



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Análisis de envíos a la plataforma de
Suministro de Información Inmediata de
Hacienda con herramientas de Business
Intelligence.

Trabajo Fin de Máster

Máster Universitario en Gestión de la Información

Autor: Álvaro Durá de Lamo

Tutor: César Ferri Ramírez

2018/2019

Análisis de envíos a la plataforma de Suministro de Información Inmediata de
Hacienda con herramientas de Business Intelligence.

Agradecimientos

En primer lugar, me gustaría dar las gracias a mi familia por apoyarme en todo momento durante toda mi carrera y en especial a mis padres, por su apoyo incondicional.

También me gustaría agradecer a la empresa AHORA la oportunidad de realizar este TFM utilizando parte de sus datos y el apoyo y facilidades recibidas durante estos dos años para poder compaginar los estudios y la vida profesional.

Por último, me gustaría agradecer a César Ferri su apoyo a la hora de realizar este trabajo y la dedicación y el tiempo empleados en supervisar este proyecto.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de
Hacienda con herramientas de Business Intelligence.

Resumen

El aumento de la cantidad de datos que generan las empresas día a día es exponencial. Es por ello, que actualmente, las empresas están considerando la posibilidad de obtener información de estos datos que apoye a sus decisiones empresariales. Para ello es necesario el uso de soluciones de Inteligencia de Negocio, que sean capaces de manejar gran cantidad de datos.

La Agencia Tributaria obliga a ciertas empresas, según una serie de requisitos, a enviar sus facturas de forma telemática. Mediante la plataforma SII (Sistema de Información Inmediata), las empresas pueden enviar las facturas a la Agencia Tributaria, de una en una mediante un portal WEB o varias a la vez mediante una API WEB. La empresa AHORA, empresa especializada en la fabricación de software para gestión de grandes y medianas empresas, ha desarrollado un módulo para su software AHORA ERP, que permite a las empresas que utilizan dicho software enviar facturas a la Agencia Tributaria mediante el SII, todo ello a través de un servidor intermedio dónde la empresa guarda los envíos.

Después de un año en producción de esta plataforma, la empresa AHORA ha recopilado más de 30.000 envíos y alrededor de un millón de facturas. Este trabajo final de máster aborda el análisis, diseño e implementación de una solución de Inteligencia de Negocio que permite obtener información del funcionamiento del módulo SII de AHORA ERP para mejorar la toma de decisiones de negocio.

Palabras clave: Inteligencia de negocio, SII, Agencia Tributaria, Aplicación WEB, almacén de datos.

Abstract

The increase in the amount of data generated by companies day by day is exponential. That is why, currently, companies are considering obtaining information from this data that supports their business decisions. This requires the use of Business Intelligence solutions, which are capable of handling large amounts of data.

The Tax Agency obliges certain companies, according to a series of requirements, to send their invoices electronically. Through the SII platform (Immediate Information System), companies can send invoices to the Tax Agency, one at a time through a web portal or several at the same time through a web API. The company AHORA, a company specialized in the manufacture of software for the management of large and medium-sized companies, has developed a module for its software AHORA ERP, which allows companies that use this software to send invoices to the Tax Agency through an intermediate server where the company saves the shipments.

After a year in production of this platform, the company AHORA collected more than 30,000 shipments and about one million invoices. This final master's project deals with the analysis, design and implementation of a Business Intelligence solution that allows obtaining information on the operation of the SII module of AHORA ERP to improve business decision making.

Keywords : Business Intelligence, SII, Tax Agency, Web application, data warehouse.



Análisis de envíos a la plataforma de Suministro de Información Inmediata de
Hacienda con herramientas de Business Intelligence.

Tabla de contenidos

1.	Introducción.....	1
1.1.	Inteligencia de negocio o BI.....	1
1.1.1.	Casos de éxito	2
1.2.	Suministro Inmediato de Información del IVA.....	3
1.3.	Motivación	4
1.4.	Objetivos	4
1.5.	Estructura de la memoria.	5
1.5.1.	Capítulo 2. Situación tecnológica.	5
1.5.2.	Capítulo 3. Análisis de la solución.....	5
1.5.3.	Capítulo 4. Desarrollo.....	5
1.5.4.	Capítulo 5. Despliegue.....	6
1.5.5.	Capítulo 6. Conclusión.	6
2.	Situación tecnológica	7
2.1.	Almacén de datos	7
2.1.1.	SQL Server	9
2.1.2.	MySQL.....	10
2.1.3.	Oracle DB.....	11
2.2.	Integración de datos	11
2.2.1.	Pentaho Data Integration (Kettle).....	12
2.2.2.	Microsoft SQL Server Integration Services	13
2.2.3.	Mule ESB	13
2.3.	Cubos multidimensionales OLAP.....	14
2.3.1.	Microsoft SQL Server Analysis Services.....	16
2.3.2.	Analytic Workspace Manager.....	17
2.3.3.	Pentaho Mondrian Analysis Services	18
2.4.	Visualización de datos.....	18
2.4.1.	Shiny R	18
2.4.2.	Microsoft Power BI.....	19
2.4.3.	Qlick Sense	20
2.4.4.	Tableau	21
2.5.	Empresa Ahora	22
3.	Análisis de la solución	23
3.1.	Obtención de requerimientos	23
3.2.	Tecnologías	24



Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

3.2.1.	Git	24
3.2.2.	GitLab	24
3.2.3.	SQL Server Management Studio 2017.....	25
3.2.4.	SQL Server	26
3.2.5.	SQL Server Integration Services	26
3.2.6.	SQL Server Analysis Services	26
3.2.7.	Power BI	26
3.2.8.	R.....	27
3.2.9.	Shiny R	27
3.2.10.	Shiny Dashboard	27
3.2.11.	Microsoft Visual Studio 2017	27
3.3.	Planificación	28
3.4.	Análisis funcional.....	30
4.	Desarrollo.....	34
4.1.	Almacén de datos.....	34
4.1.1.	Estudio de estructura de tablas	34
4.1.2.	Implementación de proyecto de base de datos	35
4.2.	Integración de datos	41
4.2.1.	Obtener el último envío registrado.....	42
4.2.2.	Borrar tablas intermedias.....	42
4.2.3.	Cargar tablas intermedias.	42
4.2.4.	Obtener dimensiones.....	43
4.2.5.	Obtener dimensión envío.	44
4.2.6.	Integrar tablas de hechos.	44
4.2.7.	Tarea de procesamiento de Analysis Services.	45
4.3.	Análisis de datos	45
4.3.1.	Orígenes de datos	46
4.3.2.	Vista del origen de datos.	46
4.3.3.	Cubos	47
4.3.4.	Dimensión Tiempo.	49
4.3.5.	Atributos, jerarquías y relaciones de atributos.	50
4.3.6.	Cálculos.....	50
5.	Despliegue.....	52
5.1.	Shiny R.....	52
5.1.1.	Instalación de Shiny R.....	52
5.1.1.	Estructura de la aplicación.	53



5.1.2.	Capturas Aplicación WEB Shiny R.....	54
5.2.	PowerBI	55
5.2.1.	Instalación de PowerBI	55
5.2.2.	Capturas de Aplicación WEB PowerBI.....	56
5.2.3.	Comparación entre los dos sistemas.	58
6.	Conclusión.....	59
6.1.	Dificultades	59
6.2.	Líneas futuras	59

Índice de figuras

Ilustración 1. Jerarquía del BI.	2
Ilustración 2 Esquema de estrella	8
Ilustración 3 Esquema de copo de nieve	8
Ilustración 4. Arquitectura de SQL Server	9
Ilustración 5 MySQL Workbench	11
Ilustración 6. Pentaho Data Integration	12
Ilustración 7. Mule ESB.....	14
Ilustración 8. Cubo multidimensional	15
Ilustración 9. Sistema ROLAP	15
Ilustración 10. Sistema ROLAP	16
Ilustración 11. Sistema HOLAP	16
Ilustración 12. SQL Server Analysis Services	17
Ilustración 13. Analytic Workspace Manager.....	17
Ilustración 14. Pentaho Mondrian.....	18
Ilustración 15. Shiny R.....	19
Ilustración 16. Power BI.	20
Ilustración 17. Qlick Sense.....	21
Ilustración 18. Tableau.	22
Ilustración 19. Estructura API SII AHORA	23
Ilustración 20. Git	24
Ilustración 21. GitLab	25
Ilustración 22. SQL Server Management Studio.....	26
Ilustración 23. Shiny Dashboard.....	27
Ilustración 24. Microsoft Visual Studio 2017.....	28
Ilustración 25. Metodología Scrum	29
Ilustración 26. Planificación del proyecto.	30
Ilustración 27. Mockup Pantalla principal	31
Ilustración 28. Mockup resumen de envíos.	32
Ilustración 29. Mockup resumen de facturas.....	33
Ilustración 30. Diagrama de almacén de datos.....	35
Ilustración 31. Tabla DimClaveRegimenEspecialOTrascendenciaEmitida	36
Ilustración 32. Tabla DimClaveRegimenEspecialOTrascendenciaRecibida.....	36
Ilustración 33. Tabla DimEstado.....	36
Ilustración 34. Tabla DimPais	37
Ilustración 35. Tabla DimPeriodoLiquidacion.....	37
Ilustración 36. Tabla DimTipoFacturaEmitida.	37
Ilustración 37. Tabla DimTipoFacturaRecibida.....	38
Ilustración 38. Tabla DimEnvio.	38
Ilustración 39. Tabla FactFacturaEmitida.	39
Ilustración 40. Tabla FactFacturaRecibida	40
Ilustración 41. Proyecto de base de datos y solución.	40
Ilustración 42. Tablas del proyecto DW	41
Ilustración 43. Tablas Auxiliares.....	41
Ilustración 44. Proyecto SSIS	42
Ilustración 45. Proceso de carga de tablas auxiliares.....	43

Ilustración 46. Flujo de datos de envíos pendientes	43
Ilustración 47. Obtención de dimensiones y scripts.	44
Ilustración 48. Proceso de integración de facturas	44
Ilustración 49. Esquema Proceso de Integración.....	45
Ilustración 50. Proyecto SSAS multidimensional.	46
Ilustración 51. Asistente de conexión	46
Ilustración 52. Asistente vista del origen de datos.....	47
Ilustración 53. Asistente de cubo. Medidas.....	48
Ilustración 54. Asistente de cubo. Dimensiones.	48
Ilustración 55. Cubo y Dimensiones.....	49
Ilustración 56. Resumen asistente de dimensión fechas.	49
Ilustración 57. Jerarquías	50
Ilustración 58. Relaciones de atributos.....	50
Ilustración 59. Cálculos.	51
Ilustración 60. R y R Studio.	53
Ilustración 61. Página principal Shiny R.....	54
Ilustración 62. Página envíos Shiny R.....	54
Ilustración 63. Pagina facturas Shiny R.....	55
Ilustración 64. Página principal PowerBI	56
Ilustración 65. Empresas PowerBI.....	57
Ilustración 66. Facturas PowerBI.....	57

Análisis de envíos a la plataforma de Suministro de Información Inmediata de
Hacienda con herramientas de Business Intelligence.

1. Introducción

En este primer apartado se dará a conocer al lector el Trabajo de Fin de Máster en Gestión de la Información realizado. En la primera sección se hablará sobre el contexto actual relacionado con el proyecto. El segundo apartado hablará sobre la motivación que ha llevado a realizar este trabajo. Seguidamente, se definen los objetivos a alcanzar con el proyecto y, por último, se detallará la estructura completa de la memoria.

1.1. Inteligencia de negocio o BI.

La tecnología avanza día a día de manera exponencial, esto hace que cada vez sea más fácil y barato para las empresas almacenar grandes cantidades de datos, como pueden ser los datos de facturación, de fabricación, etc....

Hasta hace poco, para las empresas estos **datos** no eran más que valores dentro de sus sistemas que carecían de sentido en sí mismos. Es decir, un dato por ejemplo podría ser el valor de una venta de un artículo en un momento concreto, *500* por ejemplo.

Cuando se dispone de suficientes datos, se pueden transformar en **información**, dónde se dota a los datos de un contexto específico. Por ejemplo, si se tienen todas las ventas que se han realizado durante un mes, se pueden sumar todos los valores y dividir entre la cantidad de ventas obteniendo que *el valor medio de una venta en el mes de junio es 435,82*.

Estas dos visiones acerca de los datos se engloban dentro de los **sistemas tradicionales de información**. Estos son sistemas en los que se desarrollan informes estáticos a partir de los datos y la información.

El siguiente paso es dotar de **conocimiento** a esta información, dónde gracias a esto se puede utilizar para generar informes que ayuden a los usuarios en la toma de decisiones. Por ejemplo, se puede comparar la media de ventas de un mes con la del mes anterior y decir que *en el mes de junio el valor de la venta media está un 5% por encima del mes anterior*.

Es este momento es donde entran en juego las distintas herramientas y procesos de Inteligencia de Negocio, nombrado de ahora en adelante BI. Estas herramientas permiten a partir de la información y el conocimiento aumentar la **competitividad** de las empresas en el mercado.

Como se ha comentado anteriormente, los datos que almacenan las compañías son cada vez más y esto junto a las tecnologías que apoyan y permiten los distintos procesos de BI, está haciendo que el valor de las empresas, las estrategias a seguir y las tomas de decisiones estén directamente proporcionadas a la cantidad de datos almacenados y la información generada a partir de estos.

El conocimiento y la competitividad comprenden lo que hoy en día conocemos como BI.



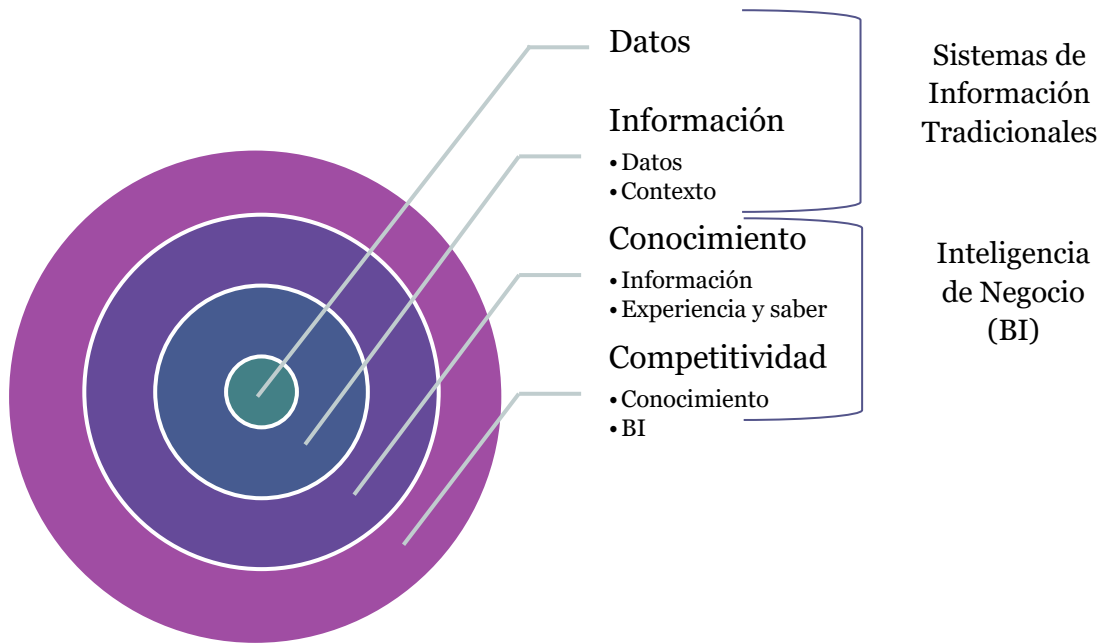


Ilustración 1. Jerarquía del BI.

1.1.1. Casos de éxito

Hoy en día todas las grandes empresas y cada vez más las pequeñas y medianas empresas tienen implantado y funcionando un sistema de BI que permite agilizar los procesos, tanto internos como externos, mejorar el servicio al cliente, detectar anomalías y en resumen aumentar los beneficios de la empresa gracias al estudio de manejo de sus datos.

A continuación, se presentan dos casos de éxito de la implantación de sistemas BI:

AMAZON

Amazon utilizó tecnologías de BI para detectar el fraude de las tarjetas de crédito y débito llegando a obtener una reducción de 50% en los fraudes online. [1]

STAPLES

Esta empresa mueve semanalmente un total de 10 millones de transacciones, lo que ha permitido a través de tecnologías de BI crear campañas de marketing personalizadas a sectores de la población. [1]

NETFLIX

Con un total de más de 53 millones de clientes en todo el mundo Netflix tiene una cantidad ingente de datos y mediante técnicas de BI, son capaces de realizar tomas de decisiones con exactitud milimétrica a la hora de contratar nuevas series y/o películas, ya que a partir de cientos de variables de cada usuario, como el dispositivo, las veces que se ve una serie, si rebobinamos, la fecha exacta del visionado y un largo etc, predicen que se verá en los próximos meses por cada tipo de usuario. [2]

UPS

Una de las grandes compañías de transportes mundiales, UPS, se apoyó de la inteligencia de negocio para optimizar las rutas que sus repartidores realizaban diariamente. Gracias a esto, consiguieron reducir en 2004 y 2012 unos 25 millones de dólares al año, además de una reducción de 10 millones de galones de gas, aumentando así las ganancias de la empresa y ayudando a mejorar la calidad de vida en el planeta. [2]

TESCO

A partir de mediados de los 90, TESCO, uno de los grandes minoristas del Reino Unido, comenzó a implantar en sus sistemas el BI. Lo hizo empezando con las tarjetas de fidelización, a partir de las cuales se comenzó a procesar la información que estas tarjetas generaban para mejorar la orientación de los vales descuento, aumentando la redención de estos de un 3 a un 70%. [2]

Después de esto, consiguieron ahorrar más de 100 millones de libras mediante el procesamiento de ventas históricas y los datos meteorológicos. Con ello optimizaron el stock y los envíos de los productos a sus distintas tiendas.

1.2. Suministro Inmediato de Información del IVA

El actual sistema de gestión del Impuesto sobre el Valor Añadido, en adelante IVA, está vigente desde hace 30 años o más. Este sistema se basa en la presentación periódica de modelos fiscales, que no son más que resúmenes de las compras y ventas de una empresa en un periodo de tiempo. Esta periodicidad varía según el tipo de modelo y de empresa y puede ser mensual, trimestral, semestral o anual.

Gracias a los avances tecnológicos de hoy en día, estos posibilitan la creación de un Suministro Inmediato de Información, en adelante SII, que permiten a la Agencia Tributaria, de ahora en adelante AEAT, llevar un control casi instantáneo de los registros de facturación de las empresas.

Este sistema consiste en un servicio web, donde las empresas, alguna de manera obligatoria y otras de manera voluntaria, envían con un plazo máximo de cuatro días a partir de la contabilización de la operación, los datos relevantes sobre el IVA.

El SII comenzó su funcionamiento el 1 de julio de 2017, afectando a un 80% de las empresas, y según la propia web de la AEAT supone las siguientes ventajas: [3]

- *“Se dispondrá de información de calidad en un intervalo de tiempo suficientemente corto como para agilizar el sistema de gestión del IVA.*
- *Obtención de “Datos Fiscales”, ya que el contribuyente dispondrá en la Sede electrónica de la AEAT de un Libro Registro “declarado” y otro “contrastado” con la información de contraste procedente de terceros que pertenezcan al colectivo de este sistema o de la base de datos de la AEAT.*
- *Los contribuyentes podrán contrastar dicha información antes de la finalización del plazo de presentación de su declaración mensual de IVA.*
- *El contribuyente tendrá la posibilidad de corregir los errores cometidos en los envíos sin necesidad de ser requerido por la AEAT para ello.*



- *Disminución de los requerimientos de información por parte de la AEAT, ya que muchos de los requerimientos actuales tienen por objeto solicitar los Libros registro, las facturas o datos contenidos en las mismas para comprobar determinadas operaciones.*
- *Moderniza y estandariza la forma de llevar los tradicionales Libros Registro de IVA.*
- *Reducción de las obligaciones formales, suprimiendo la obligación de presentación de los modelos 347, 340 y 390.*
- *Reducción de los plazos de realización de las devoluciones, al disponer la AEAT de la información en tiempo casi real y de mayor detalle sobre las operaciones.*
- *Los contribuyentes incluidos en el nuevo sistema verán ampliados en 10 días los plazos de presentación e ingreso de sus autoliquidaciones periódicas del IVA.”*

1.3. Motivación

Actualmente las empresas están comprendiendo cada vez más la importancia del uso del BI para comprender y estudiar sus datos.

La empresa AHORA (<http://www.ahora.es/>), empresa fabricante y distribuidora de software de gestión para PYMES y grandes empresas, ha desarrollado una solución para integrar los envíos de facturas al SII de sus clientes. Este sistema está basado en el envío de las facturas desde el servidor del cliente a un servicio web intermedio, que recoge los envíos de las distintas empresas y realiza los mismos a la AEAT en nombre de estas.

Dentro de este servidor intermedio se recogen todos los días decenas de envíos que suponen cientos o miles de facturas, lo que supone una gran cantidad de datos susceptibles de convertir en conocimiento a partir de las herramientas de BI.

Además, la propia empresa tiene la necesidad de controlar los envíos que realizan cada empresa por temas de facturación y de obtener una relación de los errores que se producen para optimizar la plataforma y mejorar la calidad hacia sus clientes.

Según el informe de Generación de Talento Big Data en España, “*el valor estimado de la economía del dato en Europa supuso un 1,87% del PIB de los países miembros en 2015 (272.000 millones de euros) y se prevé que alcance el 4,7% en 2020. Además, el 65% de las empresas corren el riesgo de convertirse en irrelevantes o no competitivas si no adoptan el Big Data. El mercado de Big Data en España crece un 30% cada año y empleó a 10.500 profesionales en el año 2015. Como primera advertencia, un 19% de esos puestos de trabajo se generaron fuera del país.*” [4].

Dado que dentro del máster se ven y estudian una serie de metodologías que permiten implantar soluciones de BI, la necesidad que tiene la empresa AHORA de controlar el funcionamiento de la aplicación del SII y la oportunidad laboral que presenta aprender a utilizar las herramientas de BI, motivan a la realización este este trabajo.

1.4. Objetivos

El objetivo principal de este trabajo de fin de máster consiste en crear una solución que permita a la empresa AHORA estudiar y conocer el funcionamiento de la plataforma del SII.

Dentro de este objetivo principal se plantean los siguientes objetivos secundarios:

- Estudiar y analizar el estado actual de la plataforma de SII en la empresa y obtener las distintas necesidades a partir de reuniones con la parte interesada.
- Crear a partir de estos requisitos en almacén de datos dónde almacenar toda la información que obtenemos de la plataforma intermedia.
- Generar un proceso de extracción, transformación y carga, a partir de ahora ETL, que permita obtener los datos de la plataforma origen, transformar dichos datos para su posterior inserción en el almacén de datos.
- Implementar a partir de los datos obtenidos en el almacén de datos un cubo multidimensional que permita el acceso de manera casi instantánea a los datos preprocesador.
- Crear una aplicación dónde el usuario puede realizar una visualización de los datos para ayudarle a la toma de decisiones. Dentro de este último punto, realizaremos la comparativa entre dos formas de realizar esta visualización. Una plataforma orientada al usuario menos técnico y otra plataforma donde se requiere de conocimientos de programación.

