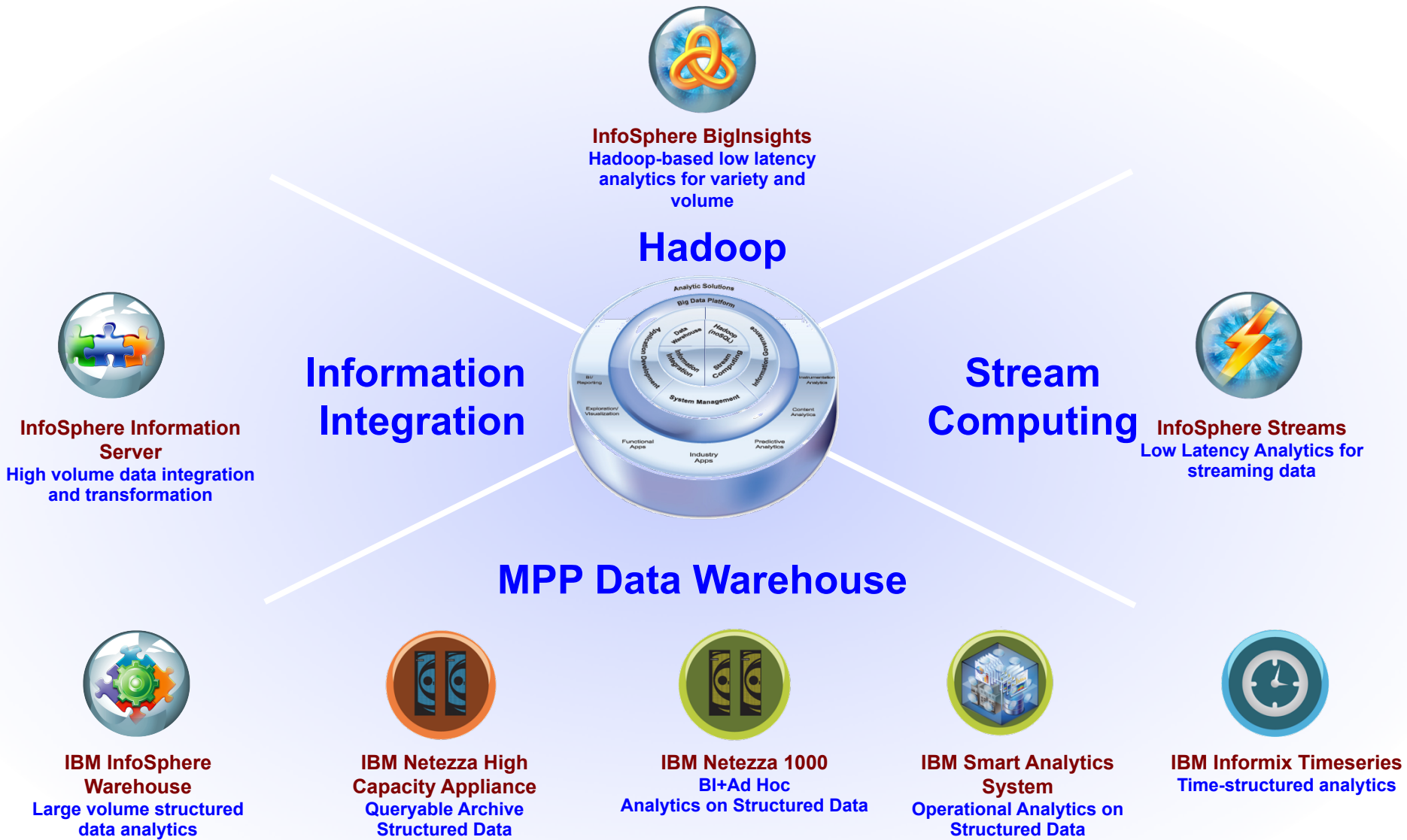


# IBM's InfoSphere BigInsights: Smart Analytics for Big Data



Ayhan Önder

# The IBM Big Data Platform



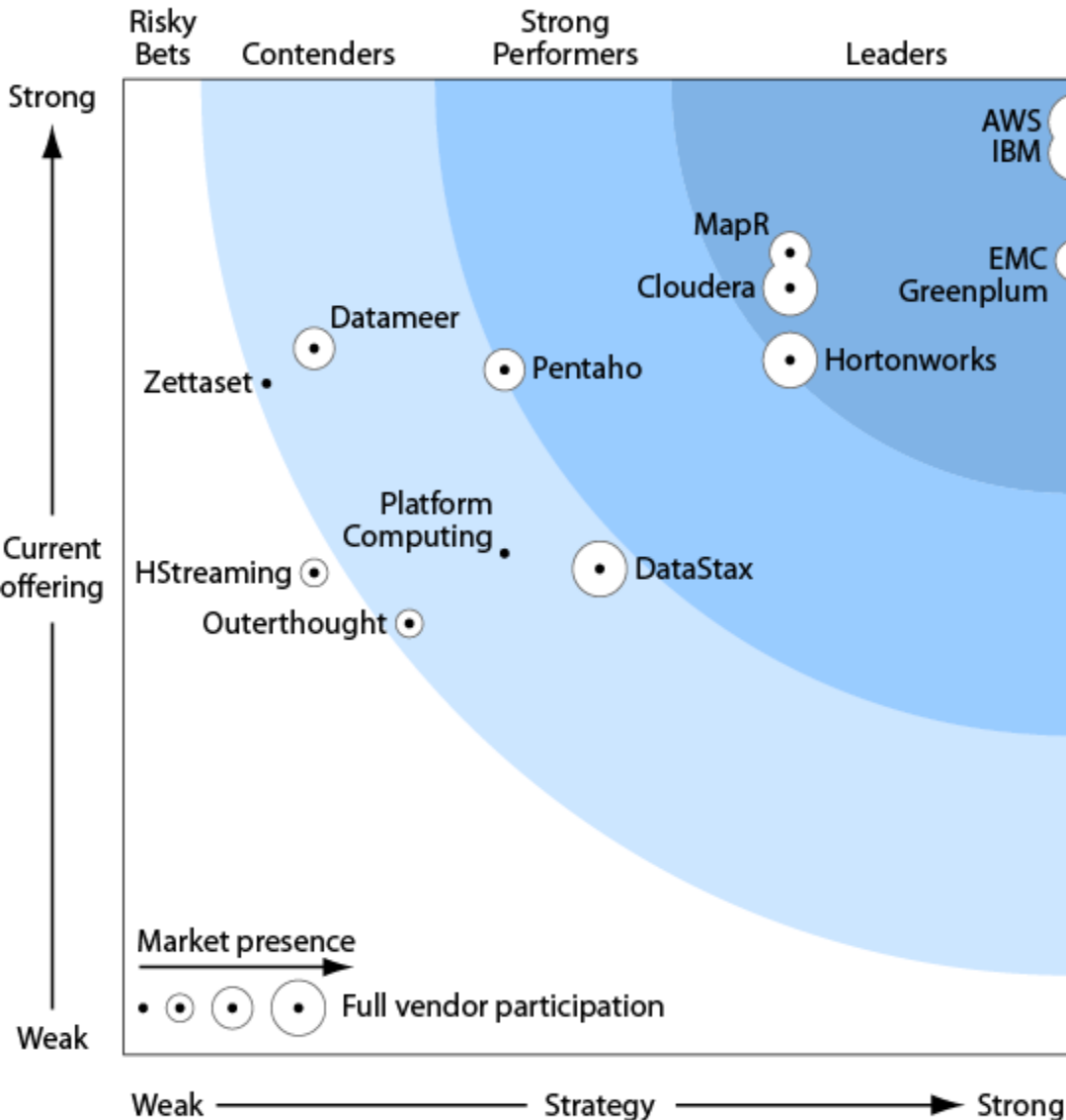


# BigInsights Summary

- **BigInsights = analytical platform for persistent “Big Data”**
  - Based on open source & IBM technologies
  - Managed like a start-up . . . .
  
- **Distinguishing characteristics**
  - Built-in analytics . . . . *Enhances business knowledge*
  - Enterprise software integration . . . . *Complements and extends existing capabilities*
  - Production-ready platform with tooling for analysts, developers, and administrators. . . . *Speeds time-to-value; simplifies development and maintenance*
  
- **IBM advantage**
  - Combination of software, hardware, services and advanced research



# First Ever Forrester Wave on Hadoop



***“IBM has the deepest Hadoop platform and application portfolio. IBM, an established EDW vendor, has its own Hadoop distribution; an extensive professional services force working on Hadoop projects; extensive R&D programs developing Hadoop technologies; connections to Hadoop from its EDW.”***

*–The Forrester Wave™: Enterprise Hadoop Solutions, 1Q12*


# Big Database Vendors Adopt Hadoop

## Microsoft climbs onto Hadoop bandwagon

Joint Microsoft-Hortonworks effort aims to deliver a Hadoop distribution for Windows Server and Windows Azure

By Jaikumar Vijayan

October 12, 2011 03:40 PM ET

 Add a comment

 Like

6

 +1

0

Computerworld - Microsoft is the latest of the world's top IT vendors to climb aboard the Hadoop 'big data' bandwagon.

