

Análisis, diseño e implementación de un screener de acciones de empresas mediante un *Sistema de Inteligencia de Negocio*



Alumno:
Aridany Ramírez Rodríguez
Grado de Ingeniería Informática

Área:
Business Intelligence

Consultor:
Xavier Martínez Fontes

FECHA 14 de Junio de 2017



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

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Análisis, diseño e implementación de un screener de acciones de empresas mediante un Sistema de Inteligencia de Negocio</i>
Nombre del autor:	<i>Aridany Vicente Ramírez Rodríguez</i>
Nombre del consultor/a:	<i>Xavier Martínez Fontes</i>
Fecha de entrega (mm/aaaa):	<i>06/2017</i>
Titulación:	<i>Grado de Ingeniería Informática</i>
Área del Trabajo Final:	<i>Business Intelligence</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave:	<i>Business Intelligence, ETL, cubo OLAP, inteligencia de negocio, data warehouse</i>

Resumen del Trabajo:

Los mercados financieros son espacios donde se realizan intercambios de activos financieros. El objetivo de estos espacios físicos o virtuales es poner en contacto a compradores y vendedores, y establecer un precio justo a los activos financieros que se intercambian.

Los mercados están formados por múltiples participantes, entre ellos, los inversores institucionales, que habitualmente son instituciones financieras (bancos, compañías de seguros, fondos de inversión, etc) que invierten en nombre de otros inversores, e inversores minoristas, que son personas físicas (ahorradores) que invierten a título individual.

Las bolsas de valores son mercados de capitales organizados y regulados, donde se realizan operaciones de compraventa de acciones, entre otros activos. Según los datos del Banco Mundial, en el año 2016, en todo el mundo existían 43.192 empresas cuyas acciones cotizaban en un mercado bursátil.

Un ahorrador que quisiera comprar acciones, puede encontrarse con una enorme dificultad para tomar una decisión de inversión debido al gran número de empresas existentes y la idoneidad de estas para su estrategia de inversión.

Este trabajo de fin de grado pretende aplicar las técnicas y desarrollos tecnológicos propios del Business Intelligence (BI), para ayudar a los inversores particulares a acotar el universo de acciones en las que invertir, mediante una herramienta de análisis que le permita aplicar una serie de condiciones o restricciones que deben cumplir las empresas seleccionadas.

Abstract:

Financial markets are spaces where exchanges of financial assets take place. The objective of these physical or virtual spaces is to put in contact buyers and sellers, and establish a fair price to the financial assets that are exchanged.

The markets are made up of multiple players, including institutional investors, usually financial institutions (banks, insurance companies, investment funds, etc.) that invest on behalf of other investors, and retail investors, which are individuals (savers) investing individually.

Stock exchanges are organized and regulated capital markets, where shares are traded, among other assets. According to The World Bank data, in 2016, there were 43,192 companies in the world whose shares were listed on a stock market.

An private investor who would like to buy stocks may find it extremely difficult to make an investment decision because of the large number of existing companies and the suitability of these for their investment strategy.

This work aims to apply the techniques and technological developments of Business Intelligence (BI), to help private investors to reduce the universe of possible stocks to invest in, through an analysis tool that allows him to apply a series of conditions or restrictions to be fulfilled by the selected companies.

Índice de contenidos

1. Introducción.....	10
1.1. Contexto y justificación del trabajo.....	10
1.2. Objetivos del trabajo.....	13
1.3. Enfoque y método elegido.....	14
1.4. Planificación del trabajo.....	14
1.5. Breve resumen de productos obtenidos.....	15
2. Inteligencia de negocio.....	17
2.1. ¿Qué es la inteligencia de negocio?.....	17
2.2. El ciclo de vida Kimball.....	18
3. Elección de herramientas.....	21
3.1. PostgreSQL.....	21
3.2. Pentaho Community Edition.....	22
4. Los inversores y la búsqueda de la rentabilidad en los mercados.....	23
4.1. El análisis técnico.....	24
4.1.1. Las herramientas del análisis técnico cuantitativo.....	25
4.1.2. Indicadores de volatilidad.....	26
4.1.3. Indicadores de tendencia.....	26
4.1.4. Indicadores de momento.....	28
4.1.5. Indicadores de volumen.....	29
4.1.6. Elección de indicadores técnicos.....	29
4.2. El análisis fundamental.....	30
4.2.1. Los ratios financieros en el análisis fundamental.....	30
4.2.2. Ratios de rentabilidad.....	31
4.2.3. Ratios de liquidez.....	32
4.2.4. Ratios de solvencia.....	33
4.2.5. Ratios bursátiles.....	33
4.2.6. Elección de ratios financieros.....	34
5. Análisis de las fuentes de información.....	35
5.1. Cotizaciones históricas (Yahoo Finance).....	35
5.2. Informes trimestrales/anuales de las empresas.....	36
5.3. Otras fuentes de información.....	41
5.3.1. Empresas de los mercados Nasdaq, Nyse y Amex.....	41
5.3.2. Lista ISO de países y estados/provincias.....	41
5.3.3. Lista de símbolos (ticker).....	42
6. Diseño del almacén de datos.....	43

6.1. Modelado dimensional.....	43
6.1.1. Seleccionar proceso de negocio.....	44
6.1.2. Establecer la granularidad.....	44
6.1.3. Identificar las dimensiones.....	45
6.1.4. Identificar los hechos.....	45
6.2. Detalle de las dimensiones.....	46
6.2.1. Dimensión Tiempo.....	46
6.2.2. Dimensión Empresa.....	47
6.3. Modelado físico.....	48
7. Implementación del almacén de datos.....	57
7.1. Procesos ETL.....	57
7.1.1. Carga del data warehouse.....	57
7.1.2. Validación de la base de datos.....	58
7.1.3. Carga de datos del área de “stage”.....	58
7.1.4. Carga de las dimensiones.....	59
7.1.5. Carga de los hechos.....	59
7.1.6. Descargar datos financieros.....	60
7.1.7. Carga tabla de referencias.....	61
7.1.8. Carga tabla SUB.....	62
7.1.9. Carga tabla NUM.....	63
7.1.10. Carga de empresas.....	64
7.1.11. Carga de los códigos CIK.....	65
7.1.12. Carga de cotizaciones históricas.....	66
7.1.13. Carga de la dimensión tiempo.....	67
7.1.14. Carga de la dimensión empresa.....	68
7.1.15. Carga de la tabla de hechos financieros fundamentales.....	69
7.1.16. Carga de la tabla de hechos técnicos.....	71
7.2. Resultados obtenidos en la carga de la ETL.....	74
7.3. Análisis OLAP.....	75
7.3.1. Definición de cubos OLAP.....	76
7.3.2. Saiku Analytics.....	78
7.4. Herramienta de análisis: Screener de acciones.....	81
7.4.1. Diseño.....	82
7.4.2. Implementación.....	83
8. Trabajo futuro.....	88
9. Conclusiones.....	90
10. Glosario.....	92
11. Anexo I: Scripts creación BBDD.....	94

Índice de ilustraciones

Ilustración 1: Diagrama del ciclo de vida Kimball.....	19
Ilustración 2: Fórmula para el cálculo del índice de fuerza relativa (RSI).....	28
Ilustración 3: Información financiera en la Securities Exchange Commission (SEC).....	38
Ilustración 4: Web scraping: Obtención del código CIK, a partir del ticker.....	42
Ilustración 5: Tabla de hechos técnicos.....	45
Ilustración 6: Tabla de hechos fundamentales.....	46
Ilustración 7: Modelo físico.....	49
Ilustración 8: Modelo físico: Área stage.....	49
Ilustración 9: Job principal de la ETL.....	57
Ilustración 10: Job validación de la base de datos.....	58
Ilustración 11: Job carga del área de "stage".....	59
Ilustración 12: Job carga de las dimensiones.....	59
Ilustración 13: Job carga de las tablas de hechos.....	60
Ilustración 14: Transacción: descarga de información financiera de la SEC.....	61
Ilustración 15: Transacción: carga de la tabla de referencias.....	61
Ilustración 16: ETL: Tabla stg_data.....	62
Ilustración 17: Transacción: carga del conjunto de datos SUB.....	62
Ilustración 18: ETL: Tabla stg_sub.....	63
Ilustración 19: Transacción: carga del conjunto de datos NUM.....	63
Ilustración 20: ETL: Tabla stg_num.....	64
Ilustración 21: Transacción: descarga de las empresas del AMEX, NASDAQ y NYSE.....	64
Ilustración 22: ETL: Tabla stg_companies.....	65
Ilustración 23: Transacción: obtención de códigos CIK.....	66
Ilustración 24: Transacción: descarga de las cotizaciones históricas.....	66
Ilustración 25: ETL: Tabla stg_quotes.....	67
Ilustración 26: Transacción: carga de la dimensión tiempo.....	68
Ilustración 27: ETL: Dimension tiempo (d_time).....	68
Ilustración 28: Transacción: carga de la dimensión empresa.....	69
Ilustración 29: ETL: Dimension empresa (d_empresa).....	69
Ilustración 30: Transacción: carga de la tabla de hechos financieros fundamentales.....	70
Ilustración 31: ETL: Tabla hechos fundamentales (f_fundamental).....	71
Ilustración 32: Transacción: carga de la tabla de hechos técnicos I.....	72
Ilustración 33: Transacción: carga de la tabla de hechos técnicos II.....	72
Ilustración 34: Transacción: carga de la tabla de hechos técnicos III.....	73
Ilustración 35: ETL: Tabla hechos técnicos (f_tecnico).....	74
Ilustración 36: Schema Workbench.....	76
Ilustración 37: Cubo OLAP: Captura Schema Workbench I.....	77
Ilustración 38: Cubo OLAP: Captura Schema Workbench II.....	78

Ilustración 39: Saiku Analytics: Sectores de la economía.....	79
Ilustración 40: Saiku Analytics: Detalle sector sanitario.....	80
Ilustración 41: Saiku Analytics: Detalle industria: Otras farmaceuticas.....	80
Ilustración 42: Saiku Analytics: Información financiera ABC, CAH, MCK.....	81
Ilustración 43: Screener: Mockup.....	82
Ilustración 44: Pentaho Community Dashboard Editor (CDE).....	83
Ilustración 45: Screener de acciones.....	84
Ilustración 46: Screener de acciones: Desplegables con autocompletado.....	85
Ilustración 47: Screener de acciones: Cotizaciones y datos fundamentales.....	87

Índice de tablas

Tabla 1: Planificación del proyecto.....	15
Tabla 2: Resumen de los productos obtenidos.....	16
Tabla 3: Conjunto de datos: Submission (SUB).....	39
Tabla 4: Conjunto de datos: Number (NUM).....	40
Tabla 5: Conjunto de datos: Tag (TAG).....	40
Tabla 6: Dimensión tiempo.....	47
Tabla 7: Dimensión empresa.....	48
Tabla 8: Modelo físico dimensiones: Dimensión tiempo (d_time).....	50
Tabla 9: Modelo físico dimensiones: Dimensión empresa (d_company).....	51
Tabla 10: Modelo físico hechos: Técnicos (f_tecnico).....	52
Tabla 11: Modelo físico hechos: Fundamentales (f_fundamental).....	53
Tabla 12: Modelo físico área stage: Tabla stg_data.....	53
Tabla 13: Modelo físico área stage: Tabla stg_sub.....	54
Tabla 14: Modelo físico área stage: Tabla stg_num.....	55
Tabla 15: Modelo físico área stage: Tabla stg_companies.....	55
Tabla 16: Modelo físico área stage: Tabla stg_quotes.....	56
Tabla 17: Modelo físico área stage: Tabla stg_sma.....	56
Tabla 18: Resultado de la ejecución de la ETL.....	75

1. Introducción

1.1. Contexto y justificación del trabajo

Los mercados financieros son espacios donde se realizan intercambios de activos financieros. El objetivo de estos espacios físicos o virtuales es poner en contacto a compradores y vendedores, y establecer un precio justo a los activos financieros que se intercambian.

Los mercados están formados por múltiples participantes, entre ellos, los inversores institucionales, que habitualmente son instituciones financieras (bancos, compañías de seguros, fondos de inversión, etc) que invierten en nombre de otros inversores, e inversores minoristas, que son personas físicas (ahorradores) que invierten a título individual.

Los activos que intercambian los inversores, se clasifican según el tipo de renta que generan, y se distinguen dos tipos: los de renta fija y los de renta variable.

En la renta fija, los rendimientos que generará el activo se conocen de antemano. La deuda emitida por instituciones públicas (obligaciones, bonos, letras) y privadas (bonos, pagarés) son instrumentos financieros de renta fija y se caracterizan porque periódicamente generan unos intereses y a fecha de vencimiento se devuelve el capital invertido en su totalidad.

La renta variable, se asocia normalmente con las acciones de las empresas aunque en esta clasificación encajan otros tipo de activos como los productos derivados (futuros, opciones, warrants, cfd, etc). Los activos de este tipo se caracterizan porque no se garantiza el capital invertido, ni se conoce el beneficio de la inversión de antemano, pues depende de múltiples factores como la evolución de resultados de la empresa, el comportamiento del mercado, la evolución de la economía, etc.

El objetivo principal de los mercados de capitales es financiar a las empresas mediante la canalización del ahorro de las familias y empresas hacia la inversión. Con ello se pretende que los ahorradores obtengan una cierta rentabilidad y las empresas una fuente de financiación para sus proyectos.

Una de las formas que tienen las empresas para obtener financiación, es emitir acciones. Las acciones son las partes en las que se divide el capital social de la entidad y estas partes hacen a la persona poseedora de su titularidad, **accionista**. Las acciones representan el porcentaje de la empresa que le pertenece al accionista y le confiere derechos políticos, como el voto en las juntas de accionistas, y derechos económicos, como la participación en los beneficios de la empresa.

Las bolsas de valores son mercados de capitales organizados y regulados, donde se realizan operaciones de compraventa de acciones, entre otros activos. La fijación del precio de las acciones que se intercambian dependen de múltiples factores como: la situación política, la marcha de la economía en general, los resultados empresariales, el sentimiento del mercado, y cualquier noticia o evento relacionado con la empresa.

Según los datos del Banco Mundial ^[1], en el año 2016, en todo el mundo existían 43.192 empresas cuyas acciones cotizaban en un mercado bursátil.

Un inversor que quisiera comprar acciones, puede encontrarse con una enorme dificultad para tomar una decisión de compra debido al gran número de empresas existentes. Además, cada inversor tiene un perfil de inversión propio, que viene determinado por un conjunto muy amplio de factores, que pueden incluir su situación personal, familiar y laboral, su patrimonio, sus estudios, su experiencia profesional, sus creencias, sus expectativas, por nombrar sólo algunos de ellos.

De esta variedad de circunstancias, surge la necesidad de crear una estrategia de inversión propia que se ajuste a las necesidades particulares del inversor. Una estrategia de inversión, no es más que un conjunto de reglas. Algunas de estas reglas pueden: fijar un horizonte temporal de corto, medio o largo plazo; establecer que se deba realizar una diversificación en distintos sectores de la economía, etc. Entre las múltiples reglas podemos encontrar:

- Comprar acciones de compañías que se consideran que estén en fase de crecimiento y que no paguen dividendo pero que tengan un potencial de incremento de beneficios elevado.
- Invertir en acciones con alto riesgo y volatilidad que permiten conseguir altas rentabilidades. Por ejemplo: empresas del sector tecnológico o biotecnológico, etc.
- Invertir en acciones que tengan una alta rentabilidad por pago dividendos, que permita la generación de una renta periódica.

- Comprar acciones de grandes compañías, denominadas “Blue Chips”, con un negocio maduro, y una baja volatilidad.
- Invertir en empresas que se encuentran actualmente infravaloradas, (inversión en valor).
- Invertir en un fondo de inversión que represente a un sector de la economía. Por ejemplo: farmacéutico, aeronáutico, materiales básicos, etc.

El desarrollo de una estrategia propia aporta un mayor control sobre la evolución del patrimonio del inversor y su rentabilidad, y habitualmente tiene como objetivo la formación de una cartera.

Una cartera o portafolio, está formada por un conjunto inversiones en diferentes productos financieros (moneda, acciones, letras, fondos de inversión, etc). En el caso de una cartera de acciones, estará formada por un conjunto de acciones de diferentes empresas y su construcción, habitualmente, sigue el conjunto de directrices establecidas por el inversor según su estrategia.

Como se intuye de la situación expuesta, la selección de estas acciones para la formación de una cartera, implica múltiples factores y puede resultar enormemente compleja.

Para analizar la idoneidad de una empresa para formar parte de su cartera, el inversor debe analizar la información de la misma. Dicha información puede incluir información general (sector e industria al que pertenece, país de origen, capitalización, ...), información financiera (ingresos, deuda, beneficio, dividendo, ...), precios históricos en el mercado, indicadores técnicos y/o fundamentales, ratios económicos y financieros, etc.

Esta información puede ser obtenida hoy en día de innumerables fuentes: periódicos generalistas, publicaciones especializadas, blogs, foros, portales web financieros como Bloomberg, Google Finance, Yahoo Finance, etc. Ante tal cantidad de información cabe preguntarse ¿cómo seleccionar fuentes de información confiable que ofrezcan la información necesaria para la toma de decisiones?

Parece lógico pensar que la mejor fuente de información sobre una empresa, es la propia empresa. Partiendo de esta premisa y de la obligación legal de las empresas cotizadas a presentar ante los organismos reguladores de los mercados de forma periódica de diversa índole, podemos concluir que la mejor fuente de información

son los registros públicos que contienen toda la información financiera de las empresas cotizadas.

Esta información incluye los estados contables de las empresas, así como todas aquellas noticias que pueden afectar a la cotización de la misma, como desarrollo de nuevos productos o servicios, fusiones, adquisiciones, ajustes de plantillas, en otra mucha información. Por lo general, esta información se encuentra en ficheros en formato PDF, aunque es posible obtener algunos de ellos en formatos que permiten su tratamiento posterior por sistemas informatizados.

Estos sistemas habitualmente están fuera del alcance del inversor particular y en este contexto, este trabajo de fin de grado pretende aplicar las técnicas y desarrollos tecnológicos propios del Business Intelligence (BI), para ayudar a los inversores particulares a acotar el universo de acciones en las que invertir, mediante una herramienta de análisis que le permita aplicar una serie de condiciones o restricciones que deben cumplir las empresas seleccionadas.

1.2. Objetivos del trabajo

El principal objetivo del presente Trabajo de fin de Grado consiste en que el alumno adquiera los conocimientos y las capacidades necesarias para el diseño y la construcción de un sistema de inteligencia de negocio y se familiarice con algunas de las herramientas existentes para la creación de un sistema de estas características.

Para lograr este objetivo, el presente trabajo tratará de crear una prueba de concepto de las herramientas BI orientadas al inversor particular en mercados de valores, de tal forma que esta herramienta le ayude en la toma de decisiones de inversión. La herramienta debe tener las siguientes características:

- Manejo intuitivo y sencillo
- Rapidez en la respuesta
- Adaptable a distintas estrategias de inversión
- Escalable y mejorable.

Para poder dar cumplimiento a este objetivo, será necesaria la realización de las siguientes acciones:

- Analizar las posibles fuentes de información que provean a la herramienta de la información necesaria para su funcionamiento.

- Análisis de los distintos parámetros y restricciones que se pueden aplicar a la información identificada en el paso anterior y que puedan ser de aplicación en este trabajo y en concreto en la herramienta de análisis propuesta.
- Diseño e implementación de un conjunto de procesos ETL, que se encargarán de obtener, tratar y almacenar la información que provendrá de las fuentes de información identificadas anteriormente.
- Diseño e implementación de un almacén de datos que contenga un modelo multidimensional que integre la información y sirva como repositorio de información para las herramienta de análisis.
- Diseño e implementación de una herramienta analítica, que denominaremos screener de acciones, que permita al inversor particular realizar una primera selección de acciones en las que invertir, según los parámetros y restricciones definidos en el paso anterior.

1.3. Enfoque y método elegido

En este trabajo de fin de grado se utilizará como metodología de trabajo la planteada por Ralph Kimball ^[2], uno de los autores más importantes en el ámbito del Business Intelligence.

La metodología propuesta permite implementar almacenes de datos de forma incremental y además, existe una gran cantidad de documentación que ayudará a solucionar cualquier tipo de duda que pueda surgir durante el desarrollo del trabajo. Ver apartado 2.2 El ciclo de vida Kimball

1.4. Planificación del trabajo

La estimación ha sido realizada intentando ajustar dichas tareas con las fechas de las distintas PECs del TFG:

	Tarea	Inicio	Fin
PEC 1		22/02/17	13/06/17
	Elaboración del plan de proyecto	22/02/17	22/02/17
	Entrega de PEC1	13/06/17	13/06/17
PEC 2		14/03/17	17/04/17
	Definición de requisitos	14/03/17	18/03/17
	Modelado Dimensional	19/03/17	11/04/17
	Modelado Físico	12/04/17	14/04/17
	Instalación y configuración de Pentaho, PostgreSQL	15/04/17	17/04/17
	Entrega de PEC2	17/04/17	17/04/17
PEC 3		18/04/17	22/05/17
	Análisis de fuentes de información	18/04/17	22/04/17
	Diseño e implementación de la ETL	23/04/17	29/04/17
	Análisis OLAP: Implementación de cubos	30/04/17	04/05/17
	Implementación cuadro de mandos	05/05/17	22/05/17
	Entrega de PEC3	22/05/17	22/05/17
Cierre		22/05/17	14/06/17
	Elaboración de memoria	22/05/17	09/06/17
	Elaboración de presentación virtual	10/06/17	14/06/17
	Entrega final	14/06/17	14/06/17
	Cierre de proyecto	14/06/17	14/06/17
TFG	Gestión del proyecto	22/02/17	14/06/17

Tabla 1: Planificación del proyecto

1.5. Breve resumen de productos obtenidos

Para la consecución de los objetivos fijados, se realizará un proceso ETL que obtendrá la información de distintas fuentes. También se creará un **data warehouse**

que integre la información más relevante y que sirva posteriormente para la creación de una herramienta de análisis que hemos denominado screening de acciones.

A continuación, se muestra un listado de los entregables que debe producir el TFG propuesto:

Producto	Descripción
Memoria de proyecto	Documento resumen de las tareas y actividades realizadas en el ámbito del TFG.
Presentación virtual	Presentación del proyecto en formato vídeo.
Procesos ETL	Fuente de los procesos ETL generados
Cubos OLAP	Esquema mondrian generado con la definición de los cubos.
Herramienta de análisis	Fuentes de la aplicación de análisis generada con Pentaho CDE.

Tabla 2: Resumen de los productos obtenidos

2. Inteligencia de negocio

2.1. ¿Qué es la inteligencia de negocio?

El término Business Intelligence (inteligencia de negocio), que se atribuye a Howard Dresner, es un conjunto de metodologías, técnicas, herramientas y buenas prácticas que tiene el objetivo de facilitar el acceso y análisis de los datos de una organización, de tal forma que permita enriquecer el proceso de toma de decisiones mediante la respuesta a las cuestiones que se planteen y con ello mejorar la competitividad.

Para conseguir las metas propuestas por el BI, surge la necesidad de extraer, cargar, transformar, analizar y visualizar los datos de tal forma que podamos obtener información de los datos y posteriormente conocimiento. Para ello el BI dispone de una serie de componentes:

- **Fuentes de información:** serán las encargadas de nutrir de datos el almacén de datos (Data Warehouse), habitualmente son sistemas transaccionales (ERP, CRM, SCM, aplicaciones a medida, etc), aunque puede tratarse de información de otro tipo como hojas de cálculo, ficheros estructurados, información de terceros, etc.
- **Procesos ETL (Extract, Transform and Load):** de extracción, transformación y carga de datos obtenidos de las fuentes de información. Previamente a la carga de los datos en el almacén de datos, es necesario transformar, limpiar y filtrar los datos.
- **Almacén de datos (Data Warehouse):** es una colección de datos especializados en un ámbito (empresa, organización, ...) concreto que ayuda a extraer conocimiento y por tanto en la toma de decisiones.
- **Sistemas OLAP (siglas de On Line Analytical Processing):** su función es la de agilizar la consulta de grandes cantidades de datos para poder realizar cálculos y análisis sobre ellos. Existen distintos tipos de sistemas:
 - **ROLAP (Relational OLAP):** usa bases de datos transaccionales. La principal ventaja es que no tiene limitaciones en cuanto al tamaño, pero es más lento que el MOLAP

- **MOLAP (Multidimensional OLAP):** usa bases de datos multidimensionales. La ventaja principal de esta alternativa es que es muy rápida en los tiempos de respuesta
- **HOLAP (Hybrid OLAP):** Accede a los datos de alto nivel en una base de datos multidimensional y a los atómicos directamente sobre la base de datos relacional. En esencia utiliza las ventajas del ROLAP y del MOLAP.
- **Herramientas de consulta, informes, análisis:**
 - **Análisis OLAP:** las herramientas de análisis OLAP, permiten el análisis en línea, administración y ejecución de consultas, para inferir información del comportamiento del negocio.
 - **Herramientas de reporting y consultas ad-hoc:** permiten la generación de informes a medida.
 - **Cuadros de Mandos:** se pueden entender como una colección de reportes, consultas y análisis interactivos que hacen referencia a un tema en particular.
- **Minería de datos (data mining):** se trata de herramientas que tienen el fin de analizar y extraer conocimiento oculto a partir de los datos.

La inteligencia de negocio actúa como un factor estratégico para una organización, generando una potencial ventaja competitiva, pues proporciona información importante para responder a los problemas del negocio.

2.2. El ciclo de vida Kimball

El ciclo de vida Kimball es una metodología para el diseño, implementación y despliegue de un sistema de inteligencia de negocio, conocido por su traducción al inglés como data warehouse/business intelligence (DW/BI).

El ciclo de vida propuesto posee una naturaleza iterativa, donde cada iteración entrega un conjunto de datos y un conjunto de informes y aplicaciones asociadas. La construcción de sistema DW/BI requiere de múltiples iteraciones para ser completado, donde cada iteración incorpora nuevos datos y capacidades al conjunto.

La Ilustración 1 muestra como transcurre el proceso de implementación de un DW. Muestra la secuencia de tareas a realizar y establece aquellas actividades que

deben realizarse en paralelo. Algunas de estas actividades están sujetas a adaptaciones para ajustarse a las necesidades particulares de una organización. No es necesario realizar todas las tareas en todos los proyectos y por ello la realización o no de las mismas debe ser decidida en cada ocasión dependiendo de las necesidades de la entidad.

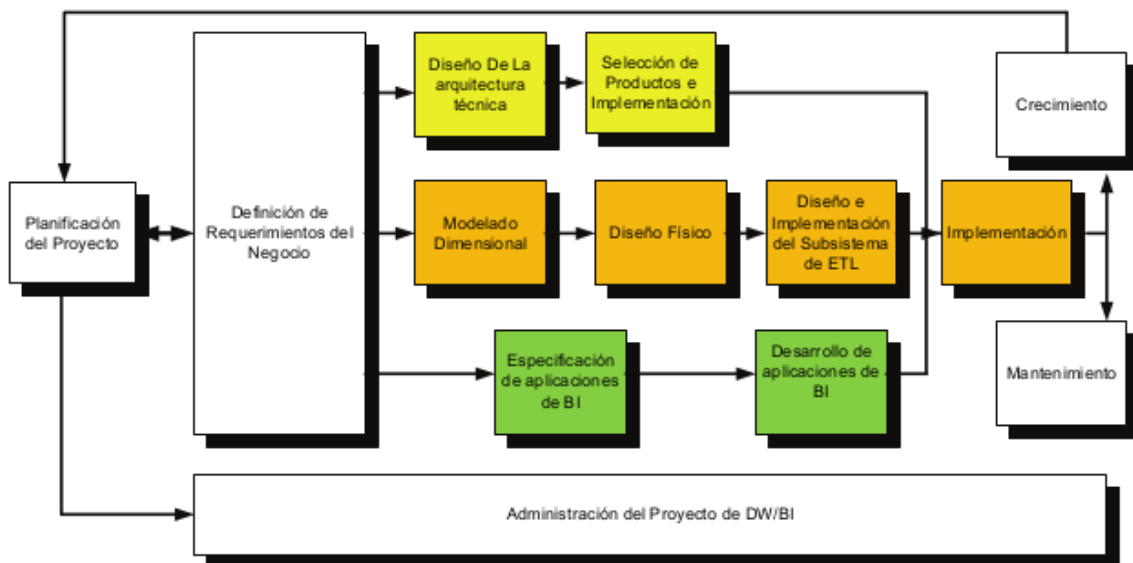


Ilustración 1: Diagrama del ciclo de vida Kimball

El ciclo de vida comienza con la planificación del proyecto, cuyo objetivo principal es determinar el propósito del proyecto de DW/BI, sus objetivos específicos, el alcance del mismo, los principales riesgos y recopilar y priorizar los requerimientos de información. El resultado de este trabajo es una lista priorizada de procesos de negocio que generan datos y que un análisis de los mismos puede proporcionar un alto valor a la organización.

