



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

Return on Information

資訊創造報酬率，迎接ROI大革新

Willyl@tw.ibm.com



大綱

- Overview

- InfoSphere BigInsight 擴充了Hadoop的可用性

- UseCases

- 徹底了解海量資料的應用說明

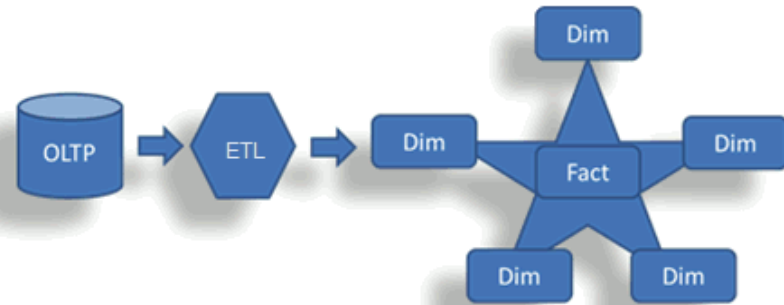


IBM引領資訊潮流不斷的變革

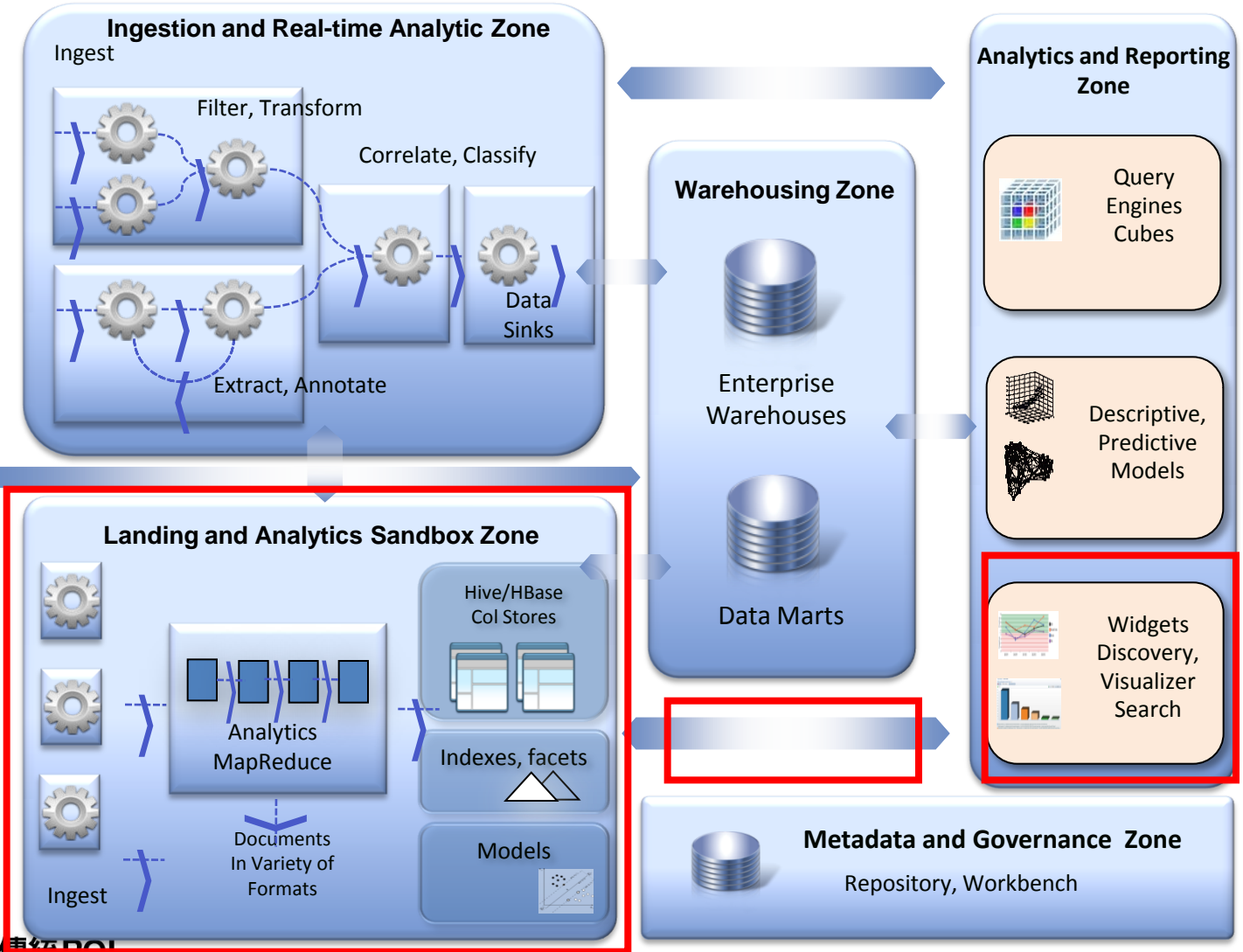
- 1960s: Navigational DBMS
 - IMS (hierarchical)
- 1970s-1980s: Relational DBMS
 - SQL
 - System R, System Z, DB2
- 1990s: Data Warehouse
 - Dimensional model, ETL, M
- **Today: Big Data**
 - **Big SQL**



Ted Codd



完整實踐海量資料平台



拼裝車和品牌車都是車，IBM提供企業級的海量資料平臺

Enterprise Capabilities

Visualization & Exploration

Development Tools

Advanced Engines

Connectors

Workload Optimization

Administration & Security

IBM-certified
Apache Hadoop

Open source
components

Platform Capabilities 平台能力

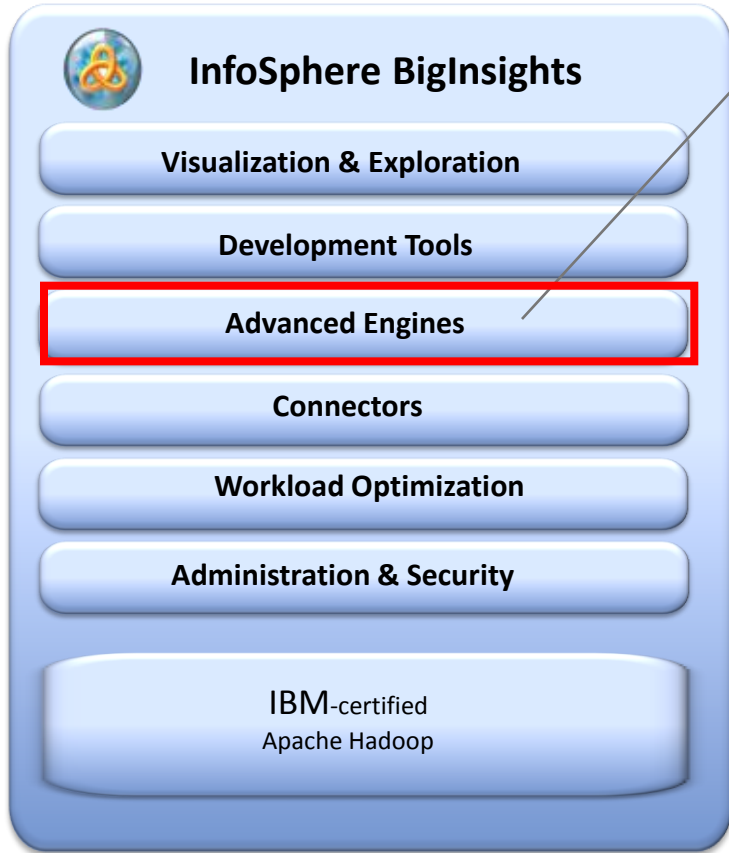
- **Built-in analytics**
- **Enterprise-grade capabilities**
- Integrated with enterprise software
- Ease of installation and management
- Reference hardware configurations
- World-class support
- **Full open source compatibility**

Business benefits

- **Quicker time-to-value**
- **Reduced operational risk**
- Enhanced business knowledge with flexible analytical platform
- Leverages and complements existing software investments



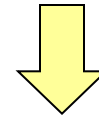
開源Hadoop不容易做到的 - 海量文本分析



Text Analytics Engine on Hadoop

- Identify and understand key information in text
- Ability to determine concepts like sentiments, intent

