

DATA WAREHOUSING

Héctor Álvarez Gómez
Felipe Vildoso Castillo



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



Ingeniería Industrial
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

MONDRIAN

Capítulo 6



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

EL ROL DE MONDRIAN EN ANALYTICS

Capítulo 6

¿QUÉ ES MONDRIAN?

Mondrian es un **motor** de analytics que acepta consultas “analíticas” y las convierte en consultas relacionales, devolviendo los datos de una forma más pertinente para ser analizados.

Para que Mondrian sea útil para los usuarios empresariales, **necesita algún tipo de interfaz y aplicación para ejecutarlo.**

MANERAS DE IMPLEMENTAR MONDRIAN

- Se puede configurar en un servidor de aplicaciones para que funcione por sí solo y proporcione servicios, por lo que **no tiene una interfaz gráfica de usuario agradable**.
- Se puede integrar en una aplicación independiente. Este enfoque es **bueno para aplicaciones personalizadas con un propósito específico** pero no es tan útil para soportar una amplia variedad de usos.

MANERAS DE IMPLEMENTAR MONDRIAN

- Se puede implementar como **parte de una aplicación web** más amplia que proporciona herramientas para los usuarios con los que trabajar. Este enfoque es el más flexible, pero también **requiere la mayor organización y configuración.**



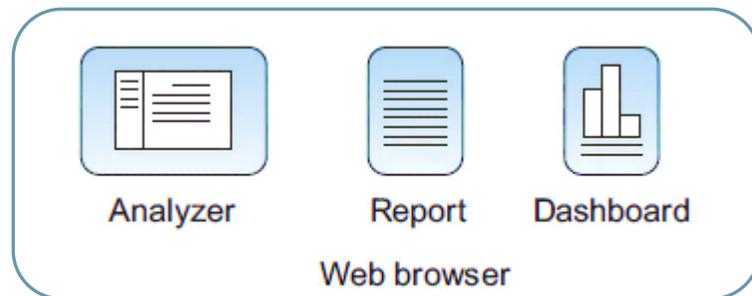
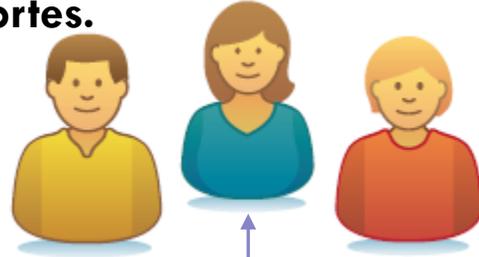
ARQUITECTURA MONDRIAN

Esta visión es simplificada, pero contiene las partes **principales de un sistema que utiliza Mondrian**.

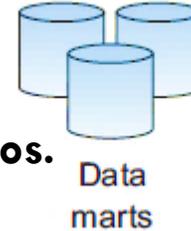
- Los usuarios interactúan utilizando herramientas basadas en la web.
- Se aceptan las consultas de estas herramientas y luego se utilizan coincidencias para generar consultas SQL.
- Se devuelven los resultados a los clientes para formatearlos y mostrarlos a los usuarios.

ARQUITECTURA MONDRIAN

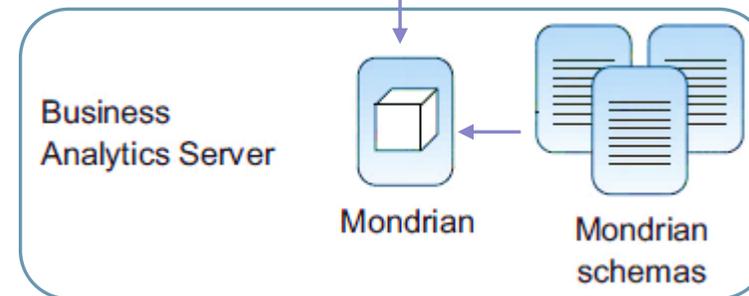
1. Los usuarios hacen análisis y generan reportes.



