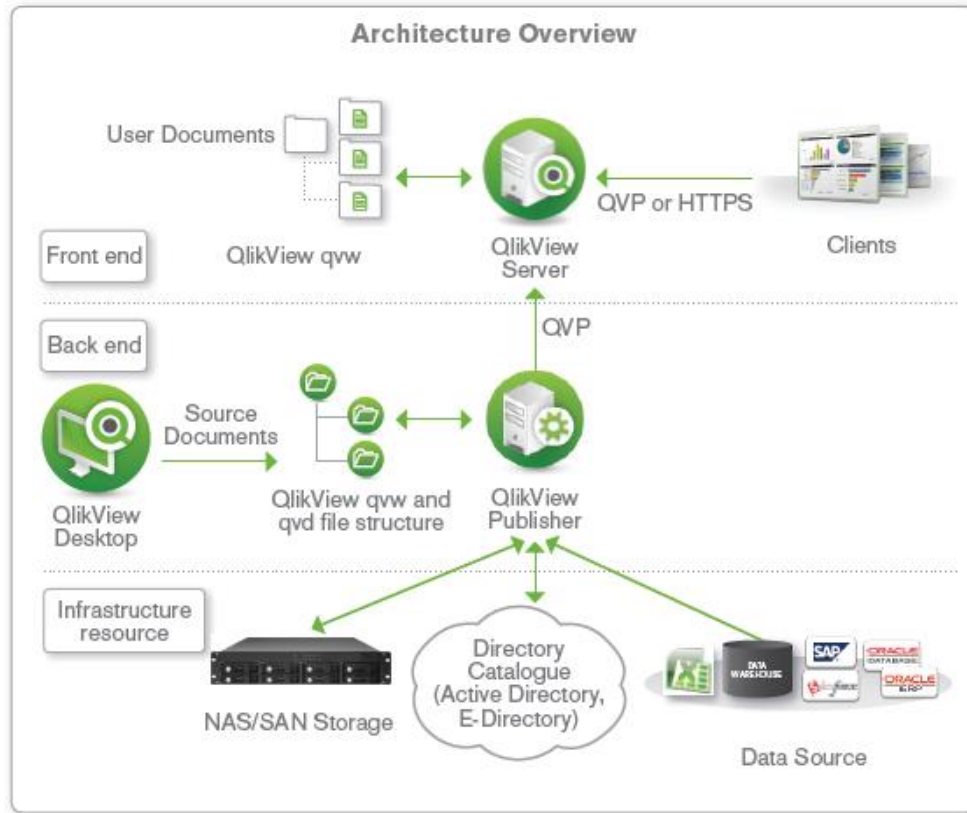


Figura 40. Arquitectura Qlik. Fuente: www.globustrainings.com

El **Developer** se encargará de realizar la extracción y transformación de los datos, ofreciendo también la interfaz gráfica (el cuadro de mandos). Por su parte **Server** es el punto de acceso para el usuario vía web, así como el motor analítico y el punto de intermediación entre el servicio de servidor y el cliente. Por último, el **Publisher** tiene como función clave la distribución de los documentos. Sobre este último punto mencionar la posibilidad de añadir servidores adicionales sobre el primero, trabajando en forma de clúster. Esta función es exclusiva para las funciones de Publisher y permiten modificar múltiples versiones del documento a la vez.

El sistema permite ejecutar programas y aplicaciones independientes en cada uno de los nodos pudiendo así distribuir mejor la carga de trabajo, mejorar los procesos y optimizar el uso de la memoria. Por su parte, el uso de estos sistemas en clúster supone una medida adicional de seguridad en términos de funcionamiento, y es que, de fallar un servidor, los restos asumirían su carga de trabajo garantizando el mantenimiento del servicio.

3.2.3 Aspectos técnicos

3.2.3.1 Arrastrar y soltar

El hilo conductor de trabajo con la herramienta *Qlik* es *arrastrar y soltar*, como sucede con la mayoría de las herramientas existentes en mercado. Sin embargo, tiene también cierta relevancia el uso de los menús y submenús, especialmente a la hora de configurar los gráficos realizados, para la elección por ejemplo de variables y dimensiones. Bajo esta filosofía trabajar con la aplicación resulta cómodo y rápido, siendo muy natural su uso, y pudiendo conseguir buenos resultados con tan solo algunas acciones, **Figura 41**.

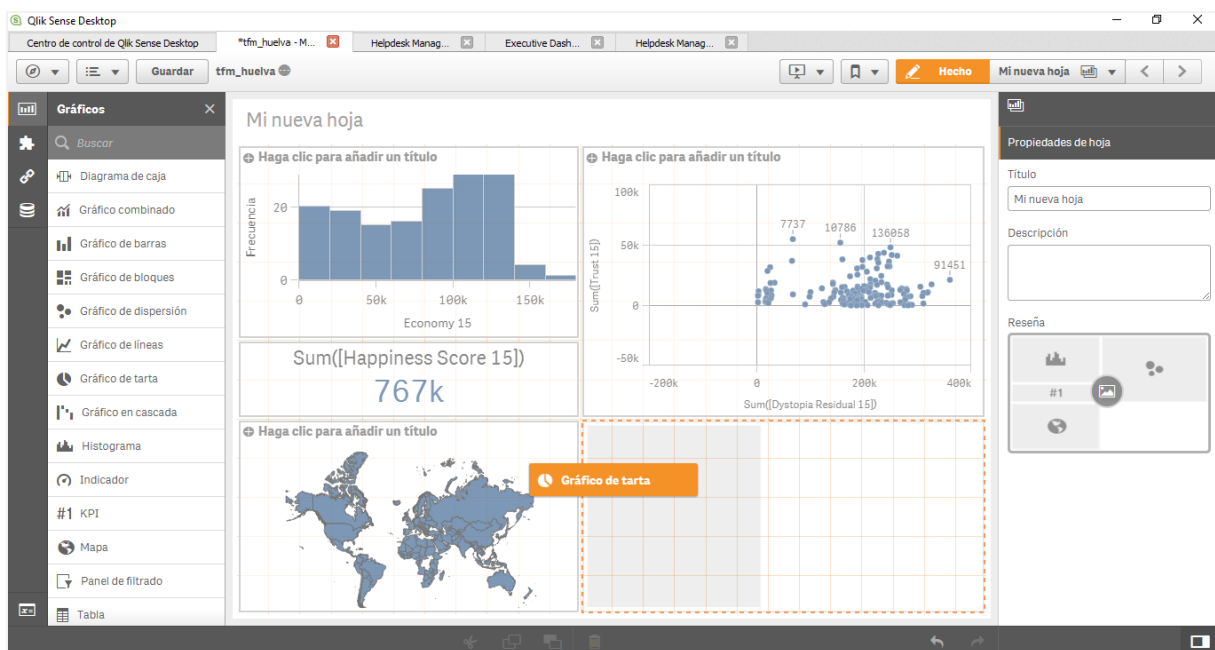


Figura 41. Interfaz Qlik Desktop. Fuente: elaboración propia

3.2.3.2 Gestión de la memoria

Qlik permite realizar ingestión y consolidación de información desde múltiples fuentes de datos. Sin embargo, la característica diferenciadora viene dada por la capacidad que tiene la herramienta para determinar qué cantidad de esa información deja en disco y qué cantidad en memoria, pudiendo ser alojada completamente en la segunda. En el caso de optar por la segunda opción contaríamos con una capacidad de procesamiento muy alta, algo que notaríamos especialmente al realizar procesos complejos.

Las circunstancias en las que será aconsejable el uso del modelo 100% en memoria serán aquellas en las que:

1. Todos los datos necesarios entran en la memoria del servidor
2. El usuario requiere de utilizar información agregada o resumida, con un determinado nivel de detalle, pero lo va a realizar solo para un limitado periodo de tiempo que permita manejarlo.

Será aconsejable un planteamiento *direct-discovery* mixto cuando:

1. Los datos no entran físicamente en la memoria
2. Las tablas requeridas por el usuario a nivel de registro son muy grandes
3. El ancho de banda disponible no permite la comunicación correcta con el server.

3.2.3.3 Asociatividad

Relacionado con el punto anterior se encuentra esta otra característica, que supone otro fuerte punto a favor de la misma. La **Asociatividad** hace referencia al motor de datos asociativo con el que cuenta *Qlik* y que, a diferencia de otras suites que utilizan una estructura fija ETL (extracción, transformación y carga), dispone de un proceso que le permite realizar una carga selectiva, haciendo que todos sus indicadores lo hagan automáticamente, no siendo necesario aplicar ningún tipo de jerarquía.

3.2.4 Funcionalidades de usuario

3.2.4.1 Fuentes y conexión a datos

Qlik permite la conexión a fuentes de datos de muy distinta índole, desde archivos, conectores, un *data-market* propio (que comentaremos posteriormente) y conectores especializados, **Figura 42**.

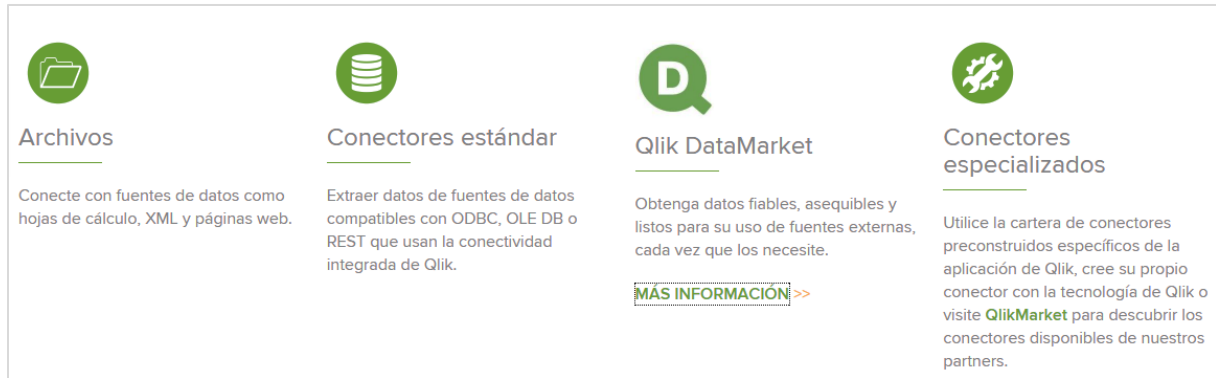


Figura 42. Fuentes de datos en Qlik. Fuente: www.qlik.com

Los archivos admitidos como fuentes de datos para Qlik quedan recogidos en la siguiente **Figura 43** de forma resumida.

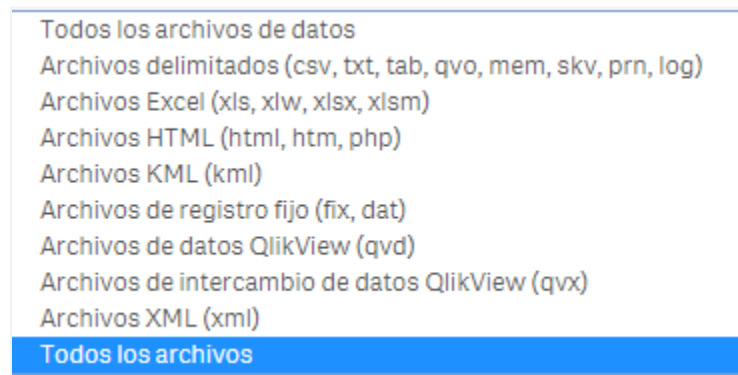


Figura 43. Archivos admitidos Qlik. Fuente: elaboración propia

Qlik dispone de múltiples opciones de acceso a datos, pudiendo combinarse varias para trabajar en un mismo proyecto:

- **Archivo:** Excel, csv, xml, texto etc.
- **Bases de datos:** Oracle, SQL Server, MySQL, PostgreSQL, Teradata, IBM DB2 etc.
- **Librerías SAS:** mediante conexión de tipo ODCB
- **Nube & Big Data:** Hadoop, Spark, Microsoft Native y opcionalmente mediante conectores externos disponibles en el Qlik Market para Hive, Hbase, Cassandra, Amazon Redshift, Neo4J, MongoDB, HP Vertica, Cloudera...
- **Cubos OLAP**
- **Otros:** páginas webs, redes sociales, repositorios.

3.2.4.2 Interfaz

La interfaz de Qlik resulta atractiva y agradable, aunque en comparación alguna de la competencia puede resultar más compleja visualmente. Esto es debido a que cuenta, a simple vista, con un mayor número de botones y opciones que hacen que el usuario primerizo pueda sentirse abrumado en un primer momento de contacto con la aplicación. La estructura general de la interfaz en lo que a creación de análisis y gráficos se refiere está compuesta por el **centro de control, apps, hojas y menús interactivos**.

El **centro de control, Figura 44**, supone el menú principal de *Qlik Desktop*. En él veremos todos los proyectos, en los que estamos trabajando, denominados Apps.

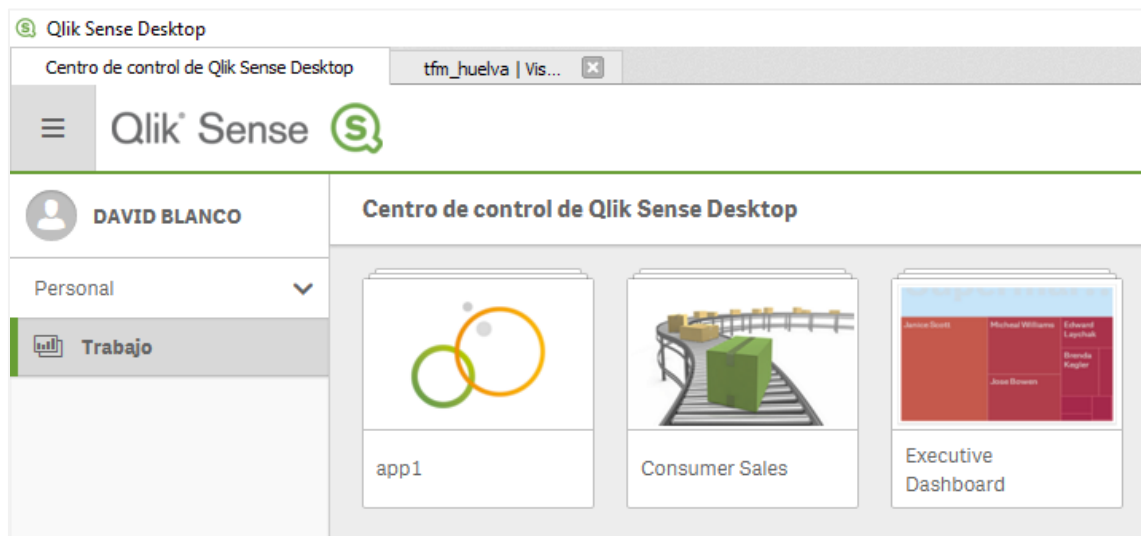


Figura 44. Centro de control. Fuente: elaboración propia

Las **Apps** vienen a ser los proyectos generales, por ejemplo, si estamos analizando la evolución del departamento financiero en un período dado, generaremos una App con este nombre en la cual integraremos todos los gráficos y análisis relacionados con este tema. Estos elementos se desarrollarán en lo denominado como *Hojas*.

Las **hojas, Figura 45**, funcionarán como nuestro tablero de trabajo, el espacio en el que realizaremos la parte analítica y gráfica. Desde ellas accederemos a los datos, previamente cargados, y dispondremos de los menús con los que se interactúa para generar las visualizaciones.

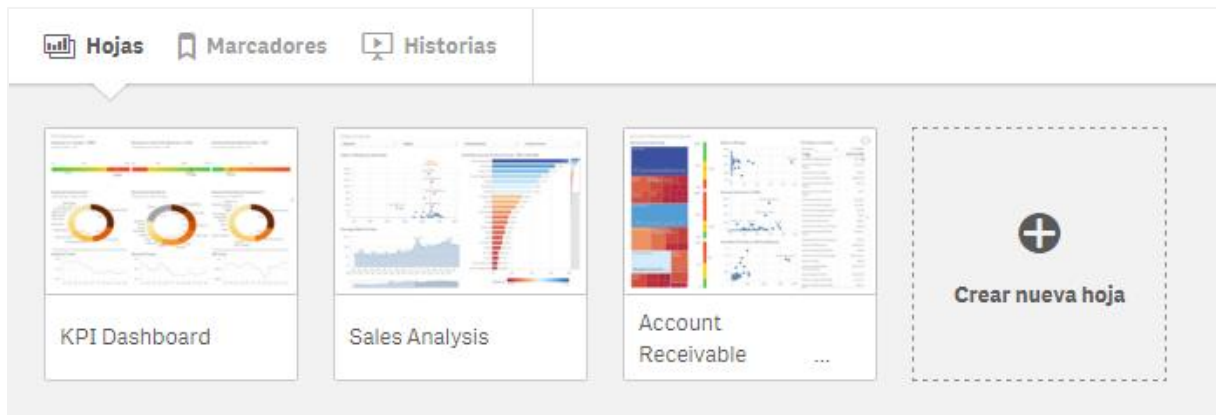


Figura 45. Hojas en Qlik. Fuente: elaboración propia

En este caso por ejemplo tendríamos 3 hojas, la primera de ellas un dashboard con los KPI de la empresa, una segunda con un análisis de ventas y el tercero una *cuenta de cobros*¹⁷. Con estos tres elementos tendríamos una *app* contable/financiera completa.

Una vez entramos en alguna de estas hojas nos encontramos con un esquema parecido al de la herramienta *Tableau*, aunque con diferencias. En la parte superior un menú interactivo de pestañas que nos permitirá tener varias *apps* abiertas a la vez y pasar de una a otra con agilidad, algo que con lo que no cuenta *Tableau*, **Figura 46**.

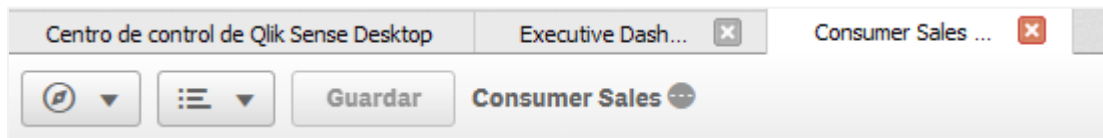


Figura 46. Pestañas. Fuente: elaboración propia

Justo debajo otro menú contextual con botones que nos darán acceso a muchas opciones útiles, **Figura 47**.

¹⁷ Documento contable opcional para que aquellas personas que se encuentren en régimen de tributación simplificado y por ende no obligadas a la expedición de facturas de IVA puedan hacerlo si lo desean.