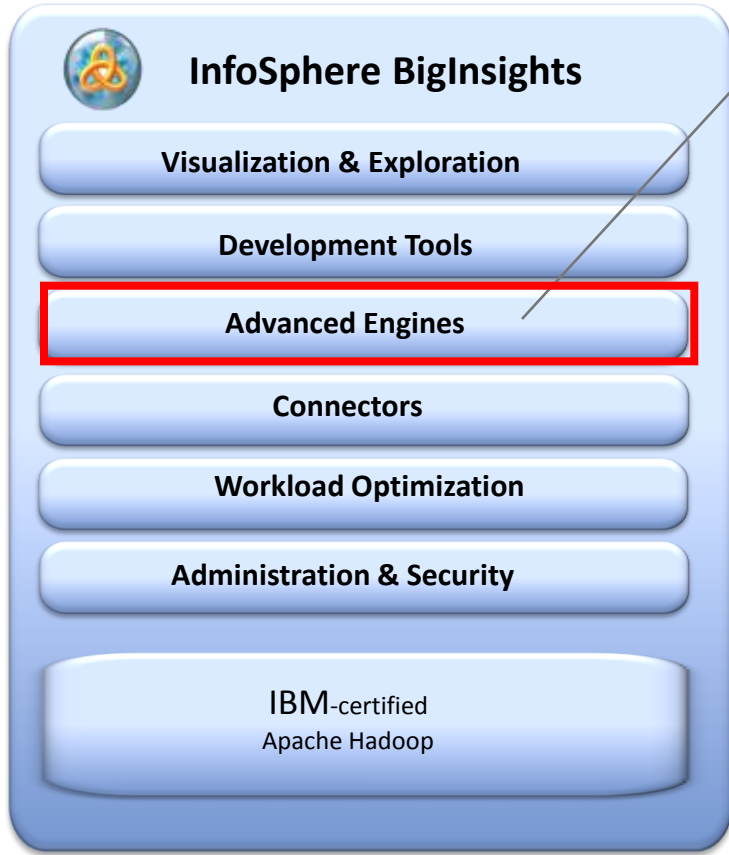
Football **World Cup 2010**, one team distinguished themselves well, losing to the eventual champions 1-0 in the Final. Early in the second half, **Netherlands' striker, Arjen Robben**, had a breakaway, but the **keeper for Spain, Iker Casillas** made the save. **Winger Andres Iniesta** scored for **Spain** for the win.



World Cup 2010 Highlights		
Name	Position	Country
Arjen Robben	Striker	Netherlands
Iker Casillas	Keeper	Spain
Andres Iniesta	Winger	Spain

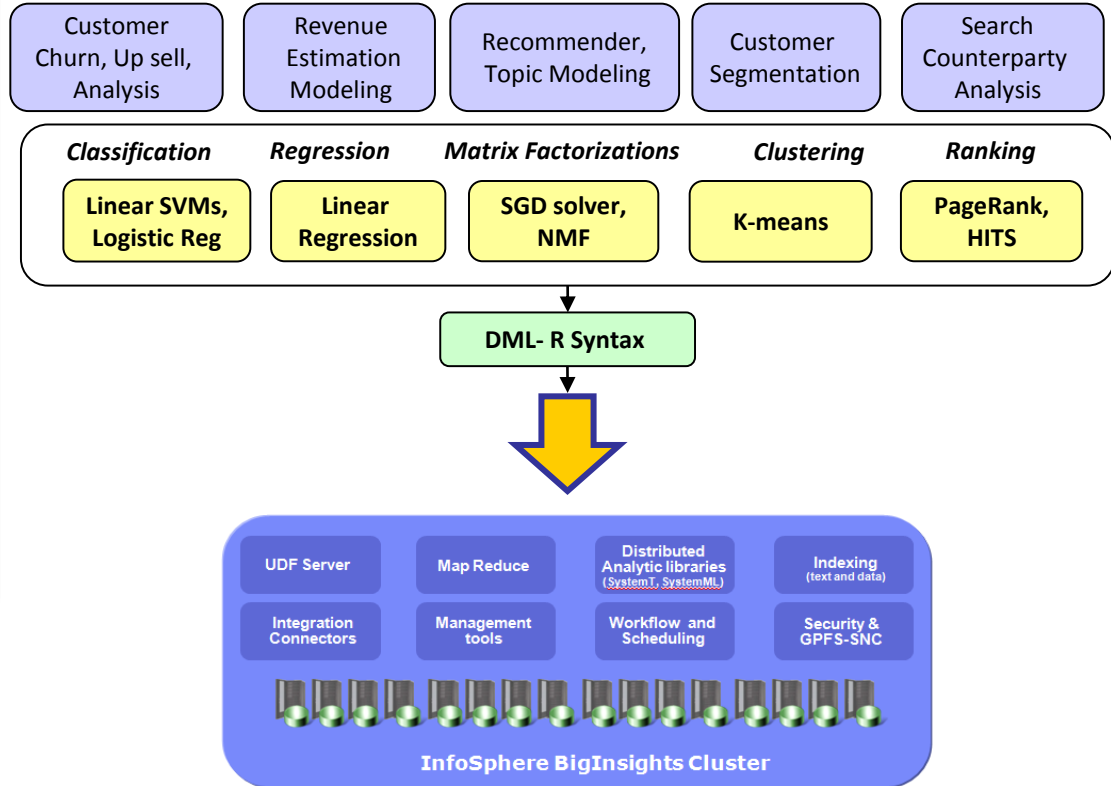


開源Hadoop不容易做到的 - 機器學習模型




Machine Learning Engine on Hadoop

- R-like syntax for ease of adoption
- Automatically generate **Map Reduce** execution plans



開源Hadoop不容易做到的 - 超友善的海量試算表



InfoSphere BigInsights

Visualization & Exploration

Development Tools

Advanced Engines

Connectors

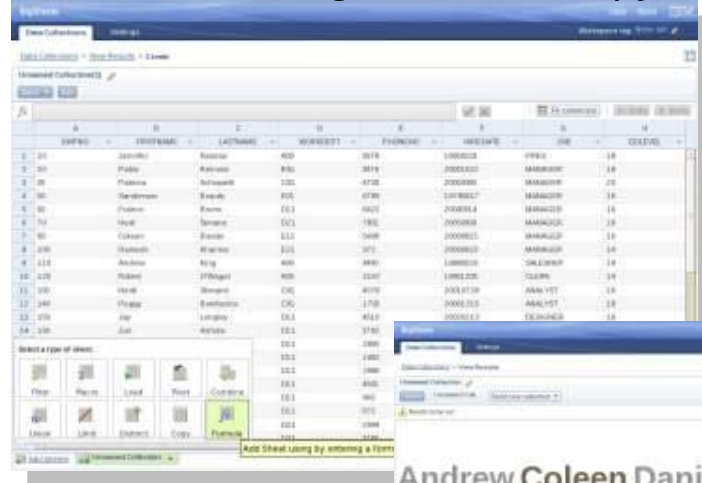
Workload Optimization

Administration & Security

**IBM-certified
Apache Hadoop**

BigSheets Visualization and Exploration

- **Web-based** analysis and visualization
- Familiar **spreadsheet-like** interface
- Define and manage data discovery jobs



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革新

Return on Information

開源Hadoop不容易做到的—海量密集資料的分散式 平行運算架構



Adaptive MapReduce

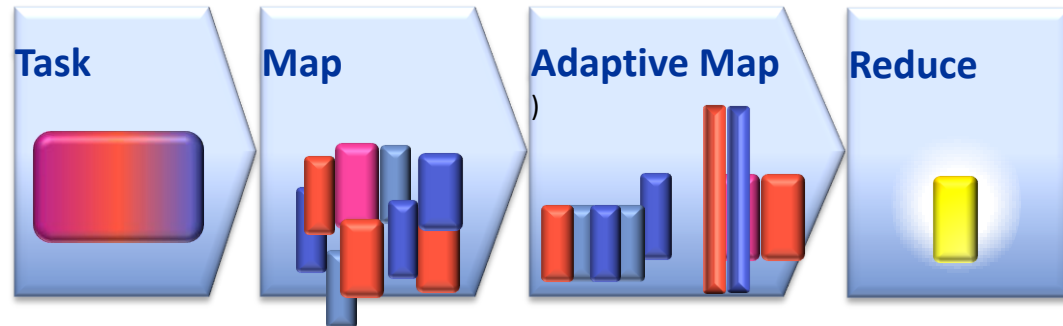
- Optimize execution time of multiple small jobs
- Performance gains of **30%** by reducing MR overheads