4. Mondrian envía consultas SQL y retorna los resultados.



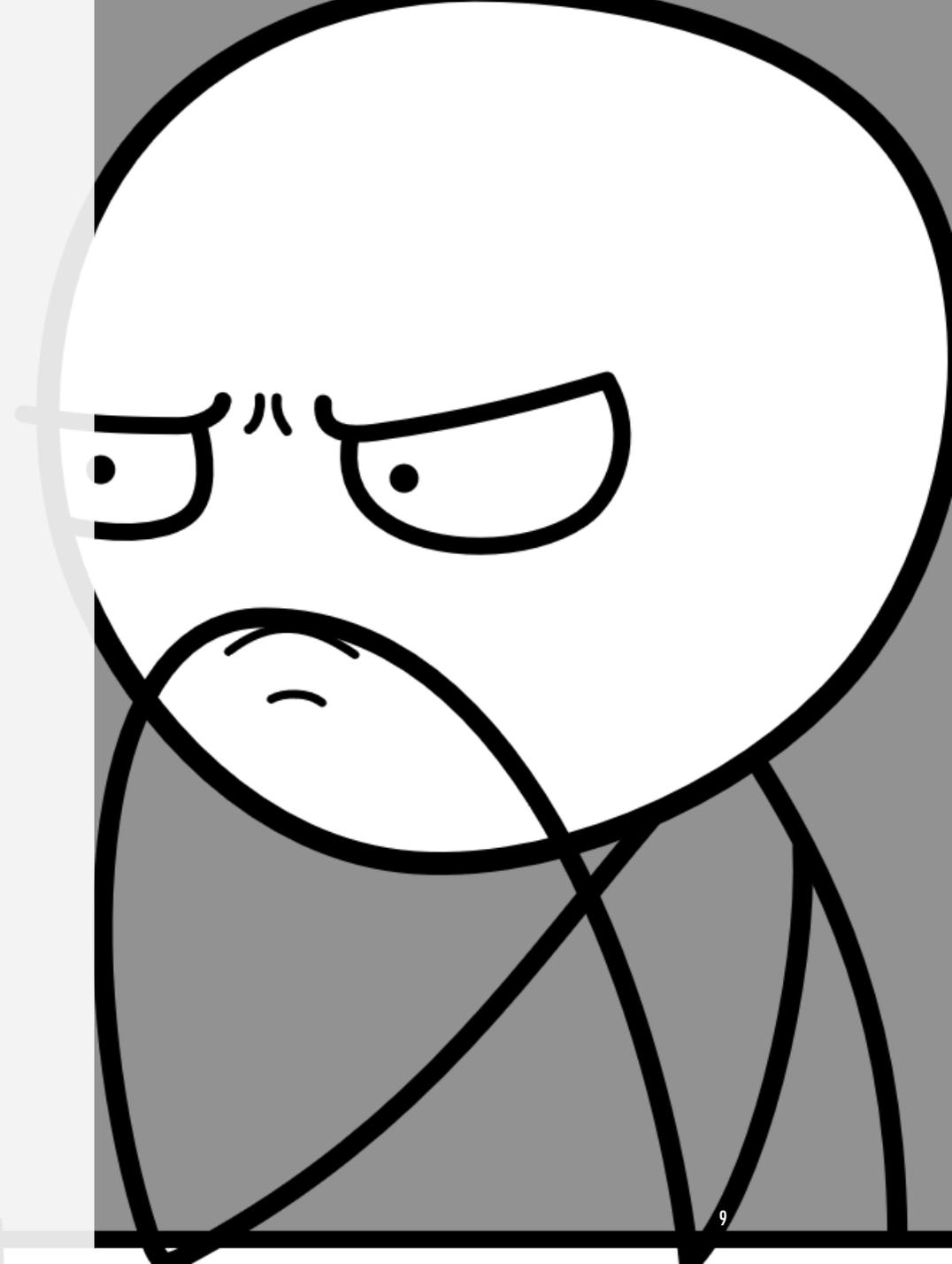
3. Mondrian usa un esquema para mapear desde la lógica del cubo al esquema físico.



2. Mondrian recibe consultas MDX para procesar y enviar resultados.

ARQUITECTURA MONDRIAN

En síntesis, Mondrian ha sido diseñado para ser **fácil de integrar, y flexible para una variedad de usuarios e interfaces.**





Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

IMPLEMENTACIONES DE MONDRIAN

Capítulo 6

CORRER Y USAR MONDRIAN

Mondrian es un **motor analítico** en lugar de una aplicación o servidor, por lo que necesita ejecutarse dentro de otra aplicación o servidor.

Como ya vimos, esto hace que Mondrian sea altamente flexible ya que puede reutilizarse dentro de una amplia variedad de aplicaciones y escenarios, dependiendo de las necesidades del negocio.

¿Qué herramienta usaremos en el curso?

JASPERREPORTS SERVER

JasperReports Server es un servidor de informes autónomo e integrable. Permite proporcionar **informes y análisis** que se integran en una aplicación web o móvil y funciona como un **centro de información central** para las empresas.