El siguiente paso (definición de los requerimientos del negocio) consiste en seleccionar los procesos de negocio indicados como más prioritarios y obtener toda la información posible sobre ellos, haciendo hincapié en comprender los detalles asociados a los datos del proceso. Este detalle debe incluir atributos, definiciones, reglas de negocio, y por último el conjunto de análisis o aplicaciones que serán construidas utilizando como base el conjunto de datos identificados.

Una vez obtenido el nivel de detalle necesario, el ciclo de vida avanza a la fase de diseño donde el ciclo se divide en tres caminos paralelos:

- En el camino superior de la Ilustración 2, se debe realizar el diseño de la arquitectura tecnológica necesaria para el DW/BI. El objetivo principal es identificar las herramientas necesarias para dar cumplimiento a las necesidades identificadas
- En el camino central, se realiza en un primer momento el modelado dimensional, una obtenido se realiza el modelado físico que dará soporte al anterior y por último se diseña e implementa el proceso de extracción, transformación y carga (ETL).
- En camino inferior, trata de identificar el conjunto de informes y herramientas de análisis que permitirán dar el valor añadido objetivo del proyecto. Este camino tiene dos fases, por un lado de fase de diseño de los informes y herramientas de análisis y por último, una fase de implementación. Esta última fase probablemente se deba acometer cerca de la finalización de la implementación de la ETL, pues es necesario disponer de datos en el sistema para completar esta tarea.

Una vez que las tareas de implementación están completadas, las herramientas, informes y aplicaciones creadas son desplegadas y se vuelve a la tarea inicial del ciclo de vida, terminando una iteración del ciclo de vida.

3. Elección de herramientas

La elección de las herramientas utilizadas para la elaboración de este trabajo, fue realizada previa a la concepción del mismo. Los motivos de la elección fueron por un lado poseer este alumno cierta experiencia profesional con PostgreSQL, y por otro lado, aprovechar una formación introductoria a Pentaho Community Edition recibida previamente. De esta forma, este trabajo servirá para profundizar en el conocimiento de ambas herramientas.

También se tuvo en cuenta que las dos herramientas propuestas son libres y permiten su uso sin restricciones. Además, ambas poseen amplias comunidades de usuarios que le dan soporte y una extensa documentación que ayudará al desarrollo del trabajo.

A continuación se realizará una breve descripción de las herramientas.

3.1. PostgreSQL

PostgreSQL^{[3][4]} es un sistema de gestión de bases de datos relacional (SGDB), distribuido bajo licencia PostgreSQL (similar a la licencia BSD). Esta licencia permite acceder libremente a su código fuente.

Este sistema de gestión de base de datos, utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará al resto y el sistema continuará funcionando.

PostgreSQL posee todas las características habituales de un SGBD, pero además incluye características avanzadas como:

- Permite accesos simultáneos a la misma tabla sin necesidad de bloqueos. Cada usuario tiene una visión consistente de los últimos datos consolidados, después de realizar una operación de commit.
- Herencia de tablas
- Integridad referencial, mediante claves extranjeras.

- Sistema de transacciones anidadas (savepoints)
- Implementación de funciones en distintos lenguajes de programación: C/C++, Java, etc.
- Permite la creación de tipos de datos definidos por el usuario.

3.2. Pentaho Community Edition

Pentaho Community Edition^{[5][6]} es una suite de herramientas para business intelligence que incluye ETLs, análisis, metadatos y generación de informes (reporting). Esta solución es totalmente libre bajo la licencia GNU GPLv2^[7].

A continuación se describen brevemente las herramientas que forman parte de esta suite para Business Intelligence:

- **Pentaho Data Integration:** También denominada como Kettle, permite la implementación de los procesos de extracción, transformación y carga (ETL), que permite preparar los datos extraídos de las fuentes de información para los procesos posteriores de análisis, reporting o minería de datos.
- **Pentaho Analysis Services:** Provee un servidor Mondrian, que es un servidor OLAP (procesamiento analítico en línea) escrito en Java. Esta solución, es compatible con el lenguaje multidimensional MDX.
- **Metadata Editor:** Permite añadir una capa adicional de información a las fuentes de datos. Habitualmente se usa para facilitar la utilización de los datos por parte de los usuarios de las herramientas de reporting.
- **Schema Workbench:** Herramienta gráfica para crear los esquemas OLAP para análisis.
- **Pentaho Reporting:** Conjunto de herramientas para la creación de informes (reporting) .
- **Pentaho Data Mining:** Suite de software que permite aplicar algoritmos de “machine learning”, aprendizaje automático y minería de datos. Esta funcionalidad se basa en Apache Weka.
- **Pentaho Dashboard:** Es una plataforma que permite la creación de cuadros de mandos. Pentaho Dashboard permite ver informes, gráficos interactivos y los cubos creados con las herramientas Pentaho Report Designer.

4. Los inversores y la búsqueda de la rentabilidad en los mercados

La inversión en los mercados financieros es una actividad que ha sido profundamente estudiada durante años desde todos los ángulos posibles y por personas con todo tipo de formación y experiencia profesional. La recompensa para aquellos que logran descifrar las complejidades de la materia son enormes, y sin embargo, son muy pocos los logran tener una rentabilidad aceptable en el largo plazo.

El precio de los activos financieros que cotizan en las bolsas de todo el mundo, varían minuto a minuto. A grandes rasgos y simplificando la cuestión, el motivo de estas fluctuaciones se debe al intercambio que se produce entre compradores y vendedores fijando un precio para los activos en cada momento. Este precio está influenciado, como veremos más adelante, por múltiples factores.

Los ahorradores que deciden invertir en los mercados financieros buscan obtener la mayor rentabilidad posible para sus ahorros. Una de las formas más evidentes de alcanzar esta rentabilidad, consiste en comprar barato y vender caro. Para lograr, este a priori sencillo objetivo, es necesario conocer el valor real de los activos en los que se invierte.

Muchas son las personas que buscan la fórmula o el método más seguro para predecir el movimiento de los mercados, el mejor activo para comprar o vender y el momento adecuado para realizar la compra o la venta, quizás animados por el éxito de algunos inversores y especuladores que han adquirido las habilidades necesarias para construir su fortuna y que por ello han alcanzado cierta notoriedad, incluso fama.

Con los años, todos estos esfuerzos para descifrar el funcionamiento de los mercados financieros, ha generado dos tendencias bien diferenciadas para determinar si el precio de un activo está infravalorado o sobrevalorado: el análisis técnico y el análisis fundamental. Estas dos tendencias tienen visiones radicalmente opuestas sobre cómo responder a las preguntas de los inversores.

4.1. El análisis técnico

Los analistas técnicos consideran que sólo requieren de un gráfico donde se pueda observar la evolución del precio del activo con respecto al tiempo, para formar una opinión de inversión. Estos analistas no necesitan ni siquiera conocer el nombre de la empresa o a qué se dedica, el gráfico ya proporciona toda la información necesaria para tomar la decisión correcta.

El análisis técnico ^[8], podemos decir que es el análisis de la variación del precio de un activo a través del uso de gráficos, con el objetivo de predecir el futuro movimiento de dicho activo. Este tipo de análisis se basa en las tres siguientes premisas:

- a) **El mercado lo descuenta todo:** Quizás esta es la premisa más importante del análisis técnico y considera que todo aquello que pueda afectar al precio de un activo, como por ejemplo: cualquier aspecto relacionado con la empresa (resultados trimestrales y/o anuales, desarrollo de nuevos productos o servicios, demandas judiciales, la competencia, entre otros muchos factores) la economía en general, la política, o de cualquier otra índole, el mercado lo descuenta, es decir, se refleja de forma inmediata en el precio del activo, y por lo tanto, sólo es necesario un gráfico para obtener la información necesaria para tomar una decisión sobre un activo cualquiera.
- b) **El precio se mueve en tendencias:** El principal objetivo del análisis técnico es la identificación de tendencias en los estados iniciales de su formación con el objetivo de actuar en la dirección de la tendencia identificada. Para que sea posible, es necesario que se cumpla que el precio se mueve en tendencias tal y como indica la premisa.
- c) **La historia se repite:** Un aspecto fundamental dentro del análisis técnico trata sobre la psicología del ser humano. El análisis de los gráficos realizado durante años, ha identificado una serie de formas, figuras o patrones que se repiten y que revelan el sentimiento alcista o bajista de los participantes en el mercado. Como estos patrones han funcionado bien en el pasado, se asume que seguirán haciéndolo bien en el futuro, debido a la tendencia en la psicología humana que tiende a evitar el cambio. Este hecho es conocido como la profecía del auto-cumplimiento del análisis técnico.

Dentro del análisis técnico, a su vez, han surgido dos ramas el chartismo y el análisis cuantitativo, que comparten muchas similitudes, pero también diferencias sustanciales.

Por un lado tenemos a aquellos que realizan un análisis basado en la interpretación solamente de los gráficos, también conocidos como chartistas. Estos analistas utilizan herramientas como soportes, resistencias, tendencias, canales, formación de figuras o patrones técnicos, retrocesos de fibonacci, etc. Este tipo de técnicas tienen un alto grado de subjetividad, pues dos chartistas pueden llegar a conclusiones diferentes interpretando un mismo gráfico. En consecuencia, puede resultar bastante complejo automatizar un sistema basado en el chartismo, debido a esta componente subjetiva.

Por otro lado, tenemos a aquellos que realizan un análisis cuantitativo. Según la Wikipedia^[9], “el **análisis cuantitativo**, el cual se basa en estudiar una serie numérica de datos mediante indicadores matemáticos y estadísticos, esto es, consiste en aplicar fórmulas a los precios y a los volúmenes de los valores con la finalidad de facilitar la toma de decisiones de inversión y de conocer los precios en situaciones determinadas. La ventaja de este análisis es la eliminación de la subjetividad del análisis gráfico o chartista”.

De estas dos ramas dentro del análisis técnico, nos centraremos en el análisis cuantitativo por considerar que se ajusta a un proyecto como el planteado en este TFG.

4.1.1. Las herramientas del análisis técnico cuantitativo

Como hemos visto, el análisis técnico cuantitativo trata de eliminar la subjetividad en la interpretación de los gráficos mediante el empleo de métodos matemáticos y estadísticos. Los analistas cuantitativos disponen una amplia gama de herramientas a su disposición, que a menudo se denominan indicadores técnicos y que permiten obtener información para el estudio de un activo.

Los indicadores, en función del tipo de información que aportan, pueden clasificarse en indicadores de volatilidad, de tendencia, de momento y de volumen.

4.1.2. Indicadores de volatilidad

La volatilidad^[10] “es una medida de la frecuencia e intensidad de los cambios del precio de un activo o de un tipo definido como la desviación estándar de dicho cambio en un horizonte temporal específico”.

Habitualmente se usa para identificar o cuantificar el riesgo asociado a un activo. Por tanto, la volatilidad es un parámetro negativo, pues indica que existe incertidumbre o riesgo asociado a dicho activo.

Indicador ATR (Rango medio verdadero)

Se basa en el cálculo del True Range para determinar qué tan volátil es un mercado. El concepto de True Range podemos decir que se calcula como el valor más alto entre las tres opciones siguientes: valor absoluto del máximo menos el último cierre, valor absoluto del mínimo menos el último cierre o diferencia entre el máximo y el mínimo.

Bandas de bollinger

Se trata de un indicador de análisis técnico ideado por John Bollinger. Está compuesto de tres líneas cuyo movimiento nos indican la volatilidad de un activo en un determinado tiempo.

Por lo general, se calcula una media móvil de 20 períodos. A continuación hay que obtener dos desviaciones, una usando los niveles mínimos y otra usando los niveles máximos de la sesión. En algunos casos el precio se aleja mucho de las líneas del indicador. Cuando esto sucede a la vez que se registra una volatilidad baja, se puede leer una señal de inversión de tendencia inminente. Otra utilidad de las bandas de Bollinger es establecer niveles de soporte y resistencia.

4.1.3. Indicadores de tendencia

Estos indicadores también son conocidos como indicadores seguidores de tendencia, pues informan si el precio de un activo sigue una tendencia. También pueden informar sobre la fortaleza o debilidad de dicha tendencia.

Se suelen utilizar para detectar cuándo se inicia una tendencia y cuándo se está debilitando.

Media Móvil Simple (SMA)

La media móvil simple, se define como la media aritmética de los últimos N precios de cierre de un determinado activo. Las medias móviles son una versión suavizada y retrasada del precio. A mayor periodo, mayor suavizado y reacción más lenta ante el movimiento del precio.

Su uso habitual consiste en la comparación de dos medias móviles de distinto período, mediante la sustracción del valor de una media lenta (mayor periodo) al de una media rápida (menor período). Valores positivos de esta diferencia suelen indicar movimientos alcistas y negativos movimientos bajistas.

Media Móvil Exponencial (EMA)

La media móvil exponencial, es una versión de la simple, que ofrece unos valores más suavizados del precio a lo largo del tiempo. Se calcula como la media exponencial de los últimos N períodos de un activo, dando más importancia a los últimos períodos.

Su uso es similar al explicado para las medias móviles simples.

Indicador MACD

El indicador de convergencia/divergencia de medias móviles (moving average convergence divergence, MACD) es uno de los indicadores más utilizados en el análisis técnico. El indicador MACD es un indicador que muestra la distancia entre una media móvil exponencial (EMA) rápida y una media móvil exponencial lenta. O lo que es lo mismo, muestra la convergencia/divergencia de dos medias móviles exponenciales.

Es usado tanto para señalar la tendencia. como para mostrar la fuerza del movimiento.

4.1.4. Indicadores de momento

Los indicadores de momento muestran la fuerza o debilidad con la que se está moviendo el precio de un activo. Se utilizan como indicadores adelantados de cambio de tendencia.

Momentum

El indicador momentum muestran la diferencia de precios entre el cierre de hoy y el de hace N días. Este indicador permanece positivo mientras se mantiene una tendencia alcista, y negativo cuando es bajista. Un cambio al alza se interpretar como una señal para comprar, y una caída bajo cero como una señal de venta.

Indicador RSI

El índice de fuerza relativa^[11] es un indicador tipo oscilador utilizado en análisis técnico que muestra la fuerza del precio mediante la comparación de los movimientos individuales al alza o a la baja de los sucesivos precios de cierre. Se calcula mediante las siguientes fórmulas:

$$RSI = 100 - \frac{100}{1 + RS}$$

$$RS = \frac{\text{media de los } n \text{ días con cierre positivo}}{\text{media de los } n \text{ días con cierre negativo}}$$

Ilustración 2: Fórmula para el cálculo del índice de fuerza relativa (RSI)

El valor de RSI oscila entre 0 y 100. Si toma un valor por encima de 70, indica una sobrevaloración del mercado y es posible que tenga lugar un techo en los precios. Si desciende por debajo de 30 es muy probable que se desarrolle un suelo en los precios, pues el mercado está siendo infravalorado.

La evolución de la tendencia del RSI es parecida a la de los precios y, por tanto, cuando se produce una divergencia entre ambas se está produciendo un aviso de cambio de tendencia.

4.1.5. Indicadores de volumen

Los indicadores de volumen son utilizados para mostrar el volumen de operaciones de compra-venta en un activo en particular. Este tipo de indicadores son, por lo general, muy útiles para confirmar cambios de tendencia.

Balance de volúmenes

El indicador OBV (On Balance Volume) se utiliza para determinar el flujo del volumen negociado de una acción. La hipótesis en la que se basa el indicador es que los movimientos en el volumen negociado pueden preceder a los movimientos del precio.

OBV es un indicador cuyo valor a lo largo del tiempo depende de las variaciones del precio y del volumen negociado. Un aumento en el volumen negociado implica el interés de muchos inversores en la compra del activo y probablemente, su encarecimiento debido al aumento de la demanda. Por el contrario, un descenso en el volumen negociado (escaso interés inversionista en el activo) el precio tenderá a caer debido a la escasa demanda.

4.1.6. Elección de indicadores técnicos

En los apartados anteriores, hemos visto sólo algunos de entre muchos, indicadores técnicos habituales en el análisis cuantitativo. Como se observa existe una gran cantidad de indicadores lo que nos obliga a realizar una selección para el desarrollo de este TFG y definir el alcance del proyecto.

Por este motivo, se ha escogido para su implementación las medias móviles simples. En nuestro caso, utilizaremos las SMA de 20, 50, 200 y 1000 sesiones que se podrán utilizar en la herramienta de análisis propuesta para encontrar empresas cuyas cotizaciones actuales se encuentren cerca de estas medias móviles. El resto de indicadores pueden ser objeto de un nuevo TFG que amplíe las capacidades de esta herramienta.

Por otro lado, también se utilizarán los precios máximos y mínimos de cierre de las últimas 52 semanas, como medida de soporte o resistencia. Se han seleccionado estos dos parámetros, adicionalmente a todos los indicadores mostrados, pues son

muy usados para obtener una referencia de lo caro o barato que se encuentra un activo en un momento dado.

4.2. El análisis fundamental

Los analistas fundamentales toman sus decisiones en base al estudio de toda la información relativa a los sectores de la economía, a la economía en general, y de las empresas. Para ello se pueden realizar distintos tipos de análisis:

- Análisis macro-económico
- Análisis sectorial
- Análisis cualitativo de la empresa
- Análisis cuantitativo de la empresa
 - Estados financieros de la empresas
 - Cuenta de resultados
 - Generación de fondos
- Estudio de los ratios financieros
- Valoración

Estos análisis generan información que es procesada y a partir de ella se toma una decisión de inversión, que puede resultar en la compra de un activo, o en su venta.

De todos los tipos de análisis que pueden realizarse dentro del análisis fundamental, nos centraremos en el análisis cuantitativo y en concreto en los ratios financieros por considerarse que serán de utilidad en un proyecto como el planteado en este TFG.

4.2.1. Los ratios financieros en el análisis fundamental

Los ratios financieros^[12]: también denominados indicadores financieros, son medidas que permiten analizar el estado actual o pasado de una organización, en función a niveles óptimos definidos para ellas.

Los ratios financieros y bursátiles se utilizan como una herramienta de análisis y comparación entre distintas compañías, sectores o mercados y se suelen utilizar para conocer la situación de la empresa en el pasado de manera que podamos concluir sobre su fortaleza o debilidad de cara al futuro.

En los siguientes apartados se realizará una breve introducción de algunos ratios financieros.

4.2.2. Ratios de rentabilidad

Estos ratios financieros comparan las ganancias de un periodo con determinadas partidas de la cuenta de resultados y el balance. Permiten tener una visión de la eficiencia en la gestión de la empresa, es decir, la forma en que los directivos han utilizado los recursos, pues ofrecen respuestas más completas acerca de cómo está siendo manejada la empresa.

Margen bruto

Gross profit margin o margen bruto de beneficio, es usado para analizar si la compañía está realizando un uso adecuado de los materiales asociados a la producción. Un valor alto de este ratio se considera óptimo.

$$\text{Gross profit margin} = \frac{\text{Gross profit}}{\text{Net sales (Revenues)}}$$

Margen operativo

Operating profit margin o margen operativo de beneficio, mide que cantidad de ingresos quedan después de restar los costes de operación de la empresa. Indica cuánto de rentables son las operaciones de la compañía.

$$\text{Operating profit margin} = \frac{\text{Operating profit}}{\text{Net Sales (Revenues)}}$$

Margen neto

Net profit margin o margen neto de beneficio, indica la rentabilidad total obtenida por unidad vendida. Aglutina todos los conceptos por los que la empresa obtiene ingresos o genera gastos. Se considera que un valor elevado de este ratio es muy positivo.

$$\text{Net profit margin} = \frac{\text{Net Income}}{\text{Net Sales (Revenues)}}$$

ROA

La rentabilidad de los activos totales mide la capacidad de los activos de generar beneficios para la empresa.

$$ROA = \frac{\text{Net Income}}{\text{Total assets}}$$

ROI

El retorno de la inversión, es similar al concepto del ROA aunque en este caso solo se considera la rentabilidad de una inversión concreta y los activos necesarios para conseguirla.

ROE

La rentabilidad sobre recursos propios, es la medida de rentabilidad por excelencia para el accionista ya que al final es la rentabilidad sobre sus propios recursos. Los recursos propios representan la inversión o pertenencia del accionista en la empresa y la capacidad de generar mayor o menor rentabilidad sobre éstos es lo que hará más atractiva la inversión.

$$ROE = \frac{\text{Net Income}}{\text{Shareholders Equity}}$$

4.2.3. Ratios de liquidez

Se considera que la liquidez de una empresa es la capacidad de ésta para cumplir con sus obligaciones a corto plazo. Los ratios de liquidez intentan medir esta capacidad comparando los activos más líquidos con las obligaciones a corto plazo.

En general, un alto nivel de cobertura de los activos líquidos con respecto a las obligaciones a corto plazo es considerada como buena. Una cobertura baja o negativa puede ser un signo de problemas para la empresa.

Current ratio

Este ratio mide la capacidad de la empresa de hacer frente a sus obligaciones de más corto plazo con respecto a los activos corrientes.

$$\text{Current ratio} = \frac{\text{Current assets}}{\text{Current liabilities}}$$

Quick ratio

Versión más conservadora del Current ratio, pues sólo utiliza para su cálculos los activos más líquidos y que por tanto, se pueden convertir en efectivo más rápido.

$$\text{Quick ratio} = \frac{\text{Cash Equivalents} + \text{Marketable Securities} + \text{Accounts Receivables}}{\text{Current liabilities}}$$

4.2.4. Ratios de solvencia

Los ratios de solvencia o ratios de apalancamiento, miden la capacidad de una empresa de mantener sus operaciones de forma indefinida. Permite comparar los niveles de deuda con el patrimonio, los activos y las ganancias. Estos ratios ayudan a identificar posibles problemas para cumplir con las obligaciones a largo plazo de una compañía en concreto.

Debt

El ratio de deuda se calcula dividiendo todas las obligaciones de la empresa entre sus activos y se utiliza como medida del apalancamiento de la sociedad.

$$\text{Debt ratio} = \frac{\text{Total liabilities}}{\text{Total assets}}$$

Debt to equity

Este ratio indica la proporción entre el patrimonio de la entidad y la deuda utilizada para financiar los activos de la empresa.

$$\text{Debt equity ratio} = \frac{\text{Total liabilities}}{\text{Shareholders Equity}}$$

4.2.5. Ratios bursátiles

Los ratios bursátiles son usados por los inversores para realizar una estimación de la posible rentabilidad de una inversión. Permiten comparar el precio de las empresas cotizadas con respecto a las ganancias o los dividendos.

Earnings per share

Representa la parte proporcional de los beneficios que le corresponde a una acción.

$$\text{Earnings per share} = \frac{\text{Net Income}}{\text{Average Outstanding shares}}$$

Payout

El payout es el porcentaje del beneficio que una empresa reparte entre sus accionistas mediante el pago de dividendos.

$$\text{Payout} = \frac{\text{Annual dividend}}{\text{Earning per share}}$$

Price to earnings ratio

El ratio precio-beneficio o PER indica cuántas veces se está pagando el beneficio neto anual de una empresa determinada al comprar una acción de ésta. Un PER más elevado implica que los inversores están pagando más por cada unidad de beneficio.

$$\text{Price/Earnings ratio} = \frac{\text{Stock price per share}}{\text{Earnings per share}}$$

Dividend yield

La rentabilidad por dividendo muestra, en porcentaje, la relación existente entre los dividendos por acción repartidos por una sociedad en el último año y el precio de ese título.

$$\text{Dividend yield} = \frac{\text{Annual dividend per share}}{\text{Stock price per share}}$$

Price to book value

Este ratio muestra cuántas veces está pagando el inversor por el patrimonio de la empresa

$$\text{Price/Book value ratio} = \frac{\text{Stock price per share}}{\text{Shareholders equity per share}}$$

4.2.6. Elección de ratios financieros

Hemos visto sólo algunos de los ratios financieros que disponen los analistas fundamentales para realizar sus análisis. Para su implementación en el ámbito de este TFG, se han escogido los ratios ROA, ROE, Debt y Debt to equity ratio.

El resto de ratios pueden ser objeto de un nuevo TFG que amplíe las capacidades de esta herramienta.

5. Análisis de las fuentes de información

En este apartado vamos a ver las fuentes de información sobre las que se desarrollará este trabajo. Por un lado, veremos como obtener la información financiera periódica que las empresas deben presentar a los organismos reguladores y por otro lado, veremos como obtener el histórico de cotizaciones de las empresas a través de la consulta de un servicio web gratuito.

5.1. Cotizaciones históricas (Yahoo Finance)

Como hemos visto en el apartado 4.1 El análisis técnico, los indicadores técnicos utilizan series de precios de las empresas cotizadas para realizar los cálculos. Por este motivo se ha utilizado un servicio gratuito de Yahoo para obtener las cotizaciones históricas de las empresas a analizar.

Yahoo! Finance es servicio de Yahoo, que ofrece información financiera, incluyendo cotizaciones de bolsa, índices bursátiles, comunicados de prensa corporativos y financieros, entre otros servicios.

Yahoo Finance API es un servicio limitado por el número de peticiones que permite, que proporciona a los desarrolladores una forma de obtener las cotizaciones históricas de las empresas a través de una api de consulta mediante la invocación de llamada a una url a la que se le pasan distintos parámetros, indicando las empresa que se quieren consultar y un rango de fechas:

Parámetro	Descripción
s	Símbolo de la empresa a consultar
a	Día de inicio
b	Mes de inicio
c	Año de inicio
d	Día de fin
e	Mes de fin
f	Año de fin

Un ejemplo de llamada al servicio sería:

```
http://ichart.finance.yahoo.com/table.csv?
s=AAPL&a=01&b=05&c=2007&d=12&e=05&f=2017&g=d&ignore=.csv
```

El resultado esta invocación, serían las cotizaciones de la empresa Apple (AAPL) desde el día 1 de mayo de 2017 hasta el 12 de mayo de 2017, los dos últimos parámetros indican que queremos descargar un fichero formato csv con el resultado:

```
Date,Open,High,Low,Close,Volume,AdjClose
2017-05-12,154.699997,156.419998,154.669998,156.100006,32221800,156.100006
2017-05-11,152.449997,154.070007,152.309998,153.949997,27255100,153.949997
2017-05-10,153.630005,153.940002,152.110001,153.259995,25805700,152.629999
2017-05-09,153.869995,154.880005,153.449997,153.990005,39130400,153.357009
2017-05-08,149.029999,153.699997,149.029999,153.009995,48752400,152.381026
2017-05-05,146.759995,148.979996,146.759995,148.960007,27327700,148.347687
2017-05-04,146.520004,147.139999,145.809998,146.529999,23371900,145.927668
2017-05-03,145.589996,147.490005,144.270004,147.059998,45697000,146.455488
2017-05-02,147.539993,148.089996,146.839996,147.509995,45352200,146.903635
2017-05-01,145.100006,147.199997,144.960007,146.580002,33602900,145.977465
```

El fichero generado contiene los para cada uno de los días consultados, los precios de apertura, máximos, mínimos, cierre, el volumen y el cierre ajustado.

