



Sistemas de Información Geográfica y su relación con la Inteligencia de Negocios

Jose Aguilar

Unviersidad de Los Andes

Sistemas de Información Geográfica



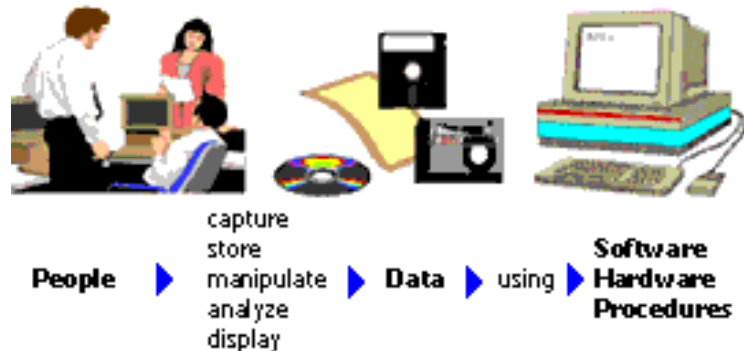
- Forman parte de los sistemas de información computarizados, conformados por conjuntos de programas que representan y gestionan grandes volúmenes de información.
- A diferencia de otros sistemas de información, los SIG permiten gestionar datos geo-referenciados.

Espacializados de acuerdo a un sistema de coordenadas geográficas o planas.



Sistemas de Información Geográfica

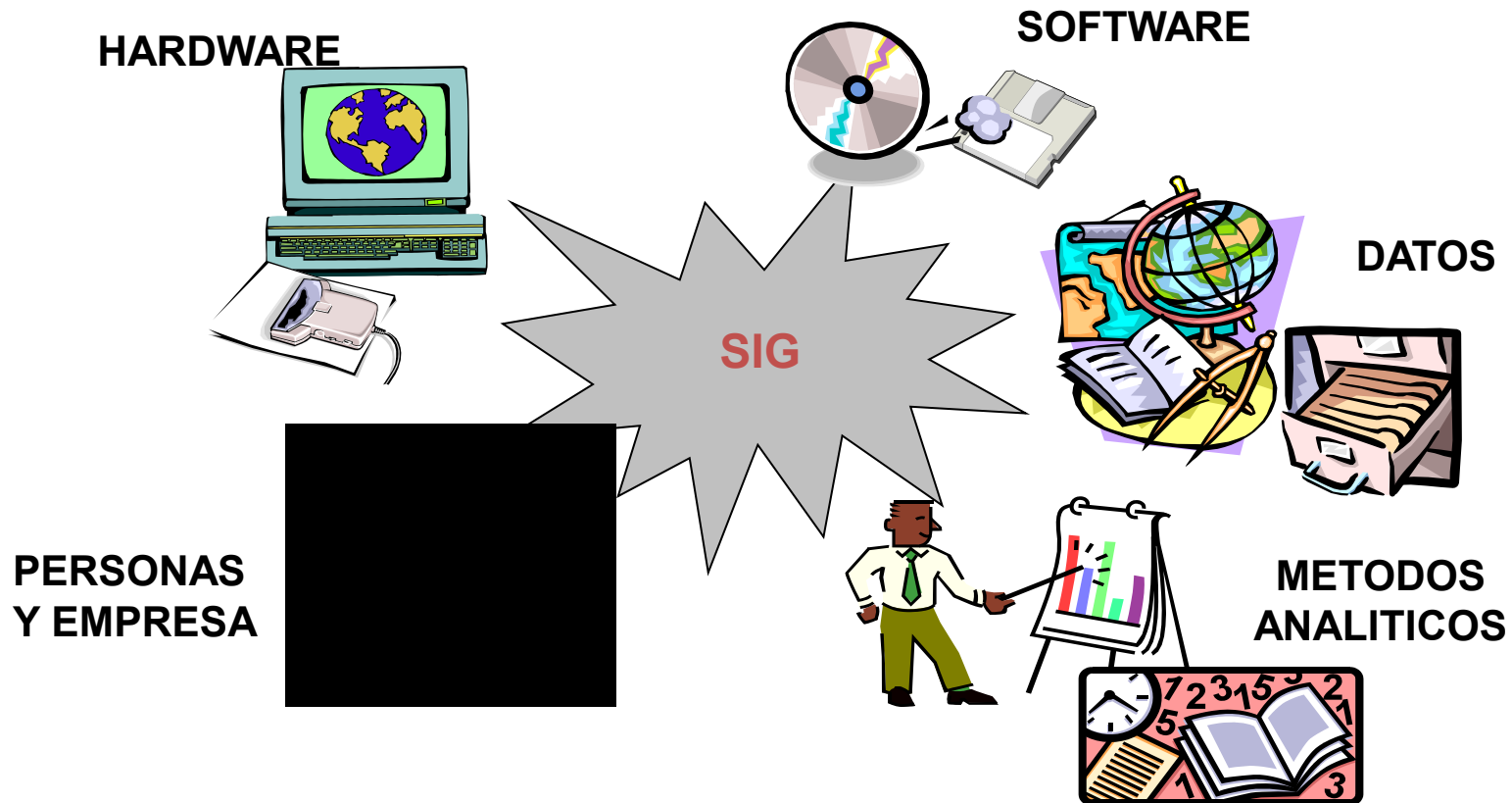
- Es un conjunto de procesos manuales o computarizados usados para almacenar y manipular datos geográficamente referenciados
- Es un apoyo para la toma de decisiones con el propósito de resolver problemas que consideren el medio ambiente



Sistemas de Información Geográfica



Combina cinco componentes:



todos organizados para analizar, manipular, procesar, almacenar , generar y visualizar todo tipo de información referenciada geográficamente.

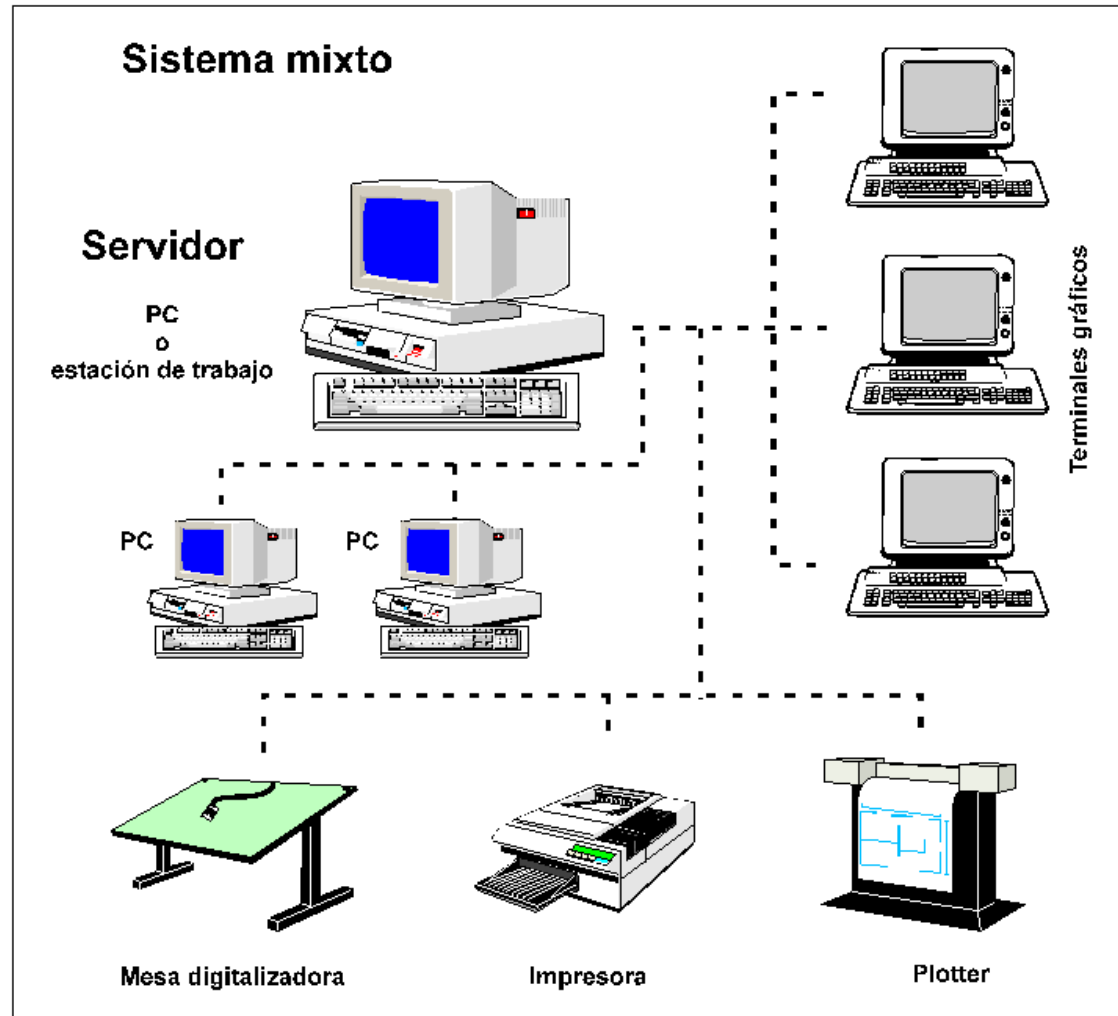
Sistemas de Información Geográfica



- Mientras otros Sistemas de Información contienen sólo datos alfanuméricos (nombres, direcciones, números de cuenta, etc.), las bases de datos de un S.I.G. han de contener además la **delimitación espacial** de cada uno de los objetos geográficos.
- Los SIG **trabajan a la vez con ambas partes** de información: su representación espacial y sus atributos temáticos asociados.
- Trabajan con cartografía y con bases de datos a la vez, uniendo ambas partes y constituyendo con todo ello una sola **base de datos geográfica**.