INSTALANDO Y USANDO JASPER

JasperReports Server Community Edition (v6.3.0)

Published: 2016-Jul-07

[Full Release Notes](#)

1. Descargar

Download

- README_releaseNotes.txt
- jasperreports-server-cp-6.3.0-windows-x64-installer.exe
- jasperreports-server-cp-6.3.0-src.zip
- jasperreports-server-cp-6.3.0-osx-x64-installer.app.zip
- jasperreports-server-cp-6.3.0-linux-x86-installer.run
- jasperreports-server-cp-6.3.0-linux-x64-installer.run
- jasperreports-server-cp-6.3.0-bin.zip
- jasperreports-server-ce-javadoc.zip
- checksum.txt
- Jaspersoft-Studio-User-Guide.pdf
- JasperReports-Server-User-Guide.pdf
- Jaspersoft-OLAP-User-Guide.pdf
- JasperReports-Server-Security-Guide.pdf
- JasperReports-Server-REST-API-Reference.pdf
- JasperReports-Server-CP-Upgrade-Guide.pdf
- JasperReports-Server-CP-Install-Guide.pdf
- JasperReports-Server-CP-Admin-Guide.pdf

2. Video de instalación
<https://youtu.be/M4lbwGJKaw>

3. Explorar la herramienta

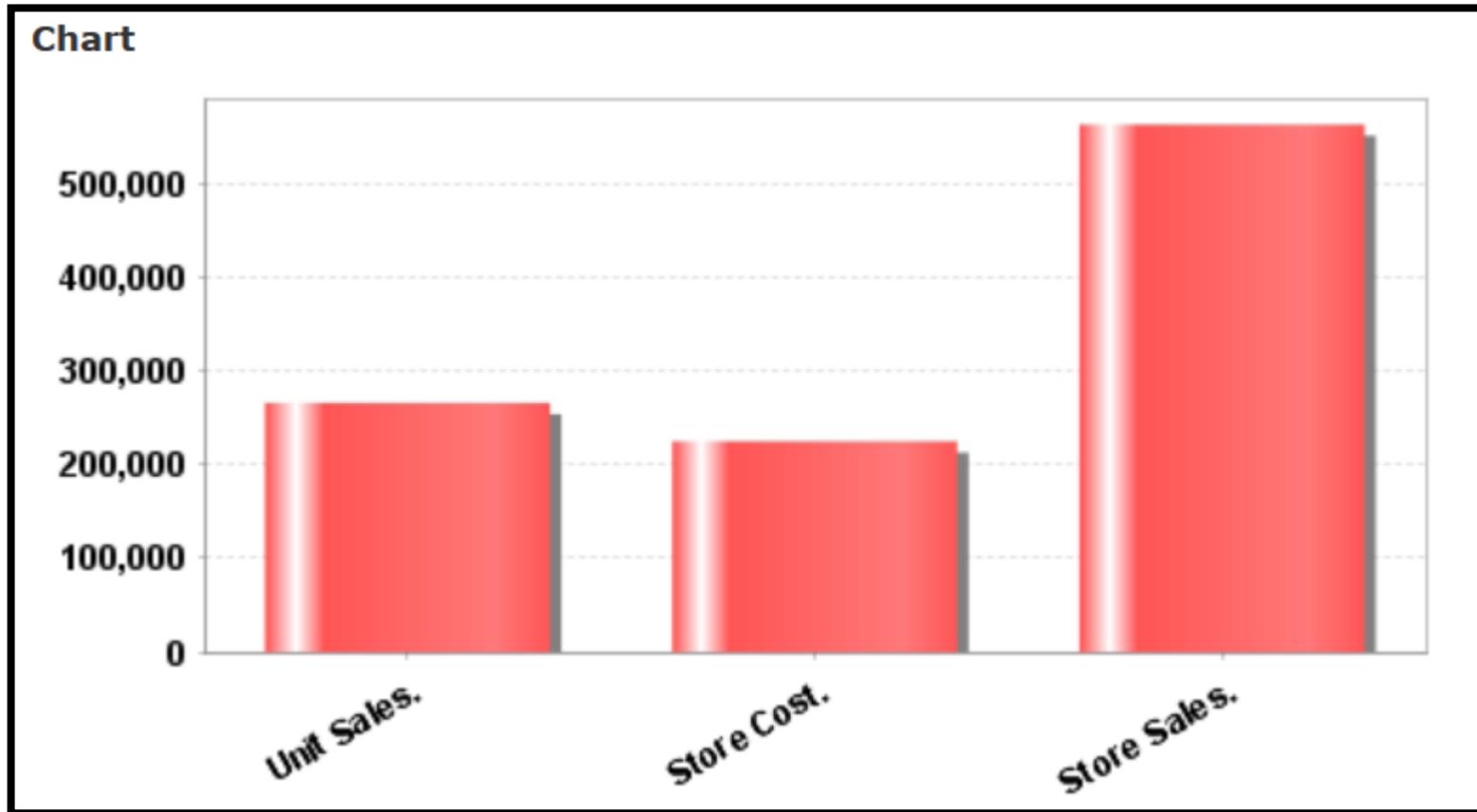
FOODMART MONDRIAN SALES REPORT

FoodMart Sales		
Store State	Unit Sales (\$)	# Customers
Store CA	\$74,748.00	2,716
Q1		
Customer CA	\$16,890.00	1,110
Q2		
Customer CA	\$18,052.00	1,173
Q3		
Customer CA	\$18,370.00	1,196
Q4		
Customer CA	\$21,436.00	1,332
Store OR	\$67,659.00	1,037
Q1		
Customer OR	\$19,287.00	767
Q2		
Customer OR	\$15,079.00	723
Q3		
Customer OR	\$16,940.00	717
Q4		
Customer OR	\$16,353.00	741
Store WA	\$124,366.00	1,828
Q1		
Customer WA	\$30,114.00	1,104
Q2		
Customer WA	\$29,479.00	1,077
Q3		
Customer WA	\$30,538.00	1,113
Q4		
Customer WA	\$34,235.00	1,188
Totals	\$266,773.00	5,581

FOODMART SAMPLE ANALYSIS VIEW

Dimensions						Measures		
Promotion Media	Media Type	Product	Product Family	Product Department	Product Category	Unit Sales	Store Cost	Store Sales
<input type="checkbox"/> All Media		<input checked="" type="checkbox"/> All Products				266,773	225,627.23	665,238.13
