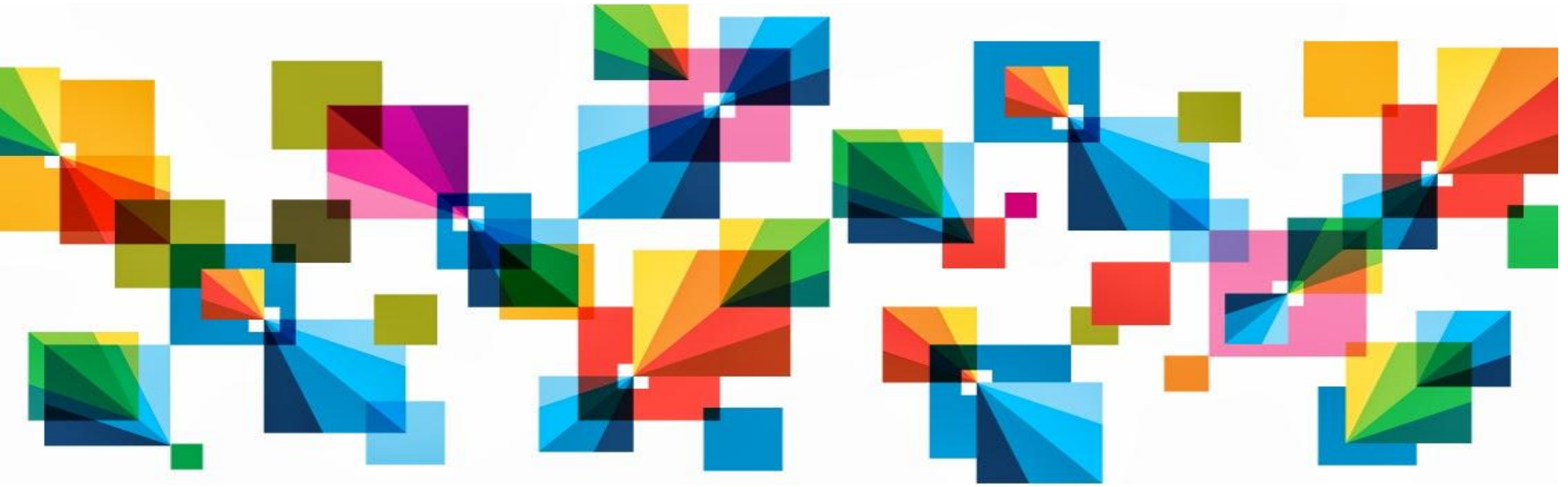
Amplia tu Estrategia de Análisis

Modernización del Almacén de Datos



SOF Quiere Ahorrar Dinero y Cumplir con sus Requisitos Regulatorios

Tengo una gran cantidad de datos históricos en mi almacén de datos relacional que tengo que mantener por razones regulatorias.

Me gustaría que hubiera un lugar más económico para almacenar esos datos.

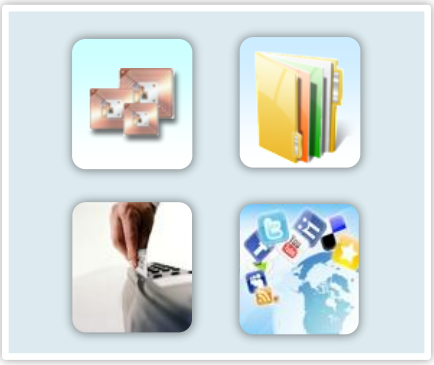
CTO

Debe considerar la construcción de un almacén de datos moderno. Le voy a enseñar cómo.

IBM

Modernización de Almacenamiento de Datos Definido

Integrar las tecnologías de “big data” con el entorno de su almacén de datos para obtener nuevas perspectivas al negocio y optimizar su infraestructura de almacén de datos.

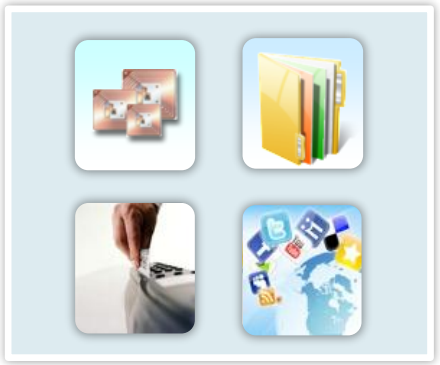


Nuevos tipos de información

Optimice la infraestructura del almacén de datos existente

Reduzca el costo total de propiedad del almacén de datos

La Modernización del Almacén de Datos Aborda Muchos Desafíos

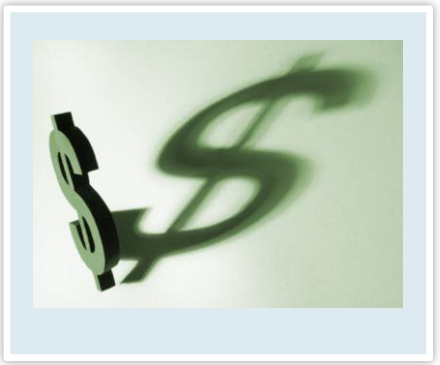


Nuevos tipos de información

- Aproveche nuevos fuentes de datos para perspectivas no antes posible
  - Internet de cosas
  - Datos sociales
- Integración con datos en movimiento
- Perspectivas en minutos y horas no días y semanas