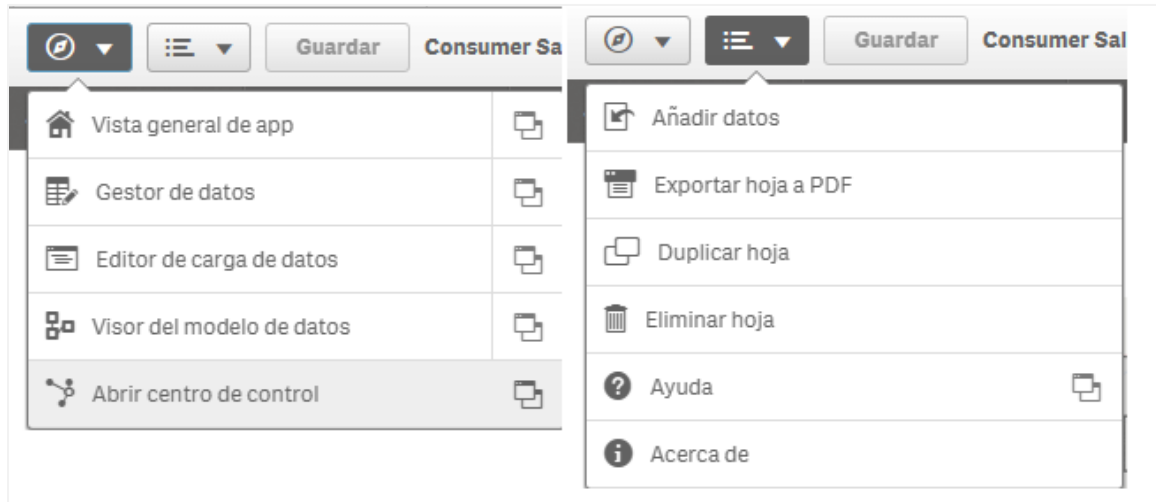


Figura 47. Botonera interactiva Qlik. Fuente: elaboración propia

El primero de ellos denominado **Navegación** con funciones relacionadas con los ingesta y edición de datos y el segundo llamado simplemente **Menú** nos otorga acceso a algunas opciones básicas para las propias hojas.

Encontramos también un botón extra que nos ofrece una pequeña funcionalidad y es poder contar con un pequeño espacio donde alojar algo de información sobre la hoja con la que estamos trabajando, la ruta en la que se encuentra alojada y la posibilidad de personalizarlo mediante el uso de un logo y un color. Esto puede resultar especialmente útil cuando trabajamos a nivel colectivo o con varias hojas y queremos identificar fácilmente el proyecto en el que nos encontramos, **Figura 48.**

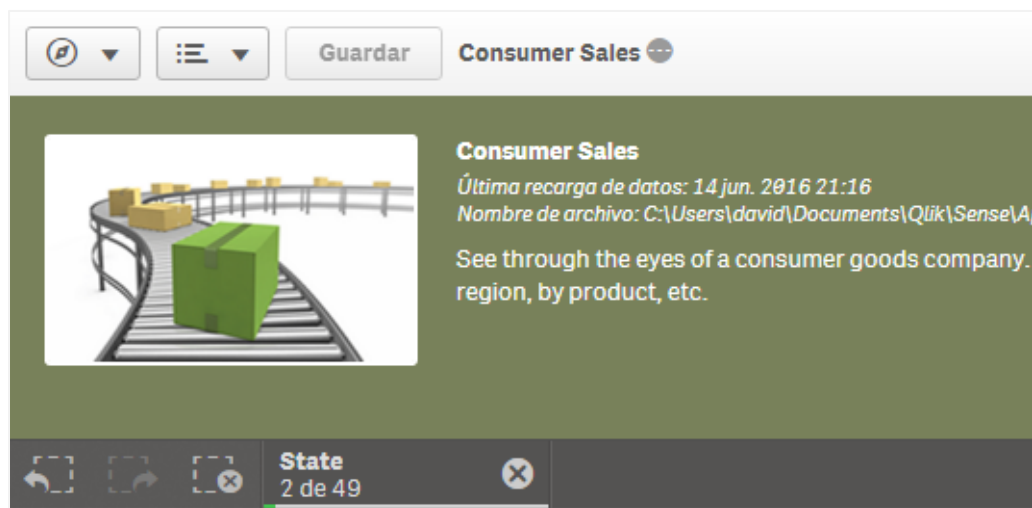


Figura 48. Información de hoja. Fuente: elaboración propia

En la última parte de esta barra de botones encontramos las siguientes funcionalidades, **Figura 49**.

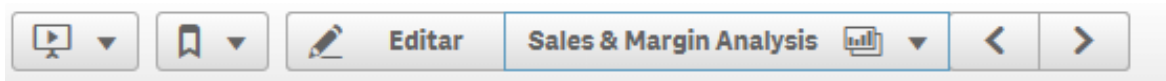


Figura 49. Botonera. Fuente: elaboración propia

De izquierda a derecha encontramos un primer botón, denominado **Historias**. Esta opción nos permite generar una presentación interactiva con las hojas de las aplicaciones que hemos desarrollado. **Marcadores** nos permitirá marcar para facilitar la identificación de aquellos puntos de especial interés. **Editar** nos dará acceso a las funciones de creación y edición de los gráficos. Una vez ejecutado aparecerá un nuevo menú lateral en la parte izquierda que nos permitirá acceder a **Gráficos, Objetos Personalizados, Elementos Maestros y Campos**.

En esta herramienta no elegimos las variables y dimensiones con las que queremos trabajar antes de generar los gráficos, sino que serán los gráficos, una vez que los hemos situado sobre el espacio de trabajo, los que nos pedirán introducir las dimensiones que queremos analizar, **Figura 50**.



Figura 50. Menú dimensión y medida sobre gráfico. Fuente: elaboración propia

Esta forma de trabajo resulta a priori más lenta y menos visual para el usuario que otras alternativas, por tener que navegar por varios menús para hacer una selección que podría realizarse con otra función arrastrar. Se observa también la presencia de un nuevo menú lateral que aparece al empezar a trabajar con un gráfico, **Figura 51**.

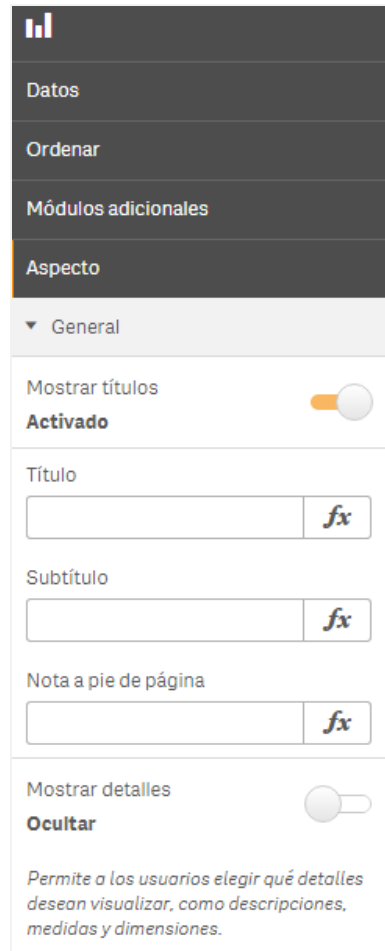


Figura 51. Menú lateral Qlik. Fuente: elaboración propia

3.2.4.3 Visualizaciones

Las opciones de visualización disponibles en la versión desktop son múltiples, como se observa en la **Figura 52**.

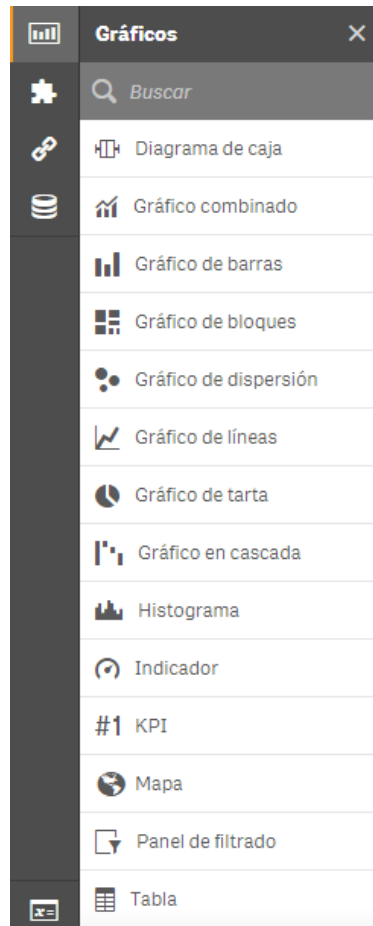


Figura 52. Menú visualizaciones Qlik. Fuente: elaboración propia

Tanto en las versiones **View** como **Sense** se encuentran disponibles diferentes tipos de **gráficos, de la caja, combinado, barras, etiqueta, mapas, KPI etc.** La metodología requerida para su utilización es la anteriormente indicada, arrastrar y soltar al área de trabajo, para posteriormente indicar dentro del propio gráfico las medidas y dimensiones. El proceso, respecto al mismo proceso con otras herramientas es menos intuitivo. En línea general hay una sensación de menor cuidado con el aspecto gráfico y la asistencia al usuario que con otras opciones. Por su parte, además, encontramos que solo se encuentra en la versión por defecto de **QlikView** la posibilidad de añadir líneas de tendencia sobre los gráficos originales. Algunos de los gráficos realizables se pueden observar en las siguientes **Figura 53 y Figura 54.**

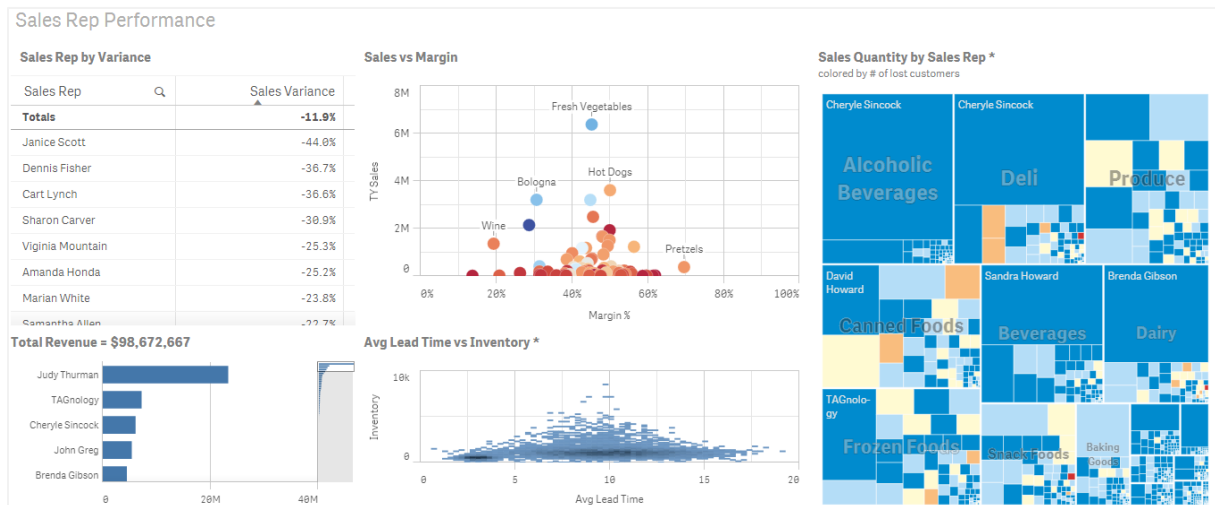


Figura 53. Muestra gráficos Qlik 1. Fuente: elaboración propia

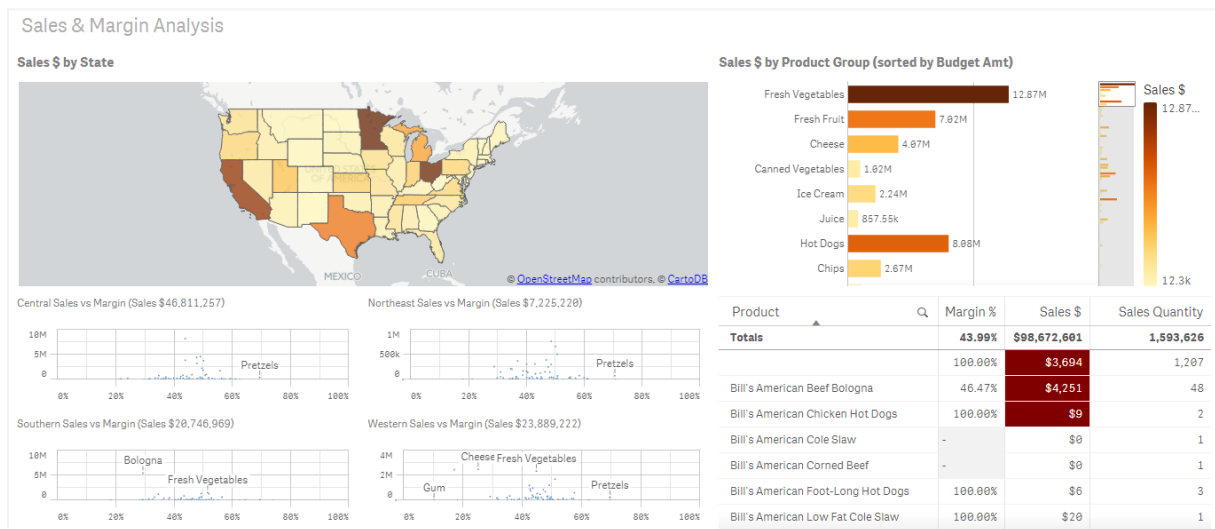


Figura 54. Muestra gráficos Qlik 2. Fuente: elaboración propia

3.2.4.4 Creación de funciones y variables.

Una vez creadas nuestras *apps* el programa nos pedirá cargar los datos con los que deseamos trabajar. Una vez cargados este desglosará la información contenida, generando las variables/dimensiones. Es probable que en muchas ocasiones necesitemos crear nuevas variables a partir de estas. El procedimiento para hacerlo resulta, una vez más, de mayor complejidad que con otras herramientas, y es que para poder crear una nueva columna tendremos en primer lugar que dirigirnos al menú de navegación y cargar el gestor de datos, desplazarnos hasta edición de tabla. Posteriormente y una vez allí, desplazarnos a la opción “añadir campo” y “campo calculado”.

Sin embargo, encontramos como puntos positivos que el menú que nos permite la creación de variables es más apropiado para usuarios *legos* puesto que no necesita que este escriba la fórmula para la creación si no quiere hacerlo, pudiendo guiarse por los menús interactivos de los que dispone, **Figura 55**.

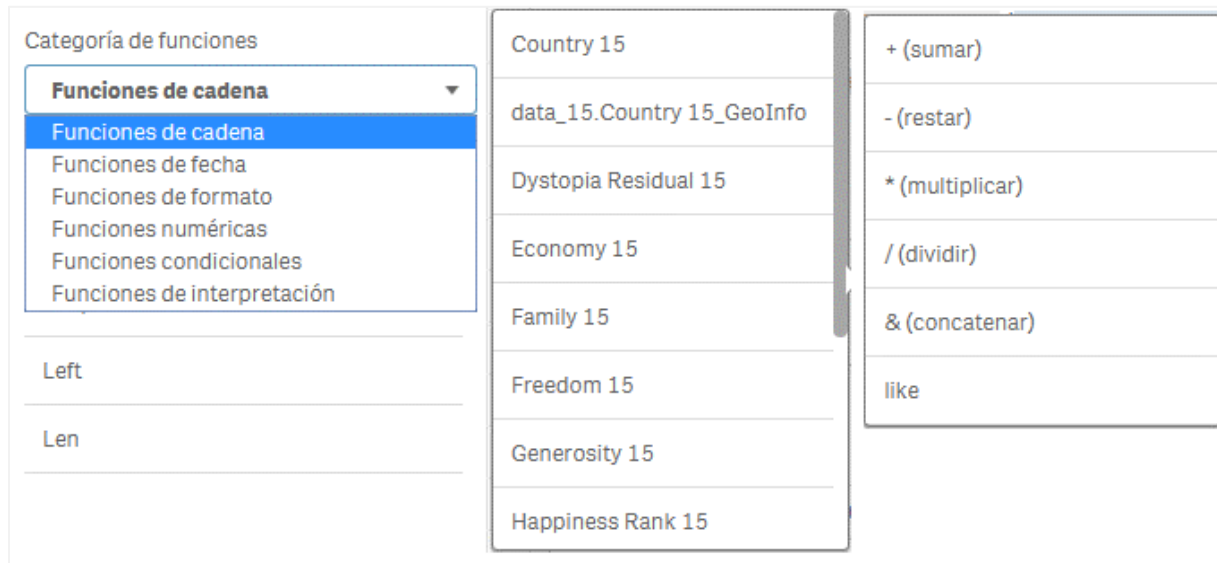


Figura 55. Menú creación variables. Fuente: elaboración propia

Por otro lado, además, dispone de una **previsualización de la variable** que se va a generar que nos permite comprobar, antes de ejecutarla, si hemos generado lo que necesitábamos, **Figura 56**.

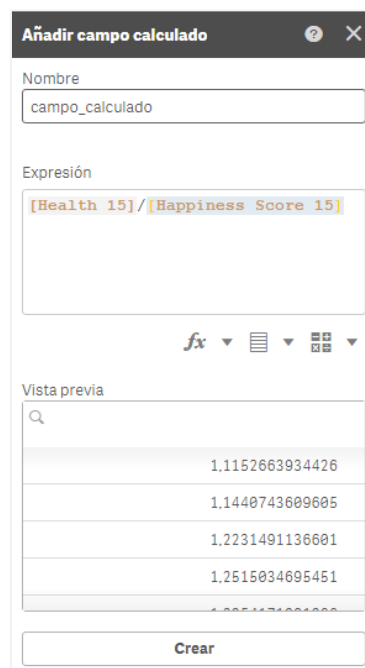


Figura 56. Añadir campo calculado. Fuente: elaboración propia

3.2.4.5 Predicciones, clúster y R

La versión de escritorio, que es en la que nos centramos en la comparativa, tiene la capacidad de realizar predicciones (forecasting), algo que no permite hacer la herramienta en su versión *View*. Se trata de una tarea no accesible para usuarios sin conocimientos técnicos, pues no dispone de un módulo o plantilla predefinida con la que hacerla, sino que será necesario que escribamos una secuencia con la que lo hagamos. La misma situación encontramos para la realización de clústeres.

En lo que respecta a la integración con R encontramos que Qlik no dispone de una integración nativa. Tienen un proyecto aún en fase de desarrollo para integrarlo en el futuro. Si es cierto que a través de la comunidad de usuarios encontramos opciones para realizar esta tarea a través de capas creadas por terceros que permiten realizar la conexión. Los más populares para hacer esto son las plataformas *ezdatamunch* y *sistel*, **Figura 57**.



Figura 57. Logos ezdatamunch y sistel. Fuente: www.ezdatamunch.com y www.sistel.com

3.2.4.6 Exportación de datos

La exportación se realiza de forma muy fácil, pudiendo hacerse para gráficos y dashboards en formatos PNG y PNG, así como los valores en los formatos csv, xls, xlsx. Permite también exportar el origen y fuente de datos, para que se puedan cargar fácilmente.

3.2.4.7 Datamarket

Contar con buena y suficiente información es un elemento indispensable para poder realizar estudios fiables que nos lleven a conclusiones útiles y veraces. *Qlik* ha desarrollado un espacio, denominado como *Qlik Datamarket*, que como su nombre indica nos permitirá adquirir de forma sencilla datos externos con los que enriquecer nuestros datos internos, **Figura 58**.