All Media	Bulk Mail	<input checked="" type="checkbox"/> All Products				4,320	3,740.95	9,349.07
	Cash Register Handout	<input checked="" type="checkbox"/> All Products				6,697	5,715.67	14,321.33
	Daily Paper	<input checked="" type="checkbox"/> All Products				7,738	6,559.23	16,479.81
	Daily Paper, Radio	<input checked="" type="checkbox"/> All Products				6,891	5,668.77	14,169.42
	Daily Paper, Radio, TV	<input checked="" type="checkbox"/> All Products				9,513	8,055.22	20,173.97
	In-Store Coupon	<input checked="" type="checkbox"/> All Products				3,798	3,263.11	8,162.46
	No Media	<input checked="" type="checkbox"/> All Products				195,448	165,214.85	414,026.92
	Product Attachment	<input checked="" type="checkbox"/> All Products				7,544	6,306.24	15,898.25
	Radio	<input checked="" type="checkbox"/> All Products				2,454	2,087.51	5,213.61
	Street Handout	<input checked="" type="checkbox"/> All Products				5,753	4,856.54	12,192.90
	Sunday Paper	<input checked="" type="checkbox"/> All Products				4,339	3,673.86	9,092.89
	Sunday Paper, Radio	<input checked="" type="checkbox"/> All Products				5,945	5,027.31	12,551.96
	Sunday Paper, Radio, TV	<input checked="" type="checkbox"/> All Products				2,726	2,341.58	5,819.33
	TV	<input type="checkbox"/> All Products					3,607	3,116.40
<input type="checkbox"/> All Products		<input type="checkbox"/> Drink				332	283.18	721.07
		Drink	<input type="checkbox"/> Alcoholic Beverages			76	70.99	182.38
			Alcoholic Beverages	<input checked="" type="checkbox"/> Beer and Wine		76	70.99	182.38
			<input checked="" type="checkbox"/> Beverages			188	167.00	419.14
			<input checked="" type="checkbox"/> Dairy			68	45.19	119.55
		<input checked="" type="checkbox"/> Food				2,563	2,238.13	5,573.27
	<input checked="" type="checkbox"/> Non-Consumable				712	595.08	1,491.87	