Enhanced System Scheduler

- Rule based job scheduler to help manage SLA
- Integrated with Platform Computing for more advanced workflow management

Enhanced Split-able Compression

- Optimized compress and decompress performance to reduce overhead of inter-node communications

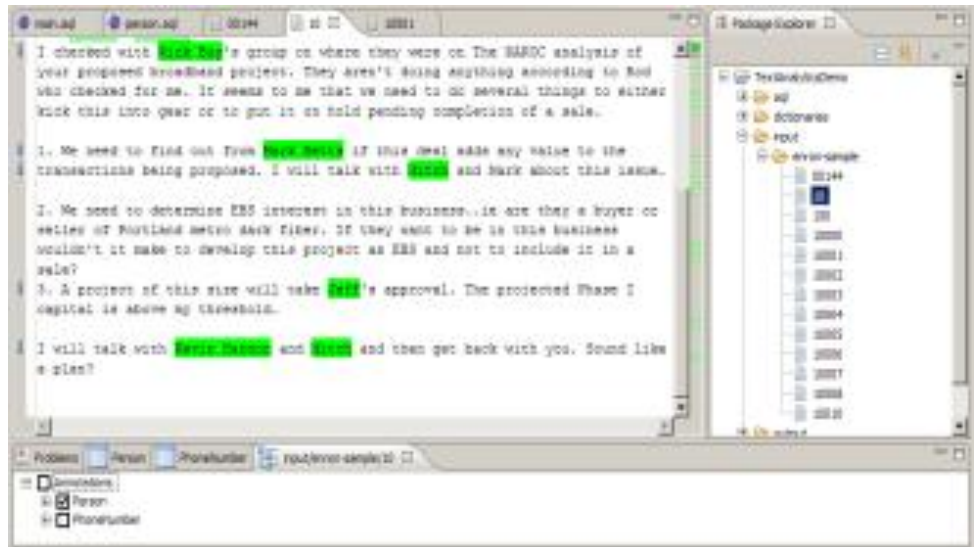


開源Hadoop不容易做到的 - 高整合度的開發環境



Development Environment

- **Eclipse based** development environment for building and deploying Big Data applications
- Reduce coding and debugging time by up to 30%
- Plug-in for Text analytics, MapReduce programming, Jaql development, and Hive query development

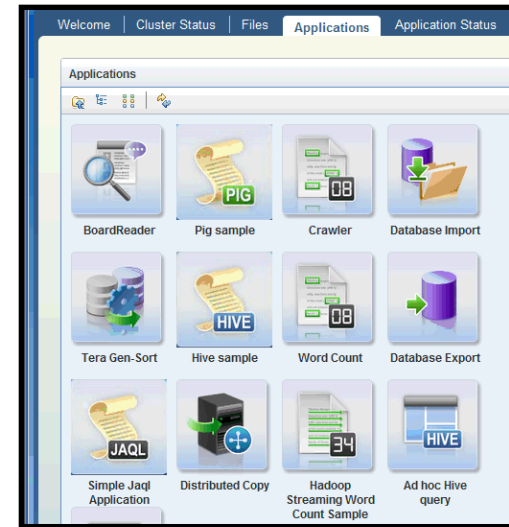


開源Hadoop不容易做到的 - 完備的企業級管理功能



Web Console

- Start / stop services
- Run / monitor jobs (applications)
- Explore / modify file system
- Built in Apps make it easy to do common tasks



Performance and HA

- Reference configurations
- HA solutions

Security

- LDAP authentication and reverse-proxy
- Administrators restrict access to authorized users
- Role based access



開源Hadoop不容易做到的—擁有DBMS的特性



Database Connectors

Batch Import/Export

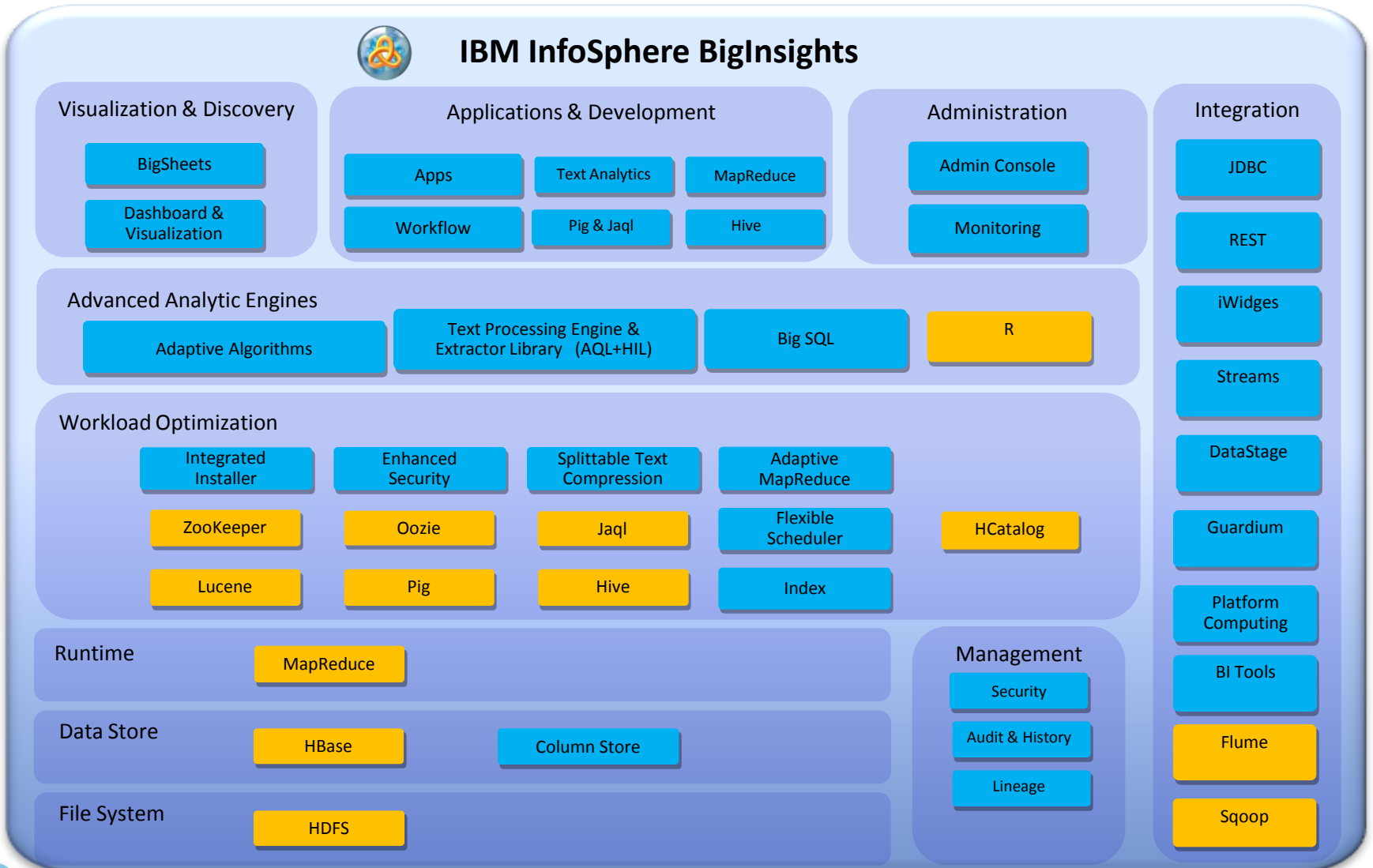
- High speed parallel transformation and load via Data Stage
- JDBC/SQL import/export from BigInsights via Java MR or JAQL
- SQOOP import/export

Query based SQL Access

- **BigSQL** SQL 92/99 compliance
- DBC/SQL interfaces from BigInsights
- HIVE/JDBC/SQL query from report tools
- JDBC/SPL real-time data pull from Streams



InfoSphere BigInsight使Hadoop變成企業真正可用



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革新

Open Source

IBM

大綱

- Overview
 - InfoSphere BigInsight 擴充了Hadoop的可用性
- UseCases
 - 徹底了解海量資料的應用說明



InfoSphere Biginsight 實現海量資料的應用



Big Data Exploration

Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external data sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革命