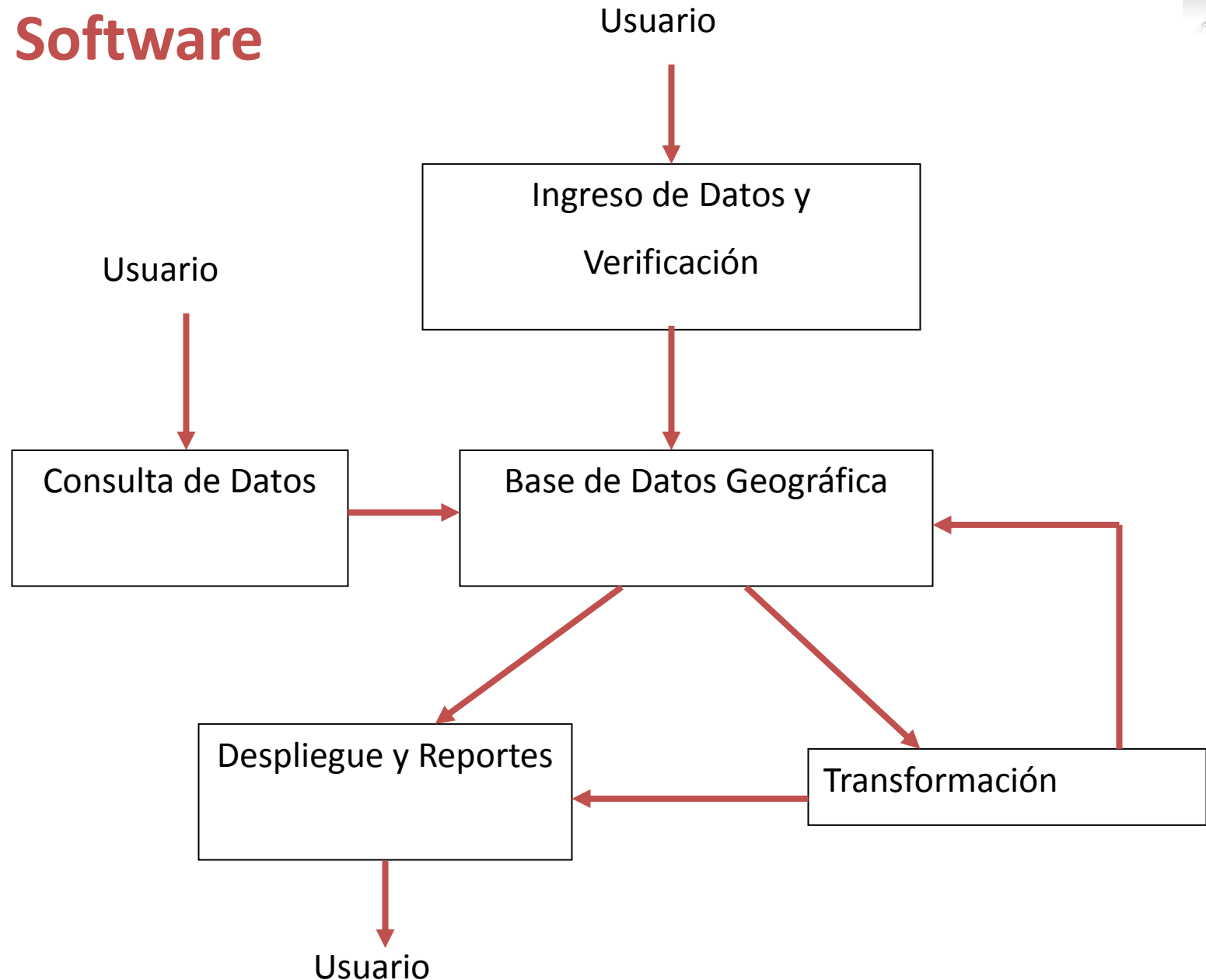
Sistemas de Información Geográfica

Hardware



Sistemas de Información Geográfica

Software

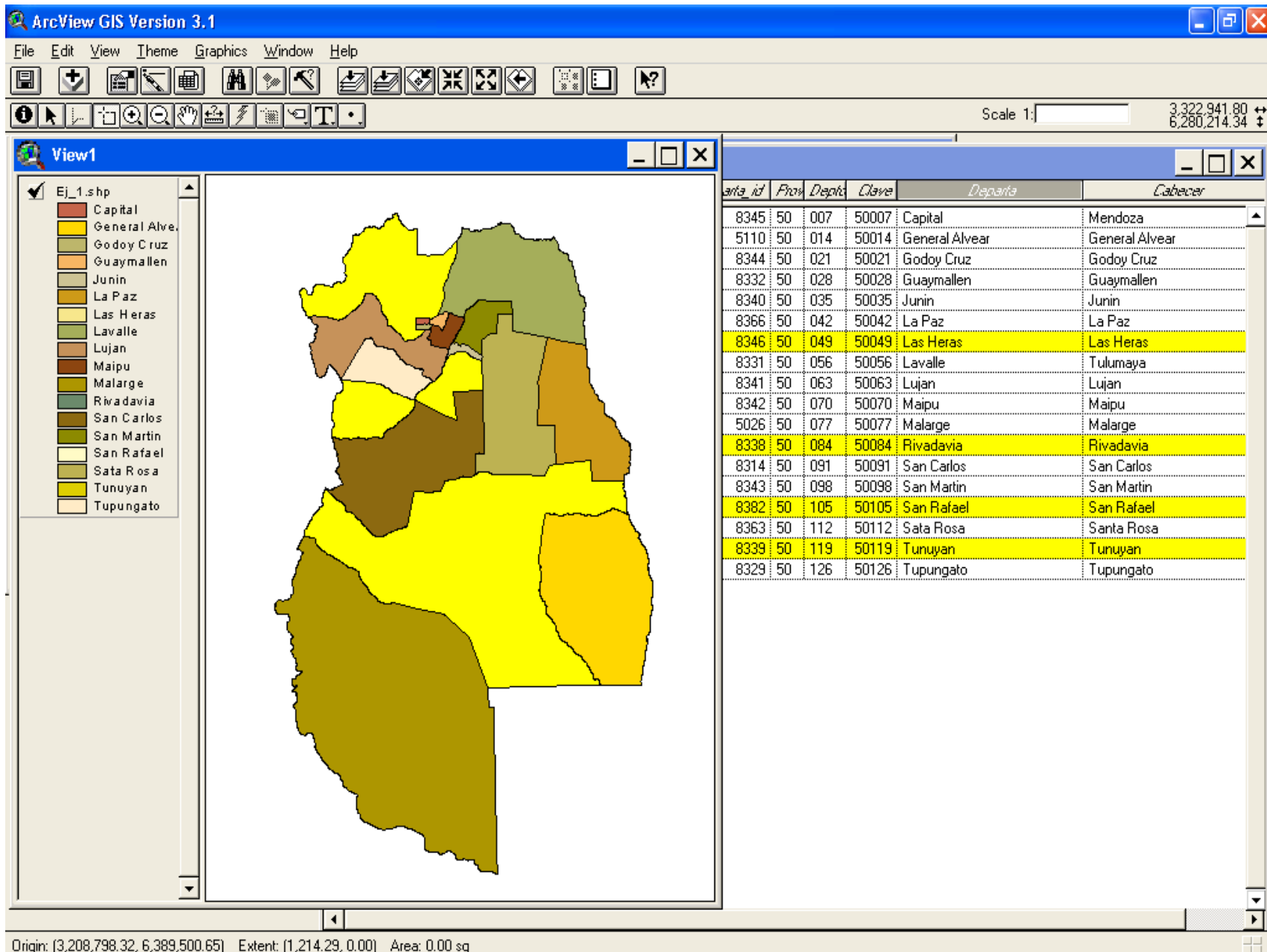


Sistemas de Información Geográfica

Software



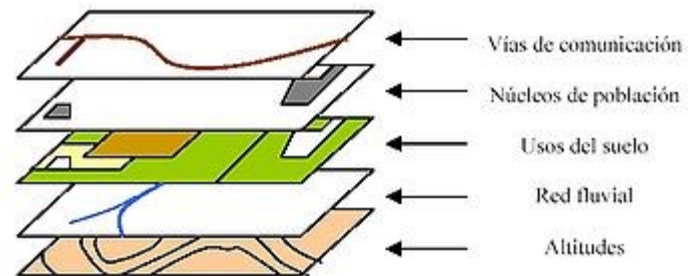
- Herramientas para la entrada y manipulación de información geográfica.
- Un sistema de manejo de base de datos.
- Herramientas que permitan realizar búsquedas geográficas, análisis y visualización.



Funcionamiento de los S.I.G



El SIG funciona como una base de datos con información geográfica (datos alfanuméricos) que se encuentra asociada por un identificador común a los objetos gráficos de un mapa digital. De esta forma, señalando un objeto se conocen sus atributos e, inversamente, preguntando por un registro de la base de datos se puede saber su localización en la cartografía.



Sistemas de Información Geográfica



Principales problemas que puede resolver un SIG:

- **Localización:** preguntar por las características de un lugar concreto.
- **Tendencia:** comparación entre situaciones temporales o espaciales distintas de alguna característica.
- **Rutas:** cálculo de rutas óptimas entre dos o más puntos.
- **Pautas:** detección de pautas espaciales.
- **Modelos:** generación de modelos a partir de fenómenos o actuaciones simuladas.

A nivel espacial, un SIG puede



Busca analizar fenómenos del territorio por medio de mapas con datos:

- Localización
- Distribución
- Asociación
- Interacción
- Evolución espacial.

A nivel espacial: LOCALIZACIÓN



- Entidades (atributos) que tienen una ubicación específica en el espacio geográfico.
- Se basa en el *espacio absoluto*: sitio (latitud y longitud) específico y fijo de emplazamiento sustentado en una topografía y condiciones geográficas particulares locales.
- De igual forma se basa en el *espacio relativo* posición cambiante respecto de otros sitios con los cuales se pueden establecer vínculos formales.

A nivel espacial: DISTRIBUCIÓN ESPACIAL E INTERACCIÓN ESPACIAL



- Conjunto de entidades de un mismo tipo que se reparten de determinada manera sobre el espacio geográfico, generalmente de manera heterogénea.
- Estructura un espacio relacional en el cual las localizaciones (sitios), distancias y vínculos resultan fundamentales en la diferenciación de espacios funcionales.

A nivel espacial: ASOCIACIÓN ESPACIAL



- Semejanzas entre distintas distribuciones espaciales.
- Poder encontrar diferentes áreas con homogeneidades propias permite poner límites al espacio geográfico.

Este trazado de límites sobre un mapa como resultado de la combinación de distribuciones y asociaciones espaciales se transforman en herramientas de planificación



SIG Y MODELADO DE DATOS

- En un SIG se deben modelar datos.
- La representación de la información geográfica se realiza estructurando datos espaciales
- Esta estructura de datos (cualitativos y/o cuantitativos), catalogada con variables de un territorio es la que provee información clave organizada, para la toma de decisiones
- Usan los programas SIG para hacer modelos de territorio.

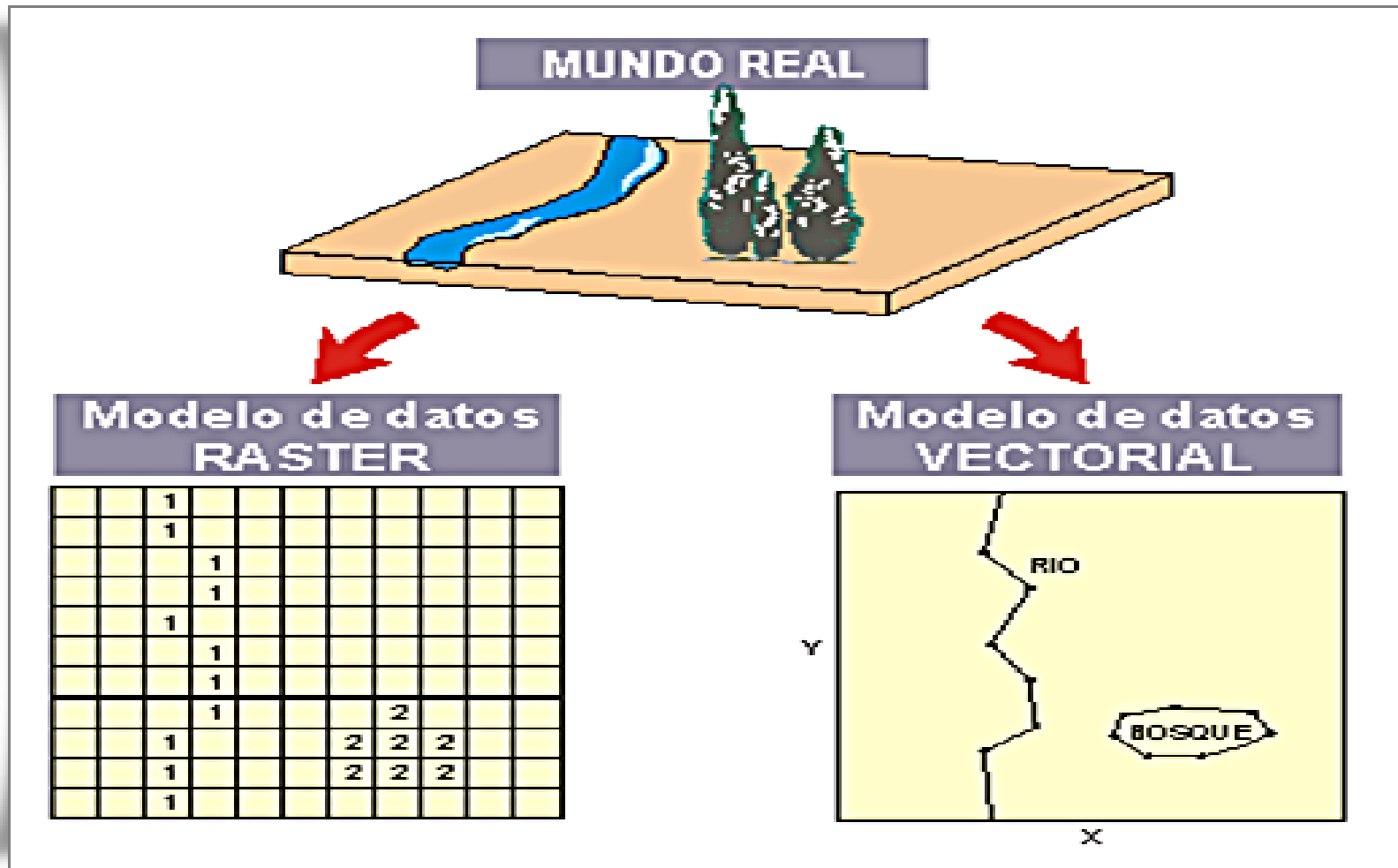


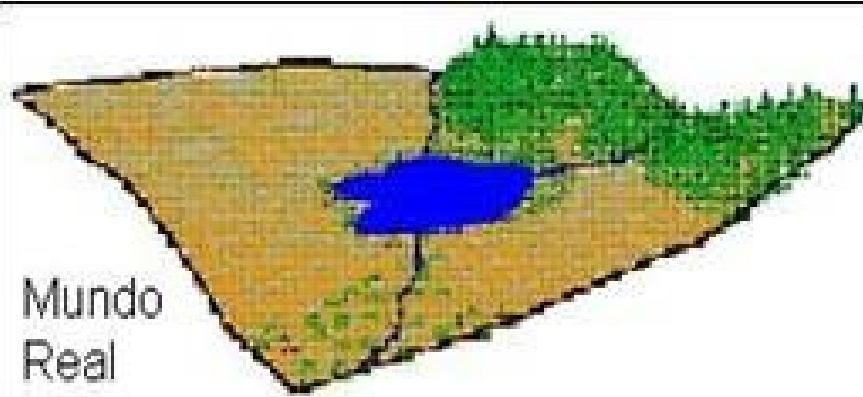
Datos espaciales básicos

- **Vectoriales:** Utilizan vectores, elementos puntuales, lineales o poligonales para delimitar a los objetos geográficos.
- **Raster:** utilizan una retícula regular para documentar los elementos geográficos que tienen lugar en el espacio. Representan variables continuas.