FOODMART SAMPLE ANALYSIS VIEW



SCHEMA MONDRIAN

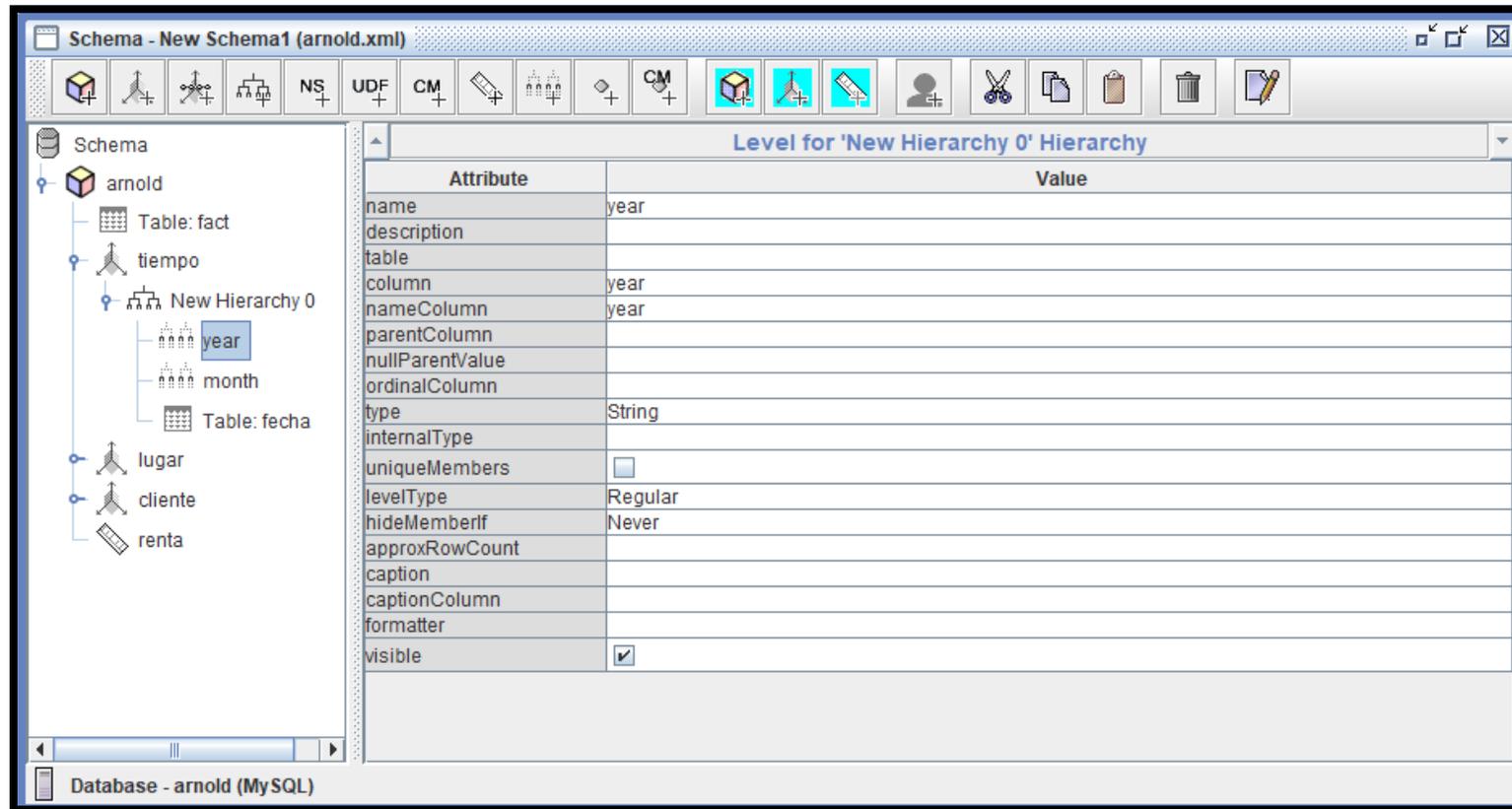
Mondrian, como un motor OLAP, presenta los datos de forma multidimensional: **el contenido son hechos** de datos que los analistas de negocios quieren conocer, como ventas e inventario, y **las dimensiones son atributos sobre los datos para análisis**, como almacén, geografía, demografía del cliente.

Hay que definir el schema en un archivo xml.

EJEMPLO SCHEMA MONDRIAN V3

```
<Schema name="New Schema1">
  <Cube name="arnold" visible="true" cache="true" enabled="true">
    <Table name="fact" alias="">
    </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_fecha" name="tiempo">
      <Hierarchy name="New Hierarchy 0" visible="true" hasAll="true" primaryKey="id">
        <Table name="fecha" alias="">
        </Table>
        <Level name="year" visible="true" column="year" nameColumn="year" uniqueMembers="false">
        </Level>
        <Level name="month" visible="true" column="month" nameColumn="month" uniqueMembers="false">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_lugar" name="lugar">
      <Hierarchy name="New Hierarchy 0" visible="true" hasAll="true">
        <Table name="lugar" alias="">
        </Table>
        <Level name="pais" visible="true" column="pais" uniqueMembers="false">
        </Level>
        <Level name="habitacion" visible="true" column="habitacion" uniqueMembers="false">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_cliente" name="cliente">
      <Hierarchy name="New Hierarchy 0" visible="true" hasAll="true">
        <Table name="usuarios" alias="">
        </Table>
        <Level name="New Level 0" visible="true" column="id" uniqueMembers="false">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="renta" column="renta" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

PODEMOS USAR SCHEMA WORKBENCH



Nota: Solo funciona para Mondrian 3.

USAMOS PARA CONSULTAR MDX

MultiDimensional eXpressions (MDX) proporciona una **sintaxis especializada para consultar y manipular los datos multidimensionales almacenados en cubos OLAP**. Ha sido adoptado por una amplia mayoría de proveedores OLAP y se ha convertido en el estándar.