## Oracle Jumps on Hadoop Bandwagon with Cloudera

By Jennifer LeClaire

January 10, 2012 12:10PM

Follow NewsFactor on [twitter!](#)

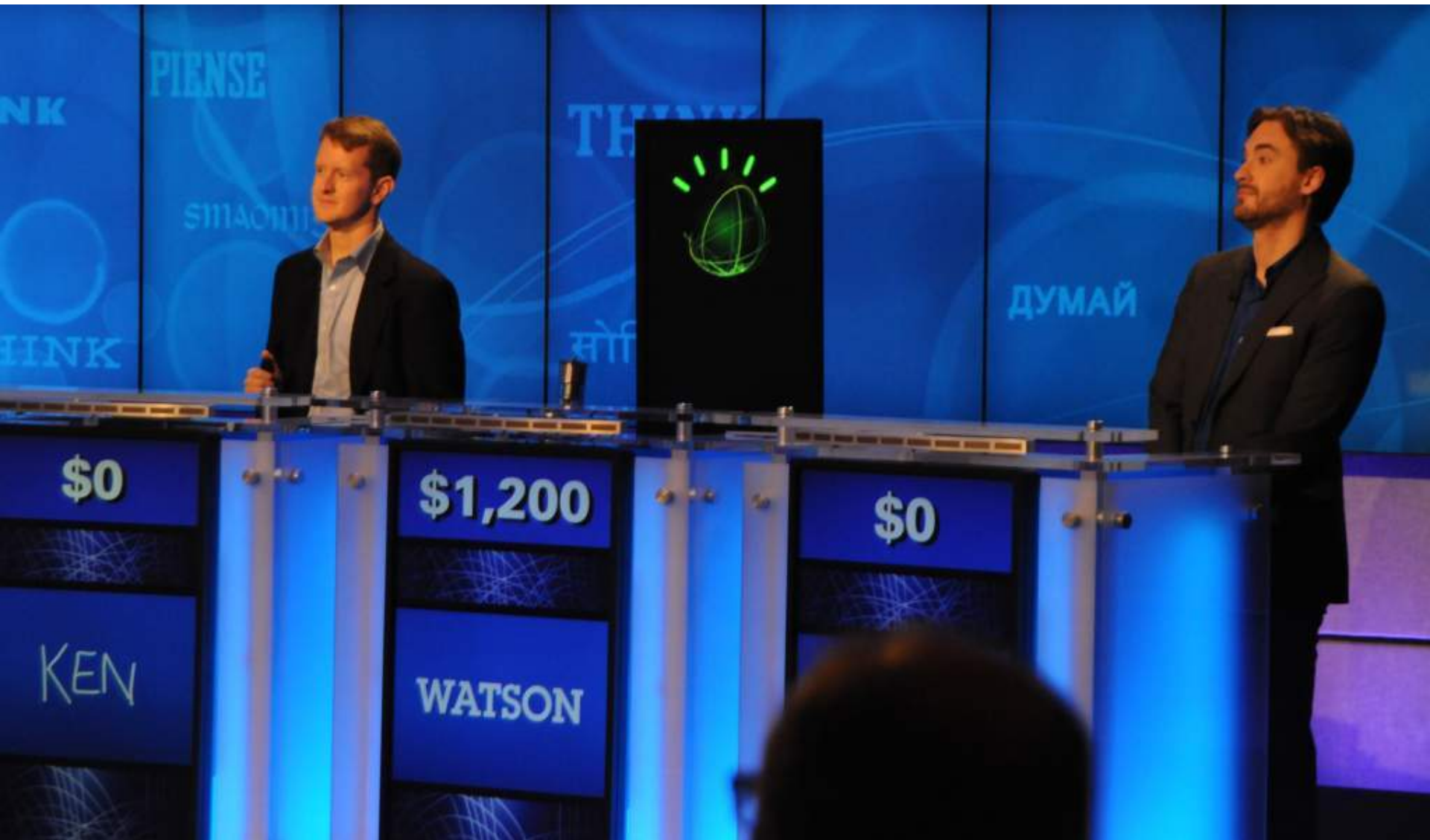
 SHARE 

Charles King, principal analyst at Pund-IT, said the Oracle-Cloudera partnership is a win-win. As he sees it, Oracle is coming late to the Hadoop-based Big Data party with Cloudera. He points to EMC's Greenplumb acquisition and IBM's Netezza acquisition, as well as HP's Vertica acquisition in this space as evidence.

## Teradata taps Hortonworks to improve Hadoop story

By Derrick Harris | Feb. 21, 2012, 6:33am PT | No Comments

# IBM Watson



# InfoSphere BigInsights

## Platform for **volume, variety, velocity** -- **V<sup>3</sup>**

- Enhanced Hadoop foundation

## Analytics for **V<sup>3</sup>**

- Text analytics & tooling

## Usability

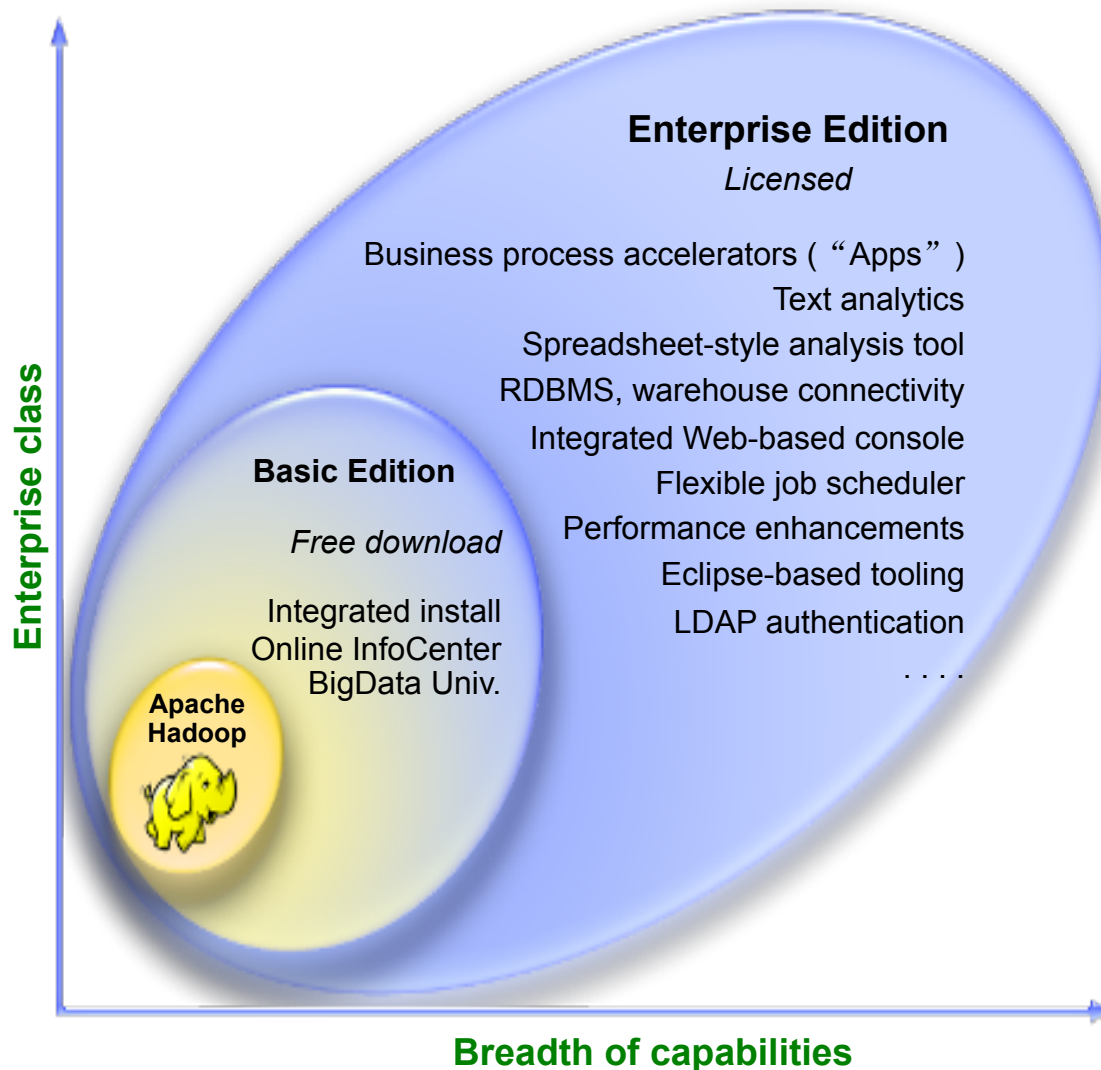
- Web console
- Integrated install
- Spreadsheet-style tool
- Ready-made “apps”