Figura 58. Qlik DataMarket. Fuente: www.qlik.com

A través del menú de **conexión a datos** encontramos las distintas opciones disponibles, y entre ellas destaca la opción del *Datamarket*, **Figura 59**.

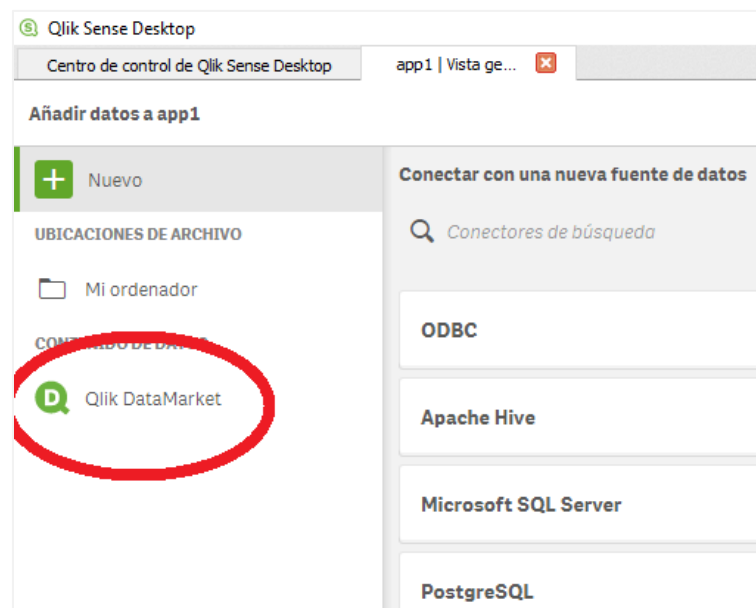


Figura 59. Captura Qlik DataMarket. Fuente: elaboración propia

Al entrar en el datamarket nos encontramos con una serie de categorías de datos a los que podemos tener acceso, siendo los siguientes:

1. **Datos esenciales:** cotizaciones, datos demográficos, sociedad, economía, climatología
2. **Datos esenciales (libres):** mismos que anterior, pero accesibles sin pago alguno
3. **Reportes financieros:** información financiera con distintas periodicidades
4. **Stocks e índices:** información histórica de stocks e índices de todo el mundo

5. **Cotización de monedas:** precios de cotización de más de 160 monedas de todo el mundo.
6. **World weather:** información meteorológica de todo el globo

Con tan solo unos clics podemos tener acceso a información, como por ejemplo en este caso, **Figura 60**, el tipo de cambio Euro/Dólar para el periodo 2010-2014. Esta información puede ser utilizada de forma conjunta con nuestros datos internos, enriqueciendo el análisis y por ende los resultados.

DateTime	Exchange rate
2010-04-02T23:00:00Z	1.35035
2010-04-05T23:00:00Z	1.3483
2010-04-06T23:00:00Z	1.3394
2010-04-07T23:00:00Z	1.3339
2010-04-08T23:00:00Z	1.3358

Figura 60. Captura tipos de cambio. Fuente: elaboración propia

3.2.5 Seguridad, volumetría y Big Data

La plataforma cuenta con medidas de seguridad, permitiendo la restricción de acceso, modificación y visualización. Esto se puede configurar, en este caso compartiendo a través del **Publisher**, pudiendo configurarse la adición/eliminación de hojas, elementos sueltos, copiar/clonar objetos, acceso a propiedades. Todo esto se realizará mediante un sistema de permisos, protegido a su vez con un usuario/contraseña.

Para facilitar esta tarea pueden generarse los denominados *roles de usuario o grupo* que permitirá establecer configuraciones predeterminadas para tipo de usuarios o grupos de usuario y asignar de forma automática esta configuración a los participantes de las apps.

La aplicación cuenta con un sistema para mejorar el rendimiento al **tratar volúmenes de datos grandes**, es a través de las denominadas *extracciones*. Esto consiste en tratar un subconjunto de los datos, con objeto de mejorar el rendimiento. Para ello se basan en concepto de *modelos de columna*. Este sistema de extracciones puede suponer una gran diferencia, logrando con ello trabajar con bases de datos comprimidas con ratios de hasta 8 a 1, por lo que, por ejemplo, se puede

llegar a trabajar con un dataset de 1TB utilizando para ello un equipo que cuente tan “solo” 125GB de memoria RAM. Esta capacidad puede explicar en buena medida por qué empresas de gran tamaño, que hacen uso de fuentes de datos muy grandes, hacen uso de ella.

Bajo el uso de *Direct Discovery* Qlik amplía sus capacidades, permitiéndole realizar la asociación de grandes conjuntos de datos con los que tenga almacenados en memoria. Esto se traduce en una mejora sustancial en los procesos de análisis de tipo asociativo, especialmente cuando tratamos con volúmenes de datos muy grandes. Pese a ello, no podemos obviar que nos encontramos ante una herramienta de *Business Intelligence*, por lo que pese a que podamos acceder a fuentes de datos denominadas como *Big Data* no podemos aplicar las técnicas propias de este tipo de tecnología.

Aunque si permite la conexión a fuentes de datos muy grandes, y pese a que hemos visto la capacidad de Qlik para trabajar en múltiples clústers para mejorar la capacidad de colaboración, no cuenta con esto mismo para realizar programación paralela en múltiples nodos, y por tanto ejecutar procesos propios de tipo *Big Data*.

3.3 PowerBI

PowerBi es el nombre otorgado por Microsoft, Ilustración X, a su suite de Business Intelligence.

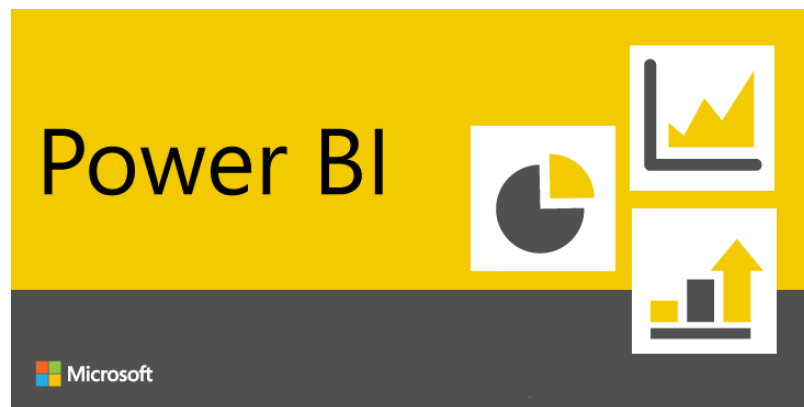


Figura 61. Logo PowerBI. Fuente: www.powerbi.microsoft.com

Se trata de un proyecto que tiene sus orígenes en los complementos de *BI* originalmente contenidos en *Microsoft Excel*, *PowerQuery*, *PowerPivot* y *Power View*. Estas herramientas fueron el germen de la suite, ofreciendo en su momento las funcionalidades que hoy encontramos agrupadas en PowerBi, amén de otras características que se han ampliado y añadido, tales como la

conectividad a nivel usuario/empresa y las opciones de seguridad. La aplicación sale al mercado en julio de 2015 y desde entonces ha ido desarrollándose, hasta ser en la actualidad una de las mejores utilidades existentes.

3.3.1 Introducción y versiones

3.3.1.1 PowerBI Desktop y Premium

Versión de escritorio de la herramienta, que permite la creación de gráficos e informes de manera fácil y dinámica. Cuenta además con un sistema que permite tanto capturar los datos desde la fuente que deseemos, como tratarlos y modificarlos. Su interfaz nos recuerda mucho a la suite de office, tanto por disposición de menús como por el estilo de los iconos y botones, **Figura 62**.

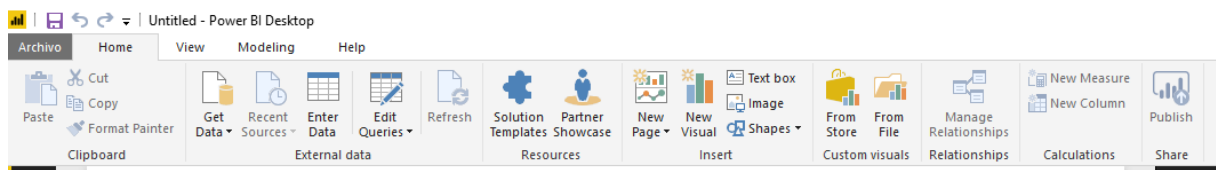


Figura 62. Estilo de interfaz PowerBI. Fuente: elaboración propia

Esta versión **Desktop** es totalmente gratuita, siendo necesario tan solo un pequeño registro en la web (información personal básica y correo electrónico, sin necesidad de verificación) los únicos requisitos para obtenerla. Cuenta con una serie de limitaciones que son cubiertas en la versión **Premium** del programa.

Conexión a más de 70 orígenes de datos	✓	✓
Publicación en Web	✓	✓
Uso compartido entre iguales	✗	✓
Exportación a PowerPoint, Excel y CSV	✓	✓
Distribución empresarial		
Aplicaciones	✗	✓
Suscripciones de correo electrónico	✗	✓
API y controles insertados	✗	✓
Colaboración		
Áreas de trabajo de la aplicación	✗	✓
Análisis en Excel y en Power BI Desktop	✗	✓

Figura 63. Limitaciones versión free y premium. Fuente: www.powerbi.microsoft.com

3.3.1.2 PowerBI Mobile

Versión móvil de la suite, que viene a suponer un medio para visualizar y compartir los informes previamente realizados. Está disponible para dispositivos *Microsoft*, *Android* y *iOS*. Resaltar que mantiene la interactividad de los informes generados, pudiendo seleccionar por ejemplo países en un mapa o visualizar etiquetas al hacer clic sobre gráficos.

En la *Play Store*, **Figura 64**, vemos cuenta con más de cien mil descargas, así como buena ratio de calificaciones.

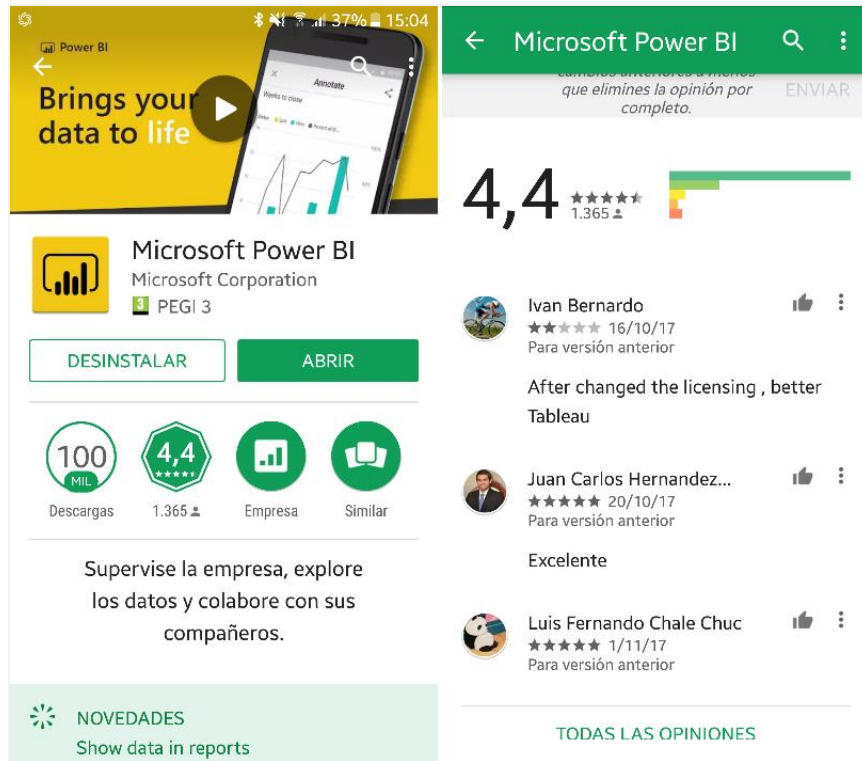


Figura 64. PowerBI Mobile Play Store. Fuente: elaboración propia

Posteriormente pasamos a probar la experiencia de uso, donde comprobamos la aplicación es un visor de informes previamente generado que funciona muy bien, siendo además muy intuitivo, con un diseño cuidado y agradable.

3.3.1.3 PowerBI Embedded

Se trata de una solución muy interesante propuesta por parte de *Microsoft*, que permite a empresas y organizaciones, interesadas en las funcionalidades y soluciones que aporta *PowerBi*, integrar las funciones de esta en sus propias aplicaciones. De esta forma, una empresa que cuente con su propio software de gestión puede incluir las funciones de la suite sin necesidad de usar ambas aplicaciones a la vez. De hecho, no resulta necesario estar suscrito a *PowerBi Services*, aunque si contar con una cuenta *Azure*¹⁸. Como observamos en la siguiente **Figura 65**, la demo demuestra cómo una compañía cualquiera podría incluir dentro de su propia plataforma la herramienta.

¹⁸ *Microsoft Azure* es una plataforma creada por *Microsoft* y alojada completamente en la nube, prestando distintos servicios, desde procesamiento de datos hasta infraestructura para comunicación segura.

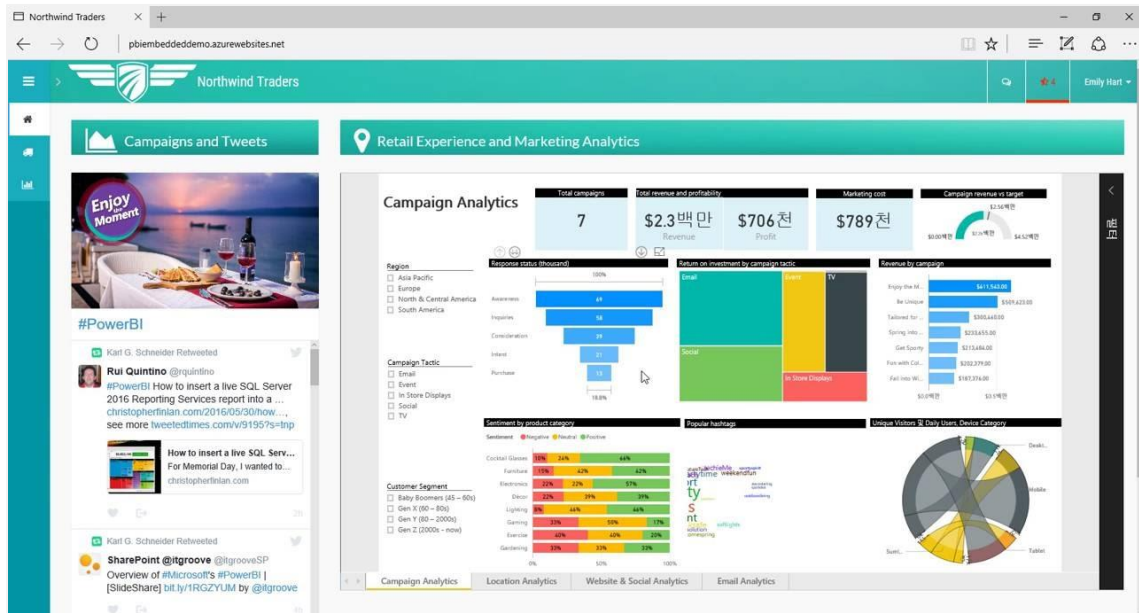


Figura 65. PowerBI Embedded. Fuente: www.powerbi.microsoft.com

3.3.1.4 Servidor de informes

El **Servidor de Informes** es la solución propuesta por **Microsoft** para añadir las funcionalidades de conectividad y nube. Se trata, más que de una versión propia, de un complemento, y tanto es así como se trata de una funcionalidad incluida por defecto en la versión **Premium** de la edición de escritorio. Este servidor de informes nos permite, haciendo uso de cualquier navegador web, organizar y mostrar los informes generados con la versión de escritorio, y también editar y crear otros nuevos.

Permite además personalizar con nuestro **branding** el servidor de informes. De esta forma, al acceder, podemos visualizar el entorno como si de una aplicación personalizada para nuestra organización fuera, incluyendo nuestros temas, colores y logos corporativos.

3.3.2 Arquitectura

Para hacer funcionar la herramienta será necesario contar con el trabajo conjunto de dos clústeres, el primero dedicado al **front-end**, gestionando las conexiones a datos y autenticaciones de usuario, mientras que el segundo ocupándose del **back-end**, ejecutando todas las acciones realizadas por el usuario.

En lo que respecta al clúster *front-end*, destacar que para la administración de identidades se vale del servicio *Azure Active Directory*¹⁹, que permite controlar el acceso de las aplicaciones y los permisos de usuario. Utiliza también el *Administrador de Tráfico de Azure*²⁰ para el control de descargas y contenido. Por último, a través del *Azure Content Delivery Network*²¹, gestiona y distribuye los archivos necesarios según las características geográficas del usuario. Todo esto queda reflejado en el siguiente esquema, **Figura 66**.

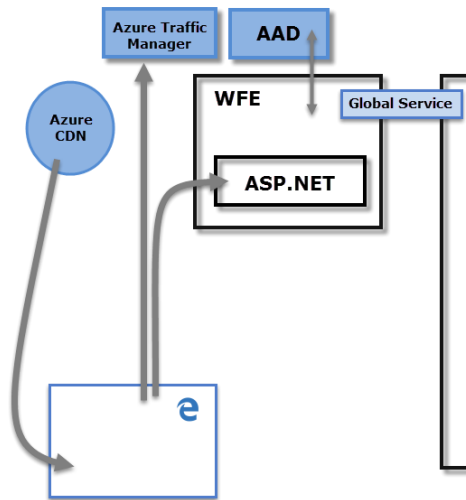


Figura 66. PowerBI front-end. Fuente: www.powerbi.microsoft.com

Por su parte, el segundo clúster, dedicado al *back-end*, se encarga de la gestión de la parte operativa del servicio, **Figura 67**. Este servicio controla los conjuntos de datos a los que se accede y almacena, así como su actualización, los paneles de usuario, la elaboración de informes etc.

¹⁹ Servicio de administración de identidades de Microsoft.

²⁰ Servicio para controlar la distribución del tráfico de los usuarios para puntos de conexión de servicio en distintos centros de datos, de Microsoft.

²¹ Servicio para facilitar que una aplicación alojada en la nube entregue contenido de manera rápida y confiable, almacenando en caché los datos y acelerando el contenido, de Microsoft.