```
select {[Measures].[Unit Sales], [Measures].[Store Cost],  
[Measures].[Store Sales]} ON COLUMNS,  
{([Promotion Media].[All Media], [Product].[All Products])} ON ROWS  
from [Sales]
```

Importante: En Clase Auxiliar verán como hacer el schema, como subirlo a JasperReports Server, y como consultarlo con MDX.

¿DÓNDE ESTÁN LOS DATOS?

Como se ha visto, Mondrian presenta los datos como cubos con múltiples dimensiones para el análisis. Pero éste **no almacena los datos**; simplemente proporciona una vista lógica de los datos físicos.

Mondrian es un motor OLAP relacional (ROLAP), lo que significa que los datos **se almacenan en una base de datos relacional** y traduce consultas MDX a consultas SQL para el RDBMS particular que utiliza para el almacenamiento.

ALGUNAS VENTAJAS

- Puedes **elegir una base de datos optimizada** para los tipos de consultas que normalmente realizarán los analistas.
- Se tienen **todas las capacidades de copia de seguridad**, los DBA ya están acostumbrados a usarlas.
- Puedes **cambiar a una base de datos diferente en el futuro** y seguir utilizando Mondrian (suponiendo que se admita la nueva base de datos).

El análisis de arrastrar y soltar (drag-and-drop) proporciona **mucha visión a los usuarios empresariales**, ya que pueden realizar rápidamente una amplia variedad de análisis sin conocer un lenguaje de consulta ni comprender los detalles de la base de datos subyacente.

SCHEMA MONDRIAN V4

Manual:

http://mondrian.pentaho.com/head/documentation/schema.php#Schema_element