Optimización de la infraestructura AD existente

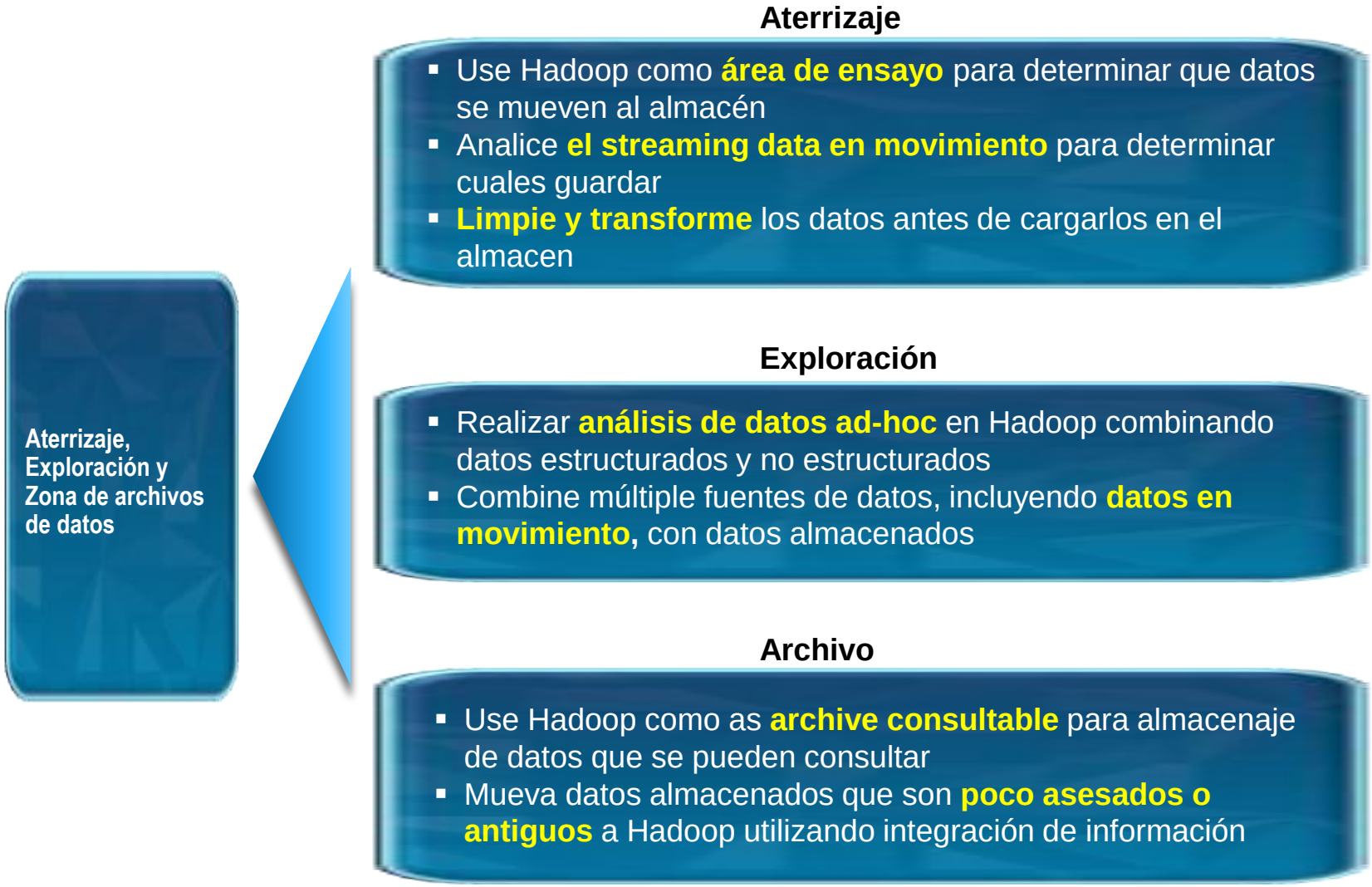
- Optimice la infraestructura del almacén mediante la descarga del pre-procesamiento
- Optimice el almacenamiento de datos a través del almacén y Hadoop
- Mas rápido ciclo de vida de informes
- Visión empresarial comprensivo de todos los datos



Reduce el costo total de propiedad del AD

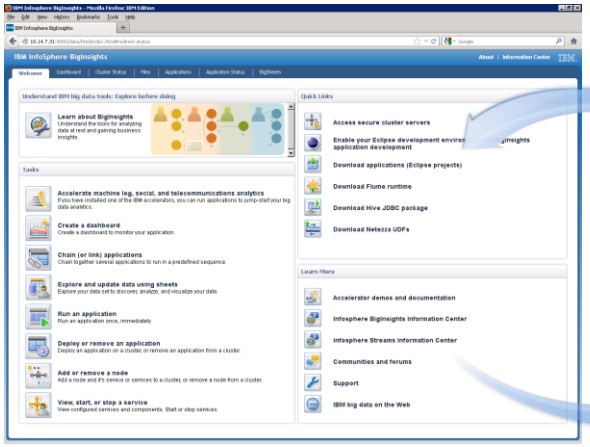
- Migrar datos rara vez usados a Hadoop
- Incremente la optimización del HW del AD alrededor de los datos actuales
- Mejorar el almacenamiento de registros y recuperación (cumplimiento con regulaciones)
- Reducir costos alrededor de la plataforma de integración de datos

Tres Escenarios Comunes de Modernización del Almacén de Datos



PureData Systems for Analytics y BigInsights: Fundación de un Almacén de Datos Moderno

InfoSphere BigInsights



PureData System for Analytics

Proporciona…

- Habilidad de fácilmente mover datos entre los dos sistemas
- Habilidad de capturar y procesar nuevos tipos de datos
- Habilidad de consultar todos tipos de datos usando habilidades SQL existentes
- Entorno optimizado de almacén de datos moderno

Service Oriented Finance Quiere Empezar a Crear su Almacén de Datos Moderno

Aquí esta una buen manera de empezar que…

1. **Ahorra dinero**
2. **Libera recursos en su almacén existente**
3. **Utiliza sus habilidades de SQL existentes**