## Enterprise Class

- Storage, security, cluster management

## Integration

- Connectivity to DB2, Netezza, JDBC databases



# BigInsights Content

Function	Version	Basic Edition	Enterprise Edition
Integrated Install		Inc	Inc
Hadoop (including common utilities, HDFS, MapReduce framework)	1.0.0	Inc	Inc
Jaql (programming / query language)	0.5.2	Inc	Inc
Pig (programming / query language)	0.9.1	Inc	Inc
Flume (data collection/aggregation)	0.9.4	Inc	Inc
Hive (data summarization/querying)	0.8.0	Inc	Inc
Lucene (text search)*	3.3.0	Inc	Inc
Zookeeper (process coordination)	3.3.4	Inc	Inc
Avro (data serialization)	1.5.1	Inc	Inc
HBase (real time read/write)	0.90.5	Inc	Inc
Oozie (workflow/ job orchestration)	2.3.1	Inc	Inc
Online documentation		Inc	Inc
Capability to integrate with JDBC sources through general-purpose Jaql module		Inc	Inc
Capability to integrate with DB2, InfoSphere Warehouse (DB2 UDF samples to submit jobs, and read results from BigInsights)		Inc	Inc



# BigInsights Content (cont' d)

Function	Basic Edition	Enterprise Edition
Capability to integrate with R (Jaql module to invoke R statistical capabilities from BigInsights)	n/a	<b>Inc</b>
Capability to integrate with Netezza, DB2 LUW with DPF from Jaql	n/a	<b>Inc</b>
LDAP authentication and additional security features	n/a	<b>Inc</b>
Integrated Web Console*	n/a	<b>Inc</b>
Integrated workflow capabilities and flexible job scheduler	n/a	<b>Inc</b>
Platform performance enhancements (Adaptive MapReduce, efficient processing of compressed text files, large-scale text indexing, etc.)	n/a	<b>Inc</b>
Text analytics	n/a	<b>Inc</b>
Eclipse plugins for text analytic development, Jaql, Hive, Java	n/a	<b>Inc</b>
Ready-made “apps” for data import/export, Web crawl, Boardreader., etc.	n/a	<b>Inc</b>
Web-based application catalog	n/a	<b>Inc</b>
Spreadsheet-like analytical tool	n/a	<b>Inc</b>
IBM support	<b>Opt</b>	<b>Inc</b>
InfoSphere Streams (limited use license)	n/a	<b>Inc*</b>
Unlimited storage	n/a	<b>Inc*</b>

# BigInsights Basic Edition

Open Source

IBM

## Infrastructure

Integrated installer

ZooKeeper

Jaql

Pig

Oozie

HBase

Hive

Lucene

MapReduce

HDFS

## Connectivity

JDBC

Flume

# BigInsights Enterprise Edition

Open Source

IBM

Optional IBM and partner offerings

**Analytics and discovery**

- Text processing engine and library
- BigSheets

**“Apps”**

- Web Crawler
- Boardreader
- Distrib file copy
- ...
- DB export
- DB import
- Hive query
- Pig query
- Jaql query

**Administrative and development tools**

**Web console**

- Monitor cluster health
- Add / remove nodes
- Start / stop services
- Inspect job status
- Inspect workflow status
  - Deploy apps
  - Launch apps / jobs
- Work with distrib file system
- Work with spreadsheet interface
- Support REST-based API
- ...

**Eclipse plug-ins**

- Text analytics
- MapReduce programming
  - Jaql development
- Hive query development

**Infrastructure**

- Integrated installer
- Enhanced security
- ZooKeeper
- Oozie
- Lucene
- Flexible scheduler
- Jaql
- HBase
- Adaptive MapReduce
- GPFS (beta)
- Pig
- Hive
- MapReduce
- HDFS

**Connectivity**

- JDBC
- Flume
- Netezza
- Streams\*
- DB2
- Streams
- R

# BigInsights: Value Beyond Open Source



## ■ Technical differentiators

- Built-in analytics
  - Text processing engine, annotators, Eclipse tooling
  - Interface to project R (statistical platform)
- Enterprise software integration (DBMS, warehouse)
- Spreadsheet-style analytical tool for analysts
- Ready-made business process accelerators
- Integrated installation of supported open source and IBM components
- Web Console for administration and application access
- Platform enrichment: additional security, performance features, . . .
- Standard IBM licensing agreement and world-class support
- More to come in future releases!

## ■ Business benefits

- Quicker time-to-value due to IBM technology and support
- Reduced operational risk
- Enhanced business knowledge with flexible analytical platform
- Leverages and complements existing software assets