```
<Schema name="Arnold4" metamodelVersion="4.0">
  <PhysicalSchema>
    <Table name="fact"/>
    <Table name="usuarios"/>
    <Table name="fecha"/>
  </PhysicalSchema>
  <Cube name="renta">
    <Dimensions>
      <Dimension name="Time" table="fecha" key="Id">
        <Attributes>
          <Attribute name="Year" column="year">
            <Key> <Column name="year"/> </Key>
          </Attribute>
          <Attribute name="Month" column="month">
            <Key> <Column name="month"/> </Key>
          </Attribute>
          <Attribute name="Id" column="id">
            <Key> <Column name="id"/> </Key>
          </Attribute>
        </Attributes>
        <Hierarchies>
          <Hierarchy name="Yearly" hasAll="false">
            <Level attribute="Year"/>
            <Level attribute="Month"/>
          </Hierarchy>
        </Hierarchies>
      </Dimension>
    </Dimensions>
    <MeasureGroups>
      <MeasureGroup name="ventas" table="fact">
        <Measures>
          <Measure name="Renta" column="renta" aggregator="sum">
          </Measures>
        <DimensionLinks>
          <ForeignKeyLink dimension="Time" foreignKeyColumn="id">
          </DimensionLinks>
        </MeasureGroup>
      </MeasureGroups>
    </Cube>
  </Schema>
```

INSTALANDO SAIKU

1. Descargar

2. Descomprimir

3. Abrir start-saiku.bat

Name	Date modified	Type	Size
data	2017-06-05 4:55 PM	File folder	
repository	2017-05-03 2:15 PM	File folder	
tomcat	2017-05-03 2:15 PM	File folder	
debug-start-saiku.bat	2017-05-03 2:15 PM	Windows Batch File	1 KB
debug-start-saiku.sh	2017-05-03 2:15 PM	Shell Script	1 KB
LICENSE	2017-05-03 2:15 PM	File	12 KB
README	2017-05-03 2:15 PM	File	2 KB
RELEASE_NOTES	2017-05-03 2:15 PM	File	2 KB
set-java.bat	2017-05-03 2:15 PM	Windows Batch File	1 KB
set-java.sh	2017-05-03 2:15 PM	Shell Script	1 KB
start-saiku.bat	2017-05-03 2:15 PM	Windows Batch File	1 KB
start-saiku.sh	2017-05-03 2:15 PM	Shell Script	1 KB
stop-saiku.bat	2017-05-03 2:15 PM	Windows Batch File	1 KB
stop-saiku.sh	2017-05-03 2:15 PM	Shell Script	1 KB

Para dejar de usarlo ejecutar stop-saiku.bat.
Si trabajan con Linux o macOS deben usar los .sh.

USANDO SAIKU

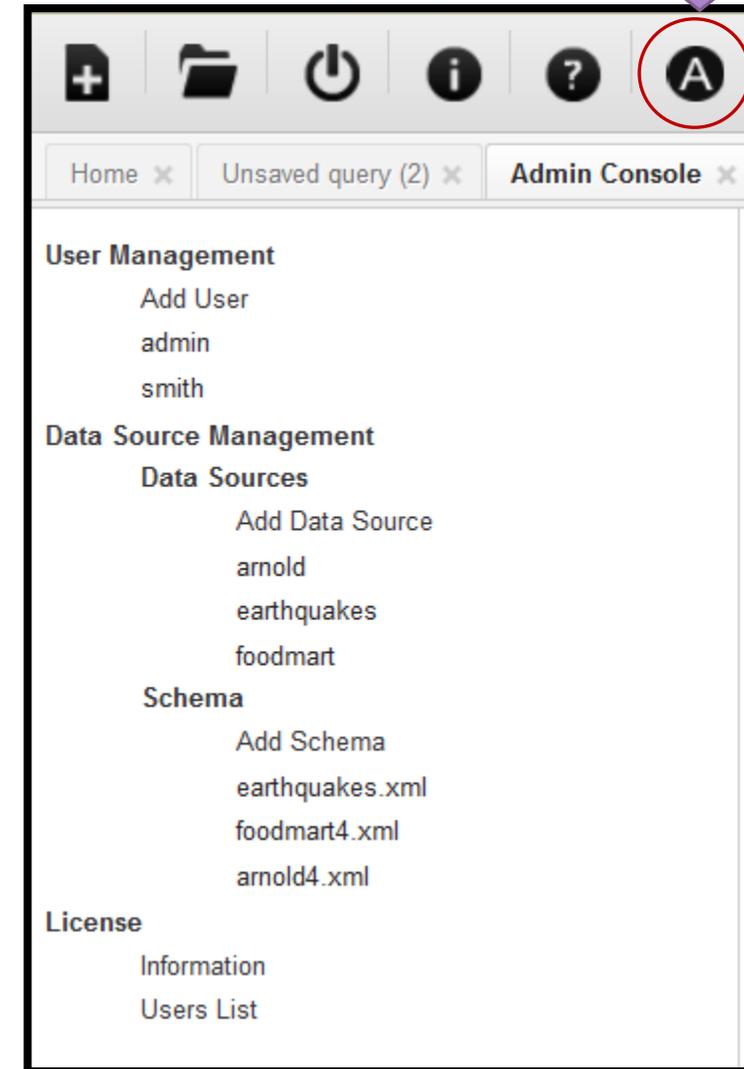
Primero deben agregar el Schema (*add schema*).

Luego, *Add Data Source*. Con los siguientes parámetros no deberían tener problemas:

- URL: `jdbc:mysql://localhost:3306/nombredb`
- Jdbc Driver: `org.mysql.jdbc.Driver`

En *Add Data Source*, usando *Refresh Cache* actualizan los cambios, pero **siempre** tienen que remover y volver a subir el schema.