Return on Information

InfoSphere Biginsight 實現海量資料的應用



Big Data Exploration

Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external data sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine

顛覆企業傳統ROI Improved business results

資訊革新論壇 The New ROI

資訊創造報酬率，迎接ROI大革命



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



開箱即用的海量試算表(BigSheets)讓探索海量資料變得容易

• Data Collection

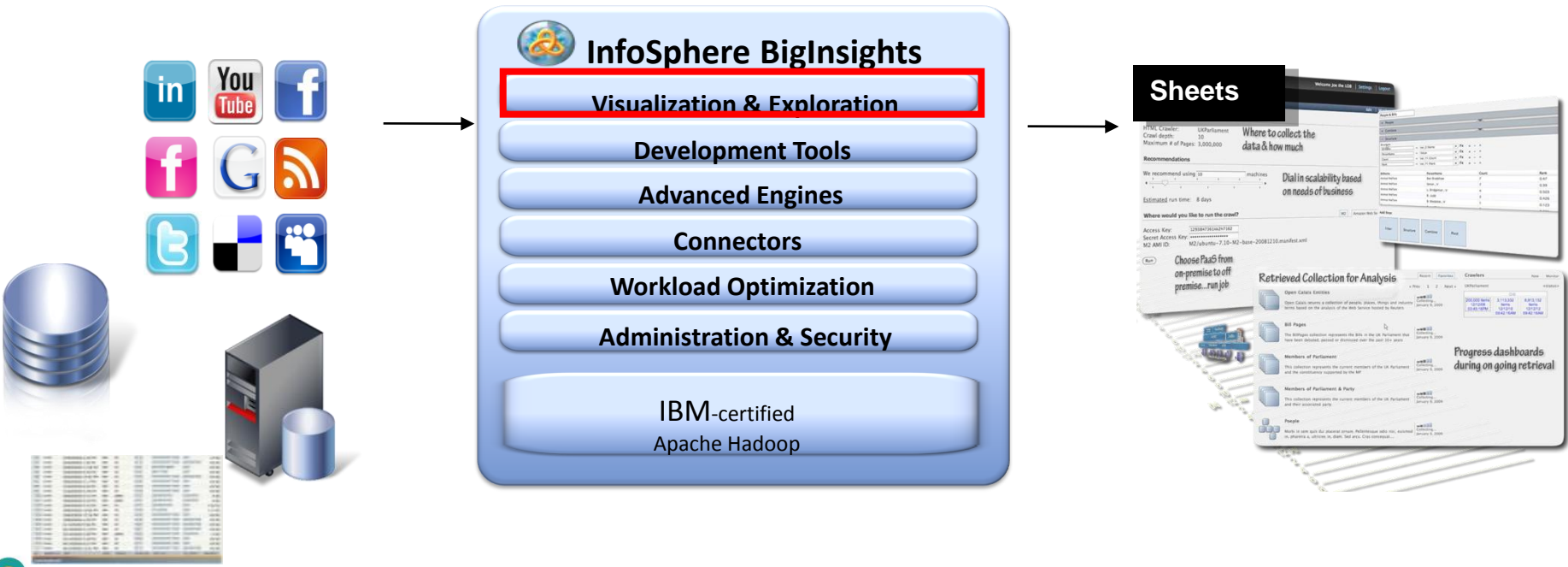
- WebCrawler app
- DBMS import app
- BoardReader app
- Flume

▪ Data Storage

- Distributed file system
- Web based file browser and administration

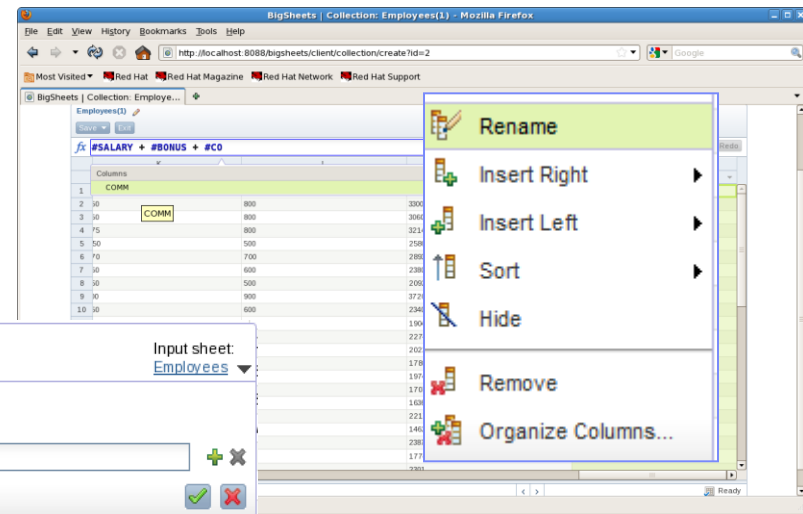
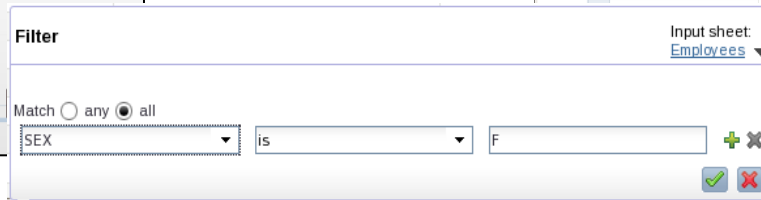
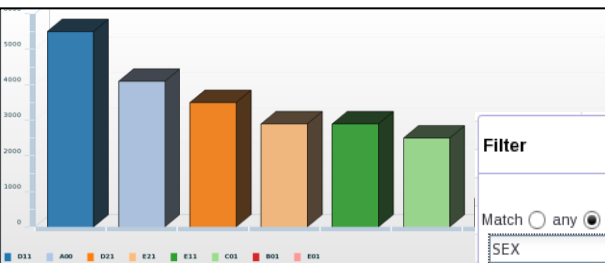
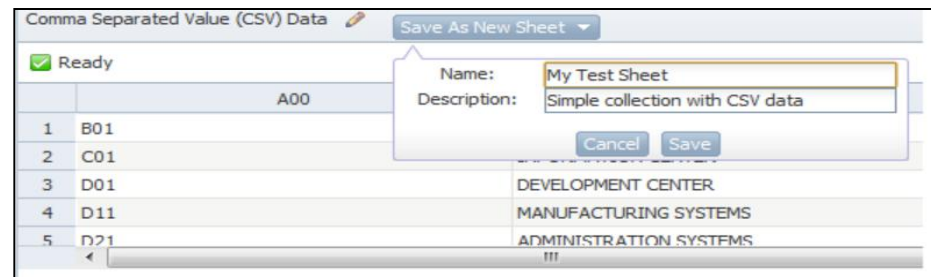
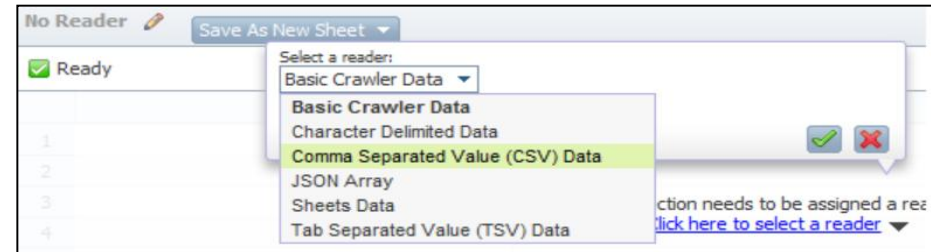
▪ Data exploration, manipulation and analysis

- BigSheets



三個步驟輕易完成海量資料探索

1. **Locate data of interest**
2. **Work with built-in “sheets” editor**
 - Add / delete columns
 - Filter data
 - Specify formulas to compute new values
 - Apply built-in or custom functions
3. **Visualizing Results**



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革命



InfoSphere Biginsight 實現海量資料的應用



Big Data Exploration

Find, visualize, understand all big data to improve decision making



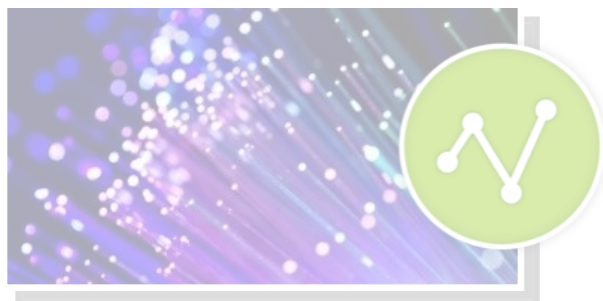
Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external data sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine

顛覆企業傳統ROI Improved business results

資訊革新論壇 The New ROI

資訊創造報酬率·迎接ROI大革命



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency



開箱即用的社群分析(Social Data Analytics)加速探索客戶的喜好

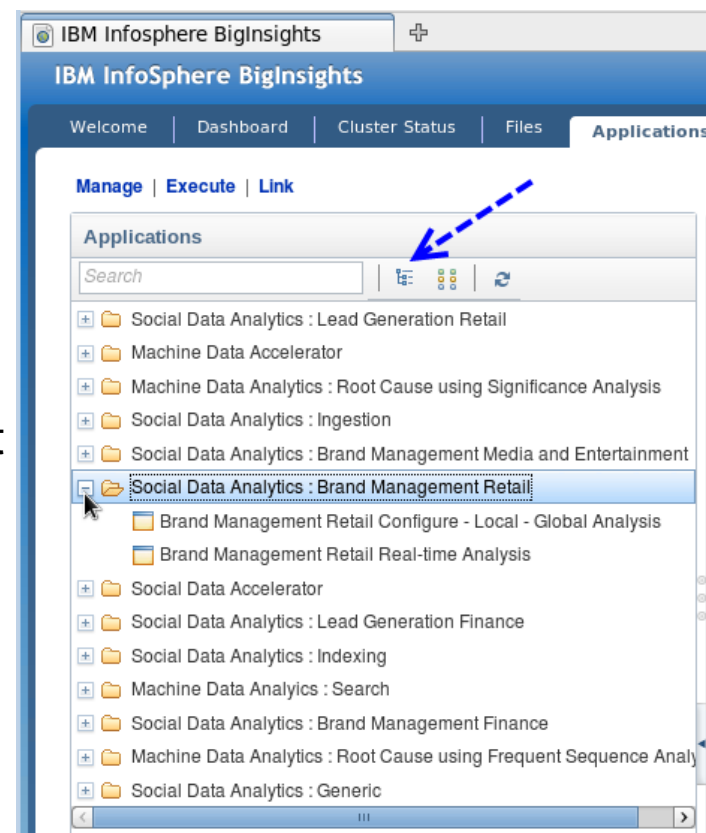
■ Outputs/Measures

- Buzz
- Sentiment
- Intent to buy/start service
- Intend to attend/see

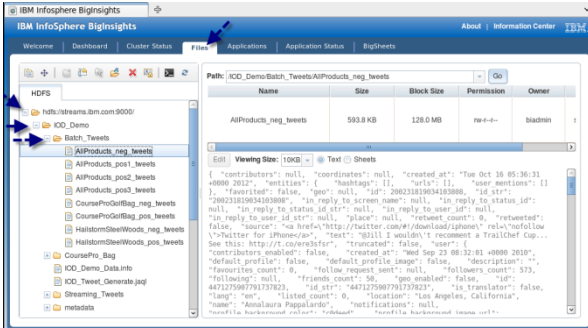
■ Example use cases

- Retail – Lead generation, Brand management
- Financial – Lead generation ,Brand management
- Media & Entertainment: Brand management
- Generic

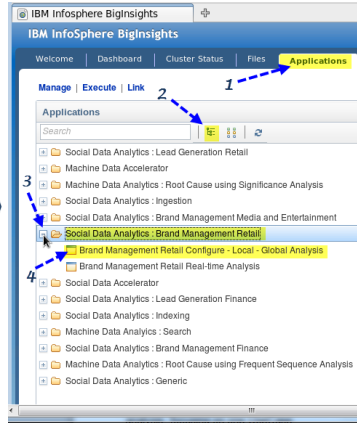
■ Visualization using BigSheets



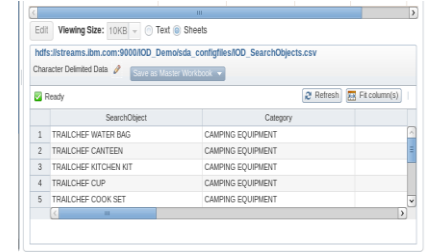
利用SDA快速實現品牌喜好度的分析



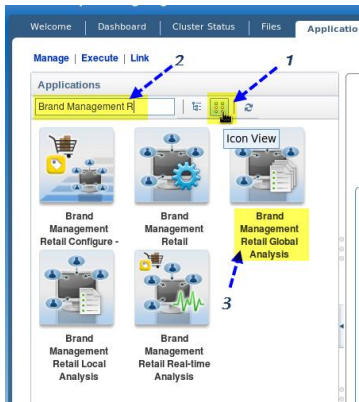
Import Data with appropriate keywords (Decahose + Powertrack + Boardreader blogs+ forums)



Pick Applications e.g. Brand Management Retail



Provide inputs for SDA using the right schema for retail



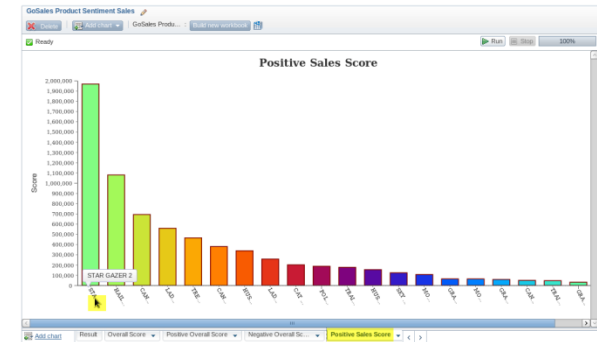
Run Config-Local-Global Analysis

顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革命

Status	Job Status	Job ID	Type	Start Time	End Time	Job ID	Job Def
✓	SUCCEEDED	000000-1210123555	java	Oct 17, 2012 4:20:00 PM	Oct 17, 2012 4:20:39 PM	job_201210162347_0001	
✓	SUCCEEDED	000000-1210123555	java	Oct 17, 2012 4:20:39 PM	Oct 17, 2012 4:20:54 PM	job_201210162347_0002	
✓	SUCCEEDED	000000-1210123555	java	Oct 17, 2012 4:20:54 PM	Oct 17, 2012 4:21:08 PM	job_201210162347_0003	
✓	SUCCEEDED	000000-1210123555	java	Oct 17, 2012 4:21:08 PM	Oct 17, 2012 4:21:21 PM	job_201210162347_0004	
✓	processTwitterDecision	000000-1210123555	switch	Oct 17, 2012 4:21:22 PM	Oct 17, 2012 4:21:22 PM		
✓	HL_sample_base_sender2	000000-1210123555	switch	Oct 17, 2012 4:21:22 PM	Oct 17, 2012 4:21:22 PM		



View visualizations using Dashboards and Collections

InfoSphere Biginsight 實現海量資料的應用



Big Data Exploration

Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external data sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency

開箱即用的機器事件分析(Machine Data Analytics)解決系統難解問題

▪ **Business requirement**

- Understand and anticipate expensive outages
- Reduce downtime

▪ **Solution Overview**

- Ingest data at rest
- Extractors for **common IT log formats**
 - Syslogs
 - WebSphere
 - Web access
 - Datapower
 - CSV (or other delimited)
 - Generic – customizable
- Use faceted **search across multiple system logs** to find events of interest
- Sessionize log data
 - Sequences of events based on a partitioning key for a particular window of time
 - Events from one or two sets of log data that are joined (target event, and potential influencing event)
- Analyze the sessions
 - Identify **frequent sequences** leading to an error condition
 - Identify which events have the **most significant impact** on an error condition

顛覆企業傳統ROI

資訊革新論壇 Return on Information **The New ROI**

資訊創造報酬率，迎接ROI大革命



透過MDA使用時間區間關聯事件 (Time Range)

Search

Search for : *
 Filtered by: [Range/ 2012-03-28 18:52:05.000- 2012-03-28 18:52:07.000](#)