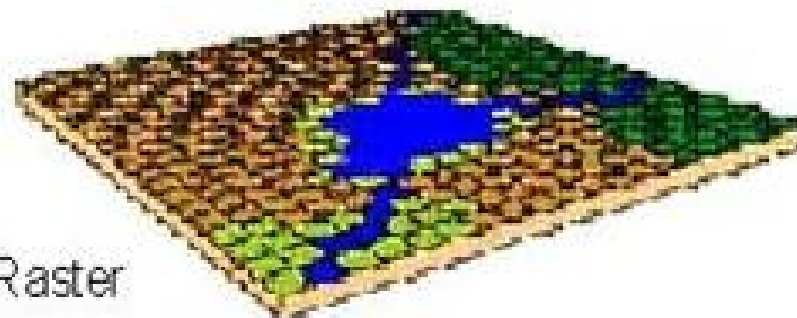


Datos espaciales básicos





Mundo
Real



Raster



Vector

Datos espaciales básicos: raster

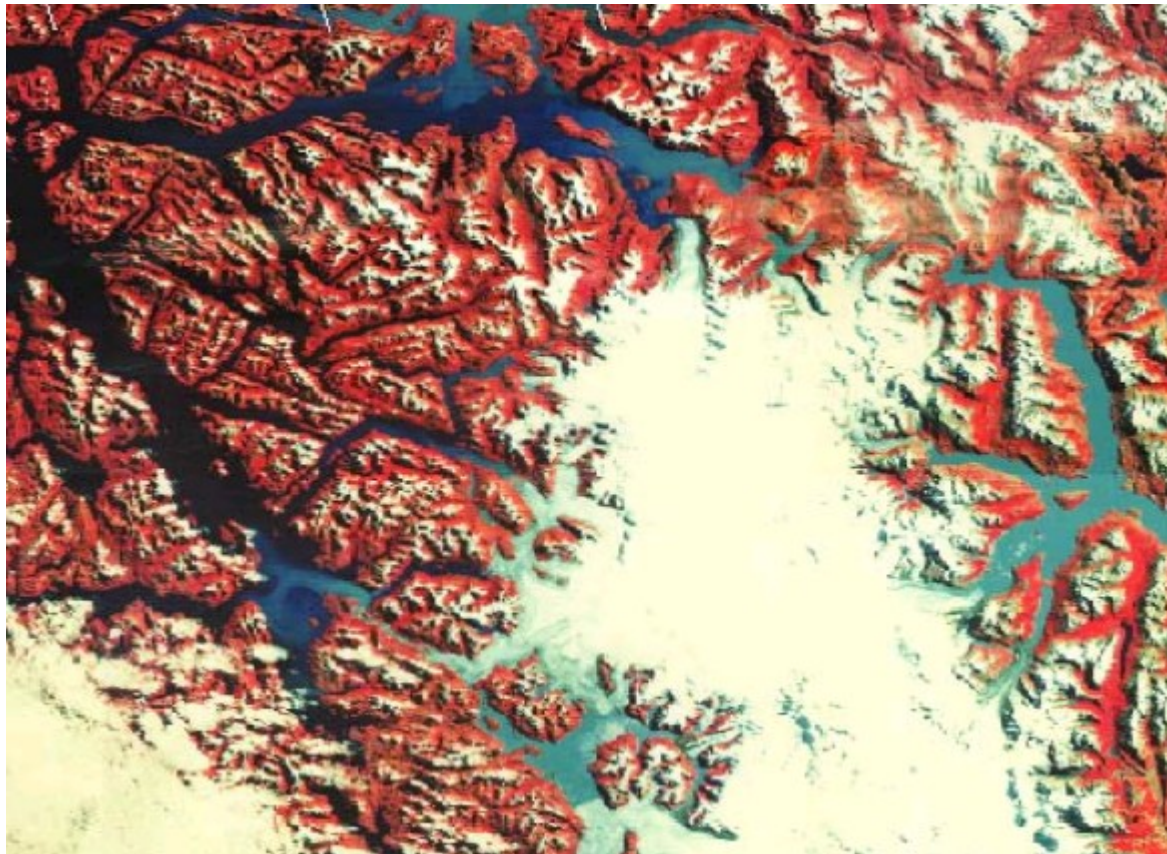


- Pueden ser
 - Imágenes de satélites (MODIS, LANDSAT, METEOSAT, SPOT, SAC-D).
 - Imágenes radar (en tierra o en altura)
 - Altimetría
 - Fotografía aérea

Datos espaciales básicos: raster



- imagen de satélite



R A S T E R



Image © 2006 MDA EarthSat
Image © 2006 DigitalGlobe

Google

Datos espaciales básicos: vectoriales



- Conjunto de vectores definidos por pares de coordenadas relativas a algún sistema de coordenadas.

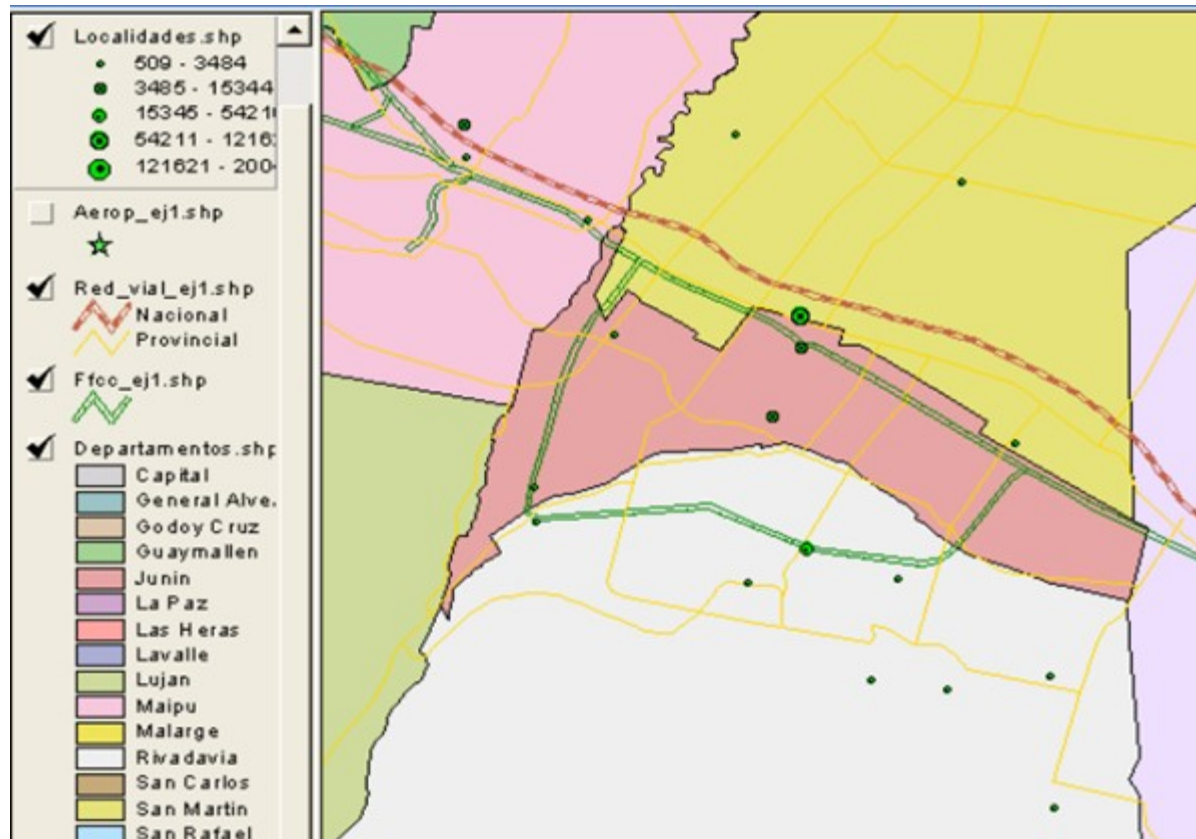
Ejemplos

- Un par de coordenadas (x,y) determinan un objeto geográfico punto.
- Dos puntos generan una línea.
- Por medio de una agrupación de líneas forman polígonos.

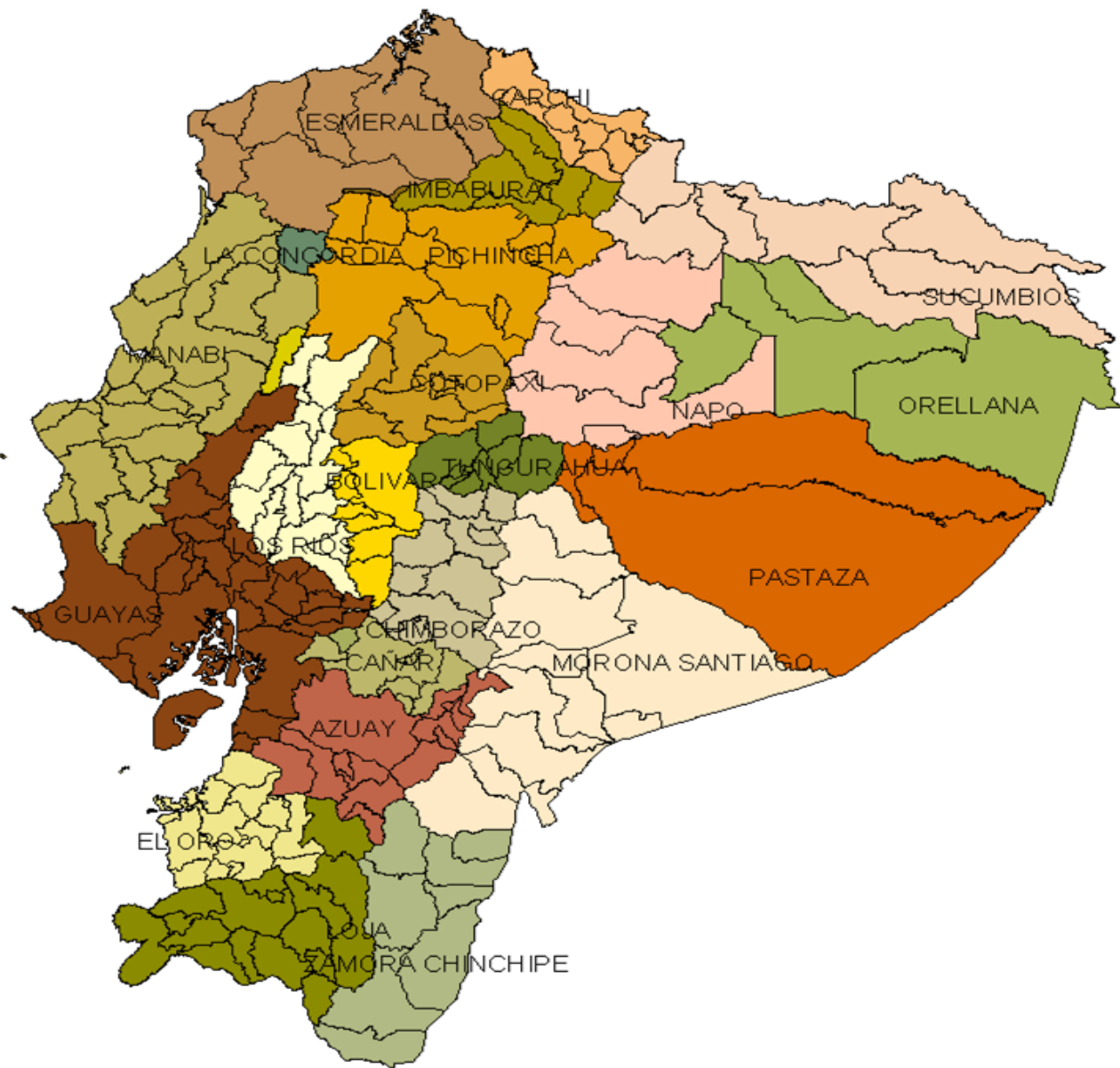
Datos espaciales básicos: Datos vectoriales



- Ejemplo de imagen vectorial

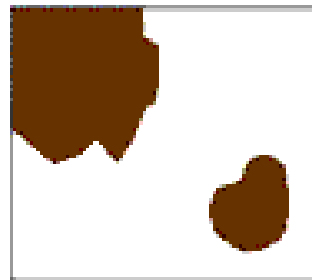
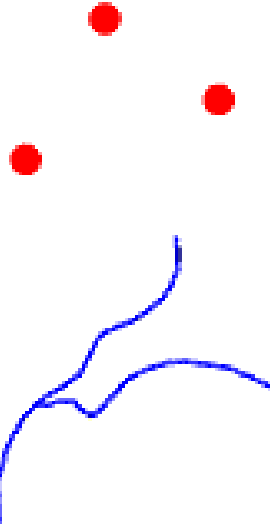


V E C T O R I A L





Vector

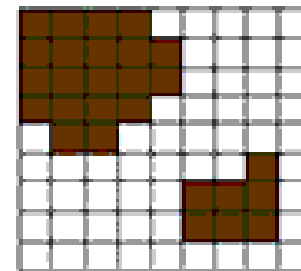
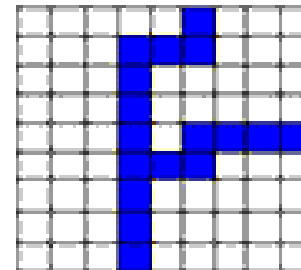
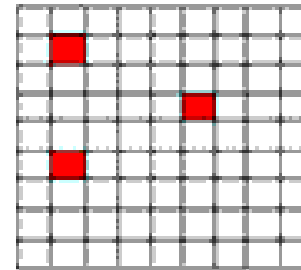


Points

Lines

Areas

Raster





	Vectores	Raster
Características	•Estructura de datos más compacta pero más compleja •Eficiente soporte de relaciones topológicas •Es más apto para el soporte gráfico •Es ineficiente para representar la alta variabilidad espacial •La manipulación y realce de imágenes no es hecho con efectividad en los dominios vectoriales •Las operaciones de superposición son más difíciles de implementar. •Más apto para pequeñas cantidades de datos •Sensitivo al tiempo de los datos de entrada y manejo •Alto grado de exactitud geométrica •Alto grado de representación gráfica	•Simple pero más compacta estructura de datos •Las operaciones de superposición son mas sencillas y eficientemente implementadas •La alta variación espacial es representada eficientemente •Variación geográfica continua •Favorece la manipulación eficiente y realce de imágenes digitales •Rápida captura de datos, frecuentemente de dispositivos de entrada de datos directamente •La salida puede ser menos placentera visualmente por que tiene apariencia de bloques más que de líneas suavizadas; puede ser mejorada pero hace excesivos los requisitos de almacenamiento.



	Vectores	Raster
Fuentes de datos	•Manual o digitalización automatizada de una copia impresa •Geometría de coordenadas (COGO) •Datos de terceros (DLG, DRG, TIGER,/ line) •GPS/Topografía •Mediciones fotogramétricas.	•Percepción remota •Ortofotografías digitales •Scanners •Camaras CCD
Aplicaciones	•Planeación y Aplicaciones Emergentes •Información de mapas basados en parcelas •Análisis de redes lineal y modelado •Infraestructura /Manejo de bienes	•Ambiente y manejo de recursos •Mapeo de ortofotos •Modelos de terreno •Uso de suelo/Cubierta de terreno •Estimación de producción de cultivos

Fuente: Varma, A. 2002. Data Sources and Measurement Technologies for Modeling.



Tipos de consultas para IN

- Localización ¿Qué hay en ?
- Condición ¿Dónde sucede que ?
- Tendencias ¿Qué ha cambiado ?
- Rutas ¿Cuál es el camino optimo ?
- Pautas ¿Qué pautas existen ?
- Modelos ¿Qué ocurriría si ?