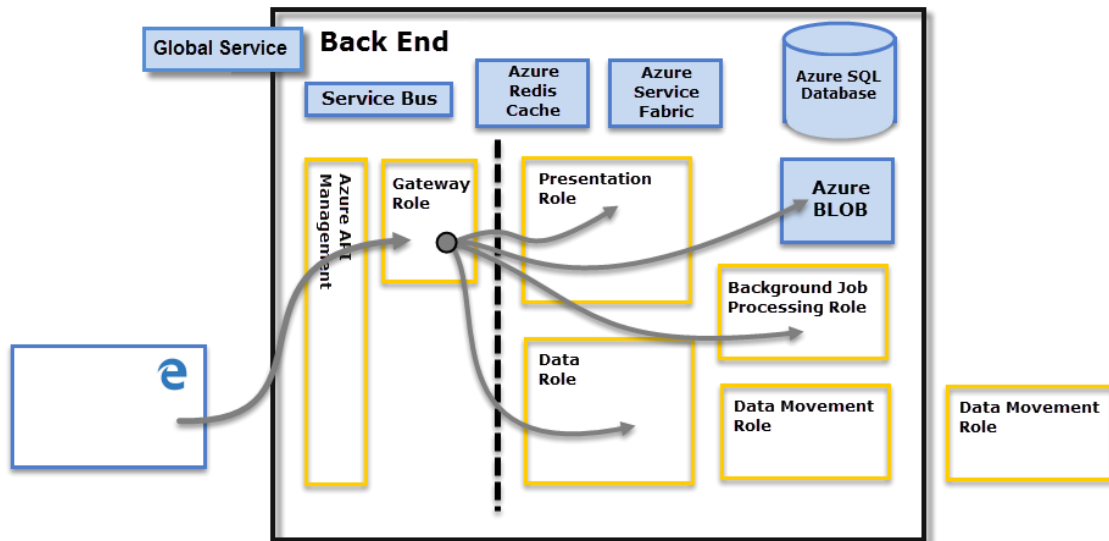


Figura 67. PowerBI back-end. Fuente: www.powerbi.microsoft.com

3.3.3 Aspectos técnicos

3.3.3.1 Arrastrar y soltar

Una vez más encontramos el concepto clave en el que se basan la mayoría de las herramientas de *BI* como elemento fundamental de funcionamiento y uso de la aplicación. Se trata de la fórmula *arrastrar y soltar*, que observamos ofrece buenos resultados en términos de usabilidad.

3.3.3.2 DAX

DAX es el acrónimo en inglés de *Data Analysis Expressions*, es el nombre dado a la biblioteca de funciones, operadores y constantes que pueden ser utilizadas en varias de las aplicaciones de *Microsoft*, como son el *SQL Server Analysis Services*, *Power Pivot Excel* y *Power BI Designer*.

Crear una nueva variable o dimensión en función de las obtenidas mediante la carga de datos puede resultar sencillo si lo que buscamos es algo simple, como por ejemplo una suma de valores. Sin embargo, cuando buscamos crear variables más complejas la situación puede complicarse, y es ahí donde *DAX* puede marcar la diferencia. Para ejemplificarlo, supongamos que queremos comparar la cifra de ventas del trimestre pasado frente a las del actual trimestre. Para esto, tendremos que realizar una búsqueda filtrada, una operación que puede simplificarse utilizando algunas funciones preconfiguradas de *DAX*. En nuestro caso quedaría lo indicado en la siguiente **Figura 68**.

La fórmula debe tener el siguiente aspecto: > **Previous Quarter Sales = CALCULATE(SUM(Sales[SalesAmount]), PREVIOUSQUARTER(Calendar[DateKey]))**

Figura 68. Ejemplo estructura sentencias DAX. Fuente: elaboración propia

Ya tendríamos nuestra nueva variable, pudiendo utilizar el mismo filtro con los datos restantes y realizar con ello un gráfico comparativo de las ventas del pasado trimestre frente al actual, **Figura 69.**

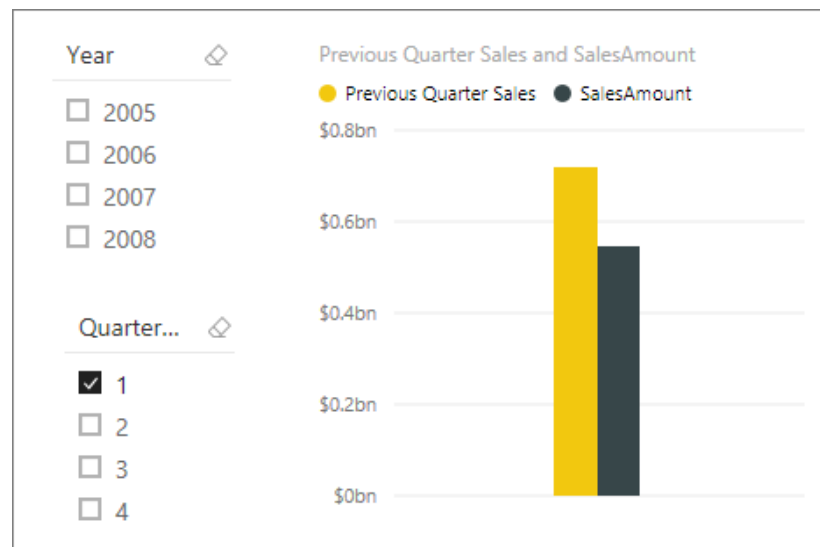


Figura 69. Gráfico DAX ejemplo. Fuente: elaboración propia

3.3.3.3 Ampliación de funciones mediante APIs

PowerBI incluye una API denominada **REST** que proporciona acceso a recursos tales como datos, tablas y filas. Esta es la API que nos permite integrar las funciones de la herramienta dentro de nuestras aplicaciones. Algunas de las funciones admitidas son la configuración de nuevas fuentes de datos, copiar conjuntos de datos, verificación de credenciales, etc.

3.3.4 Funcionalidades de usuario

3.3.4.1 Fuentes y conexión a datos

El proceso de conexión a datos se realiza de forma muy sencilla, pudiendo realizarse en un primer momento cuando ejecutamos nuestro espacio de trabajo por primera vez, mediante el menú interactivo a través de la opción *get data*, **Figura 70.**

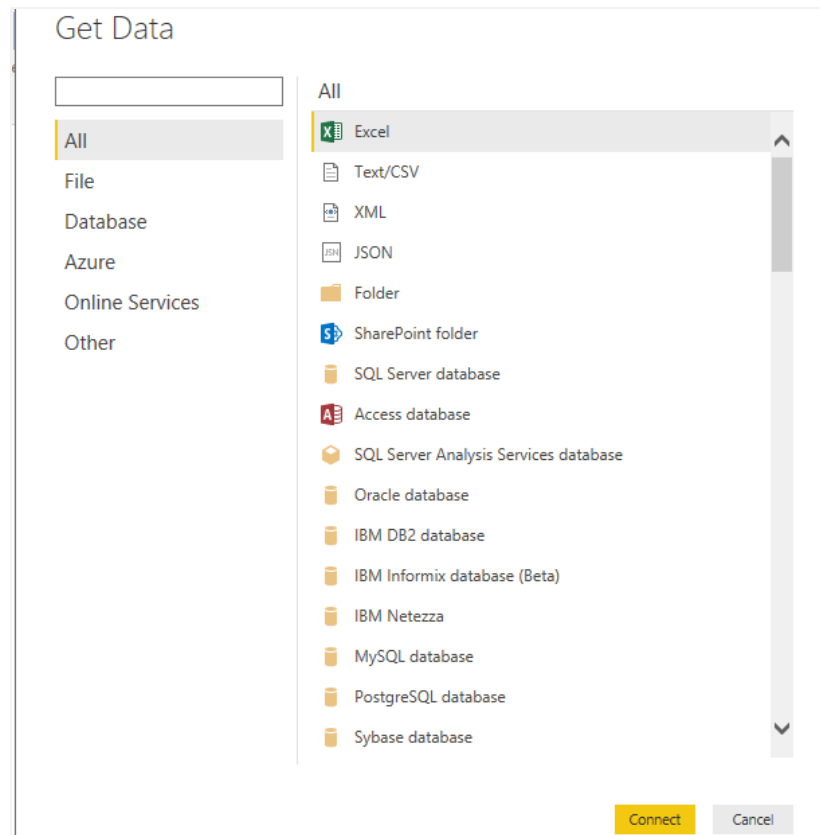


Figura 70. Menú ingestión datos. Fuente: elaboración propia

Si tras ejecutar la aplicación, mientras trabajamos con ella, quisiéramos añadir nuevos datos lo haríamos mediante el botón que se encuentra en uno de los menús de la interfaz, **Figura 71**.

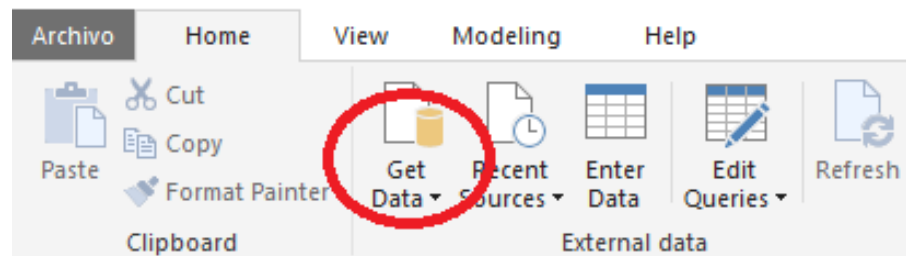


Figura 71. Botón Get data en menú. Fuente: elaboración propia

Las fuentes de datos permitidas por *PowerBI* son las siguientes:

- **Fichero:** texto, csv, xls, xlsx, json, sharepoint folder.
- **Bases de datos:** SQL Server, Access, SQL Server Analysis Services, Oracle, IBM DB2, Informix IBM (beta), Netezza, MySQL, PostgreSQL, Sybase, Teradata, SAP HANA, SAP Busines Warehouse, Amazon Redshift, Impala, Google BigQuery, Snowflake

- **Azure**
- **Servicios online:** Sharepoint Online, Microsoft Exchange Online, Dynamics 365, Common Data Service, Microsoft Azure Consumption Insights, Visual Studio Team Services, Salesforce, Google Analytics, appFigures, comScore Digital Analytix, Facebook, GitHub, Kusto, MailChimp, Mixpanel, Planview Enterprise, Projectplace, QuickBooks Online, Smartsheet, SparkPost, SQL Sentry, Stripe, SweetLQ, Troux, Twilio, tyGraph, Webtrends, Zendesk
- **Otros:** Sharepoint, Web, Odata, Active Directory, Microsoft Exchange, Hadoop, Spark, Scripts R, ODBC, OLE DB,

3.3.4.2 Interfaz

Los **informes** son el elemento básico de trabajo sobre el cuál se construirán los gráficos e informes, denominados visualizaciones. Cada informe estará a su vez compuesto de varias hojas, pudiendo ser esta gráficos simples o conjuntos de gráficos. El aspecto básico de la interfaz en la vista de informe se muestra en la **Figura 72**.

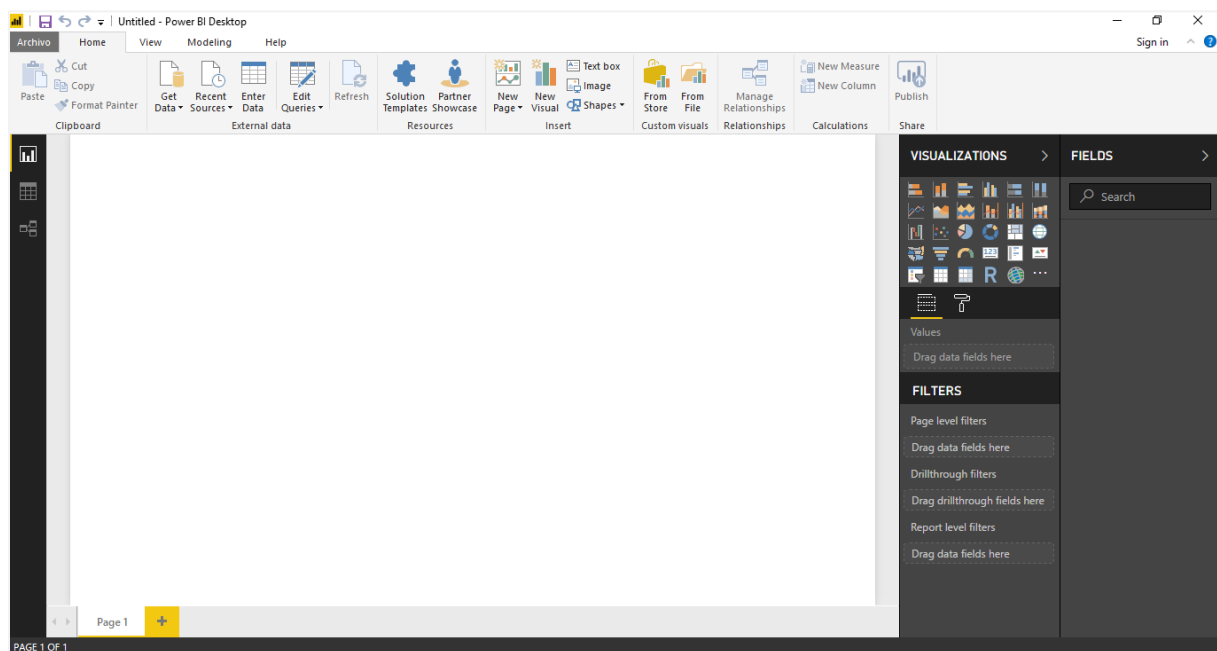


Figura 72. Interfaz PowerBI. Fuente: elaboración propia

Las **hojas** suponen los distintos tapices sobre los que vamos a crear las visualizaciones, se pueden seleccionar en la parte inferior izquierda de la interfaz.

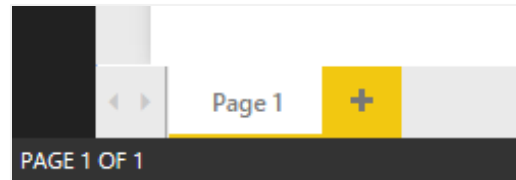


Figura 73. Hojas y menú interactivos. Fuente: elaboración propia

La estructura de la interfaz se compone de los siguientes elementos. En la parte superior encontramos las barras de opciones y menú interactivo del estilo de la suite de *Office*. En ellos encontramos pestañas que nos darán acceso a las funciones básicas de herramienta.

En la parte izquierda se sitúa una botonera que nos dará acceso a tres elementos importantes. El primero de ellos nos lleva a la pestaña de **Hojas** que ya se ha comentado anteriormente. La segunda pestaña nos lleva a **Data**, el espacio destinado a la visualización y edición de los datos previamente cargados, **Figura 74**.

Country 15	Region 15	Happiness Rank 15	Happiness Score 15	Standard Error 15	Economy 15	Family 15	Health 15	Freedom 15	Trust 15	Generosity 15
Albania	Southern Asia	153	3575	3084	31982	30285	30335	23414	9719	365
Albania	Central and Eastern Europe	95	4959	5013	87867	80434	81325	35733	6413	1427
Algeria	Middle East and Northern Africa	68	5605	5099	93929	107772	61766	28579	17383	782
Angola	Sub-Saharan Africa	137	4033	4758	75778	8604	16683	10384	7122	1234
Argentina	Latin America and Caribbean	30	6574	4612	105351	124823	78723	44974	8484	1145
Armenia	Central and Eastern Europe	127	435	4763	76821	77711	7229	19847	39	785
Australia	Australia and New Zealand	10	7284	4083	133358	130923	93156	65124	35637	4356
Austria	Western Europe	13	72	3751	133723	129704	89042	62433	18676	3308
Azerbaijan	Central and Eastern Europe	80	5212	3363	102389	93793	64045	3703	16065	779
Bahrain	Middle East and Northern Africa	49	596	5412	132376	121624	74716	45492	306	1796
Bangladesh	Southern Asia	109	4694	3077	39753	43106	60164	4082	12569	2122
Belarus	Central and Eastern Europe	59	5813	3938	103192	123289	73608	37938	1909	1104
Belgium	Western Europe	19	6937	3595	130782	128566	89667	5845	2254	222
Benin	Sub-Saharan Africa	155	334	3656	28665	35386	3191	4845	801	182
Bhutan	Southern Asia	79	5253	3225	77042	110395	57407	53206	15445	4799
Bolivia	Latin America and Caribbean	51	589	5642	68133	97841	5392	57414	88	2053
Bosnia and Herzegovina	Central and Eastern Europe	96	4949	6913	83223	91916	79081	9245	227	2480
Botswana	Sub-Saharan Africa	128	4332	4934	99355	110464	4776	49495	12474	1046
Brazil	Latin America and Caribbean	16	6983	4076	98124	123287	69702	49049	17521	1457
Bulgaria	Central and Eastern Europe	134	4218	4828	101216	110614	76649	30587	872	1192
Burkina Faso	Sub-Saharan Africa	152	3587	4324	25812	85188	27125	39493	12832	2174
Burundi	Sub-Saharan Africa	157	2905	8658	153	41587	22396	1185	10062	1972
Cambodia	Southeastern Asia	145	3819	5069	46038	62736	61114	66246	7247	4035
Cameroon	Sub-Saharan Africa	133	4752	4678	4725	88767	23402	49309	5786	2061

Figura 74. Pestaña de Data en PowerBI. Fuente: elaboración propia

La tercera nos permite visualizar las **conexiones existentes** entre las dimensiones contenidas en nuestras fuentes de datos.

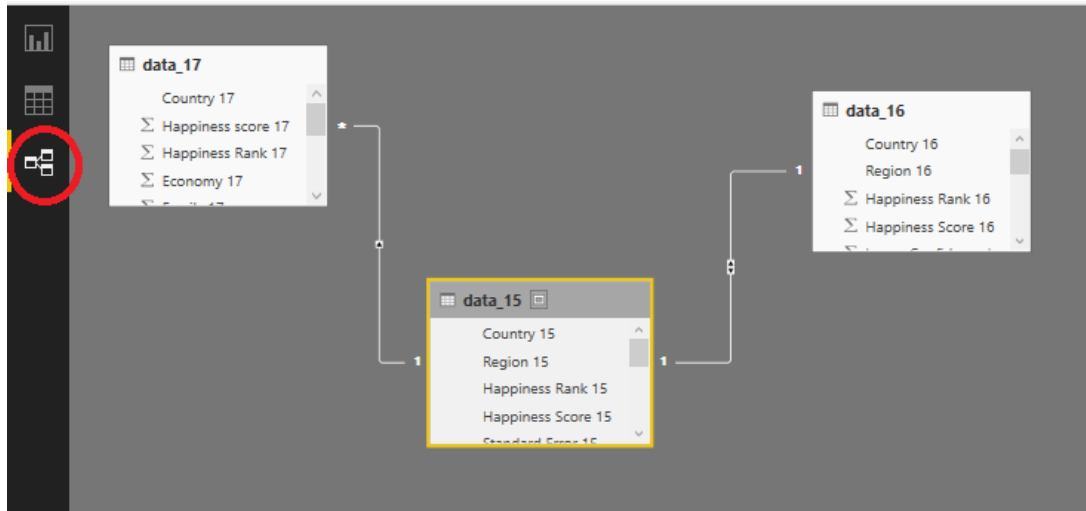


Figura 75. Pestaña de conexiones existentes en PowerBI. Fuente: elaboración propia

3.3.4.3 Visualizaciones

En la parte derecha de las hojas de trabajo encontramos un menú interactivo de gran utilidad compuesto por tres subsecciones; **visualizaciones, filtros y campos**.