Por último, otro objetivo principal, pero a su vez ligado a todo el trabajo, es el aprendizaje de las metodologías, procesos y herramientas necesarios para llevar a cabo la implantación de un sistema de BI.

1.5. Estructura de la memoria.

La memoria del trabajo de fin de máster de gestión de la información “Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence”, se encuentra estructurada de la siguiente manera:

1.5.1. Capítulo 2. Situación tecnológica.

En este capítulo se estudian y explican las diferentes alternativas tecnológicas que podemos encontrar en el mercado para cada una de las partes en las que se divide el proyecto de BI. Además, se hablará de la empresa AHORA, de su trayectoria y de los distintos productos que desarrolla actualmente.

1.5.2. Capítulo 3. Análisis de la solución.

Dentro de este apartado crearemos un análisis de la solución de BI a partir de los requerimientos y requisitos proporcionados por la empresa AHORA. Dentro de este análisis encontraremos también las tecnologías escogidas, la planificación del proyecto y un análisis técnico de la solución.

1.5.3. Capítulo 4. Desarrollo.

En el capítulo de desarrollo, se expondrá como se ha desarrollado la solución, desde crear el almacén de datos a obtener un cubo multidimensional.

1.5.4. Capítulo 5. Despliegue.

En el penúltimo capítulo se realizará el despliegue del desarrollo mediante dos plataformas de visualización de datos, creando una comparativa entre ellas.

1.5.5. Capítulo 6. Conclusión.

En este último capítulo se engloban las conclusiones obtenidas del trabajo, las dificultades encontradas a la hora de crear el trabajo y los posibles trabajos futuros.

2. Situación tecnológica

En el siguiente capítulo se explicarán las distintas herramientas que se pueden utilizar para realizar cada una de las partes en la que consta el proyecto. Se verán también las tecnologías a utilizar transversales al proyecto, que no son clasificables en ninguna de las secciones anteriores.

Por último, se dará a conocer la empresa AHORA, empresa junto con la cual se ha realizado el proyecto, propietaria de los datos y usuario final de la aplicación generada en este trabajo.

2.1. Almacén de datos

Se define como almacén de datos, en adelante DW, cómo el resultado de aplicar una serie de técnicas para recopilar y administrar datos de una o diversas fuentes para proporcionar información significativa.

Es el almacenamiento de gran cantidad de información destinada al diseño de consultas y análisis en lugar de procesos transaccionales.

A diferencia de las bases de datos relacionales tradicionales, dónde los datos se normalizan, es decir, un campo de una tabla contiene un índice que se relaciona con otra tabla, reduciendo así el tamaño de la base de datos, en los DW los datos se desnormalizan, replicando datos en varias tablas, esto hace que aumente el tamaño, pero aumenta la velocidad en el acceso a los datos.

Las principales características de un sistema de almacén de datos para BI según Kimball y Ros [5] son:

- Un sistema DW debe hacer que la información sea accesible con facilidad
- La información debe ser consistente en un sistema de DW.
- Un sistema DW debe adaptarse a los cambios.
- Un sistema DW debe presentar la información en una línea de tiempo.
- Un sistema DW debe almacenar la información de forma segura.
- Un sistema DW es la base para mejorar la toma de decisiones.

El DW aporta beneficios a los usuarios a la hora de mejorar y comprender el rendimiento de su organización.

Para realizar un DW se tiene que diferenciar entre tablas de hechos y tablas de dimensiones. Las tablas de hechos son aquellas que contienen los datos que son susceptibles de ser medibles, suelen responder a la pregunta ¿Cuánto? Las tablas de dimensiones por el contrario son los datos que nos ofrecen la posibilidad de filtrar, agrupar y definitiva dar conocimiento a los valores. Responden a la pregunta de ¿Cómo? ¿Qué? ¿Dónde? ¿Cuándo?

Un DW se puede generar basándose en dos esquemas, el esquema de estrella o el esquema de copo de nieve. En el esquema de estrella, las tablas de hechos son el centro del esquema y de ellas solo existe un nivel de relación con las tablas de dimensiones.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

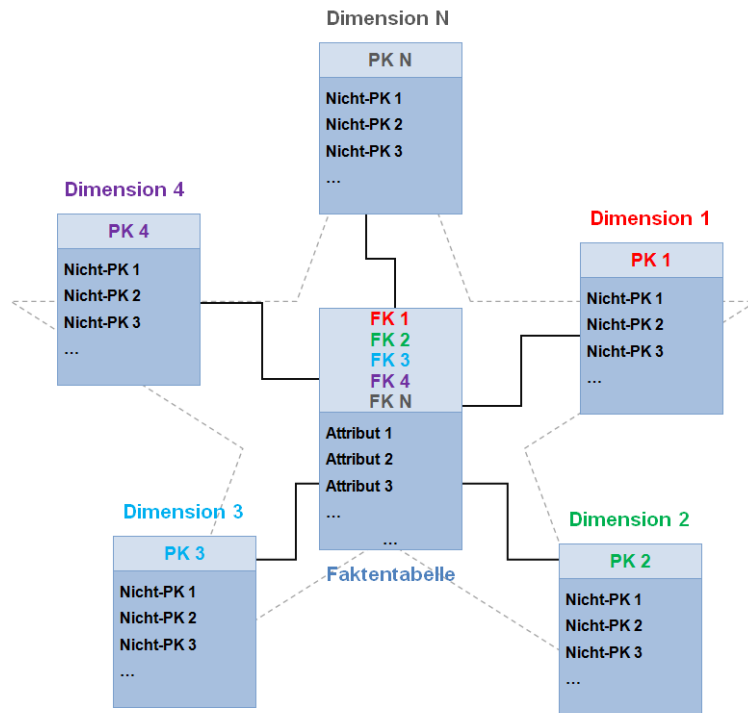


Ilustración 2 Esquema de estrella¹

En cambio, en el esquema de copo de nieve, las tablas de hechos se relacionan con tablas de dimensiones que a su vez se relacionan con una o más tablas de dimensiones en N niveles.

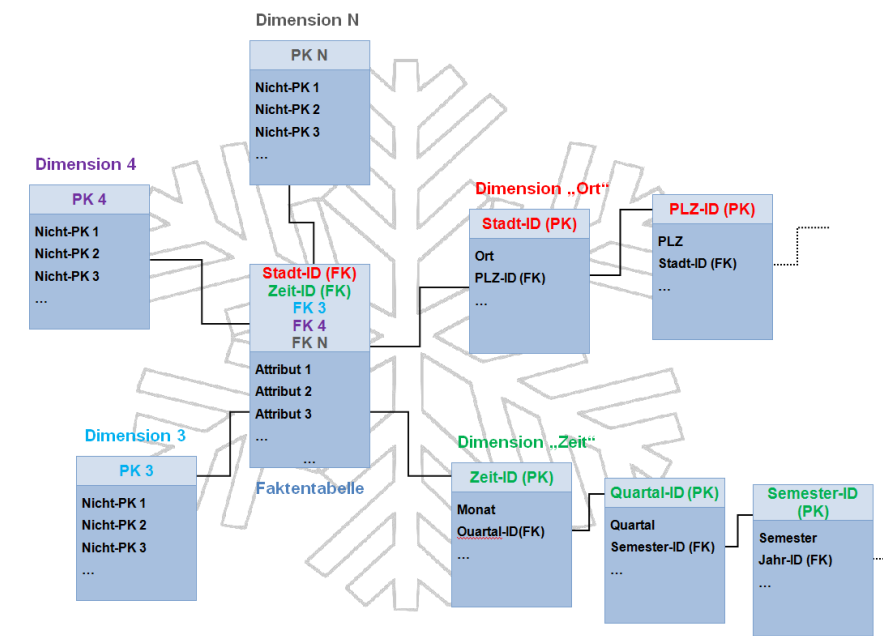


Ilustración 3 Esquema de copo de nieve²

¹ https://de.wikipedia.org/wiki/Datei:Star_Schema.png

² https://de.wikipedia.org/wiki/Datei:Snowflake_schema.png

A continuación, se ven las diferentes tecnologías que podemos utilizar para generar un DW.

2.1.1. SQL Server

SQL Server es un sistema de gestión de base de datos, desarrollado y distribuido por Microsoft.

Al igual que otros sistemas de gestión de base de datos, SQL Server está desarrollado por encima de SQL, un lenguaje programación estándar para interactuar con bases de datos. SQL Server está ligado a Transact SQL o T-SQL, una implementación de Microsoft sobre SQL que añade una serie de instrucciones de programación propietarias.

La plataforma de SQL Server ha sido exclusiva del sistema operativo Windows, hasta que en 2016 Microsoft lanzó una versión compatible con los sistemas operativos basados en Linux.

2.1.1.1. Arquitectura de SQL Server

El siguiente diagrama muestra la arquitectura de SQL Server:

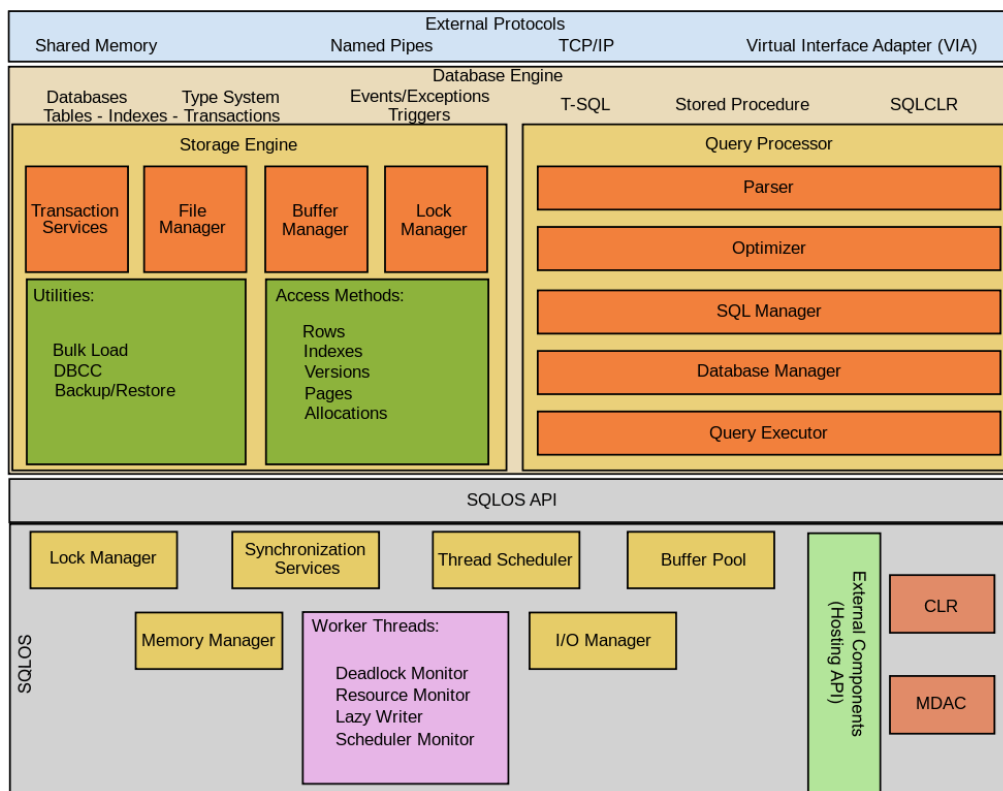


Ilustración 4. Arquitectura de SQL Server³

³ http://udayarumilli.com/wp-content/uploads/2013/06/062513_0610_SQLServerAr3.png



SQL Server se puede dividir en 2 componentes principales, el motor de base de datos y SQLOS o sistema operativo de SQL.

Motor de base de datos

El componente principal de SQL Server es su motor de base de datos. El motor de base de datos se compone de un motor relacional que se encarga de realizar y procesar las consultas a la base de datos y un motor de almacenamiento administra los archivos, las páginas, los índices, etc....

SQLOS

Bajo el motor relacional y el motor de almacenamiento se encuentra el SQLOS o Sistema Operativo de SQL Server.

SQLOS proporciona muchos servicios de sistema operativo como pueden ser la gestión de la memoria o de los sistemas de entrada salida.

2.1.1.2. Ediciones de SQL Server

SQL Server proporciona cuatro tipos de versiones para su instalación:

- SQL Server Developer Edition, creada para entornos de desarrollo y de test.
- SQL Server Express Edition, creada para pequeñas instalaciones, solo permite trabajar con 1GB de memoria y 10GB de almacenamiento.
- SQL Server Enterprise Edition, que incluye todas las características de SQL Server.
- SQL Server Standart Edition, que es una versión intermedia entre la versión Express y Enterprise.

La herramienta de gestión más utilizada y creada también por Microsoft es Microsoft SQL Management Studio.

2.1.2. MySQL

MySQL es el sistema de gestión de bases de datos de código abierto más utilizado en el mundo. Cuenta con un sistema de doble licencia, uno de código abierto y otro comercializado por Oracle.

Es un sistema de gestión de base de datos desarrollado en C y C++ y es uno de los sistemas más utilizados para la programación web.

Una de sus principales características es que, al ser código abierto, se puede modificar el código fuente para satisfacer las necesidades para cada problema. [6]

Existen versiones para los sistemas operativos Linux, Windows y Mac.

La herramienta estándar para su gestión es MySQL WorkBench.

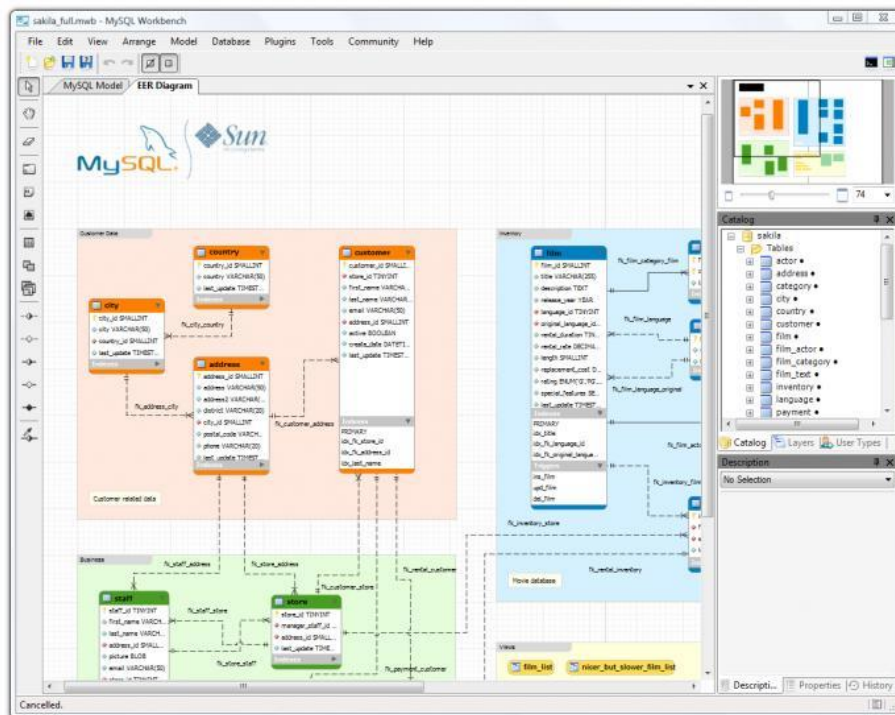


Ilustración 5 MySQL Workbench⁴

2.1.3. Oracle DB

Oracle database, Oracle DB, es un sistema de gestión de bases de datos relacionales creado por Oracle Corporation. Desarrollado en 1977 por Lawrence Ellison, es uno de los sistemas de gestión de base de datos más fiables y valorados.

Oracle DB es competidor directo de SQL Server para las bases de datos en el mundo empresarial. La estructura de Oracle DB y SQL Server son bastantes similares.

Oracle DB cuenta con versiones tanto para Windows, como Linux y Mac y se distribuye en los siguientes formatos:

- Enterprise Edition: Ofrece todas las características, incluyendo un mayor rendimiento y seguridad.
- Standart Edition: Contiene las características principales para usuario que no necesitan una versión *Enterprise*.
- Express Edition: versión ligera y gratuita disponible solo para Linux y Windows.
- Oracle Lite: versión solo para dispositivos móviles.

2.2. Integración de datos

La integración de datos o procesos ETL (*Extact, transform and load*) son procesos divididos en tres subprocesos:

⁴ <http://www.tecnopedia.net/wp-content/uploads/2009/04/workbench.png>



Extracción

En los procesos de extracción, se recogen datos de N orígenes de datos, creando la conexión necesaria a cada uno de ellos.

Transformación

Una vez se obtienen los datos, a estos se les da el formato necesario, se limpian los valores nulos e innecesario y se preparan para ser insertado en el destino de datos.

Carga

En este proceso se insertan los datos en el/los destino/s de datos utilizados para su posterior análisis y visualización.

2.2.1. Pentaho Data Integration (Kettle)

Basado en Java, se trata de una herramienta multiplataforma que pertenece a la suite de herramientas de Pentaho y que permite generar procesos ETL de forma gráfica.

Su interfaz gráfica se llama Spoon y es la que permite generar todas las transformaciones y los trabajos necesarios para realizar el proceso ETL de manera completa. [7]

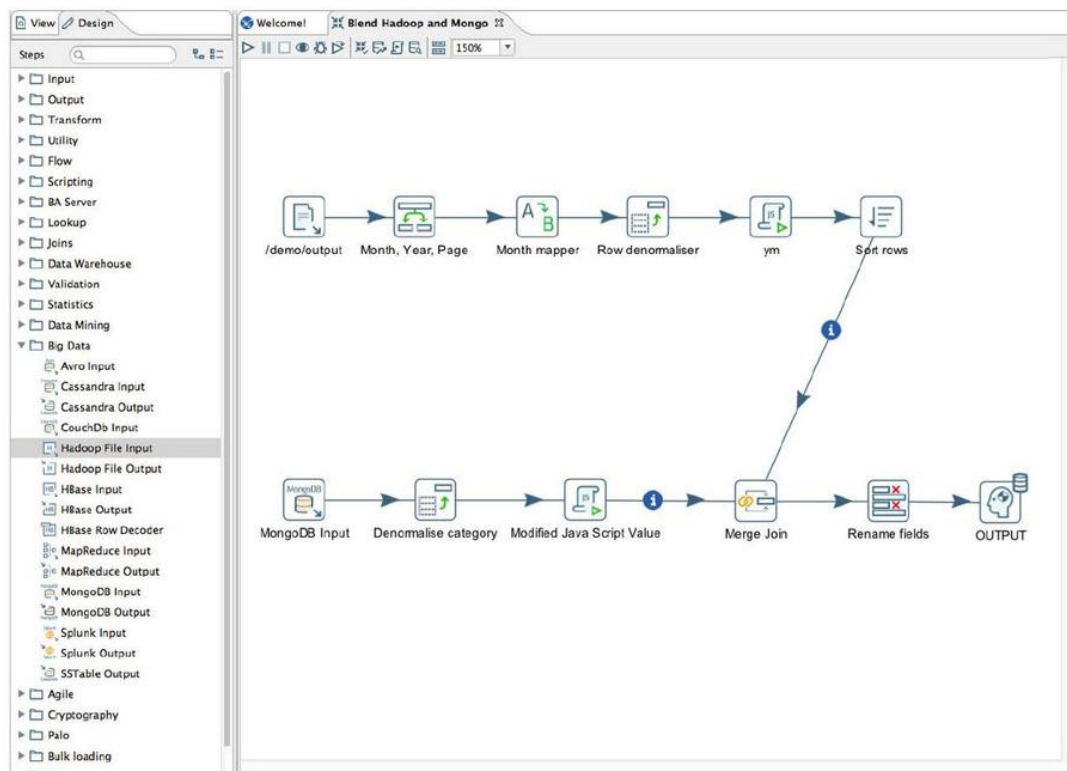


Ilustración 6. Pentaho Data Integration⁵

⁵ <http://www.know-bi.be/images/uploads/global/pentaho-data-integration.jpg>

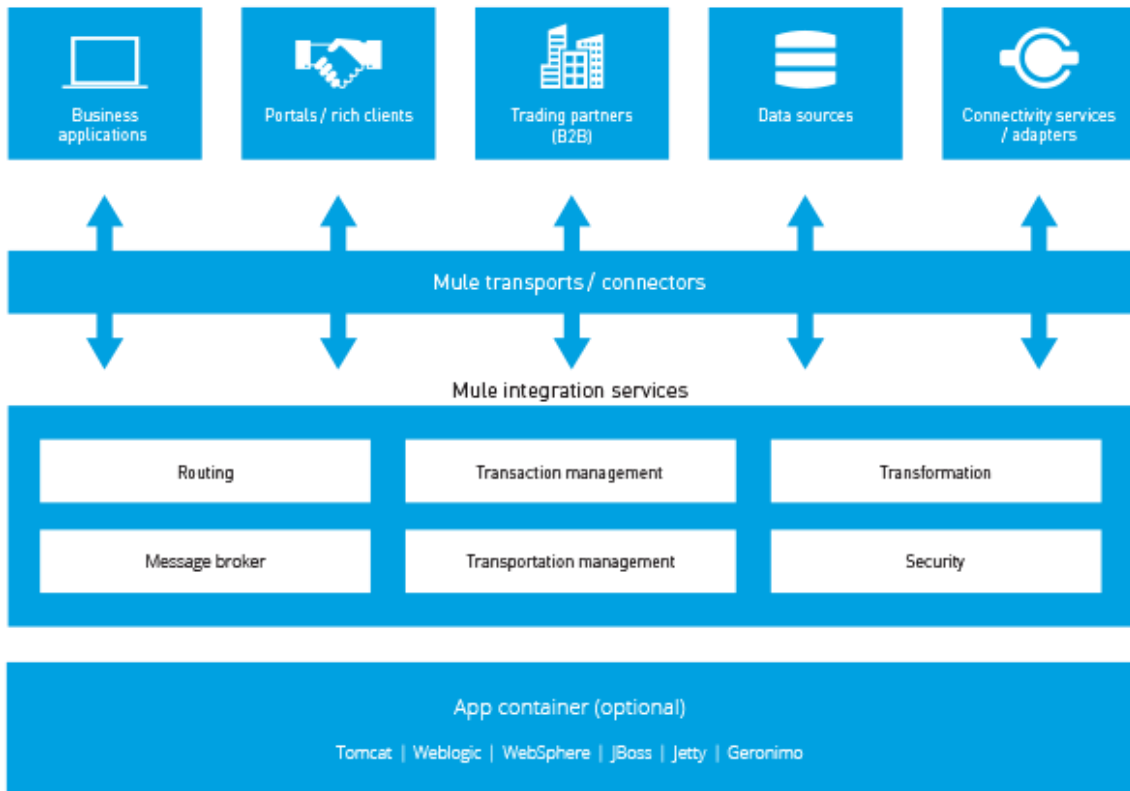


Ilustración 7. Mule ESB⁶

2.3. Cubos multidimensionales OLAP

Otra parte importante dentro de los sistemas de BI son los cubos multidimensionales OLAP (Online Analytical Processing) o procesamiento analítico online.