PostgreSQL y PostGIS



- Es un **módulo** que añade soporte de objetos geográficos a la base de datos objeto-relacional PostgreSQL, convirtiéndola en una **base de datos espacial** para su utilización en SIG.

PostGIS ha sido certificado en 2006 por el Open Geospatial Consortium (OGC) lo que garantiza la interoperabilidad con otros sistemas .



APLICACIONES DE LOS S.I.G



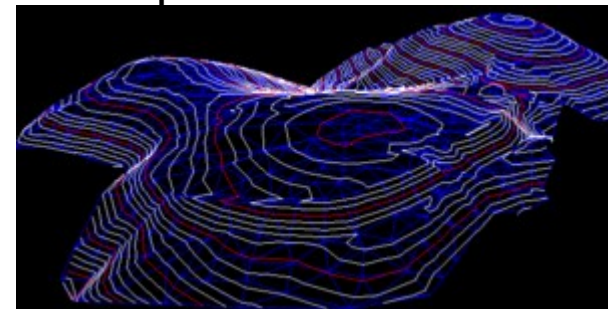
Impacto ambiental



Planificación urbana



Planificación



Cartografía

Aplicaciones de los SIG en procesos de IN

<i>Organization</i>	<i>GIS Application</i>
Pepsi Cola Inc., Super Value, Acordia Inc.	Used in site selection for new Taco Bell and Pizza Hut restaurants; combining demographic data and traffic patterns
CIGNA (health insurance)	Uses GIS to answer such questions as, How many CIGNA-affiliated physicians are available within an 8-mile radius of a business?
Western Auto (a subsidiary of Sears)	Integrates data with GIS to create a detailed demographic profile of store's neighborhood to determine the best product mix to offer at the store
Sears, Roebuck & Co.	Uses GIS to support planning of truck routes
Health maintenance organizations	Tracks cancer rate and that of other diseases to determine expansion strategy and allocation of expensive equipment in their facilities
Wood Personnel Services (employment agencies)	Maps neighborhoods where temporary workers live to locate marketing and recruiting cities
Wilkening & Co. (consulting services)	Designs optimal sales territories and routes for their clients, reducing travel costs by 15 percent
CellularOne Corporation	Maps its entire cellular network to identify clusters of call disconnects and to dispatch technicians accordingly
Sun Microsystems	Manages leased property in dozens of places worldwide
Consolidated Rail Corporation	Monitors the condition of 20,000 miles of railroad track and thousands of parcels of adjoining land
Federal Emergency Management Agency	Assesses the damage of hurricanes, floods, and other natural disasters by relating videotapes of the damage to digitized maps of properties
Toyota (and other car manufacturers)	Combines GIS and GPS as a navigation tool Directs drivers to destinations via the best route



Inteligencia de Negocio

PENTAHO

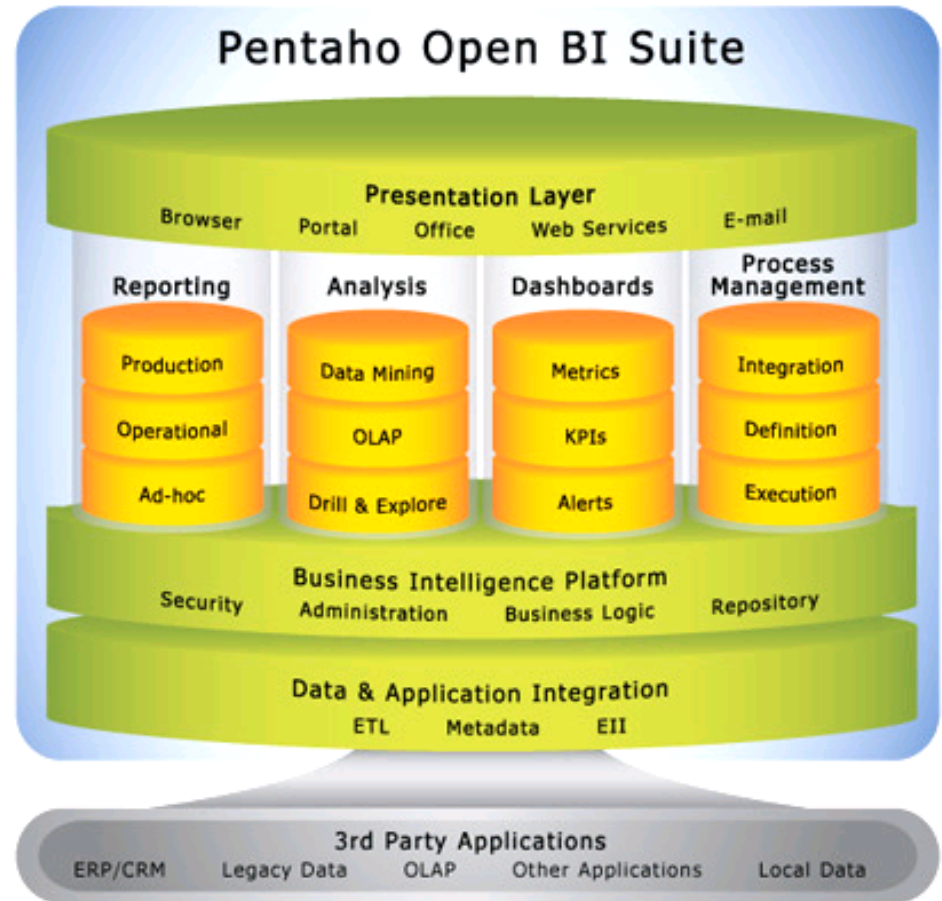
Jose Aguilar
CEMISID, Escuela de Sistemas
Facultad de Ingeniería
Universidad de Los Andes
Mérida, Venezuela

Pentaho: La solución Open Source Business Intelligence

Plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basados en procesos.

Pentaho incluye:

- Reporting
- Análisis
- Dashboards
- Data Mining



PENTAHO



- Pentaho ha sido desarrollado desde el año 2004.
- Open Source:
 - Tiene una comunidad de usuarios
 - Corre bien bajo múltiples plataformas (Windows, Linux, Macintosh, Solaris, etc.
- Tiene un paquete completo para presentar informes, ETL para la gestión de almacenamiento de datos (data warehousing), minería de datos, OLAP y dashboard.

Bajo Java

PENTAHO



Velocidad de desarrollo



Marketing dashboard in less than 1 day



2 weeks time to market



8 weeks time to market



Fully rolled out in budget in 4 months

Innovación y escalabilidad



Music files from 20,000 sources



Operational reports at all 1000 retail stores



Analyzing buying patterns of 5 million members



Analytics on 500,000 patients records

Servicio a cliente



"... better functionality and more support"



"... top-notch professional support"



"Pentaho support is as good as its software"



"... a great partner through every phase of our project"

Valor total



75% lower acquisition costs



€350K+ cost saving



Less than 1 month ROI



"...ROI was almost immediate."

PENTAHO

cuenta con:

- Jfree Report: Motor para reportes
- Mondrian: Motor para cubos
- Weka: Motor para minería de datos
- Plataforma Web para publicar y visualizar la información

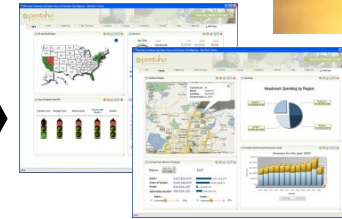


Pero además



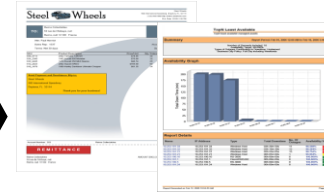
Dashboards

ETL: Kettle (Spoon/Pan)



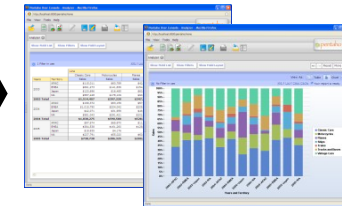
Reportes

Report Designer



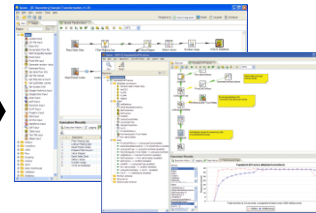
Análisis

Weka
ETL: Kettle (Spoon/Pan)



Datos

High Cube designer, WorkBench
ETL: Kettle (Spoon/Pan)



Data Mining

Weka



Componentes independientes



http://localhost:8080/pentaho/Login;jsessionid=E17F50339D3098831AC80B0422DD8EA3



Google



Pentaho Business Intelligence



Página



Herram



What's New At Pentaho



Login

Valid Users:

Joe ([es_85] Admin) ▼

User:

joe

Password:

••••••••

Login

Reset

Version: Pentaho BI Platform 1.7.1.1117



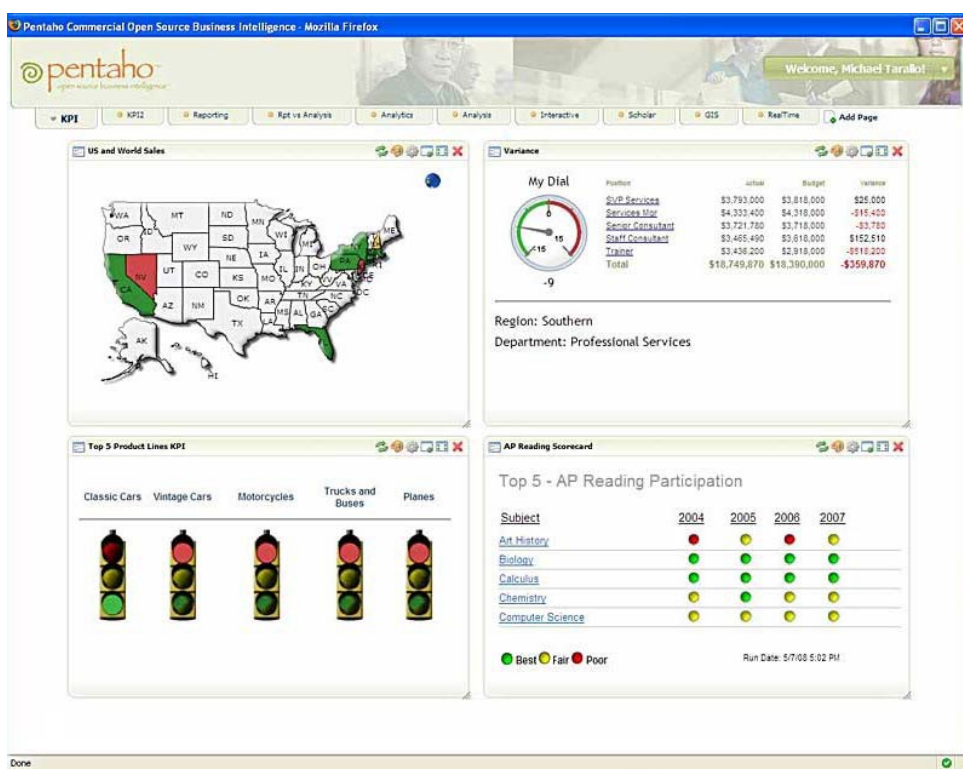
Supplied free of charge with no support, no certification, no maintenance, no warranty and no indemnity by Pentaho or its Certified Partners. [Click here](#) for Support, Certification, or Indemnification. Copyright © 2004 - 2008 Pentaho Corporation, All rights reserved.

Pentaho Dashboard

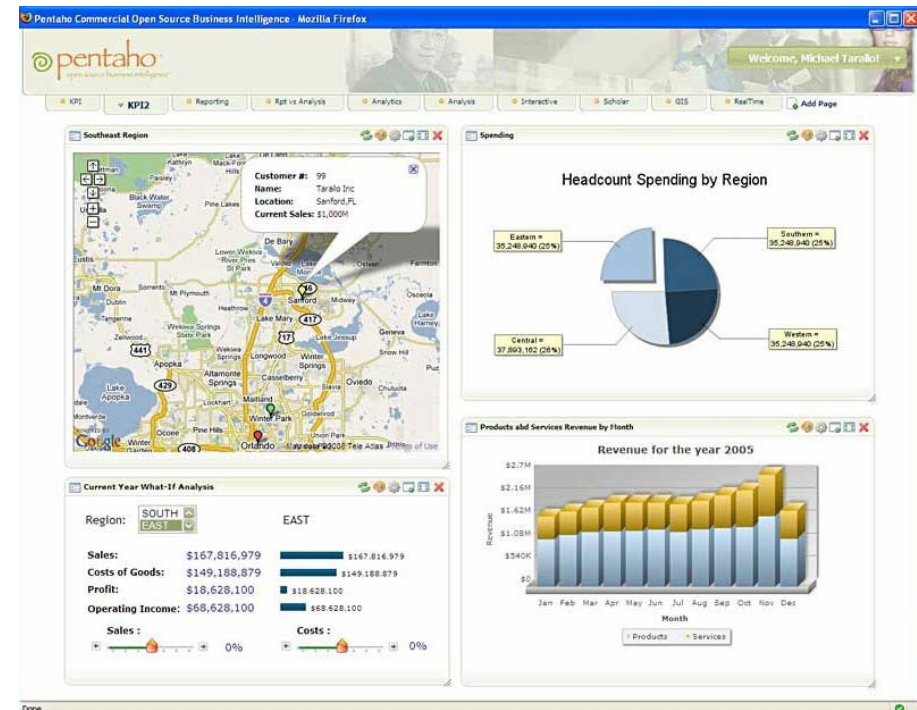


- Brinda a los usuarios la información crítica que necesitan para comprender y mejorar el desempeño organizacional.
- Proporciona una visión inmediata del rendimiento individual, departamental, o de la empresa, mediante la entrega de métricas claves a través de una interfaz visual atractiva e intuitiva.