IBM

Paso 1

Mueva datos mas antiguos a BigInsights

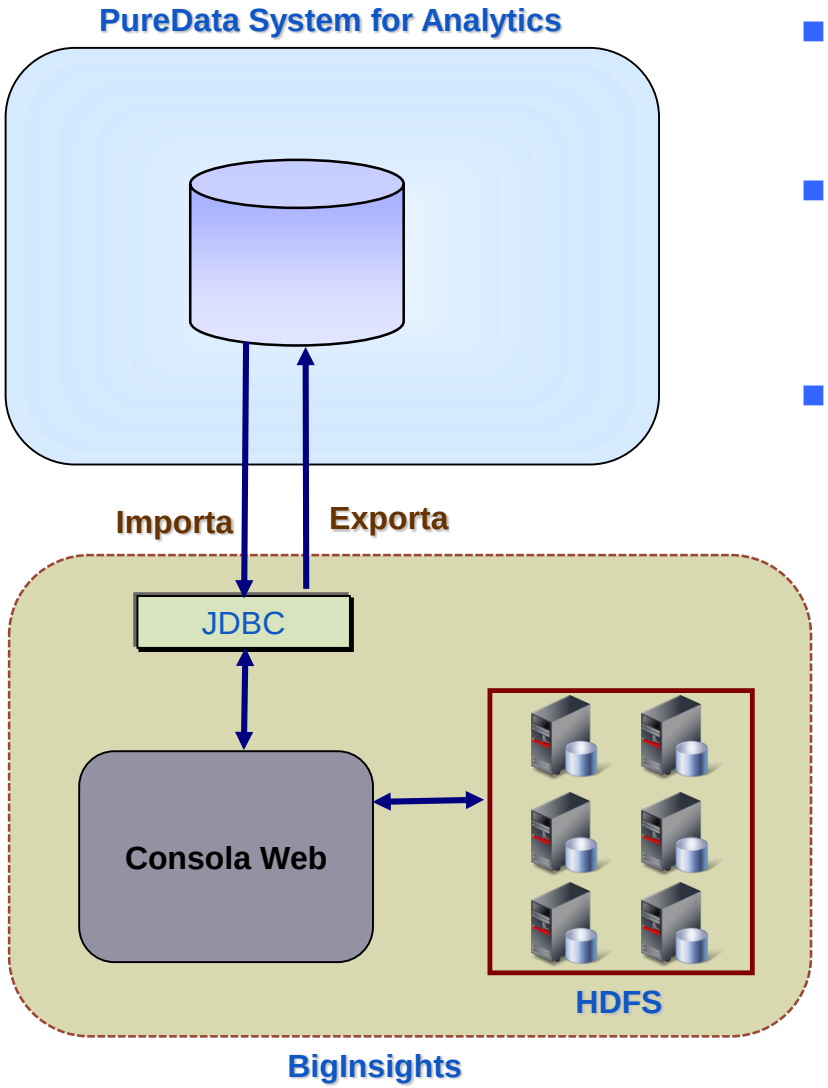
Paso 2

Define el esquema en términos SQL

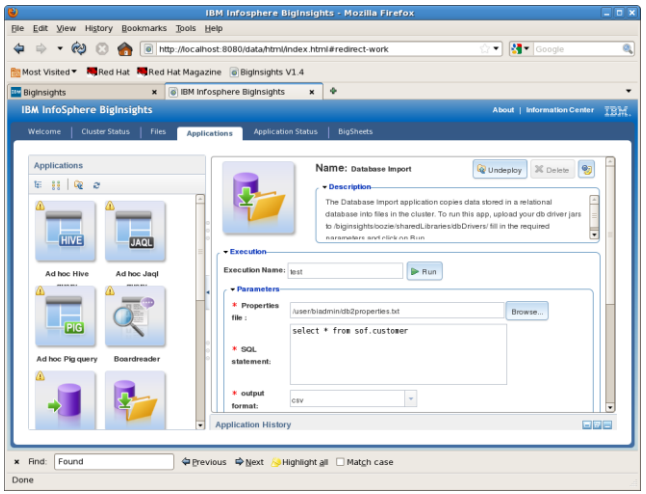
Paso 3

Empiece acezar archivos que se pueden consultar!

Usando BigInsights Console y JDBC – Exporta e Importa

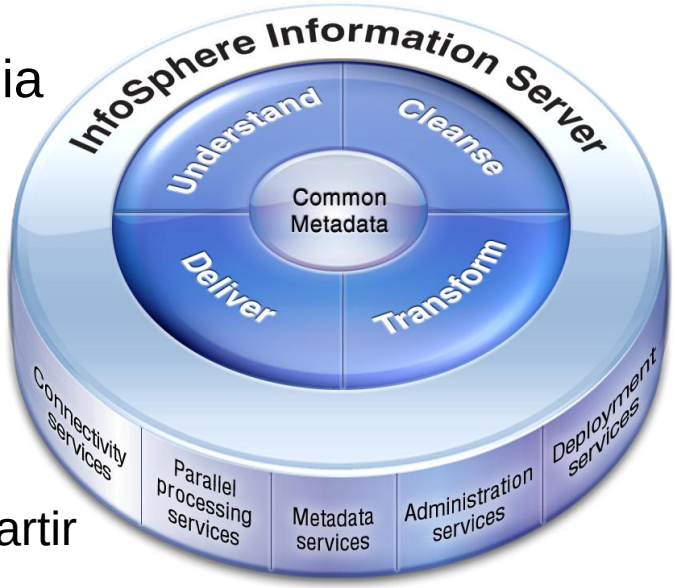


- Parámetros JDBC especificados en la consola de la aplicación
- La importación carga HDFS con datos relacionales – formato es CSV o JSON
- La exportación escribe datos de HDFS en una tabla relacional – Formato HDFS debe ser CSV o JSON



El "Information Server" Proporciona una Solución Integrada Completa de Clase Empresarial

- Se usan las mismas herramientas de diseño graficas independiente de la fuente de los datos
- Herramientas probadas con años de uso y amplia difusión
- Escalabilidad lineal probada
- Acelera la productividad, reduce los costos y riesgos
  - Metadatos compartidos fomentan colaboración
  - Maximiza reutilización – construir una vez y compartir
  - Cienes de componentes pre-construidos
- Se conecta a cualquier fuente de datos
  - Base de datos, aplicaciones, archivos, colas de mensajes, Hadoop, NoSQL