Visualizaciones, Figura 76, nos dará acceso a los tipos de gráfico disponibles en *PowerBi*, como son los gráficos de barra, líneas, círculos, árboles de decisión, mapa, etc.

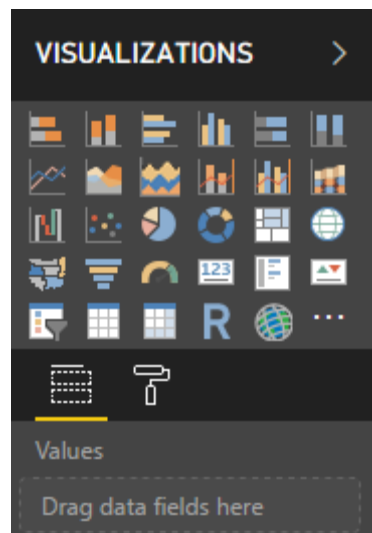


Figura 76. Menú visualizaciones. Fuente: elaboración propia

Los **filtros, Figura 77**, nos permiten seleccionar qué dimensiones y para qué las queremos usar en cada tipo de gráfico. Esta sección se amplía además en función del tipo de gráfico seleccionado.

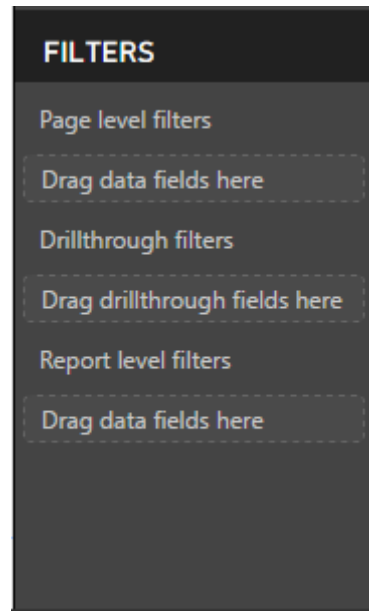


Figura 77. Menú filtros. Fuente: elaboración propia

Ya se han mencionado las múltiples opciones de visualización disponibles en la herramienta, desde gráficos de barra hasta mapas de calor. Con estas visualizaciones se pueden afrontar presentaciones para ámbitos muy distintos de la empresa, desde la gestión de ventas, comercial, producción, marketing etc. Algunos ejemplos de los resultados que podemos conseguir se muestran a continuación, **Figura 78**, **Figura 79** y **Figura 80**.

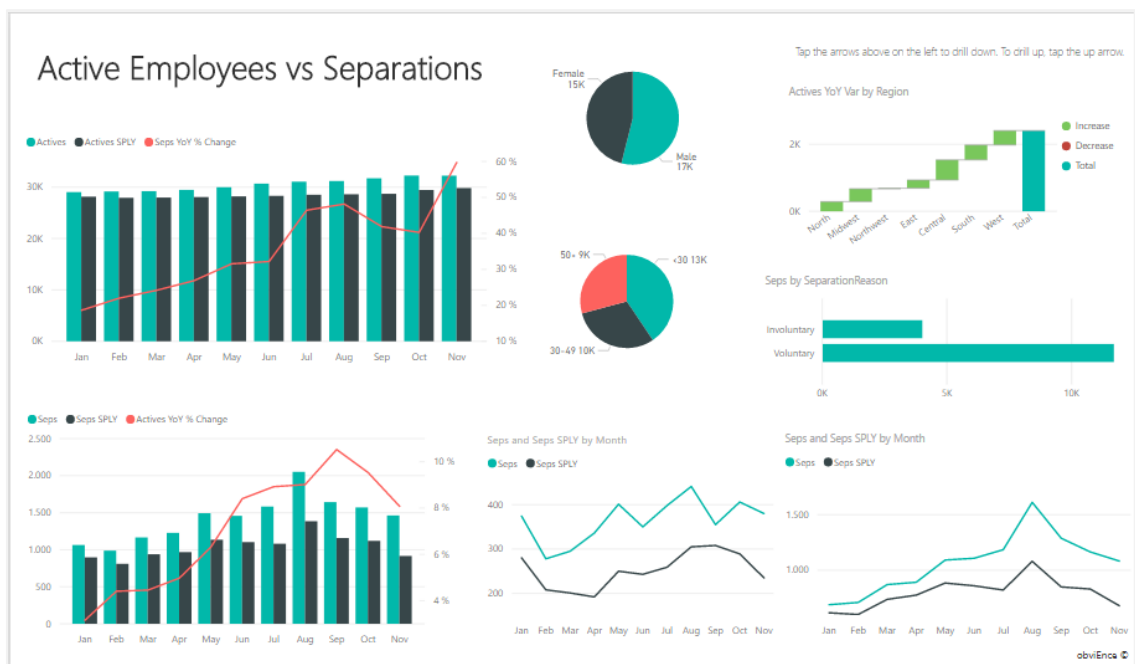


Figura 78. Ejemplo visualizaciones 1. Fuente: elaboración propia



Figura 79. Ejemplo visualizaciones 2. Fuente: elaboración propia

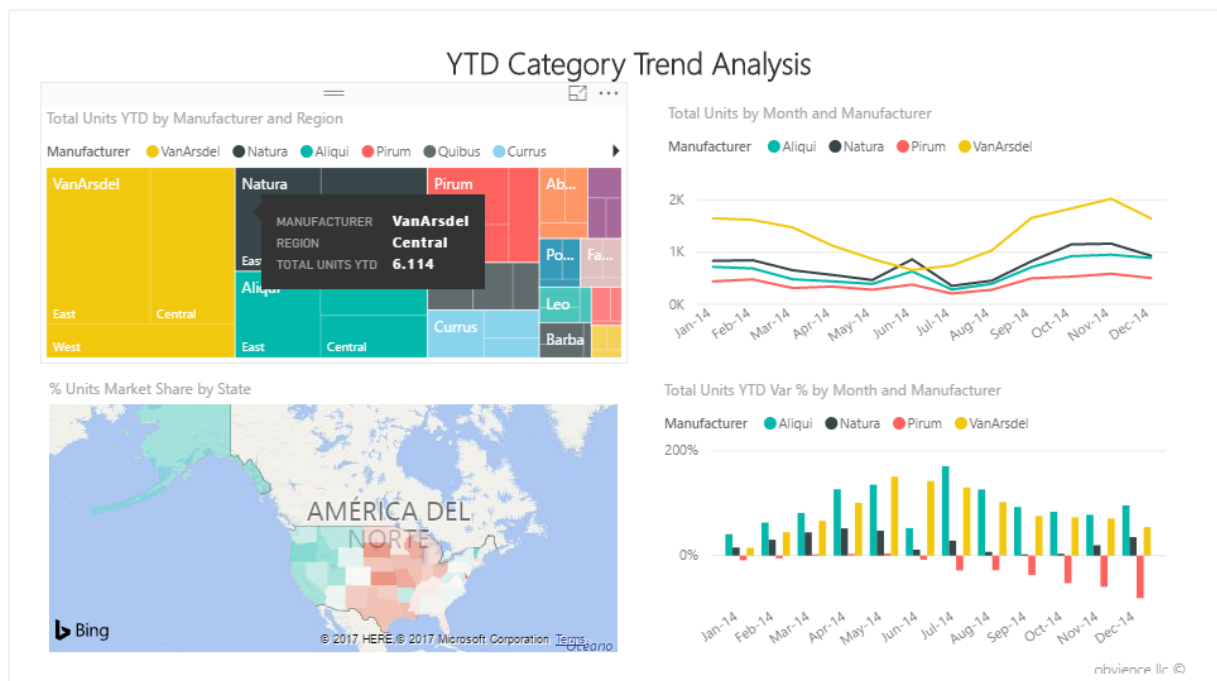


Figura 80. Ejemplo visualizaciones 3. Fuente: elaboración propia

3.3.4.4 Creación de funciones y variables.

Una vez creadas nuestras *apps* el programa nos pedirá cargar los datos con los que deseamos trabajar. Una vez cargados este desglosará la información contenida, generando las

variables/dimensiones. Como ya hemos visto en el caso de la herramienta anterior es probable que en muchas ocasiones necesitemos crear nuevas variables a partir de estas. El procedimiento para hacerlo resulta, una vez más, de mayor complejidad que con otras herramientas. Y es que, para poder crear una nueva columna, tendremos que dirigirnos al menú de navegación y cargar el gestor de datos, desplazarnos hasta edición de tabla, para una vez allí desplazarnos a la opción “añadir campo” y “campo calculado”.

Sin embargo, encontramos como puntos positivos que el menú que nos permite la creación de variables es más apropiado para usuarios *legos* puesto que no necesita que este escriba la fórmula para la creación si no quiere hacerlo, pudiendo guiarse por los menús interactivos de los que dispone. En caso de que sí fuera preciso, hemos visto como la herramienta se vale de DAX para facilitar esta ejecución.

3.3.4.5 Predicciones, clúster y R

La creación de herramientas para hacer predicciones queda exclusivamente disponible para la versión *Premium* del programa, no pudiendo testearlas con la versión que se dispone para este estudio. Esto puede solventarse haciendo uso de los scripts que podemos desarrollar con *R*. En la versión ampliada del programa encontramos un botón extra en el área de filtros, denominado *Analytics*, que nos dará acceso a una serie de funciones de utilidad, como son la creación de:

- Líneas constantes para ejes X e Y
- Líneas mínimas y máximas
- Promedios
- Medianas
- Percentiles
- Previsiones/predicciones/forecasting

PowerBI cuenta con integración nativa para scripts de *R*, pudiendo hacer uso de los mismos de la misma forma que realizamos cualquier tipo de análisis. Para esto necesitamos tener instalado *Rstudio* en nuestro equipo, desde donde crearemos y editaremos los scripts que luego integraremos en la herramienta, **Figura 81**.

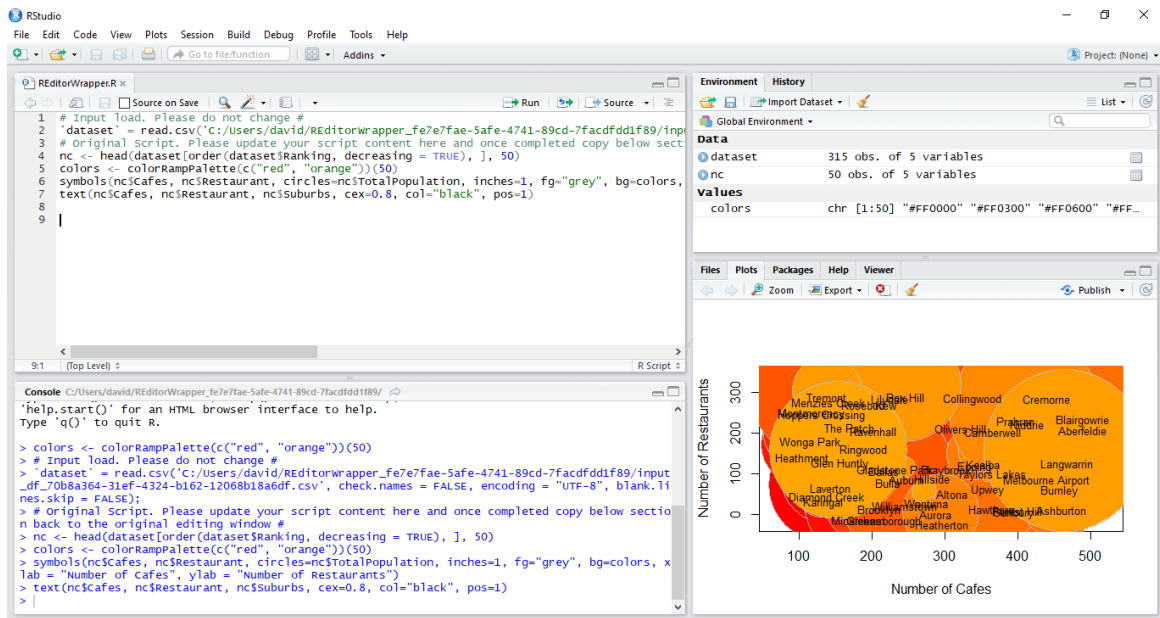


Figura 81. RScript Studio. Fuente: elaboración propia

Este script una vez ejecutado en *PowerBI* quedaría como sigue, **Figura 82**.

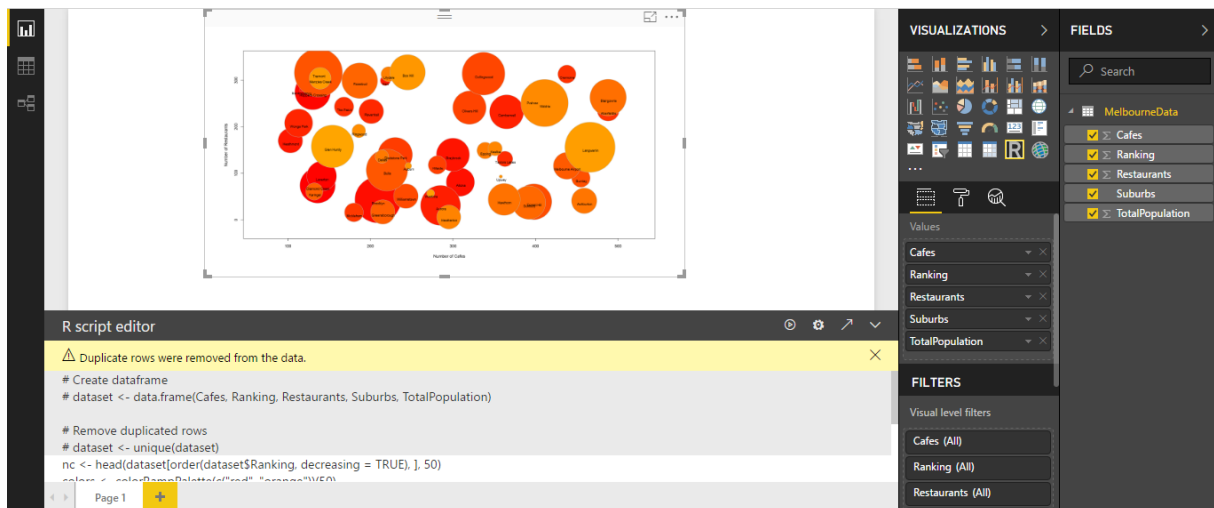


Figura 82. Visualización creada a partir de script R en PowerBI. Fuente: elaboración propia

Vemos que en la parte superior quedaría el objeto visual generado mediante el script, situado en la parte inferior, mientras que a la derecha tendríamos las dimensiones generadas. El editor de scripts R cuenta con una botonera que nos permite, ejecutar el script, acceso a la configuración y edición. Al hacer clic en el botón de edición el programa nos mandará automáticamente al editor de *RStudio*.

3.3.4.6 Exportación de datos

La exportación de los datos con los que hemos trabajado puede ser exportados con facilidad a través de la opción **export data**, permitiéndose exclusivamente en formato csv.

3.3.5 Seguridad, volumetría y Big Data

En lo que respecta a la **seguridad**, *PowerBi* cuenta con un sistema de administración que permite otorgar permisos tanto a nivel de usuario como a nivel de grupo. Cuenta además con el sistema de *roles de usuario* que permite crear configuraciones de permisos predeterminadas para facilitar la administración del sistema. Se utiliza el servicio *Azure Active Directory*, **Figura 83**, para realizar la autenticación de usuarios en el momento de autenticarnos en el servicio mediante la dirección de correo electrónico facilitada.



Figura 83. Azure Active Directory Logo. Fuente: elaboración propia

Tanto la conexión como las solicitudes cliente/servidor se administran mediante la *Puerta de Enlace* controlada por el servicio *Azure API Management* que viene a interactuar con los distintos servicios del sistema. El servicio de *HDInsights* permite además la autorización de grupos de usuario, añadiendo una capa extra de seguridad, funcionando bajo las directrices HIPAA, PCI y SOC.

En lo referente a la **volumetría** ya hemos comentado la capacidad que tiene la aplicación para acceder a bases de datos de tipo **Big Data**, como son Spark y Azure. Se ha visto también su capacidad para unir datos provenientes de fuentes distintas, pudiendo con ello visualizar y tratar volúmenes de datos elevados. Además de esto, apoyándose de **Azure HDInsight** las posibilidades de *PowerBI* se disparan. Este servicio permite crear clústeres optimizados para *Hadoop*, *Spark*; *hive*, *Hbase Storm*, *Kafka* y *Microsoft R Server*.

Con todo ello *PowerBI* está capacitado, conectándose de manera nativa, para la realización de procesos de distribución paralela en múltiples clústers, como son los procesos de ETL, ciencia de datos (ejemplo: MapReduce), y está preparado para analizar información proveniente de la IoT, que requiere de una capacidad muy elevada en cuanto a la rapidez y carga de trabajo necesaria. Permite además ejecutar entornos de desarrollo como *Visual Studio*, *Eclipse* e *IntelliJ*. Según sus desarrolladores la paralelización conseguida puede llegar a controlar datos con una tasa mil veces superior a la que se alcanzaría con el lenguaje R OS.

Trabajar con IoT supone un desafío. Los sensores que se encuentran en los dispositivos móviles capturan una enorme cantidad de datos, con una tasa de cadencia muy alta, que requiere de procedimientos ágiles que permitan capturar la información y tratarla rápidamente. A través de Kafka, Storm y Spark Streaming la plataforma promete ejecutar todo esto en tiempo real, siguiendo el *esquema* de la siguiente **Figura 84**.