Pentaho Dashboard



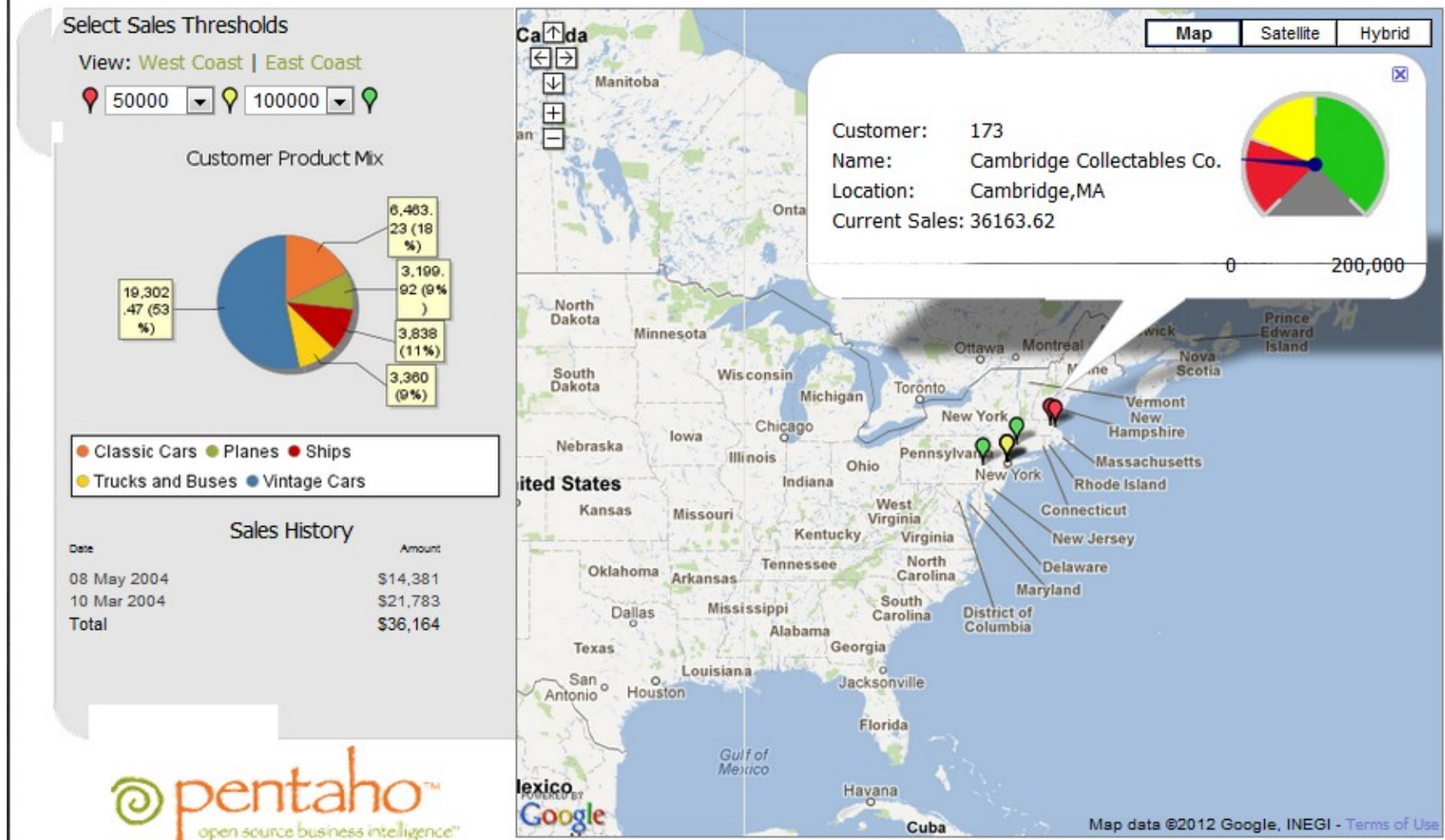
Interactive



Pentaho Dashboard: Geo Location



Pentaho Google Maps Dashboard



Reporting



Region: Central

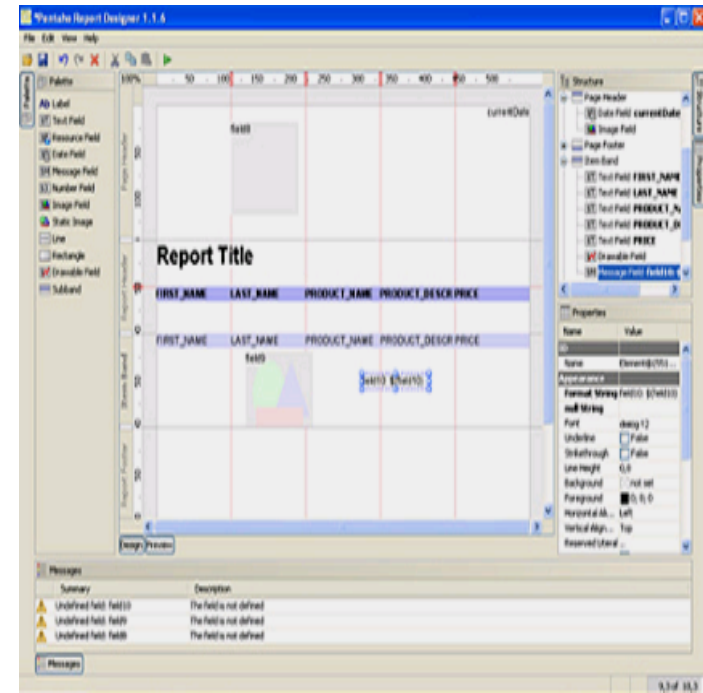
Department	Position	Actual	Budget	Variance
Executive Management				
	SVP Partnerships	\$367,415	\$392,100	\$24,685
	SVP WW Operations	\$476,000	\$725,887	\$249,887
	SVP Strategic Development	\$383,242	\$403,405	\$20,163
	CEO	\$549,625	\$522,250	-\$27,375
	Total	\$1,776,282	\$2,043,642	\$267,360

Department	Position	Actual	Budget	Variance
Finance				
	Controller	\$570,373	\$577,070	\$6,697
	Payroll	\$367,415	\$432,100	\$64,685
	Administrative Assistant	\$827,861	\$760,990	-\$66,871
	IS	\$570,759	\$577,346	\$6,587
	CFO	\$770,272	\$719,855	-\$50,417
	Total	\$3,106,680	\$3,067,361	-\$39,319

Department	Position	Actual	Budget	Variance
Human Resource				
	Sexual Harassment	\$530,473	\$538,570	\$8,097
	EOE	\$530,207	\$538,380	\$8,173
	HR Generalists	\$856,190	\$771,225	-\$84,965
	HR Training	\$397,473	\$443,570	\$46,097
	Administration	\$549,625	\$552,250	\$2,625
	SVP HR	\$574,896	\$570,300	-\$4,596
	Total	\$3,438,863	\$3,414,295	-\$24,568

Department	Position	Actual	Budget	Variance
Marketing & Communication				
	Graphics	\$782,375	\$728,500	-\$53,875
	Writer	\$405,985	\$459,650	\$53,665
	Analyst Relations	\$383,375	\$443,500	\$60,125
	Press Relations	\$497,296	\$524,872	\$27,576
	CMO	\$827,861	\$760,990	-\$66,871
	Product Marketing Mgr	\$693,531	\$665,040	-\$28,491

La solución de Reporting que plantea Pentaho, incluida dentro de su suite, es **Jfree Report**.



Reporting



Permite a las organizaciones fácilmente acceder, formatear y entregar reportes a los empleados, clientes y socios.

Product Lines by Territory - Report Pack

Shipped Orders by Territory - Report Pack

Product Lines within Territory

Territory: APAC -

Product Line	Total	Quantity Ordered
Ships	\$38,393.48	428
Trains	\$9,907.07	139
Classic Cars	\$411,956.24	3,852
Trucks and Buses	\$145,665.69	1,380
Planes	\$121,426.20	1,330
Vintage Cars	\$364,538.92	3,897
Motorcycles	\$189,818.23	1,852
Total	\$1,281,705.83	12,878

Territory: EMEA -

Product Line	Total	Quantity Ordered
Trucks and Buses	\$500,978.82	4,655
Trains	\$138,506.18	1,544
Planes	\$491,654.86	5,513
Ships	\$427,264.68	4,808
Vintage Cars	\$851,080.90	9,660

Shipped Orders by Territory

Countries are counted per Territory

Quantity Ordered & Total Revenue are being Totalled

Territory : NA

Country	Quantity Ordered	Total
USA	32,923	3,372,204.3
Canada	2,293	224,078.56
2	35,216	\$3,596,282.86

Territory : Japan

Country	Quantity Ordered	Total
Singapore	1,524	172,989.68
Philippines	961	94,015.73
Japan	1,842	188,167.81
Hong Kong	596	48,794.36
4	4,923	\$503,957.58

Territory : EMEA

Country	Quantity Ordered	Total
UK	4,584	428,472.21
Switzerland	1,078	117,713.56
Sweden	1,239	135,043.08

Steel

Customer

Muscle Machine Inc

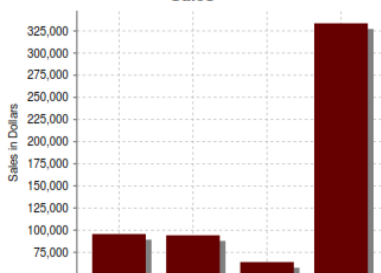
Mini Gifts Distributors Ltd.

Online Diecast Creations Co.

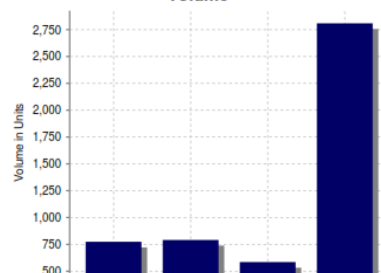
Other Customers

Total **\$ 587,428** **4,959** **56.79%**

Sales



Volume



Mondrian permite

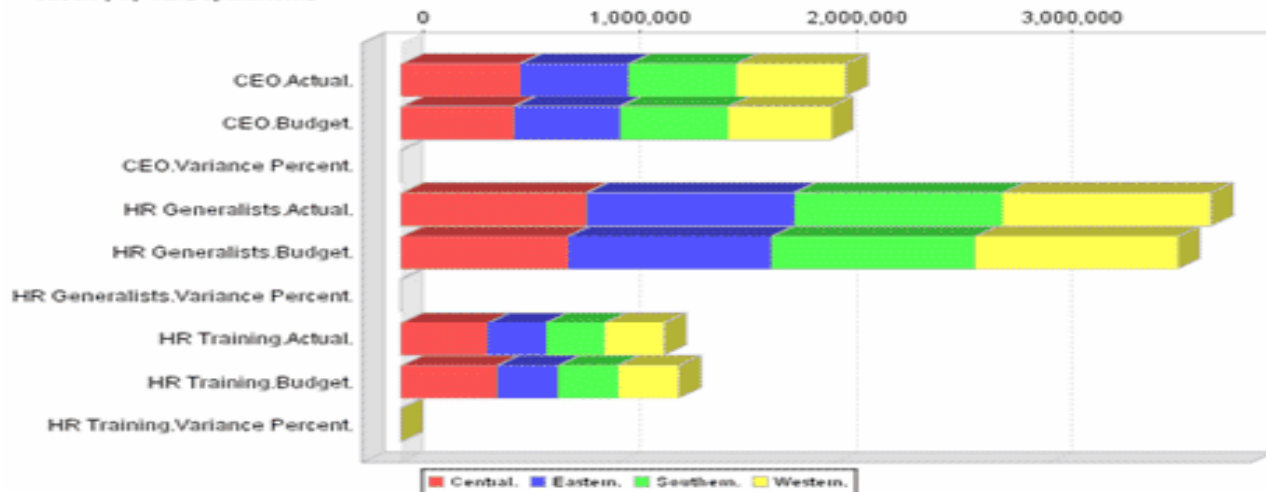
- Alto rendimiento, análisis interactivo de volúmenes grandes o pequeños de la información
- Exploración "Dimensional" de datos, por ejemplo, el análisis de las ventas por línea de productos, por regiones, para un período de tiempo
- Convertir consultas multidimensionales en (MDX) en SQL para hacer consultas
- Consultas de alta velocidad
- Cálculos avanzados utilizando expresiones de cálculo del lenguaje MDX

Mondrian permite

		Region			
Positions	Measures	Central	Eastern	Southern	Western
CEO	Actual	\$549,625.00	\$500,000.00	\$500,000.00	\$500,000.00
	Budget	\$522,250.00	\$488,750.00	\$498,750.00	\$478,750.00
	Variance Percent	-5.24%	-2.30%	- .25%	-4.44%
HR Generalists	Actual	\$856,190.00	\$961,000.00	\$961,000.00	\$961,000.00
	Budget	\$771,225.00	\$940,158.00	\$940,158.00	\$938,158.00
	Variance Percent	-11.02%	-2.22%	-2.22%	-2.43%
HR Training	Actual	\$397,473.00	\$271,200.00	\$271,200.00	\$271,200.00
	Budget	\$443,570.00	\$279,674.00	\$279,674.00	\$277,674.00
	Variance Percent	10.39%	3.03%	3.03%	2.33%

Slicer: [(All)=All Departments]

Slicer: (All)=All Departments





OLAP en Pentaho

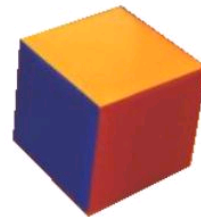
Permite realizar todas las **típicas funcionalidades de un sistema OLAP** a una gran velocidad

Hace uso de diversas Bases de Datos:

- Oracle
- DB2
- SQL-Server
- **MySQL**
- **Postgre**



OLAP en Pentaho



Pentaho OLAP

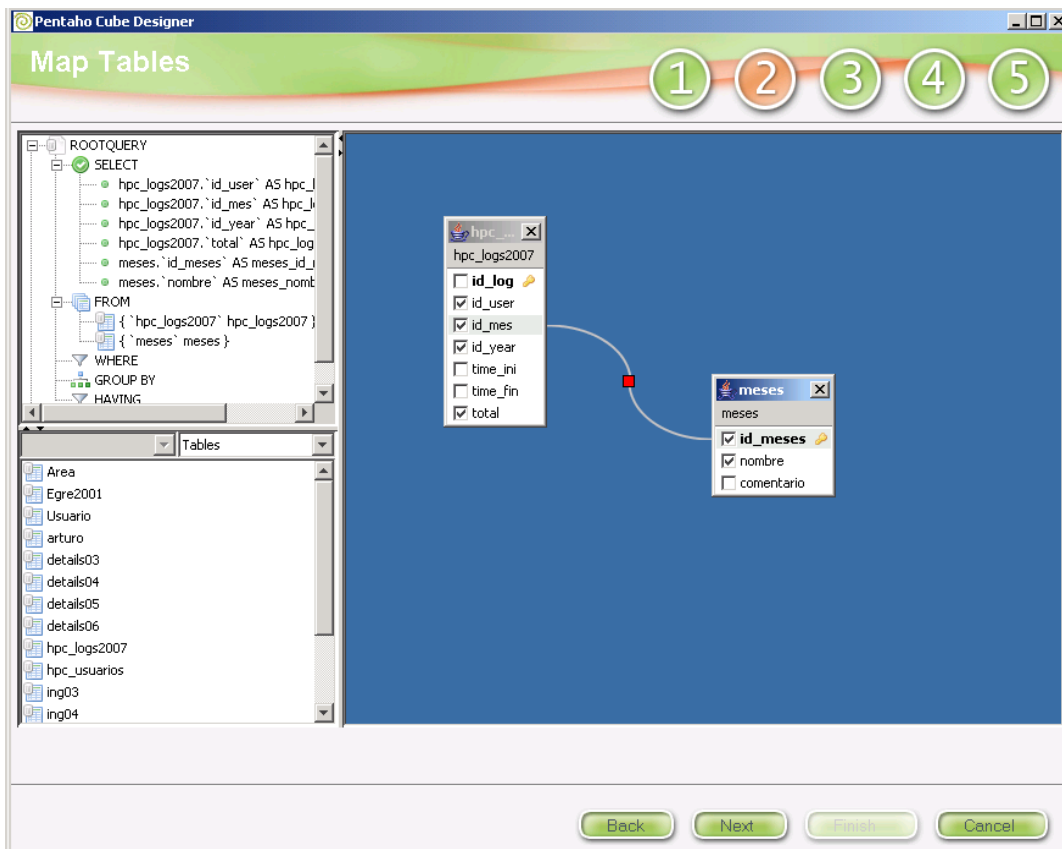
Puntos Acumulados



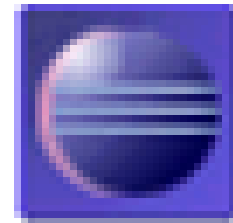
Permite diseñar cubos dinámicos, para un análisis de la información mucho más rápida y detallada, que apoya la Toma de Decisiones dentro de una organización



OLAP en Pentaho



Cube Designer
incorpora “arrastrar y
soltar” para un
manejo más fácil.





OLAP en Pentaho

Create Dimensions

1
2
3
4
5

Create Dimensions

Create dimensions from the source fields using "Add New Dimension" button. You can add levels by using the Arrow buttons. Once dimensions are created dimension/hierarchy/level to view/change the attributes at the bottom section. Click "Add Property" button to add member properties for a level

Source fields

hpc_logs2007.id_user
 hpc_logs2007.id_mes
 hpc_logs2007.id_year
 hpc_logs2007.total
 meses.id_meses
meses.nombre

Add New Dimension

==>

<==

Dimensions

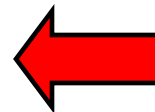
- hpc_logs2007.id_user
 - hpc_logs2007.id_user
 - hpc_logs2007.id_user
- hpc_logs2007.id_year
 - hpc_logs2007.id_year
 - hpc_logs2007.id_year
- meses.nombre
 - meses.nombre
 - meses.nombre

Attributes and Properties

Add Property

Remove Property

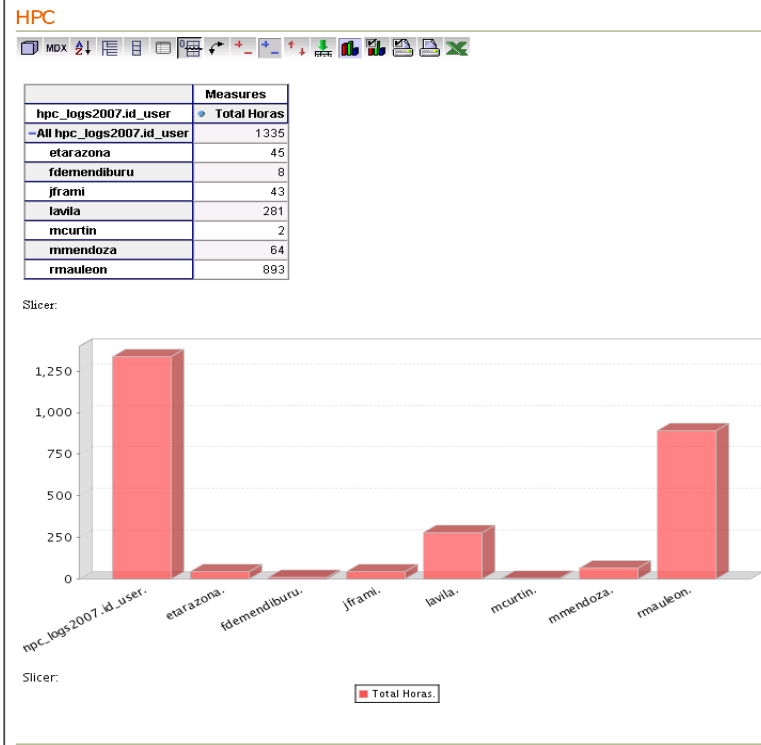
Property	Value
name	meses.nombre
nameColumn	meses.nombre
uniqueMembers	false



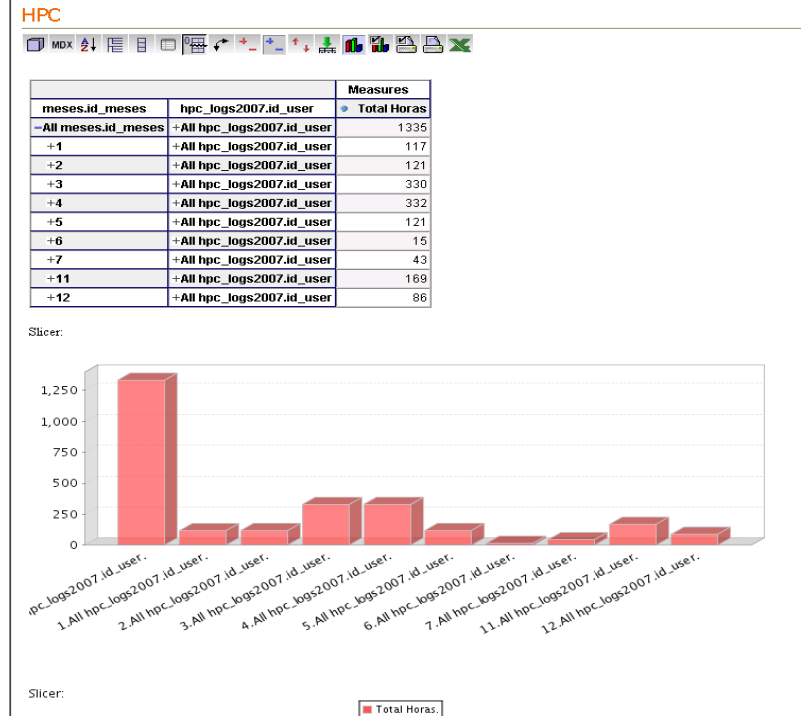
Crea las
dimensiones del
Cubo que luego
seran mostradas en
la Suite de Pentaho



Gráficas de Resultados



Estadística por Usuario



Estadística por Mes



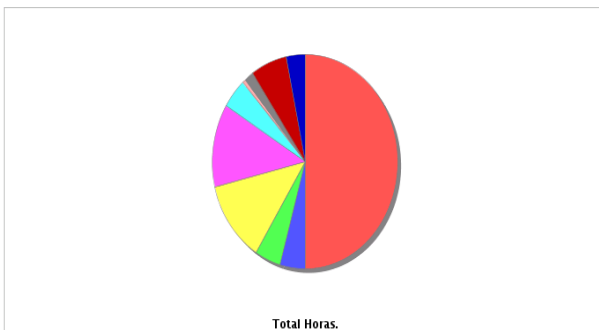
Gráficas de Resultados

HPC



meses.id_meses	hpc_logs2007.id_user	Measures
-All meses.id_meses	+All hpc_logs2007.id_user	Total Horas
+1	+All hpc_logs2007.id_user	117
+2	+All hpc_logs2007.id_user	121
+3	+All hpc_logs2007.id_user	330
+4	+All hpc_logs2007.id_user	332
+5	+All hpc_logs2007.id_user	121
+6	+All hpc_logs2007.id_user	15
+7	+All hpc_logs2007.id_user	43
+11	+All hpc_logs2007.id_user	169
+12	+All hpc_logs2007.id_user	86

Slicer:



Slicer:

● All meses.id_meses.All hpc_logs2007.id_user
 ● 1.All hpc_logs2007.id_user
 ● 2.All hpc_logs2007.id_user
 ● 3.All hpc_logs2007.id_user
 ● 4.All hpc_logs2007.id_user
 ● 5.All hpc_logs2007.id_user
 ● 6.All hpc_logs2007.id_user
 ● 7.All hpc_logs2007.id_user
 ● 11.All hpc_logs2007.id_user
 ● 12.All hpc_logs2007.id_user

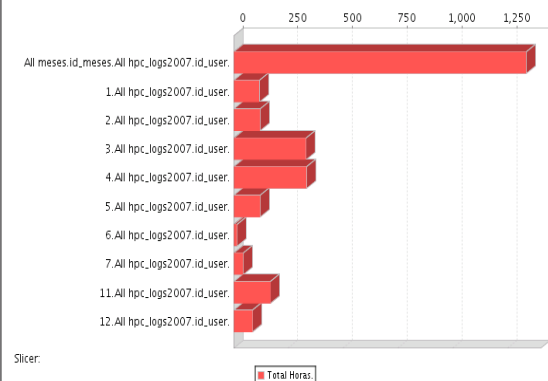
*Todos los gráficos
y estadísticas se
generan
dinamicamente,
sólo con solicitar
la información los
gráficos van
cambiando*

HPC



meses.id_meses	hpc_logs2007.id_user	Measures
-All meses.id_meses	-All hpc_logs2007.id_user	Total Horas
+1	+All hpc_logs2007.id_user	117
+2	+All hpc_logs2007.id_user	121
+3	+All hpc_logs2007.id_user	330
+4	+All hpc_logs2007.id_user	332
+5	+All hpc_logs2007.id_user	121
+6	+All hpc_logs2007.id_user	15
+7	+All hpc_logs2007.id_user	43
+11	+All hpc_logs2007.id_user	169
+12	+All hpc_logs2007.id_user	86

Slicer:



Slicer:

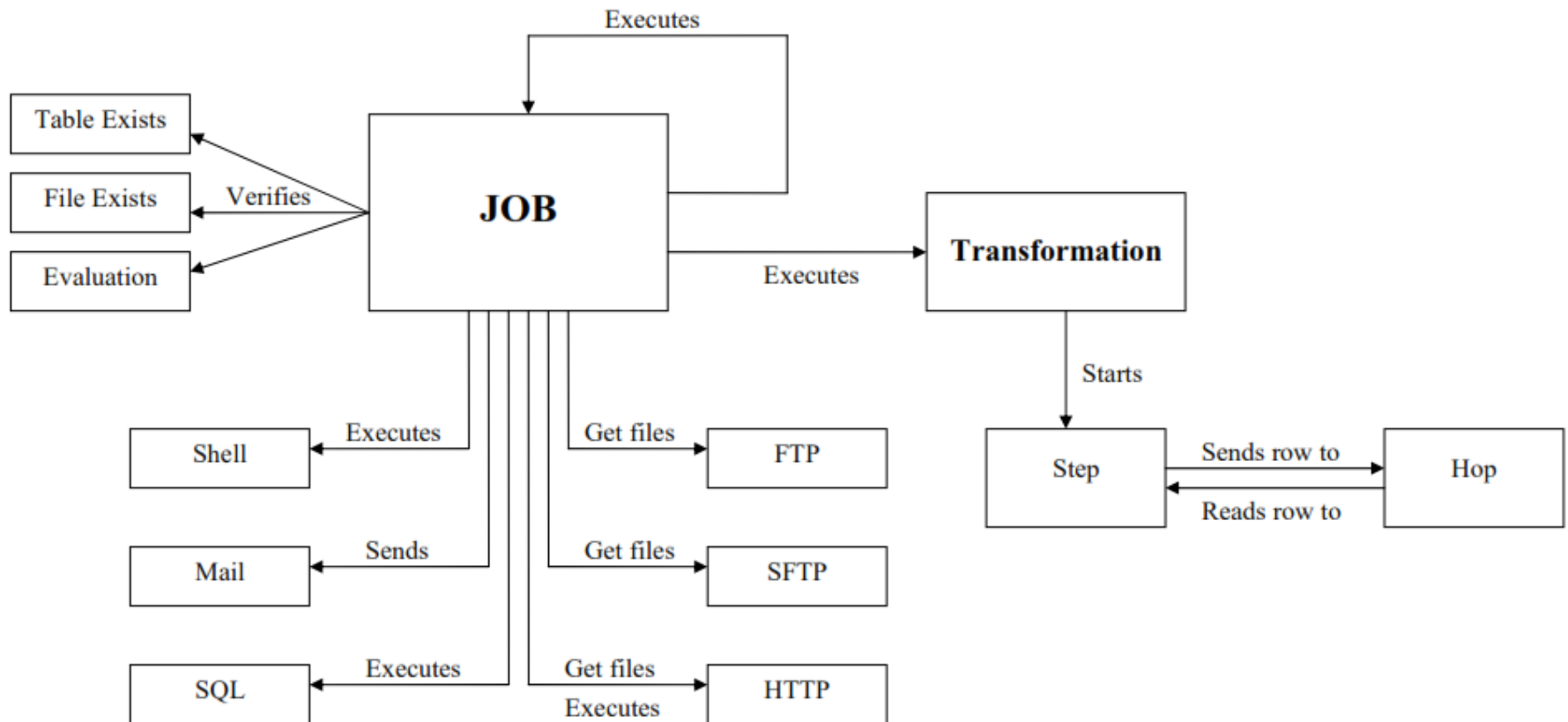
Kettle

- tiene capacidades de acceder, limpiar e integrar datos desde cualquier lugar de la organización.
- ofrece una poderosa mecanismo de extracción, transformación y carga (ETL), con un diseño intuitivo y una arquitectura escalable basada en estándares probados.