5.2. Informes trimestrales/anuales de las empresas

En el apartado 4.2 El análisis fundamental, se han introducido algunos ratios financieros que se utilizan en el ámbito del análisis fundamental. El cálculo de estos ratios requiere conocer los estados contables de las empresas objeto de análisis.

La U.S. Securities And Exchange Commission de los Estados Unidos, conocida comúnmente por sus siglas SEC, es una agencia federal del Gobierno de los Estados Unidos dedicada a la protección de los inversores y a regular y supervisar los mercados financieros y sus participantes. Todos los países con mercados financieros regulados poseen un órgano similar. Por ejemplo, en España este organismo es la Comisión Nacional del Mercado de Valores (CNMV).

Una de las principales funciones de estos organismos es la supervisión de los participantes en las bolsas y mercados financieros y especialmente de las entidades con valores negociables (acciones). A través de leyes y regulaciones estas instituciones obligan a las empresas a publicar información importante sobre su

actividad, con el objetivo de proteger a los inversores y de esta forma se consigue promover la transparencia y evitar el fraude.

En el ámbito de este trabajo vamos a centrar el foco de atención en la información sobre la situación financiera de las empresas que deben publicar de forma periódica. En el caso de la SEC, las empresas deben cumplimentar unos formularios con una estructura estándar conocidos como los formularios 10-K y 10-Q. El formulario 10-K contiene un informe anual en el que se detallan los estados contables de la empresa: la cuenta de resultados, el balance de situación y el estado de flujos de efectivo (en inglés, income statement, balance sheet y el cash flow statement), además de otra información como los comentarios de los gestores sobre la marcha de la empresa, sobre la competencia, posibles riesgos y cualquier otra información que se considere relevante. El formulario 10-Q es una versión similar al anterior pero con información del último trimestre.

Como hemos visto en la introducción, existen miles de empresas que cotizan en mercados regulados y por lo tanto, están obligadas a presentar esta información puntualmente. Una enorme cantidad de información de vital importancia a la hora de analizar las empresas. Con el propósito de facilitar la automatización del intercambio de información financiera, en 1998 Charles Hoffman, experto contable y auditor, propone la creación de un lenguaje basado en XML, llamado **XBRL** (*Extensible Business Reporting Language*) que pretende estandarizar la información financiera.

En la actualidad muchas de las empresas que tienen la obligación de presentar la información periódica, incluye también la versión en XBRL de los informes. La SEC con el objeto de facilitar aún más el tratamiento de esta información ha creado un conjunto de datos, en formato texto, extraídos de los documentos XBRL reportados y que contienen la información más relevante de los informes. Estos conjunto de datos son publicados de forma trimestral y descargados libremente para su uso.



- ABOUT THE SEC
- Commissioners
- Contact
- Data
- Division and Offices
- Regional Offices
- Division and Office Directors
- Forms
- Privacy and Security
- Reports and Publications
- SEC Docket
- Securities Laws
- Upcoming Events
- What We Do

Data



Financial Statement Data Sets



The data sets below provide selected information extracted from exhibits to corporate financial reports filed with the Commission using eXtensible Business Reporting Language (XBRL). The information is presented without change from the "as filed" annual and quarterly financial reports submitted by each registrant. The data is presented in a flattened format to help users analyze and compare it. The data sets also contain additional fields including a company's Standard Industrial Classification to facilitate the data's use.

Data sets will be updated quarterly. Data contained in documents filed after 5:30PM Eastern on the last business day of a quarter will be included in the subsequent quarterly posting.

Annual and Quarterly Financial Statements (PDF, 175 kb) provides documentation of scope, organization, file formats and table definitions.

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
2017	2017q1.zip 34 mb			
2016	2016q1.zip 37 mb	2016q2.zip 31 mb	2016q3.zip 31 mb	2016q4.zip 31 mb
2015	2015q1.zip 41 mb	2015q2.zip 40 mb	2015q3.zip 33 mb	2015q4.zip 34 mb
2014	2014q1.zip 40 mb	2014q2.zip 37 mb	2014q3.zip 38 mb	2014q4.zip 33 mb
2013	2013q1.zip	2013q2.zip	2013q3.zip	2013q4.zip

Ilustración 3: Información financiera en la Securities Exchange Commission (SEC)

Este conjunto de datos está formado por cuatro subconjuntos de datos:

- **SUB (Submission data set):** Este conjunto de datos presenta un registro por cada XBRL registrado en el periodo. La información contenida es extraída de la base de datos de la SEC. A continuación se presentan los campos más importantes del conjunto de datos:

NOMBRE	DESCRIPCIÓN	FORMATO	TAMAÑO	OBLIGATORIO
adsh (*)	Número de control asignado por la SEC al formulario enviado	Alfanumérico	20	Sí
cik	Código asignado por la SEC a la empresa	Numérico	10	Sí
name	Denominación de la empresa	Alfanumérico	150	Sí
sic	Código de clasificación industrial	Numérico	4	No
countryba	Código ISO 3166-1 del país de la sede social de la empresa	Alfanumérico	2	Sí
stprba	El estado o provincia de la sede social de la empresa, si el país es US o CA	Alfanumérico	2	No
cityba	Ciudad de la sede social de la empresa	Alfanumérico	30	Sí
zipba	Código postal de la sede social de la empresa	Alfanumérico	10	No
bas1, bas2	Dirección de la sede social de la empresa	Alfanumérico	40	No
baph	Teléfono de la empresa	Alfanumérico	12	No
fye	Fecha fin del año fiscal	Alfanumérico (mmdd)	4	Sí
form	Tipo de formulario	Alfanumérico	10	Sí
período	Fecha del balance de situación	Fecha (yymmddd)	8	Sí
fy	Año fiscal	Fecha (yyyy)	4	Sí
fp	Período fiscal	Alfanumérico (FY, Q1, Q2, Q3, Q4)	2	Sí

Tabla 3: Conjunto de datos: Submission (SUB)

(*): Estos campos forman clave

- **NUM (Number data set):** es un conjunto de datos numérico, que representa cada una de las cantidades presentadas en los documentos XBRL. A continuación se presentan los campos más importantes del conjunto de datos:

NOMBRE	DESCRIPCIÓN	FORMATO	TAMAÑO	OBLIGATORIO
adsh (*)	Número de control asignado por la SEC	Alfanumérico	20	Sí
tag (*)	Código asociado a la descripción del valor en el conjunto de datos TAG	Alfanumérico	256	Sí
version (*)	Versión asociada a la descripción del valor en el conjunto de datos TAG	Alfanumérico	20	Sí
ddate (*)	Fecha final para el dato	Fecha (yyyymmdd)	8	Sí
qtrs (*)	El trimestre asociado al valor	Numérico	8	Sí
uom (*)	Unidad de medida	Alfanumérico	20	Sí
value	El valor numérico.	Numérico (28,4)	16	No
footnote	Texto o notas asociadas al valor en el registro original	Alfanumérico	512	No

Tabla 4: Conjunto de datos: Number (NUM)

(*): Estos campos forman clave

- **TAG (Tag data set):** Este conjunto de datos define cada una de las entradas de los conjuntos de datos y contiene los nombres y las descripciones usadas en los conjuntos de datos. A continuación se presentan los campos más importantes del conjunto de datos:

NOMBRE	DESCRIPCIÓN	FORMATO	TAMAÑO	OBLIGATORIO
tag (*)	Código asociado a la descripción de un valor	Alfanumérico	256	Sí
version (*)	Versión asociada a la descripción de un valor	Alfanumérico	20	Sí
tlabel	Nombre del atributo	Alfanumérico	512	o
doc	Descripción detallada del atributo	Alfanumérico	2048	No

Tabla 5: Conjunto de datos: Tag (TAG)

(*): Estos campos forman clave

- **PRE (Presentation data set):** proporciona información sobre cómo se presenta la información en los estados financieros. No se entra en más detalle sobre este conjunto de datos, pues no aplica a este trabajo.

5.3. Otras fuentes de información

5.3.1. Empresas de los mercados Nasdaq, Nyse y Amex

El estudio de los datos obtenidos de los registros de la SEC, muestra la existencia de información de un gran número de empresas. Por ello se ha decidido seleccionar la información financiera de aquellas empresas que cotizan en los grandes mercados americanos, en este caso Nasdaq, Nyse y Amex.

El portal web del NASDAQ proporciona unos ficheros excel con las empresas que cotizan en estos mercados. Además se incluye el símbolo utilizado para identificar las acciones de cada empresa, la industria y el sector al que pertenece, información que incorporaremos al DW/BI.

<http://www.nasdaq.com/screening/company-list.aspx>

Ejemplo del formato de los datos obtenidos:

```
"Symbol", "Name", "LastSale", "MarketCap", "IPOyear", "Sector", "industry",  
"XXII", "22nd_Century_Group_Inc", "1.89", "$171.79M", "n/a", "Consumer-Non-  
Durables", "Farming/Seeds/Milling",  
"BRN", "Barnwell Industries, Inc.", "2.1", "$17.38M", "n/a", "Energy", "Oil & Gas  
Production",  
"BGSF", "BG_Staffing_Inc", "16.875", "$147.49M", "n/a", "Technology", "Profession  
al Services",
```

5.3.2. Lista ISO de países y estados/provincias

La ISO 3166 es el estándar internacional de normalización, publicado por la Organización Internacional de Normalización (ISO), que proporciona códigos para los nombres de países y otras dependencias administrativas.

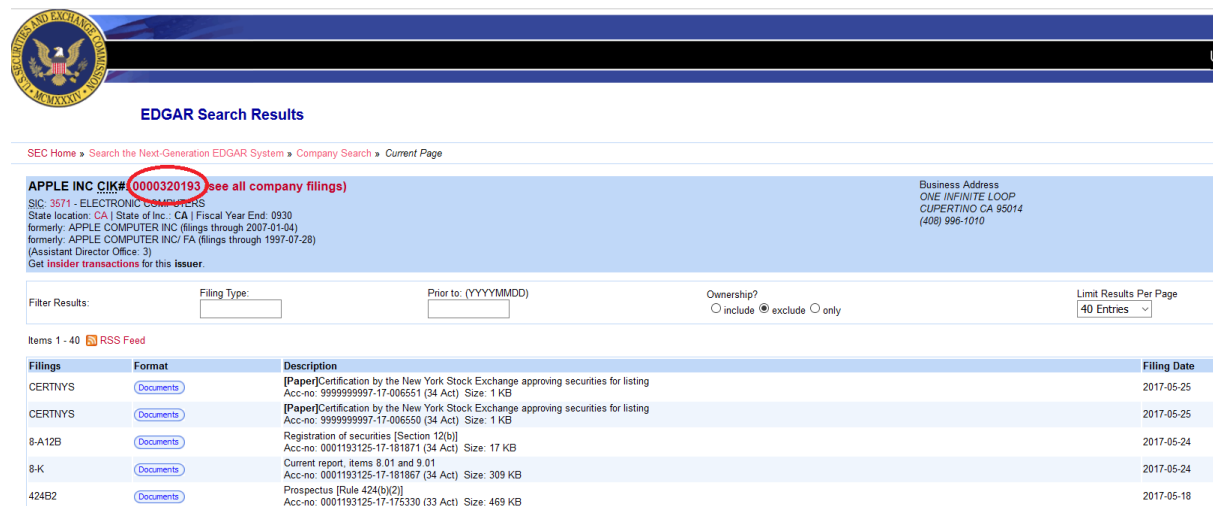
La ISO 3166-1 proporciona los códigos de los países y la ISO 3166-2 los códigos de las dependencias administrativas (estados, provincias, etc) para cada país.

Esas listas serán utilizadas para completar la información de las empresas, pues la SEC proporciona el código ISO del país y del estado. La lista permitirá obtener la descripción de los países y estados.

5.3.3. Lista de símbolos (ticker)

Para relacionar los datos obtenidos de una misma empresa obtenidos por distintas fuentes, hemos identificado que se necesitan relacionar su símbolo de cotización (ticker), con el código CIK utilizado por la SEC.

Para ello, se utilizarán técnicas de “web scraping”, para obtener el código CIK, mediante la realización de una consulta a un buscador web que proporciona la SEC, que a partir del ticker nos devuelve toda la información disponible para esa empresa, entre la que se encuentra el código CIK.



EDGAR Search Results

SEC Home » Search the Next-Generation EDGAR System » Company Search » Current Page

APPLE INC CIK# 0000320193 (see all company filings)

SIC: 3571 - ELECTRONIC COMPUTERS
 State location: CA | State of Inc.: CA | Fiscal Year End: 0930
 formerly: APPLE COMPUTER INC. (filings through 2007-01-04)
 formerly: APPLE COMPUTER INC/ FA (filings through 1997-07-28)
 (Assistant Director Office: 3)
 Get insider transactions for this issuer.

Business Address
 ONE INFINITE LOOP
 CUPERTINO CA 95014
 (408) 996-1010

Filter Results: Filing Type: Prior to: (YYYYMMDD) Ownership? include exclude only Limit Results Per Page: 40 Entries

Items 1 - 40 [RSS Feed](#)

Filings	Format	Description	Filing Date
CERTNYS	Documents	[Paper] Certification by the New York Stock Exchange approving securities for listing Acc-no: 999999997-17-006551 (34 Act) Size: 1 KB	2017-05-25
CERTNYS	Documents	[Paper] Certification by the New York Stock Exchange approving securities for listing Acc-no: 999999997-17-006650 (34 Act) Size: 1 KB	2017-05-25
8-A12B	Documents	Registration of securities [Section 12(b)] Acc-no: 0001193125-17-181871 (34 Act) Size: 17 KB	2017-05-24
8-K	Documents	Current report, items 8.01 and 9.01 Acc-no: 0001193125-17-181867 (34 Act) Size: 309 KB	2017-05-24
424B2	Documents	Prospectus [Rule 424(b)(2)] Acc-no: 0001193125-17-175330 (33 Act) Size: 469 KB	2017-05-18

Ilustración 4: Web scraping: Obtención del código CIK, a partir del ticker

6. Diseño del almacén de datos

6.1. Modelado dimensional

El modelado dimensional es una técnica de diseño de bases de datos para almacenes de datos. Este modelado está destinado a ayudar a las consultas de los usuarios sobre los datos, mejorando la comprensión de los mismos. Otra característica importante de este modelado es la rapidez de ejecución de las consultas, pues los almacenes suelen contener elevados volúmenes de información y procesar toda esta información puede ser un proceso complejo.

El modelado dimensional, se basa en crear un esquema en estrella, que está formada por una tabla de hechos que contiene los datos para el análisis, principalmente valores numéricos que se pueden agregar y que representan algún tipo de evento de negocio que se produce a lo largo del tiempo, rodeada por un conjunto de tablas de dimensiones que contienen categorías y descriptores de los hechos.

Este tipo de diseño en estrella permite acceder de forma simple y óptima tanto a datos agregados como de detalle, además de permitir el diseño de sistemas multidimensionales sobre una base de datos relacional, de uso más habitual. Otro aspecto importante tiene relación con el mejor rendimiento y velocidad de consulta que permite, pues es posible indexar de forma individual las tablas de dimensiones sin repercutir en el rendimiento general.

El método utilizado para el diseño del modelo dimensional es un proceso de cuatro etapas, cuyo resultado será un esquema de estrella con una tabla de hechos y sus tablas de dimensiones. A continuación se describen las fases del proceso:

- **Seleccionar un proceso de negocio:** El modelado dimensional requiere seleccionar uno de los procesos de negocio sobre el que construir el DW.
- **Establecer la granularidad para el proceso:** Una vez seleccionado el proceso de negocio, es necesario establecer el nivel de detalle que se desea disponer.

- **Identificar las dimensiones:** En esta fase se deben identificar las dimensiones, teniendo en cuenta el nivel de granularidad establecido en la fase anterior.
- **Identificar los hechos:** La última fase consiste en la identificación de las tablas de hechos y las medidas que formarán un registro de dicha tabla.

Con la definición del problema que hemos realizado vamos a desarrollar en el siguiente apartado el proceso de modelado descrito. Formalmente se debería realizar un proceso iterativo, donde en cada iteración se trabaja con un sólo proceso de negocio, pero por simplicidad se realizará en una sola iteración.

6.1.1. Seleccionar proceso de negocio

El objetivo de este TFG es crear un sistema DW/BI que permita al inversor particular facilitar la toma de decisiones de inversión. Hemos visto que existen dos tendencias en el análisis de las inversiones en los mercados de valores: el análisis técnico y el análisis fundamental. Partiendo de esta base podemos definir los procesos seleccionados.

Por un lado, el análisis técnico requiere de la información de las cotizaciones de las empresas para realizar sus análisis. Podemos decir que el proceso de negocio en este caso son las cotizaciones.

Por otro lado, el análisis fundamental utiliza la información financiera reportada por las empresas para realizar sus análisis. Por ello el proceso de negocio serán la propia información financiera de las empresas.

6.1.2. Establecer la granularidad

Los datos técnicos varían minuto a minuto, pero consideramos que las cotizaciones diarias tienen un nivel de detalle suficiente y por ello una fila de la tabla de hechos se corresponderá con los datos técnicos de una empresa para un día en concreto.

Los datos fundamentales son publicados de forma trimestral y por tanto, una entrada de la tabla de hechos se corresponderá con los datos fundamentales de una empresa en un trimestre concreto.

6.1.3. Identificar las dimensiones

Una vez fijada las granularidades, podemos identificar la dimensiones. En nuestro caso tendremos las siguientes dimensiones: tiempo y empresa. Ver detalle en apartado 6.2 Detalle de las dimensiones

6.1.4. Identificar los hechos

El último paso consiste en determinar las hechos y sus medidas correspondientes que aparecerán en las tablas de hechos.

En el caso del análisis técnico, los hechos serán los precios de apertura, máximo, mínimo, cierre y volumen de las cotizaciones diarias de una empresa. Por otro lado, hemos seleccionado algunos indicadores (ver 4.1.6 Elección de indicadores técnicos) que se corresponderán con nuestras medidas, quedando el esquema en estrella de la siguiente forma:

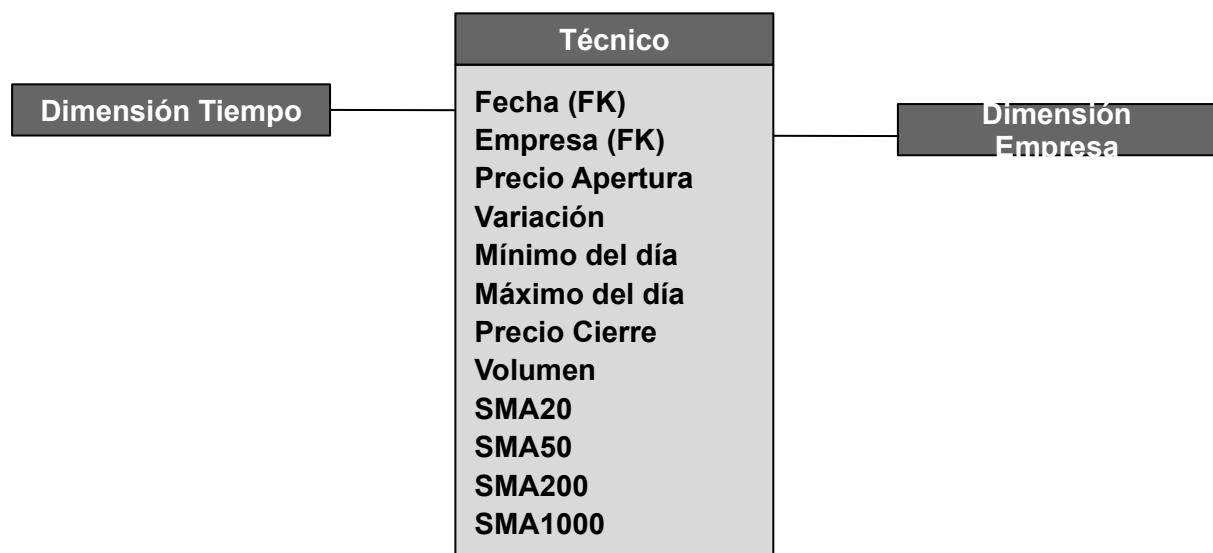


Ilustración 5: Tabla de hechos técnicos

Para el análisis fundamental, los hechos se corresponderán con información financiera seleccionada extraída de la cuenta de resultados, balance de situación y/o los flujos de efectivo. Estos hechos han sido elegidos teniendo en cuenta las necesidades de los ratios financieros seleccionados (ver 4.2.6 Elección de ratios financieros) como medidas. A continuación se muestra el esquema en estrella con los hechos y medidas:

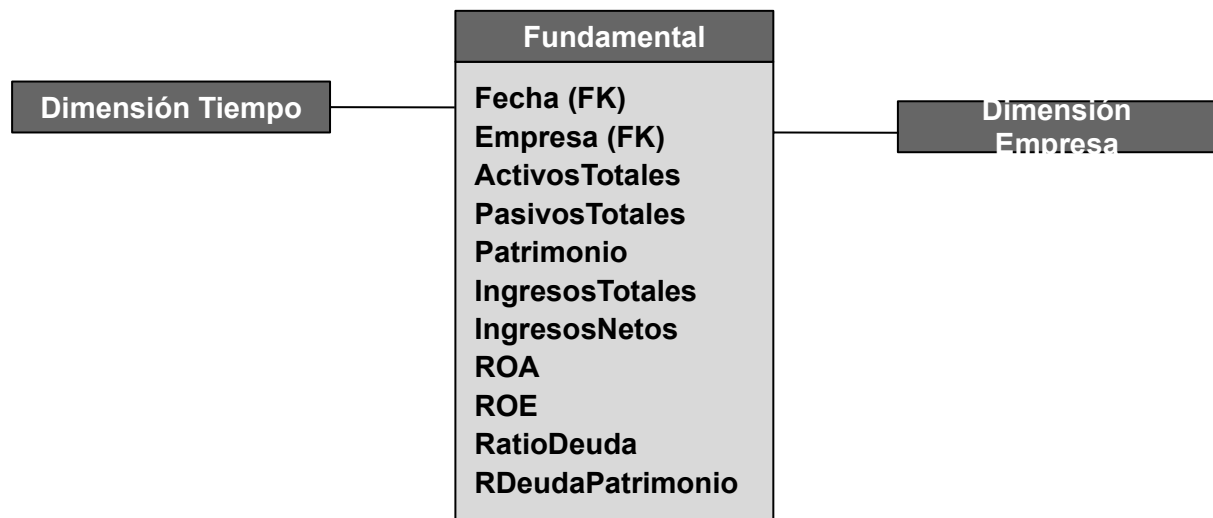


Ilustración 6: Tabla de hechos fundamentales

6.2. Detalle de las dimensiones

Para cada una de las dimensiones que se han identificado en la tercera fase, anteriormente descrita, indicaremos todos los atributos que contendrá, especificando la siguiente información:

- **Nombre del atributo:** El nombre usado en el ámbito del negocio.
- **Definición del atributo:** Pequeño resumen que permita adquirir cierto conocimiento sobre el atributo
- **Tipo de dato:** indica el tipo de dato: entero, string, fecha, etc.
- **Valores de ejemplo:** Posibles valores de ejemplo que contendrá el atributo.

6.2.1. Dimensión Tiempo

Nombre del atributo	Descripción del atributo	Tipo	Valores de ejemplo
Identificador	Número entero que representa una fecha	Integer	20170101
Año	Entero que representa el año	Integer	2017
Mes	Entero que representa el número de mes	Integer	01
Día del mes	Entero que representa el número de día del mes	Integer	01
Fecha	Representa una fecha	Date	01/01/2017
Descripción media de fecha	Texto que describe la fecha en un formato de tamaño medio	Varchar	01-ene-2017
Descripción completa de fecha	Texto que describe la fecha en un formato completo	Varchar	domingo 1 de enero de 2017
Descripción del día de la semana	Descripción del día de la semana	Varchar	domingo
Descripción del mes	Descripción del mes	Varchar	enero
Descripción del trimestre	Descripción del trimestre	Varchar	Q1
Descripción del año y trimestre	Descripción del año y trimestre	Varchar	2017-Q1
Descripción del año y mes	Descripción del año y mes	Varchar	2017-01

Tabla 6: Dimensión tiempo

6.2.2. Dimensión Empresa

Nombre del atributo	Descripción del atributo	Tipo	Valores de ejemplo
Denominación de la empresa	Nombre de la empresa	Varchar	BAYERISCHE MOTOREN WERKE AG
Ticker	Identificador de la empresa en el mercado financiero en el que cotiza	Varchar	BMW
CIK	en un código asignado por la SEC y que identifica de forma única a una empresa	Varchar	1090872
Sector	Sector de la economía al que pertenece la empresa	Varchar	Bienes de consumo
Industria	Industria dentro del sector al que pertenece la empresa	Varchar	Fabricante de vehículos
Mercado	Mercados en el que cotizan las acciones de la empresa	Varchar	DAX
País	País donde se encuentra la sede social de la empresa	Varchar	Alemania
Estado	Estado/Provincia donde se encuentra la sede social de la empresa	Varchar	Múnich

Tabla 7: Dimensión empresa

6.3. Modelado físico

Una vez presentado el modelado dimensional realizado, debemos diseñar el modelo lógico que le dará soporte. Recordemos que Pentaho Community Edition es un sistema ROLAP, que permite la utilización de herramientas OLAP sobre una base de datos relacional. Por ello vamos a realizar un modelado entidad-relación, con algunas adaptaciones para mejorar el rendimiento de las consultas.

La principal adaptación consiste en la eliminación de las claves primarias y extranjeras del modelo, aunque en la imagen siguiente se han dibujado para mejorar la comprensión del mismo. Esta decisión se ha tomado, siguiendo las directrices en numerosas fuentes para el diseño físico, que con el objetivo de agilizar las consultas eliminan toda clave. Aunque a priori, pueda parecer que esta decisión elimina la integridad referencial del modelo, no es así, pues el proceso ETL es el encargado de mantener esta integridad y que el sistema funciona como se espera.

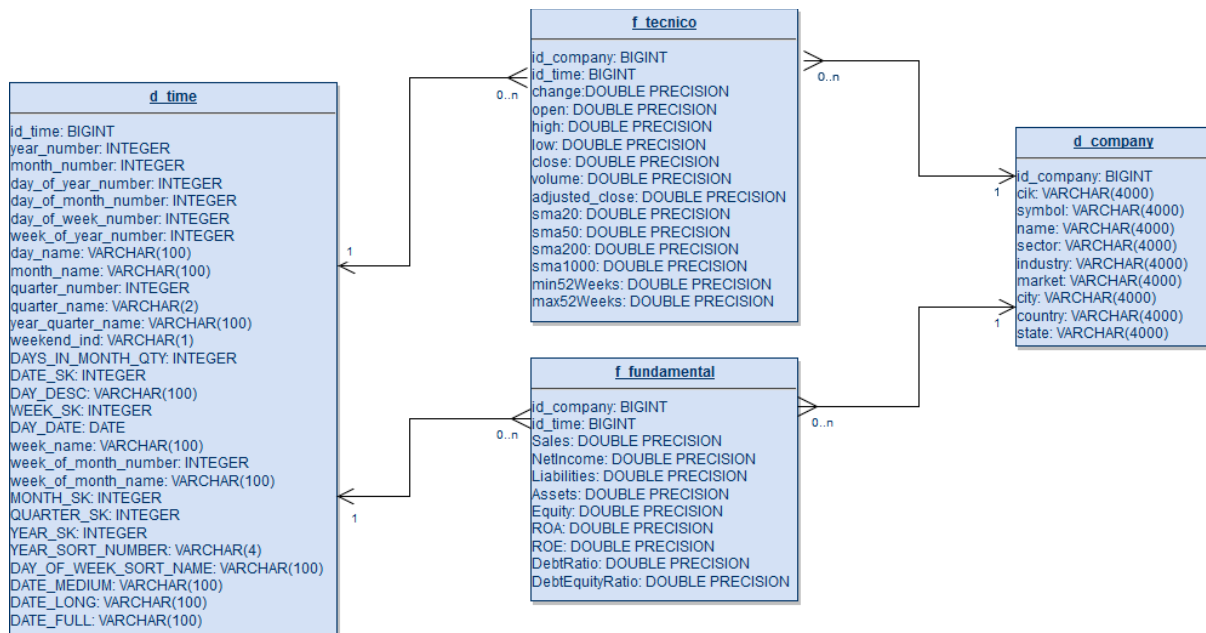


Ilustración 7: Modelo físico

También, se ha diseñado un área de stage donde almacenar la información tal cual se obtiene de las distintas fuentes de información. Por otro lado, se utilizará una tabla para agilizar los cálculos a realizar para las medidas de la tabla de hechos técnicos.