# BigInsights and the data warehouse



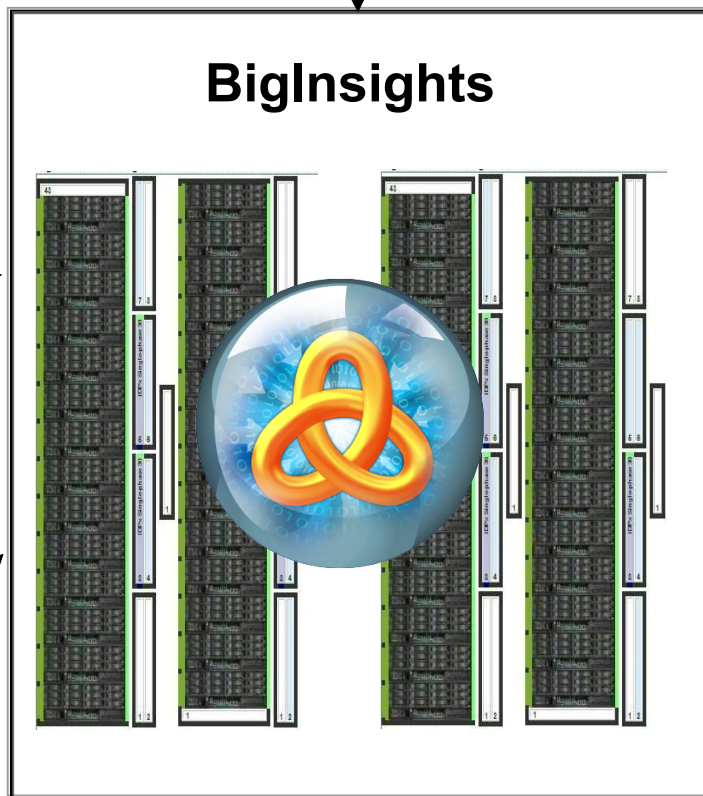
Big Data analytic applications



Traditional analytic tools

facebook

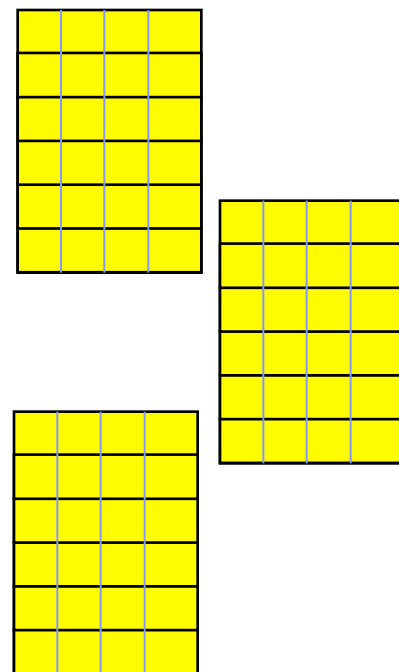
twitter



BigInsights

*Filter Transform Aggregate*

Data warehouse



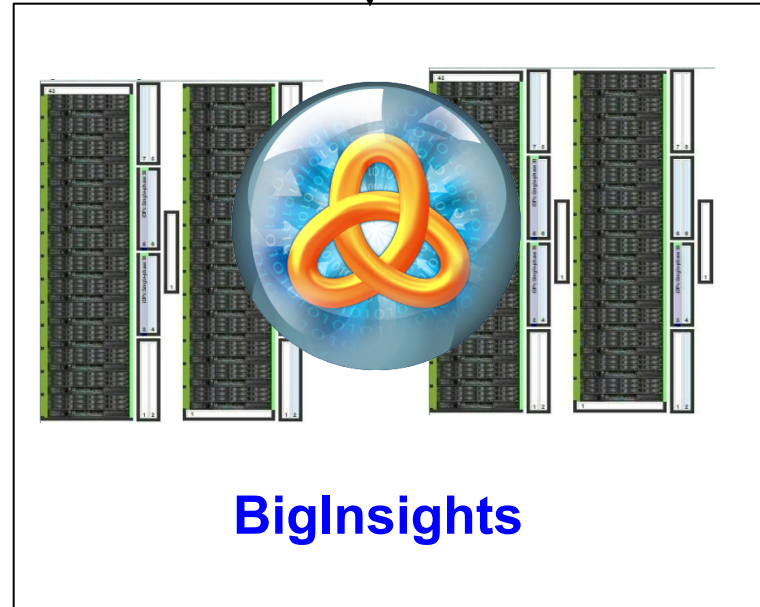
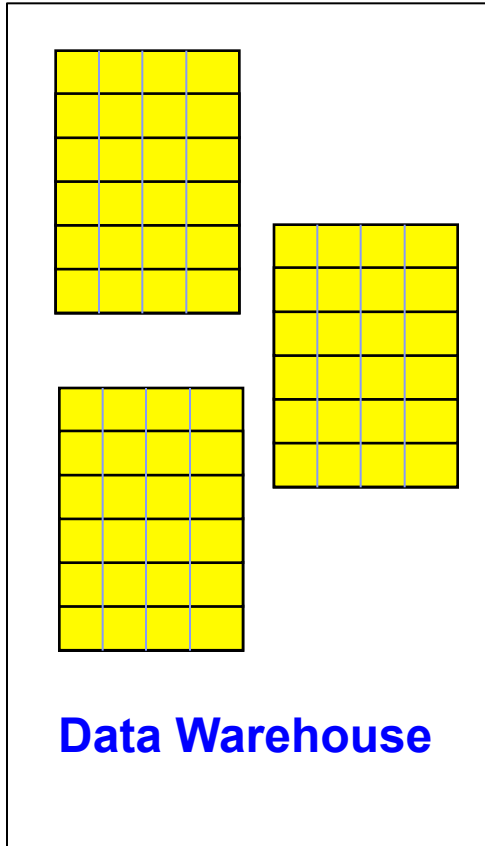
# BigInsights and the data warehouse



Traditional analytic tools



Big Data analytic applications



- *Query-ready archive for “cold” warehouse data*

# Growing Ecosystem of Solutions

## IBM Solutions



### Cognos Consumer Insights

Social media analytics solution that uses BigInsights

### IBM Content Analytics

Unlock valuable business insight from unstructured data.

### InfoSphere DataStage

High performance extract, transform, and load from multiple data sources, including BigInsights

## Partner Solutions



... with more to come

# A Closer Look at BigInsights . . . .





# About the BigInsights Platform



- **Flexible, enterprise-class support for processing large volumes of data**
  - Based on Google’s MapReduce technology
  - Inspired by Apache Hadoop; compatible with its ecosystem and distribution
  - Well-suited to batch-oriented, read-intensive applications
  - Supports wide variety of data
  
- **Enables applications to work with thousands of nodes and petabytes of data in a highly parallel, cost effective manner**
  - CPU + disks = “node”
  - Nodes can be combined into clusters
  - New nodes can be added as needed without changing
    - Data formats
    - How data is loaded
    - How jobs are written

# The MapReduce Programming Model

- **"Map" step:**
  - Input split into pieces
  - Worker nodes process individual pieces in parallel (under global control of the Job Tracker node)
  - Each worker node stores its result in its local file system where a reducer is able to access it
- **"Reduce" step:**
  - Data is aggregated ("reduced" from the map steps) by worker nodes (under control of the Job Tracker)
  - Multiple reduce tasks can parallelize the aggregation

# What is the Hadoop Distributed File System?

- HDFS stores data across multiple nodes
- HDFS assumes nodes will fail, so it achieves reliability by replicating data across multiple nodes
- The file system is built from a cluster of *data nodes*, each of which serves up blocks of data over the network using a block protocol specific to HDFS.

# How To Create MapReduce Jobs

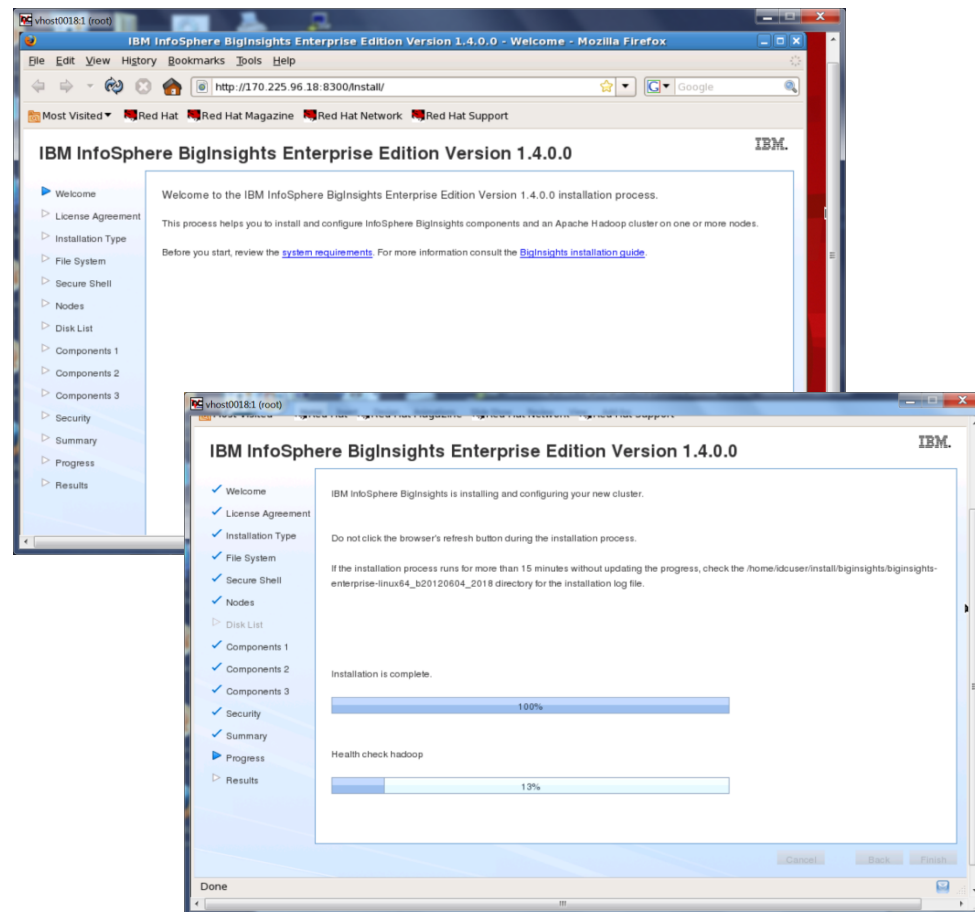
- **MapReduce development in Java**
- **Pig**
  - Open source language / Apache sub-project
- **Hive**
  - Open source language / Apache sub-project
  - Provides a SQL-like interface to Hadoop
- **Jaql**
  - IBM Research Invented query language
  - Very useful for loosely structured data
- . . .

## So What Does This Result In?

- Easy To Scale
- Fault Tolerant and Self-Healing
- Data Agnostic
- Extremely Flexible

# Web Installation Tool

- Seamless process for single node and cluster environments
- Integrated installation of all selected components
- Post-install validation of IBM and open source components



*No need to iteratively download, configure, and test multiple open source projects and their pre-requisite software.*

# Web Console

- **Manage BigInsights**
  - Inspect system health
  - Add / drop nodes
  - Start / stop services
  - Run / monitor jobs (applications)
  - Explore / modify file system
  - . . .
  