También llamados cubos multidimensionales, son estructuras que permiten el análisis de bases de datos relacionales de manera ágil, rápida y robusta. [8]

⁶ http://www.goldona.cz/ckphotos/1457520498_Mule%20ESB%20schema.png

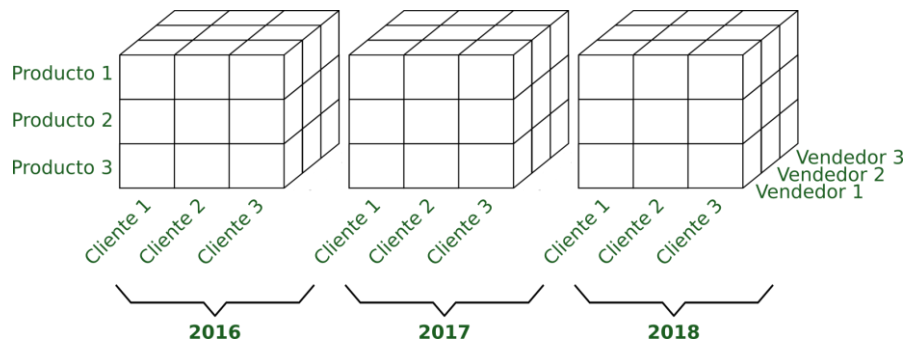


Ilustración 8. Cubo multidimensional⁷

Podemos diferenciar los cubos multidimensionales en tres sistemas [9]:

ROLAP

En el sistema ROLAP, las diferentes particiones de agregaciones de datos que genera el cubo se guardan como vistas dentro del origen de datos relacional. Esto hace que el tamaño del sistema multidimensional el tiempo de procesamiento y las consultas multidimensionales sean más lentas que con los otros dos sistemas, pero, sin embargo, consigue que las consultas se realicen sobre los datos en tiempo real y disminuye en gran cantidad el almacenamiento necesario.

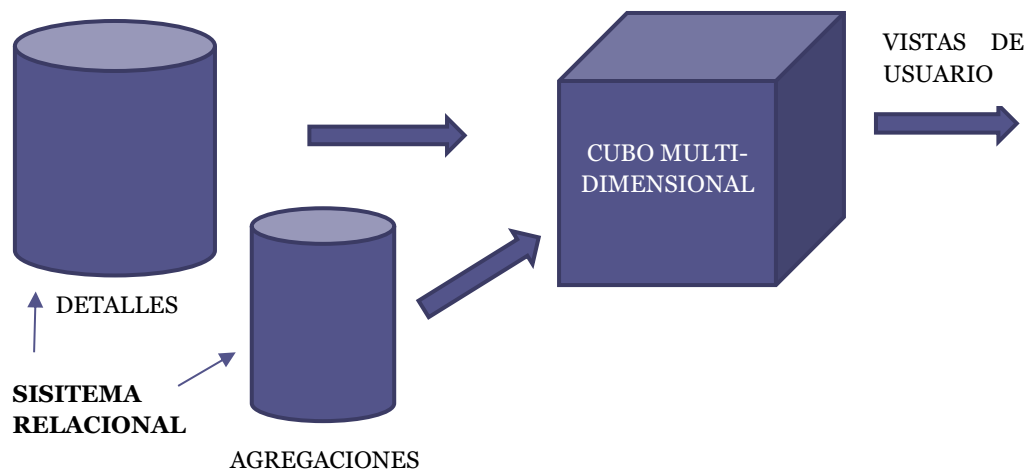


Ilustración 9. Sistema ROLAP

MOLAP

El sistema MOLAP crea las agregaciones de datos y los detalles en el cubo multidimensional, de esta manera se obtiene una gran ventaja en cuanto a la velocidad de acceso a dato y en el tiempo de procesamiento. Por otra parte, en el sistema MOLAP el tamaño de almacenamiento aumenta considerablemente y el proceso de integración de datos resulta más costoso.

⁷ http://troyanx.com/Hefesto/img_013.png



Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Se almacena una copia íntegra de la base de datos relacional en el cubo multidimensional, además, los datos multidimensionales están desconectados de la base de datos relacional por los datos están actualizados con respecto a la última vez que se procesó el cubo.

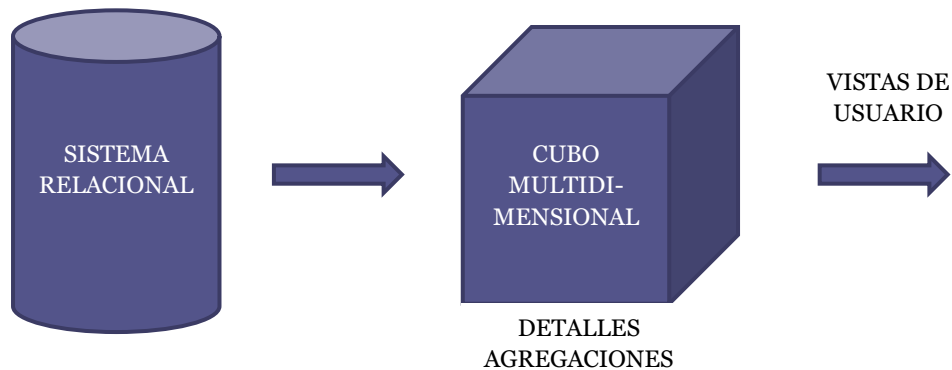


Ilustración 10. Sistema ROLAP

HOLAP

El sistema HOLAP es una mezcla entre los sistemas ROLAP y MOLAP. Al igual que en el sistema MOLAP, en el sistema HOLAP los datos agregados se almacenan en el cubo multidimensional, mientras que no se realiza una copia de la base de datos, si no los accesos a los datos relacionales no agregados están en el sistema relacional original.

Este sistema reduce el tamaño del sistema de datos multidimensional, pero ralentiza el acceso a datos no agregados. Por tanto, en un sistema HOLAP el tamaño de las particiones almacenadas es más pequeño que en un sistema MOLAP, pero a su vez el acceso a datos es más rápido que en un sistema ROLAP.

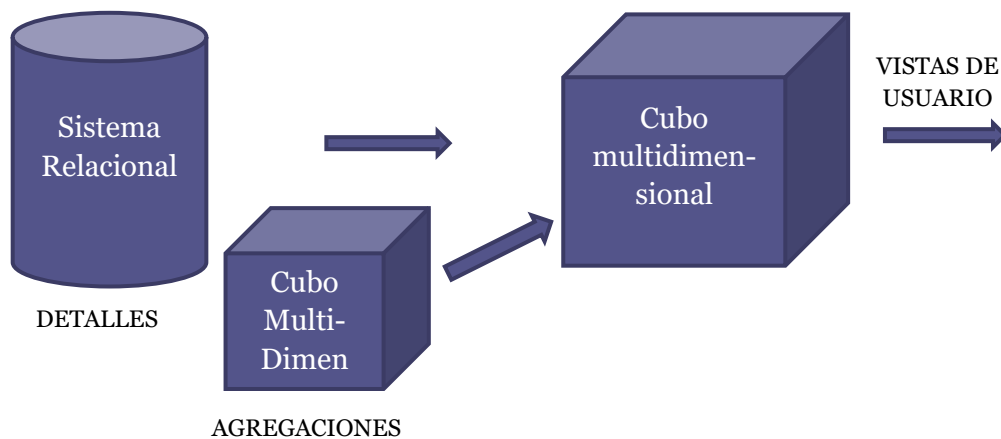


Ilustración 11. Sistema HOLAP

2.3.1. Microsoft SQL Server Analysis Services

Integrada dentro de la solución SSDT (SQL Server Data Tools), Microsoft SQL Server Analysis Services es una plataforma que contiene una serie de herramientas útiles para la generación de procesos de BI. Entre estas herramientas se encuentra la funcionalidad que permite generar bases de datos multidimensionales a partir de orígenes de datos relacionales. [10]

Microsoft SQL Server Analysis Services se instala con las herramientas integradas en la instalación de SQL Server Enterprise.

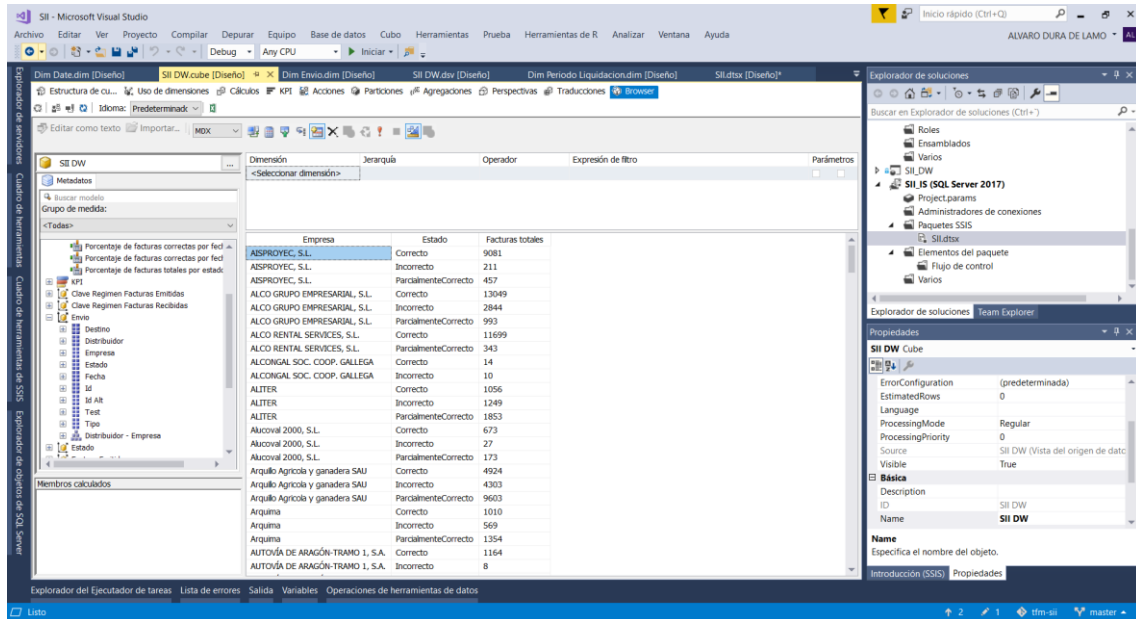


Ilustración 12. SQL Server Analysis Services

2.3.2. Analytic Workspace Manager

Analysis Workspace Manager es la herramienta para desarrollar, crear y administrar cubos multidimensionales en almacenes de datos de Oracle. Provista de una interfaz gráfica, permite generar los cubos multidimensionales de forma visual. [11]

Analysis Workspace Manager se encuentra dentro de la solución de herramientas Oracle OLAP.

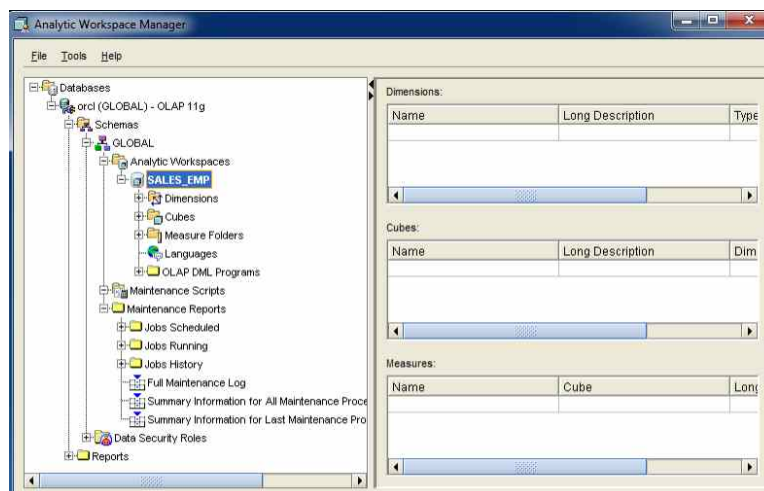


Ilustración 13. Analytic Workspace Manager⁸

⁸ <https://mperkolaba.files.wordpress.com/2012/10/1.jpg?w=300&h=190&zoom=2>



2.3.3. Pentaho Mondrian Analysis Services

Pentaho Mondrian es un motor OLAP escrito en Java. Este motor ejecuta consultas escritas en lenguajes MDX, leyendo los datos de bases de datos relacionales y mostrando los datos con formato multidimensional mediante una API de Java. [12]

Pentaho Mondrian, al contrario que las otras dos herramientas anteriores, es de código abierto.

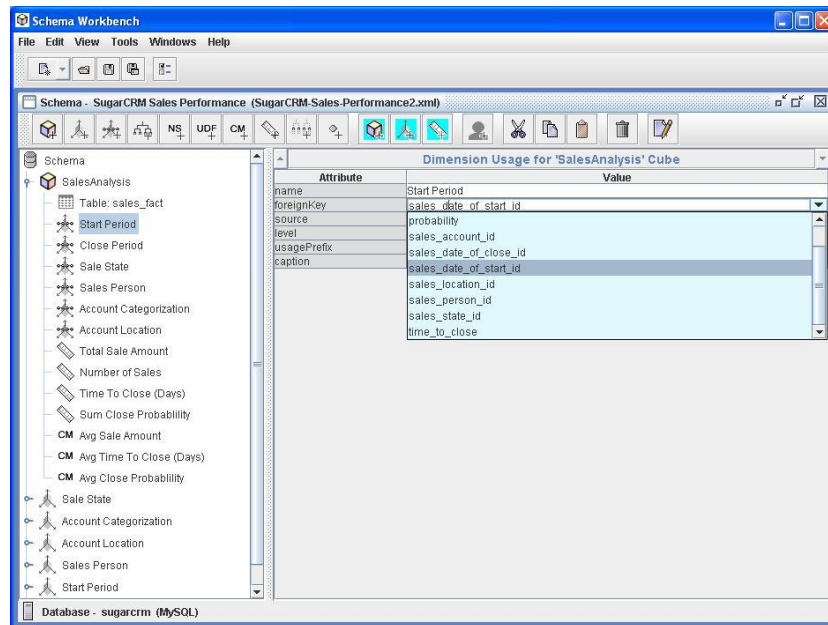


Ilustración 14. Pentaho Mondrian⁹

2.4. Visualización de datos

Una de las partes más importantes en los sistemas de BI es la visualización de los datos. De nada sirve generar un cubo multidimensional si después no se presentan los datos de la manera adecuada.

A continuación, vamos a ver una serie de herramientas que permiten generar aplicaciones para la visualización de los datos.

2.4.1. Shiny R

Shiny R es un paquete de código abierto de R que proporciona un poderoso a la vez que elegante entorno web para desarrollar aplicaciones web usando el lenguaje de programación R.

Shiny ayuda a convertir proyectos de análisis en aplicaciones web interactivas sin necesidad de conocimientos de programación en HTML, CSS y JavaScript.

⁹ https://mondrian.pentaho.com/documentation/images/workbench_database_validation.png

A diferencia de las herramientas que veremos a continuación, Shiny R requiere conocimientos previos de programación en R, pero proporciona al desarrollador una mayor posibilidad de personalización de la aplicación.

Es una herramienta de código abierto y gratuita.

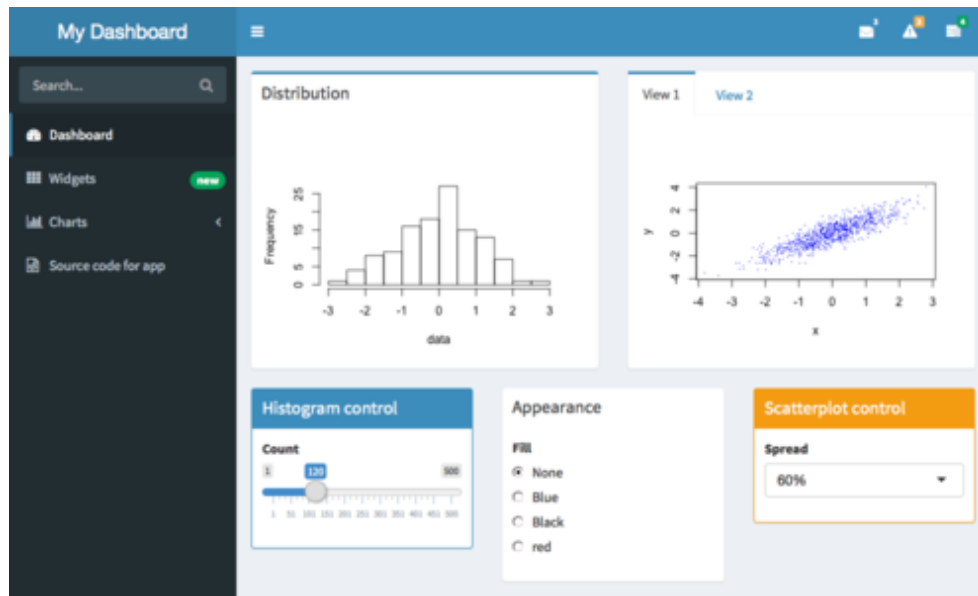


Ilustración 15. Shiny R.¹⁰

2.4.2. Microsoft Power BI

Microsoft Power BI es una herramienta de inteligencia de negocio que proporciona a usuarios comerciales sin conocimientos técnicos la posibilidad de agregar, analizar, visualizar y compartir información.

La interfaz de usuario de Power BI es intuitiva para usuarios familiarizados con Excel y tiene una gran integración con otras herramientas de Microsoft, lo que hace que se una herramienta con una curva baja de aprendizaje. [13]

Power BI tiene 3 niveles de comercialización: Power BI Desktop que es gratuita, Power BI Pro que cuesta 9.99€ al mes por usuario y Power BI Premium, dónde el precio depende del tamaño de los desarrolladores y usuario.

¹⁰ <https://steemitimages.com/640x0/http://rstudio.github.io/shinydashboard/images/dashboard.png>

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

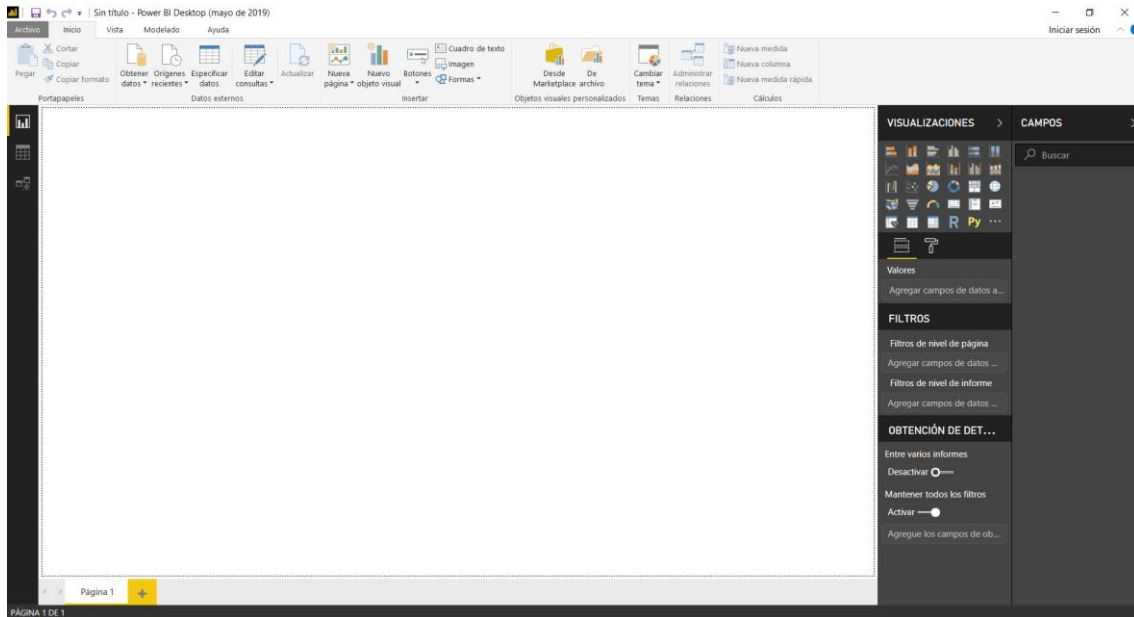


Ilustración 16. Power BI.

2.4.3. Qlick Sense

Qlick Sense nace en 2014 como la hermana sencilla de Qlick View, ambas de la empresa Qlick.

Qlick Sense se creó como competencia a las nuevas tecnologías de BI del mercado, creando una plataforma más fácil de utilizar, con adaptación para tecnologías móviles y pensada en el usuario final.

Al igual que Power BI, Qlick Sense se distribuye en 3 versiones: Qlick Sense Cloud Basic, gratuita para usuarios principiantes, Qlick Sense Cloud Bussines, con un coste de 15 USD al mes por usuario, está destinada a equipos relativamente pequeños y Qlick Sense Enterprise, que varía de precio según el tipo de empresa. [14]

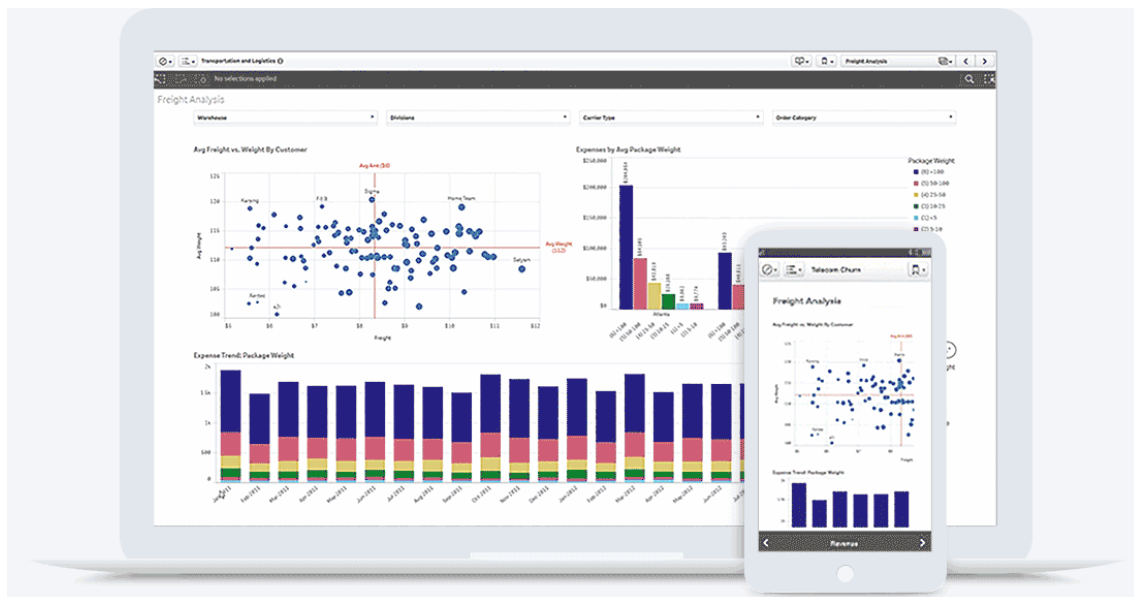


Ilustración 17. Qlick Sense.¹¹

2.4.4. Tableau

Tableau es una suite de herramientas BI desarrollada por Tableau Software a partir de 2003.