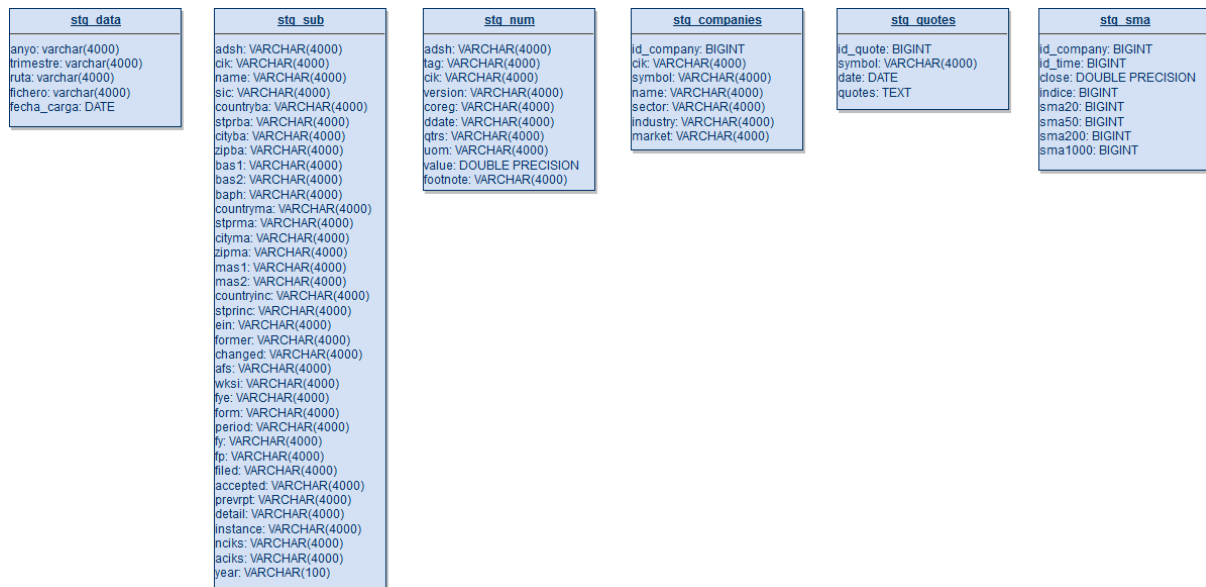


Ilustración 8: Modelo físico: Área stage

A continuación se detallan los campos para cada una de las tablas creadas.

D_TIME			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_time	Identificador de la fecha tiene con formato YYYYMMDD	BIGINT	
year_number	Valor entero que representa el año	INTEGER	
month_number	Valor entero que representa el mes	INTEGER	
day_of_year_number	Valor entero que representa el día del año	INTEGER	
day_of_week_number	Valor entero que representa el día de la semana	INTEGER	
week_of_year	Valor entero que representa la semana del año	INTEGER	
date_medium	Descripción de la fecha en formato medio. Ejemplo: 01-ene-2017	VARCHAR	100
date_long	Descripción de la fecha en formato largo. Ejemplo: 1 de enero de 2017	VARCHAR	100
date_full	Descripción de la fecha en formato entero. Ejemplo: domingo 1 de enero de 2017	VARCHAR	100

Tabla 8: Modelo físico dimensiones: Dimensión tiempo (d_time)

(*): Por extensión de los campos de omiten algunos de ellos

STG_COMPANY			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_company	Identificador de la empresa	BIGINT	
cik	Código asignado por la SEC a la empresa	VARCHAR	4000
symbol	Símbolo que identifica las acciones de la empresa (ticker)	VARCHAR	4000
name	Nombre de la empresa	VARCHAR	4000
sector	Sector de la economía a la que pertenece	VARCHAR	4000
industry	Industria a la que pertenece	VARCHAR	4000
market	Mercado en el que cotizan sus acciones	VARCHAR	4000
city	Ciudad de la empresa	VARCHAR	4000
country	País de la empresa	VARCHAR	4000
state	Estado/Provincia al que pertenece la empresa	VARCHAR	4000

Tabla 9: Modelo físico dimensiones: Dimensión empresa (d_company)

F_TECNICO			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_company	Identificador de la empresa	BIGINT	
id_time	Identificador de fecha	BIGINT	
change	Diferencia entre el precio de cierre y el precio de cierre anterior	DOUBLE	
open	Precio de apertura	DOUBLE	
high	Precio máximo de la sesión	DOUBLE	
low	Precio mínimo de la sesión	DOUBLE	
close	Precio de cierre	DOUBLE	
volume	Volumen de acciones negociadas	DOUBLE	
adjusted_close	Precio de cierre ajustado	DOUBLE	
sma20	Media móvil de las 20 sesiones	DOUBLE	
sma50	Media móvil de las 50 sesiones	DOUBLE	
sma200	Media móvil de las 200 sesiones	DOUBLE	
sma500	Media móvil de las 500 sesiones	DOUBLE	
sma1000	Media móvil de las 1000 sesiones	DOUBLE	
min52Weeks	Precio mínimo de las últimas 52 semanas	DOUBLE	
max52Weeks	Precio máximo de las últimas 52 semanas	DOUBLE	

Tabla 10: Modelo físico hechos: Técnicos (f_tecnico)

F_FUNDAMENTAL			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_company	Identificador de la empresa	BIGINT	
id_time	Identificador de fecha	BIGINT	
sales	Ventas del periodo	DOUBLE	
Net Income	Beneficio del periodo	DOUBLE	
Liabilities	Pasivos totales	DOUBLE	
Assets	Activos totales	DOUBLE	
Equity	Patrimonio total	DOUBLE	
ROA	Return in assets	DOUBLE	
ROE	Return on equity	DOUBLE	
DebtRatio	Ratio de deuda	DOUBLE	
DebtEquityRatio	Ratio de deuda con respecto al patrimonio	DOUBLE	

Tabla 11: Modelo físico hechos: Fundamentales (f_fundamental)

STG_DATA			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
anyo	Año asociado a la información del fichero	VARCHAR	4000
trimestre	Trimestre asociado a la información del fichero	VARCHAR	4000
ruta	Ruta local donde se encuentra el fichero	VARCHAR	4000
fichero	Nombre del fichero	VARCHAR	4000
fecha_carga	Fecha de carga del fichero en el DW	DATE	

Tabla 12: Modelo físico área stage: Tabla stg_data

STG_SUB			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
adsh	Número de control asignado por la SEC al formulario enviado	VARCHAR	4000
cik	Código asignado por la SEC a la empresa	VARCHAR	4000
name	Denominación de la empresa	VARCHAR	4000
sic	Código de clasificación industrial	VARCHAR	4000
countryba	Código ISO 3166-1 del país de la sede social de la empresa	VARCHAR	4000
stprba	El estado o provincia de la sede social de la empresa, si el país es US o CA	VARCHAR	4000
cityba	Ciudad de la sede social de la empresa	VARCHAR	4000
zipba	Código postal de la sede social de la empresa	VARCHAR	4000
bas1, bas2	Dirección de la sede social de la empresa	VARCHAR	4000
baph	Teléfono de la empresa	VARCHAR	4000
fye	Fecha fin del año fiscal	VARCHAR	4000
form	Tipo de formulario	VARCHAR	4000
período	Fecha del balance de situación	VARCHAR	4000
fy	Año fiscal	VARCHAR	4000
fp	Período fiscal	VARCHAR	4000

Tabla 13: Modelo físico área stage: Tabla stg_sub

(*): Por extensión de los campos de omiten algunos de ellos

STG_NUM			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
adsh	Número de control asignado por la SEC	VARCHAR	4000
cik	Código asignado por la SEC a la empresa	VARCHAR	4000
tag	Código asociado a la descripción del valor en el conjunto de datos TAG	VARCHAR	4000
version	Versión asociada a la descripción del valor en el conjunto de datos TAG	VARCHAR	4000
coreg	Empresa matriz, si la tiene	VARCHAR	4000
ddate	Fecha final para el dato	VARCHAR	4000
qtrs	El trimestre asociado al valor	VARCHAR	4000
uom	Unidad de medida	VARCHAR	4000
value	El valor numérico.	VARCHAR	4000
footnote	Texto o notas asociadas al valor en el registro original	VARCHAR	4000

Tabla 14: Modelo físico área stage: Tabla stg_num

STG_COMPANIES			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_company	Identificador de la empresa	BIGINT	
cik	Código asignado por la SEC a la empresa	VARCHAR	4000
symbol	Símbolo que identifica las acciones de la empresa (ticker)	VARCHAR	4000
name	Nombre de la empresa	VARCHAR	4000
sector	Sector de la economía a la que pertenece	VARCHAR	4000
industry	Industria a la que pertenece	VARCHAR	4000
market	Mercado en el que cotizan sus acciones	VARCHAR	4000

Tabla 15: Modelo físico área stage: Tabla stg_companies

STG_QUOTES			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_quote	Identificador	BIGINT	
symbol	Símbolo que identifica las acciones de la empresa (ticker)	VARCHAR	4000
date	Fecha de carga de los datos en el DW	DATE	
quotes	Resultado de la llamada a la API Yahoo Finance, contiene el histórico de cotizaciones de la empresa	TEXT	

Tabla 16: Modelo físico área stage: Tabla stg_quotes

STG_SMA			
NOMBRE	DESCRIPCIÓN	TIPO	TAMAÑO
id_company	Identificador de la empresa	BIGINT	
id_time	Símbolo que identifica las acciones de la empresa (ticker)	BIGINT	
close	Precio de cierre para el día id_time	DOUBLE	
indice	Valor temporal utilizado para el cálculo de las medias móviles	BIGINT	
sma20	Media móvil de 20 sesiones	BIGINT	
sma50	Media móvil de 50 sesiones	BIGINT	
sma200	Media móvil de 200 sesiones	BIGINT	
sma500	Media móvil de 500 sesiones	BIGINT	
sma1000	Media móvil de 1000 sesiones	BIGINT	

Tabla 17: Modelo físico área stage: Tabla stg_sma

7. Implementación del almacén de datos

7.1. Procesos ETL

En este apartado describiremos el proceso ETL generado para crear el DW. Para su descripción utilizaremos un enfoque **Top-down** (“de arriba abajo”), empezaremos mostrando las tareas (jobs) de alto nivel para a continuación entrar en el detalle de cada una de las subtareas y transformaciones que las forman.

7.1.1. Carga del data warehouse

Este es el job principal de la ETL y es el encargado de lanzar los jobs secundarios. Al iniciarse ejecuta un job de comprobación de la base de datos, después realizar la carga del área de stage, que carga datos utilizados posteriormente para la carga de las tablas de dimensiones y hechos. A continuación se lanzan las cargas de las tablas de dimensiones y hechos. Si los pasos intermedios se han ejecutado el proceso finaliza de forma correcta. En el caso, de que alguno de los pasos intermedio fallase se aborta la carga completa.

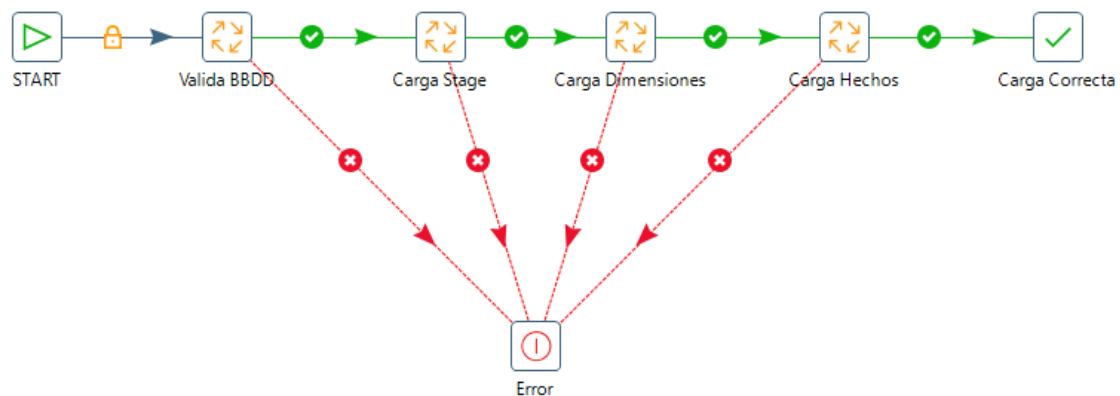


Ilustración 9: Job principal de la ETL

7.1.2. Validación de la base de datos

Este job es el encargado de asegurar que las tablas en base de datos necesarias para la carga de la ETL y por tanto requeridas por el sistema DW/BI, existen. En caso contrario, lanzará los scripts sql de creación de tablas correspondiente.

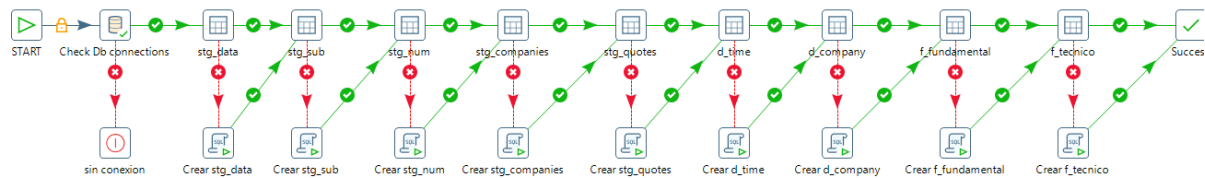


Ilustración 10: Job validación de la base de datos

7.1.3. Carga de datos del área de “stage”

Este job se encarga de la carga del área de stage, que obtendrá los datos desde la distintas fuentes de información y la depositará en tablas en la base de datos para su posterior utilización por parte de los procesos de carga de las tablas de dimensiones y hechos.

En primer lugar, se descargan los datos fundamentales de los servidores de la SEC y se almacenan como fichero en el equipo. A continuación se carga una tabla que ha sido denominada de referencias y que es utilizada a modo índice para los ficheros descargados en el paso anterior y de esta forma permitir cargas incrementales de los datos.

Seguidamente se cargan las tablas con los datos fundamentales, obtenidos anteriormente, haciendo uso de la tabla de referencias. Una vez completado, se lanza la carga de las empresas.

Debido al alto volumen de información disponible en los registros de la SEC se ha decidido, sólo cargar en el DW/BI la información financiera de las empresas que cotizan en las bolsas de Estados Unidos. Estas bolsas con el NASDAQ, NYSE y AMEX.

Por último, se obtienen las cotizaciones históricas para estas empresas.

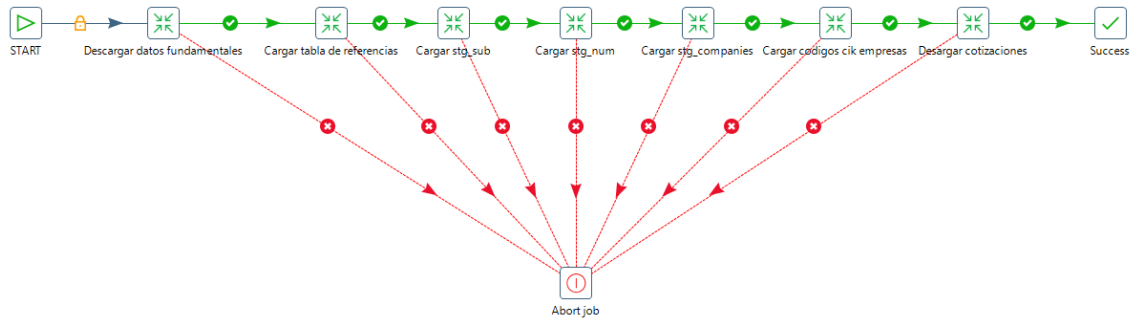


Ilustración 11: Job carga del área de "stage"

7.1.4. Carga de las dimensiones

El job que se muestra a continuación, ejecuta las transformaciones que se encargan de la carga de las dimensiones.

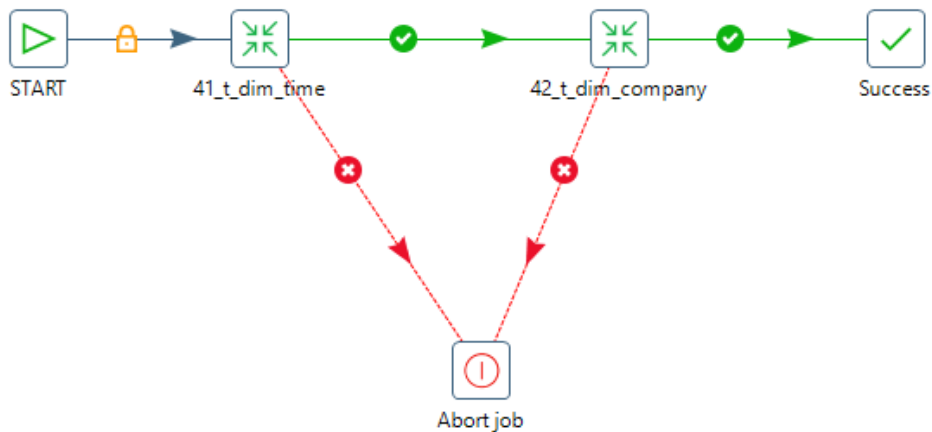


Ilustración 12: Job carga de las dimensiones

7.1.5. Carga de los hechos

El siguiente job ejecuta las transformaciones que se encargan de la carga de las tablas de hechos.

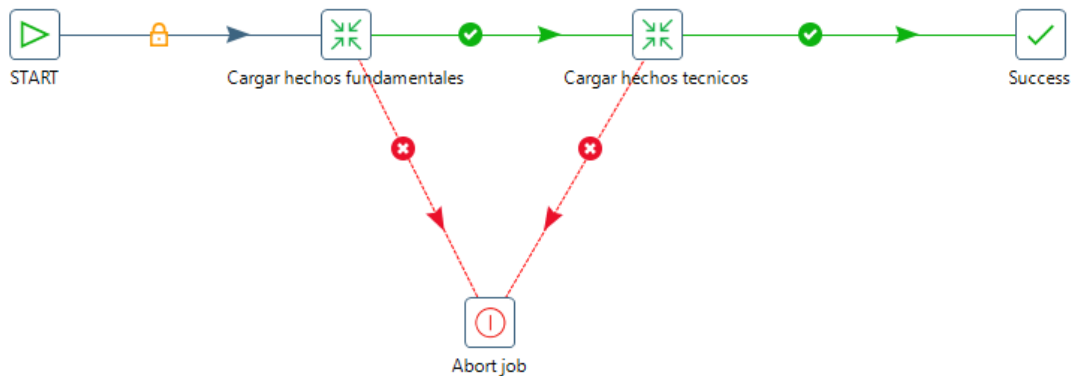


Ilustración 13: Job carga de las tablas de hechos

7.1.6. Descargar datos financieros

La siguiente transformación obtiene la información financiera de la página web de la Securities Exchange Commission (SEC). Ver apartado 5.2.

Esta transformación a grandes rasgos, genera las urls de los archivos zip localizados en la web de la SEC y las rutas locales donde se almacenarán los ficheros. Se ha utilizado Apache VFS para acceder directamente a los ficheros que contienen los archivos comprimidos y facilitar así su obtención.

A continuación se muestra un ejemplo que obtiene el fichero sub.txt contenido en el comprimido 2009q4.zip

```
zip:https://www.sec.gov/files/dera/data/financial-statement-data-sets/2009q4.zip!/sub.txt
```

Se ha decidido guardar en local los ficheros descargados para permitir la carga incremental de los datos. Como veremos en el siguiente apartado, los ficheros descargados se guardan en una tabla de referencias que servirá a tal objetivo.

Una vez generadas las rutas locales y las urls para obtener los ficheros, se utiliza lanzan procesos en paralelo para obtener los distintos ficheros y de esta forma acelerar la ejecución de la transformación.

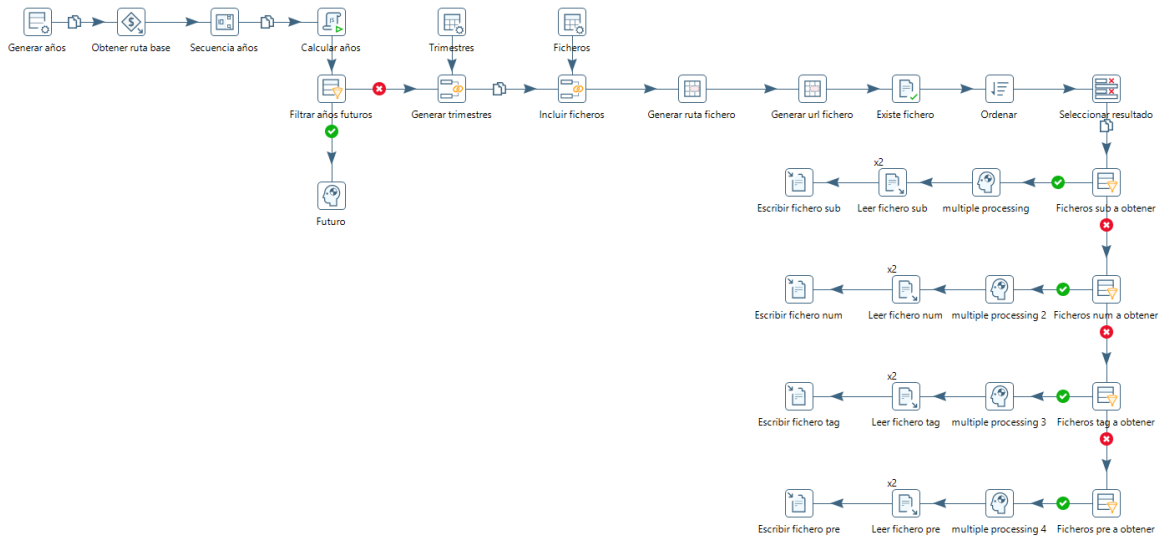


Ilustración 14: Transacción: descarga de información financiera de la SEC

7.1.7. Carga tabla de referencias

Como hemos indicado anteriormente, esta transformación crea un registro en base de datos por cada fichero descargado. Para ello, genera las posibles rutas de los ficheros, comprueba si existen y en tal caso los persiste en una tabla del área de stage, que será posteriormente utilizada por otra transformación para realizar la carga de los datos en BBDD.

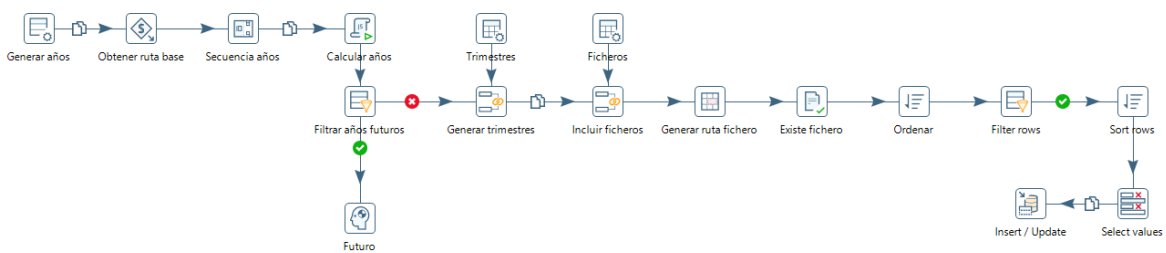


Ilustración 15: Transacción: carga de la tabla de referencias

Resultado obtenido:

stg_data Enter a SQL expression to filter results (use Ctrl+Space)					
	anyo	trimestre	ruta	fichero	fecha_carga
1	2017	q1	file:///C:/Users/dari/Workspace/pentaho/2017/q1/tag.txt	tag.txt	2017-05-19
2	2017	q1	file:///C:/Users/dari/Workspace/pentaho/2017/q1/pre.txt	pre.txt	2017-05-19
3	2017	q1	file:///C:/Users/dari/Workspace/pentaho/2017/q1/num.txt	num.txt	2017-05-19
4	2017	q1	file:///C:/Users/dari/Workspace/pentaho/2017/q1/sub.txt	sub.txt	2017-05-19
5	2016	q1	file:///C:/Users/dari/Workspace/pentaho/2016/q1/tag.txt	tag.txt	2017-05-19
6	2016	q1	file:///C:/Users/dari/Workspace/pentaho/2016/q1/pre.txt	pre.txt	2017-05-19
7	2016	q2	file:///C:/Users/dari/Workspace/pentaho/2016/q2/tag.txt	tag.txt	2017-05-19
8	2016	q2	file:///C:/Users/dari/Workspace/pentaho/2016/q2/pre.txt	pre.txt	2017-05-19
9	2016	q3	file:///C:/Users/dari/Workspace/pentaho/2016/q3/tag.txt	tag.txt	2017-05-19
10	2016	q3	file:///C:/Users/dari/Workspace/pentaho/2016/q3/pre.txt	pre.txt	2017-05-19
11	2016	q4	file:///C:/Users/dari/Workspace/pentaho/2016/q4/tag.txt	tag.txt	2017-05-19
12	2016	q4	file:///C:/Users/dari/Workspace/pentaho/2016/q4/pre.txt	pre.txt	2017-05-19
13	2016	q2	file:///C:/Users/dari/Workspace/pentaho/2016/q2/num.txt	num.txt	2017-05-19
14	2016	q4	file:///C:/Users/dari/Workspace/pentaho/2016/q4/num.txt	num.txt	2017-05-19
15	2016	q3	file:///C:/Users/dari/Workspace/pentaho/2016/q3/num.txt	num.txt	2017-05-19
16	2016	q1	file:///C:/Users/dari/Workspace/pentaho/2016/q1/sub.txt	sub.txt	2017-05-19
17	2016	q2	file:///C:/Users/dari/Workspace/pentaho/2016/q2/sub.txt	sub.txt	2017-05-19
18	2016	q3	file:///C:/Users/dari/Workspace/pentaho/2016/q3/sub.txt	sub.txt	2017-05-19
19	2016	q4	file:///C:/Users/dari/Workspace/pentaho/2016/q4/sub.txt	sub.txt	2017-05-19
20	2016	q1	file:///C:/Users/dari/Workspace/pentaho/2016/q1/num.txt	num.txt	2017-05-19

Ilustración 16: ETL: Tabla stg_data

7.1.8. Carga tabla SUB

Esta transformación leerá de la tabla de referencias los ficheros sub.txt descargados y no tratados anteriormente, y los guarda en la tabla correspondiente del área de stage. Una vez finalizado el proceso se actualiza la tabla de referencias, mecanismo que posibilita la carga incremental citada anteriormente.