- **Launch applications**
  - Spreadsheet-like analysis tool
  - Pre-built applications (IBM supplied or user developed)
  
- **Publish applications**
  
- **Leverage community resources**

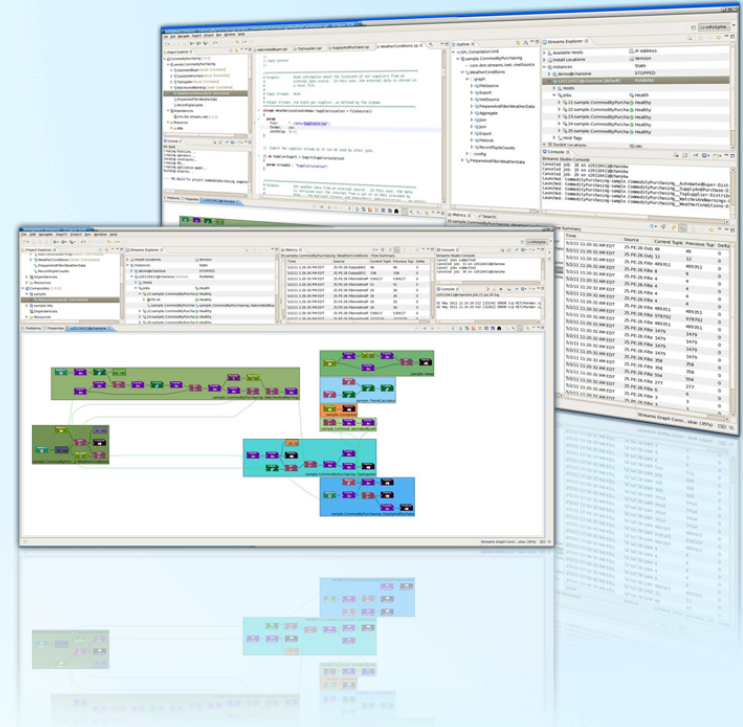
Type	Total Tasks	Successful Tasks	Failed Tasks	Killed Tasks	Warning Tasks	Pending Tasks	Start Time	End Time
map	2	1	0	1	0	0	N/A	Nov 22, 2011 8:50 PM
map	1	1	0	0	0	0	Nov 22, 2011 8:50 PM	Nov 22, 2011 8:50 PM
reduce	0	0	0	0	0	0	N/A	N/A

# Ease of Use for Developers and Users



## End-user visualization

Data exploration, crawling, and analytics



## Development Environment

Familiar coding and tooling environment, testing, and optimization



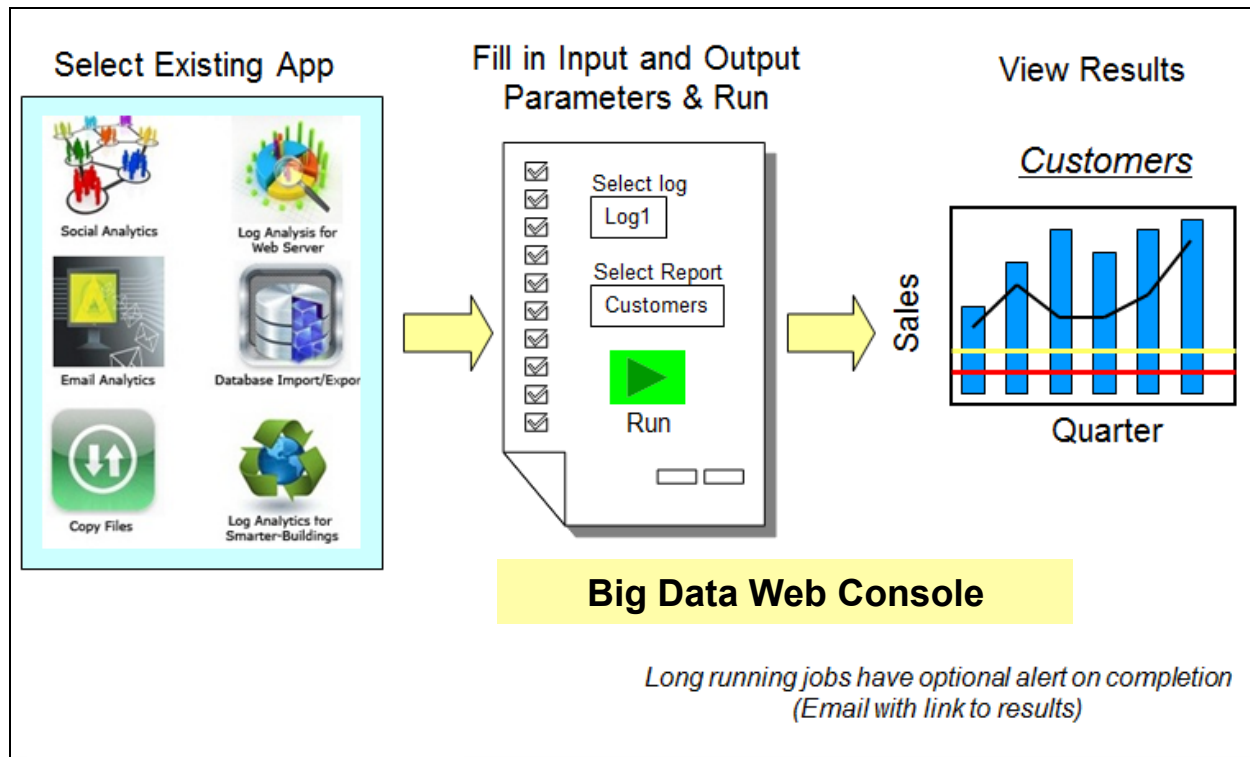
# Big Data Application Ecosystem

**Data integration scenario:**  
 Pre-defined work flows simplify loading data from various sources

- Work flows can be configured, deployed, executed and scheduled

**Application scenarios (web log, email, social media, ...):**

- Samples provide starting point, speed time to value

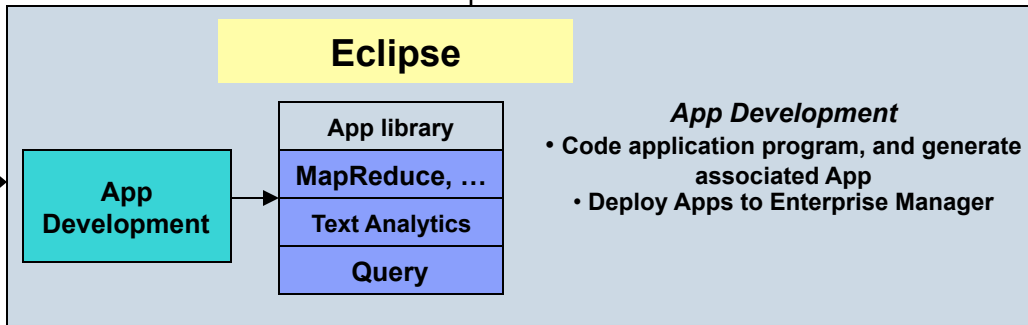


## Big Data Web Console

*Long running jobs have optional alert on completion (Email with link to results)*

**Development tooling:**

- Text analytics
- MapReduce
- Query languages
- ...



# Spreadsheet-like Analysis

- Web-based analysis and visualization

- Spreadsheet-like interface

- Define and manage long running data collection jobs
- Analyze content of the text on the pages that have been retrieved

The screenshot displays the IBM Watson interface with a data table, a menu for selecting sheet types, and a pie chart showing language coverage.

Language	PostSize	PostTitle	Published	
15	English	3148	What the "n@&*" Happened to Comics? Strong Women, Strong Girls and <Keyword>IBM Watson<Keyword>, a superbrain	2012-03-27 10:00:53
16	English	8176	<Keyword>IBM Watson<Keyword> - Tucson Watch Event for Jeopardy Day 1	2011-02-15 05:13:51
17	English	21567	<Keyword>IBM Watson<Keyword> - Research Team Answers Your Questions	2011-02-23 20:13:00
18	English	14056	<Keyword>IBM Watson<Keyword> and the Future of Work	2011-10-16 02:48:32
19	English	5374	<Keyword>IBM Watson<Keyword> Could Have A Career in Health Care	2012-02-15 08:22:08
20	Russian	5540	Data Mining / СынOPSIS<Keyword>IBM Watson<Keyword> - условия знания 2-го курса медицинского вуза	2011-05-27 09:44:26
21	English	1589	Cligroup Hires <Keyword>IBM Watson<Keyword> to Work on Wall Street	2012-03-07 01:31:39
22	Russian	1428	СынOPSIS<Keyword>IBM Watson<Keyword> - поможет онкологам в диагностике и лечении рака	2012-03-26 04:30:00
23	Indonesian	22944	<Keyword>IBM Watson<Keyword>	2011-02-23 07:31:00
24	English	6535	Fresh off 'Jeopardy!' victory, <Keyword>IBM Watson<Keyword> lands health care job	2011-03-07 08:26:19
25	English	16094	Network with <Keyword>IBM Watson<Keyword> - the #FollowFriday	2011-02-18 08:00:46
26	English	6421	<Keyword>IBM Watson<Keyword> Goes to School to Engage Next Generation of Innovators	2011-03-31 14:37:39
27	English	4990	<Keyword>IBM Watson<Keyword> - Moves Beyond Jeopardy to the Doctor's Office	2011-10-25 01:37:43
28	English	11271	<Keyword>IBM Watson	
29	English	10341	<Keyword>IBM Watson	
30	Romanian	1253	Supercomputerul <Key	
31	English	5670	Elemental queries <Key	
32	English	5879	<Keyword>IBM Watson	
33	English	7283	<Keyword>IBM Watson	
34	Bulgarian	1290	СынOPSIS<Keyword>IBM Watson	
35	English	7858	<Keyword>IBM Watson	
36	English	11861	<Keyword>IBM Watson	