Time Range Search

Filter by Range

From: 2012-03-28 18:52:05
 To: 2012-03-28 18:52:07
 Last 1 Hour

Filter by Facet

- batchId (>=391)
- CodesAndValues (>=62)
- ConnReceived (>=9)
- ConnReplied (>=33)
- DateTime (>=391)
- EventTypeWS (>=7)
- ID (>=12)
- IPAddress (>=206)
- LogDate (>=391)
- LogDateTime (>=391)
- LoggerApp (>=329)
- LoggerDomain (>=261)
- LoggerLevel (>=384)
- LogTime (>=391)
- logType (>=391)
 - datapower (384)
 - was (7)

Time Graph

Results

"2012-03-28 18:52:05.000 GMT" ← **Normalized Time: 2012-03-28 18:52:05.000 GMT**
 LogLevel : info
 LogApp : audit
 logType : datapower
 Tid : 180687
[Less..](#)
 Raw Log: Wed Mar 28 2012 11:52:05 [audit][info] : tid(180687); (SYSTEM.CO:*:172.21.162.7); source-http 'FSH_Helloworld' - Conn

"2012-03-28 18:52:05.000 GMT"
 LogLevel : info
 LogDomain : NCAL
 LogApp : latency
 logType : datapower
 Tid : 404524999
[More..](#)

"2012-03-28 18:52:05.000 GMT"
 LogLevel : info
 LogDomain : National
 LogApp : latency
 logType : datapower
 Tid : 372836257
[More..](#)

"2012-03-28 18:52:05.000 GMT"
 LogLevel : notice



透過MDA使用關鍵字關聯事件(patterns matching)

IBM Introsphere BigInsights

Search

Search for : "Estimates"
 Filtered by: [Range/ 2012-03-28 18:52:05.000- 2012-03-28 18:52:07.000](#)

Wildcard Search → "Estimates" Found 11

Filter by Range

From: 2012-03-28 18:52:00
 To: 2012-03-28 18:52:59
 Last 1 Hour

Time Graph

Filter by Facet

- CodesAndValues (>=1)
 - 200 [OK] (1)
- DateTime (>=11)
- EventTypeWS (>=5)
- LogDate (>=11)
- LogDateTime (>=11)
- LoggerApp (>=2)
- LoggerDomain (>=2)
- LoggerLevel (>=6)
- LogTime (>=11)
- logType (>=11)**
 - datapower (6)
 - was (5)
- nodeName (>=11)
- postalCode (>=1)
- serverName (>=11)**
 - esbcdcpdp030 (6)
 - CAMyEstimates_003 (5)

Results

```

"2012-03-28 18:52:05.000 GMT"
  LogLevel : info
  LogDomain : National
  LogApp : crypto
  logType : datapower
  Tid : 732975504
  Less.
  Raw Log: Wed Mar 28 2012 11:52:05 [National][crypto][info] valcred(VC_CA_MyEstimates); tid(732975504); certificate validation succeeded for /C=US/postalCode=94588/ST=CA/L=Pleasanton/streetAd
  Plaza, Oakland, CA/O=Kaiser Permanente/OU=EIS/OU=Hosted by Kaiser Foundation Health Plan, Inc./OU=Comodo InstantSSL/CN=soa-services1.kp.org' against 'VC_CA_MyEstimates'

"2012-03-28 18:52:05.000 GMT"
  LogLevel : info
  URLBackside Server : https://soa-services1.kp.org:9457/CA_MyEstimates_Estimate/Estimate
  logType : datapower
  Tid : 732975504
  Less.
  Raw Log: Wed Mar 28 2012 11:52:05 [National][ws-proxy][info] wsgw(WSP_Estimate); tid(732975504); Using Backside Server. https://soa-services1.kp.org:9457/CA_MyEstimates_Estimate/Estimate

"2012-03-28 18:52:06.226 GMT"
  Eventtype : 0
  ThreadID : 00000127
  Shortname : SystemOut
  logType : was
  Less.
  Raw Log: [3/28/12 11:52:06.226 PDT] 00000127 SystemOut O 11:52:06.225 INFO HistoricalEstimateSummaryProcessor:58 - before Utils.getNewOracleConnection execution time(ms):0
          
```



透過MDA獨家串聯session分析找出root cause

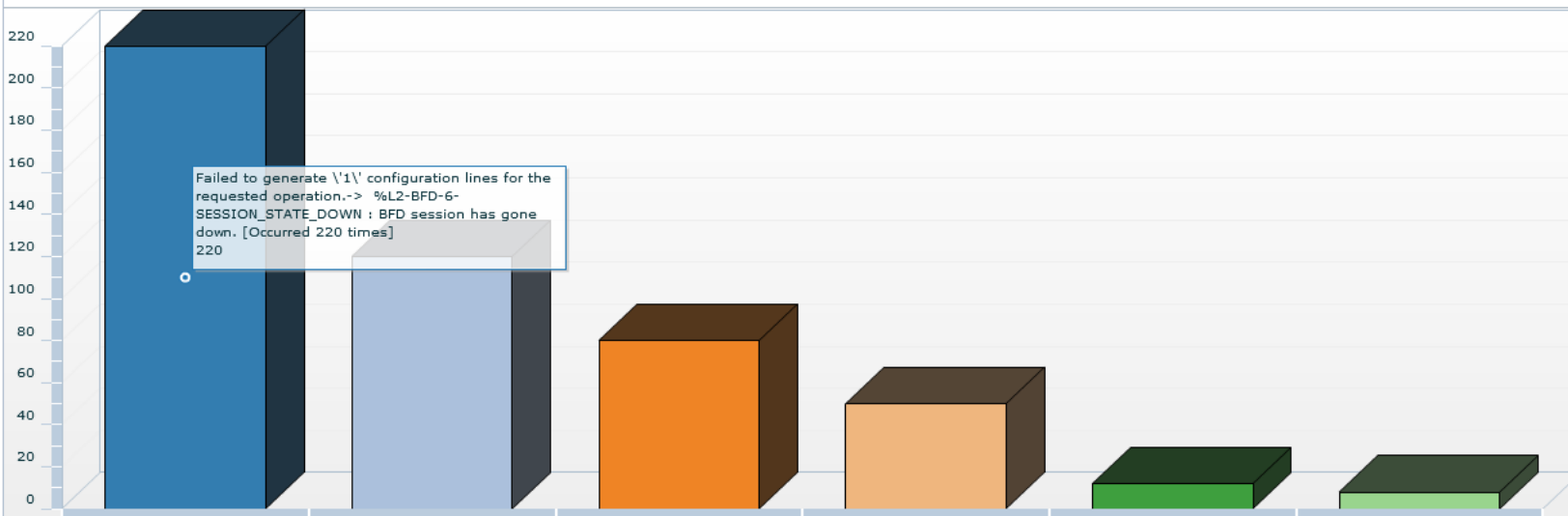
Data Collections > View Results

Pattern subsequences around errors

Delete | Pattern subsequ... : Build new collection

Ready

Run Stop 100%



- Failed to generate '\1\' configuration lines for the requested operation. -> %L2-BFD-6-SESSION_STATE_DOWN : BFD session has gone down. [Occurred 220 times]
- Failed to generate '\1\' configuration lines for the requested operation. -> %L2-BFD-6-SESSION_STATE_DOWN : BFD session has gone down. -> session recovery failed. [Occurred 120 times]
- Authentication to privilege level 15 failed by ipngn on vty2 -> Failed to generate '\1\' configuration lines for the requested operation. -> %L2-BFD-6-SESSION_STATE_DOWN : BFD session has gone down. [...]
- Failed to generate '\1\' configuration lines for the requested operation. -> Failed to generate '\1\' configuration lines for the requested operation. [Occurred 50 times]



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革新

Return on Information

InfoSphere Biginsight 實現海量資料的應用



Big Data Exploration

Find, visualize, understand all big data to improve decision making



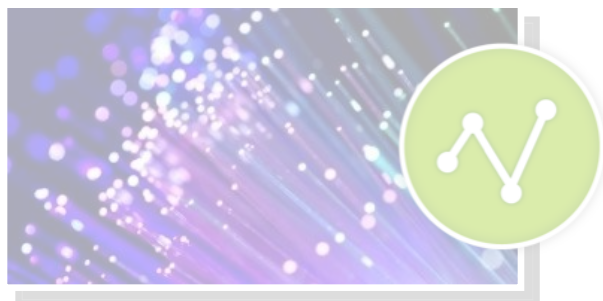
Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external data sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine