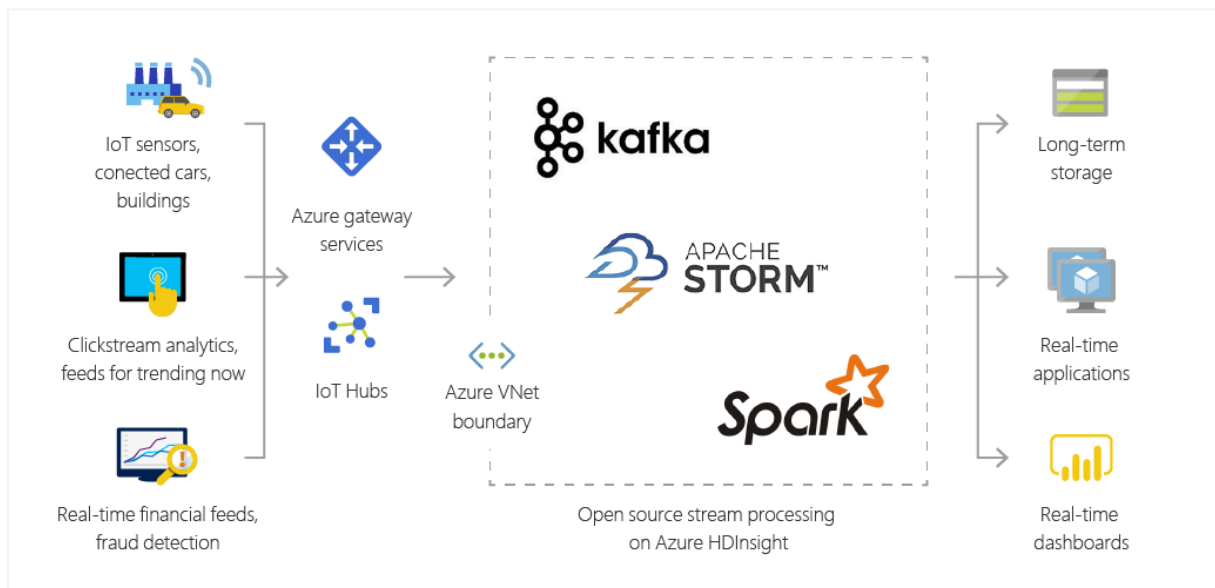


Figura 84. Kafka, Apache Storm y Spark en Streaming con Azure HDInsight. Fuente: www.azure.microsoft.com

Las utilidades que tienen los clústeres accesibles a través de *HDInsight* y de las que nos podremos beneficiar por tanto con el uso de *PowerBI* son, entre otras:

- **Ambari:** aprovisionamiento, administración, supervisión y utilidades de clústeres.
- **Avro** (biblioteca de Microsoft .NET para Avro): serialización de datos para el entorno de Microsoft .NET.
- **Hive y HCatalog:** consultas tipo SQL y una capa de administración de almacenamiento y de tablas.
- **Mahout:** para aplicaciones de aprendizaje automático escalables.
- **MapReduce:** marco heredado para el procesamiento distribuido y la administración de recursos de Hadoop. Consulte [YARN](#).
- **Oozie:** administración de flujo de trabajo.
- **Phoenix:** capa de base de datos relacional sobre HBase.
- **Pig:** scripts más sencillos para transformaciones de MapReduce.
- **Sqoop:** importación y exportación de datos.
- **Tez:** permite a los procesos con muchos datos ejecutarse de forma eficaz a escala.
- **YARN:** administración de recursos que forma parte de la biblioteca básica de Hadoop.
- **ZooKeeper:** coordinación de procesos en sistemas distribuidos.

Figura 85. Componentes y utilidades en clústeres HDInsight. Fuente: www.docs.microsoft.es

3.4 Pentaho

Pentaho BI Suite es el nombre de la solución de *Business Intelligence* desarrollada por la compañía *Hitachi Data Systems*, **Figura 86**. Fundada por *Richard Daley* en el año 2004, originaria de Estados Unidos, se compone de una serie de herramientas que vienen a cubrir las necesidades de cualquier organización en materia de inteligencia de negocio.



Figura 86. Logo Pentaho. Fuente: www.pentaho.com

3.4.1 Introducción y versiones

El nombre completo de esta suite es ***Pentaho: Data Integration, Business Analytics and Big Data***. Ya por su nombre nos viene a dar una idea de la intención de la empresa desarrolladora de generar un producto integral, que venga a dar respuesta a la totalidad, según ellos mismos dicen, de las necesidades relativas a inteligencia empresarial que se puedan tener. Su principal característica

diferenciadora, el haberse desarrollado bajo la filosofía *Open Source*²², desarrollado bajo lenguaje *Java*²³.

A diferencia de lo que ocurre habitualmente, en este caso un mismo archivo instalador nos dará acceso a todas las versiones del programa disponibles, pudiendo instalar todas o seleccionar específicamente la que queremos, **Figura 87**.

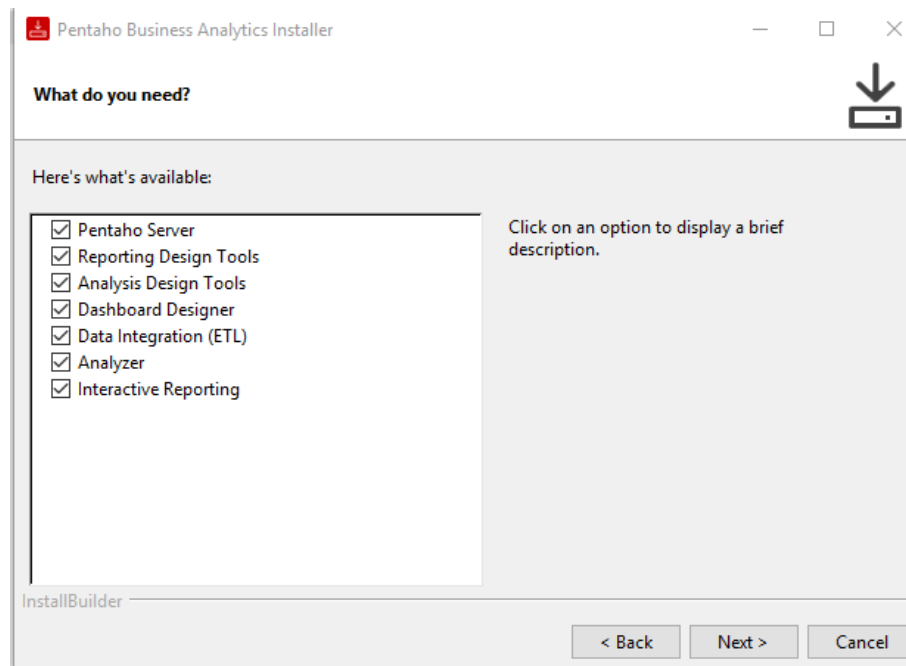


Figura 87. Opciones de instalación Pentaho. Fuente: elaboración propia

Esta herramienta tiene una particularidad respecto de las vistas anteriormente, y es la forma en que el usuario interactúa con ella, algo en lo que ahondaremos en el apartado de la **Arquitectura**. Esta interacción, en vez de realizarse a través de una aplicación de escritorio se producirá mediante el uso de una interfaz a la que accederemos utilizando un navegador web de nuestra elección, para, por ejemplo, realizar los análisis de inteligencia de negocio.

²² Software en el que el propietario de los derechos de uso y código fuente permite su utilización de forma pública.

²³ Lenguaje de programación orientado a objetos.

3.4.1.1 Aggregation Designer

El **Pentaho Aggregation Designer**, **Figura 88**, es una herramienta de la suite *Pentaho* desde la versión 2.0 de la misma. Esta herramienta sirve para la gestión, creación y manejo de tablas agregadas²⁴, con lo que se optimiza el rendimiento de las consultas de la herramienta de análisis. La herramienta proporciona a través de una interfaz sencilla el conjunto de instrucciones necesarias para generar estas tablas, incorporando además un asesor inteligente para facilitar el proceso de *Definición de datos (DDL)* y *Manipulación de Datos (DML)*.

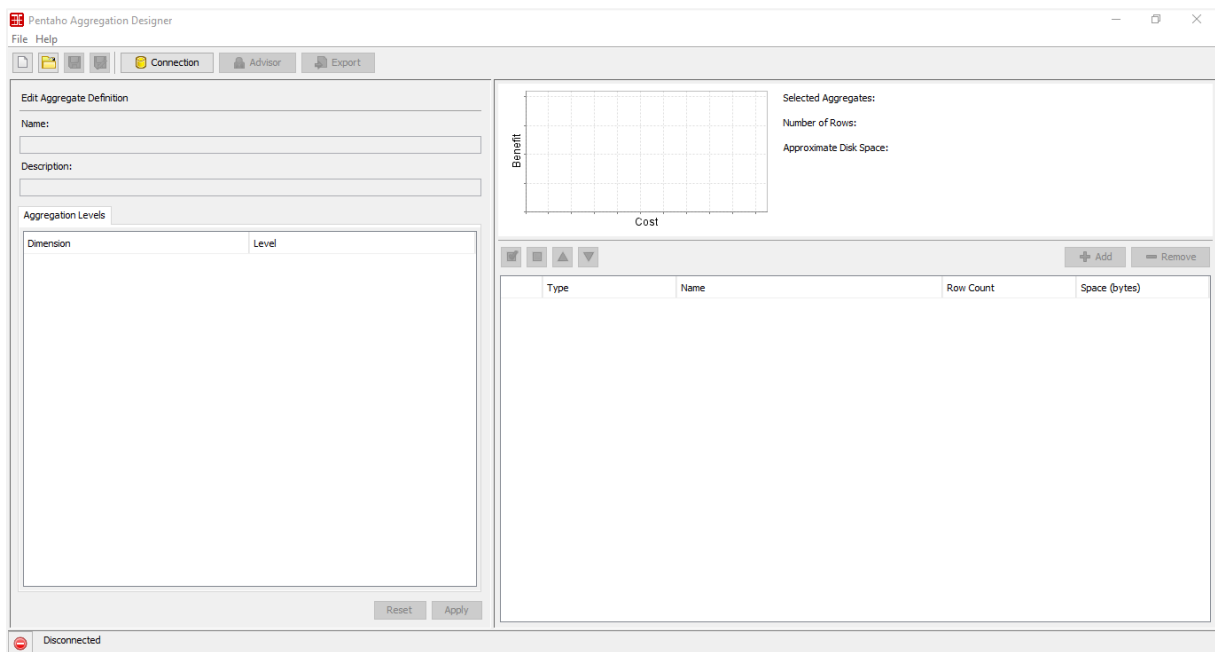


Figura 88. Interfaz Pentaho Aggregation Designer. Fuente: elaboración propia

3.4.1.2 Pentaho Data Integration

Pentaho Data Integration (PDI), **Figura 89**, anteriormente conocido como **Kettle** y actualmente denominado como **Spoon** (por el nombre del motor gráfico) es una herramienta muy poderosa incluida en la suite. Su función principal, facilitar los procesos *ETL* (*extracción, transformación y carga de datos*), valiéndose para ello de una *GUI* (*interfaz gráfica de usuario*).

²⁴ Las tablas agregadas son un proceso que permite, a partir de las tablas de datos originales que cuentan con un elevado nivel de detalle, crear otras más manejables con los indicadores tipo, reduciendo los procesos y agilizando las consultas.

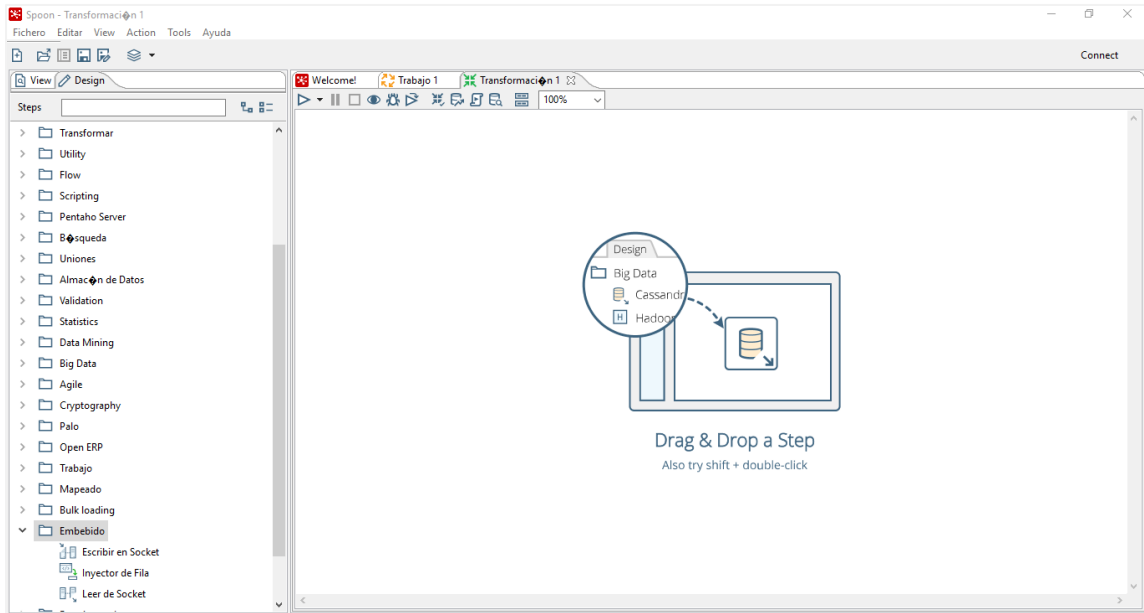


Figura 89. Interfaz PDI. Fuente: elaboración propia

3.4.1.3 Pentaho Report Designer

Report Designer es una funcionalidad que nos permite generar informes profesionales, **Figura 90**.

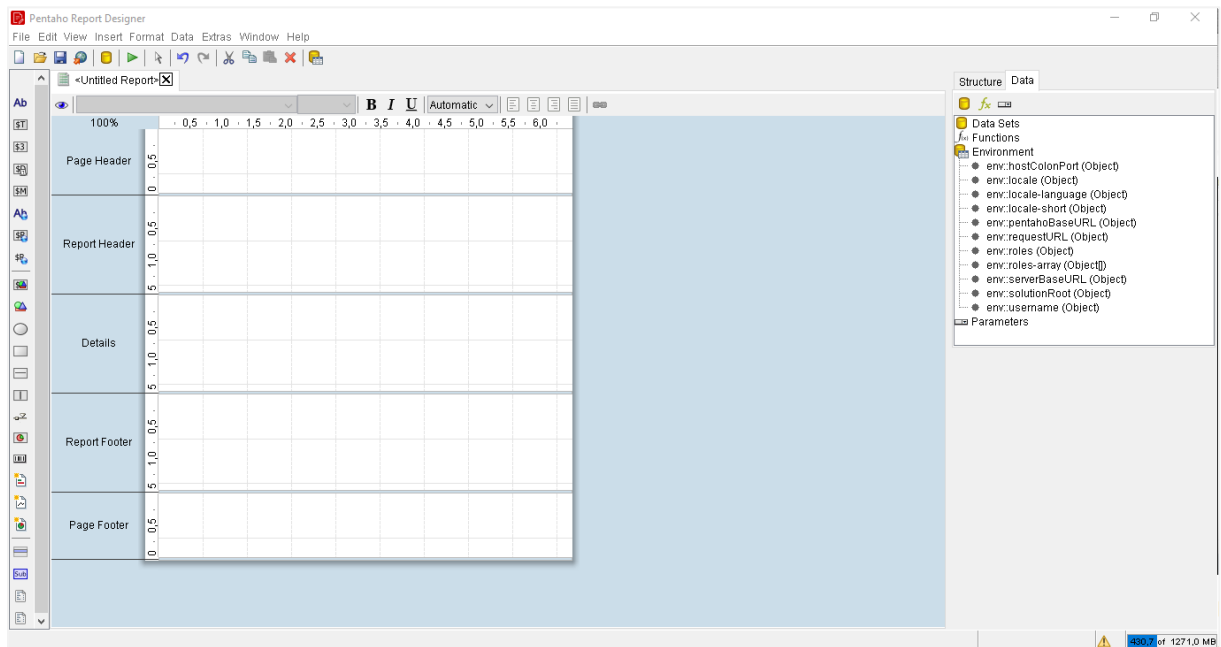


Figura 90. Interfaz Pentaho Report Designer. Fuente: elaboración propia

El espacio de trabajo está compuesto de un elemento central, donde vemos los distintos niveles que compondrán nuestro reporte, y menús en la parte superior y laterales para incluir, a través del sistema *arrastrar-soltar* los elementos que queramos formen parte del reporte. El aspecto general de la interfaz es anticuado, poco intuitivo, obligando al usuario a saltar por multitud de menús y opciones para hacer prácticamente cualquier cosa.

3.4.1.4 Pentaho Schema Workbench

Schema Workbench es una herramienta gráfica destinada a la construcción de esquemas de tipo *Mondrian*, un tipo de cubo de datos *OLAP*, para posteriormente ser incorporados en la herramienta de *BI*.

3.4.1.5 Pentaho Server

El servidor de **Pentaho** permite alojar todo el contenido creado por el usuario en las otras versiones, y supone el eje central para ejecutar las transformaciones y procesos de integración de datos. Para ello se vale del motor PDI anteriormente mencionado.

Este sistema permite además administrar permisos de usuarios, así como crear y gestionar roles de usuarios, para facilitar estas tareas. Mencionar su capacidad para la integración de seguridad con sistemas **LDAP** o **Active Directory**, entre otros.

Unido a la versión de *BI*, permite a través de la consola de usuario accesible desde la interfaz web la utilización de una serie de complementos:

- **Analizador:** visualiza y filtra datos
- **Informes interactivos:** crea informes basados en plantillas mediante la filosofía *arrastrar-soltar*
- **Diseñador de dashboards:** permite generar dashboards, con gráficos, tablas de datos, etc.
- **Aplicación móvil:** permite crear y gestionar contenido para móvil

3.4.1.6 Interfaz Web

Para poder acceder al resto de funcionalidades, deberemos conectarnos a un servidor. Este puede ser a nuestro propio servidor, que hayamos montado a través del **Pentaho Server**, o bien a otro alojado fuera de nuestro equipo, o localmente. Para esto último, emulando el que sería el sistema

de trabajo de una aplicación de escritorio corriente tendríamos que ejecutar un servidor en local. Para ello lanzaríamos dicho servidor como si de una aplicación se tratase, **Figura 91**.

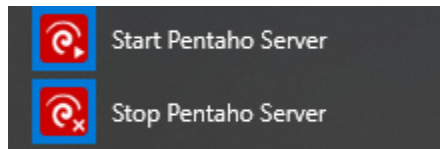


Figura 91. Start/stop Pentaho Server. Fuente: elaboración propia

Sorprende el elevado uso de memoria que hace esto, **Figura 92**.

Nombre	CPU	Memoria	Disco	Red	GPU
Commons Daemon Service Run... Pentaho Server	0,3%	2.637,8 MB	0,1 MB/s	0 Mbps	0%

Figura 92. Uso de memoria de Pentaho Server. Fuente: elaboración propia

Posteriormente se accede a la consola de usuario, haciendo uso del icono correspondiente, que nos llevará a la interfaz web con el navegador en este caso predeterminado por el usuario en el sistema. Al acceder nos encontramos con una ventana que nos pide el usuario y contraseña, para posteriormente una vez autenticados, acceder al menú principal, **Figura 93**.