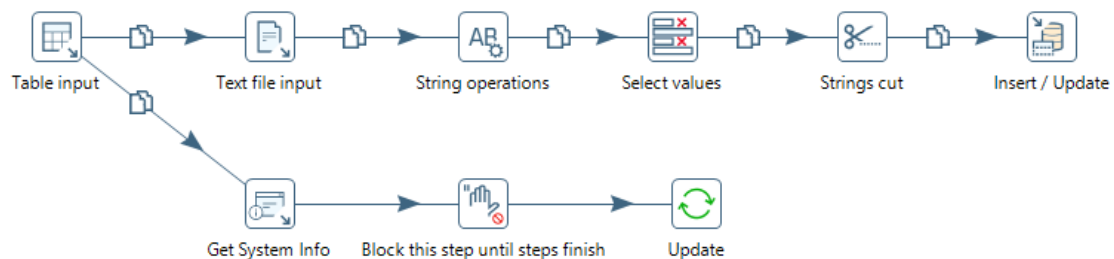


Ilustración 17: Transacción: carga del conjunto de datos SUB

Resultado obtenido:

id	adsh	stk	name	sic	countryba	stprba	cityba	zipba	bas1	bas2	bapb	countryma
1	0000038074-09-000029	0000038074	FOREST LABORATORIES INC	2834	US	NV	NEW YORK	10022	909 THIRD AVENUE	[NULL]	(212)421-785	US
2	0000029280-09-000018	0000029280	SOUTHWEST AIRLINES CO	4512	US	TX	DALLAS	75235	2702 LOVE FIELD DR	P O BOX 36611	214)7924000	US
3	0000277848-09-000066	0000277848	CSX CORP	4011	US	FL	JACKSONVILLE	32202	500 WATER STREET	15TH FLOOR	904)3932000	US
4	0000796343-09-000026	0000796343	ADOBE SYSTEMS INC	7372	US	CA	SAN JOSE	95110-2704	345 PARK AVE	[NULL]	408)3660000	US
5	0000831259-09-000047	0000831259	FREEMONT MCMORAN COPPER & GOLD INC	1000	US	AZ	PHOENIX	85004-4414	ONE NORTH CENTRAL AVENUE	[NULL]	602)3681000	US
6	0000891618-09-000150	0001164727	NEWMONT MINING CORP / DE	1040	US	CO	GREENWOOD VILLAGE	80222	6363 SOUTH FIDDLERS GREEN CIRCLE	[NULL]	303-863-7414	US
7	0000891618-09-000166	0000093410	CHEVRON CORP	2911	US	CA	SAN RAMON	94583	6001 BOLLINGER CANYON ROAD	[NULL]	925-842-1000	US
8	0000950123-09-008350	0000013610	BOWNE & CO INC	2750	US	NY	NEW YORK	10041-0006	55 WATER STREET	[NULL]	212)4245500	US
9	0000950134-09-009913	0001349436	SANDRIDGE ENERGY INC	1311	US	OK	OKLAHOMA CITY	73118	1801 NW EXPRESSWAY	SUITE 1600	405-753-5500	US
10	0001002638-09-000023	0001002638	OPEN TEXT CORP	7373	CA	ON	ONTARIO CANADA	N2L 0A1	275 FRANK TOMPA DRIVE	WATERLOO	519-889-7111	CA
11	0001031296-09-000011	0001031296	FIRSTENERGY CORP	4911	US	OH	AKRON	44308-1890	76 SOUTH MAIN ST	[NULL]	330-761-7837	US
12	0001104659-09-025790	0000883984	ICU MEDICAL INC/DE	3841	US	CA	SAN CLEMENTE	92763	951 CALLE AMANECER	[NULL]	949-366-2183	US
13	0001104659-09-029605	0000901491	PAPA JOHN'S INTERNATIONAL INC	5812	US	KY	LOUISVILLE	40299-2334	2002 PAPA JOHN'S BOULEVARD	[NULL]	502)2617272	US
14	0001104659-09-031513	0001124198	FLUOR CORP	1600	US	TX	IRVING	75039	6700 LAS COLINAS BLVD	[NULL]	469)9870000	US
15	0001193125-09-088851	0001335793	CNX GAS CORP	1311	US	PA	CANONSBURG	15317	CNX CENTER	1000 CONSOL ENERGY DRIVE	724-485-4000	US
16	0001193125-09-088859	0001070412	CONSOL ENERGY INC	1321	US	PA	PITTSBURGH	15241	C/O CONSOL INC	1800 WASHINGTON RD	724-485-4000	US
17	0001193125-09-091829	0000884905	PELXAR INC	2810	US	CT	DANBURY	06810-5113	39 OLD RIDGEBURY RD	[NULL]	203)8372000	US
18	0001193125-09-094509	0000108772	XEROX CORP	3577	US	CT	NORWALK	06856	45 GLOVER AVENUE	PO BOX 4505	203)9623000	US
19	0001193125-09-096522	0001365135	WESTERN UNION CO	7389	US	CO	ENGLEWOOD	80112	12500 EAST BELFORD AVENUE	[NULL]	(720) 332-33	US
20	0001193125-09-100014	0001316631	LIBERTY GLOBAL, INC.	4841	US	CO	ENGLEWOOD	80112	12300 LIBERTY BOULEVARD	[NULL]	720 875 5800	US
21	0001193125-09-101273	0000086144	SAFEWAY INC	5411	US	CA	PLEASANTON	94588	5918 STONERIDGE MALL RD	[NULL]	925)4673000	US

Ilustración 18: ETL: Tabla stg_sub

7.1.9. Carga tabla NUM

Al igual que la transformación anterior, se cargan los ficheros num.txt en el área de stage y se actualizan las correspondientes entradas de la tabla de referencias.

Como particularidad del conjunto de datos que carga esta transformación, indicar que habitualmente los informes de un trimestre contienen la información del mismo trimestre del año anterior para permitir la realización de comparaciones. Esta transformación elimina esta información, pues puede provocar incorporar registros duplicados que dificultarían el tratamiento posterior de los datos.

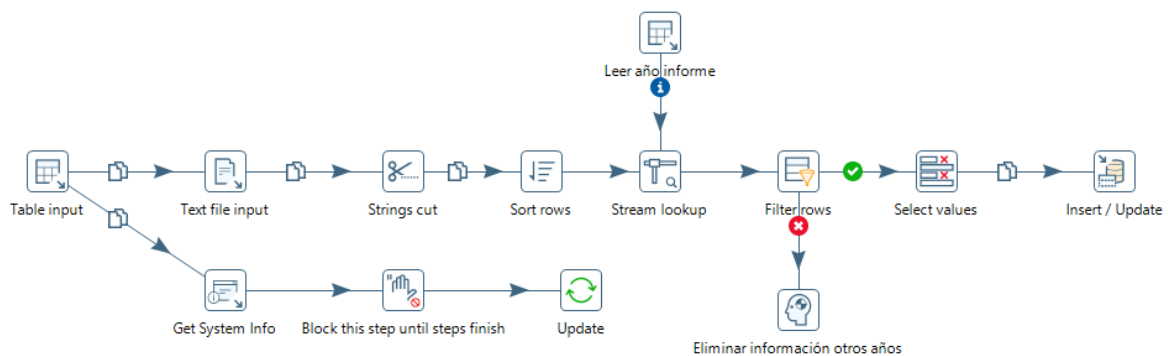


Ilustración 19: Transacción: carga del conjunto de datos NUM

Resultado obtenido:

stg_num | Enter a SQL expression to filter results (use Ctrl+Space)

	adsh	tag	cik	version	coreg	ddate	qtrs	uom	value
1	0000003453-10-000042	AccountsPayableCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	127.000.000
2	0000003453-10-000042	AccruedAndOtherLiabilities	0000003453	0000003453-10-000042	[NULL]	20100630	0	USD	88.000.000
3	0000003453-10-000042	AccumulatedDepreciationDepletionAndAmortizationPropertyPlantAndEquipment	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	1.217.000.000
4	0000003453-10-000042	AccumulatedOtherComprehensiveIncomeLossNetOfTax	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	-76.000.000
5	0000003453-10-000042	AdditionalPaidInCapitalCommonStock	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	215.000.000
6	0000003453-10-000042	Assets	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	2.461.000.000
7	0000003453-10-000042	AssetsCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	318.000.000
8	0000003453-10-000042	CashAndCashEquivalentsAtCarryingValue	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	36.000.000
9	0000003453-10-000042	CommonStockDividendsPerShareCashPaid	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,315
10	0000003453-10-000042	CommonStockValue	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	33.000.000
11	0000003453-10-000042	CostOfGoodsAndServicesSold	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	319.600.000
12	0000003453-10-000042	DebtCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	56.000.000
13	0000003453-10-000042	DeferredTaxAssetsNetCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	6.000.000
14	0000003453-10-000042	DeferredTaxLiabilitiesNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	423.000.000
15	0000003453-10-000042	DefinedBenefitPlanAssetsForPlanBenefitsNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	3.000.000
16	0000003453-10-000042	DividendsPayableAmount	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	13.000.000
17	0000003453-10-000042	EarningsPerShareBasic	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,7
18	0000003453-10-000042	EarningsPerShareDiluted	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,7
19	0000003453-10-000042	EmployeeRelatedLiabilitiesCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	17.000.000
20	0000003453-10-000042	IncomeBeforeTaxes	0000003453	0000003453-10-000042	[NULL]	20100630	1	USD	36.600.000
21	0000003453-10-000042	IncomeLossFromContinuingOperations	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	22.900.000
22	0000003453-10-000042	IncomeLossFromContinuingOperationsPerBasicShare	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,55
23	0000003453-10-000042	IncomeLossFromContinuingOperationsPerDilutedShare	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,55
24	0000003453-10-000042	IncomeLossFromDiscontinuedOperationsNetOfTaxAttributableToReportingEntity	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	6.000.000
25	0000003453-10-000042	IncomeLossFromDiscontinuedOperationsNetOfTaxPerBasicShare	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,15
26	0000003453-10-000042	IncomeLossFromDiscontinuedOperationsNetOfTaxPerDilutedShare	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	0,15
27	0000003453-10-000042	IncomeLossFromEquityMethodInvestments	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	-200.000
28	0000003453-10-000042	IncomeTaxExpenseBenefit	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	13.700.000
29	0000003453-10-000042	InterestExpense	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	6.500.000
30	0000003453-10-000042	InventoryNet	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	59.000.000
31	0000003453-10-000042	InventoryRealEstateLandAndLandDevelopmentCosts	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	106.000.000
32	0000003453-10-000042	InvestmentIncomeNet	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	200.000
33	0000003453-10-000042	InvestmentsInAffiliatesSubsidiariesAssociatesAndJointVentures	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	283.000.000
34	0000003453-10-000042	LiabilitiesAndStockholdersEquity	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	2.461.000.000
35	0000003453-10-000042	LiabilitiesCurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	299.000.000
36	0000003453-10-000042	LiabilitiesNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	1.061.000.000
37	0000003453-10-000042	LongTermDebtNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	470.000.000
38	0000003453-10-000042	NetIncomeLoss	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	28.900.000
39	0000003453-10-000042	OperatingExpenses	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	356.500.000
40	0000003453-10-000042	OperatingIncomeLoss	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	42.400.000
41	0000003453-10-000042	OtherAssetsNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	198.000.000
42	0000003453-10-000042	OtherLiabilitiesNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	48.000.000
43	0000003453-10-000042	OtherNonoperatingIncome	0000003453	us-gaap/2009	[NULL]	20100630	1	USD	700.000
44	0000003453-10-000042	PensionAndOtherPostretirementDefinedBenefitPlansLiabilitiesNoncurrent	0000003453	us-gaap/2009	[NULL]	20100630	0	USD	120.000.000
45	0000003453-10-000042	PrepaidExpensesAndOtherAssets	0000003453	0000003453-10-000042	[NULL]	20100630	0	USD	22.000.000

Ilustración 20: ETL: Tabla stg_num

7.1.10. Carga de empresas

Como hemos indicado en el apartado 5.3.1, obtenemos las empresas que cotizan en los mercados Nasdaq, Nyse y Amex.

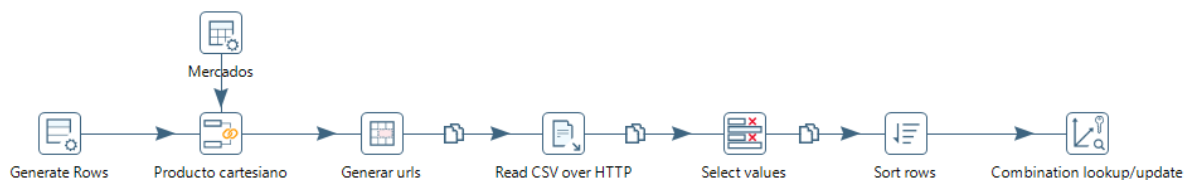


Ilustración 21: Transacción: descarga de las empresas del AMEX, NASDAQ y NYSE

Resultado obtenido:

stg_companies Enter a SQL expression to filter results (use Ctrl+Space)							
	id_company	cik	symbol	Name	sector	industry	market
1	1	0001090872	A	Agilent Technologies, Inc.	Capital Goods	Biotechnology: Laboratory Analytical Instruments	nyse
2	2	0001675149	AA	Alcoa Corporation	Capital Goods	Metal Fabrications	nyse
3	3	0001611787	AAAP	Advanced Accelerator Applications S.A.	Health Care	Major Pharmaceuticals	nasdaq
4	4	0001606180	AAC	AAC Holdings, Inc.	Health Care	Medical Specialties	nyse
5	5	0000086201	AAL	American Airlines Group, Inc.	Transportation	Air Freight/Delivery Services	nasdaq
6	6	0001555074	AAMC	Altisource Asset Management Corp	Finance	Real Estate	amex
7	7	0000008177	AAME	Atlantic American Corporation	Finance	Life Insurance	nasdaq
8	8	0000706688	AAN	Aaron's, Inc.	Technology	Diversified Commercial Services	nyse
9	9	0001158114	AAOI	Applied Optoelectronics, Inc.	Technology	Semiconductors	nasdaq
10	10	0000824142	AAON	AAON, Inc.	Capital Goods	Industrial Machinery/Components	nasdaq
11	11	0001158449	AAP	Advance Auto Parts Inc	Consumer Services	Other Specialty Stores	nyse
12	12	0001630940	AAPC	Atlantic Alliance Partnership Corp.	Consumer Services	Services-Misc. Amusement & Recreation	nasdaq
13	13	0000320193	AAPL	Apple Inc.	Technology	Computer Manufacturing	nasdaq
14	14	0001500217	AAT	American Assets Trust, Inc.	Consumer Services	Real Estate Investment Trusts	nyse
15	15	0001015647	AAU	Almaden Minerals, Ltd.	Basic Industries	Precious Metals	amex
16	16	0001259995	AAV	Advantage Oil & Gas Ltd	Energy	Oil & Gas Production	nyse
17	17	0001135185	AAWW	Atlas Air Worldwide Holdings	Transportation	Transportation Services	nasdaq
18	18	n/a	AAXJ	iShares MSCI All Country Asia ex Japan Index Fund	n/a	n/a	nasdaq
19	19	0001069183	AAXN	Axon Enterprise, Inc.	Capital Goods	Ordinance And Accessories	nasdaq
20	20	0000825313	AB	AllianceBernstein Holding L.P.	Finance	Investment Managers	nyse
21	21	0001486299	ABAC	Aoxin Tianli Group, Inc.	Consumer Non-Durables	Farming/Seeds/Milling	nasdaq
22	22	0000881890	ABAX	ABAXIS, Inc.	Capital Goods	Industrial Machinery/Components	nasdaq
23	23	0001091587	ABB	ABB Ltd	Consumer Durables	Electrical Products	nyse
24	24	0001551152	ABBV	AbbVie Inc.	Health Care	Major Pharmaceuticals	nyse
25	25	0001140859	ABC	AmerisourceBergen Corporation (Holding Co)	Health Care	Other Pharmaceuticals	nyse
26	26	0000351569	ABCB	Ameris Bancorp	Finance	Major Banks	nasdaq
27	27	0001466815	ABCD	Cambium Learning Group, Inc.	Consumer Services	Publishing	nasdaq
28	28	0001157377	ABCO	The Advisory Board Company	Consumer Services	Other Consumer Services	nasdaq
29	29	0001578620	ABDC	Alcentra Capital Corp.	n/a	n/a	nasdaq
30	30	n/a	ABE	Aberdeen Emerging Markets Smaller Company Opportunities Fund I	n/a	n/a	amex
31	31	0000318306	ABEO	Abeona Therapeutics Inc.	Health Care	Major Pharmaceuticals	nasdaq
32	32	0000318306	ABEOW	Abeona Therapeutics Inc.	Health Care	Major Pharmaceuticals	nasdaq
33	33	0001565025	ABEV	Ambev S.A.	Consumer Non-Durables	Beverages (Production/Distribution)	nyse
34	34	0001144980	ABG	Asbury Automotive Group Inc	Consumer Durables	Automotive Aftermarket	nyse
35	35	0001652866	ABIL	Ability Inc.	Finance	Business Services	nasdaq
36	36	0000907654	ABIO	ARCA biopharma, Inc.	Health Care	Biotechnology: In Vitro & In Vivo Diagnostic Substances	nasdaq
37	37	0000771497	ABM	ABM Industries Incorporated	Finance	Diversified Commercial Services	nyse
38	38	0000815094	ABMD	ABIOMED, Inc.	Health Care	Medical/Dental Instruments	nasdaq
39	39	0001253986	ABR	Arbor Realty Trust	Consumer Services	Real Estate Investment Trusts	nyse
40	40	0001253986	ABR^A	Arbor Realty Trust	n/a	n/a	nyse
41	41	0001253986	ABR^B	Arbor Realty Trust	n/a	n/a	nyse
42	42	0001253986	ABR^C	Arbor Realty Trust	n/a	n/a	nyse
43	43	0001253986	ABRN	Arbor Realty Trust	Consumer Services	Real Estate Investment Trusts	nyse
44	44	0000001800	ABT	Abbott Laboratories	Health Care	Major Pharmaceuticals	nyse
45	45	0001023364	ABTL	Autobyte Inc.	Technology	Computer Software: Programming, Data Processing	nasdaq

Ilustración 22: ETL: Tabla stg_companies

7.1.11. Carga de los códigos CIK

En esta tarea, para el conjunto de empresas obtenidos en el paso anterior, obtenemos los códigos CIK necesarios para poder combinar la información obtenida por distintas fuentes.

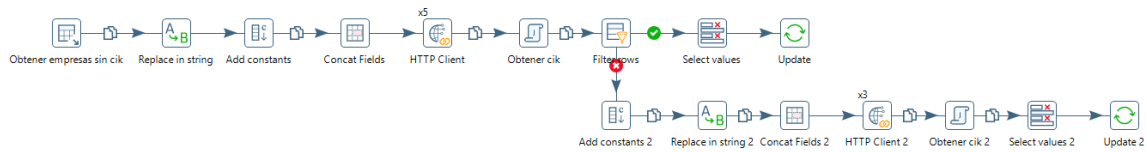


Ilustración 23: Transacción: obtención de códigos CIK

7.1.12. Carga de cotizaciones históricas

Utilizamos la api de Yahoo Finance para obtener las cotizaciones históricas de las empresas que cotizan en los mercados americanos, obtenidas en una transformación anterior.

El resultado de la invocación a la api, es un fichero csv que es almacenado en base de datos antes de ser procesado con el objetivo de solicitar sólo aquella información de la que no se disponga, debido al alto volumen de información descargada. De esta forma también podemos realizar múltiples ejecuciones de la ETL sin sobrecargar el servicio.

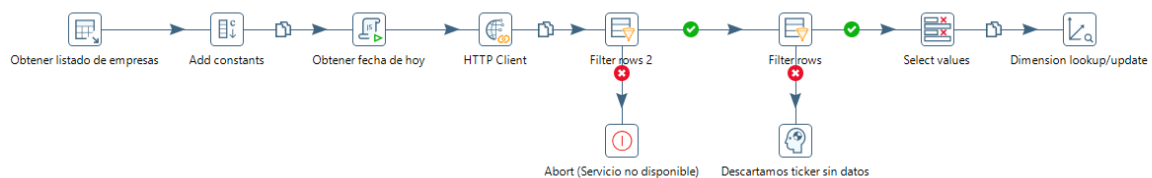


Ilustración 24: Transacción: descarga de las cotizaciones históricas

Resultado obtenido:

stg_quotes Enter a SQL expression to filter results (use Ctrl+Space)									
	id_quote	symbol	date	quotes	version	date_from	date_to		
1	1	AAO	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
2	2	AAOI	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
3	3	AAAP	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
4	4	AAMC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
5	5	AAME	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
6	6	AAON	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
7	7	AAN	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
8	8	AAT	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
9	9	AAXN	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
10	10	AA	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
11	11	AAV	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
12	12	AAWW	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
13	13	AAJ	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
14	14	A	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
15	15	ABDC	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
16	16	ABBV	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
17	17	AAL	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
18	18	AAPC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
19	19	AB	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
20	20	ABCD	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
21	21	ABR	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
22	22	AAU	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
23	23	ABAX	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
24	24	ACBI	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
25	25	ABE	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
26	26	ABCO	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
27	27	AAP	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
28	28	AAPL	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
29	29	ABRN	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
30	30	ABCB	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
31	31	ABG	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
32	32	ACHC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
33	33	ABMD	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
34	34	ABB	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
35	35	ABAC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
36	36	ABUS	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
37	37	ABC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
38	38	ACAD	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
39	39	ABTL	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
40	40	ACC	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
41	41	ACP	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
42	42	ABIO	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
43	43	ABIL	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
44	44	ACIU	[NULL]	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		
45	45	ABM	2017-05-25	Date,Open,High,Low,Close,Volume,Adj Close	1	2017-05-16	2199-12-31		

Ilustración 25: ETL: Tabla stg_quotes

7.1.13. Carga de la dimensión tiempo

Para la carga de la dimensión tiempo, se ha utilizado una transformación estándar disponible en la documentación de Pentaho, a la cual, se han realizado algunos ajustes, como por ejemplo, que las fechas en formato texto se generen según el idioma del sistema o añadir descripciones cortas, medias y largas de la fecha en formato texto.

Esta transformación generará cerca de 40.000 registros que se corresponderán con las fechas desde el 1 de enero de 1950 hasta el 7 de julio de 2059 aproximadamente.

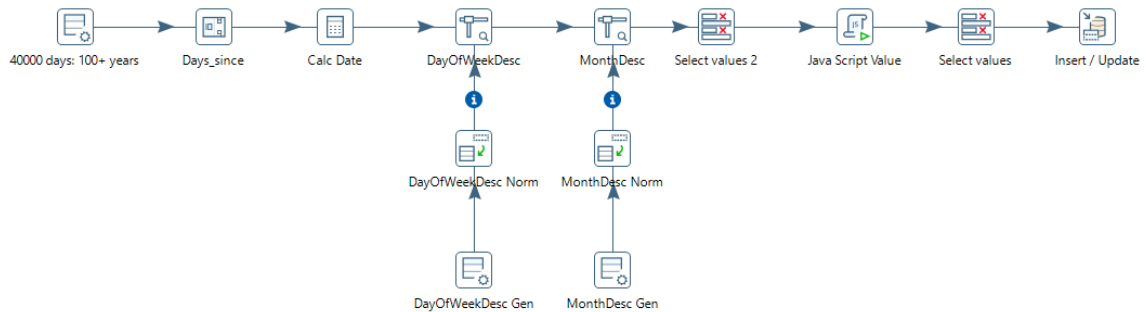


Ilustración 26: Transacción: carga de la dimensión tiempo

Resultado obtenido:

id_time	year_number	month_number	day_of_year_number	day_of_month_number	day_of_week_number	week_of_year_number	day_name	month_name	quarter_number	quarter_name	year_quart
1	19.500.101	1950	1	1	1	52	Domingo	Enero	1	Q1	Q11950
2	19.500.103	1950	3	3	3	1	Martes	Enero	1	Q1	Q11950
3	19.500.104	1950	4	4	4	1	Miércoles	Enero	1	Q1	Q11950
4	19.500.105	1950	5	5	5	1	Jueves	Enero	1	Q1	Q11950
5	19.500.106	1950	6	6	6	1	Viernes	Enero	1	Q1	Q11950
6	19.500.107	1950	7	7	7	1	Sábado	Enero	1	Q1	Q11950
7	19.500.108	1950	8	8	1	1	Domingo	Enero	1	Q1	Q11950
8	19.500.109	1950	9	9	2	2	Lunes	Enero	1	Q1	Q11950
9	19.500.110	1950	10	10	3	2	Martes	Enero	1	Q1	Q11950
10	19.500.111	1950	11	11	4	2	Miércoles	Enero	1	Q1	Q11950
11	19.500.112	1950	12	12	5	2	Jueves	Enero	1	Q1	Q11950
12	19.500.113	1950	13	13	6	2	Viernes	Enero	1	Q1	Q11950
13	19.500.114	1950	14	14	7	2	Sábado	Enero	1	Q1	Q11950
14	19.500.115	1950	15	15	1	2	Domingo	Enero	1	Q1	Q11950
15	19.500.116	1950	16	16	2	3	Lunes	Enero	1	Q1	Q11950
16	19.500.117	1950	17	17	3	3	Martes	Enero	1	Q1	Q11950
17	19.500.118	1950	18	18	4	3	Miércoles	Enero	1	Q1	Q11950
18	19.500.119	1950	19	19	5	3	Jueves	Enero	1	Q1	Q11950
19	19.500.120	1950	20	20	6	3	Viernes	Enero	1	Q1	Q11950
20	19.520.101	1952	1	1	3	1	Martes	Enero	1	Q1	Q11952
21	19.500.121	1950	21	21	7	3	Sábado	Enero	1	Q1	Q11950
22	19.500.122	1950	22	22	1	3	Domingo	Enero	1	Q1	Q11950
23	19.500.123	1950	23	23	2	4	Lunes	Enero	1	Q1	Q11950
24	19.500.124	1950	24	24	3	4	Martes	Enero	1	Q1	Q11950
25	19.500.125	1950	25	25	4	4	Miércoles	Enero	1	Q1	Q11950
26	19.500.126	1950	26	26	5	4	Jueves	Enero	1	Q1	Q11950
27	19.500.127	1950	27	27	6	4	Viernes	Enero	1	Q1	Q11950
28	19.500.128	1950	28	28	7	4	Sábado	Enero	1	Q1	Q11950
29	19.500.129	1950	29	29	1	4	Domingo	Enero	1	Q1	Q11950
30	19.500.130	1950	30	30	2	5	Lunes	Enero	1	Q1	Q11950
31	19.500.131	1950	31	31	3	5	Martes	Enero	1	Q1	Q11950
32	19.500.201	1950	2	32	1	4	Miércoles	Febrero	1	Q1	Q11950
33	19.500.202	1950	2	33	2	5	Jueves	Febrero	1	Q1	Q11950
34	19.500.203	1950	2	34	3	6	Viernes	Febrero	1	Q1	Q11950
35	19.500.204	1950	2	35	4	7	Sábado	Febrero	1	Q1	Q11950
36	19.500.205	1950	2	36	5	1	Domingo	Febrero	1	Q1	Q11950
37	19.500.206	1950	2	37	6	2	Lunes	Febrero	1	Q1	Q11950
38	19.500.207	1950	2	38	7	3	Martes	Febrero	1	Q1	Q11950
39	19.500.208	1950	2	39	8	4	Miércoles	Febrero	1	Q1	Q11950
40	19.500.209	1950	2	40	9	5	Jueves	Febrero	1	Q1	Q11950
41	19.500.210	1950	2	41	10	6	Viernes	Febrero	1	Q1	Q11950
42	19.500.211	1950	2	42	11	7	Sábado	Febrero	1	Q1	Q11950
43	19.500.212	1950	2	43	12	1	Domingo	Febrero	1	Q1	Q11950
44	19.500.213	1950	2	44	13	2	Lunes	Febrero	1	Q1	Q11950

Ilustración 27: ETL: Dimensión tiempo (d_time)

7.1.14. Carga de la dimensión empresa

Para realizar la carga de la dimensión empresa, se leen todos los registros de la tabla stg_companies y se relacionan con la tabla stg_num, del área de stage. Se

completa la información con la obtención de la descripción del país y el estado al que pertenecen.