Actualmente es una de las herramientas más utilizadas junto con las otras 2 herramientas destinadas de usuario no programador que hemos visto.

Al igual que las herramientas anteriores, también podemos crear vistas tanto para escritorio como para dispositivos móviles. [15]

Tableau se distribuye en tres versiones:

- Tableau Creator: creada para desarrolladores, la licencia cuesta 70 € por usuario y mes e incluye licencias para Tableau Desktop, Tableau Prep Builder y una licencia para Tableau Server y Tableau Online.
- Tableau Explorer: creada para usuarios expertos que utilizan Tableau para realizar consultas experimentadas y cuesta 35 € por usuario y mes con acceso a la licencia de Tableau Server.
- Tableau Viewer: creada para usuarios menos experimentados, con acceso a consulta de los paneles principales por un coste de 12€ al mes.

¹¹ <https://www.decideo.com/photo/art/default/25490530-26738767.jpg?v=1536899142>



Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

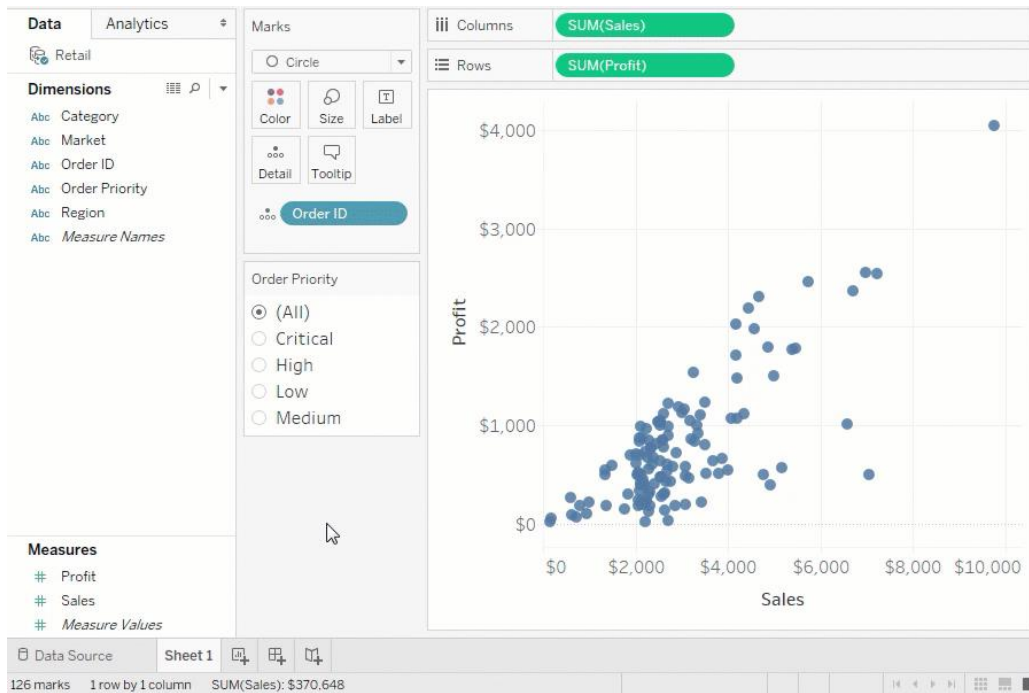


Ilustración 18. Tableau.¹²

2.5. Empresa Ahora

AHORA es una empresa de desarrollo Software con más de 25 años de actividad. Actualmente cuenta con una plantilla de alrededor a 100 personas, centradas en dos productos.

Ahora ERP. Es un software de gestión de empresa implantado en más de 1000 clientes y que cuenta con una red de distribución de 55 partners en territorio nacional.

Flexygo. Herramienta RAD para la generación de aplicaciones Web en un entorno de alta rentabilidad y basada en las últimas tendencias tecnológicas. Esta herramienta tiene clientes tan potentes como el Banco de Santander, Mapfre y Garrigues abogados, etc.

Dentro de estas actividades, el proyecto que engloba este trabajo de fin de máster se enmarca como el primer eslabón en la aplicación de tecnologías BI dentro de la empresa, en el marco del SII, donde se gestiona el envío a hacienda de las facturas de empresas.

¹² <https://cdns.tblsft.com/sites/default/files/pages/answerdeeperquestions.png>

3. Análisis de la solución

Es este capítulo se detallará un análisis completo de la solución, dónde se mostrará los requerimientos y peticiones del cliente, la empresa AHORA en este caso, después se mostrarán las distintas tecnologías a utilizar en el proyecto, seguido de la planificación utilizada para desarrollar la solución y por último un análisis técnico.

3.1. Obtención de requerimientos

La empresa AHORA tiene implementada una solución para proporcionar a sus clientes la posibilidad de enviar sus facturas al SII. Puesto que para realizar peticiones a servicios pertenecientes a organismos públicos es necesaria una autenticación mediante certificados, la empresa decidió generar un servidor intermedio dónde todas las empresas que utilicen el sistema pueden enviar las facturas a este servidor y la empresa AHORA realice el envío en nombre de las empresas usuarias de la solución. De esta manera se obtiene una trazabilidad mayor de los envíos realizados por cada empresa y se evitan los pasos para conseguir el certificado de la Agencia Tributaria.

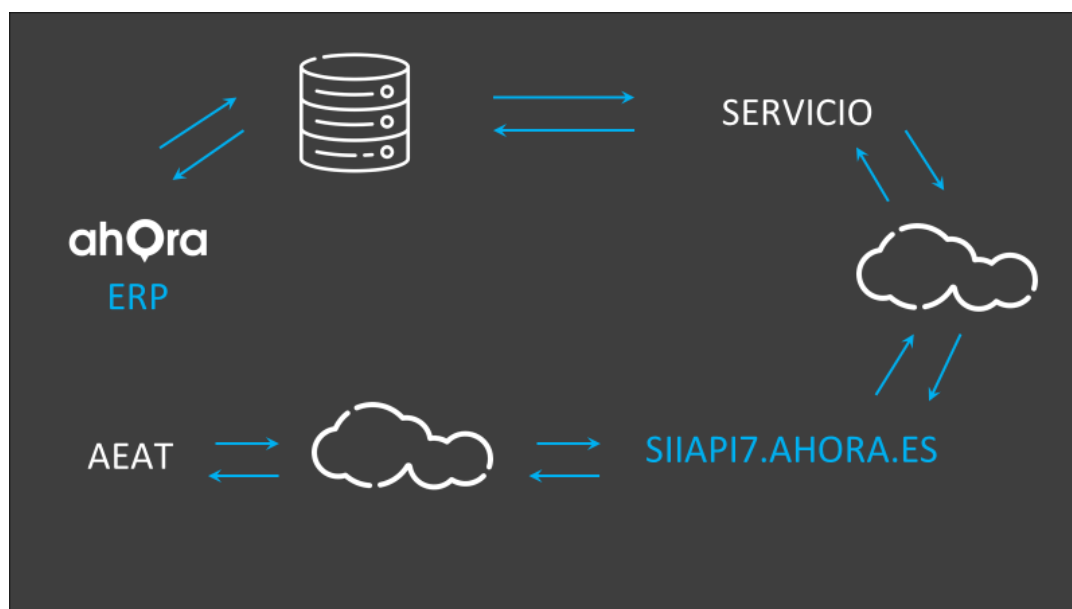


Ilustración 19. Estructura API SII AHORA

Dentro del servidor intermedio de AHORA se guardan cientos de envíos diarios lo que supone miles de facturas. Debido a las especificaciones y requerimientos del SII, los envíos se realizan en formato XML, indicando las propiedades necesarias para identificar y describir cada factura.

El problema que tienen es que debido a la gran cantidad de datos que suponen todos estos envíos en XML, no es posible realizar consultas relacionales con un tiempo razonable de respuesta.

La empresa AHORA quiere poder estudiar el comportamiento de cada empresa, revisando los envíos que realizan y sus respectivas facturas. También quieren saber el

rendimiento de la solución, estudiando los errores que se pueden producir en los envíos.

3.2. Tecnologías

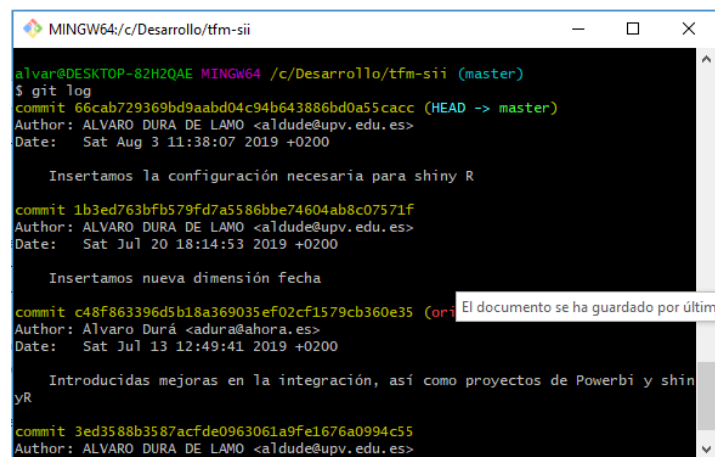
A continuación, vamos a ver las tecnologías a utilizar para realizar la solución. El principal motivo por el que se van a utilizar las siguientes tecnologías es que son herramientas que se utilizan diariamente en la empresa AHORA y con las que tanto el autor como los posibles desarrolladores futuros estarán familiarizados.

3.2.1. Git

Git es el sistema de gestión de versiones moderno más utilizado hoy en día. Desarrollado inicialmente en 2005, Git es proyecto maduro de código abierto y de mantenimiento activo. Es compatible con varios sistemas operativos y diversos entornos de desarrollo. [16]

Git no es válido solo para mantener código, sino que también podremos mantener documentación y otros tipos de archivos.

Se utilizará Git para tener el código versionado.



```
MINGW64; c:/Desarrollo/tfm-sii
alvar@DESKTOP-82H2QAE MINGW64 /c/Desarrollo/tfm-sii (master)
$ git log
commit 66cab729369bd9aab04c94b643886bd0a55cacc (HEAD -> master)
Author: ALVARO DURA DE LAMO <a1dude@upv.edu.es>
Date: Sat Aug 3 11:38:07 2019 +0200

    Insertamos la configuración necesaria para shiny R

commit 1b3ed763bf579fd7a5586bbe74604ab8c07571f
Author: ALVARO DURA DE LAMO <a1dude@upv.edu.es>
Date: Sat Jul 20 18:14:53 2019 +0200

    Insertamos nueva dimensión fecha

commit c48f863396d5b18a369035ef02cf1579cb360e35 (ort... El documento se ha guardado por últim
Author: Alvaro Durá <adura@ahora.es>
Date: Sat Jul 13 12:49:41 2019 +0200

    Introducidas mejoras en la integración, así como proyectos de Powerbi y shin
yR

commit 3ed3588b3587acfde0963061a9fe1676a0994c55
Author: ALVARO DURA DE LAMO <a1dude@upv.edu.es>
```

Ilustración 20. Git

3.2.2. GitLab

GitLab es una plataforma completa de *DevOps*. Dentro de esta aplicación se puede mantener un servidor de archivos Git, además de facilitar la gestión de código en Git en los equipos de desarrollo. Por otra parte, permite gestionar tareas mediante el uso de tableros y crear procesos de CI/CD (Integración y despliegue continuos). [17]

Se usará esta aplicación para crear el servidor del gestor de versiones de Git.

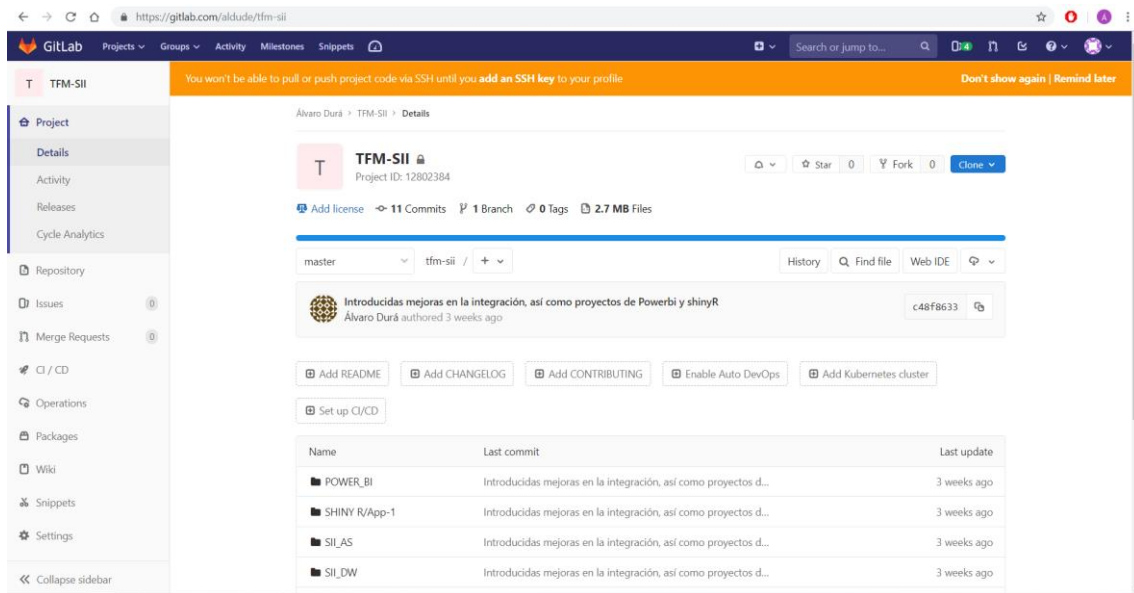


Ilustración 21. GitLab

3.2.3. SQL Server Management Studio 2017

SQL Server Management Studio es un entorno de desarrollo integrado para gestionar cualquier infraestructura de SQL, desde SQL Server hasta bases de datos de SQL en Azure. Este entorno proporciona herramientas para configurar, monitorizar y administrar instancias de SQL Server y bases de datos. [18]

SQL Server Management Studio se puede usar para consultar, diseñar y administrar bases de datos y almacenes de bases de datos. Además, este entorno es gratuito.

Se utilizará esta herramienta para gestionar nuestras bases de datos relacionales y los almacenes de datos que desarrollaremos.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

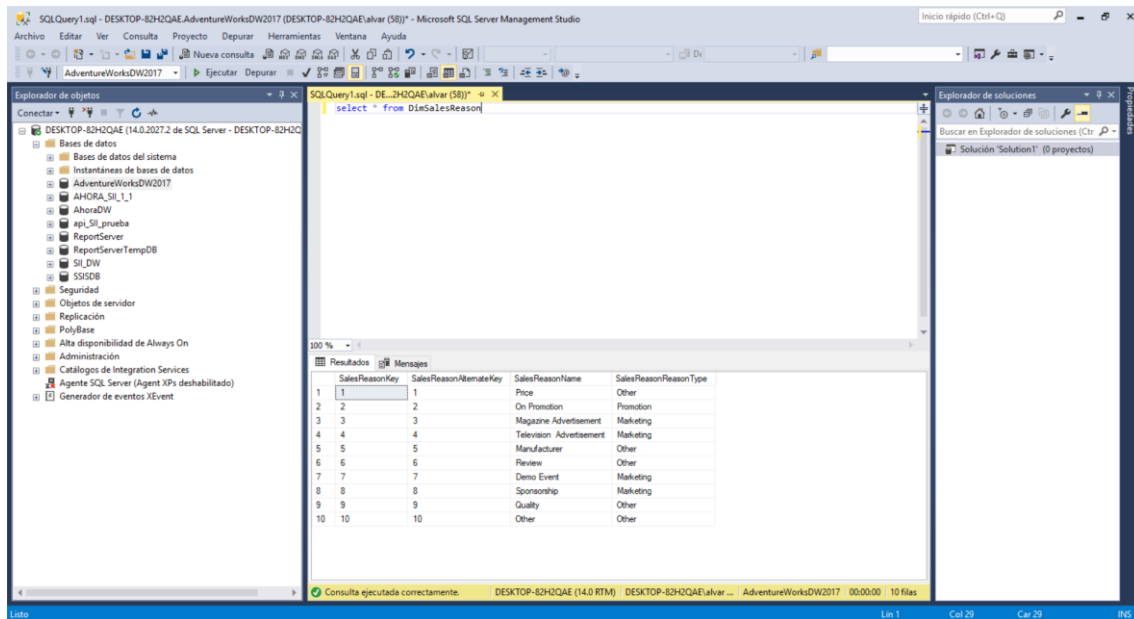


Ilustración 22. SQL Server Management Studio

3.2.4. SQL Server

Se usará el motor de bases de datos relacionales de Microsoft SQL Server.

Se utilizará la versión para desarrolladores ya que es gratuita para entornos de desarrollo como es el caso.

3.2.5. SQL Server Integration Services

Se utilizará el componente de Microsoft SQL Server, Integration Services, para generar trabajos de migración de datos.

Mediante esta herramienta obtendremos los datos desde la base de datos relacional, transformaremos los datos y los insertaremos en el almacén de datos.

3.2.6. SQL Server Analysis Services

La herramienta Analysis Services, incluida dentro de los componentes de Microsoft SQL Server, permite crear un cubo multidimensional a partir del almacén de datos.

3.2.7. Power BI

Se usará Power BI para crear las gráficas a partir del cubo multidimensional de manera sencilla a sin necesidad de conocimientos de programación. Esto hace que sea más fácil crear las gráficas, pero añade más complejidad a la hora de crear el cubo multidimensional.

Power BI pertenece al conjunto de herramientas de Microsoft para crear soluciones de inteligencia de negocio.

3.2.8. R

R es un lenguaje de programación y un entorno para computación estadísticas y gráficas. Es un proyecto de GNU de software libre. [19].

Se ha utilizado como lenguaje de programación para crear la visualización de las gráficas.

3.2.9. Shiny R

Shiny R es un paquete de código abierto programado en el lenguaje de programación R, que se usará para crear una aplicación web donde los usuarios podrán consultar el estado de la aplicación mediante gráficas y tablas.

Es necesario tener nociones de programación al contrario que en Power BI, pero permite una mayor personalización.

3.2.10. Shiny Dashboard

Shiny Dashboard es un paquete basado en R y Shiny R que permite generar un panel de control y de información de manera rápida y sencilla.

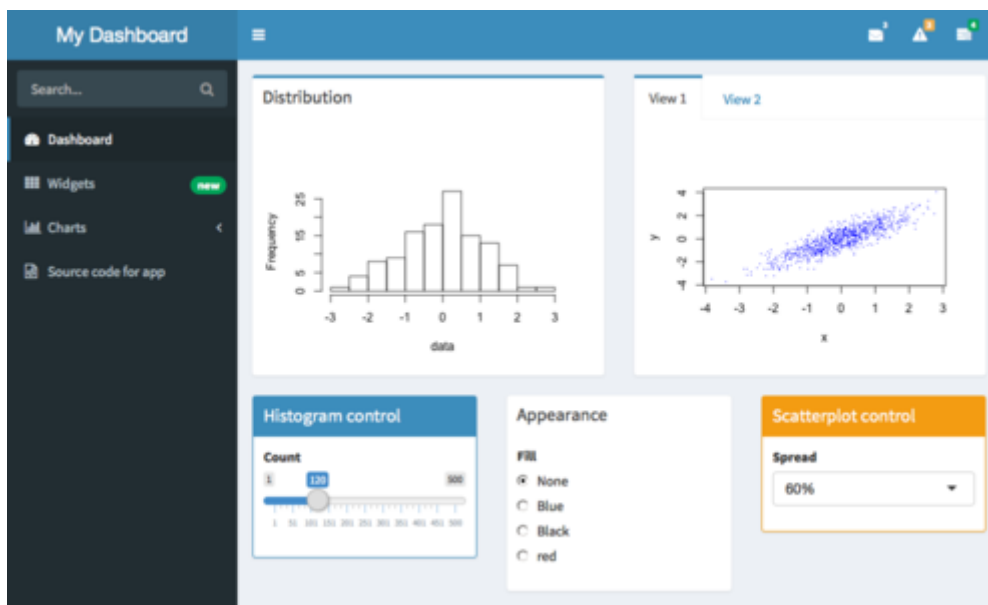


Ilustración 23. Shiny Dashboard¹³

3.2.11. Microsoft Visual Studio 2017

Visual Studio 2017, Microsoft Visual Studio 2017o VS 2017es un entorno de programación creado por Microsoft. Este entorno permite desarrollar en un número muy elevado de lenguajes y tecnologías.

¹³ <https://steemitimages.com/640x0/http://rstudio.github.io/shinydashboard/images/dashboard.png>

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Se usará en este proyecto para crear la solución dónde se integrará el proyecto de base de datos para el almacén de datos, el proyecto de SQL Server Integration Services para el proceso de ETL y el proyecto de SQL Server Analysis Services para la creación del proyecto del cubo multidimensional.

Existen tres versiones de esta plataforma, la versión Community, gratuita y dirigida a estudiantes y desarrolladores de software libre, la versión Professional, con un precio de 45 dólares al mes por usuario y enfocada a pequeños equipos de desarrollo y por último la versión Enterprise, con un precio de 250 dólares al mes por usuario y dirigida a grandes empresas de desarrollo de software.

En este proyecto se utilizará la versión Community, ya que es la que mejor se ajusta a nuestras necesidades.

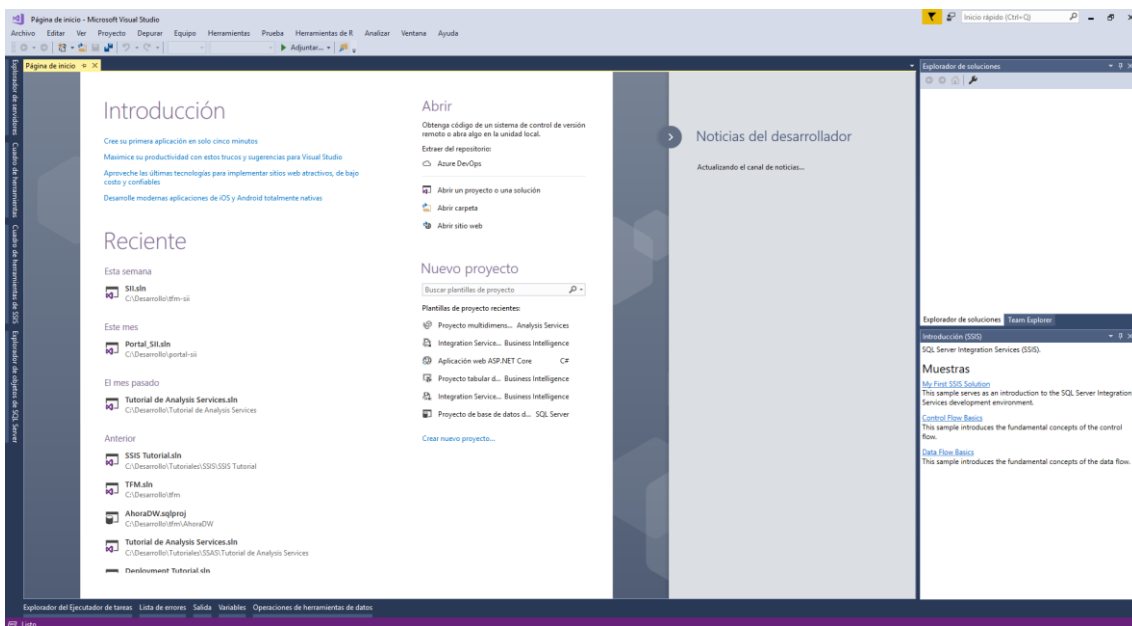


Ilustración 24. Microsoft Visual Studio 2017

3.3. Planificación

En este subcapítulo se expondrá la planificación seguida para realizar el proyecto.