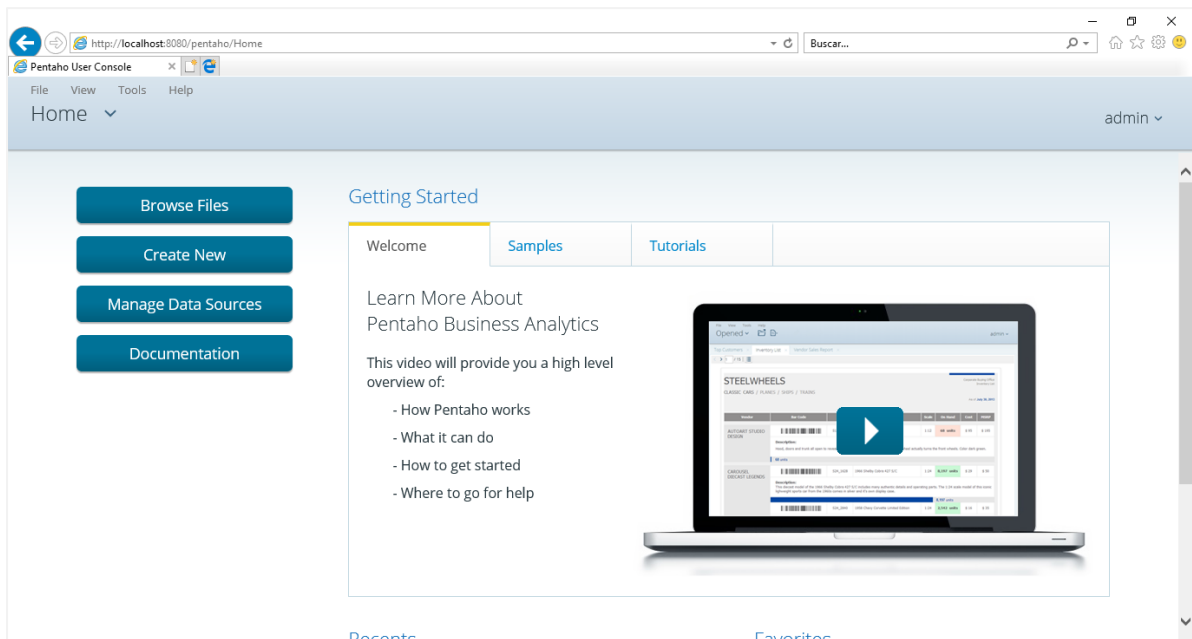


Figura 93. Interfaz principal consola web Pentaho. Fuente: elaboración propia

3.4.2 Arquitectura

La arquitectura de la herramienta se distribuye en 3 niveles o capas. El primero de ellos el nivel o **Capa de Usuario/cliente** a la que se accede a través de dos vías. La primera, toda la operativa que se realiza dentro del servidor y a la que accede el usuario a través de la interfaz web. En ella podremos realizar procesos analíticos para descubrir datos, generación de dashboards y reportes interactivos, **Figura 94**.



Figura 94. Consola de usuario, opciones. Fuente: www.pentaho.com

La segunda vía, está compuesta por aquellas aplicaciones que se realizan fuera del entorno del servidor, en las aplicaciones que complementan la aplicación. Estas son el **Pentaho Data Integration**, **Pentaho Report Designer** y el **WorkBench**.

El segundo nivel lo compone la **Capa de Servidor**, que se encarga de realizar los cálculos de los procesos analíticos la integración de datos y a los repositorios de contenido, **Figura 95**. Se encargará por tanto también de la carga de datos desde las distintas fuentes posibles.



Figura 95. Servidor analítico. Fuente: www.pentaho.com

Estas fuentes compondrán la tercera capa, las fuentes de datos, **Figura 96**.

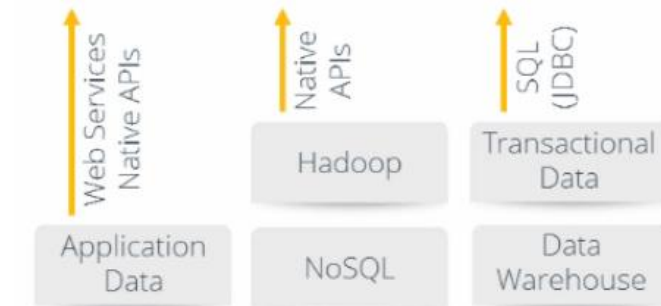


Figura 96. Fuentes de datos para el servidor. Fuente: www.pentaho.com

3.4.3 Aspectos técnicos

Pentaho destaca por ser una plataforma **Open Source**. Cuenta con una comunidad de usuarios muy potente que están continuamente desarrollando mejoras y extensiones para la plataforma, **Figura 97.**



Figura 97. Logo Pentaho Open Source. Fuente: www.bigdataanalyticsnews.com

Dispone de un sistema para trabajar con informes multidimensionales mediante cubos **OLAP**, así como herramientas integradas para ejecutar procesos de minería de datos con los que detectar patrones.

A diferencia del resto de herramientas, a la suite de **Pentaho** tiene un modelo de funcionamiento en el que se le da mayor peso al uso web. Tanto es así que, para obtener acceso a la interfaz de usuario, salvo para las aplicaciones anteriormente mencionadas, se debe hacer acceso de un navegador web que nos servirá de puente con el servidor con el que trabajemos, tanto si este se encuentra fuera de nuestro equipo o a nivel local.

Permite la integración de datos desde distintas fuentes, a través de su herramienta **Kettle ETL** que incorpora métodos para la migración de datos, o incluso la integración con externos. La gran ventaja que supone el uso de **Kettle** es que no se requiere de conocimientos técnicos para realizar

estas tareas, debiendo tan solo especificar lo que el usuario busca a través de métodos preconfigurados, realizando tan solo un esquema del mismo y configurándolo debidamente.

3.4.4 Funcionalidades de usuario

3.4.4.1 Fuentes y conexión a datos

La conexión a datos disponible permite múltiples opciones, pudiendo ser estas de distinta naturaleza. En primer lugar, encontramos las **Bases de Datos Relacionales**, donde encontramos:

- MySQL
- PostgreSQL
- SQLite
- Microsoft SQL Server
- Oracle
- IBM DB2
- Teradata
- Azure SQL Database
- Azure SQL Server

También es posible la conexión a bases de dato de tipo analítico, algo especialmente útil en el análisis de largos y complejos datasets, como, por ejemplo:

- HPE Vertica
- Greenplum
- Amazon Redshift
- SAP HANA
- Teradata
- Netezza

Encontramos también conexión a bases de dato de tipo NOSQL, *Ilustración X*, y framework Hadoop:

- Cassandra
- MongoDB

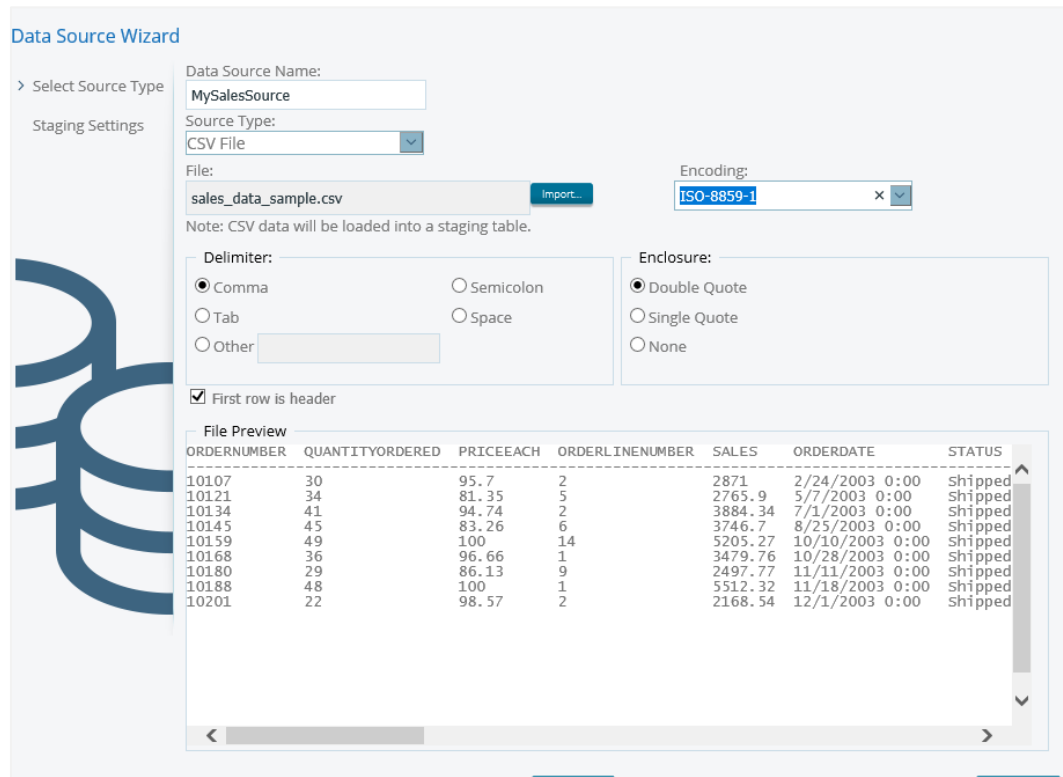
- Apache HBase
- CouchDD
- Cloudera
- Hortonworks
- MapR
- Amazon EMR
- Microsoft Azure HDInsight
- Hive
- Impala
- Apache Spark

Por último, *Pentaho* admite la conexión a múltiples fuentes, como archivos y aplicaciones de negocio:

- JSON
- XML
- Amazon S3
- Excel
- Text Files
- CSV Files
- RSS Feeds
- Business Applications (ERP, CRM)
- Avro
- Splunk
- Google Analytics
- HL7 data
- Java Message Service
- Fixed-width Files

La conexión a estas fuentes se realiza de forma muy sencilla, a través del ***Data Source Wizard***, funcionalidad a la que accedemos a través del menú interactivo principal y que nos guiará en el proceso de conexión hasta la fuente deseada, *Ilustración X*, mostrándonos además una

previsualización de los datos cargados y posteriormente la posibilidad de realizar modificaciones en los mismos antes de empezar a analizarlos.



Data Source Wizard

> Select Source Type

Staging Settings

Data Source Name:

Source Type:

File:

Encoding:

Note: CSV data will be loaded into a staging table.

Delimiter:

☒ Comma ☐ Semicolon

☐ Tab ☐ Space

☐ Other

Enclosure:

☒ Double Quote ☐ Single Quote

☐ None

☒ First row is header

File Preview

ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS
10107	30	95.7	2	2871	2/24/2003 0:00	shipped
10121	34	81.35	5	2765.9	5/7/2003 0:00	shipped
10134	41	94.74	2	3884.34	7/1/2003 0:00	shipped
10145	45	83.26	6	3746.7	8/25/2003 0:00	shipped
10159	49	100	14	5205.27	10/10/2003 0:00	shipped
10168	36	96.66	1	3479.76	10/28/2003 0:00	shipped
10180	29	86.13	9	2497.77	11/11/2003 0:00	shipped
10188	48	100	1	5512.32	11/18/2003 0:00	shipped
10201	22	98.57	2	2168.54	12/1/2003 0:00	shipped

Figura 98. Data Source Wizar. Fuente: elaboración propia

3.4.4.2 Interfaz

Revisamos en este apartado los elementos que conforman la interfaz principal de usuario cuando se accede a la misma a través de la interfaz web y que hemos visto anteriormente. Esta primera ventana nos dará acceso a los menús rápidos, así como a los elementos recientes y favoritos.

En la parte superior encontramos en la barra de menú que siempre se encuentra presente en pantalla y nos da acceso a los elementos típicos, como creación de nuevo contenido, vistas, herramientas y ayuda. Encontramos justo debajo de esta barra un menú que nos permitirá navegar por más opciones, en este caso indicando como *Home*.

Tanto con este botón como a través de la vista principal podemos acceder a la opción de **Búsqueda de Archivos**, que nos dará acceso a un menú navegador de las carpetas, para seleccionar los

ficheros con los que queremos trabajar, **Figura 99**, así como una serie de acciones permitidas para cada tipo de archivo a través del menú en la parte derecha.

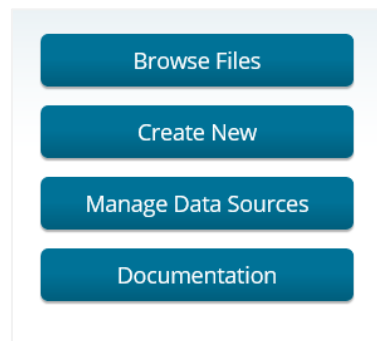


Figura 99. Menú interactivo principal. Fuente: elaboración propia

Ya hemos visto el primer elemento del menú rápido, el botón **Home**. El segundo elemento del menú rápido nos dará acceso a la función **Opened** la cual nos llevará a visualizar los espacios de trabajo con los que estamos trabajando, cada uno de ellos separado conformando una pestaña independiente, **Figura 100**.

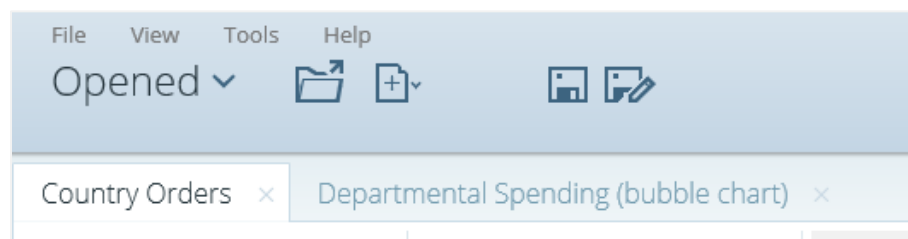


Figura 100. Menú rápido Opened. Fuente: elaboración propia

A través de la opción **Browse Files** tendremos acceso al menú encargado de controlar todos los elementos generados a través de la aplicación, **Figura 101**.

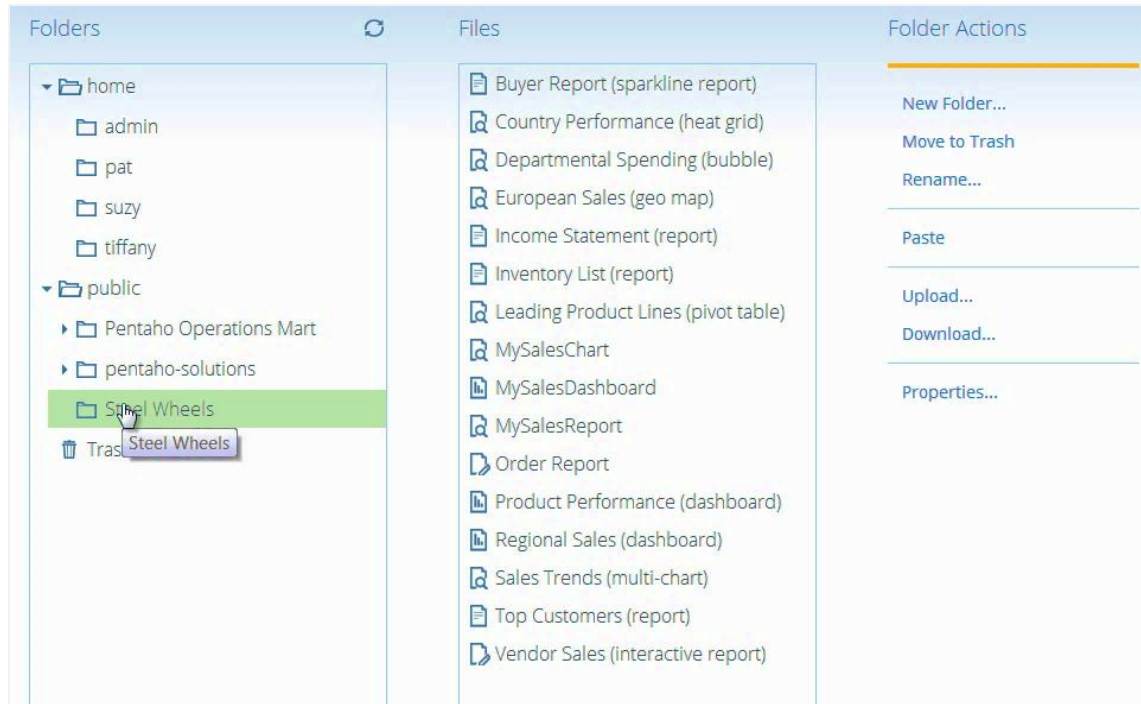


Figura 101. Menú rápido, Browse Files. Fuente: elaboración propia

En él encontramos tres columnas principales. La primera de ellas corresponde a **Folders** que nos dará acceso a las carpetas donde se encuentran alojados los archivos, y donde vemos se sitúan las dos carpetas raíces, **home** y **public**. En la primera alojaremos los ficheros de trabajo personal, mientras que en la segunda situamos los espacios de trabajo colaborativos. En la columna central de **Files**, aparecerán los ficheros incluidos en cada una de las carpetas y en la tercera y última, **Folder Actions**, las acciones disponibles para cada una de ellas.

La opción **Schedules**, **Figura 102**, nos da acceso a las funciones de programación de ejecución de informes en intervalos regulares, así como establecer envíos de informe mediante correo electrónico de forma automatizada, pausar o eliminar un programa etc.

Schedules ▾ admin ▾

Manage Schedules

Schedule Name ▾	Repeats	Source File	Output Location	Last Run	Next Run	Created By	Status
UpdateAuditData	Every 30 minutes at 10:10:55 PM	/public/pentaho-operations-mart/update_audit_mart_data/U	/home/admin	-	Nov 21, 2017 6:10:55 PM	admin	Paused
FillAuditMartTime	Every 10 seconds at 10:10:55 PM	/public/pentaho-operations-mart/update_audit_mart_data/F	/home/admin	-	-	admin	Paused
PentahoSystemVersionCh	Every day at 1:26:26 PM	PentahoSystemVersionCheck	-	Nov 21, 2017 1:26:26 PM	Nov 22, 2017 1:26:26 PM	admin	Normal

Figura 102. Menú rápido Schedules. Fuente: elaboración propia

La última pestaña de **Administration**, **Figura 103**, nos permite acceder a distintas opciones útiles de administración, como control de acceso para usuarios, permisos, roles, licencias etc.

Users & Roles

Authentication >

Mail Server >

Licenses >

Settings >

Users / Roles

Manage Users Manage Roles System Roles

Users: + x

admin
pat
suzy
tiffany

Password:

Edit...

Roles

Available:
Business Analyst
Power User
Report Author

Selected:
Administrator

Figura 103. Menú rápido Administration. Fuente: elaboración propia

Una vez visualizadas las distintas opciones, pasamos a revisar las partes de las que se compone el espacio de trabajo para formular los informes y analíticas, **Figura 104**.