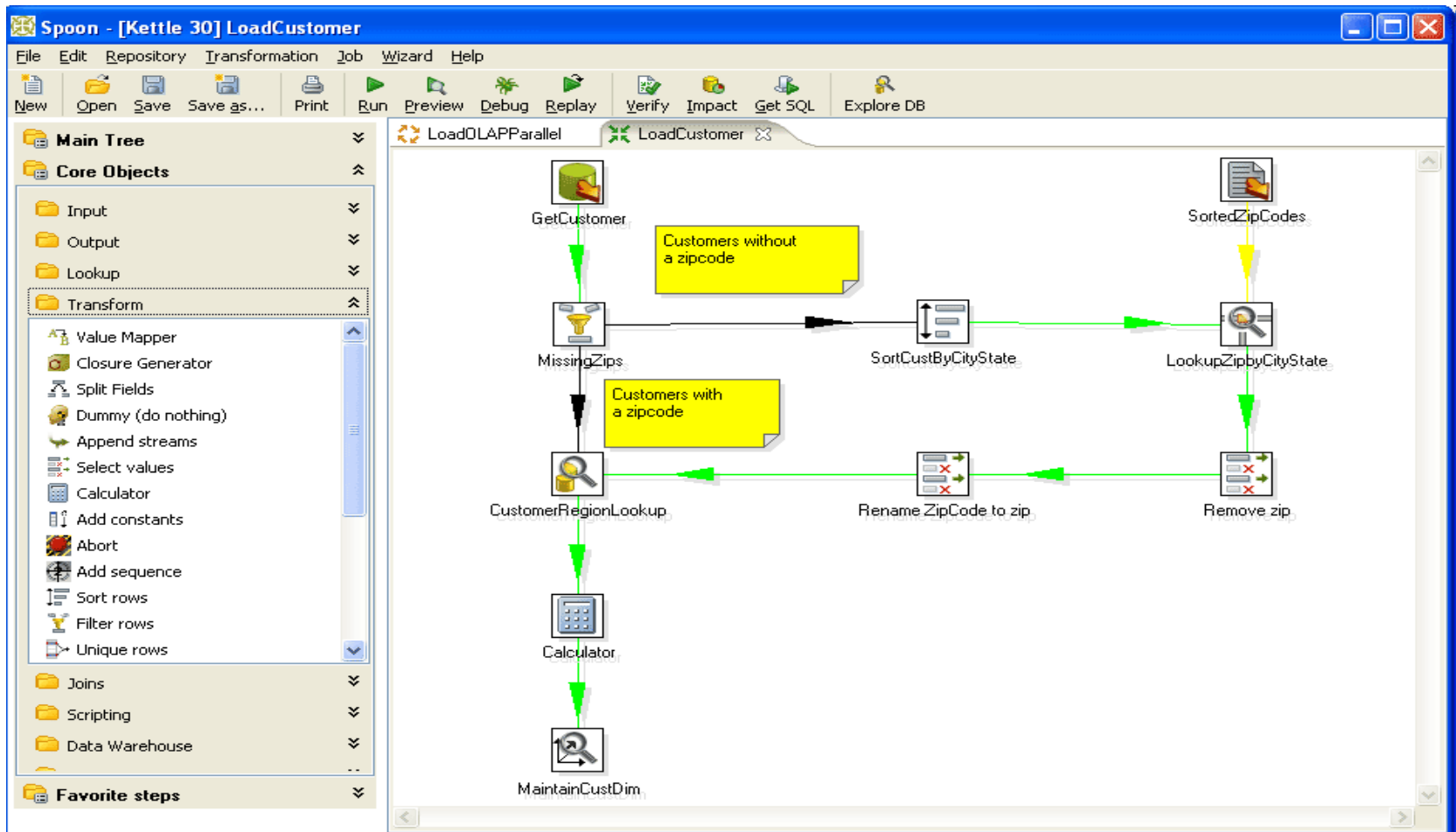
Suite

- Kitchen : Job Execution
- Spoon : GUI Designer
- Carte : Cluster Web Server
- Pan : Transformation Execution

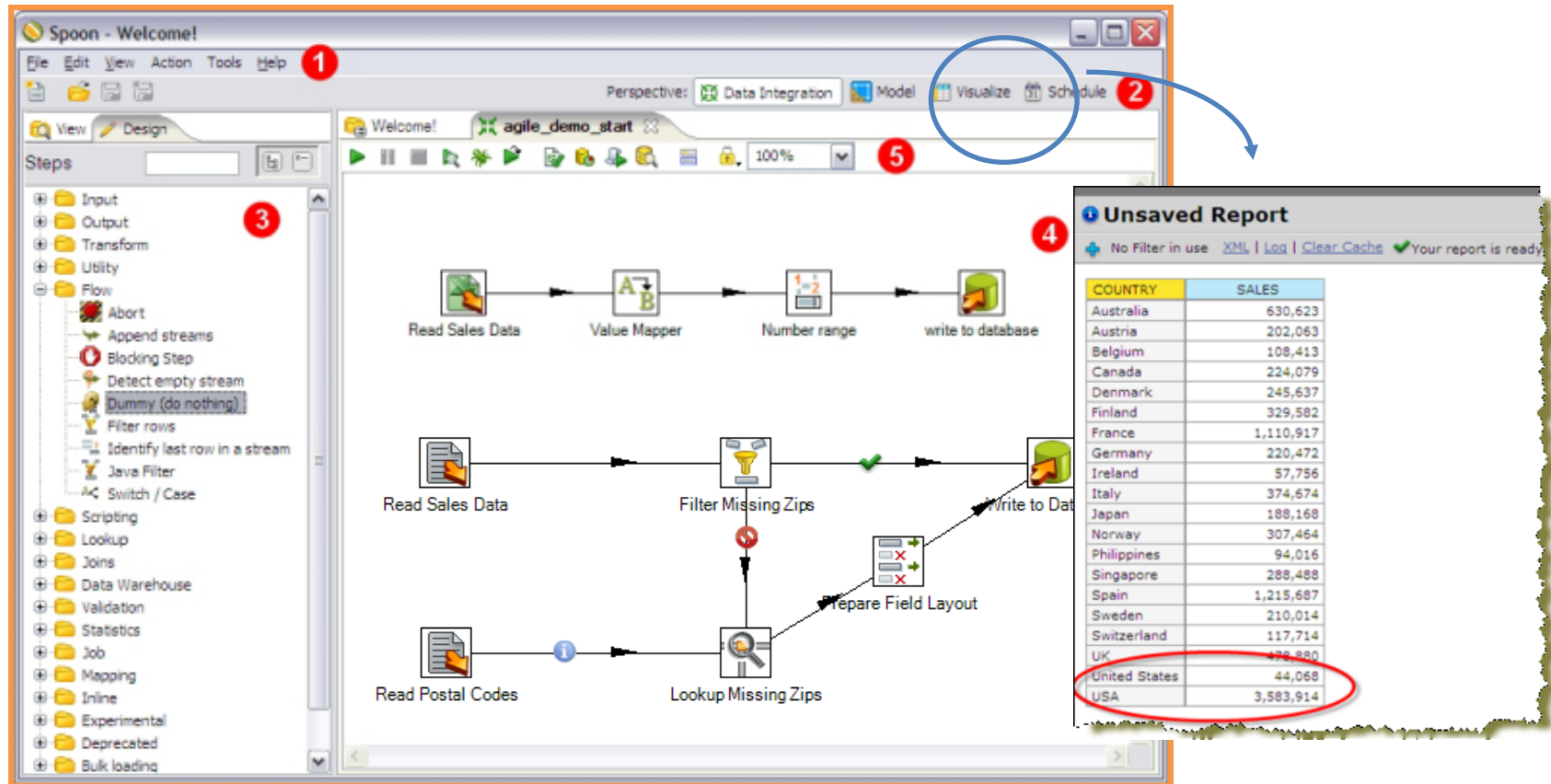
Modelo de Kettle



Kettle



Integración de datos

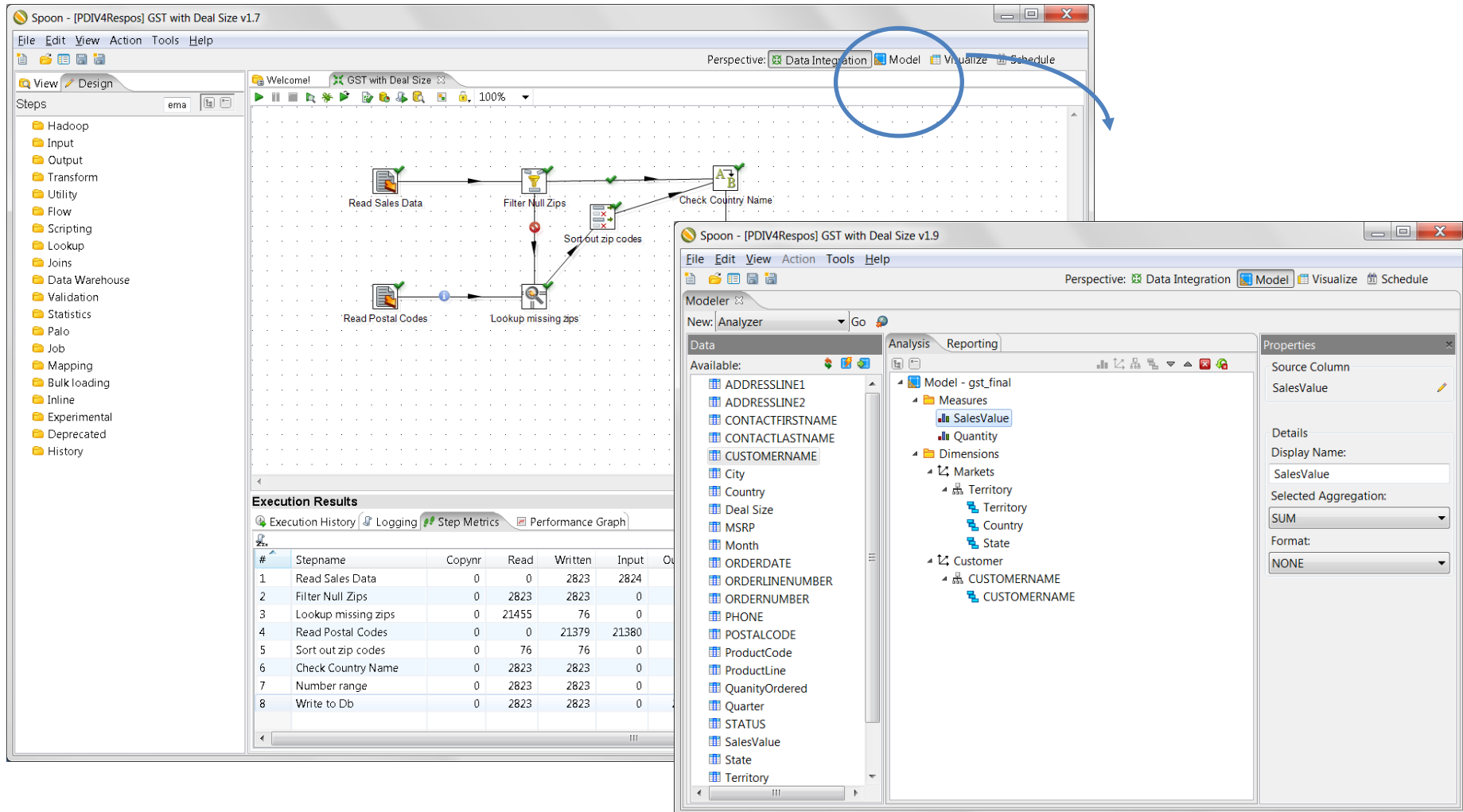


The screenshot displays the Pentaho Spoon interface for data integration. The main workspace shows a job design with two parallel paths. The top path consists of 'Read Sales Data', 'Value Mapper', 'Number range', and 'write to database'. The bottom path consists of 'Read Postal Codes', 'Lookup Missing Zips', 'Filter Missing Zips', 'Repare Field Layout', and 'Write to Dat'. The 'Filter Missing Zips' step is highlighted with a red circle and a blue arrow pointing to the 'Visualize' button in the top toolbar. The 'Visualize' button is also circled in blue. The 'Steps' panel on the left lists various transformation and job steps, with 'Dummy (do nothing)' selected. The 'Unsaved Report' window on the right shows a table of sales data by country, with the 'United States' and 'USA' rows circled in red.