Para realizar el trabajo se ha seguido la metodología ágil Scrum.

Una metodología de trabajo para el desarrollo de software es un marco de trabajo que se usa para estructurar, planificar y controlar el proceso de desarrollo de un proyecto. [20]

Dentro de una metodología de trabajo se puede diferenciar entre dos tipos, metodologías ágiles y metodologías robustas o tradicionales.

Las metodologías ágiles permiten gestionar un proyecto de manera dinámica. Las especificaciones iniciales pueden ir variando, lo que permite variar las especificaciones e ir ajustándolas mientras el proyecto avanza. En cambio, las metodologías tradicionales el proyecto se cierra a principio del desarrollo y se genera una planificación cerrada, que no cambia hasta que se termina el proyecto.

Dentro de las metodologías ágiles tenemos Scrum. Esta metodología se basa en la generación de pequeños equipos autogestionados, que trabajan mediante *sprints*, con pequeños desarrollos que no duran más de una o dos semanas. [21]

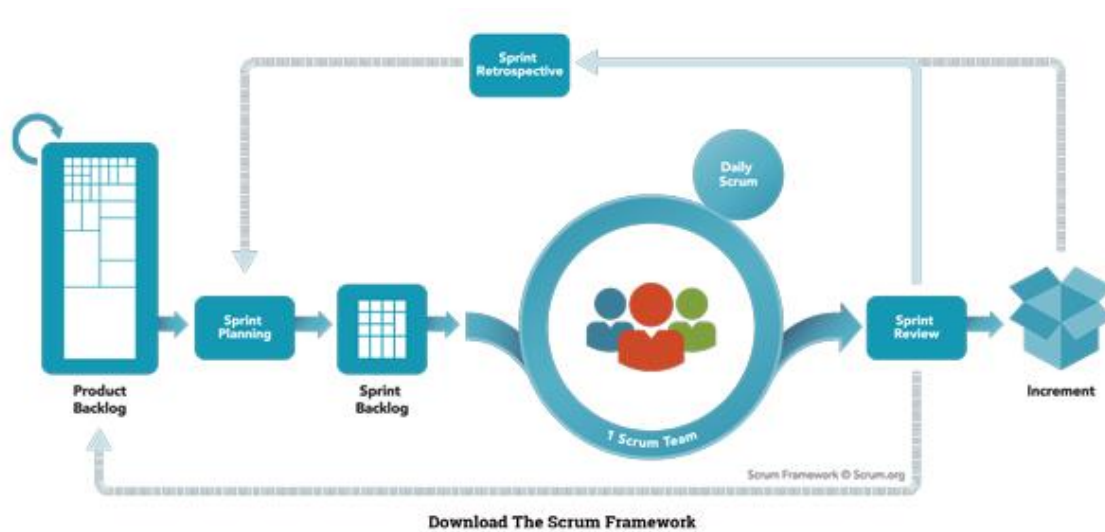


Ilustración 25. Metodología Scrum¹⁴

En esta metodología, se divide la totalidad del proyecto en pequeñas tareas y con ellas se conforma un *backlog*. Una vez se tiene definido el *backlog*, se escogen una serie de tareas que se realizarán en el *sprint*, creando el *sprint backlog*. Durante el tiempo de duración del *sprint*, los miembros del equipo realizan una reunión diaria llamada *daily*, donde exponen sus problemas e intentan resolverlos. Por último, se realiza una revisión del *sprint*, llamada *review*, y se pueden crear nuevas tareas para el *sprint backlog*.

En este caso se utilizará la metodología de trabajo Scrum, con la salvedad que el equipo de trabajo estará compuesto por una persona, pero se gestionará el proyecto en *sprints* de dos semanas.

¹⁴ http://provexilon.com/wp-content/uploads/2017/11/Scrum-framework_en-768x384.jpg

Número de Sprint	Descripción
1	Se estudiarán las aplicaciones a utilizar, adquiriendo experiencia y soltura a la hora de utilizarlas.
2	Se realizarán varias reuniones con los responsables financieros y de producto de la empresa AHORA para obtener los requerimientos necesarios para crear la aplicación.
3	Se creará un análisis funcional, dónde se especificará la funcionalidad que ha de tener la solución y otro análisis técnico, dónde se crearán los distintos esquemas y procesos que se han de crear.
4	Se implementará el almacén de datos dónde se insertarán los datos una vez procesados.
5	Se desarrollará la primera parte del proceso de ETL, dónde obtendremos los datos de las distintas dimensiones especificadas anteriormente.
6	Se desarrollará la segunda parte del proceso de ETL, dónde obtendremos los datos para las tablas de hechos.
7	Se implementará el cubo multidimensional que permitirá la consulta rápida de los datos almacenados.
8	Se programará la aplicación de consulta de datos mediante PowerBI.
9	Se desarrollará la aplicación de consulta de datos mediante Shiny R.
10	Se generará la documentación necesaria para la presentación del proyecto.

Ilustración 26. Planificación del proyecto.

3.4. Análisis funcional

A continuación, se va a mostrar un análisis funcional dónde se detallará la funcionalidad que debe tener la aplicación.

Este análisis deja de lado la parte técnica de la aplicación y muestra cómo se debe comportar la misma frente al usuario.

La aplicación se compondrá de varios elementos que informarán de la situación actual de la solución, como pueden ser gráficas, tablas, KPIs, etc.

La aplicación constará de 3 pantallas, una pantalla principal con un resumen global de la aplicación, otra pantalla con los detalles de los envíos y una última pantalla con los detalles de las facturas.

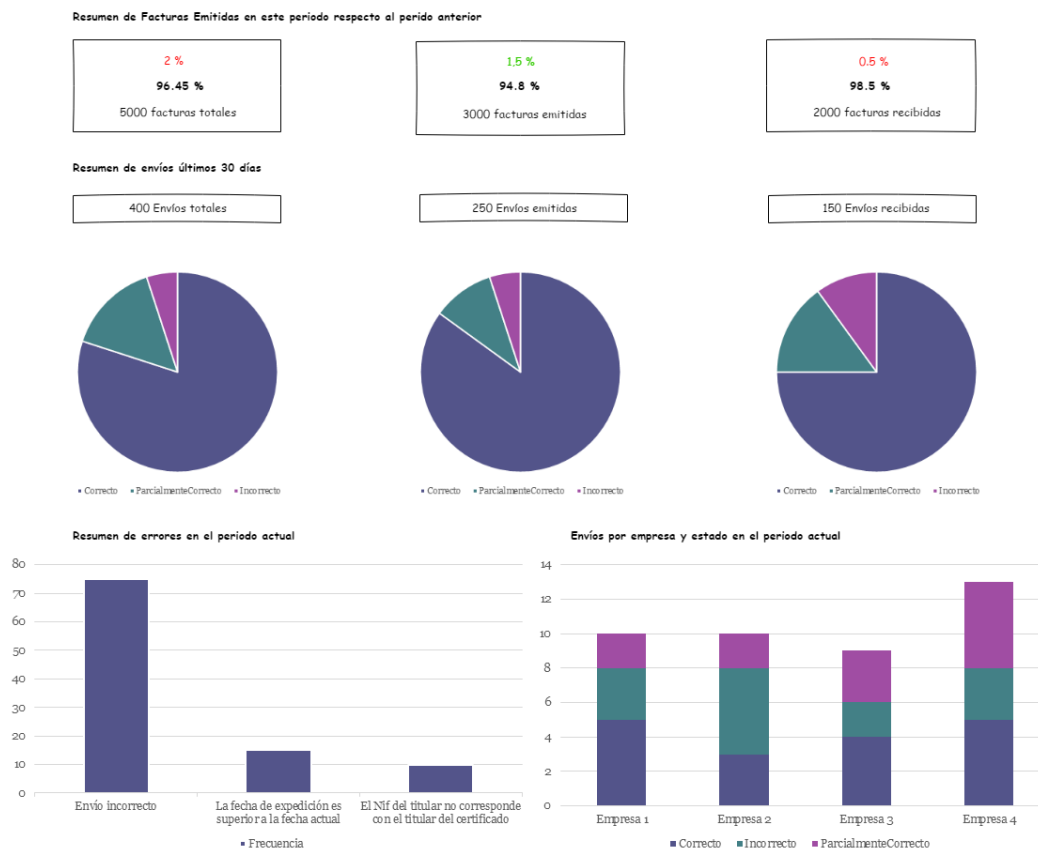


Ilustración 27. Mockup Pantalla principal

Esta primera pantalla se dividirá en cuatro secciones, en la primera sección se podrá ver un resumen de las facturas enviadas en este periodo con respecto al periodo anterior, considerando que un periodo se establece hasta el día 15 del mes siguiente. Se encontrarán 3 medidores de rendimiento que mostrarán el número de facturas en el periodo actual, el porcentaje de facturas correctas del total de las enviadas en ese periodo y la comparativa con mismo porcentaje en el periodo anterior, teniendo en cuenta que, si ha aumentado el porcentaje, se coloreará de verde y si, por el contrario, este periodo el porcentaje de facturas correctas es menor, se coloreará en rojo.

Estos indicadores de rendimiento se realizarán para el total de las facturas, las facturas emitidas y facturas recibidas.

La segunda sección mostrará un resumen de los envíos de los últimos 30 días. Aquí se verán tres subsecciones, las tres con la misma estructura, pero una para el total de envíos, otra para los envíos de facturas recibidas y otra para los envíos de facturas emitidas. Se visualizarán el recuento de envíos y debajo un gráfico de tarta, donde se verán los porcentajes de números de envíos por estado (Correcto, Parcialmente Correcto, Incorrecto).

En la tercera sección se mostrará el resumen de errores de facturas en el periodo actual, donde se podrá ver la descripción del error y la frecuencia de aparición.

En la cuarta sección y última, se podrá ver un resumen de los envíos realizados por empresa y estado, mostrando un gráfico de barras acumulativo, separando por empresa y acumulando por estado.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Filtros

07-07-2019 a 07-08-2019 Distribuidor: Distribuidor 2 Empresa: Empresa 1

Resumen de envíos

TOTAL ENVÍOS 2.631 Envíos	TOTAL ENVÍOS EMITIDAS 1.486 Envíos	TOTAL ENVÍOS RECIBIDAS 1.145 Envíos
PORCENTAJE ENVÍOS CORRECTOS 81,99 %	PORCENTAJE ENVÍOS EMITIDAS CORRECTOS 80,15 %	PORCENTAJE ENVÍOS RECIBIDAS CORRECTOS 84,50 %

Envíos por empresa y estado

Empresa	Correcto	ParcialmenteCorrecto	Incorrecto
Empresa 1	130	3	10
Empresa 2	130	3	10
Empresa 3	130	3	10
Empresa 4	130	3	10
Empresa 5	130	3	10
Empresa 6	130	3	10
Empresa 7	130	3	10
Empresa 8	130	3	10
Empresa 9	130	3	10
Empresa 10	130	3	10

Ilustración 28. Mockup resumen de envíos.

En la segunda pantalla se podrá estudiar el comportamiento de los envíos. Para ello se dividirá en tres secciones. Una primera sección que permitirá mediante fechas todos los componentes de la pantalla. Por defecto, este filtro vendrá precargado con el último mes.

La segunda sección de la pantalla contendrá seis subsecciones. Tres subsecciones superiores y tres subsecciones inferiores. En la primera subsección se mostrarán el total de envíos, los envíos totales de facturas emitidas y los envíos totales de facturas recibidas. En la segunda subsección se encontrarán los porcentajes del total de envíos correctos, el porcentaje del total de envíos correctos de facturas emitidas y el porcentaje del total de envíos correctos de facturas recibidas.

En la tercera y última sección se encontrará una tabla con los envíos de cada empresa divididos en correctos, parcialmente correctos e incorrectos. Además, se podrá filtrar para buscar empresas en concreto.

Filtros

07-07-2019 a 07-08-2019 Distribuidor: Distribuidor 2 Empresa: Empresa 1

Resumen de facturas

TOTAL FACTURAS 116.919 Facturas	TOTAL FACTURAS EMITIDAS 86.608 Facturas	TOTAL FACTURAS RECIBIDAS 30.311 Facturas
PORCENTAJE FACTURAS CORRECTAS 92,89 %	PORCENTAJE FACTURAS EMITIDAS CORRECTAS 95,57 %	PORCENTAJE FACTURAS RECIBIDAS CORRECTAS 93,58 %

Facturas por tipo y fecha

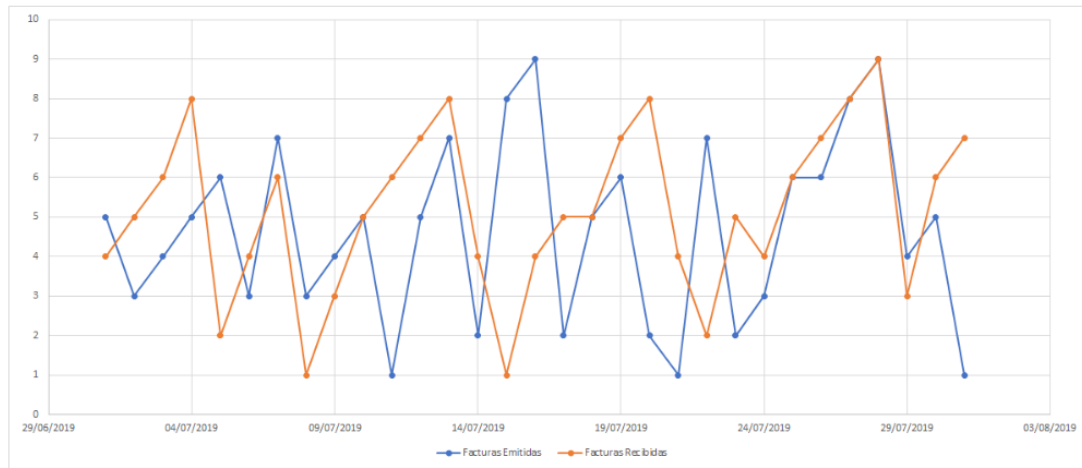


Ilustración 29. Mockup resumen de facturas.

En la tercera pantalla se observará un resumen de las facturas enviadas por fechas, empresas y distribuidores y se dividirá en tres secciones.

La primera sección mostrará filtros que afectarán a toda la pantalla y permitirá filtrar por fecha de envío, distribuidor y empresa.

La segunda sección contará con dos subsecciones, divididas en tres a su vez. La primera subsección incluirá el total de facturas enviadas, el total de facturas emitidas y el total de facturas recibidas. La segunda subsección contendrá el porcentaje de facturas correctas, el porcentaje de facturas emitidas correctas y el porcentaje de facturas recibidas correctas.

La tercera y última sección se compondrá de una gráfica que mostrará las facturas emitidas por fecha y tipo.



4. Desarrollo.

En el capítulo siguiente se explicará cómo se ha desarrollado el proyecto, viendo las diferentes fases en las que se ha generado.

Para mantener el proyecto junto, cohesionado y correctamente versionado, es necesario crear una solución de Visual Studio que agrupe los distintos proyectos que engloban esta primera fase de tratamiento de datos. Esta solución se generará cuando se cree el primer proyecto.

4.1. Almacén de datos

Se detallarán los pasos a seguir para desarrollar el almacén de datos que posteriormente permitirá guardar los datos a partir del origen de datos.

4.1.1. Estudio de estructura de tablas

Se generará la estructura de tablas basándose en la estructura de estrella, ya que mediante este esquema aumentaremos la velocidad de lectura y cálculo. Este esquema penaliza el tamaño, ya que, al desnormalizar la base de datos, aumenta la cantidad de datos en las tablas de dimensiones, pero comparado con el tamaño y número de filas de las tablas de hechos, este aumento es prácticamente nulo en comparación con el tamaño total.

Para poder generar el sistema de almacén de datos es necesario realizar un estudio de la estructura de tablas que ayudará en su posterior programación.

El sistema que se necesita pasar al almacén de datos se basa en una serie de envíos al servicio web creado por la empresa AHORA que contienen distintas facturas que posteriormente se envían a la AEAT.

El análisis de este sistema es el siguiente:

- Un envío es realizado por una empresa, en una fecha y horas concretas. Puede ser de facturas emitidas o de facturas recibidas, además de contener un identificador único, que sirve para identificarlo durante todo el proceso de envío. También se identifica si el envío se realiza al servicio web de pruebas o real y a que destino fiscal, AEAT, Navarra, Bizkaia, Gipuzkoa, Araba o Canarias. Por último, el servicio web devuelve el estado del envío, que puede ser Correcto, Parcialmente Correcto o Incorrecto.
- Un envío puede contener una o muchas facturas, pero todas de un mismo tipo y destino.
- Una empresa además pertenece a un distribuidor que a su vez puede tener varias empresas asociadas.
- Una factura emitida contiene una clave de régimen especial o trascendencia y un tipo de factura. Además, puede ser exenta, no exenta, sujeta, no sujeta, de entrega y de prestación de servicios.
- Una factura recibida contiene una clave de régimen especial o trascendencia, un tipo de factura y una fecha de operación. Además, puede ser de inversión de

sujeto pasivo y desglose IVA y tiene asociado una cuota deducible y una fecha de registro contable.

- Ambos tipos de facturas contienen un estado, que puede ser correcta, parcialmente correcta o incorrecta, el país de la contraparte de la factura, el periodo de liquidación de la factura, el envío al que pertenecen, la fecha de expedición de la factura y la fecha de operación. Además, tienen asociado el importe total de la factura.

4.1.2. Implementación de proyecto de base de datos

Para satisfacer los requisitos anteriores se genera la siguiente estructura de base de datos:

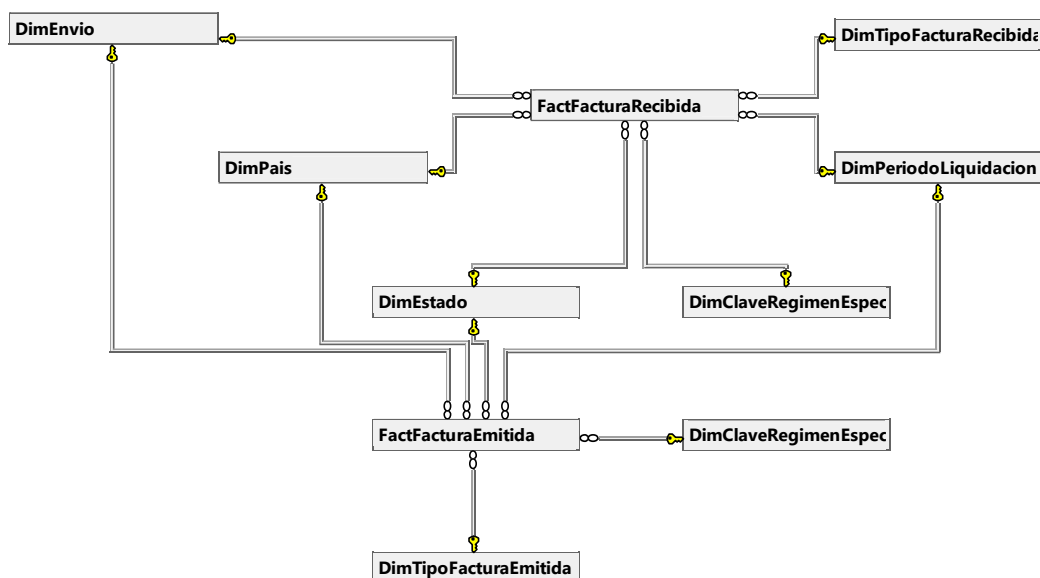


Ilustración 30. Diagrama de almacén de datos

4.1.2.1. Dimensiones

A continuación, se definirán las diferentes tablas que contendrán los datos de las dimensiones del almacén de datos. Todas ellas tendrán definidas una clave autonumérica que las identificará y una clave alternativa que identifica el registro en el origen de datos relacional.

La dimensión fecha se generará mediante la herramienta de Analysis Services y no será necesario generar tabla.

DimClaveRegimenEspecialOTrascendenciaEmitida

En esta dimensión se identifica el tipo de régimen del IVA o una operación con trascendencia tributaria para las facturas emitidas.

Su relación es de uno a muchos con respecto a la tabla de hechos de Facturas Emitidas y se compone de las claves comunes a todas las dimensiones y de una columna de descripción.



Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
ClaveRegimenEspecialOTrascendencia	nvarchar (255)			

Ilustración 31. Tabla DimClaveRegimenEspecialOTrascendenciaEmitida

DimClaveRegimenEspecialOTrascendenciaRecibida

En esta dimensión se identifica el tipo de régimen del IVA o una operación con trascendencia tributaria para las facturas recibidas.

Se relaciona con la tabla de hechos de Facturas Recibidas de forma uno a muchos y además de las claves de dimensiones tiene un campo de descripción.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
ClaveRegimenEspecialOTrascendencia	nvarchar (255)			

Ilustración 32. Tabla DimClaveRegimenEspecialOTrascendenciaRecibida

DimEstado

Esta dimensión contiene los estados que puede tener una factura. Esta dimensión se crea desnormalizada, ya que tiene el estado global de la factura y la descripción del error de la factura, en caso de que este exista. Además, se crean dos errores nuevos para identificar las facturas sin error o las facturas cuyo error sea heredado desde el envío.