**Select a type of sheet:**

- Filter
- Macro
- Load**
- Pivot
- Combine
- Union
- Intersection
- Complement
- Limit
- Distinct
- Copy
- Formula

**IBM Watson coverage by language**

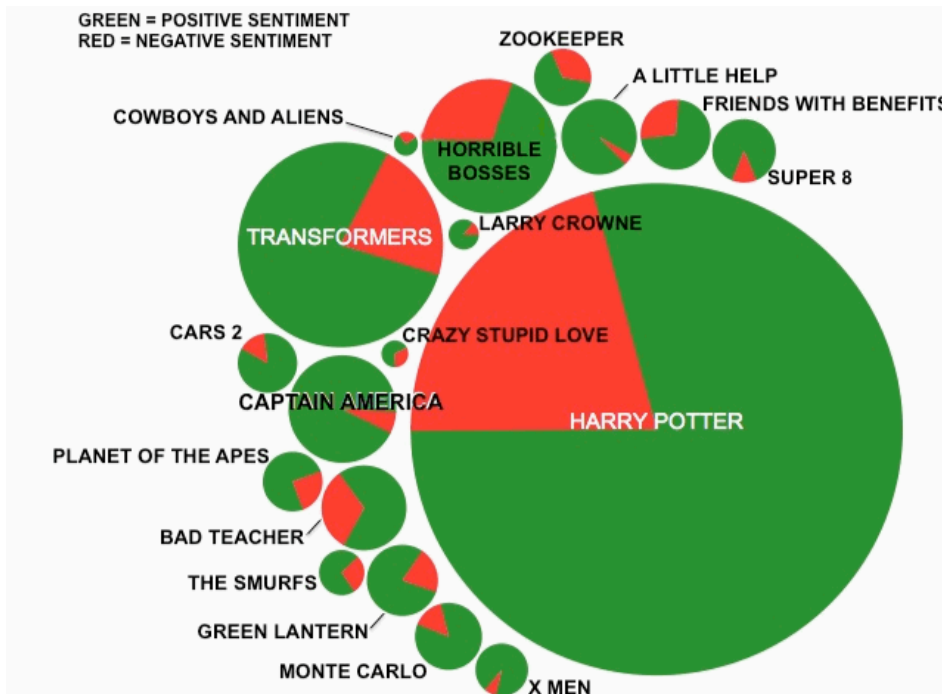
Language	Coverage
English	78.9%
Russian	4.2%
Other languages	16.9%

**Top 10 UK sites with IBM Watson coverage**

- itbriefing.net
- rss.feedsportal.com
- www.computerweekly.com
- www.computerworlduk.com
- www.supercomputingonline.com
- www.techeye.net
- www.telcaraware.com
- www.thespoof.com
- www.vadvert.co.uk
- www.zdnet.co.uk

# Big Data Made Easy for the Little Guy

- USC's Film Forecaster correctly predicted a clamor for "Hangover 2" that resulted in \$100 million opening over Memorial Day weekend
  - Looked at 250K-500K Tweets and broke down positive and negative messages using a lexicon of 1700 words



*The Film Forecaster sounds like a big undertaking for USC, but it really came down to one communications masters student who **learned Big Sheets in a day**, then pulled in the tweets and analyzed them - Ryan Kim*

# Quick start sample applications

- **Resuable software assets based on customer engagements**
  - Useful for starting point for various applications
  - Accessible through Web console
- **Available assets**
  - Data export (from relational DBMS, files, HBase)
  - Data import (from relational DBMS, files)
  - Web crawler
  - Boardreader.com support (Web forum search engine)
  - Ad hoc queries for Jaql, Hive, Pig
  - TeraGen-TeraSort, WordCount sample applications



# Running Applications from the Web Console

Welcome | Cluster Status | Files | **Applications** | Application Status

## Applications

BoardReader | Pig sample | Crawler | Database

Tera Gen-Sort | Hive sample | Word Count | Database

Simple Jaql Application | Distributed Copy | Hadoop Streaming Word Count Sample | Ad h...

**Name:** Ad hoc Hive query

**Description**  
You can use the ad hoc Hive query to create your own customiz...

**Execution**  
Execution Name:

**Parameters**  
Hive query:  

```
CREATE EXTERNAL TABLE student (
  NAME STRING,
  AGE INT,
  GPA FLOAT
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE
LOCATION '/tmp/mystudents';
```

**Advanced Settings**

**Application History**

Status	Execution Name	Progress
No filter applied		
✓	Default Execution	100%

# Build a Big Data Program – Map Reduce example

## Eclipse based development tools For JAQL, Hive, Java MapReduce, Text Analytics

The screenshot illustrates the Eclipse IDE environment for developing a Java MapReduce program. The 'New' wizard is open, showing the 'Java MapReduce Program' option selected. The Project Explorer shows the file structure for 'WordCount.java'. The editor displays the code for 'WordCount.java'. A context menu is open over the code, with 'Java MapReduce' selected. The Console window shows the output of the program execution.

```

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    //conf.set("mapred.job.tracker", "local");
    //conf.set("fs.default.name", "file://local/path");
    String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
    if (otherArgs.length != 2) {
        System.err.println("Usage: wordcount <in> <out>");
        System.exit(2);
    }
    Job job = new Job(conf, "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
}

```

```

<terminated> New_configuration [Java MapReduce] /home/hadoop/java/ibm-java-i386-60/bin/javaw (Apr 25, 2012 3:49:17 PM)
12/04/25 15:49:21 INFO input.FileInputFormat: Total input paths to process : 2
12/04/25 15:49:22 INFO mapred.JobClient: Running job: job_local_0001
12/04/25 15:49:22 INFO util.ProcessTree: setsid exited with exit code 0
12/04/25 15:49:22 INFO mapred.Task: Using ResourceCalculatorPlugin : org.apache.hadoop.util.LinuxResourceCalcu
12/04/25 15:49:22 INFO mapred.MapTask: io.sort.mb = 100
12/04/25 15:49:25 INFO mapred.JobClient: map 0% reduce 0%
12/04/25 15:49:25 INFO mapred.MapTask: data buffer = 79691776/99614720
12/04/25 15:49:25 INFO mapred.MapTask: record buffer = 262144/327680
12/04/25 15:49:27 INFO mapred.MapTask: Starting flush of map output
12/04/25 15:49:30 INFO mapred.MapTask: Finished spill 0
12/04/25 15:49:30 INFO mapred.Task: Task:attempt_local_0001_m_000000_0 is done. And is in the process of commit
12/04/25 15:49:31 INFO mapred.LocalJobRunner:
12/04/25 15:49:31 INFO mapred.LocalJobRunner:

```

# Develop Hive with the SQL Editor and view results

The screenshot displays an IDE interface with the following components:

- Data Source Explorer:** Shows a tree view of database connections, including 'Hive JDBC (Hive v. 0.0.0)' with sub-items for 'default', 'Authorization IDs', and 'Schemas'.
- SQL Editor:** Contains the following SQL code:
 

```

      CREATE TABLE ratings (
        userid INT,
        movieid INT,
        rating INT,
        unixtime STRING)
      ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
      STORED AS TEXTFILE;

      LOAD DATA LOCAL INPATH 'data/data.file'
      OVERWRITE INTO TABLE ratings;

      Select * from ratings;
      
```
- SQL Results:** Shows the execution status as 'Terminated' and a table of results. The table has columns: Status, Operation, Date, Connection Profile, userid, movieid, rating, and unixtime. The results are as follows:
 

Status	Operation	Date	Connection Profile	userid	movieid	rating	unixtime
Terminated	Select * fro	May 25, 20:	Hive JDBC	196	242	3	881250949
				186	302	3	891717742
				22	377	1	878887116
				244	51	2	880606923
				166	346	1	886397596
				298	474	4	884182806
				115	265	2	881171488
				253	465	5	891628467
				305	451	3	886324817