Unsaved Report

COUNTRY	SALES
Australia	630,623
Austria	202,063
Belgium	108,413
Canada	224,079
Denmark	245,637
Finland	329,582
France	1,110,917
Germany	220,472
Ireland	57,756
Italy	374,674
Japan	188,168
Norway	307,464
Philippines	94,016
Singapore	288,488
Spain	1,215,687
Sweden	210,014
Switzerland	117,714
UK	458,880
United States	44,068
USA	3,583,914

Integración de datos: crear metadatos



The screenshot displays the Pentaho Spoon interface for a data integration job titled "[PDIV4Respos] GST with Deal Size v1.7". The job is designed in the "Data Integration" perspective, showing a sequence of steps: Read Sales Data, Filter Null Zips, Sort out zip codes, Check Country Name, Read Postal Codes, Lookup missing zips, and Write to Db. A blue circle highlights the "Model" tab in the top right corner, with an arrow pointing to a detailed view of the job's metadata.

The detailed view shows the "Model" tab for the job, displaying a list of available data fields and a tree structure of the job's metadata. The "Available" list includes fields such as ADDRESSLINE1, ADDRESSLINE2, CONTACTFIRSTNAME, CONTACTLASTNAME, CUSTOMERNAME, City, Country, Deal Size, MSRP, Month, ORDERDATE, ORDERLINENUMBER, ORDERNUMBER, PHONE, POSTALCODE, ProductCode, ProductLine, QuantityOrdered, Quarter, STATUS, SalesValue, State, and Territory.

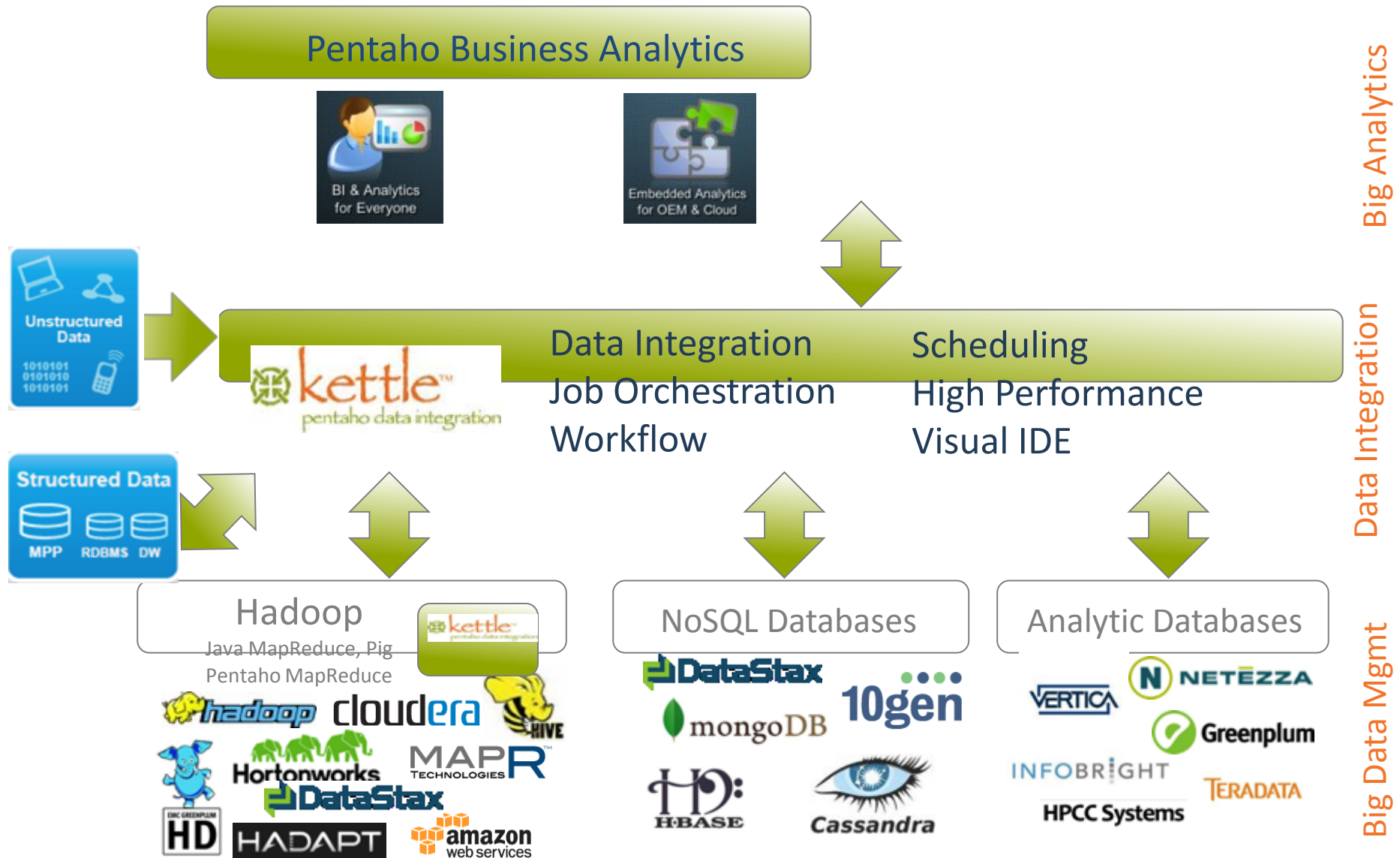
The "Analysis" tab shows a tree structure of the job's metadata, including Measures (SalesValue), Dimensions (City, Country, State, Customer), and a list of fields (CUSTOMERNAME, CUSTOMERNAME).

The "Properties" tab shows the "Source Column" set to "SalesValue" and the "Selected Aggregation" set to "SUM".

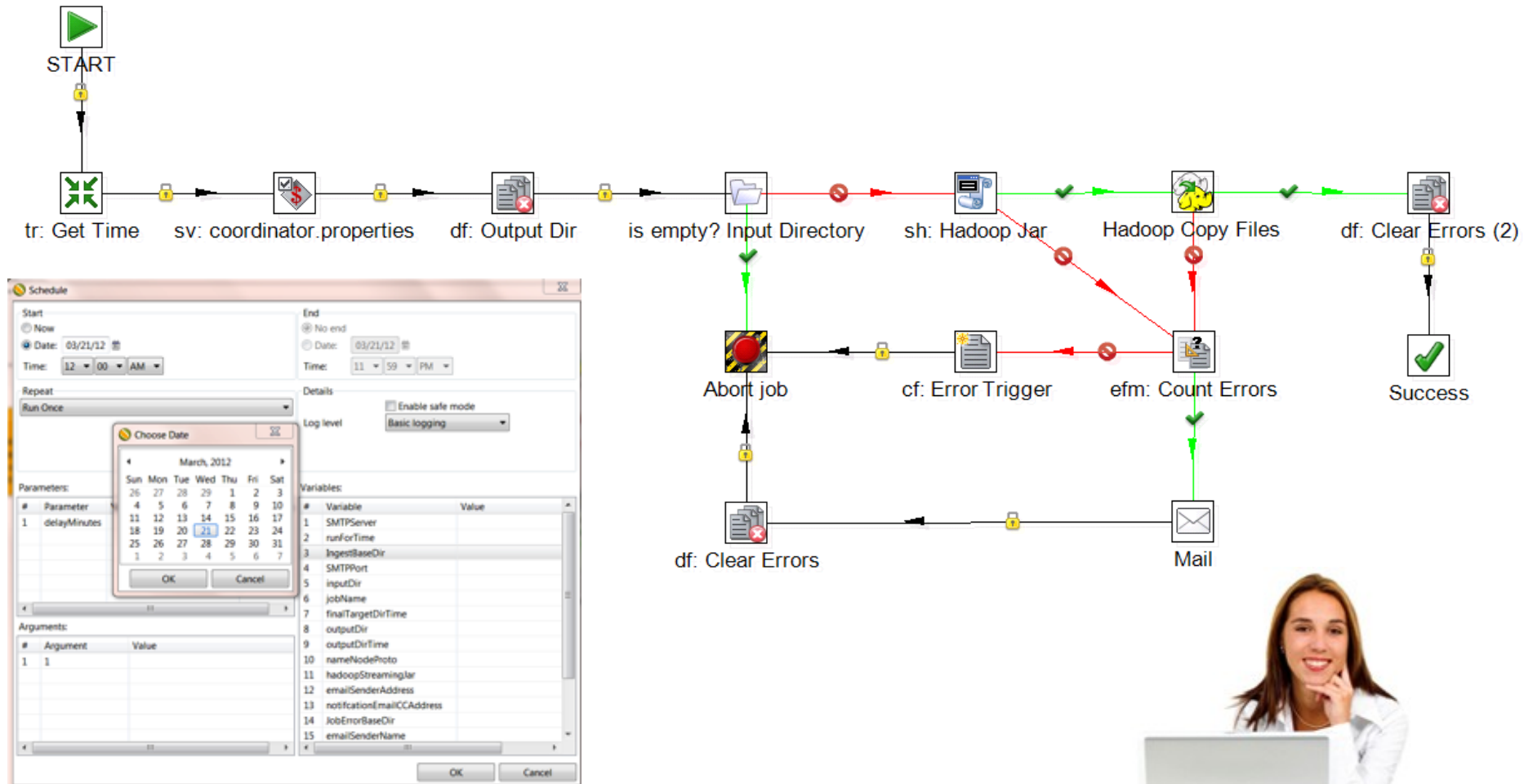
The "Execution Results" tab shows a table of execution metrics for each step in the job.

#	Stepname	Copynr	Read	Written	Input	Output
1	Read Sales Data	0	0	2823	2824	
2	Filter Null Zips	0	2823	2823	0	
3	Lookup missing zips	0	21455	76	0	
4	Read Postal Codes	0	0	21379	21380	
5	Sort out zip codes	0	76	76	0	
6	Check Country Name	0	2823	2823	0	
7	Number range	0	2823	2823	0	
8	Write to Db	0	2823	2823	0	

Pentaho y Big Data



Orquestación virtual de tareas con diferentes fuentes de datos



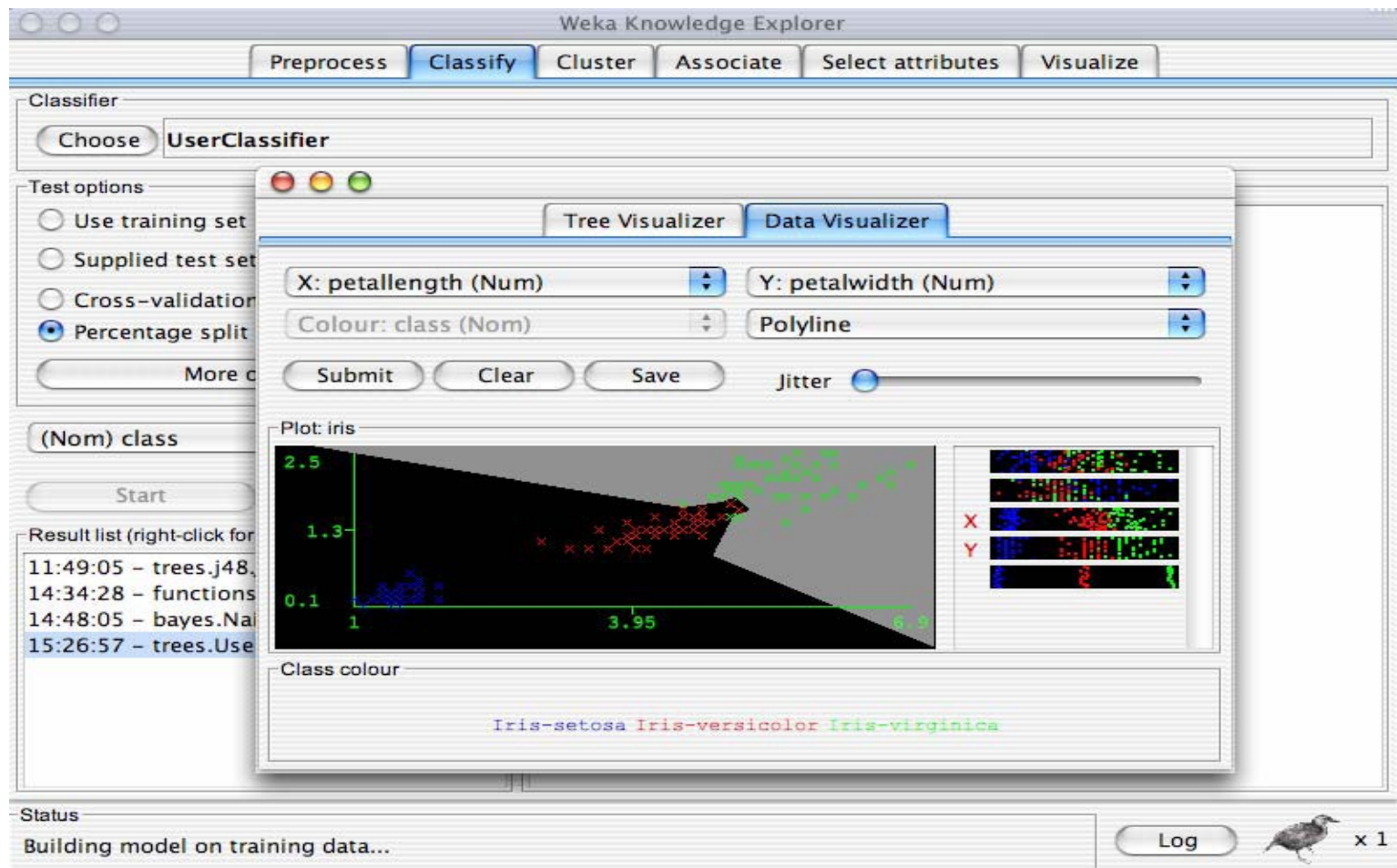
Planificación



Weka

- Descubrir relaciones ocultas en los datos y descubrir indicadores de desempeño futuro.
- Análisis predictivo y exploración de correlaciones en los datos para mejorar el desempeño organizacional

Weka



Weka