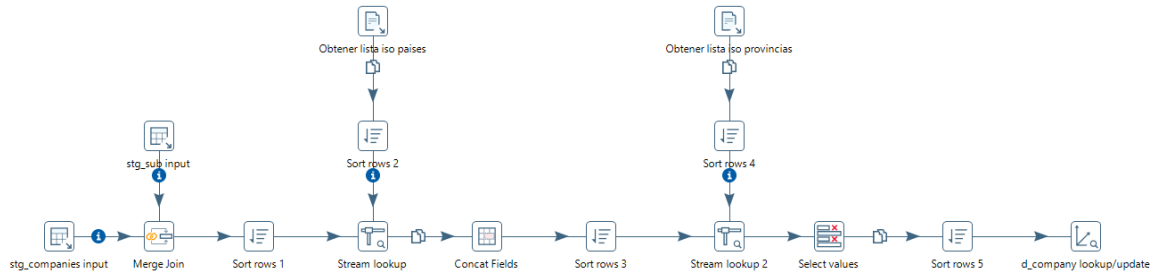


Ilustración 28: Transacción: carga de la dimensión empresa

Resultado obtenido:

d_company	id_company	cik	symbol	name	sector	industry	market	city	country	state
2	1	0001096872	A	Agilent Technologies, Inc.	Capital Goods	Biotechnology: Laboratory Analytical Instruments	nyse	SANTA CLARA	Estados Unidos de América	California
3	2	0001675149	AA	Alcoa Corporation	Capital Goods	Metals Fabrications	nyse	NEW YORK	Estados Unidos de América	New York
4	3	0001626180	AAC	AAC Holdings, Inc.	Health Care	Medical Specialties	nyse	BRENTWOOD	Estados Unidos de América	Tennessee
5	5	0000006201	AAL	American Airlines Group, Inc.	Transportation	Air Freight/Delivery Services	nasdaq	FORT WORTH	Estados Unidos de América	Texas
6	4	0000006201	AAL	American Airlines Group, Inc.	Transportation	Air Freight/Delivery Services	nasdaq	FT WORTH	Estados Unidos de América	Texas
7	7	0001555074	AAMC	Aitsource Asset Management Corp	Finance	Real Estate	amex	CHRISTIANSTED	Islas Virgenes de los Estados Unidos de América	[NULL]
8	6	0001555074	AAMC	Aitsource Asset Management Corp	Finance	Real Estate	amex	FREDERIKSTED	Islas Virgenes de los Estados Unidos de América	[NULL]
9	8	0000008177	AAME	Atlantic American Corporation	Finance	Life Insurance	nasdaq	ATLANTA	Estados Unidos de América	Georgia
10	9	0000706688	AAN	Aaron's, Inc.	Technology	Diversified Commercial Services	nyse	ATLANTA	Estados Unidos de América	Georgia
11	10	0001158114	AAOI	Applied Optoelectronics, Inc.	Technology	Semiconductors	nasdaq	SUGAR LAND	Estados Unidos de América	Texas
12	11	0000824142	AAON	AAON, Inc.	Capital Goods	Industrial Machinery/Components	nasdaq	TULSA	Estados Unidos de América	Oklahoma
13	12	0001158449	AAP	Advance Auto Parts Inc	Consumer Services	Other Specialty Stores	nyse	ROANOKE	Estados Unidos de América	Virginia
14	13	0001630940	AAPC	Atlantic Alliance Partnership Corp.	Consumer Services	Services-Misc. Amusement & Recreation	nasdaq	NEW YORK	Estados Unidos de América	New York
15	14	0000320193	AAPL	Apple Inc.	Technology	Computer Manufacturing	nasdaq	CUPERTINO	Estados Unidos de América	California
16	15	0001500217	AAT	American Assets Trust, Inc.	Consumer Services	Real Estate Investment Trusts	nyse	SAN DIEGO	Estados Unidos de América	California
17	16	0001135185	AAWW	Atlas Air Worldwide Holdings	Transportation	Transportation Services	nasdaq	PURCHASE	Estados Unidos de América	New York
18	17	0001099183	AAVN	Axon Enterprise, Inc.	Capital Goods	Ordinance And Accessories	nasdaq	SCOTTSDALE	Estados Unidos de América	Arizona
19	18	0000825319	AB	AllianceBernstein Holding L.P.	Finance	Investment Managers	nyse	NEW YORK	Estados Unidos de América	New York
20	19	0001486299	ABAC	Aoson Tianji Group, Inc.	Consumer Non-Durables	Farming/Seeds/Milling	nasdaq	WUJIAN CITY, HUBEI PROVINCE	China	[NULL]
21	20	0000881890	ABAX	ABAXIS, Inc.	Capital Goods	Industrial Machinery/Components	nasdaq	UNION CITY	Estados Unidos de América	California
22	21	0001091587	ABB	ABB Ltd	Consumer Durables	Electrical Products	nyse	ZURICH SWITZERLAND	Suiza	[NULL]
23	22	0001551152	ABBV	AbbVie Inc.	Health Care	Major Pharmaceuticals	nyse	NORTH CHICAGO	Estados Unidos de América	Illinois
24	23	0001140859	ABC	AmerisourceBergen Corporation (Holding Co)	Health Care	Other Pharmaceuticals	nyse	CHESTERBROOK	Estados Unidos de América	Pennsylvania
25	24	0000351569	ABCB	Ameris Bancorp	Finance	Major Banks	nasdaq	MOULTRIE	Estados Unidos de América	Georgia
26	25	0001466815	ABCD	Cambium Learning Group, Inc.	Consumer Services	Publishing	nasdaq	DALLAS	Estados Unidos de América	Texas
27	26	0001157377	ABCO	The Advisory Board Company	Consumer Services	Other Consumer Services	nasdaq	WASHINGTON	Estados Unidos de América	District of Colu
28	28	0000318296	ABEOW	Abeona Therapeutics Inc.	Health Care	Major Pharmaceuticals	nasdaq	DALLAS	Estados Unidos de América	Texas
29	27	0000318296	ABEO	Abeona Therapeutics Inc.	Health Care	Major Pharmaceuticals	nasdaq	DALLAS	Estados Unidos de América	Texas
30	29	0001144980	ABG	Asbury Automotive Group Inc	Consumer Durables	Automotive Aftermarket	nyse	DULUTH	Estados Unidos de América	Georgia
31	30	0001652866	ABIL	Ability Inc.	Finance	Business Services	nasdaq	TEL AVIV	Israel	[NULL]
32	32	0000907654	ABIO	ARCA biopharma, Inc.	Health Care	Biotechnology: In Vitro & In Vivo Diagnostic Substances	nasdaq	WESTMINSTER	Estados Unidos de América	Colorado
33	31	0000907654	ABIO	ARCA biopharma, Inc.	Health Care	Biotechnology: In Vitro & In Vivo Diagnostic Substances	nasdaq	BROOMFIELD	Estados Unidos de América	Colorado
34	33	0000771497	ABM	ABM Industries Incorporated	Finance	Diversified Commercial Services	nyse	NEW YORK	Estados Unidos de América	New York
35	34	0000815294	ABMD	ABIMED, Inc.	Health Care	Medical/Dental Instruments	nasdaq	DAWERS	Estados Unidos de América	Massachusetts
36	35	0001253986	ABR	Arbor Realty Trust	Consumer Services	Real Estate Investment Trusts	nyse	UNIONDALE	Estados Unidos de América	New York
37	36	0001253986	ABR^A	Arbor Realty Trust	n/a	n/a	nyse	UNIONDALE	Estados Unidos de América	New York
38	37	0001253986	ABR^B	Arbor Realty Trust	n/a	n/a	nyse	UNIONDALE	Estados Unidos de América	New York
39	39	0001253986	ABRN	Arbor Realty Trust	Consumer Services	Real Estate Investment Trusts	nyse	UNIONDALE	Estados Unidos de América	New York
40	38	0001253986	ABR^C	Arbor Realty Trust	n/a	n/a	nyse	UNIONDALE	Estados Unidos de América	New York
41	40	0000001800	ABT	Abbott Laboratories	Health Care	Major Pharmaceuticals	nyse	ABBOTT PARK	Estados Unidos de América	Illinois
42	41	0001023364	ABTL	Autobytel Inc.	Technology	Computer Software: Programming, Data Processing	nasdaq	IRVINE	Estados Unidos de América	California
43	42	0001640361	ABTX	Alliance Bancshares, Inc.	Finance	Major Banks	nasdaq	HOUSTON	Estados Unidos de América	Texas
44	43	0001447028	ABUS	Arbutus Biopharma Corporation	Health Care	Major Pharmaceuticals	nasdaq	BURNABY	Canada	British Columbi
45	44	0000756894	ABX	Barrick Gold Corporation	Basic Industries	Precious Metals	nyse	TORONTO ONTARIO CANA	Canada	Ontario

Ilustración 29: ETL: Dimension empresa (d_empresa)

7.1.15. Carga de la tabla de hechos financieros fundamentales

La carga de la tabla de hecho que contiene la información financiera de las empresas, parte de la tabla stg_num. Esta tabla contiene toda la información de los distintos estados contables. Para el desarrollo de este TFG, sólo cargaremos

aquellos que se requieren para el cálculo de las medidas que hemos seleccionado implementar (Net Income, Sales, Liabilities, Assets, Equity).

La estructura de la tabla stg_num, consiste en un registro para cada hecho de una empresa, necesitamos transformar esta información para disponer de ella en columnas. Para ello, realizaremos de una desnormalización de la fuente de origen y a continuación con los datos obtenidos realizaremos los cálculos de las medidas y persistiremos en BBDD.



Ilustración 30: Transacción: carga de la tabla de hechos financieros fundamentales

Resultado obtenido:

	id_company	id_time	sales	netincome	liabilities	assets	equity	roa	roe	debratio	debttequityratio
1	116	20.110.630	824.210.000	3.589.000	260.061.000	359.388.000	99.327.000	0,9986421361	3,6133176276	72,3621823767	261,8230692561
2	116	20.110.930	801.690.000	9.026.000	245.234.000	353.587.000	108.353.000	2,5526956591	8,3301800596	69,3560566424	226,3287587792
3	116	20.111.231	[NULL]	[NULL]	268.158.000	378.840.000	110.682.000	[NULL]	[NULL]	70,7839721254	242,2778771616
4	116	20.120.331	877.489.000	6.575.000	295.277.000	412.534.000	117.257.000	1,5938080255	5,6073411395	71,5764033995	251,8203604049
5	116	20.120.630	831.474.000	5.386.000	221.392.000	344.035.000	122.643.000	1,56553839	4,3916081635	64,3515921345	180,5174367881
6	116	20.120.930	795.525.000	8.263.000	257.749.000	388.655.000	130.906.000	2,1260500958	6,3121629261	66,318199946	196,8962461614
7	116	20.121.231	876.476.000	7.567.000	283.643.000	419.501.000	135.858.000	1,8038097645	5,5697861002	67,6143799419	208,7790192701
8	116	20.130.331	953.501.000	8.015.000	273.945.000	417.818.000	143.873.000	1,9182993552	5,5708854337	65,5656290538	190,4075121809
9	116	20.130.630	965.874.000	6.330.000	277.942.000	427.217.000	149.275.000	1,4816826109	4,2404957294	65,05874064	186,1946072685
10	116	20.130.930	1.060.770.000	7.156.000	319.176.000	474.680.000	155.504.000	1,507541923	4,6018108859	67,2402460605	205,2525980039
11	116	20.131.231	968.096.000	109.000	293.397.000	448.082.000	154.685.000	0,0243259046	0,0704657853	65,4784168969	189,6738531148
12	116	20.140.331	949.189.000	5.363.000	305.492.000	464.612.000	159.120.000	1,1542964883	3,3704122675	65,7520683926	191,9884364002
13	116	20.140.630	1.159.931.000	3.975.000	320.977.000	483.144.000	162.167.000	0,8227360787	2,4511768732	66,4350586989	197,9299117576
14	116	20.140.930	1.173.970.000	3.855.000	290.806.000	455.900.000	165.094.000	0,8455801711	2,3350333749	63,7872340426	176,1457109283
15	116	20.141.231	849.736.000	-6.670.000	183.317.000	340.814.000	157.497.000	-1,957079228	-4,2350013016	53,78798993	116,3939630596
16	116	20.150.331	555.573.000	3.097.000	159.857.000	319.524.000	159.667.000	0,9692542657	1,9396619214	50,0297317259	100,1189976639
17	116	20.150.630	600.558.000	4.340.000	176.494.000	339.572.000	163.078.000	1,2780794647	2,6613031801	51,9754278916	108,2267381253
18	116	20.150.930	439.893.000	-308.000	126.959.000	288.801.000	161.842.000	-0,1066478302	-0,1903090669	43,9607203576	78,4462624041
19	116	20.151.231	348.255.000	-8.404.000	90.705.000	243.215.000	152.510.000	-3,4553789857	-5,5104583306	37,2941636001	59,4747885385
20	116	20.160.331	250.531.000	1.430.000	84.903.000	237.915.000	153.012.000	0,6010549986	0,9345672235	35,6862745098	55,487804878
21	116	20.160.630	293.163.000	3.404.000	96.038.000	251.526.000	155.488.000	1,3533392174	2,1892364684	38,1821362404	61,7655381766
22	116	20.160.930	256.877.000	-2.153.000	80.504.000	232.911.000	152.407.000	-0,924387427	-1,4126647726	34,5642756246	52,8217207871
23	116	20.161.231	256.877.000	1.430.000	95.560.000	246.872.000	151.312.000	0,5792475453	0,945067146	38,7083184808	63,1542772549
24	286	20.150.930	1.061.000.000	-197.000.000	3.565.000.000	3.229.000.000	-336.000.000	-6,1009600496	58,630952381	110,4056983586	-1.061,0119047619
25	286	20.151.231	[NULL]	-102.000.000	3.521.000.000	3.109.000.000	-412.000.000	-3,2807976841	24,7572815534	113,2518494693	-854,6116504854
26	286	20.160.331	832.000.000	-109.000.000	3.484.000.000	2.981.000.000	-503.000.000	-3,6564911104	21,6699801193	116,8735323717	-692,6441351889
27	286	20.160.630	1.027.000.000	69.000.000	3.729.000.000	3.316.000.000	-413.000.000	2,0808202654	-16,7070217918	112,4547647768	-902,9055690073
28	286	20.160.930	1.307.000.000	-406.000.000	3.231.000.000	3.616.000.000	385.000.000	-11,2278761062	-105,4545454545	89,3528761062	839,2207792208
29	286	20.161.231	[NULL]	-3.000.000	2.905.000.000	3.321.000.000	416.000.000	-0,0903342367	-0,7211538462	87,4736525143	698,3173076923
30	358	20.161.231	1.882.500.000	299.800.000	8.794.700.000	15.956.200.000	7.161.500.000	1,8788934709	4,1862738253	55,1177598676	122,8052782238
31	3.185	20.100.630	398.900.000	28.900.000	1.360.000.000	2.461.000.000	1.101.000.000	1,1743193824	2,6248864668	55,2620885819	123,5240609282
32	3.185	20.100.930	444.300.000	25.700.000	1.309.000.000	2.441.000.000	1.132.000.000	1,0528471938	2,2703180212	53,6255632937	115,6360424028
33	3.185	20.110.331	405.600.000	5.200.000	1.381.000.000	2.516.000.000	1.135.000.000	0,2066772655	0,4581497797	54,8887122417	121,6740088106
34	3.185	20.110.630	488.200.000	18.700.000	1.423.000.000	2.556.000.000	1.133.000.000	0,7316118936	1,6504854369	55,6792964476	125,5957634598
35	3.185	20.110.930	440.200.000	8.700.000	1.414.000.000	2.559.000.000	1.145.000.000	0,3399765533	0,7598253275	55,2559593591	123,4934497817
36	3.185	20.111.231	460.200.000	1.600.000	520.000.000	1.643.000.000	1.123.000.000	0,0973828363	0,142475512	31,6494217894	46,3045414069
37	3.185	20.120.331	405.000.000	3.800.000	1.447.000.000	2.574.000.000	1.127.000.000	0,1476301476	0,3371783496	56,216006216	128,3939662822
38	242	20.120.630	72.400.000	-4.400.000	543.600.000	1.424.900.000	881.300.000	-0,3087935996	-0,4992624532	38,1500456172	61,6816067173
39	272	20.110.630	62.036.000	20.157.000	1.423.593.000	1.771.975.000	348.382.000	1,1375442656	5,785890201	80,3393388733	408,6298947707
40	272	20.111.231	64.607.000	20.634.000	1.412.507.000	1.771.307.000	358.800.000	1,1649025268	5,7508361204	79,7437711249	393,6753065775
41	272	20.120.331	63.660.000	19.482.000	1.413.859.000	1.772.990.000	359.131.000	1,0988217644	5,424761438	79,7443301993	393,6889324508
42	272	20.120.930	48.642.000	18.856.000	1.406.605.000	1.765.477.000	358.872.000	1,0680399688	5,2542410665	79,6728023078	391,9517265209
43	272	20.121.231	48.491.000	617.157.000	1.149.657.000	1.481.810.000	332.153.000	41,6488618649	185,8050356312	77,5846431054	346,1227205535
44	272	20.130.331	48.775.000	14.162.000	1.149.741.000	1.485.263.000	335.522.000	0,9535011644	4,2208856647	77,4099267268	342,672313589
45	272	20.130.630	47.302.000	13.139.000	1.136.389.000	1.469.572.000	333.183.000	0,894069838	3,9434785088	77,3278886642	341,0705228058

Ilustración 31: ETL: Tabla hechos fundamentales (f_fundamental)

7.1.16. Carga de la tabla de hechos técnicos

Para la carga de la tabla de hechos técnico se ha optado por crear un proceso en varias fases que permita la ejecución en paralelo, para acelerar la carga de la ETL.

Para ello, en la primera etapa se leen los datos a tratar del área de stage (stg_quotes) y se lanza la ejecución en paralelo.

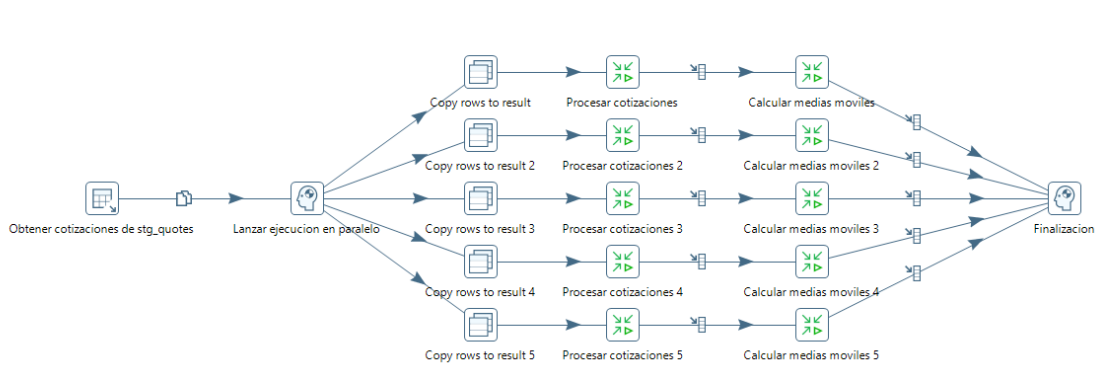


Ilustración 32: Transacción: carga de la tabla de hechos técnicos I

La siguiente etapa, vuelca a un fichero temporal el resultado de la llamada a la api de Yahoo Finance, realizada en la carga del área stage, y almacena los datos en la tabla de hechos técnicos. El proceso se ha realizado de esta forma pues se ha comprobado que resulta más eficiente, que otras soluciones probadas.

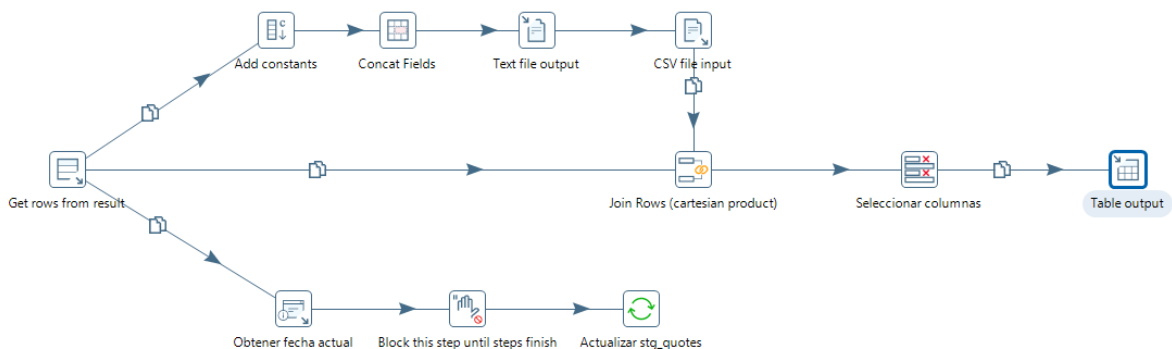


Ilustración 33: Transacción: carga de la tabla de hechos técnicos II

Una vez cargados los hechos técnicos, se procede al cálculo de las medidas seleccionadas, para ello y con el objetivo nuevamente de acelerar el proceso, se precarga en la una tabla del área de stage unos índices que se utilizarán posteriormente para lanzar unas sentencias sql que calculen las medidas.

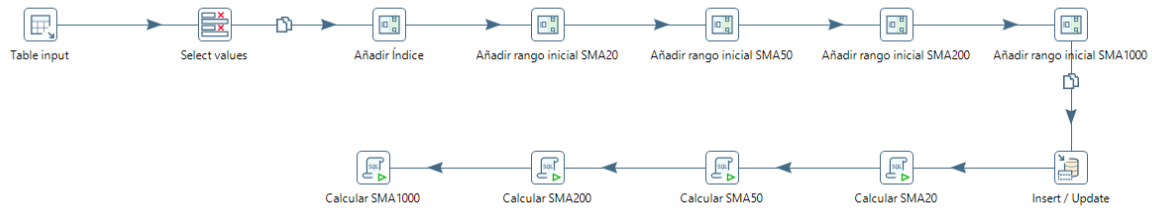


Ilustración 34: Transacción: carga de la tabla de hechos técnicos III

A continuación, se muestra a modo de ejemplo, la consulta sql de actualización que se ejecuta para el cálculo de la media móvil de 20 sesiones, dónde se puede observar la utilización de la tabla stg_sma para agilizar la tarea, pues se delega el cálculo en la propia BBDD.

```

update f_tecnico ft
set sma20 = (
  select
    (case when sma.sma20 < 1 then 0
     else (
       select sum(q.close)/20 q
       from stg_sma q
       where q.indice >= sma.sma20 and q.indice <= sma.indice
       and q.id_company = sma.id_company
       and q.id_time <= sma.id_time)
     end)
  from stg_sma sma
  where sma.id_company = ft.id_company
  and sma.id_time = ft.id_time
)
where ft.id_company = ${id_company};
  
```

Resultado obtenido:

f_tecnico | Enter a SQL expression to filter results (use Ctrl+Space)

	id_company	id_time	change	open	high	low	close	volume	adjusted_close	sma20	sma50	sma200	sma1000
1	989	20.120.606	0,08	2,18	2,34	2,17	2,21	8.900	11,05	2,4125	3,0616	3,1769	3,847200483
2	989	20.120.607	0	2,25	2,37	2,21	2,21	16.400	11,05	2,346	3,0222	3,1702	3,836720229
3	989	20.120.608	0,06	2,24	2,28	2,16	2,27	4.100	11,350001	2,318	2,9838	3,16315	3,82702999
4	989	20.120.611	-0,12	2,31	2,44	2,15	2,15	19.900	10,750001	2,297	2,9456	3,15555	3,817699761
5	989	20.120.612	0,15	2,17	2,39	2,17	2,3	10.200	11,5	2,2805	2,911	3,14815	3,808379529
6	989	20.120.613	-0,03	2,26	2,33	2,24	2,27	4.700	11,350001	2,2655	2,7338	3,14065	3,798799293
7	989	20.120.614	0,09	2,21	2,48	2,21	2,36	7.400	11,799999	2,2685	2,8414	3,1317	3,788659044
8	989	20.120.615	0,18	2,4	2,6	2,23	2,54	34.700	12,7	2,2945	2,8156	3,1232	3,779198805
9	989	20.120.618	-0,06	2,51	2,58	2,43	2,48	11.000	12,4	2,313	2,7876	3,11485	3,76986857
10	989	20.120.619	-0,09	2,54	2,59	2,3	2,39	9.700	11,950001	2,3255	2,762	3,10735	3,760008326
11	989	20.120.620	-0,16	2,41	2,43	2,21	2,23	15.400	11,15	2,33	2,7352	3,09905	3,750238087
12	989	20.120.621	0,02	2,22	2,28	2,21	2,25	5.100	11,25	2,3285	2,7108	3,09155	3,741537869
13	989	20.120.622	-0,06	2,26	2,26	2,17	2,19	18.100	10,950001	2,3175	2,6874	3,08345	3,733127658
14	989	20.120.625	0,1	2,16	2,3	2,15	2,29	8.600	11,45	2,3005	2,6664	3,0759	3,724817447
15	989	20.120.626	-0,07	2,29	2,29	2,2	2,22	1.200	11,100001	2,292	2,6454	3,0683	3,71613723
16	989	20.120.627	0,02	2,23	2,35	2,2	2,24	13.600	11,2	2,2805	2,6276	3,061	3,707477013
17	989	20.120.628	-0,14	2,18	2,27	2,08	2,1	18.300	10,499999	2,264	2,607	3,0525	3,697576774
18	989	20.120.629	-0,11	2,16	2,24	1,93	1,99	71.100	9,95	2,249	2,5848	3,0434	3,68781654
19	989	20.120.702	0,11	2	2,1	2	2,1	7.200	10,499999	2,246	2,5626	3,03335	3,678166306
20	989	20.120.703	0,03	2,15	2,16	2,12	2,13	4.000	10,650001	2,246	2,5374	3,02725	3,669656093
21	989	20.120.705	0,03	2,14	2,19	2,11	2,16	6.000	10,8	2,2435	2,5178	3,0198	3,661835894
22	989	20.120.706	0,06	2,11	2,22	2,08	2,22	8.300	11,100001	2,244	2,4986	3,0134	3,6543057
23	989	20.120.709	0	2,23	2,24	2,15	2,22	5.100	11,100001	2,2415	2,4754	3,00855	3,647025511
24	989	20.120.710	-0,03	2,19	2,27	2,19	2,19	5.900	10,950001	2,2435	2,4524	3,0054	3,640065329
25	989	20.120.711	-0,11	2,18	2,23	2,08	2,08	10.200	10,4	2,2325	2,427	3,0018	3,633075149
26	989	20.120.712	-0,01	2,12	2,13	2,07	2,07	4.900	10,35	2,2225	2,404	2,99815	3,626144969
27	989	20.120.713	-0,03	2,11	2,23	2,04	2,04	12.000	10,2	2,2065	2,3784	2,99455	3,619644798
28	989	20.120.716	0,05	2,08	2,12	2,05	2,09	2.200	10,45	2,184	2,3546	2,99185	3,613724638
29	989	20.120.717	0,03	2,1	2,17	2,09	2,12	5.600	10,6	2,166	2,334	2,99045	3,607244466
30	989	20.120.718	0	2,15	2,15	2,1	2,12	1.800	10,6	2,1525	2,311	2,98945	3,600764294
31	989	20.120.719	-0,02	2,11	2,16	2,08	2,1	3.800	10,499999	2,146	2,2874	2,9897	3,594774133
32	989	20.120.720	0,02	2,1	2,13	2,09	2,12	3.200	10,6	2,1395	2,259	2,99045	3,58823972
33	989	20.120.723	-0,08	2,09	2,1	2,04	2,04	6.300	10,2	2,132	2,2432	2,99025	3,582573806
34	989	20.120.724	0,08	2,09	2,14	2,08	2,12	6.700	10,6	2,1235	2,242	2,9885	3,576533643
35	989	20.120.725	-0,07	2,13	2,13	2,04	2,05	2.700	10,25	2,115	2,2226	2,9821	3,568773448
36	989	20.120.726	0,05	2,06	2,12	2,03	2,1	4.800	10,499999	2,108	2,2132	2,97775	3,561883269
37	989	20.120.727	0	2,12	2,12	2,05	2,1	4.200	10,499999	2,108	2,2092	2,96895	3,555613102
38	989	20.120.730	0,42	2,76	3,24	2,49	2,52	650.500	12,600001	2,1345	2,2192	2,9635	3,549792936
39	989	20.120.731	-0,22	2,53	2,62	2,29	2,3	60.000	11,5	2,1445	2,223	2,9556	3,543872772
40	989	20.120.801	-0,08	2,33	2,4	2,19	2,22	36.200	11,100001	2,149	2,2246	2,94805	3,538672624
41	989	20.120.802	0,03	2,22	2,37	2,21	2,25	31.400	11,25	2,1535	2,2268	2,9413	3,532562457
42	989	20.120.803	0	2,27	2,32	2,25	2,25	12.500	11,25	2,155	2,2262	2,93595	3,526312287
43	989	20.120.806	0,03	2,26	2,37	2,26	2,28	12.100	11,4	2,158	2,2236	2,93185	3,520602128
44	989	20.120.807	0,14	2,28	2,54	2,28	2,42	36.400	12,100001	2,1695	2,2194	2,9299	3,514791964
45	989	20.120.808	0,02	2,45	2,53	2,37	2,44	14.100	12,200001	2,1875	2,2204	2,9246	3,508911798

Ilustración 35: ETL: Tabla hechos técnicos (f_tecnico)

7.2. Resultados obtenidos en la carga de la ETL

En este apartado resumiremos brevemente el resultado de la ejecución de la ETL realizada en mayo de 2017.