La tabla contiene los campos identificativos de las dimensiones, el error global de la factura y la descripción del error y su relación será de uno a uno tanto con la tabla de hechos de facturas emitidas, como con la tabla de hechos de facturas recibidas.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
Estado	nvarchar (50)	No		
DescripcionErrorRegistro	nvarchar (1000)	No		

Ilustración 33. Tabla DimEstado

DimPais

Dimensión que contiene los países a los que va destinada una factura. Estos países pertenecen al destinatario de la factura en caso de facturas emitidas y al emisor de la factura en caso de facturas recibidas. Se relaciona de forma uno a muchos con las tablas de hechos de las facturas y además de las claves de facturas de dimensiones, tiene una columna de descripción.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
Pais	nvarchar (50)	No		

Ilustración 34. Tabla DimPais

DimPeriodoLiquidacion

En esta dimensión se identifica el periodo de liquidación, compuesto por el periodo o mes natural y el ejercicio contable al que pertenece una factura. Su relación es de uno a muchos con las tablas de hechos de facturas, dónde la clave alternativa es la concatenación del identificador del ejercicio más el del periodo.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
Periodo	nvarchar (2)	No		
Ejercicio	nvarchar (4)	No		

Ilustración 35. Tabla DimPeriodoLiquidacion

DimTipoFacturaEmitida

Dimensión que contiene el tipo al que puede pertenecer una factura emitida. Contiene además de las claves de las dimensiones un campo de descripción.

Tiene una relación de uno a muchos con la tabla de hechos de facturas emitidas.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
TipoFactura	nvarchar (255)	No		

Ilustración 36. Tabla DimTipoFacturaEmitida.

DimTipoFacturaRecibida

Dimensión que contiene el tipo al que puede pertenecer una factura recibida. Contiene además de las claves de dimensiones un campo de descripción.

Tiene una relación de uno a muchos con la tabla de hechos de facturas recibidas.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
TipoFactura	nvarchar (255)	No		

Ilustración 37. Tabla DimTipoFacturaRecibida.

DimEnvio

Dimensión que contiene el envío al que pertenece una factura. Esta dimensión se considera una dimensión degenerada, ya que también contiene medidas y se pueden estudiar.

Su relación es de uno a muchos tanto con la tabla de hechos de facturas emitidas como con la tabla de hechos de facturas recibidas.

Además de los identificadores de dimensiones, contiene el campo de identificador único del envío, la empresa que realiza el envío, el distribuidor al que pertenece la empresa, el destino al que va dirigida, el tipo de factura del que se trata, el estado del envío, la fecha en la que se realiza el envío y si el envío se ha enviado al servicio web de pruebas o real.

Esta tabla es desnormalizada, ya que contiene campos que deben pertenecer a una tabla de datos maestros, pero se desnormalizan para mantener el esquema de estrella.

Columna	Tipo	Nulo	Predeterminado	Enlace a
Id	int	No	Identity (1,1)	
IdAlt	int	No		
UID	uniqueidentifier	No		
Empresa	nvarchar (60)	No		
Distribuidor	nvarchar (255)	No		
Destino	nvarchar (255)	No		
Tipo	nvarchar (255)	No		
Estado	nvarchar (255)	No		
Fecha	date	No		
Test	bit	No		

Ilustración 38. Tabla DimEnvio.

4.1.2.2. Hechos

En el siguiente subcapítulo se definirán las tablas de hechos necesarias para estudiar las facturas tanto emitidas como recibidas.

FactFacturaEmitida

En esta tabla de hechos se guardarán los datos de las facturas emitidas por las distintas empresas.

Columna	Tipo	Nulo	Enlace a
Id	int	No	
IdAlt	int	No	
IdPeriodoLiquidacion	int	No	DimPeriodoLiquidacion (Id)
FechaExpedicion FacturaEmisor	date	No	
IdTipoFactura	int	No	DimTipoFacturaEmitida (Id)
FechaOperacion	date	No	
IdClaveRegimen EspecialOTrascendencia	int	No	DimClaveRegimenEspecial oTrascendenciaEmitida (Id)
ImporteTotal	decimal (38,14)	No	
IdPaisContraparte	int	No	DimPais (Id)
Entrega	bit	No	
PrestacionServicios	bit	No	
Sujeta	bit	No	
Exenta	bit	No	
NoExenta	bit	No	
IdEstado	int	No	DimEstado (Id)

Ilustración 39. Tabla FactFacturaEmitida.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

FactFacturaRecibida

La tabla de hechos de facturas recibidas tendrá los datos referentes a las facturas recibidas.

Columna	Tipo	Nulo	Enlace a
Id	int	No	
IdAlt	int	No	
IdPeridoLiquidacion	int	No	DimPeriodoLiquidacion (Id)
FechaExpedicion FacturaEmisor	date	No	
IdTipoFactura	int	No	DimTipoFacturaRecibida (Id)
FechaOperacion	date	No	
IdClaveRegimen EspecialOTrascendencia	int	No	DimClaveRegimenEspecial oTrascendenciaRecibida (Id)
ImporteTotal	decimal (38,14)	No	
IdPaisContraparte	int	No	DimPais (Id)
InversionSujetoPasivo	bit	No	
DesgloseIVA	bit	No	
FechaRegContable	bit	No	
CuotaDeducible	bit	No	
IdEstado	int	No	DimEstado (Id)

Ilustración 40. Tabla FactFacturaRecibida

Una vez estructuradas todas las tablas, se crea un proyecto de base de datos y al ser el primer proyecto, se genera también la solución.

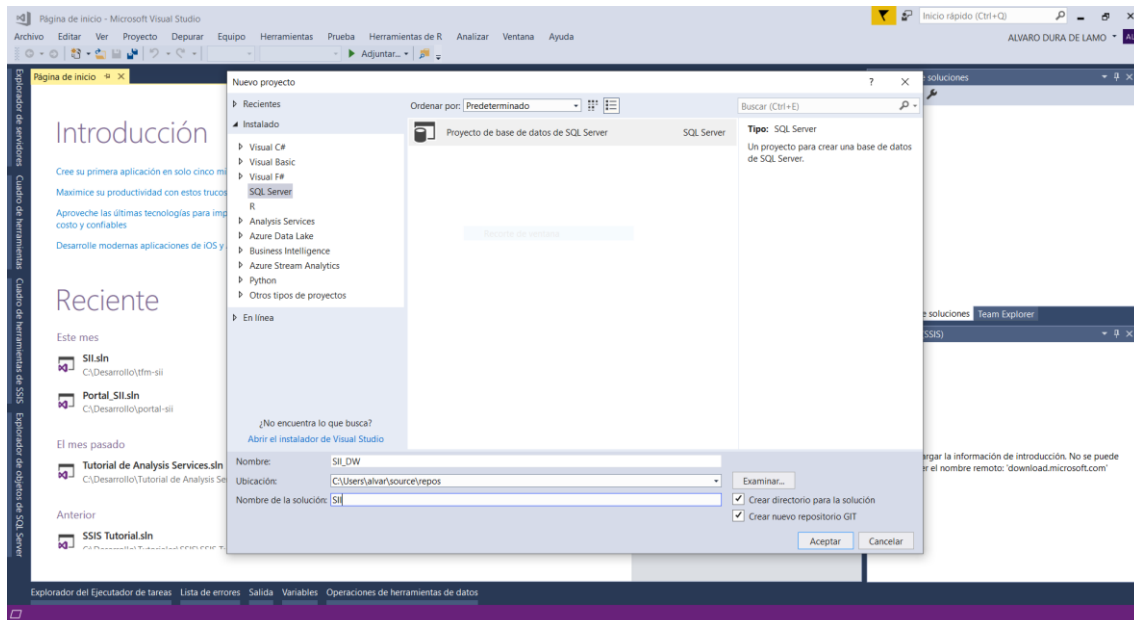


Ilustración 41. Proyecto de base de datos y solución.

El siguiente paso es agregar un archivo de tabla de base de datos al proyecto por cada una de las tablas que se han definido anteriormente.

Para ello se crean dos carpetas nuevas dentro del proyecto, una llamada Dimensiones y otro Hechos, que contendrán las tablas de dimensiones y hechos respectivamente.

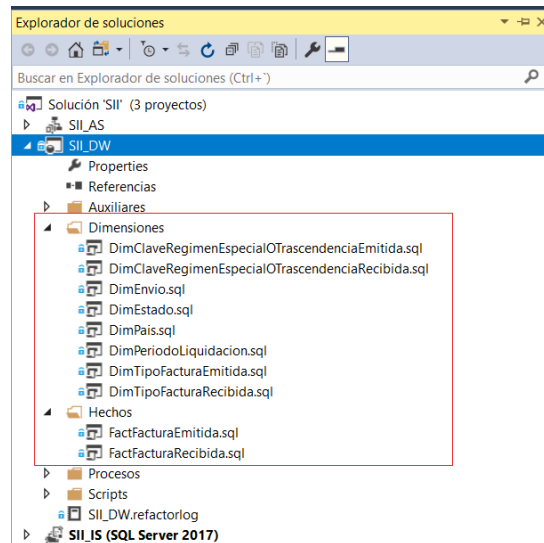


Ilustración 42. Tablas del proyecto DW

4.2. Integración de datos

Mediante este proceso se obtendrán los datos desde el origen de datos relacional y se insertarán en el almacén de datos.

En este proceso se deben utilizar una serie de tablas intermedias o auxiliares dónde volcar los datos desde el origen de datos en una primera instancia y realizar sobre estas tablas las tareas de transformación y limpieza de datos. Mediante este proceso, se libera de carga el origen de datos y se evitan interbloqueos y/o suspensión de procesos.

Se crea una nueva carpeta en el proyecto de base de datos y se inserta en ella los archivos de tabla de base de datos para las tablas auxiliares.

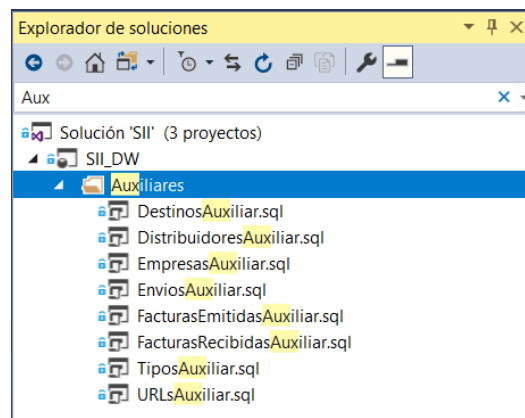


Ilustración 43. Tablas Auxiliares.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Una vez creadas las tablas auxiliares, se crea un nuevo proyecto de SQL Server Integration Services dentro de la solución previamente creada.

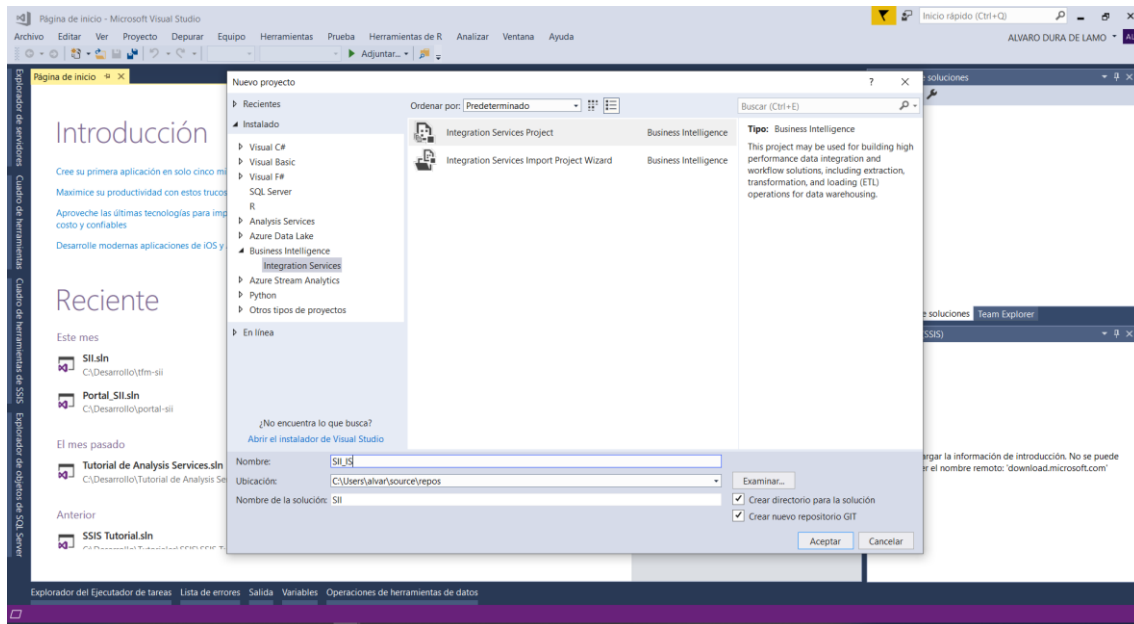


Ilustración 44. Proyecto SSIS

A partir de aquí, se han de generar una serie de procesos, que seguirán un flujo de trabajo, para obtener los datos desde el origen de datos, transformarlos e insertarlos en la estructura de tablas previamente creada.

4.2.1. Obtener el último envío registrado.

Se obtendrá el último envío registrado en el almacén de datos, se guardará en una variable y se utilizará posteriormente para obtener los envíos a partir de este.

Inicialmente se barajó la posibilidad de volver a crear el destino de datos e insertar de nuevo todos los registros, pero dada la gran cantidad de datos, el tipo de origen XML de estos, el tiempo de procesamiento era muy elevado, por lo que, al ser datos estáticos, que una vez generados no varían, se decidió obtener sólo los envíos no procesados.

Este proceso se realiza mediante un proceso de Ejecutar Tarea SQL, que ejecuta una consulta SQL en un origen de datos. Además, se crea una variable que obtiene el resultado y lo almacena para su posterior uso.

4.2.2. Borrar tablas intermedias.

En este proceso, se realizará una limpieza de los datos de las tablas intermedias utilizadas para cargar los datos desde el origen. Para este proceso, se utiliza también el proceso de Ejecutar Tarea SQL, pero esta vez se ejecuta un script de Transact SQL creado previamente.

Estos Scripts se guardan en una nueva carpeta en el proyecto de base de datos llamada Scripts.

4.2.3. Cargar tablas intermedias.

Mediante este proceso, se obtiene los datos del origen de datos y se guardan en las tablas intermedias o auxiliares.

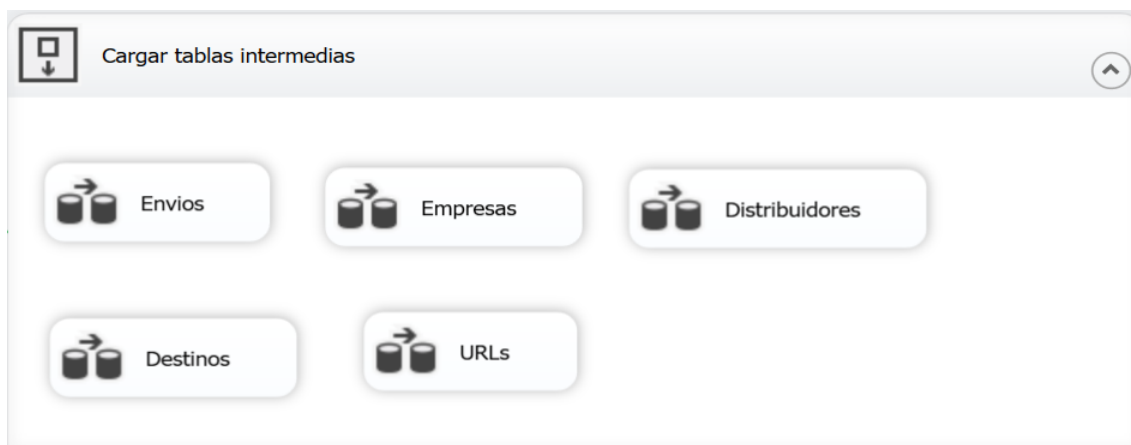


Ilustración 45. Proceso de carga de tablas auxiliares

Para ello se generaron cinco flujos de datos, que contienen dos procesos, uno que obtiene los datos y otro que los inserta.

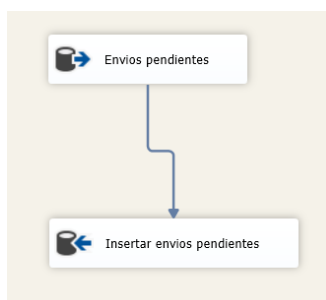


Ilustración 46. Flujo de datos de envíos pendientes

En el caso del flujo de datos de envíos, solo se obtienen los envíos pendientes, mientras que en los otros casos se obtienen todos los datos.

4.2.4. Obtener dimensiones.

El siguiente proceso es el de obtención de dimensiones.

Para realizar las distintas transformaciones de los datos de las tablas intermedias, se intentó utilizar los flujos de datos que proporciona SSIS, pero al utilizar columnas de tipo XML, los procesos se sobrecargaban y nunca terminaban. Es por esto que se optó por realizar las transformaciones mediante proceso de Transact SQL.

Para cada transformación se genera un Script de base de datos en el proyecto anterior y se ejecuta mediante tareas de Ejecutar Tarea SQL.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

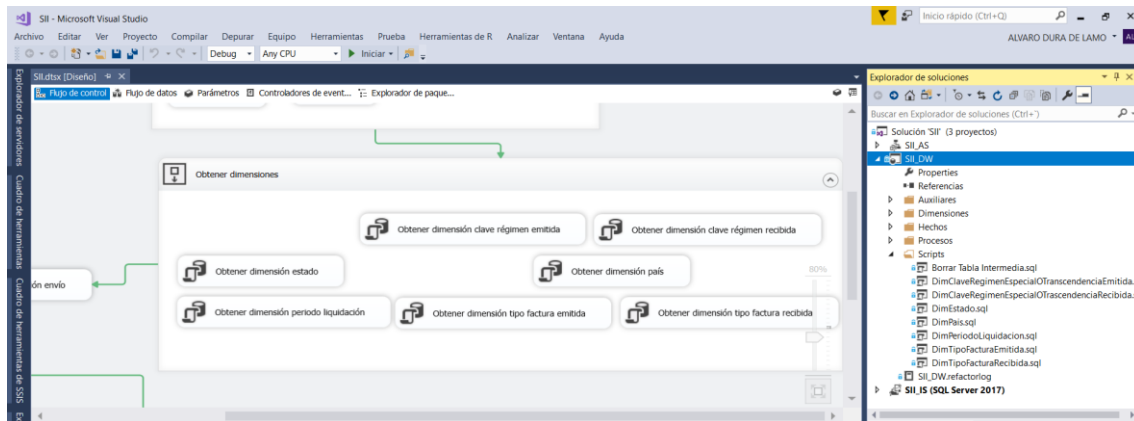


Ilustración 47. Obtención de dimensiones y scripts.

4.2.5. Obtener dimensión envío.

Una vez generadas todas las transformaciones de las tablas de dimensiones, se decide extraer el proceso más costoso, el de obtención de la dimensión envío y realizarlo de manera individual en el siguiente paso, para utilizar toda la potencia de procesamiento en un solo paso.

4.2.6. Integrar tablas de hechos.

Una vez insertadas las dimensiones, se transforman y se insertan los datos de las facturas en los envíos, procesando los campos XML de cada envío, ya que dependen de las tablas de dimensiones para algunos de sus campos. Este proceso, al igual que en las dimensiones, se realiza mediante procesos de bases de datos.

Se ha de tener en cuenta que el proceso de lectura de los XML de las facturas es costoso por defecto, pero en este caso, se han de leer los datos de cinco Web Services distintas, una por cada Hacienda, y cada una de ellas cuenta con su propio esquema XSD, lo que eleva el coste de esta tarea aún más.

Aquí se puede observar una de las ventajas de utilizar Integration Services y es que de manera sencilla se puede ejecutar de manera paralela dos procesos independientes.

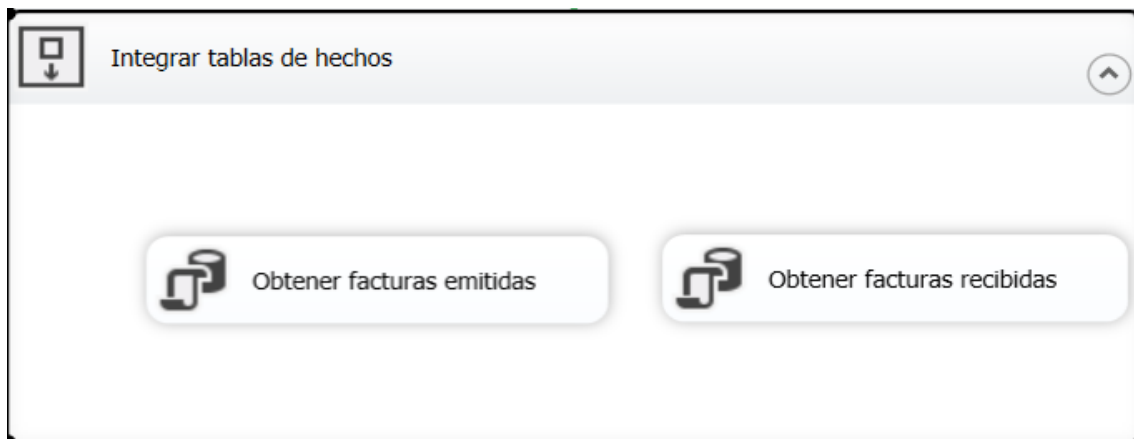


Ilustración 48. Proceso de integración de facturas

4.2.7. Tarea de procesamiento de Analysis Services.

Una vez obtenidos los datos desde el origen de datos e insertado en el almacén de datos, se debe procesar el cubo mediante una tarea de procesamiento de Analysis Services. Esto automatiza el procesamiento del cubo cada vez que se ejecute el proceso de integración.

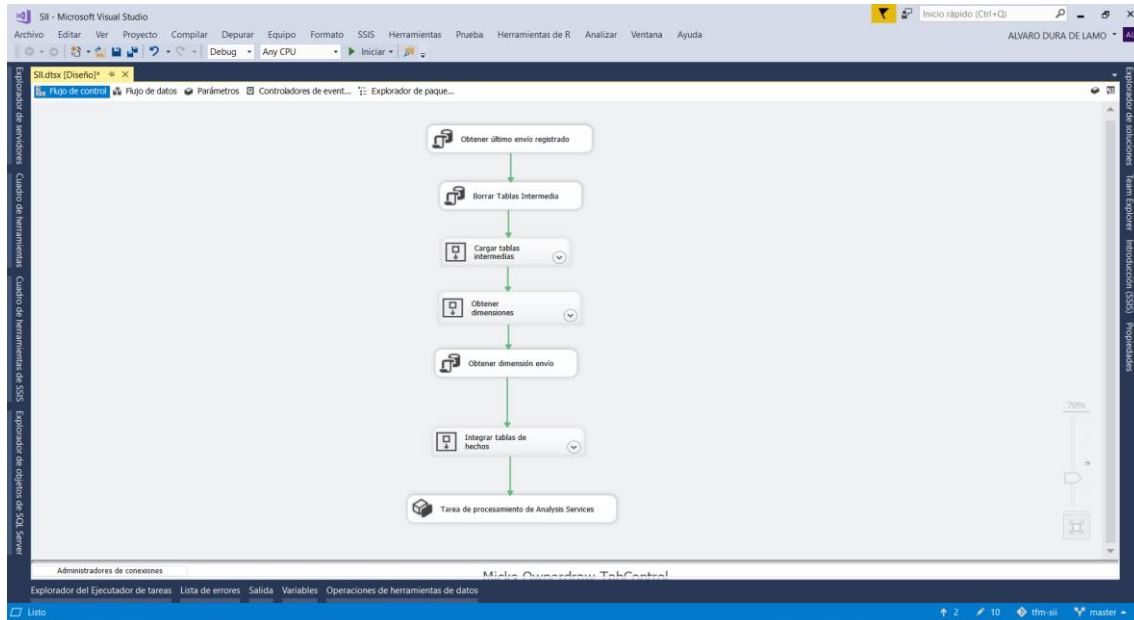


Ilustración 49. Esquema Proceso de Integración

4.3. Análisis de datos

Una vez se ha poblado con datos el almacén de datos, es necesaria la creación de un cubo multidimensional para la posterior explotación de los datos, reduciendo los tiempos de consultas en las aplicaciones.