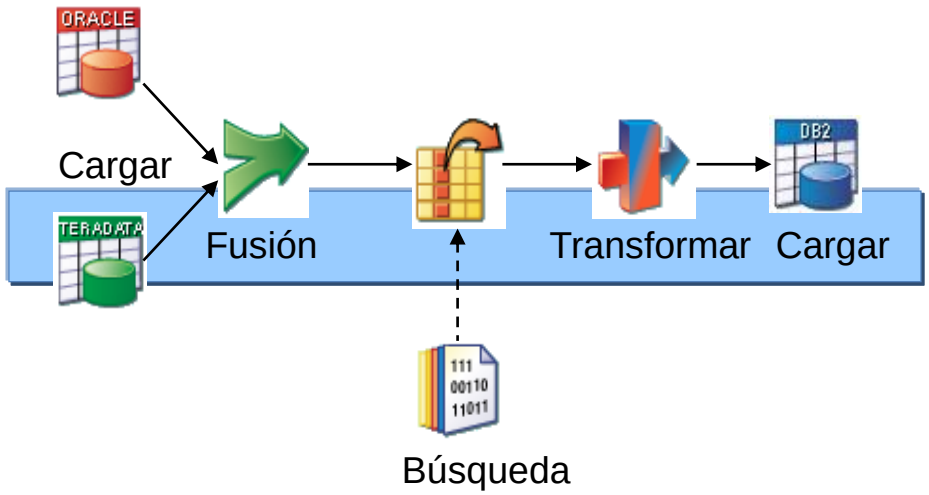


Step 1

Mueva datos mas antiguos a BigInsights

DataStage Hace Fácil La Integración de Todos Las Fuentes de Datos

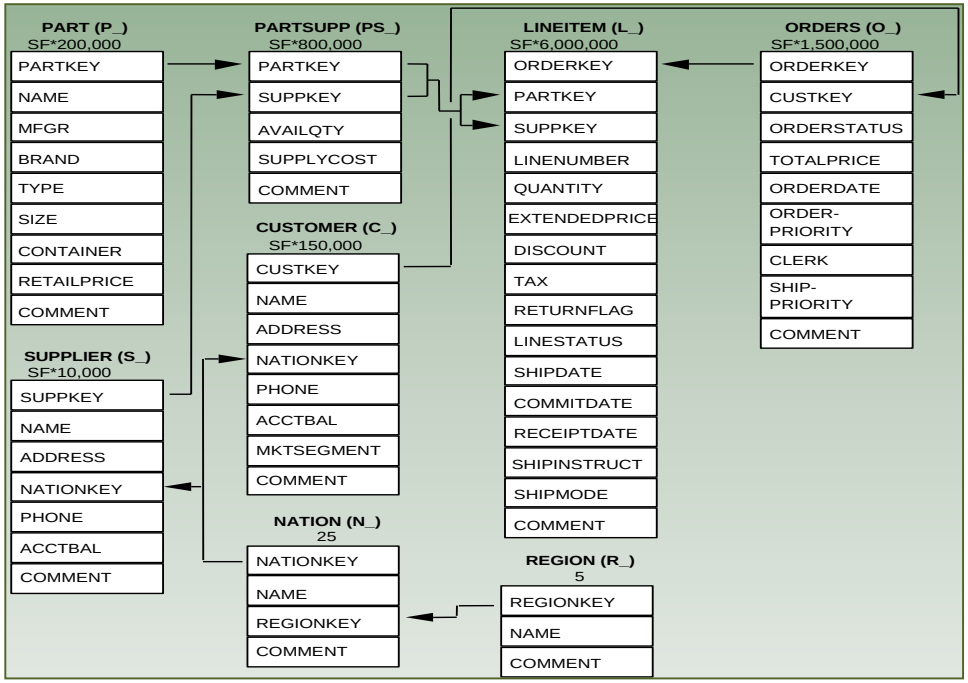
Arrastre, suelte, y configure



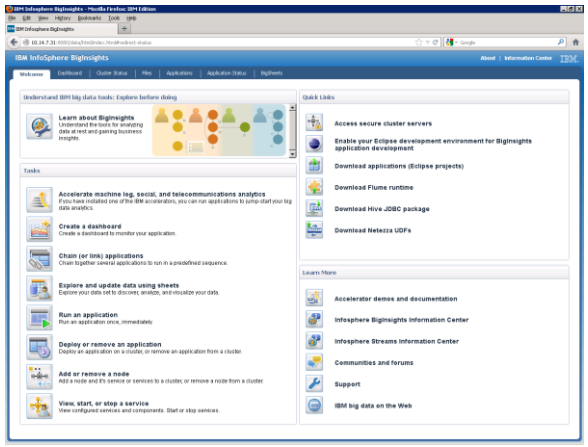
- Productivo
  - Diseño grafico de los flujos de datos
    - Se centran en el flujo en lugar de la implementación
  - Numerosos componentes pre-construidos
- Escalable
  - Aprovecha procesamiento paralelo
- Reduce el riesgo y el costo
  - Enfoque modular maximiza la reutilización de los componentes
  - Metadatos compartidos mejora la colaboración
- Una herramienta integra todas las fuentes de datos

Paso 2: Definir Esquema Existente en BigInsights

Esquema Relacional Existente



InfoSphere BigInsights

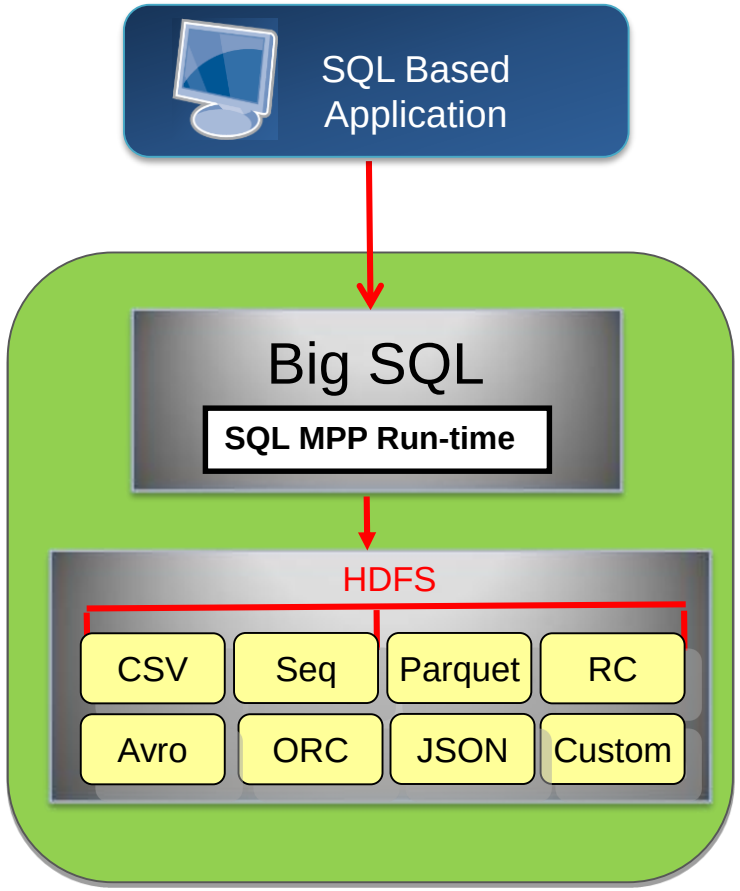


Paso 2

Define el esquema en términos SQL

Big SQL V3.0: Moviendo SQL en Hadoop al Siguiente Nivel

- Masivamente paralelo motor SQL en Hadoop
  - Diseñado desde el principio para baja latencia y alto rendimiento
- Soporte SQL comprensivo
  - El mismo SQL que utiliza en su almacén de datos debe ejecutar con pocas o ninguna modificación
  - Soporte completo para sub-consultas
  - Todas operaciones “JOIN” estándares
  - Procedimientos almacenados / Funciones definidas por usuario
- Soporta todos los formatos de archivos modernos