Click en Admin Console



Cubos 

HR

Medidas Añadir

- ▼ HR
 - Org Salary
 - Count
 - Number of Employees
- ▼
 - Employee Salary
 - Avg Salary

Dimensiones

- ▶ Actual VS Budget
- ▼ Department
 - (All)
 - Department Description
 - (All)
 - Department Description
- ▶ Employee
- ▶ Store
- ▼ Time
 - Year
 - Quarter
 - Month

Medidas ▼

Org Salary

Columnas ▼

Year Time

Filas ▼

Department

Department Description

Filtro ▼

Year	1997	1998
Department Description	Org Salary	Org Salary
1	\$2,376.00	\$3,540.00
2	\$428.76	\$716.40
3	\$739.80	\$882.00
4	\$72.36	\$140.40
5	\$832.68	\$1,165.20
11	\$5,984.28	\$50,475.70
14	\$874.80	\$7,017.20
15	\$9,190.80	\$81,041.16
16	\$5,765.23	\$10,809.80
17	\$5,873.04	\$52,435.26
18	\$5,641.52	\$10,642.95
19	\$1,652.40	\$13,254.70

Información: 20:17 / 3 x 14 / 0.01s

Cubos

HR

Medidas Añadir

- HR
 - Org Salary
 - Count
 - Number of Employees
- Employee Salary
- Avg Salary

Dimensiones

- Actual VS Budget
- Department
 - (All)
 - Department Description
 - (All)
 - Department Description
- Employee
- Store
- Time
 - Year
 - Quarter
 - Month



Medidas

Org Salary

Columnas

Year Time

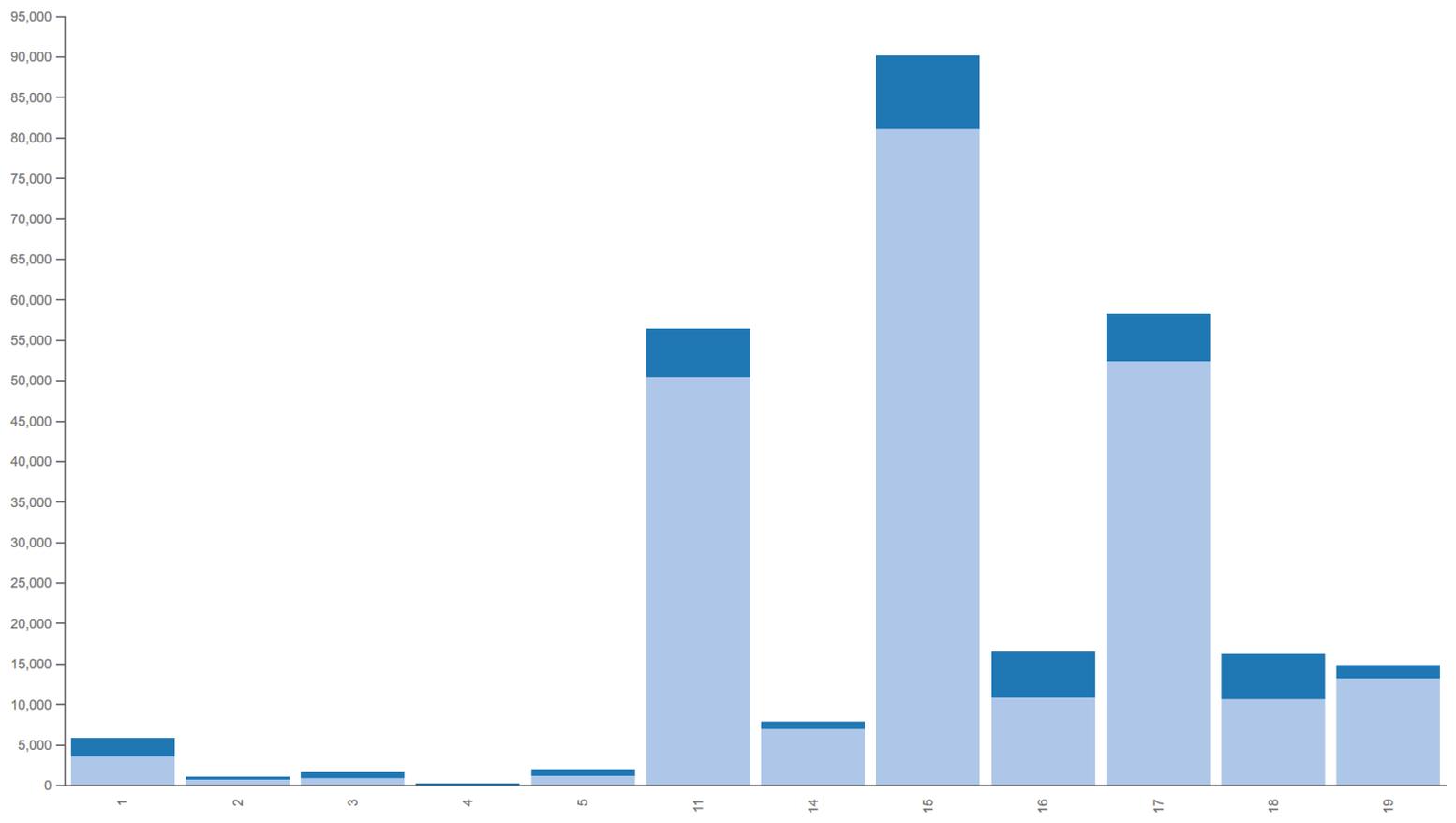
Filas

Department Department Description

Filtro

Información: 20:17 / 3 x 14 / 0.01s

1997 ~ Org Salary 1998 ~ Org Salary



Exportar

- Bar chart
- Line chart
- Table
- Area chart
- Line with markers
- Area with markers
- Scatter plot
- Pie chart
- Donut chart
- Gauge
- Settings
- Network
- Map
- Star

OTROS PRODUCTOS MONDRIAN

Nombre	Descripción
Pentaho Analyzer	Interfaz de usuario empresarial que proporciona un análisis interactivo con tablas y gráficos.
Pentaho Reporting	Herramienta que crea informes usando datos de Mondrian.
Community Dashboard Framework	Popular <i>framework</i> de código abierto para la creación de <i>dashboards</i> .
Saiku	Herramienta gratuita de código abierto que proporciona análisis interactivo con tablas y gráficos.



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

ALTERNATIVAS A MONDRIAN

Capítulo 6



cube.js

Una completa solución analítica de código abierto: incluye un **kit de desarrollo para frontend de visualización agnóstica** y una **API respaldada por una infraestructura de servidores analíticos**.

<https://cube.dev/>

¿Cómo partir? (Se necesitan conocimientos de NodeJS y Javascript)

<https://cube.dev/docs/getting-started>



fcfm

Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



DATA WAREHOUSING

Héctor Álvarez Gómez
Felipe Vildoso Castillo