Para comenzar se ha de generar un proyecto de SQL Server Analysis Services Multidimensional mediante Visual Studio y agregarlo a la solución creada previamente.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

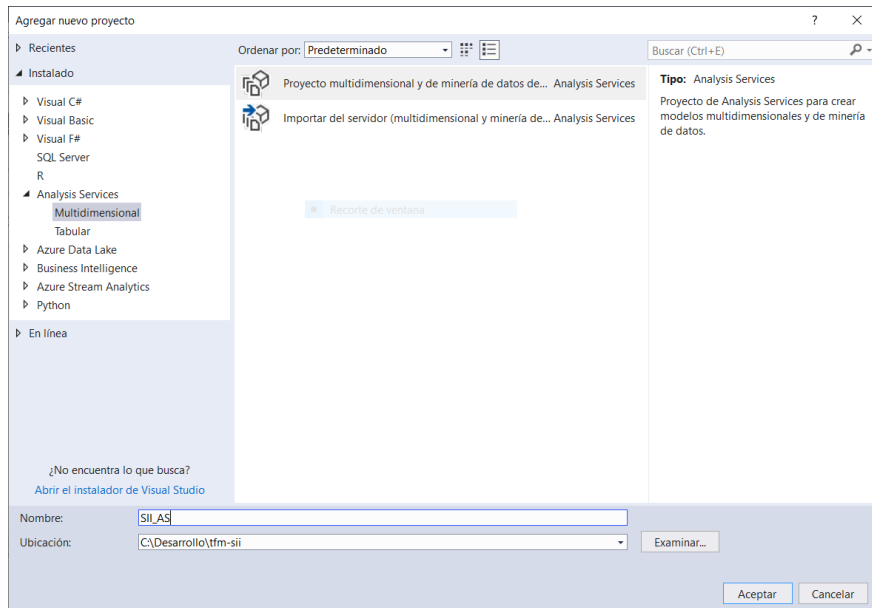


Ilustración 50. Proyecto SSAS multidimensional.

A continuación, con la estructura de tablas bien creada, se deben seguir una serie de pasos para construir el cubo multidimensional.

4.3.1. Orígenes de datos

El primer paso es indicar al proyecto cual/es será/n los orígenes de datos. Para ello Visual Studio provee de un asistente dónde paso a paso se indicará de que tipo es el origen de datos, que nombre tiene, dónde está alojada y la manera de acceder a él. En definitiva, es una asistente para construir una cadena de conexión.

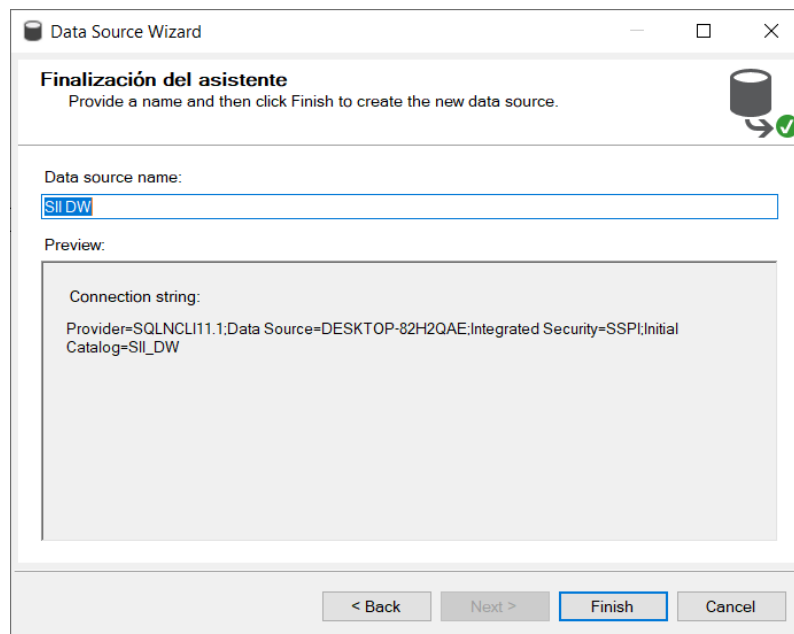


Ilustración 51. Asistente de conexión

4.3.2. Vista del origen de datos.

Una vez se ha configurado un origen de datos, el siguiente paso es crear una vista para este origen de datos. Una vista del origen de datos contiene los metadatos que representan los objetos seleccionados de uno o más orígenes de datos y se pueden crear tantas vistas para el origen de datos como se deseen, normalmente ligadas al número de cubos a implementar, aunque también es posible genera cubos que obtengan los datos de varias vistas.

Para generar las vistas, se utiliza un asistente dónde se indica el/los origen/es de datos, se seleccionan las tablas a utilizar, normalmente seleccionando las tablas de hechos y agregando las tablas de dimensiones relacionadas y se le indica un nombre.

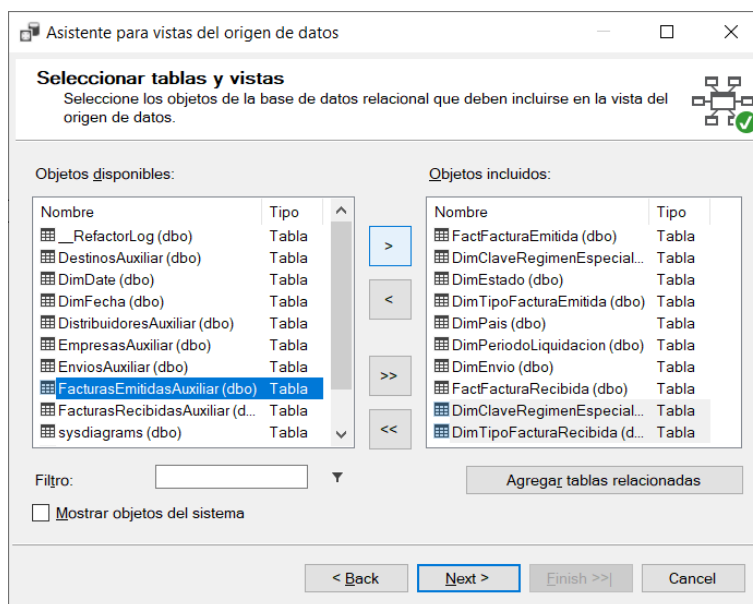


Ilustración 52. Asistente vista del origen de datos.

En este caso se seleccionan las tablas de hechos de facturas y se utiliza el botón Agregar tablas relacionadas para incluir las tablas de dimensiones asociadas.

4.3.3. Cubos

El siguiente paso una vez definidas las vistas del origen de datos es crear el cubo mediante el asistente de creación de cubos de Visual Studio.

Primero se seleccionan los grupos de medidas o tablas de hechos, que en este caso son las tablas de facturas, pero también se incluirá la tabla de envíos, ya que, a pesar de ser una tabla de dimensiones, también es susceptible de ser explotada.

Además, se deben elegir las medidas que se deseen incluir en el cubo, en este caso será el recuento de las medidas, es decir, cuantas facturas, recibidas o emitidas, y envíos. Para las facturas también se incluye el importe total y la cuota deducible para las facturas recibidas.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

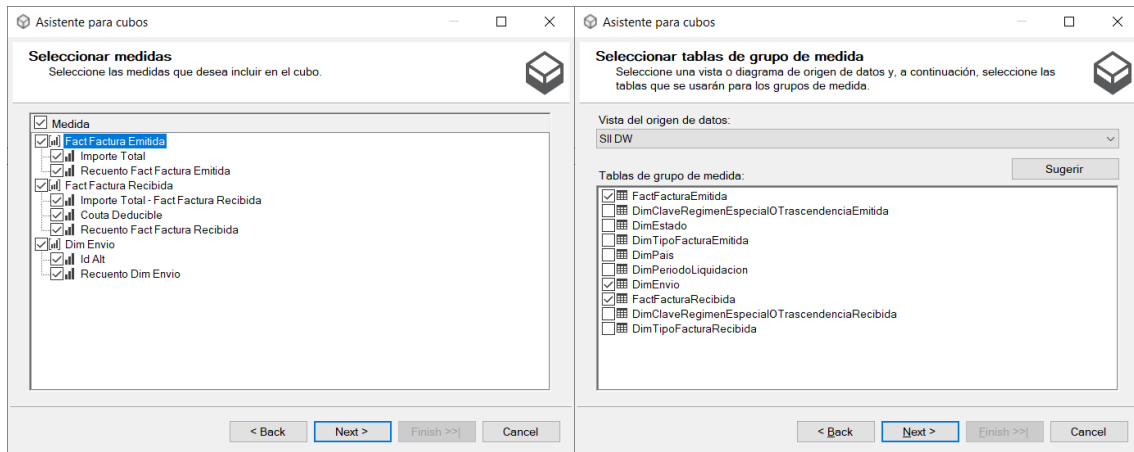


Ilustración 53. Asistente de cubo. Medidas.

El siguiente paso es seleccionar las dimensiones que se quieren agregar al cubo. Se seleccionarán todas, incluidas las tablas de hechos, y que contienen campos que se pueden utilizar para medir los propios hechos.

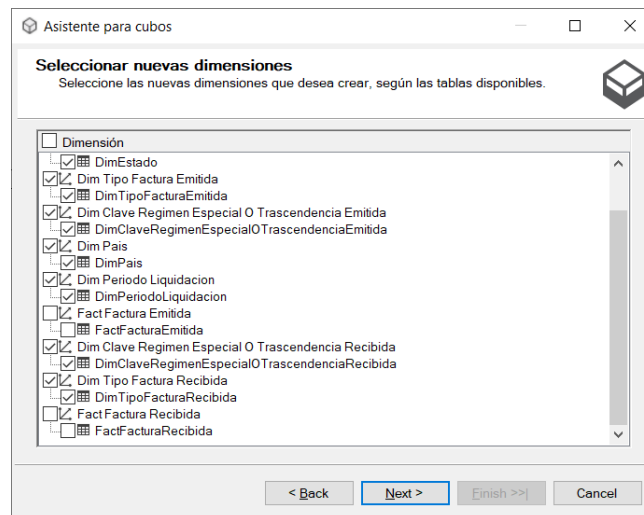


Ilustración 54. Asistente de cubo. Dimensiones.

Una vez finalizado el asistente, Visual Studio generará el cubo y las dimensiones seleccionadas anteriormente.

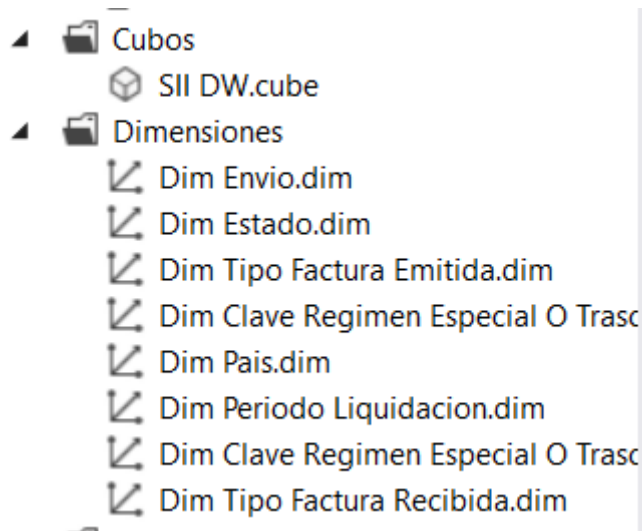


Ilustración 55. Cubo y Dimensiones.

4.3.4. Dimensión Tiempo.

Una vez generados el cubo y las dimensiones, falta generar la dimensión tiempo, ya que no se ha creado desde el origen de datos, sino que se va a generar desde el asistente de Visual Studio, creándose así de manera rápida y sencilla.

Para ello, se abre el asistente de dimensiones y se selecciona la opción de generar una tabla de tiempos en el origen de datos. Mediante esta opción, se indica el rango de fechas a crear, el primer día de la semana, que periodo de tiempos se han de generar (Año, Semestre, Trimestre, Mes, Dia, etc...)

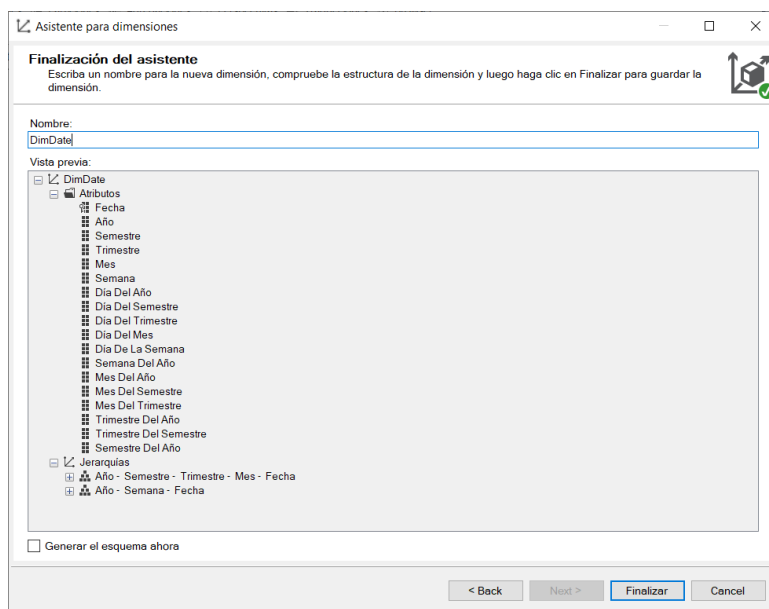


Ilustración 56. Resumen asistente de dimensión fechas.

Con todo esto el asistente genera una nueva dimensión fecha, agregando automáticamente los atributos y las jerarquías, términos que se verán más adelante.

4.3.5. Atributos, jerarquías y relaciones de atributos.

Una vez se tienen definidas las dimensiones, es necesario seleccionar los atributos y jerarquías que se encuentran dentro de cada una de ellas.

Un atributo es un elemento de la dimensión que se encuentra en la tabla de la vista del origen de datos.

Una jerarquía de dimensión ayuda y optimiza el rendimiento del cubo al mejorar las agregaciones. Consiste en definir que atributo está por encima de otro a nivel de herencia. Por ejemplo, un año está por encima de un semestre.

Con las relaciones de atributos se define la relación de un atributo con el atributo clave de la dimensión. Esta relación puede ser directa o indirecta a través de otros atributos.

Mediante la definición de relaciones y atributos, se consigue por ejemplo que, a la hora de agregar por semestres, los resultados sean por año y semestre, en vez de que sólo por semestre.

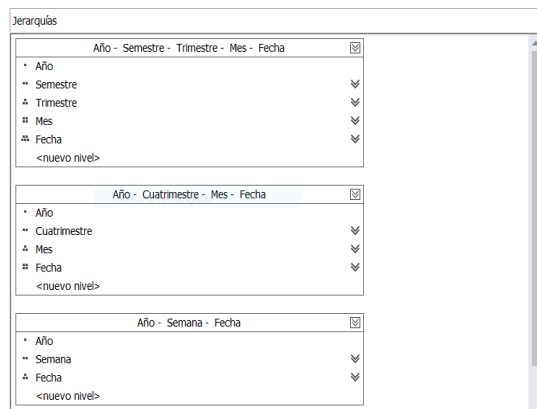


Ilustración 57. Jerarquías

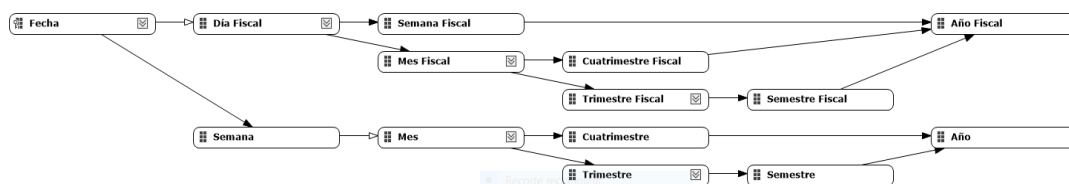


Ilustración 58. Relaciones de atributos.

4.3.6. Cálculos

En este momento ya se pueden generar consultas y obtener resultados del cubo, pero para realizar consultas complejas, como porcentajes del total o recuento total de dos medidas es necesario crear cálculos.

Estos cálculos son consultas creadas mediante el lenguaje MDX.

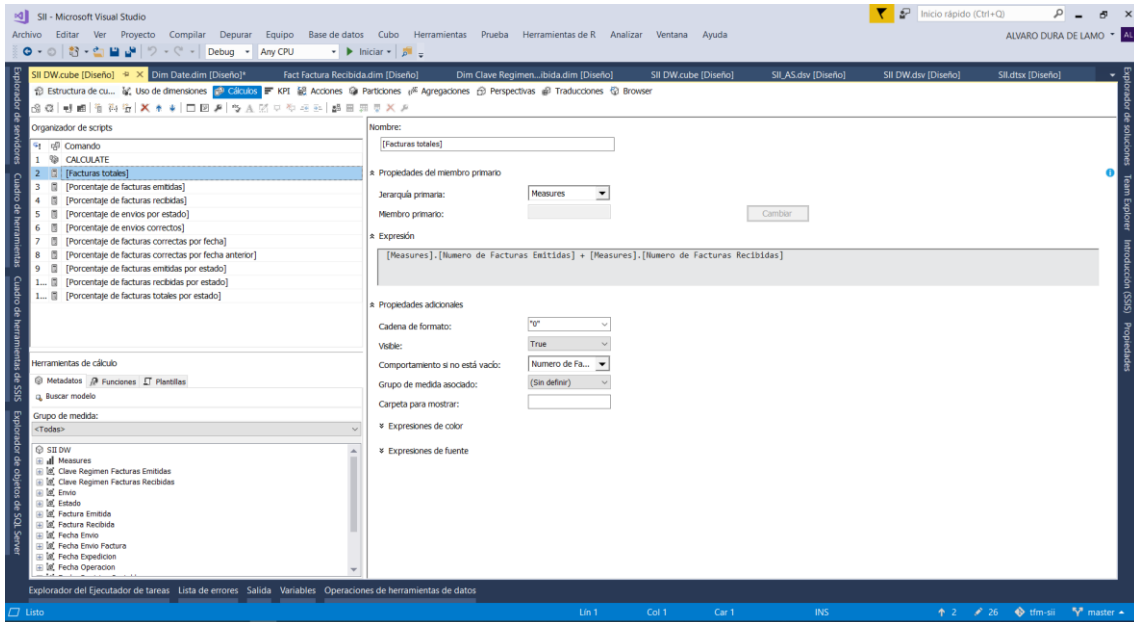


Ilustración 59. Cálculos.

En la imagen superior se puede observar el cálculo de la suma total de facturas emitidas y recibidas.

5. Despliegue

En la siguiente sección se explica cómo una vez estructurados los datos y creado el cubo multidimensional, se muestran a través de dos aplicaciones de visualización de datos.

Se ha decidido desplegar la solución sobre dos plataformas de visualización de datos diferentes con el fin de comprar la facilidad a la hora de desplegar este tipo de aplicaciones, ya que una de ellas requiere cierto nivel de programación, pero a su vez permite una mayor libertad a la hora de realizar consultas y otra permite crear elementos visuales sin necesidad de programar, pero la capacidad de realizar consultas personalizadas es limitada.

5.1. Shiny R

5.1.1. Instalación de Shiny R.

El primer paso para desarrollar la aplicación de visualización de datos mediante Shiny R es instalar el lenguaje de programación R y el entorno de programación de R Studio.

Una vez instalado, se instalan los paquetes:

5.1.1.1. Paquetes para elementos gráficos.

- **shiny**: paquete básico para crear aplicaciones web con R.
- **shinydashboard**: paquete basado en shiny que permite crear paneles de control de manera sencilla.
- **shinywidgets**: inserta una serie de elementos extras para utilizar con shiny.
- **shinydashboardPlus**: añade extensiones para el paquete shinydashboard.
- **c3**: incluye las gráficas de la librería c3 para shiny.

5.1.1.2. Paquetes para conexión con Análisis Services.

Para conectar R con el cubo multidimensional primero se intenta utilizar el paquete **olapR**. Este paquete permite la conexión de manera sencilla a el origen de datos de Análisis Services, pero solo se puede utilizar mediante la instalación de R desde Microsoft R Client, el cual instala la versión 3.5.2, siendo actualmente la versión 3.6.1 la última versión liberada.

Debido a que la versión 3.5.2 no es la actualizada, presenta incompatibilidades con los paquetes de Shiny y se descarta.

La siguiente opción es **RDCOMClient**. Este paquete permite la conexión a clientes COM, entre ellos los cubos multidimensionales de Análisis Services, pero es menos intuitivo que el anterior, por lo que requiere un nivel superior de programación, ya que hay que entrar en profundidad en el manejo de listas y vectores.

RDCOMClient es compatible con la última versión de R y por tanto se utilizará para conectar con la base de datos de Análisis Services.

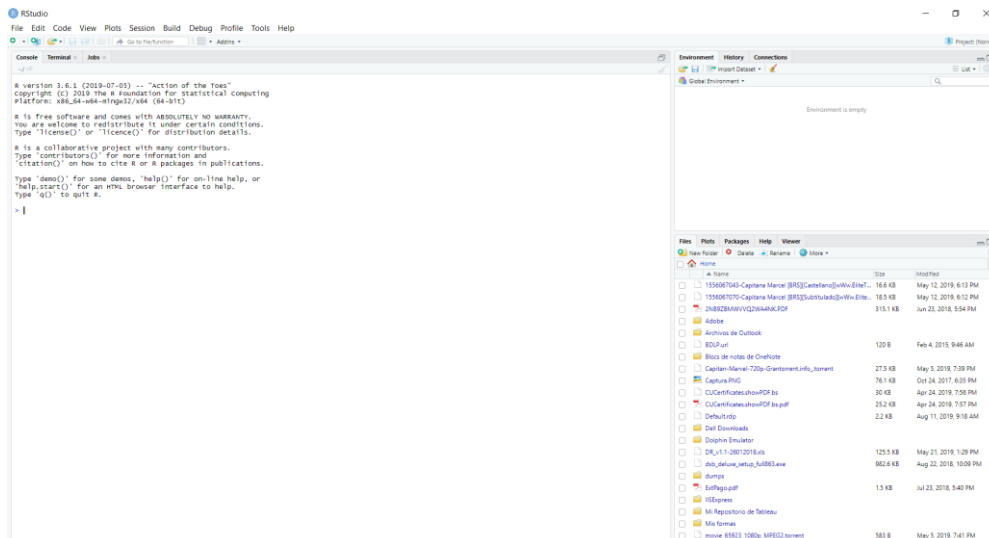


Ilustración 60. R y R Studio.