A continuación se muestra una tabla con el número de registros creados para cada una de las tablas del modelo físico de datos.

Tabla	Número de registros
stg_companies	6.716
stg_data	128
stg_num	10.222.270
stg_quotes	6.014
stg_sub	178.467
d_company	5.717

d_time	40.000
f_fundamental	85.539
f_tecnico	20.974.691

Tabla 18: Resultado de la ejecución de la ETL

Desde la web del Nasdaq, se ha obtenido 6,716 compañías que se corresponden a empresas que cotizan en las tres grandes bolsas del mercado americano: Nasdaq, Nyse y Amex.

Por otro lado, desde la web de la Securities Exchange Commission (SEC) se ha descargado información desde el segundo trimestre del año 2009 hasta el primer trimestre de 2017, ambos inclusive. Esta información ha generado 10 millones de registros en la tabla stg_num, que contiene los datos financieros de los estados contables de las distintas empresas, y que se corresponden con los 178.000 registros presentados por las mismas (tabla stg_sub).

Como se ha indicado, los registros de la SEC contienen información de todas las compañías que tienen obligación legal de presentar sus resultados ante la SEC. Esta información incluye más compañías de las que cotizan en los mercados, pues se incluye información de fondos de inversión y otras empresas no cotizadas. Al cruzar esta información, con la obtenida de Nasdaq, podemos ver que la dimensión empresa (d_company) contiene información de 5.717 empresas.

También, se han obtenido las cotizaciones históricas de 6.014 empresas (stg_quotes), que han generado cerca de 21 millones de registros en la tabla de hechos técnicos (f_tecnico).

Por último, la carga de la tabla de hechos fundamentales ha generado más de 85.000 registros. Cada registro contiene la información seleccionada para una empresa en un trimestre determinado.

7.3. Análisis OLAP

Una vez realizada la primera carga con la ETL, ya dispone el DW de datos suficientes como para poder realizar análisis sobre los mismos. Una de las opciones naturales en un BI, es la creación de cubos OLAP que permitan mediante las herramientas de análisis correspondientes, estudiar del conjunto de datos para dar respuesta a los problemas de negocio planteados. Como primer paso para poder

realizar estos análisis es necesario diseñar los cubos sobre los cuales se van a realizar las preguntas.

Pentaho utiliza Mondrian como motor OLAP (Online Analytical Processing server), que utiliza unos ficheros xml, con una estructura determinada, para definir los cubos. Para la definición de los cubos en este TFG utilizaremos una herramienta de Pentaho, denominada Schema Workbench, que facilita la tarea, pues dispone de un sistema de validación del cubo enormemente útil.

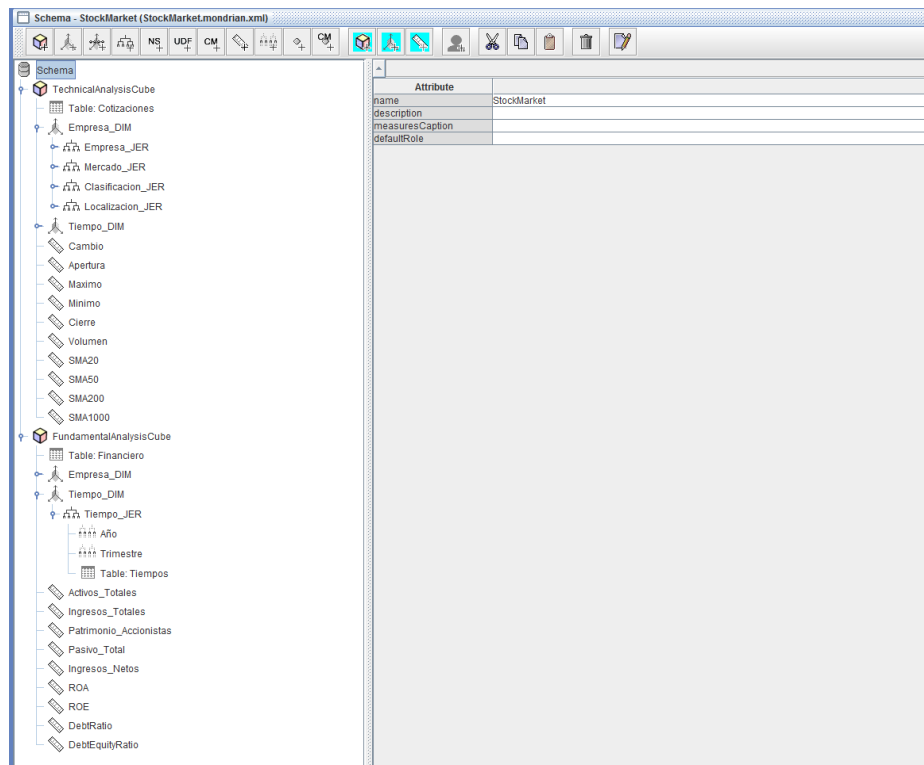


Ilustración 36: Schema Workbench

7.3.1. Definición de cubos OLAP

Se han definido dos cubos OLAP, que se han denominado Technical Analysis Cube y Fundamental Analysis Cube.

Technical Analysis Cube

Este cubo está asociado a dos dimensiones: la dimensión empresa y la dimensión tiempo. Para la dimensión empresa se han creado cuatro jerarquías: empresa, mercado, clasificación y localización.

La jerarquía empresa sólo contiene un nivel que se corresponde con el nombre de la empresa, al igual que la jerarquía mercado que también contiene un nivel, bolsa o mercado que indica el mercado en el que cotiza la empresa.

La jerarquía clasificación, está formada por dos niveles, el nivel sector y el nivel industria, que indica el sector y la industria a la que pertenece la empresa. La jerarquía localización también tiene dos niveles país y estado de la empresa. La dimensión tiempo, se ha definido con una sola jerarquía con tres niveles: año, mes y día.

Por último, se han definido las medidas: apertura, máximo y mínimo, cierre, volumen, cambio y las medias móviles de 20, 50, 200 y 1000 sesiones.

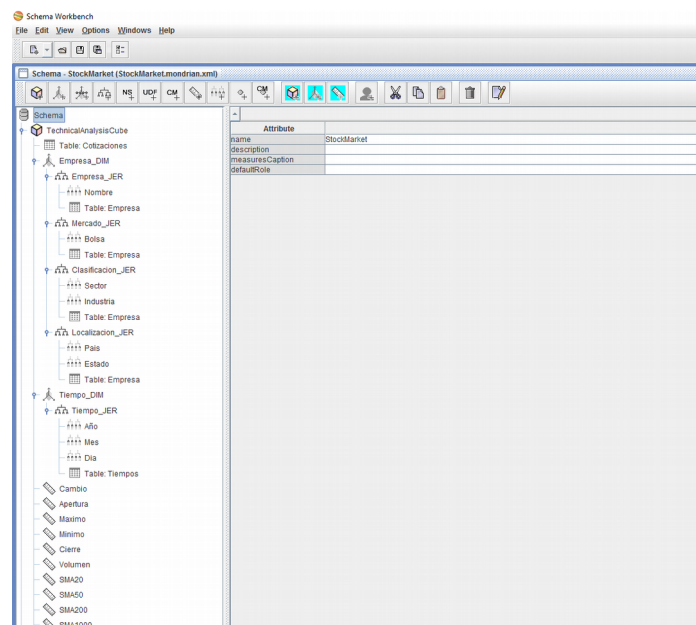


Ilustración 37: Cubo OLAP: Captura Schema Workbench I

Fundamental Analysis Cube

Al igual que el anterior, este cubo está asociado a dos dimensiones: la dimensión empresa y la dimensión tiempo. La dimensión empresa es igual a la descrita

anteriormente, la dimensión tiempo es similar, pero han definido dos niveles en lugar de tres: año y trimestre.

Las medidas definidas en este cubo son: activos, ingresos, patrimonio, pasivo, ingresos netos y los ratios financieros: ROA, ROE, Debt y Debt to equity.

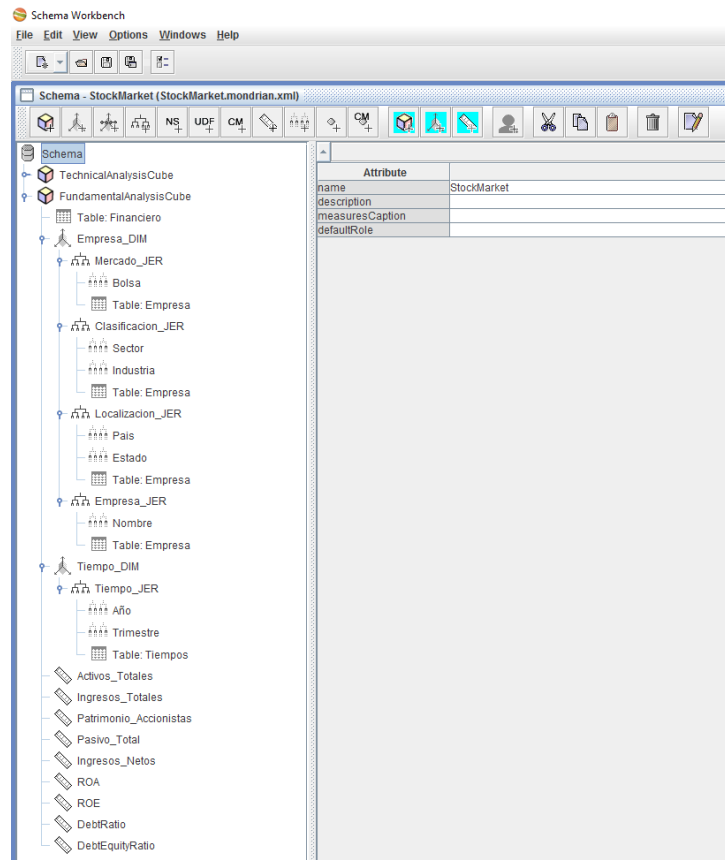


Ilustración 38: Cubo OLAP: Captura Schema Workbench II

Una vez definidos los cubos podemos cargarlos en servidor BI y realizar análisis con herramientas como JPivot o Saiku analytics.

7.3.2. Saiku Analytics

Este apartado tiene el objetivo de mostrar de forma ilustrativa, el uso de la herramienta de análisis Saiku Analytics para realizar un estudio sobre los datos del DW.

Partiremos del supuesto que somos inversores que estamos buscando ideas de inversión y para ello, vamos a generar un informe de ingresos netos (Net Income) o beneficio neto del conjunto de sectores de la economía para los últimos cinco años.

Para obtener esta información, vamos a poner en columnas el nivel año, de la jerarquía tiempo que pertenece a la dimensión tiempo y en filas, el nivel sector de la jerarquía clasificación que pertenece a la dimensión empresa y seleccionamos como medida los ingresos netos (Net Income), obteniendo el siguiente resultado:













Año	2012	2013	2014	2015	2016	
Sector	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	
Basic Industries	69.497.829,512	72.900.902,659	53.826.420,212	31.237.979,452	54.593.324,527	
Capital Goods	132.838.605,811	134.266.960,066	211.256.362,666	67.172.127,57	62.113.205,079	
Consumer Durables	25.125.545,615	26.597.365,882	28.880.905,412	27.879.111,555	33.532.661,453	
Consumer Non-Durables	98.370.024,338	100.461.188,765	98.845.968,172	95.438.704,788	104.863.654,056	
Consumer Services	56.589.724,012	62.510.792,956	60.122.331,357	52.548.432,881	63.584.505,56	
Energy	202.108.737,593	132.988.483,579	128.565.957,357	-129.273.952,786	-40.775.246,395	
Finance	58.308.518,423	103.331.123,365	107.998.212,633	90.206.679,346	108.832.416,606	
Health Care	55.216.795,606	61.150.700,414	51.763.755,083	51.564.393,722	43.185.434,553	
Miscellaneous	41.158.966,393	52.921.350,035	49.722.055,7	47.601.601,002	69.237.005,658	
Public Utilities	62.193.422,808	99.292.363,717	78.442.371,436	87.139.803,38	83.954.889,481	
Technology	92.519.667,656	99.138.705,973	100.136.698,212	84.239.487,483	88.839.384,608	
Transportation	58.482.170,549	131.181.810,526	97.458.783,261	169.873.863,561	127.582.476,474	

Ilustración 39: Saiku Analytics: Sectores de la economía

Como se puede observar a simple vista sectores como el Basic Industries (Materias primas), Capital Goods (Fabricación y distribución de bienes de consumo), Healthcare (Sanitario), Energy (Energético), Technology (Tecnológico) en los últimos han tenido un comportamiento negativo, pues sus beneficios muestran una tendencia descendente. En cambio, sectores como Consumer Durables (Bienes de consumo con una vida de tres o más años), Consumer Non-Durables (Bienes de consumo de uso diario), Consumer services (Servicios al consumidor), Finance (Financiero), Public Utilities (Suministros) y Transportation (Transportes) han tenido un buen comportamiento presentando una línea ascendente en los beneficios.

Supongamos que creemos que el sector sanitario (Healthcare) tendrá en el futuro un buen comportamiento y queremos conocer el comportamiento de las distintas industrias que pertenecen a ese sector.

Para obtener dicha información, vamos a añadir a la anterior consulta en las filas el nivel industria que también pertenece a la jerarquía clasificación de la dimensión

empresa y aplicaremos un filtro sobre el nivel sector, para sólo mostrar el sector sanitario, obteniendo el siguiente resultado:















		Año	2012	2013	2014	2015	2016	
Sector	Industria	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	
Health Care	Biotechnology: Biological Products (No Diagnostic Substances)	52.680.804,562	56.869.007,751	98.669.458,102	113.327.502,732	80.133.173,164		
	Biotechnology: Commercial Physical & Biological Research	-1.447.654,648	803.503,295	1.311.053,077	6.963.135,822	2.467.225,122		
	Biotechnology: Electromedical & Electrotherapeutic Apparatus	5.591.923,23	5.750.259,27	2.231.688,222	18.598.591,878	55.950.611,795		
	Biotechnology: In Vitro & In Vivo Diagnostic Substances	4.954.379,457	4.641.524,333	-282.472,966	837.089,317	-374.028,701		
	Hospital/Nursing Management	17.132.540,244	16.462.355,784	9.538.601,867	6.317.152,011	5.924.807,895		
	Industrial Specialties	20.229.247,548	19.473.333,672	22.742.581,738	10.390.351,497	12.688.389,291		
	Major Pharmaceuticals	65.012.755,204	74.118.580,454	43.724.639,253	36.940.003,866	21.939.004,858		
	Medical Electronics	-22.832.000	-295.326,750	12.983.750	46.825.000	66.082.000		
	Medical Specialties	171.783.053,048	158.555.400,188	122.035.626,974	116.347.344,831	118.222.471,194		
	Medical/Dental Instruments	26.654.297,695	37.902.623,496	38.212.984,281	27.579.704,477	44.319.270,418		
	Medical/Nursing Services	106.566.108,696	142.701.304,348	165.524.431,373	255.207.833,333	202.166.225,806		
	Ophthalmic Goods	25.671.571,429	42.489.285,714	37.714.714,286	26.858.666,667	32.725.000		
	Other Pharmaceuticals	174.195.625	114.873.666,667	141.514.913,043	176.464.916,667	217.295.173,913		
	Precision Instruments	-489.333,333	-47.285,714	-550.500	-5.828.358	-3.976.934,429		

Ilustración 40: Saiku Analytics: Detalle sector sanitario

Siguiendo el supuesto vamos a descartar las industrias relacionadas con la biotecnología, por considerarlas de alto riesgo. También descartamos las industrias con un comportamiento negativo en los últimos años y aquellas que muestran elevada volatilidad en sus beneficios. Estas simples reglas nos dejan las siguientes industrias a seguir estudiando: Medical/Nursing Services, Other Pharmaceuticals.

Supongamos que nos interesa explorar las empresas de la industria Other Pharmaceuticals, para obtener una visión de las empresas que pertenecen a esta industria, nos basamos en el informe anterior añadiendo en las filas el nivel empresa de la jerarquía empresa y filtramos por la industria, resultando:

			Año	2012	2013	2014	2015	2016	
Sector	Industria	Nombre	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	Ingresos_Netos	
Health Care	Other Pharmaceuticals	Aceto Corporation	4.673.250	7.771.250	5.586.500	9.903.750	5.254.750		
		AmerisourceBergen Corporation (Holding Co)	181.370.250	76.620.250	8.788.000	98.859.250	336.717.250		
		Cardinal Health, Inc.	285.850.000	94.000.000	276.000.000	342.250.000	338.000.000		
		Herbalife LTD.	117.877.250	131.871.250	63.058.666,667	84.775.000	65.000.000		
		McKesson Corporation	400.000.000	287.750.000	428.750.000	489.750.000	478.250.000		
		Nu Skin Enterprises, Inc.	55.403.000	91.229.250	47.292.250	33.251.500	34.967.000		

Ilustración 41: Saiku Analytics: Detalle industria: Otras farmacéuticas

Podemos observar que tres de las empresas sobresalen del resto: AmerisourceBergen Corporation, Cardinal Health Corporation y McKesson Corporation. El negocio de estas tres empresas es la distribución de fármacos y material médico y son las principales empresas de esta industria.

Vamos a incluir algunas medidas a la consulta (Ingresos Brutos, Ingresos Netos, Activos, Pasivo) y transponer filas por columnas para mostrar en las filas los años y en columnas las empresas y sus medidas correspondientes:

Empresa_JER - Nombre	Cardinal Health, Inc.				AmerisourceBergen Corporation (Holding Co)				McKesson Corporation			
	Ingresos_Totales	Ingresos_Netos	Activos_Totales	Pasivo_Total	Ingresos_Totales	Ingresos_Netos	Activos_Totales	Pasivo_Total	Ingresos_Totales	Ingresos_Netos	Activos_Totales	Pasivo_Total
2009	24.850.200.000	98.150.000	20.628.100.000	15.544.450.000	18.864.879.000	136.057.000	13.288.254.000	10.541.038.000	27.353.000.000	305.000.000	26.626.666.666,667	19.922.333.333,333
2010	24.652.925.000	239.025.000	21.270.650.000	15.946.300.000	19.508.000.000	161.485.250	6.428.713.250	3.508.859.250	27.743.666.666,667	399.333.333,333	28.194.000.000	21.117.500.000
2011	26.676.225.000	236.875.000	23.372.225.000	17.585.475.000	20.407.670.000	177.060.000	4.167.052.000	1.140.160.250	29.972.000.000	294.000.000	31.741.750.000	24.564.500.000
2012	26.200.625.000	285.850.000	24.591.025.000	18.264.250.000	20.100.047.000	181.370.250	12.788.540.000	10.119.473.250	30.883.500.000	400.000.000	32.810.500.000	25.461.000.000
2013	24.193.750.000	94.000.000	25.341.750.000	18.919.000.000	24.018.929.500	76.620.250	17.851.374.000	15.518.368.750	32.522.000.000	287.750.000	35.675.750.000	28.052.250.000
2014	23.482.000.000	276.000.000	25.923.250.000	19.601.000.000	30.995.341.750	8.788.000	21.857.238.000	19.892.959.250	43.490.500.000	428.750.000	53.026.750.000	44.147.750.000
2015	28.105.500.000	342.250.000	30.689.750.000	24.229.500.000	34.770.681.750	98.859.250	28.207.748.250	26.851.393.750	47.282.750.000	489.750.000	55.175.250.000	48.545.750.000
2016	31.808.750.000	338.000.000	34.401.000.000	27.875.500.000	37.077.476.250	336.717.250	32.996.267.500	30.917.157.500	49.124.500.000	478.250.000	57.664.000.000	48.835.500.000

Ilustración 42: Saiku Analytics: Información financiera ABC, CAH, MCK

Hemos visto, como mediante la utilización de la herramienta Saiku Analytics, hemos podido generar distintos informes que pueden ayudar a un inversor a responder a cuestiones que se le pueden plantear y que muestran la potencia de las herramientas de análisis aplicadas al DW/BI y en concreto al conjunto de datos obtenido en este TFG.

En el siguiente apartado, mostramos la implementación realizada de la herramienta de análisis para la búsqueda de empresas que cumplan ciertas condiciones, que se ha denominado “Screener de acciones”.

7.4. Herramienta de análisis: Screener de acciones

El objetivo de este TFG es la creación de una herramienta de análisis que permita la búsqueda de ideas de inversión, mediante la búsqueda de empresas que cumplan ciertos criterios.

Como hemos visto en el apartado 4.1 y 4.2 existen muchos indicadores del análisis técnico y ratios financieros que nos pueden ayudar en la tarea planteada. Debido a su elevado número, hemos realizado una selección de algunos para realizar su implementación en la herramienta propuesta.

7.4.1. Diseño

Uno de los objetivos de la herramienta propuesta, consiste en que sea sencilla de utilizar. Para ello, se ha diseñado una interfaz acorde, donde prima la sencillez y la facilidad de uso.

Se ha utilizado una herramienta para la realización de un mockup que muestre el diseño generado:

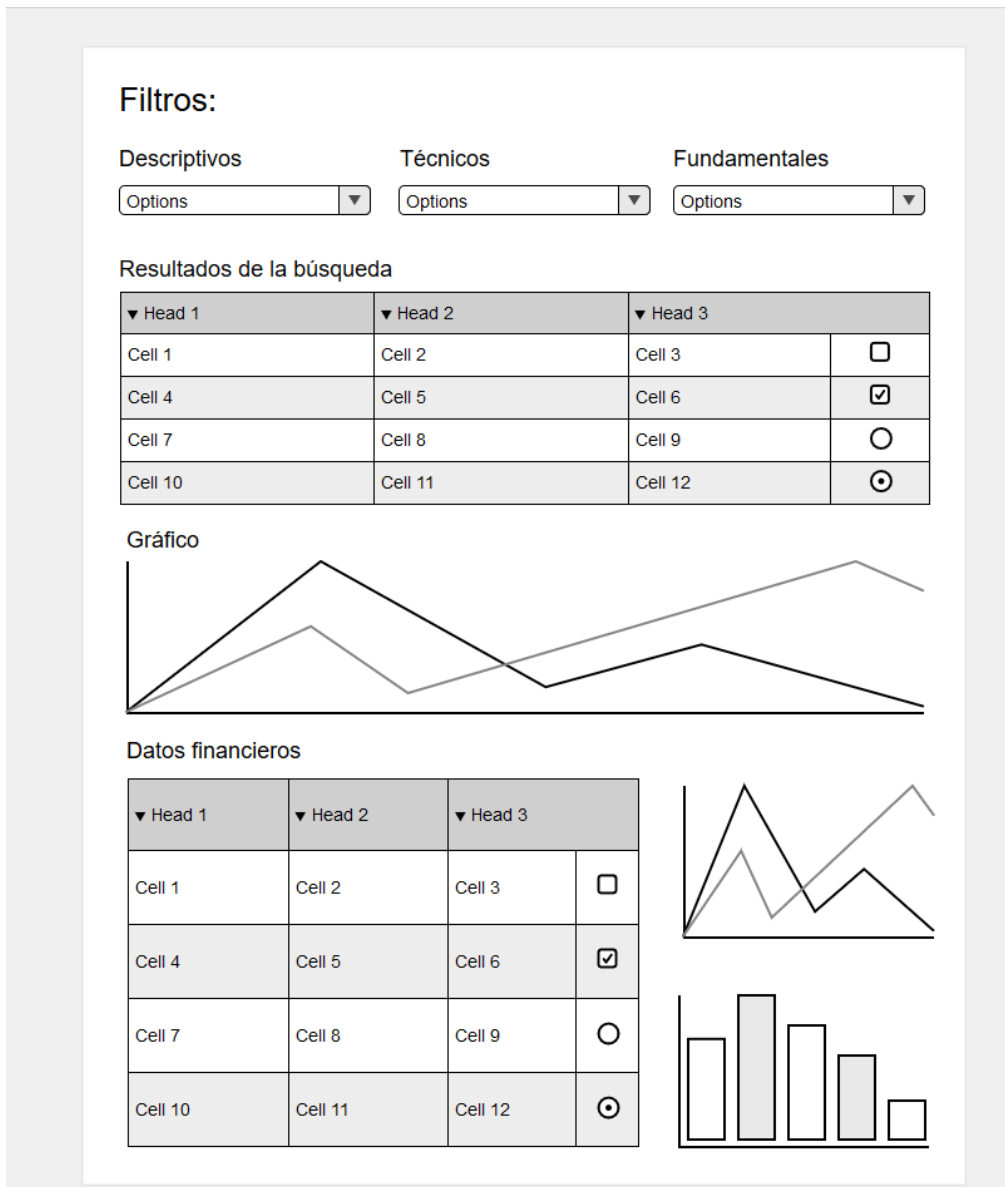


Ilustración 43: Screener: Mockup

En la parte superior se encuentran las opciones de filtrado que se podrán aplicar sobre el conjunto de datos. Los filtros estarán agrupados en tres columnas, una para los filtros que se aplican sobre los datos generales de la empresa (descriptivos), una para los filtros creados para los indicadores técnicos (técnicos) y una última, para los filtros creados para los ratios financieros (fundamentales).

En la parte central encontramos una tabla que mostrará los resultados obtenidos al aplicar los filtros seleccionados en la sección correspondiente. Pulsando sobre uno de los elementos de la tabla cargará un gráfico que muestra las cotizaciones de la empresa, es decir, la variación del precio de sus acciones en el tiempo.

En la parte final del diseño se muestra una tabla con algunas de las medidas obtenidas de la tabla de hechos fundamentales y unos gráficos generados a partir de estas medidas.

7.4.2. Implementación

Para la implementación de esta herramienta se ha utilizado Pentaho Community Dashboard Editor (CDE), que es una herramienta gráfica para crear, editar y previsualizar cuadros de mandos sobre la plataforma de Pentaho BI.