# Performance enhancements



- **Flexible job scheduler option**
  - Optimize response time for small jobs
  - Available in addition to FAIR, FIFO scheduling
  
- **Adaptive MapReduce**
  - Speeds up a class of jobs (e.g., jobs that process small files)
  - Accomplished by changing how certain MapReduce tasks executed
    - Mappers can decide at runtime to take on more work (until it doesn't make sense anymore). Communication via ZooKeeper.
  - Enabled through Jaql option, MapReduce job property setting
  
- **Efficient processing of compressed text data**
  - Use multiple Map tasks (vs Hadoop default of 1) for processing compressed text files
  - Enabled through BigInsights LZ0-based compression technology
  - Automatic with Jaql; programming option with Java MapReduce
  
- **Large scale text indexing and faceted (categorized) search**



# Comparison of general compression technologies (conducted by third party)

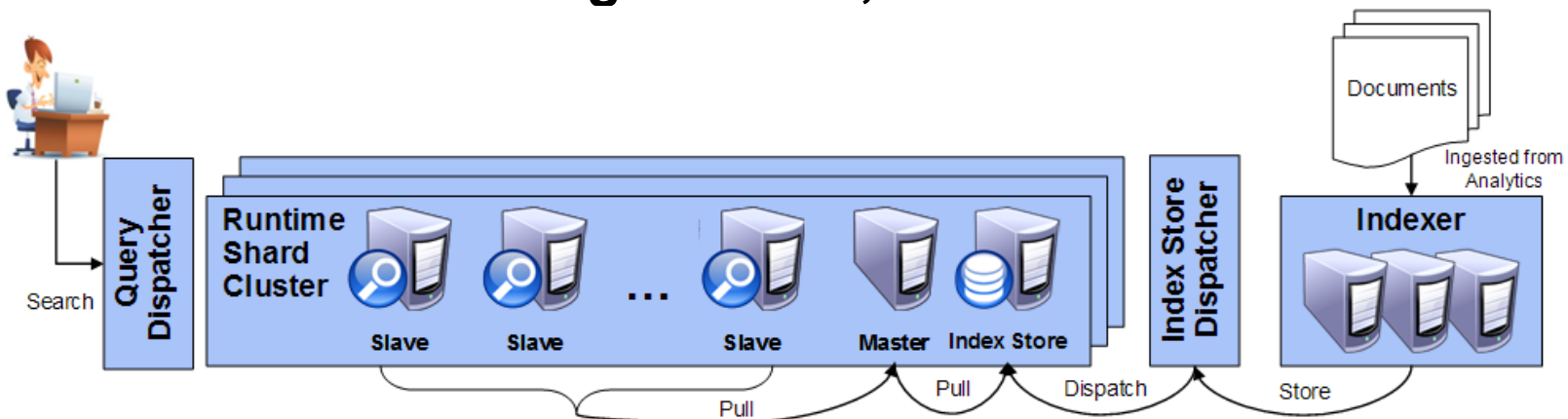
	Size (Mbytes)	Compression speed (sec)	Compression memory used (MBytes)	Decompression speed	Decompression memory used (Mbytes)
uncompressed	<b>96</b>				
gzip	<b>23</b>	<b>10</b>	<b>.7</b>	<b>1.3</b>	<b>.5</b>
bzip2	<b>19</b>	<b>22</b>	<b>8</b>	<b>5</b>	<b>4</b>
lzo	<b>36</b>	<b>1</b>	<b>1</b>	<b>.6</b>	<b>0</b>
lzm	<b>18</b>	<b>63</b>	<b>14</b>	<b>3</b>	<b>1.8</b>

Approximate values from

[http://stephane.lesimple.fr/wiki/blog/lzop\\_vs\\_compress\\_vs\\_gzip\\_vs\\_bzip2\\_vs\\_lzma\\_vs\\_lzma2-xz\\_benchmark\\_reloaded](http://stephane.lesimple.fr/wiki/blog/lzop_vs_compress_vs_gzip_vs_bzip2_vs_lzma_vs_lzma2-xz_benchmark_reloaded)

# Large Scale Indexing, Faceted Search

- Designed to improve text searches over big data
- Indexing characteristics
  - Based on Apache Lucene
  - Parallel index
    - Index operation is run in parallel, but the index is stored in one physical index.
  - Distributed index
    - Index is too large to be contained in one physical index
    - Index is distributed into shards, representing one logical index
    - Each query is evaluated against all shards
- Faceted search - categorization, drill down



# BigInsights Secure Architecture



Authentication store

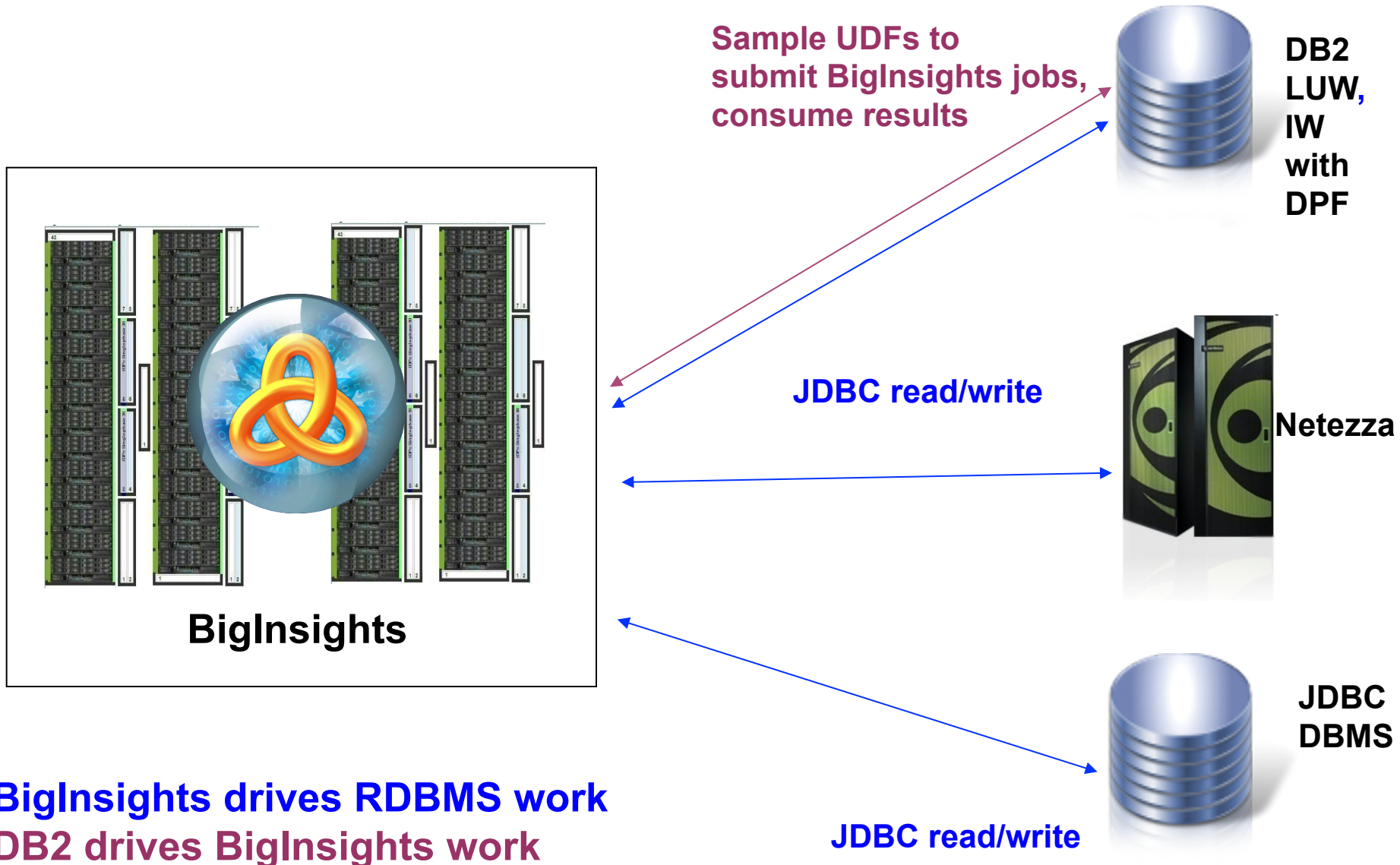
Web Console



# Security (cont' d)

- **User authentication approaches**
  - None
  - Flat file
  - LDAP
  - Pluggable Authentication Modules (PAM)
  
- **Authorization (role-based)**
  - System administrator
  - Data administrator
  - Application administrator
  - User
  
- **Credentials store**
  - Store potentially sensitive info: tokens, passwords, etc.
  - Maintained in BigInsights distributed file system
    - Public and private directories
    - Read/write via command line utility

# BigInsights Connectivity to DBMS / Warehouse



- BigInsights drives RDBMS work
- DB2 drives BigInsights work

# Ways to get started with BigInsights

## ■ In the Cloud

- Via RightScale, or directly on Amazon, Rackspace, IBM Smart Enterprise Cloud, or on private clouds.
- Pay only for the resources used.