5.1.1. Estructura de la aplicación.

Una aplicación Shiny R se estructura en 2 ficheros principales, el fichero ui y el fichero server. Además, crean 2 ficheros más de apoyo a estos dos primeros.

5.1.1.1. Fichero ui.

El fichero ui contiene la estructura *front-end* de la aplicación. En ella se declaran los elementos visuales que deben aparecer en la aplicación web, como pueden ser las gráficas, cajas de texto, *widgets*, etc....

5.1.1.2. Fichero server.

El fichero server contiene la estructura *back-end* de la aplicación. Este fichero es el encargado de rellenar con datos los controles instanciados en el fichero ui.

5.1.1.3. Fichero source.

En este fichero se crean las funciones necesarias para conectar con la base de datos de Analisis Services y se realizan las consultas pertinentes. Además, estas funciones se utilizan en el fichero server para obtener los datos.

5.1.1.4. Fichero KPI.

El fichero KPI contiene la instancia de la clase KPI. Este fichero genera un KPI y aplica las modificaciones necesarias para representarlo de manera correcta.



Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

5.1.2. Capturas Aplicación WEB Shiny R



Ilustración 61. Página principal Shiny R.

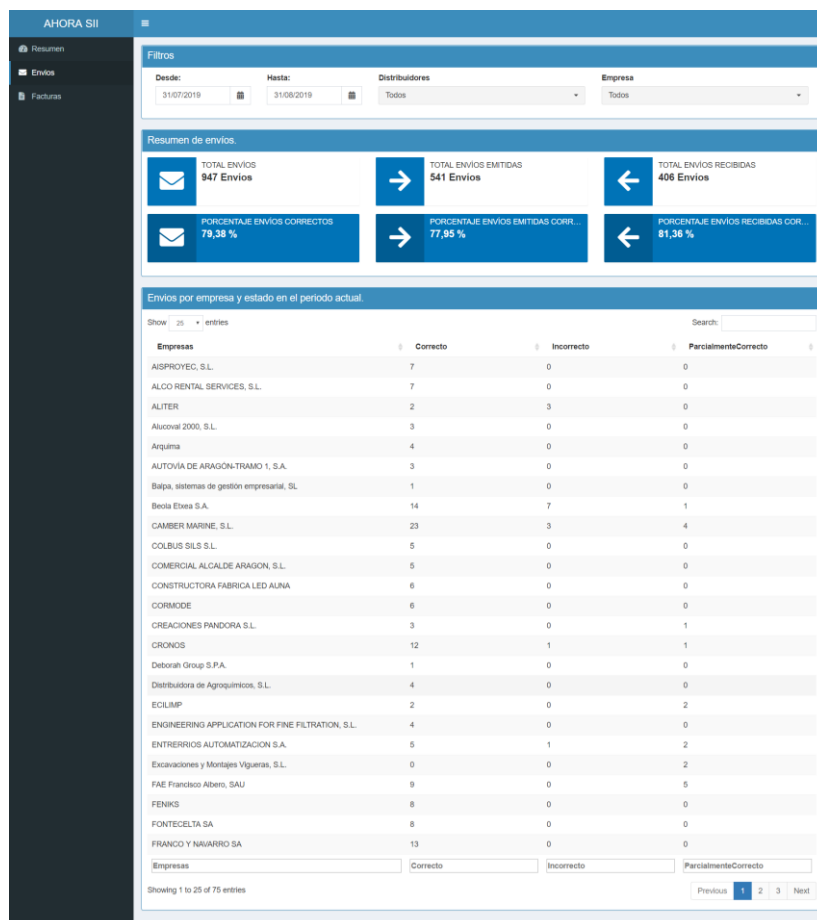


Ilustración 62. Página envíos Shiny R.

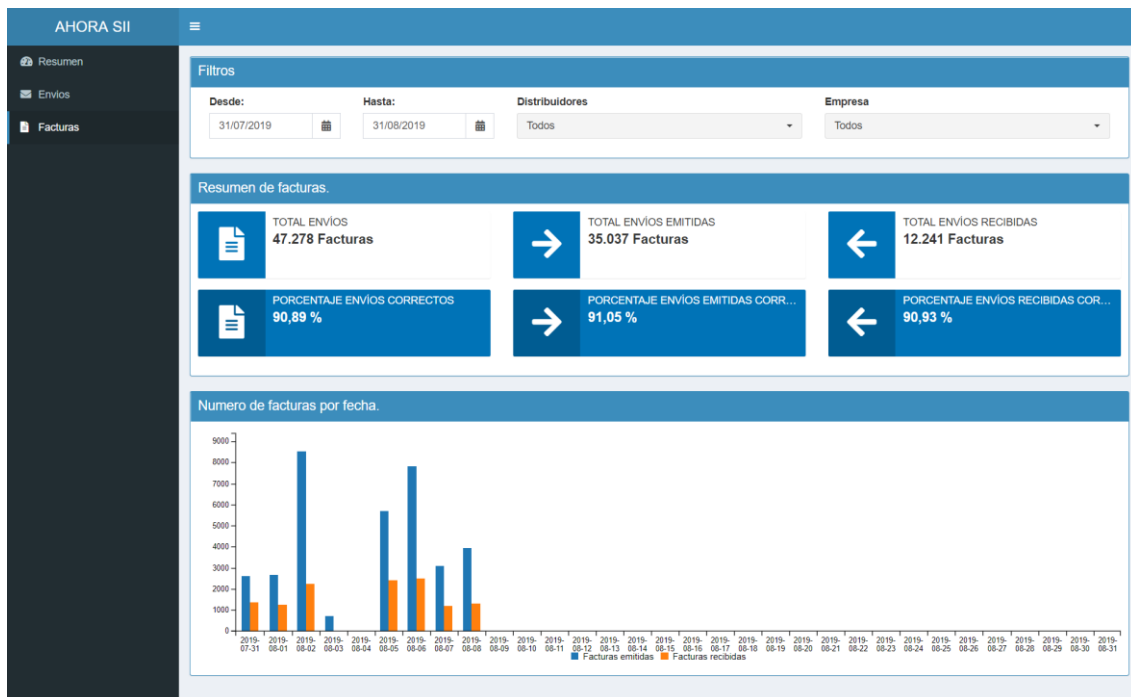


Ilustración 63. Pagina facturas Shiny R

5.2. PowerBI

5.2.1. Instalación de PowerBI

Para utilizar la herramienta de PowerBI y visualizar los informes, es necesario instalar PowerBI Desktop, en la versión de mayo de 2019 y PowerBI Server.

5.2.1.1. PowerBI Desktop

La instalación de PowerBI Desktop se realiza mediante el instalador proporcionado de manera sencilla. No es necesario instalar ningún paquete adicional como en R.

Esta aplicación permitirá conectar con el origen de datos del cubo multidimensional y generar los elementos de información para la aplicación de manera sencilla sin necesidad conocimientos de programación.

5.2.1.2. PowerBI Server

PowerBI Server permite, una vez generados los informes desde PowerBI Desktop, publicarlos en un servidor web para hacerlo accesible a los usuarios.

5.2.1.3. Cálculos adicionales

Puesto que en PowerBI no podemos programar las consultas, es necesario generar medidas calculadas previamente para poder obtener porcentajes y totales. Para ello crearemos las siguientes medidas calculadas.

Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Facturas totales

- Suma de facturas emitidas y recibidas.
- Facturas totales por periodo actual
- Facturas totales por periodo anterior

Porcentajes

- Porcentaje de envíos correctos
- Porcentaje de envíos por estado
- Porcentaje de facturas correctas por periodo
- Porcentaje de facturas correctas por periodo anterior
- Porcentaje de facturas emitidas correctas por periodo
- Porcentaje de facturas emitidas correctas por periodo anterior
- Porcentaje de facturas recibidas correctas por periodo
- Porcentaje de facturas recibidas correctas por periodo anterior

5.2.2. Capturas de Aplicación WEB PowerBI

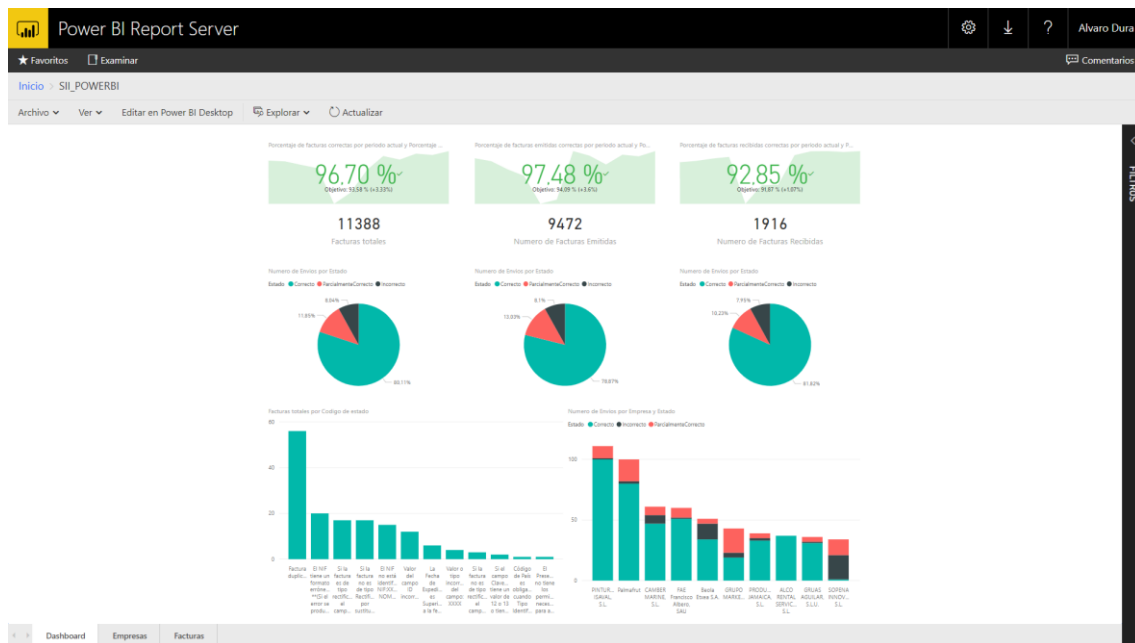


Ilustración 64. Página principal PowerBI

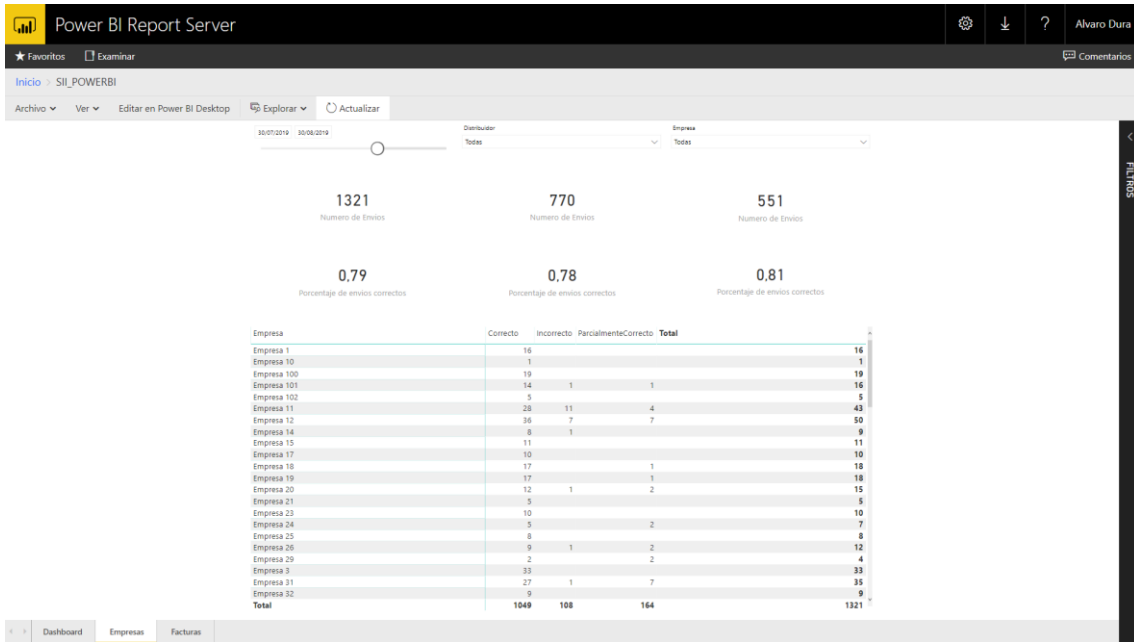


Ilustración 65. Empresas PowerBI

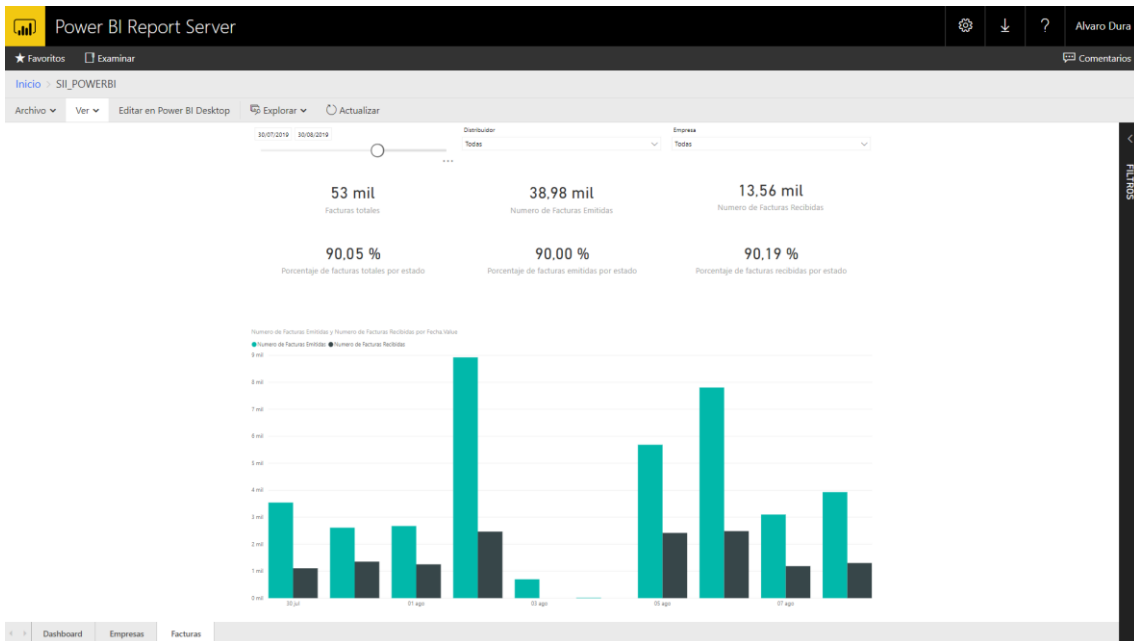


Ilustración 66. Facturas PowerBI



5.2.3. Comparación entre los dos sistemas.

En los dos sistemas se ha podido generar la aplicación web especificada en el análisis funcional, por tanto, se puede decir que las dos aplicaciones son válidas para solucionar el problema propuesto en el trabajo.

PowerBI al ser un sistema de código propietario, que además está destinado para usuario final sin grandes nociones de programación, limita las posibilidades a la hora de crear nuevos elementos gráficos y obtener los datos desde el origen, pero permite crear la aplicación web de manera sencilla, tanto los elementos gráficos como la propia aplicación.

Por otro lado, Shiny R, al estar basado en código abierto, y sobre todo la posibilidad de programar los elementos desde el propio código, permite una mayor libertad a la hora de crear las consultas al sistema de origen de datos y la oportunidad al programador de generar elementos más personalizados, pero el sistema es más costoso de implementar, ya que requiere el desarrollo de la aplicación mediante código pero además, al contrario que PowerBI Server, es necesaria la configuración de un servidor para arrancar la aplicación.

6. Conclusión

Una vez terminado el proyecto se puede decir que los objetivos planteados al inicio del proyecto han sido completados.

- Se ha estudiado y entendido la plataforma actual del SII implementada por la empresa AHORA.
- Se ha creado un almacén de datos a partir de los requisitos obtenidos de la empresa.
- Se ha generado un proceso ETL, extracción, transformación y carga, que inserta los datos necesarios en el almacén de datos.
- Se ha creado un cubo multidimensional a partir del almacén de datos, mejorando considerablemente el acceso a los datos.
- Se han implementado dos aplicaciones web para la visualización de datos, para la ayuda en la toma de decisiones y se ha creado una comparativa entre estas dos.

Por último, se ha adquirido experiencia en todas las herramientas utilizadas para generar cada una de las partes que componen este trabajo.

6.1. Dificultades

Durante la realización de este trabajo se han encontrado diversas dificultades a la hora de crear la solución, destacando:

- Los envíos guardados en XML dentro de tablas SQL Server han hecho imposible la utilización de algunas de las herramientas que ofrece Integration Services, ya que los procesos no terminaban. Por ello, se ha tenido que rehacer el proceso de integración, perdiendo una cantidad de tiempo considerable.
- A la hora de generar la aplicación web para PowerBI, la imposibilidad de generar consultas MDX contra las bases de datos multidimensionales, ha aumentado la dificultad a la hora de generar el cubo multidimensional.

6.2. Líneas futuras

Una vez terminado el proyecto, se consideran las siguientes líneas futuras:

- Aumentar el número de gráficos e informes en la aplicación web, estudiando importes, situaciones geográficas, etc....
- Generar una aplicación web con ingreso por usuarios y añadir seguridad.
- Crear sistemas de predicción que ayuden a la toma de decisiones futuras.



Análisis de envíos a la plataforma de Suministro de Información Inmediata de Hacienda con herramientas de Business Intelligence.

Bibliografía

- [1] SOLO PIENSO EN TIC, «solo pienso en TIC,» 24 Octubre 2017. [En línea]. Available: <http://www.solopiensoentic.com/casos-de-exito-de-business-intelligence/>.
- [2] Analítica de Retail, «Analítica de retail,» 10 Febrero 2019. [En línea]. Available: <http://analiticaderetail.com/casos-de-exito-business-intelligence/>.
- [3] Agencia Tributaria, «www.agenciatributaria.es,» [En línea]. Available: https://www.agenciatributaria.es/AEAT.internet/Inicio/Ayuda/Modelos_Procedimientos_y_Servicios/Ayuda_P_G417_IVA_Llevanza_de_libros_registro_SII/Informacion_general/Nuevo_sistema_de_gestion_del_IVA_basado_en_el_Suministro_Inmediato_de_Informacion.sh.
- [4] FUNDACIÓN COTEC PARA LA INNOVACIÓN, «GENERACIÓN DE TALENTO BIG DATA EN ESPAÑA,» Madrid, 2017.
- [5] R. Kimball y M. Ross, The Data Warehouse Toolkit, Indianapolis: John Wiley & Sons, Inc., 2013.
- [6] Oracle Corporation, «<https://dev.mysql.com>,» Oracle Corporation and/or its affiliates, 2019. [En línea]. Available: <https://dev.mysql.com/doc/refman/8.0/en/what-is-mysql.html>. [Último acceso: 4 8 2019].
- [7] P. Martín, «<https://www.itop.academy>,» iTop Academy, 24 Enero 2019. [En línea]. Available: <https://www.itop.academy/blog/item/que-es-pentaho-data-integration-pdi-y-para-que-sirve.html>. [Último acceso: 4 Agosto 2019].
- [8] Logicalis, «<https://blog.es.logicalis.com>,» 24 Julio 2015. [En línea]. Available: <https://blog.es.logicalis.com/analytics/cubos-olap-y-estructuras-multidimensionales-todo-lo-que-hay-que-saber>. [Último acceso: 5 Agosto 2019].
- [9] Microsoft, «<https://docs.microsoft.com>,» Microsoft, 02 Mayo 2018. [En línea]. Available: <https://docs.microsoft.com/es-es/sql/analysis-services/multidimensional-models-olap-logical-cube-objects/partitions-partition-storage-modes-and-processing?view=sql-server-2017>. [Último acceso: 5 Agosto 2019].
- [10] Microsoft, «<https://docs.microsoft.com>,» 12 Agosto 2018. [En línea]. Available: <https://docs.microsoft.com/es-es/sql/analysis-services/analysis-services?view=sql-server-2017>. [Último acceso: 5 Agosto 2019].
- [11] Oracle, «<https://www.orafaq.com>,» 21 Julio 2012. [En línea]. Available: https://www.orafaq.com/wiki/Oracle_OLAP. [Último acceso: 5 Agosto 2019].



- [12] Pentaho, «<https://mondrian.pentaho.com/>,» 1 Agosto 2006. [En línea]. Available: <https://mondrian.pentaho.com/documentation/olap.php>. [Último acceso: 5 Agosto 2019].
- [13] M. Rouse, «<https://techtarget.com/>,» TechTarget, 1 Diciembre 2018. [En línea]. Available: <https://searchcontentmanagement.techtarget.com/definition/Microsoft-Power-BI>. [Último acceso: 5 Agosto 2019].
- [14] Qlick Sense, «<https://www.qlik.com/>,» [En línea]. Available: <https://www.qlik.com/es-es/products/qlik-sense>. [Último acceso: 05 Agosto 2019].
- [15] Tableau Software, «<https://www.tableau.com/>,» [En línea]. Available: <https://www.tableau.com/products>. [Último acceso: 5 Agosto 2019].
- [16] Bitbucket, «Atlassian,» [En línea]. Available: <https://www.atlassian.com/git/tutorials/what-is-git>. [Último acceso: 6 Agosto 2019].
- [17] GitLab, «gitlab.com,» [En línea]. Available: <https://about.gitlab.com/what-is-gitlab/>. [Último acceso: 6 Agosto 2019].
- [18] Microsoft, «docs.microsoft.com,» 26 Julio 2019. [En línea]. Available: <https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-2017>. [Último acceso: 6 Agosto 2019].
- [19] The R Foundation, «www.r-project.org,» [En línea]. Available: <https://www.r-project.org/about.html>. [Último acceso: 6 Agosto 2019].
- [20] IT Knowledge portal, «www.itinfo.am,» [En línea]. Available: <http://www.itinfo.am/eng/software-development-methodologies/>. [Último acceso: 7 Agosto 2019].
- [21] Scrum.org, «www.scrum.org,» [En línea]. Available: <https://www.scrum.org/resources/what-is-scrum>. [Último acceso: 7 Agosto 2019].
- [22] V. Dertiano, «biggeek,» 22 Septiembre 2014. [En línea]. Available: <https://blog.biggeek.com/que-es-business-intelligence/>.
- [23] guru99, «https://www.guru99.com,» Guru99, [En línea]. Available: <https://www.guru99.com/data-warehousing.html>. [Último acceso: 4 Agosto 2019].