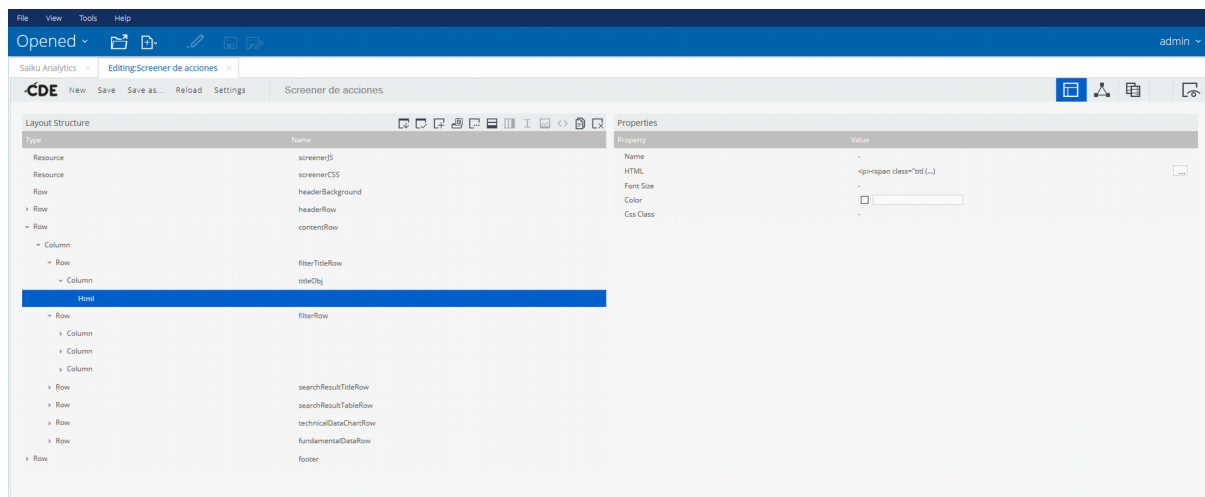
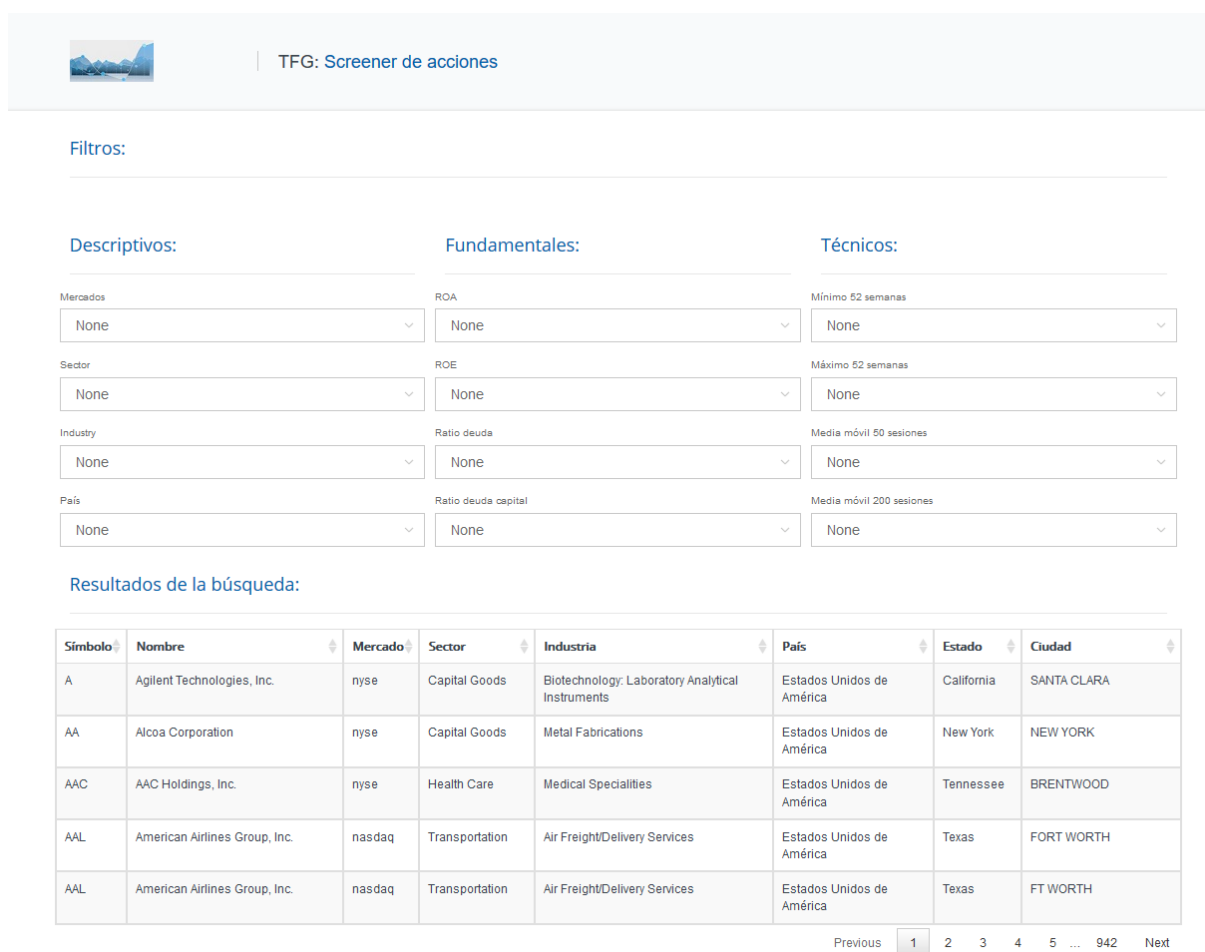


Ilustración 44: Pentaho Community Dashboard Editor (CDE)

Antes de comenzar a describir el screener desarrollado, comentar que la tabla de hechos técnicos dispone de cerca de 21 millones de registros y este motivo provoca que la realización de búsquedas sobre esta tabla, no se realice en unos tiempos de respuesta razonable. Para solventar esta circunstancia, se ha creado una vista materializada en base de datos que contiene la última cotización para cada una de las empresas. Esta pequeña adición al sistema permite mejorar drásticamente los tiempos de respuesta.

A continuación se muestra una captura del screener en funcionamiento:



TFG: Screener de acciones

Filtros:

Descriptivos:	Fundamentales:	Técnicos:
Mercados None	ROA None	Mínimo 52 semanas None
Sector None	ROE None	Máximo 52 semanas None
Industry None	Ratio deuda None	Media móvil 50 sesiones None
País None	Ratio deuda capital None	Media móvil 200 sesiones None

Resultados de la búsqueda:

Símbolo	Nombre	Mercado	Sector	Industria	País	Estado	Ciudad
A	Agilent Technologies, Inc.	nyse	Capital Goods	Biotechnology: Laboratory Analytical Instruments	Estados Unidos de América	California	SANTA CLARA
AA	Alcoa Corporation	nyse	Capital Goods	Metal Fabrications	Estados Unidos de América	New York	NEW YORK
AAC	AAC Holdings, Inc.	nyse	Health Care	Medical Specialities	Estados Unidos de América	Tennessee	BRENTWOOD
AAL	American Airlines Group, Inc.	nasdaq	Transportation	Air Freight/Delivery Services	Estados Unidos de América	Texas	FORT WORTH
AAL	American Airlines Group, Inc.	nasdaq	Transportation	Air Freight/Delivery Services	Estados Unidos de América	Texas	FT WORTH

Previous 1 2 3 4 5 ... 942 Next

Ilustración 45: Screener de acciones

En su parte superior se pueden observar los filtros implementados:

- Filtros descriptivos: mercado (Nyse, Nasdaq, Amex), sector (Basic industries, Capital Goods, ...), industria (Aerospace, Agricultural Chemical, ...), país (Alemania, Estados Unidos,...)
- Filtros fundamentales: roa, roe, ratio deuda, ratio deuda patrimonio
- Filtros técnicos: máximo y mínimo de 52 semanas, media móvil de 50 y 200 sesiones

Cada uno de los filtros es un desplegable que permite realizar la búsquedas parciales con una función de auto-completado

Filtros:

Descriptivos:

Mercados:

Sector:

Industry:

- Professional Services
- Radio And Television Broadcasting And Co...
- Retail: Computer Software & Peripheral Equ...
- Semiconductors
- Telecommunications Equipment

Fundamentales:

ROA:

ROE:

Ratio deuda:

Ratio deuda capital:

Técnicos:

Mínimo 52 semanas:

Máximo 52 semanas:

Media móvil 50 sesiones:

Media móvil 200 sesiones:

Sector	Industria	País	Estado	Ciudad
Technology	Semiconductors	Estados Unidos de América	Texas	SUGAR LAND
AAPL	Apple Inc.	Estados Unidos de América	California	CUPERTINO
ABTL	Autobyte Inc.	Estados Unidos de América	California	IRVINE
ACIA	Acacia Communications, Inc.	Estados Unidos de América	Massachusetts	MAYNARD
ACIW	ACI Worldwide, Inc.	Estados Unidos de América	New York	NEW YORK

Previous 1 2 3 4 5 ... 83 Next

Ilustración 46: Screener de acciones: Desplegables con autocompletado

En los casos de los filtros denominados técnicos y fundamentales, las opciones seleccionables se corresponden con opciones del tipo:

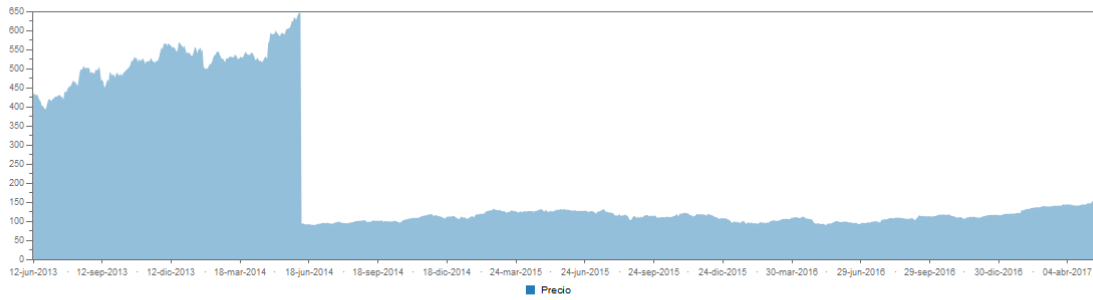
- Precio un 5,10,15% por encima del mínimo de 52 semanas

- Precio un 5,10,15% por debajo del máximo de 52 semanas
- Precio un 5,10,15% por encima de la media móvil de 50 sesiones
- Precio un 5,10,15% por encima de la media móvil de 200 sesiones
- ROA superior al 5,10,15,20,25,50,100%
- ROE positivo (>0%), negativo (<0%), superior al 5,10,15,20,25,50,100%
- Ratio deuda inferior al 25,50,75,100%
- Ratio deuda patrimonio inferior al 25,50,75,100%

A continuación de los filtros, se muestra una tabla paginada con los resultados obtenidos de la búsqueda resultante de aplicar las condiciones seleccionadas. Los datos que se muestran son información descriptiva de las empresas (símbolo, nombre, mercado, sector, industria, país, estado/provincia y ciudad) .

Esta tabla permite la selección de los registros. Al seleccionar uno de ellos se cargan los componentes que se encuentran justo debajo de la tabla: el gráfico de las cotizaciones históricas, la tabla de datos fundamentales y las gráficas generadas a partir de estos datos.

Cotización histórica:



Datos financieros:

Año	Ingresos totales	Beneficio neto	Activos totales	Recursos propios	Pasivo total
2,016	54,530 M	11,304 M	315,927 M	129,409 M	186,517 M
2,015	58,747 M	13,433 M	279,527 M	125,576 M	153,951 M
2,014	49,950 M	11,116 M	230,561 M	118,999 M	111,562 M
2,013	43,498 M	9,258 M	206,696 M	128,019 M	78,677 M
2,012	41,172 M	10,437 M	171,496 M	114,950 M	56,546 M
2,011	31,960 M	8,246 M	114,179 M	74,372 M	39,806 M
2,010	19,071 M	4,160 M	70,927 M	46,229 M	24,698 M
2,009	12,010 M	2,304 M	49,856 M	31,099 M	18,757 M

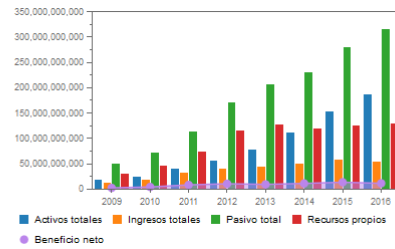
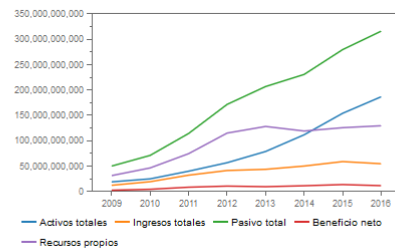


Ilustración 47: Screener de acciones: Cotizaciones y datos fundamentales

8. Trabajo futuro

El mundo de las inversiones en los mercados financieros es un campo muy extenso y que permite la ampliación de este trabajo en múltiples aspectos. Este apartado trata de plantear algunas de las posibilidades detectadas durante este trabajo:

- La ETL diseñada obtiene los datos financieros de las empresas desde unos ficheros de textos que proporciona la SEC y que son generados a partir de los ficheros XBRL presentados por la empresas de forma periódica. Una posible mejora podría tratar de obtener los datos directamente desde los ficheros en formato XBRL.
- A partir de la mejora anterior, se podría implementar un proceso que se suscriba al feed RSS, que la propia SEC proporciona, para incorporar los datos financieros prácticamente en el mismo momento de su publicación. Esta funcionalidad es interesante, pues los ficheros de texto utilizados actualmente tardan un tiempo en estar disponibles.
- En cuanto a los propios datos, se podría enriquecer el sistema proporcionando más información al mismo. En este ámbito sería posible incorporar información de fondos de inversión, de empresas que cotizan en mercados regulados por otros organismos, como por ejemplo las empresas que cotizan en los mercados españoles, supervisados por la Comisión Nacional del Mercado de Valores (CNMV).
- También sería posible incorporar mejoras en la herramienta de análisis creada, incorporando más posibilidades de búsqueda de empresas que cumplan ciertas restricciones, añadiendo algunos de los indicadores técnicos y/o ratios fundamentales expuestos en este TFG.
- La creación de cuadros de mandos para el análisis de la información financiera de una empresa en concreto, es otro área extensa que sería posible abordar en un nuevo TFG. Estos cuadros de mandos podría ofrecer una visión sobre los distintos estados contables (cuenta de resultados, balance de situación y los flujos de efectivo).

- De la misma forma, también se podría ampliar el presente trabajo mediante la creación de informes/reporting con la información financiera propuesta para el cuadro de mandos.
- Por último, otro campo a explorar sería la utilización de las técnicas y herramientas de la minería de datos (modelos predictivos, árboles de decisión, etc), al sistema generado. A modo de ejemplo, se podría plantear realizar algo similar a un análisis de la cesta de la compra, es decir, sugerir la compra de nuevos productos relacionados con los ya incorporados en la cesta. Aplicándolo al campo objeto de este TFG, se podría realizar el análisis de una cartera de inversiones para sugerir otras empresas en las que invertir. Podría utilizarse como base de conocimiento la composición de las carteras de distintos fondos de inversión.

9. Conclusiones

La realización de este trabajo de fin de grado, en mi opinión, ha cumplido con el objetivo principal planteado: la adquisición por parte de este alumno de los conocimientos y las capacidades necesarias para el diseño y la construcción de un sistema de inteligencia de negocio. El trabajo realizado ha permitido profundizar en cada una de las etapas necesarias para el desarrollo de un sistema multidimensional que permita explotar la información de un ámbito concreto con el objetivo mejorar el proceso de toma de decisiones.

En cuanto a los objetivos específicos, planteados al inicio de este documento, creo que se han alcanzado todos en mayor o menor medida, pues se ha logrado generar un producto funcional.

En cuanto al desarrollo del propio trabajo, me gustaría comentar algunos aspectos:

- La idea de proyecto inicial consistía en la creación de una herramienta de análisis para fondos de inversión, pero durante la primera entrega evolucionó hasta la solución finalmente adoptada. Este cambio surge principalmente, por la cantidad de información disponible para empresas y fondos de inversión, siendo la disponible para acciones muy superior, y ello permite tener un sistema más rico en información y potente, en cuanto a funcionalidades.

La herramienta inicialmente planteada que es una versión especializada de la solución propuesta, pues un fondo de inversión no es más que un portafolio de inversiones gestionada por un tercero.

También considero que la solución adoptada, es más generalista, versátil y permite plantear múltiples escenarios de ampliación o trabajos futuros. Por este motivo, el planteamiento inicial ha quedado como una futura línea de trabajo a explorar.

- Una vez planteada esta solución, se realizó un trabajo importante de análisis de las distintas fuentes de información. En primer lugar, se identificaron ficheros excel con la información financiera y posteriormente se encontró la referencia al formato XBRL. Este formato tiene una alta complejidad y por ello la propia SEC realiza un trabajo para reducir la complejidad de estos

documentos. Hemos aprovechado este trabajo y lo hemos utilizado para obtener toda la información financiera de las empresas.

- En cuanto a la planificación del proyecto, ha existido una desviación en el desarrollo de la ETL y en el desarrollo del cuadro de mandos a consecuencia de la anterior. Esta desviación inicial se ha originado, principalmente, por la complejidad de las estructuras de los ficheros a tratar y el tiempo de carga de algunos de los procesos, debido al alto volumen de datos generados en la primera carga.
- Para poder solventar estas complicaciones se han dedicado más horas de las planificadas en un primer momento. Completar este trabajo ha requerido de un gran esfuerzo a nivel personal, pues compatibilizar la vida familiar y profesional con este trabajo ha sido un reto importante.

Por último destacar, que la realización de este proyecto me ha servido para entrar en contacto con el mundo del análisis de los datos, con las tecnologías que se aplican, descubrir las posibilidades del BI y ampliar los conocimientos en este campo. Creo que estos conocimientos adquiridos me van a ser muy útiles para el desarrollo personal en el ámbito profesional.

10. Glosario

Business Intelligence: Conjunto de estrategias y herramientas que permiten la transformación de los datos en información y estos en conocimiento de tal forma que permita mejorar los procesos de toma de decisiones de una organización.

Data Warehouse: Almacén de datos cuya finalidad es integrar los datos existentes en los sistemas de información de las organizaciones, que permite un análisis de la información desde distintas perspectivas o dimensiones.

Dimensión: Una dimensión es una categorización de los hechos, que permiten agrupar, filtrar u ordenar la información.

Hecho: Un hecho representa los eventos o conceptos de negocio que son cuantificables y que se utilizan para realizar análisis.

Granularidad. Indica el nivel de detalle de las tablas de hechos dentro del BI.

Métrica o Indicador: Son medidas, o conceptos calculados, y con los cuales generalmente se puede operar, que resultan interesantes para el análisis y cumplir con los objetivos requeridos.

Staging área: Área intermedia de almacenamiento de datos utilizada para el procesamiento durante los procesos de ETL. Esta área se encuentra entre las fuentes de los datos y su destino.

ETL (Extract, Transform And Load): Procesos de extracción, transformación y carga de datos que nutren de información el BI.

OLAP: Solución usada en el campo de la Inteligencia Empresarial cuyo objetivo es agilizar la consulta de grandes volúmenes de datos.

ROLAP: Implementación OLAP construida sobre bases de datos relacionales.

MDX: Lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP.

Dashboard/Cuadro de mandos: Herramienta que permite mediante un sencillo manejo tener acceso a los principales indicadores de negocio.

Bibliografía

- 1: Banco Mundial, Compañías nacionales que cotizan en Bolsa, <http://datos.bancomundial.org/indicador/CM.MKT.LDOM.NO>,
- 2: Kimball, Ralph, The Data Warehouse Lifecycle Toolkit, 1998
- 3: Wikipedia, PostgreSQL, <https://es.wikipedia.org/wiki/PostgreSQL>,
- 4: PostgreSQL, Documentación PostgreSQL, <https://www.postgresql.org/docs/>
- 5: Wikipedia, Pentaho Community Edition, <https://es.wikipedia.org/wiki/Pentaho>,
- 6: Pentaho Community Edition, Documentación Pentaho Community Edition, <http://wiki.pentaho.com/display/COM/Community+Wiki+Home>
- 7: GNU.ORG, GNU General Public License, version 2, <https://www.gnu.org/licenses/old-licenses/gpl-2.0.html>
- 8: Wikipedia, Análisis Técnico, https://en.wikipedia.org/wiki/Technical_analysis,
- 9: Wikipedia, Análisis Cuantitativo, [https://es.wikipedia.org/wiki/An%C3%A1lisis_cuantitativo_\(finanzas\)](https://es.wikipedia.org/wiki/An%C3%A1lisis_cuantitativo_(finanzas)),
- 10: Wikipedia, La Volatilidad, [https://es.wikipedia.org/wiki/Volatilidad_\(finanzas\)](https://es.wikipedia.org/wiki/Volatilidad_(finanzas)),
- 11: Wikipedia, Índice de fuerza relativa, https://es.wikipedia.org/wiki/%C3%8Dndice_de_fuerza_relativa,
- 12: Wikipedia, Ratios financieros, https://es.wikipedia.org/wiki/Ratio_financiero,

11. Anexo I: Scripts creación BBDD

```
CREATE TABLE stg_data(  
  anyo varchar(4000),  
  trimestre varchar(4000),  
  ruta varchar(4000),  
  fichero varchar(4000),  
  fecha_carga DATE  
);  
CREATE INDEX idx_stg_data_lookup ON stg_data(anyo, trimestre, fichero);  
  
CREATE TABLE stg_sub(  
  adsh VARCHAR(4000),  
  cik VARCHAR(4000),  
  "name" VARCHAR(4000),  
  sic VARCHAR(4000),  
  countryba VARCHAR(4000),  
  stprba VARCHAR(4000),  
  cityba VARCHAR(4000),  
  zipba VARCHAR(4000),  
  bas1 VARCHAR(4000),  
  bas2 VARCHAR(4000),  
  baph VARCHAR(4000),  
  countryma VARCHAR(4000),  
  stprma VARCHAR(4000),  
  cityma VARCHAR(4000),  
  zipma VARCHAR(4000),  
  mas1 VARCHAR(4000),  
  mas2 VARCHAR(4000),  
  countryinc VARCHAR(4000),  
  stprinc VARCHAR(4000),  
  ein VARCHAR(4000),  
  former VARCHAR(4000),  
  changed VARCHAR(4000),  
  afs VARCHAR(4000),  
  wksi VARCHAR(4000),  
  fye VARCHAR(4000),  
  form VARCHAR(4000),  
  period VARCHAR(4000),  
  fy VARCHAR(4000),  
  fp VARCHAR(4000),  
  filed VARCHAR(4000),  
  accepted VARCHAR(4000),  
  prevrpt VARCHAR(4000),  
  detail VARCHAR(4000),  
  "instance" VARCHAR(4000),  
  nciks VARCHAR(4000),  
  aciks VARCHAR(4000),  
  "year" VARCHAR(100)  
);  
CREATE INDEX idx_stg_sub_lookup ON stg_sub(adsh, cik);  
  
CREATE TABLE stg_num(  
  anyo varchar(4000),  
  trimestre varchar(4000),  
  ruta varchar(4000),  
  fichero varchar(4000),  
  fecha_carga DATE  
);  
CREATE INDEX idx_stg_num_lookup ON stg_num(anyo, trimestre, fichero);
```

```
adsh VARCHAR(4000),
tag VARCHAR(4000),
cik VARCHAR(4000),
version VARCHAR(4000),
coreg VARCHAR(4000),
ddate VARCHAR(4000),
qtrs VARCHAR(4000),
uom VARCHAR(4000),
"value" DOUBLE PRECISION,
footnote VARCHAR(4000)
);
CREATE INDEX idx_stg_num_lookup ON stg_num(adsh, tag, cik);

CREATE TABLE stg_companies(
id_company BIGINT,
cik VARCHAR(4000),
Symbol VARCHAR(4000),
"Name" VARCHAR(4000),
Sector VARCHAR(4000),
Industry VARCHAR(4000),
Market VARCHAR(4000)
);
CREATE UNIQUE INDEX idx_stg_companies_pk ON stg_companies(id_company);
CREATE INDEX idx_stg_companies_lookup ON stg_companies(Symbol, "Name", Sector, Industry, Market);

CREATE TABLE stg_quotes (
id_quote BIGINT,
symbol VARCHAR(4000),
"date" DATE,
quotes TEXT,
version INTEGER,
date_from DATE,
date_to DATE,
CONSTRAINT stg_quotes_UNIQUE UNIQUE (id_quote)
);
CREATE INDEX idx_stg_quotes_lookup ON stg_quotes(symbol, "date");

CREATE TABLE stg_sma (
id_company BIGINT,
id_time BIGINT,
"close" DOUBLE PRECISION,
indice BIGINT,
sma20 BIGINT,
sma50 BIGINT,
sma200 BIGINT,
sma1000 BIGINT
);
CREATE INDEX idx_stg_sma_lookup ON stg_sma(id_company, id_time);

CREATE TABLE d_time (
id_time BIGINT,
year_number INTEGER,
month_number INTEGER,
day_of_year_number INTEGER,
day_of_month_number INTEGER,
day_of_week_number INTEGER,
week_of_year_number INTEGER,
```

```
day_name VARCHAR(100),
month_name VARCHAR(100),
quarter_number INTEGER,
quarter_name VARCHAR(2),
year_quarter_name VARCHAR(100),
weekend_ind VARCHAR(1),
DAYS_IN_MONTH_QTY INTEGER,
DATE_SK INTEGER,
DAY_DESC VARCHAR(100),
WEEK_SK INTEGER,
DAY_DATE DATE,
week_name VARCHAR(100),
week_of_month_number INTEGER,
week_of_month_name VARCHAR(100),
MONTH_SK INTEGER,
QUARTER_SK INTEGER,
YEAR_SK INTEGER,
YEAR_SORT_NUMBER VARCHAR(4),
DAY_OF_WEEK_SORT_NAME VARCHAR(100),
DATE_MEDIUM VARCHAR(100),
DATE_LONG VARCHAR(100),
DATE_FULL VARCHAR(100)
);
CREATE INDEX idx_d_time_lookup ON d_time(id_time);

CREATE TABLE d_company(
    id_company BIGINT,
    cik VARCHAR(4000),
    symbol VARCHAR(4000),
    "name" VARCHAR(4000),
    sector VARCHAR(4000),
    industry VARCHAR(4000),
    market VARCHAR(4000),
    city VARCHAR(4000),
    country VARCHAR(4000),
    "state" VARCHAR(4000),
    version INTEGER,
    date_to DATE,
    date_from DATE,
    CONSTRAINT id_company_UNIQUE UNIQUE (id_company)
);
CREATE INDEX idx_d_company_lookup ON d_company(cik);

CREATE TABLE f_fundamental (
    id_company BIGINT,
    id_time INTEGER,
    Sales DOUBLE PRECISION,
    NetIncome DOUBLE PRECISION,
    Liabilities DOUBLE PRECISION,
    Assets DOUBLE PRECISION,
    Equity DOUBLE PRECISION,
    ROA DOUBLE PRECISION,
    ROE DOUBLE PRECISION,
    DebtRatio DOUBLE PRECISION,
    DebtEquityRatio DOUBLE PRECISION
);
CREATE INDEX idx_f_fundamental_lookup ON f_fundamental(id_submission, id_company, id_time);
```



```
CREATE TABLE f_tecnico (  
  id_company BIGINT,  
  id_time BIGINT,  
  change DOUBLE PRECISION,  
  "open" DOUBLE PRECISION,  
  high DOUBLE PRECISION,  
  low DOUBLE PRECISION,  
  "close" DOUBLE PRECISION,  
  volume DOUBLE PRECISION,  
  adjusted_close DOUBLE PRECISION,  
  sma20 DOUBLE PRECISION,  
  sma50 DOUBLE PRECISION,  
  sma200 DOUBLE PRECISION,  
  sma1000 DOUBLE PRECISION,  
  min52Weeks DOUBLE PRECISION,  
  max52Weeks DOUBLE PRECISION  
);  
CREATE INDEX idx_f_tecnico_lookup ON f_tecnico(id_company, id_time);  
  
CREATE MATERIALIZED VIEW f_tecnico2 AS  
select t1.*  
from f_tecnico t1  
where t1.id_time = (select max(t.id_time)  
                   from f_tecnico t  
                   where t.id_company = t1.id_company)  
order by t1.id_company;  
  
CREATE UNIQUE INDEX idx_f_tecnico2_lookup ON f_tecnico2 (id_time, id_company);  
REFRESH MATERIALIZED VIEW CONCURRENTLY f_tecnico2;
```