## ■ In the Virtual Classroom

- Free Hadoop Fundamentals training course  
[www.bigdatauniversity.com](http://www.bigdatauniversity.com)

## ■ On Your Cluster

- Download Basic Edition from [ibm.com](http://ibm.com).

## ■ In the Classroom

- Enroll in the InfoSphere BigInsights Essentials course.

The screenshot shows the BigDataUniversity website homepage. The header includes the logo "BigDataUniversity" with the tagline "Learn from the industry's best" and navigation links for HOME, LEARN, DOWNLOAD, RESOURCES, JOBS, and LEARN Hadoop. There are also links for "login" and "sign up". The main content area features several promotional banners and text blocks:

- Hadoop Fundamentals:** A banner with a play button icon and the text "Start your career on Hadoop with this FREE course. Learn with hands-on exercises on a Hadoop cluster. Enroll now!"
- Why register?:** A section titled "Easy and Affordable" stating "Learning Hadoop and other Big Data technologies has never been more affordable! Many courses are FREE!"
- Latest industry trends:** A section titled "Learn from the Experts!" stating "Big Data University offers education about Hadoop and other technologies by the industry's best!"
- Learn at your Own Pace!** A section stating "Find everything right here when you need it and from wherever you are."
- Study Made Easy!** A section with a "sign up now" button and the text "for a chance to receive: FREE Books".
- Student Testimonials:** A section titled "Balázs (USA)" with a "go to sign up" button.

At the bottom of the page, there are links for "about us", "legal", "contact", and "bug reports".

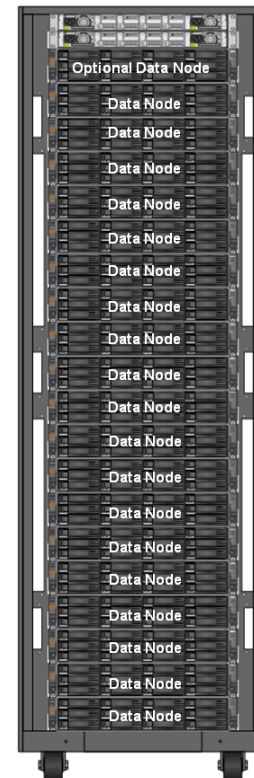
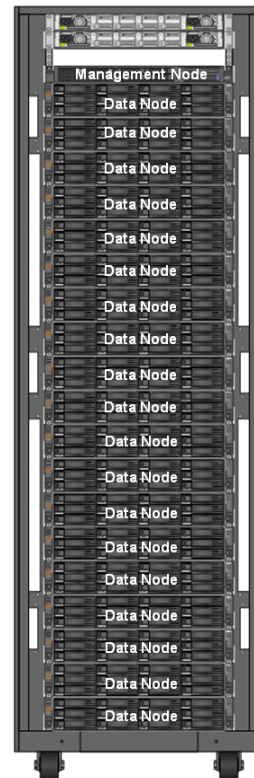
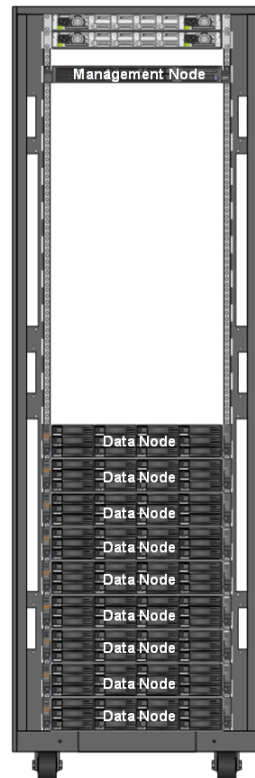
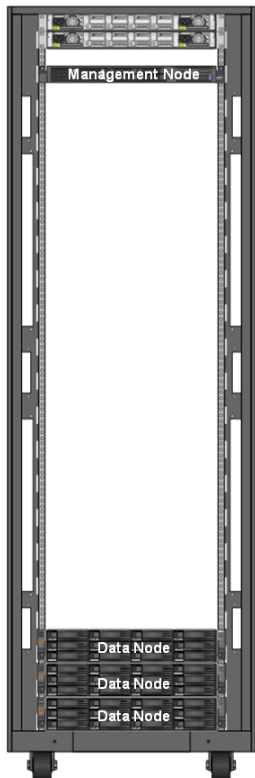
# Scalable IBM Systems for BigInsights

*Scalable with multiple entry points – start small and grow*

*Flexible CPU, memory, storage and networking choices to support varied workloads*



**Based on the IBM System x3630 M3:  
Ultra-dense,  
storage-rich server  
for Big Data**



Configuration	Starter	Half Rack	1 <sup>st</sup> Thru 4 <sup>th</sup> Full Rack*	Additional Full Racks
Usable Storage	Up to 108TB	Up to 324TB	Up to 684TB per rack	Up to 720TB per rack
User space	Up to 36TB	Up to 108TB	Up to 228TB per rack	Up to 240TB per rack

# Visit the BigInsights technical portal . . . .

- Free links to papers, demos, discussion forum, and more

- <http://www.ibm.com/developerworks/wiki/biginsights/>

BigInsights - Welcome - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://www.ibm.com/developerworks/mydeveloperworks/wikis/home? ... huawei symantec logo

Most Visited Getting Started Latest Headlines IBM

BigInsights - Welcome

## BigInsights Technical Enablement Wiki

IBM InfoSphere BigInsights  
Bring the Power of Hadoop to the Enterprise

Get up to speed on InfoSphere BigInsights, IBM's software platform designed to help firms store, manage, and analyze "big data".

### Technical materials

- Articles and white papers
- BigInsights InfoCenter

### Demos and videos

- Product demos
- IBM interviews and presentations

### Downloads

- BigInsights Basic Edition (free)
- Fix packs for BigInsights Enterprise Edition (licensed)

### Discussion Forum

- BigInsights forum on IBM developerWorks

### Education

- BigData University (free online courses)
- IBM Education (classroom, instructor-led online, and self-study courses)

### News and announcements

- Recent news reports and press releases
- Announcement letters

### Blogs and social media

Popular Tags: biginsights (7), big (5), data (5), hadoop (3), analysis (2), articles (2), ibm (2), infosphere (2), papers (2), basic (1), bigsheets (1), case (1), customer (1), demos (1)

View as list | cloud | all

Site Tools

developerWorks : Information Management : IBM InfoSphere BigInsights Forum - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ibm.com/developerworks/forums/forum.jspa?forumID=2409&c ...

Most Visited Getting Started Latest Headlines IBM

developerWorks : Infor...

Country/region [select]

IBM

Home Solutions Services Products Support & downloads My IBM

## IBM InfoSphere BigInsights Forum

Search for within IBM InfoSphere BigI... Go

This forum is a place to exchange ideas and share solutions with your peers in the IBM InfoSphere BigInsights. Active participation in the forum will allow all the participants to get the maximum return. All information in for community discussion and should not be considered IBM support.

By using this forum, you agree to abide by the [forum etiquette](#) and that you understand the [Terms of Use](#) governing this web site.

Post New Thread Watch Forum Watch List Forum Settings Help

Messages: 0 - Threads: 0 - Filter: All Threads

Threads	Author	Views	Replies	Last Post	Tags
Learn Hadoop for free and Go to Live Vegas All Expenses Paid	sanjoo@ibm.com	61	0	Sep 09, 2011 3:02:12 AM Last Post By: sanjoo@ibm.com	
"Value of BigInsights" white paper	sanjoo	422	1	Sep 07, 2011 03:15:56 PM Last Post By: sanjoo	
Online Manual Information Center - InfoSphere BigInsights	sanjoo	497	2	Aug 03, 2011 02:41:41 PM Last Post By: sanjoo	
Download UML - InfoSphere BigInsights Basic Edition	sanjoo	256	0	May 27, 2011 07:55:00 AM Last Post By: sanjoo	
Welcome to IBM InfoSphere BigInsights forum	Adeline.Bassi	365	0	Mar 28, 2011 08:08:52 AM Last Post By: Adeline.Bassi	

6.1.7 © 2011 IBM Corp. All rights reserved. See the application agreement for authorization.