顛覆企業傳統ROI Improved business results

資訊革新論壇 The New ROI

資訊創造報酬率，迎接ROI大革命

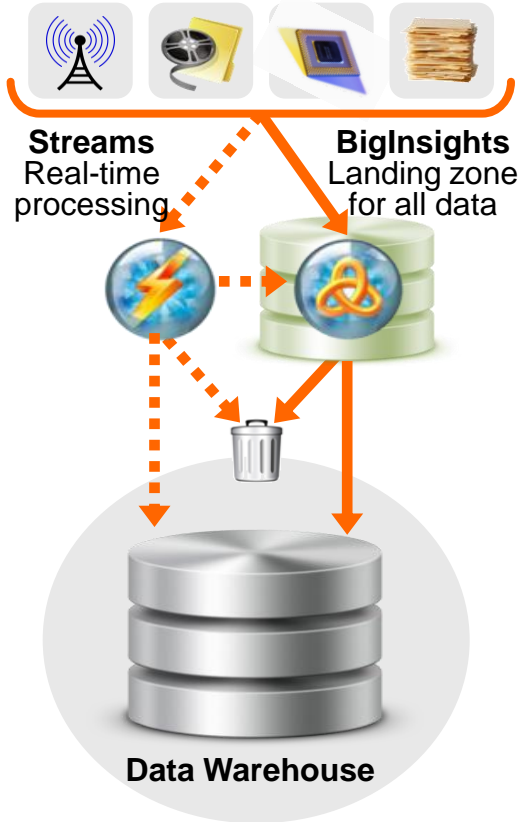


Data Warehouse Augmentation

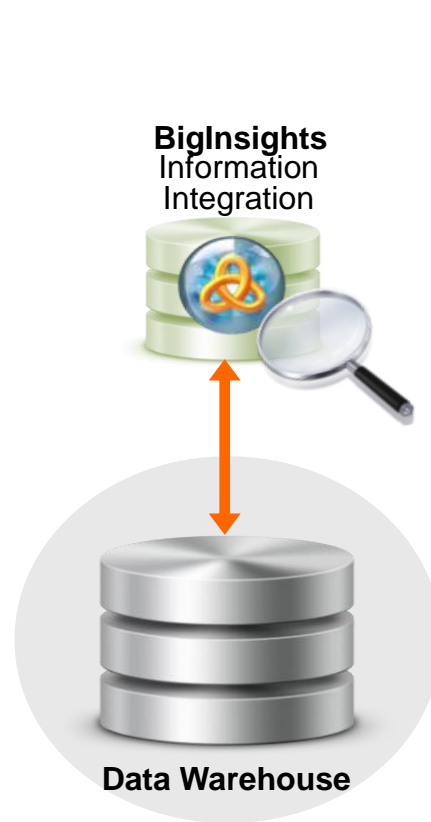
Integrate big data and data warehouse capabilities to increase operational efficiency

上下整合Big Data技術 三種整合資料倉儲的應用

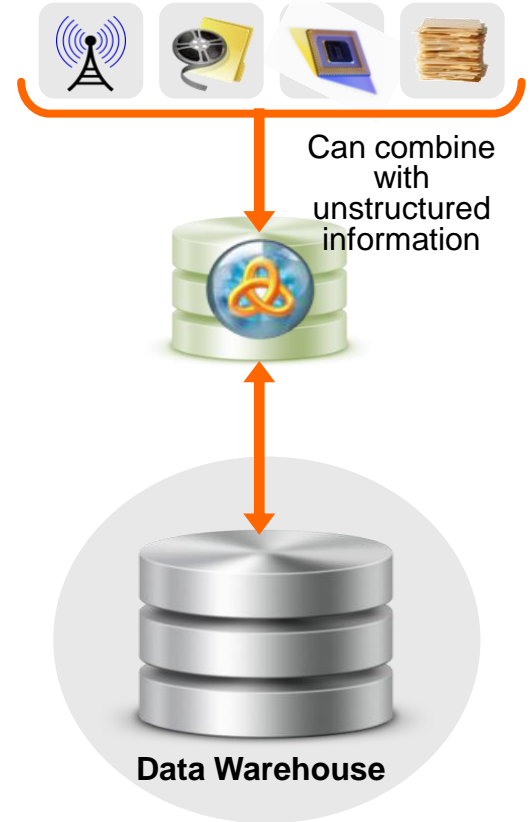
1 Pre-Processing Hub



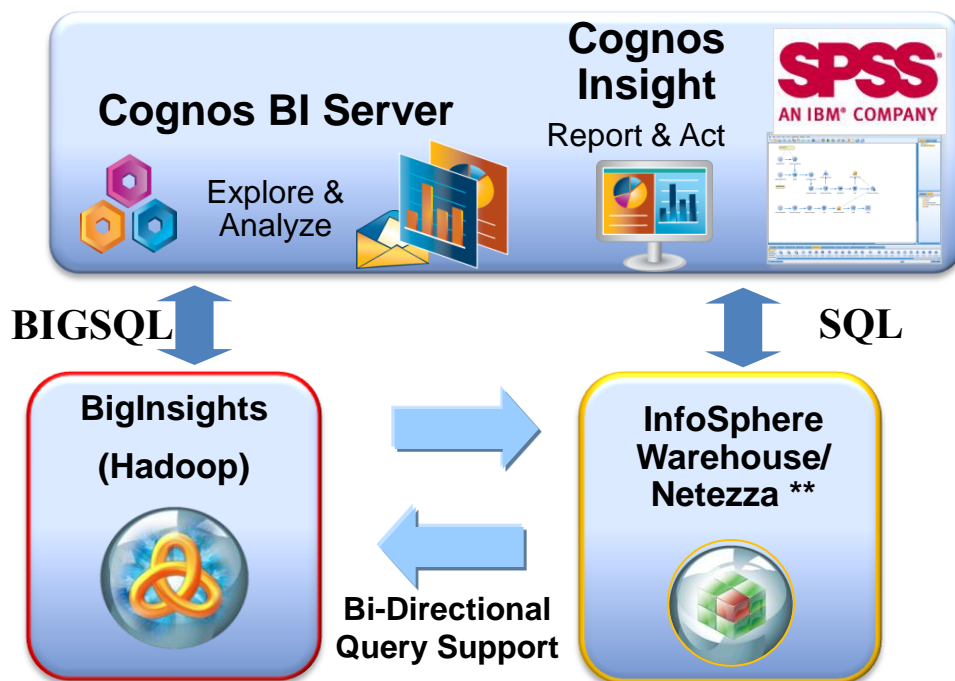
2 Query-able Archive



3 Exploratory Analysis



IBM BigSQL技術使的資料倉儲可以幾無限制的儲存資料



- Cognos BI can issue SQL Queries against data managed by BigSQL in BigInsights
- The IBM BigData platform supports bi-directional queries between BigInsights and the EDW
- Key Benefits:
 - Existing SQL based applications can leverage the BigData platform
 - EDW optimized from size and performance perspective
 - Provides cost effective and flexible big data storage and analysis

有了BigSQL就可以使用熟悉的SQL完成BigData的作業

The screenshot shows the Eclipse IDE interface. The main editor displays a SQL script:

```

1 SELECT
2   L_RETURNFLAG,
3   L_LINESTATUS,
4   SUM(L_QUANTITY) AS SUM_QTY,
5   SUM(L_EXTENDEDPRI) AS SUM_BASE_PRICE,
6   SUM(L_EXTENDEDPRI*(1-L_DISCOUNT)) AS SUM_DISC_PRICE,
7   SUM(L_EXTENDEDPRI*(1-L_DISCOUNT)*(1+L_TAX)) AS SUM_CHARGE,
8   AVG(L_QUANTITY) AS AVG_QTY,
9   AVG(L_EXTENDEDPRI) AS AVG_PRICE,
10  AVG(L_DISCOUNT) AS AVG_DISC,
11  COUNT(*) AS COUNT_ORDER
12 FROM LINEITEM
13 WHERE L_SHIPDATE <= '1998-09-02'
14 GROUP BY L_RETURNFLAG, L_LINESTATUS
15 ORDER BY L_RETURNFLAG, L_LINESTATUS
16

```

The 'Properties for BigSQL - 170.224.193.37' dialog box is open, showing the 'Big SQL JDBC Connection Properties' configuration. The 'Drivers' dropdown is set to 'IBM Big SQL JDBC Driver 2.0 Default'. The 'Optional' tab is selected, showing the following properties:

- Catalog: default
- Host: 170.224.193.37
- Port number: 7052
- User name: biadmin
- Password: [masked]
- Save password
- Connection URL: jdbc:bigsql://170.224.193.37:7052/default

Buttons for 'Test Connection', 'OK', and 'Cancel' are visible at the bottom of the dialog.