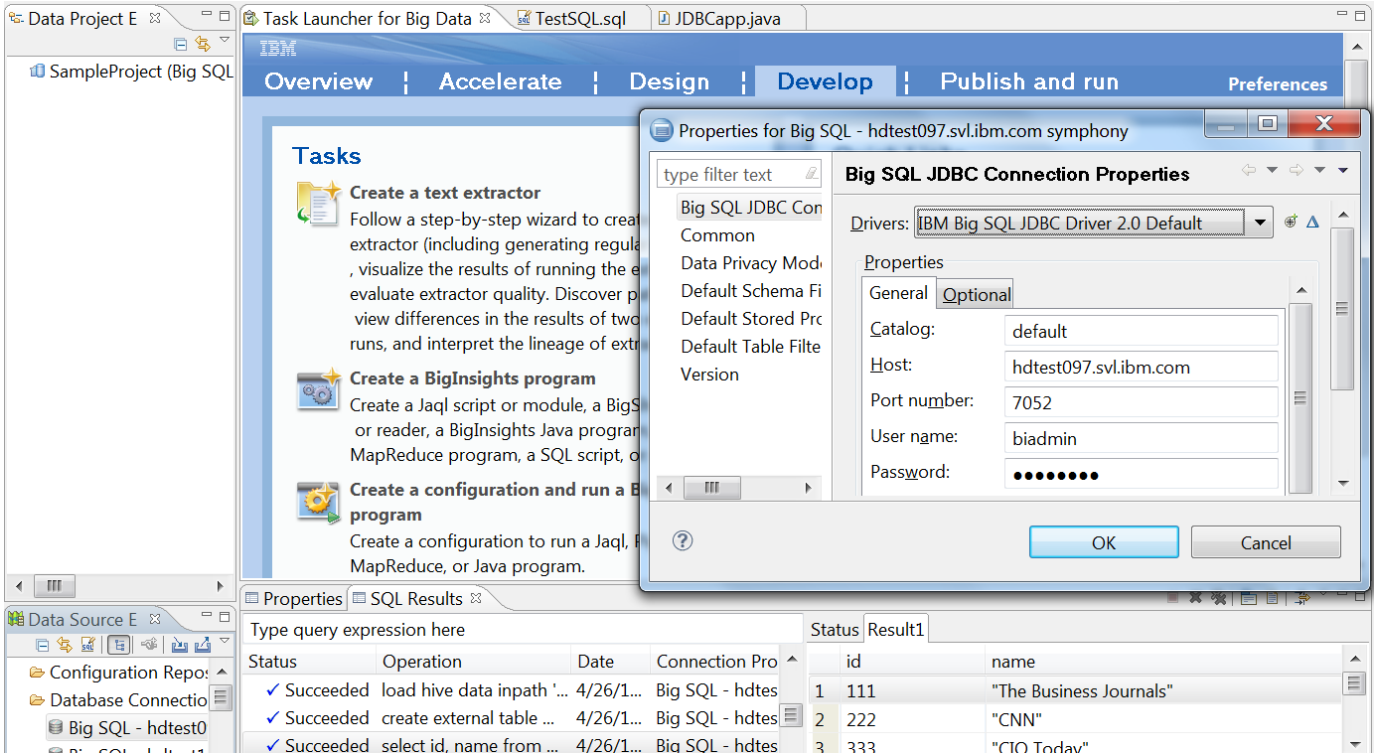


¿Por qué Queremos Utilizar SQL con Datos en Hadoop?

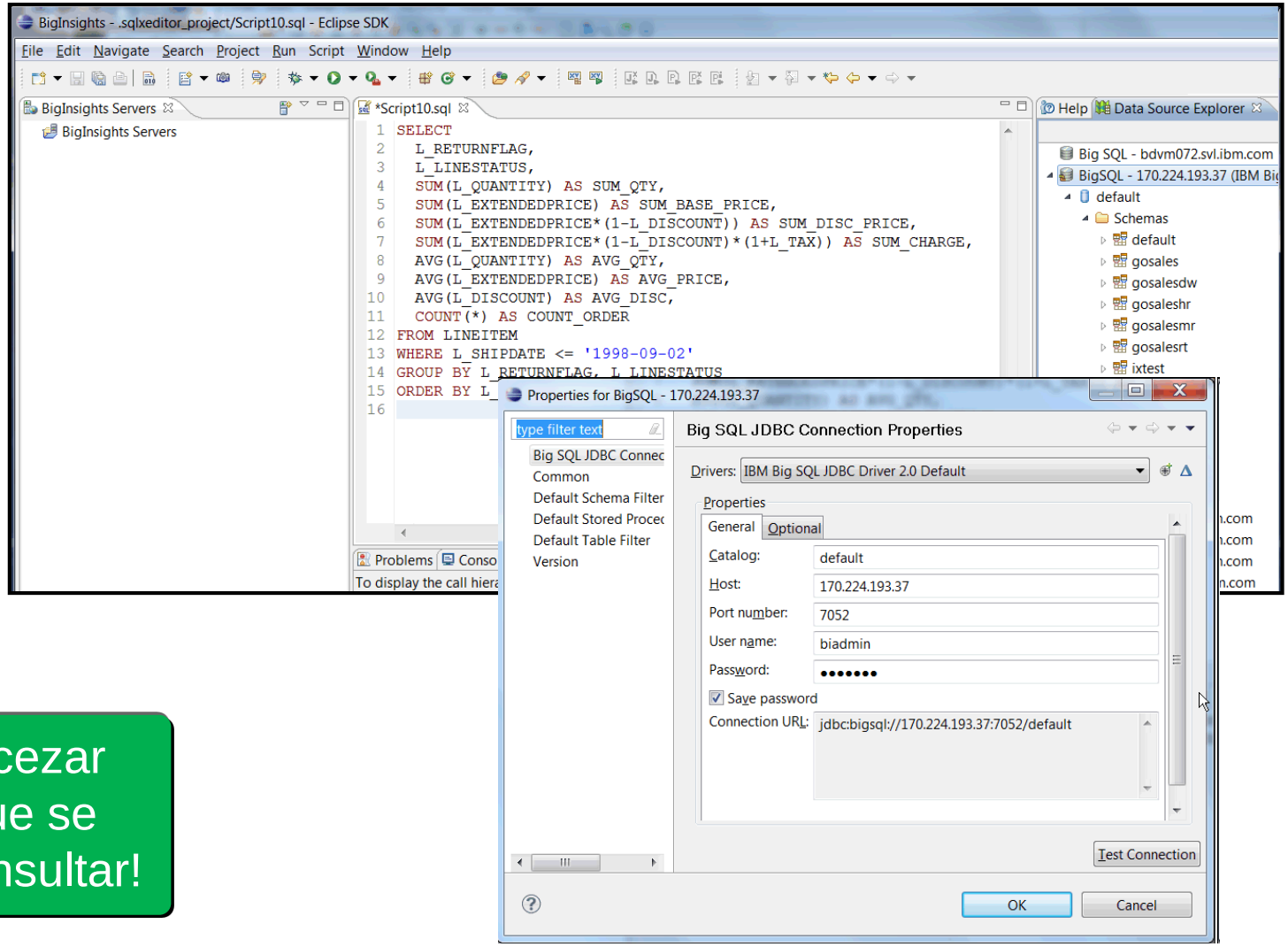
- Programación MapReduce es difícil
  - MapReduce Java API requiere conocimientos de programación
- Hadoop/MapReduce son tecnologías nuevas
  - La cantidad de experiencia es limitada
- Lenguas desconocidas (como Pig) también requieren habilidades especiales
- Compatibilidad con SQL abre los datos a un público mucho más amplio
  - Rampa de abordar fácil para profesionales de Hadoop para SQL
- Soporte para SQL abre los datos a una amplia variedad de herramientas de SQL
  - Cognos, JDBC, ODBC