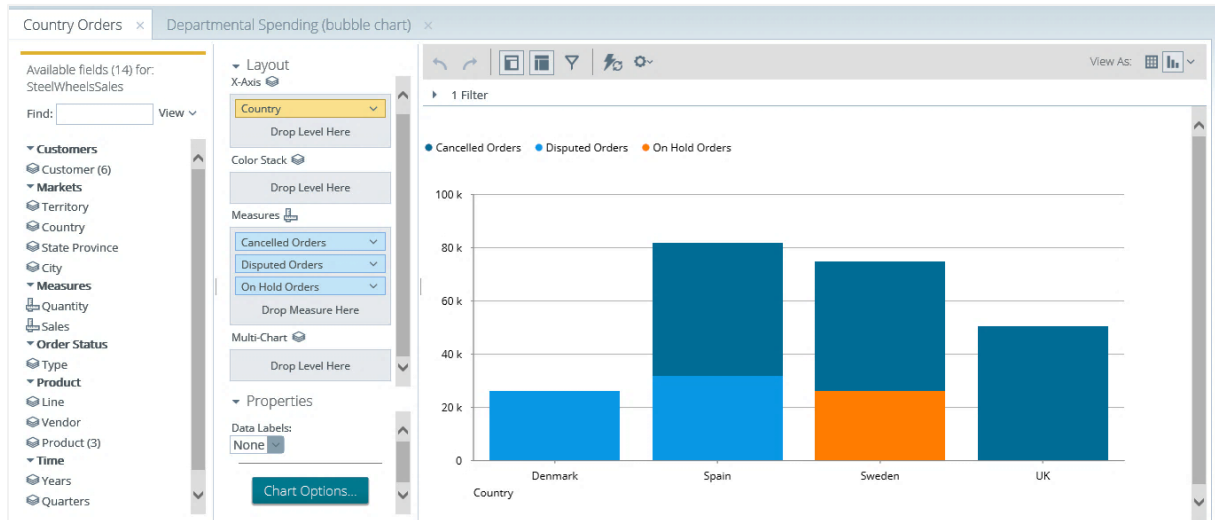


Figura 104. Vista general espacio de trabajo. Fuente: elaboración propia

De izquierda a derecha encontramos en primer lugar con la columna de dimensiones y medidas que hemos cargado desde la fuente de datos. En este caso vemos variables como dimensión, tipo, vendedor, producto etc. En segundo lugar, la columna de **Layout** donde encontramos una serie de cajas para situar los elementos que queremos analizar, como estas dimensiones que mencionamos o los elementos que deben aparecer en la leyenda de los ejes, marcas de color y demás aspectos necesarios para generar el gráfico. Por último, aparecería el espacio de visualización donde se conformará nuestro gráfico, en este caso un gráfico de barras por colores en función de países que cuenta con leyenda y en su parte superior un menú de opciones extra, que nos permitirá cambiar entre vistas de querer cambiar este último.

3.4.4.3 Visualizaciones

Las visualizaciones disponibles para la elaboración de informes en Pentaho no son demasiado abundantes, pero resultan suficientes para la mayor parte de los análisis que queramos realizar,

Figura 105.



Figura 105. Vistas de gráfico disponibles en Pentaho. Fuente: elaboración propia

Algunos ejemplos de los resultados que podemos conseguir utilizando las opciones disponibles, **Figura 106.**



Figura 106. Ejemplos gráficos con Pentaho. Fuente: elaboración propia

Combinando estos elementos se pueden conseguir *Dashboards* interactivos. A su vez, la plataforma permite también al usuario la creación de reportes profesionales mediante su herramienta **Pentaho Report Design** que ya hemos mencionado.

3.4.4.4 Creación de funciones y variables, predicciones, clúster y R

Agrupamos en este caso estos dos puntos, pues el tratamiento que hace de *Pentaho* de estas funciones difiere bastante de lo que ocurre con las otras herramientas analizadas. Cuando cargamos los datos desde la fuente definida, sea esta una base de datos externa, interna o un archivo local, se generarán una serie de dimensiones con las que trabajamos para realizar nuestras analíticas, informes etc. Desde la propia interfaz de trabajo para la ejecución de procesos no podemos generar nuevas variables, ni tampoco modificar las ya existentes (más allá de cambiar su nombre en el proceso de carga anteriormente visto). Para la ejecución de estos procesos es necesario acceder a los datos a través del *PDI*, desde donde sí existen multitud de procesos preconfigurados de transformaciones de datos, como sumas, divisiones, separaciones, mapeos, etc. Esto puede observarse en la captura siguiente, **Figura 107.** Esto supone sin duda alguna una enorme traba para muchos usuarios, puesto que el uso de la herramienta *PDI* es bastante compleja, contando con una curva de aprendizaje mucho más elevada que la de las interfaces de *BI* habituales.

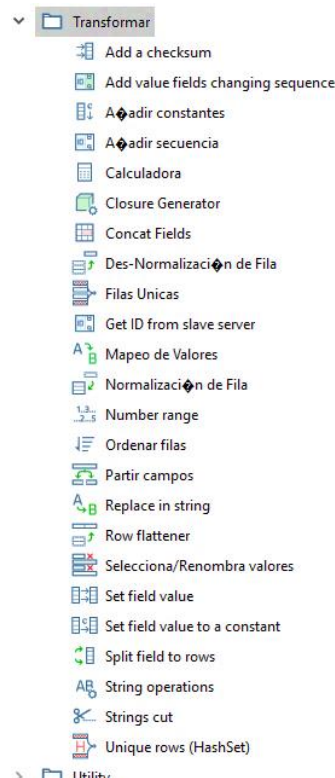


Figura 107. Opciones de transformación de datos en Pentaho PDI. Fuente: elaboración propia

Lo mismo ocurre para los procesos de predicción, estadística y R. Estos pueden realizarse, contando además con multitud de opciones para ello, pero requiere del uso de la herramienta *PDI*

que puede suponer una barrera extra para el usuario sin demasiado dominio de la misma, además de que, aun manejándolo con soltura, nos obligue a usar más tiempo del que necesitaríamos con cualquier otra alternativa de las analizadas.

3.4.4.5 Exportación de datos

La exportación de datos está permitida en formatos PDF, CSV y Excel, accesible desde el menú de trabajo interactivo.

3.4.5 Seguridad, volumetría y Big Data

En lo referente a la **seguridad**, la herramienta permite su administración directamente desde la plataforma o a través de proveedores externos. Ya hemos observado que la interfaz web cuenta con un sistema integrado para gestionar los permisos dados a usuarios, accesos y creación de roles. A nivel externo, cuenta con las siguientes funcionalidades:

- LDAP
- Single Sign-On
- Microsoft AD
- JDBC

La **volumetría** supone uno de los puntos fuertes de la plataforma, puesto que el propio servidor permite la configuración de un clúster junto con sistemas de balanceo de carga para distribuir mejor las tareas, agilizando los procesos. Por la parte del PDA, permite el manejo de grandes volúmenes de datos, permitiendo la conexión a Hadoop, Spark, Cassandra, MongoDB etc. Cuenta con sistemas propios para el tratamiento y mezcla de datos, contando con un amplio rango de conexiones y capas adaptativas, que hacen de la plataforma un elemento muy flexible ante cambios del entorno.

La posibilidad de ejecutar procesos de programación paralela en clústeres, algo especialmente útil para entornos de **Big Data**, es sin duda alguna uno de los estándares de la aplicación y elemento diferenciador respecto de buena parte de su competencia. A través del **Pentaho Data Integration** la herramienta nos dará acceso a multitud de funciones que van más allá de las propias de una herramienta con funcionalidad exclusiva para **BI**, pero de la cual puede beneficiarse enormemente.

Desde esta herramienta se pueden realizar, por ejemplo, ejecuciones EMR y Hive con Amazon, MapReduce nativo, Scripts Pig, trabajar con Spark, Sqoop, etc.

4 Conclusiones

La revisión de las herramientas de *BI* estudiadas nos ha proporcionado una buena visión general de la situación en el mercado de suites de inteligencia de negocio en el presente y cuál puede ser su evolución en el futuro.

Las empresas desarrolladoras plantean un mercado muy parejo, donde la alta competitividad y madurez del sector las obligan a desarrollar multitud de versiones de sus programas, tratando de abarcar el mayor abanico posible de usuarios. En todas las herramientas encontramos las versiones de escritorio tradicionales y alrededor de ellas un conjunto de *sub-herramientas* que vienen a cumplimentarla. En todas ellas es visible la preocupación de los desarrolladores por proveer de un ecosistema completo, donde las máximas son el trabajo colaborativo, la interconexión de aplicaciones y la búsqueda de facilidad de uso.

En cuanto a las **arquitecturas**, **Figura 108**, las estructuras de cliente servidor, con sus similitudes y diferencias, dominan claramente. En un entorno de datos creciente todas las soluciones buscan incorporar los mecanismos suficientes que garanticen su capacidad para crecer y escalar, asegurando la operatividad de la aplicación, por lo que es común el uso de múltiples clústeres en sus estructuras, que en todas las analizadas permiten la utilización de las mismas para mejorar los procesos de consulta y ejecución de procesos típicos de *BI*, pero no todas para ejecución de procesos de distribución paralela. Es en este último punto donde las herramientas *PowerBI* y *Pentaho* destacan por encima de las otras dos, permitiendo la ejecución de procesos con distribución paralela en múltiple clúster, y no solo para la visualización.



Figura 108. Orden Arquitectura. Fuente: elaboración propia

La solución *arrastrar-soltar*, **Figura 109**, tiene completamente monopolizado el mercado. En la era presente, donde lo táctil ha pasado a ocupar la mayor parte de nuestra interacción tecnológica, el uso de esta fórmula para la ejecución de los gráficos resulta fácil e intuitiva para el usuario. Si bien es cierto que todas acuden a este sistema como hilo conductor del proceso creador, algunas herramientas como *Tableau* la llevan a un punto mayor que el resto, donde. Hemos observado el gran número de espacios, denominados cajones, disponibles en ella para que el usuario arrastre, no solo los gráficos preconfigurados, sino las propias variables y dimensiones. Estas variables y dimensiones las situará, o bien en estos cajones y espacios, que pueden ser por ejemplo los representativos de los ejes de coordenadas del gráfico, como directamente los espacios para filtros. Es en esta herramienta, donde el uso de submenús resulta menos necesario, pudiendo prácticamente ejecutar cualquier gráfico sin necesidad de ellos, ganando en agilidad e intuición. Sería esta aplicación la que, a nuestro juicio, logra este efecto con mayor intensidad. En el resto de las aplicaciones esta característica tiene también un papel importante, pero resulta necesario el uso de más menús y submenús para conseguir los mismos resultados. De las cuatro herramientas estudiadas sería sin duda alguna *Pentaho* la que saca menos rendimiento a esta característica.



Figura 109. Orden Facilidad uso. Fuente: elaboración propia

La **creación de variables**, *Figura 110*, puede suponer un auténtico problema en función de lo que necesitemos y en este sentido, especialmente para el usuario sin conocimientos técnicos, todas las facilidades son pocas. En este sentido es donde *PowerBI*, tanto por el esquema seguido para realizando como por contar con DAX, lleva clara ventaja, y es que la creación de nuevas variables resulta mucho más sencilla. Se le ofrece con ella al usuario muchas ayudas mediante la predicción de acciones, por ejemplo, ver cómo debe construir la sentencia para hacer una suma antes de que termine de realizarla. Con *Qlik*, pese a que esta característica es menos potente, nos ocurre algo parecido, resultando muy intuitivo. *Tableau* pierde puntos en este aspecto, y es que puede ser necesario realizar scripts para resolver pequeños problemas, como es la gestión de datos nulos, donde hemos visto ser necesario contar con conocimientos de programación con algún *bucle* y sentencias *if/else* para sortearlo. Pero sin duda, donde mayor complejidad para el usuario encontramos es con *Pentaho*, pues esta herramienta no cuenta con funcionalidad para la creación de variables una vez cargados los datos (momento en que se crean las dimensiones). Será para esta aplicación necesaria, como hemos visto, el uso de una herramienta auxiliar, el *PDI*, cuyo grado de complejidad en el aprendizaje es mayor, amén del tiempo extra que requiere el tener que cambiar de una herramienta a otra para hacer algo, que según el usuario/estudio, puede ser necesario realizar de forma repetida.



Figura 110. Orden Facilidad con dimensiones. Fuente: elaboración propia

La **interfaz**, *Figura 111*, es un aspecto muy importante a tener en cuenta a la hora de elegir la suite adecuada para nuestra organización, y es que en buena medida dependerá de esta la curva de aprendizaje de los usuarios de la misma. En este sentido *Tableau* marca claramente la diferencia, y es que por contra de lo que ocurre con sus rivales en casi ningún momento hemos necesitado usar manuales para aprender a usarla, pudiendo ejecutar funciones complejas rápidamente de

forma intuitiva. PowerBI, pese a ser más compleja que la anterior, cuenta con la ventaja de tener un modelo parecido al de la suite de Office, por lo que salva ligeramente el escollo, además de ser bastante ágil en los procesos. Qlik se sitúa al mismo nivel que PowerBi, siendo agradable el trabajo con la misma. La interfaz de Pentaho es también eficiente, aunque en términos estéticos se encuentra menos cuidada que sus competidoras, y el hecho de trabajar en el navegador puede suponer algún problema extra por configuración. Cuenta con menús grandes y claros, con las distintas funciones desempeñadas en cada sección bien delimitada y organizada.



Figura 111. Orden Atractivo interfaz. Fuente: elaboración propia

En el apartado de **visualizaciones**, **Figura 112**, podemos decir que las tres primeras proponen un escenario muy parecido, ofreciendo una amplia variedad de gráficos y visualizaciones, no destacando ninguna por encima de las demás en cuanto a cantidad. Pentaho cuenta con menos opciones para realizar gráficos, y estos resultan menos atractivos visualmente hablando, pero cuenta con los suficientes para hacer la mayoría de las presentaciones necesarias. En lo que a diseño se refiere, si podríamos decir que tal vez el resultado final de la aplicación Qlik parezca más profesional que las otras, pero se trata de un elemento subjetivo difícilmente evaluable. Cerramos este punto con una reflexión, y es que para un usuario inexperto probablemente la capacidad de *Tableau* para hacer simple la construcción de las visualizaciones y lo “acotado” de su proceso, si bien es muy útil al principio, supone un lastre cuando el usuario se vuelve experto. Y esto es así porque no dispone de los medios para permitir al usuario experto la creación, mediante APIs de nuevos tipos de visualización, y es ahí donde un usuario maduro en la herramienta se sentirá “pobre” con *Tableau*, algo que no le ocurriría con Qlik, por ejemplo, o con PowerBI y Pentaho, más teniendo en cuenta de esta última el “extra” de funciones que contiene más allá de las propias de una herramienta de BI. En este sentido podríamos concluir que, por lo menos para las herramientas estudiadas, la simplicidad de la interfaz viene acompañada de un menor número

de opciones y posibilidades, por lo que necesariamente cuando queramos hacer acciones más complejas tendremos que pivotar hacia plataformas que también lo sean. En este sentido el orden de prelación quedaría sigue, **Figura 112**.



Figura 112. Orden Atractivo visualizaciones. Fuente: elaboración propia

Tanto en los procesos de **exportación de datos** como en las **opciones de seguridad, disponibles**, no encontramos diferencias significativas que nos hagan poder decantarnos por una opción u otras.

En el último apartado, dedicado a la **volumetría** de datos es donde, como ya hemos visto, encontramos un punto de corte muy importante. Todas las herramientas estudiadas cuentan con acceso a multitud de fuentes de datos, pudiendo combinarlas, trabajar en memoria o no, etc. Sin embargo, tan solo con *Pentaho* y *PowerBi* podemos dar un paso más allá. Estas dos aplicaciones nos permiten dar un salto cualitativo, dándonos pie a trabajar de forma directa con volúmenes de datos verdaderamente grandes, a través de sus sistemas de conexión, pero sobre todo a través de la posibilidad de ejecutar distribución paralela en múltiples clústeres, con lo que nos permiten ejecutar scripts en Pig o procesos de MapReduce, pudiendo trabajar con auténtico **Big Data**.

Se trata de un punto clave en el mundo en que nos encontramos, y cada vez más un elemento vital para el mundo que nos viene, donde la avalancha de datos será cada vez mayor y la capacidad para capturarlos y tratarlos un elemento vital. Encontramos con esto una buena justificación de por qué *Pentaho* sin estar entre las primeras herramientas para *Gartner* si está generando gran revuelo entre las comunidades de usuarios y medios especializados, **Figura 113**.



Figura 113. PowerBI y Pentaho, soluciones de BI con funcionalidad BigData. Fuente: elaboración propia.

Referencias Bibliográficas

- Amazon Web Services. (2017). ¿Qué es una base de datos relacional? – Amazon Web Services (AWS). Obtenido November 25, 2017, from <https://aws.amazon.com/es/relational-database/>
- Carlos Paramio. (2011). El concepto NoSQL, o cómo almacenar tus datos en una base de datos no relacional. Obtenido November 25, 2017, from <https://www.genbetadev.com/bases-de-datos/el-concepto-nosql-o-como-almacenar-tus-datos-en-una-base-de-datos-no-relacional>
- Ecured.cu. (2016). Arquitectura Cliente Servidor - EcuRed. Obtenido November 25, 2017, from https://www.ecured.cu/Arquitectura_Cliente_Servidor
- Gartner Inc. (2017). Business Intelligence - BI - Gartner Glosario de TI. Obtenido November 16, 2017, from <https://www.gartner.com/it-glossary/business-intelligence-bi/>
- Hewlett Packard Enterprise. (2017). ¿Qué es un OEM? – Glosario de definiciones de HPE | HPE™ España. Obtenido November 25, 2017, from <https://www.hpe.com/es/es/what-is/oem.html>
- IBM. (2017). Big Data Analytics | IBM Analytics. Obtenido November 25, 2017, from <https://www.ibm.com/analytics/hadoop/big-data-analytics>
- Luhn, H. P. (1958). A Business Intelligence System. Obtenido de <http://altaplana.com/ibmrd0204H.pdf>

Bibliografía General

- Intelligence, B., Data, B., Systems, H. D., Analytics, P., Etl, K., Integration, D., & Data, B. (2015).
2. Estudio del estado del arte de herramientas de Business Intelligence y su relación con Big Data., 10–35.
- BSA, T. S. A. (2015). ¿Por qué son tan importantes los datos? Obtenido de http://data.bsa.org/wp-content/uploads/2015/10/BSADataStudy_es.pdf
- androidphoria.com. (2017). ¿Cuántos WhatsApp se mandan al día? Obtenido en November 25, 2017, de <https://androidphoria.com/novedades/cuantos-whatsapp-se-mandan-al-dia>
- microfocus.com. (2017). How Much Data is Created on the Internet Each Day? | Micro Focus Blog. Obtenido en November 25, 2017, from <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>
- silicon.es. (2013). Dropbox registra mil millones de archivos cada día. Obtenido en November 25, 2017, de http://www.silicon.es/dropbox-registra-mil-millones-de-archivos-subidos-cada-dia-2232219?inf_by=59f66f3c681db8ba508b49c3
- El Business Intelligence y su evolución con el tiempo | Blog Conento. (2016). Obtenido en November 16, 2017, de <http://blog.conento.com/es/el-business-intelligence-y-su-evolucion-con-el-tiempo/>
- Luhn, H. P. (1958). A Business Intelligence System. Obtenido de <http://altaplana.com/ibmrd0204H.pdf>
- ESADE. (2017). BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN. Obtenido de http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf
- spaceanalytics.blogspot.com.es. (2017). Ingestión de Datos - Observatorio BI & Analytics. Obtenido October 13, 2017, from <http://spaceanalytics.blogspot.com.es/2016/09/ingestion-datos.html>

Sallam, R. L., Howson, C., Idoine, C. J., Oestreich, T. W., Richardson, J. L., & Tapadinhas, J. (2017). Magic Quadrant for Business Intelligence and Analytics Platforms ID: G00301340 Analyst(s). Obtenido de https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&mkt_tok=eyJpIjoiWmpZMU1tUTF0ell3Wm1VeCIsInQiOiIyS2dJUVhsNmtRcFZrT...