- Using existing SQL tooling against BigData
- Same setup as for existing SQL sources!!
- Support for “standard” authentication!!

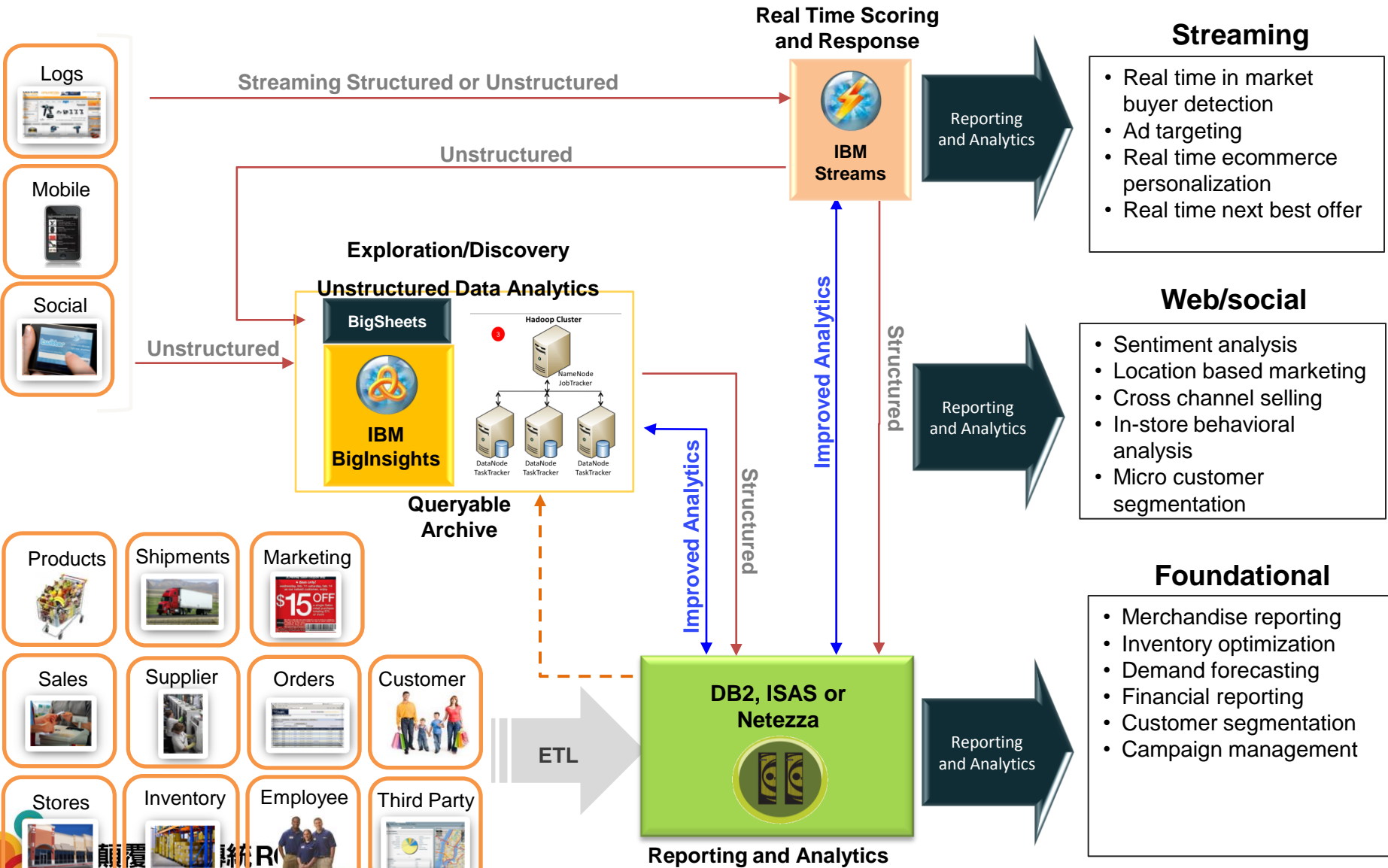
顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率，迎接ROI大革命

Return on Information

IBM 海量資料分析平台間的協同運作



- ### Streaming
- Real time in market buyer detection
 - Ad targeting
 - Real time ecommerce personalization
 - Real time next best offer

- ### Web/social
- Sentiment analysis
 - Location based marketing
 - Cross channel selling
 - In-store behavioral analysis
 - Micro customer segmentation

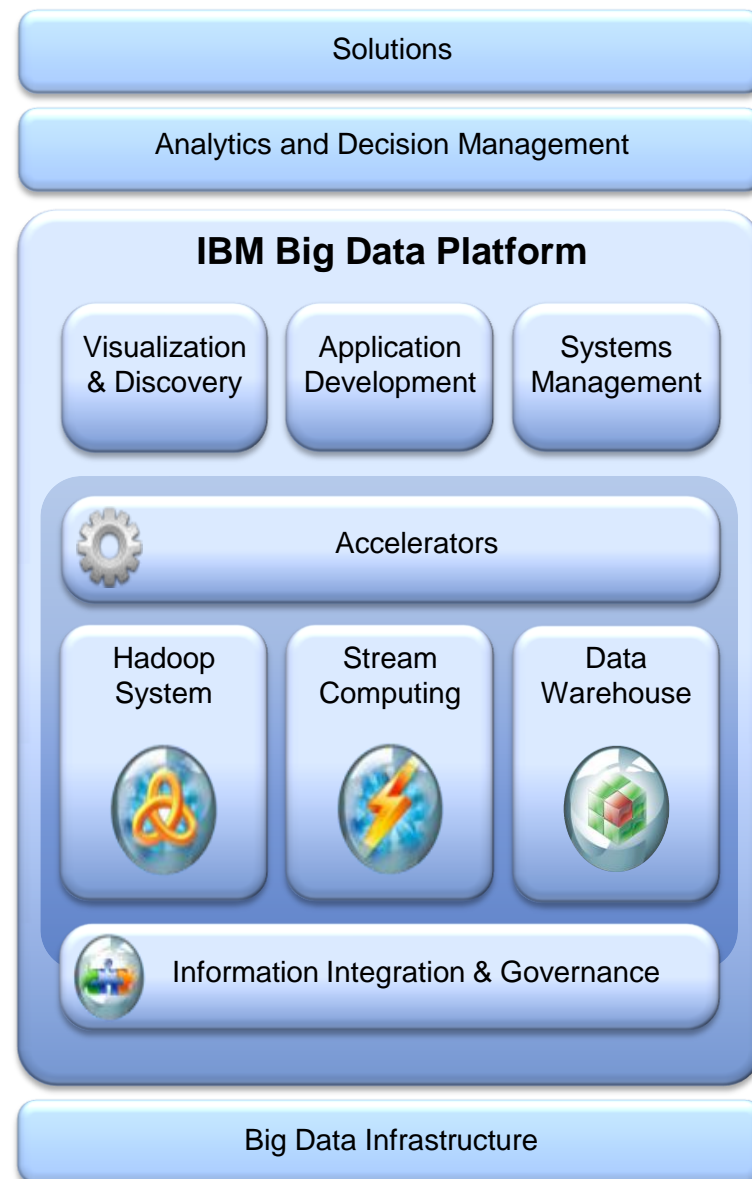
- ### Foundational
- Merchandise reporting
 - Inventory optimization
 - Demand forecasting
 - Financial reporting
 - Customer segmentation
 - Campaign management

The Big Data Platform

- Process any type of data
- Built-for-purpose engines
- Analyze data in motion
- Manage and govern data in the ecosystem
- Enterprise data integration
- Grow and evolve on current infrastructure

Visit the IBM Big Data Track
and Big Data Hub

www.ibmbigdatahub.com



顛覆企業傳統ROI

資訊革新論壇 *The New ROI*

資訊創造報酬率 · 迎接ROI大革命

Return on Information