Opciones de Invocación Proporcionados con BigInsights

- Command-line interface (JSqsh shell)
- Web-based interface (BigInsights web console)
- Eclipse (BigInsights plug-in)



DEMO: Usando Herramientas Eclipse Para el Acceso de Archivos que se Pueden Consultar

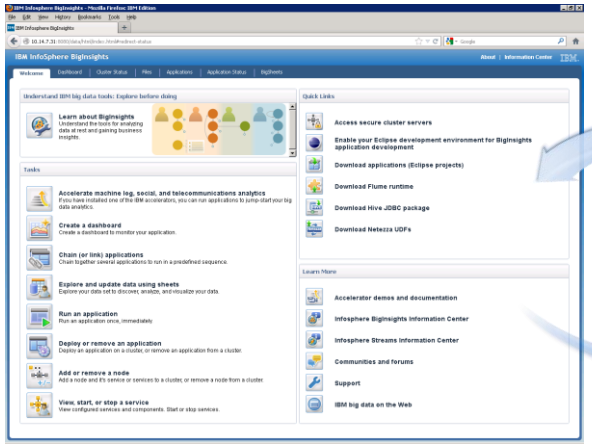


Step 3

Empiece acezar archivos que se pueden consultar!

Service Oriented Finance ha Creado un Almacén de Datos Moderno

InfoSphere BigInsights



PureData System for Analytics

Paso 1

Mueva datos mas antiguos a BigInsights

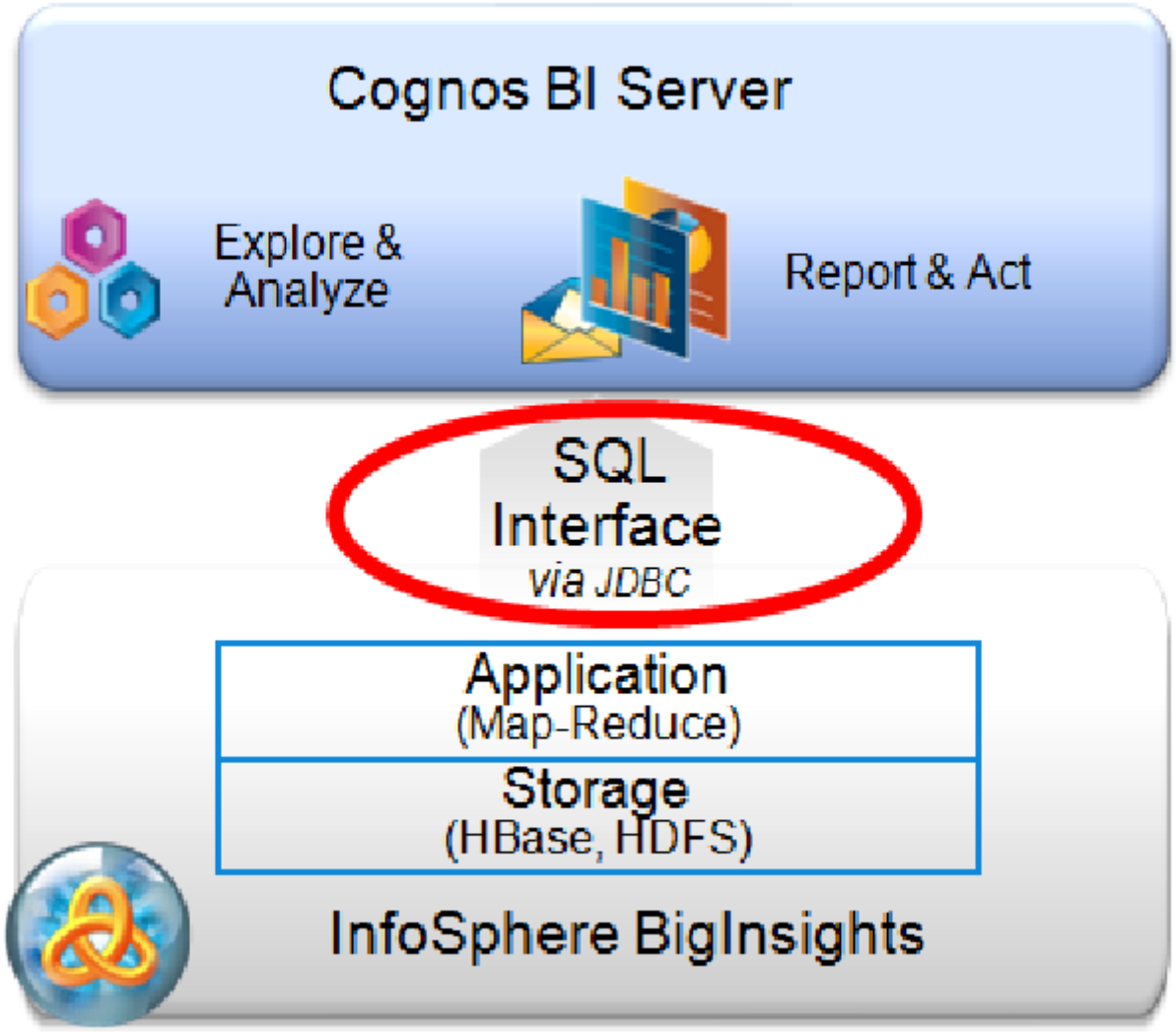
Paso 2

Define el esquema en términos SQL

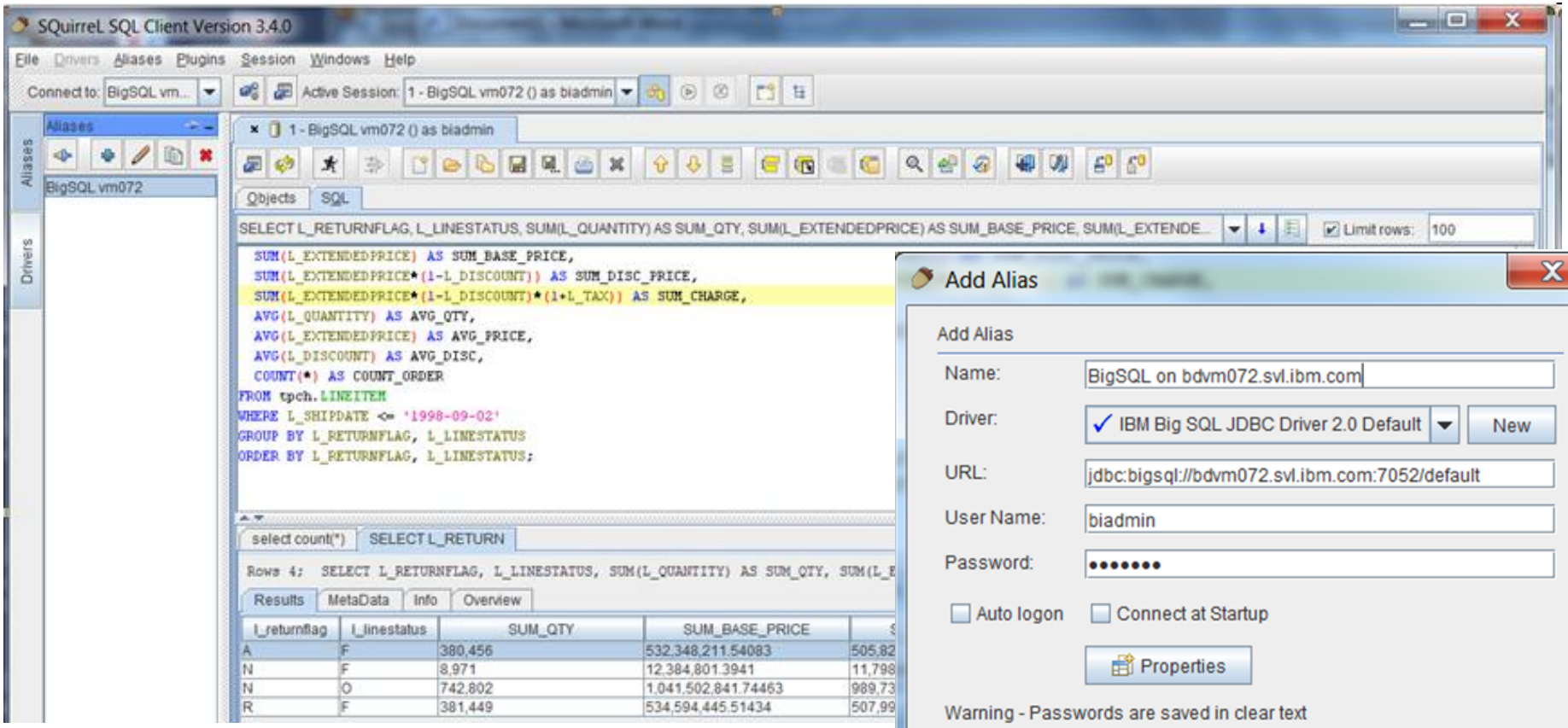
Paso 3

Empiece acezar archivos que se pueden consultar!

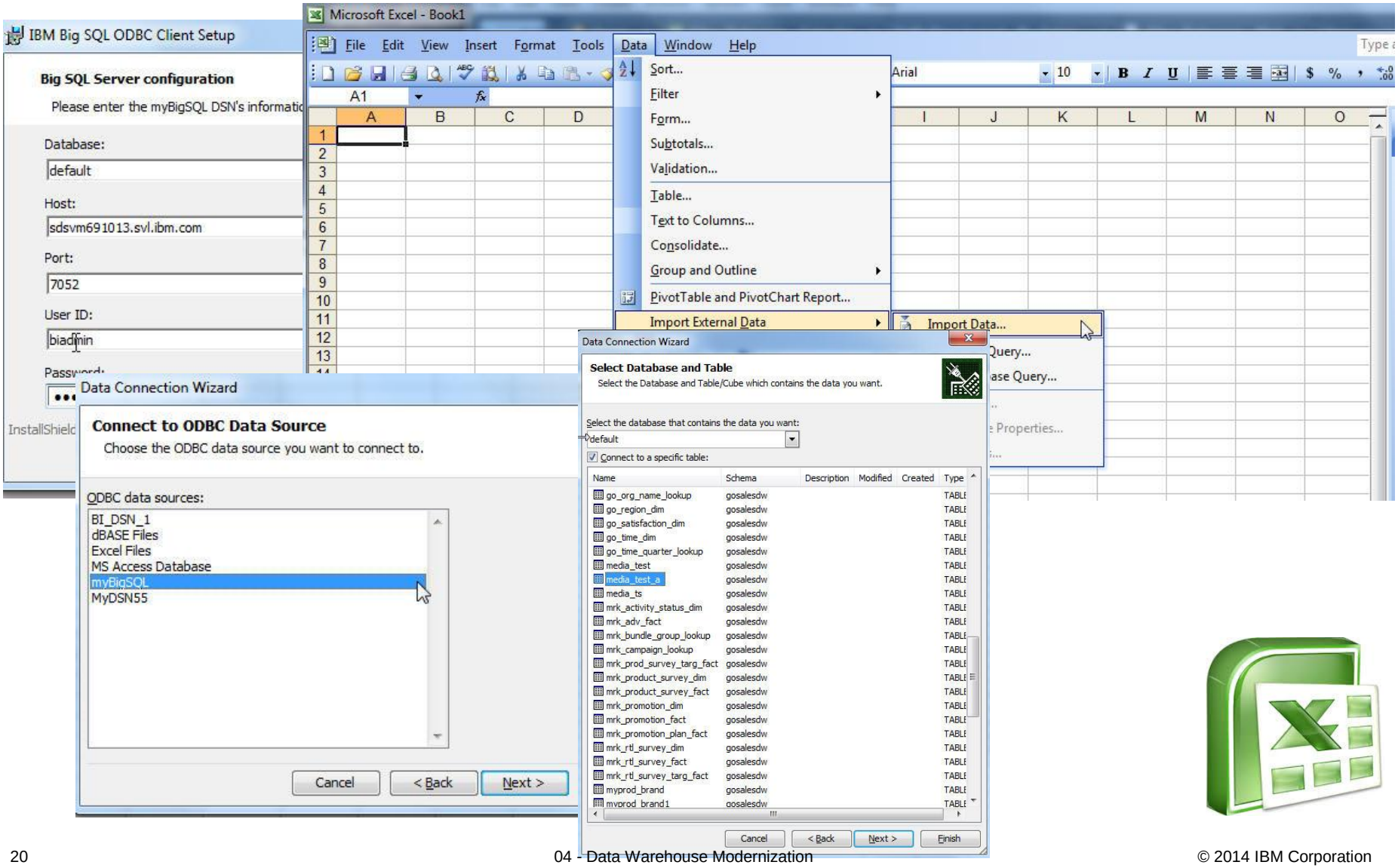
Inteligencia Empresarial Cognos



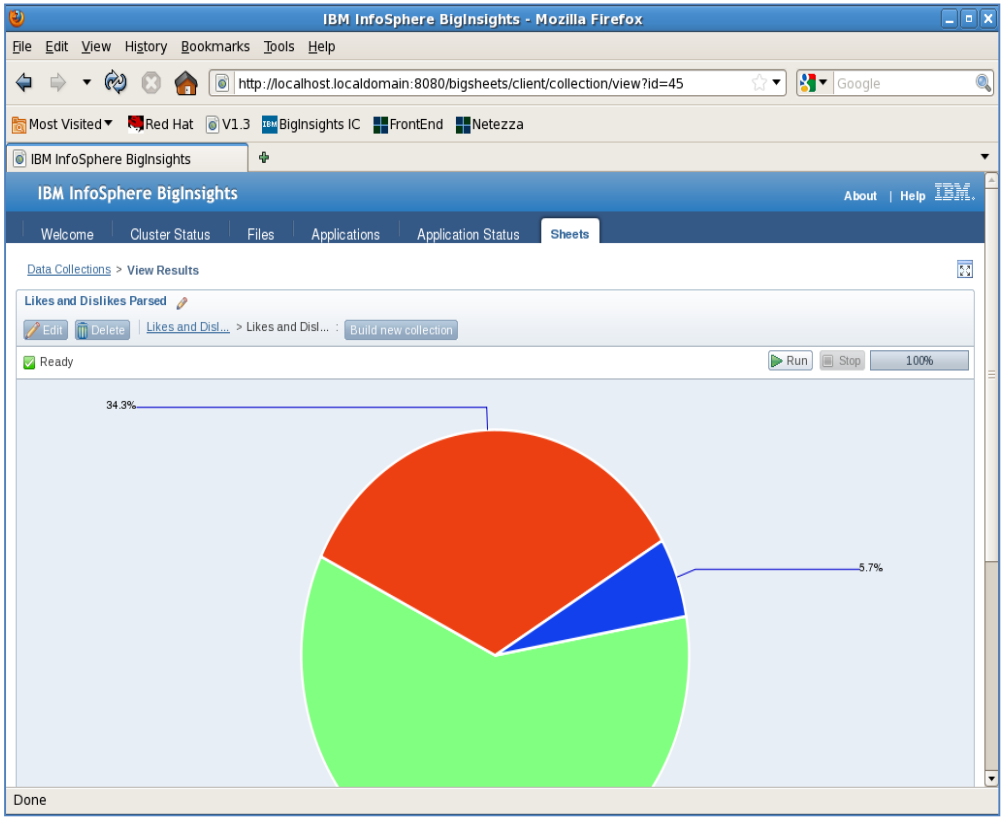
Usando Existente Herramientas Estándares Como SQuirreL SQL



MS Excel: Integración Big SQL a Través de ODBC



Datos Disponibles para Todas las Herramientas de BigInsights



- BigSheets spreadsheet y visualización
- Acelerador de datos de maquina
- Acelerador de datos de medios sociales
- Analítica de texto avanzado
- Lenguaje de consulta JAQL
- Consola de Herramientas
- Big R –Integración con lenguaje R

BigInsights con Big SQL en POWER Proporciona Consultas Big Data Mas Rápido Que Hive en x86

Consulta de Carga de Trabajo de Inteligencia Empresarial Moderno

BigInsights v3.0 en 8 S822L nodos

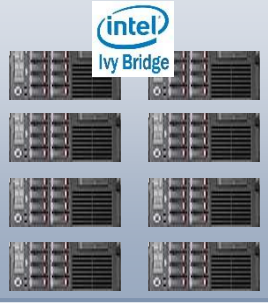
S822L, 3.3 GHz
24 núcleos
256 GB Memoria
RHEL 6.5

Tiempo para completar las consultas en 7 corrientes concurrentes

8 horas 40 min

Hive v0.12 en 8 Ivy Bridge EP nodos

Intel Ivy Bridge EP
2.7GHz
24 núcleos
256 GB Memoria
RHEL 6.4



Tiempo para completar las consultas en 7 corrientes concurrentes

97 horas 27 min

11.2x
Mas rápido

4.4x
Menor precio de rendimiento

This is an IBM internal study of a benchmark-inspired workload in a controlled laboratory environment. The IBM system under test consists of eight Power S822L data nodes (24 cores each @ 3.3 GHz) with 256 GB RAM running RHEL 6.5. The Intel-based system under test consists of eight Ivy Bridge EP data nodes (24 cores each @ x2.7 GHz) with 256 GB RAM running RHEL 6.4. Both systems each used a 10TB scale factor for the TCP-DS inspired test. Both systems were driven by a single management system (Power S822L with 24 cores, 256 GB RAM, RHEL 6.5 and Ivy Bridge EP - 24 cores, 256 GB RAM, RHEL 6.4) that were immaterial to the performance of the systems under test. Tests measured total elapsed time to complete an identical subset of 42 of the 99 queries in the TPC-DS-inspired workload. Test executed 7 simultaneous query streams. Measured time was the maximum of the 7 streams across 2 back-to-back runs. Customer applications, differences in the systems deployed and other system variations or testing conditions may produce different results. Cost analysis based on 3 year total cost of acquisition of hardware, software and support services over a 3 year period. Prices for both IBM and competitor systems based on US list prices valid as of July 2014.

Como Empezar con la Modernización del Almacén de Datos

Consiga educación:

- Descargue BigInsights Quick Start ibm.com/infosphere/quickstart
- Descargue el white paper: Data Warehouse Augmentation: the Queryable Data Store
- Visite el sitio: Data Warehouse Augmentation website
- Visite el sitio: IBMBigDataHub.com para acezar podcasts y video de este caso
- Descarge la investigación: Capitalizing on Big Data

Agenda de Hoy

Time	Topic
09:00 – 09:15 AM	Introducción: Lo Que la Analítica Big Data Puede Hacer Para Su Negocio
09:15 – 10:00 AM	Domine los Fundamentos: Analizando datos estructurados con sistemas PureData System for Analytics
10:00 – 10:15 AM	**Break**
10:15 – 11:00 AM	La Analítica de Datos no Estructurados: Análisis Big Data con Hadoop
11:00 – 11:45 AM	Amplía tu Estrategia de Análisis: Modernización de Almacén de Datos
11:45 – 12:00 PM	Resumen y Acción