# Leveraging Clouds
# for Small and Big Data

- 6 December 2011

**IM Cloud Computing Center of Competence**    IMcloud@ca.ibm.com
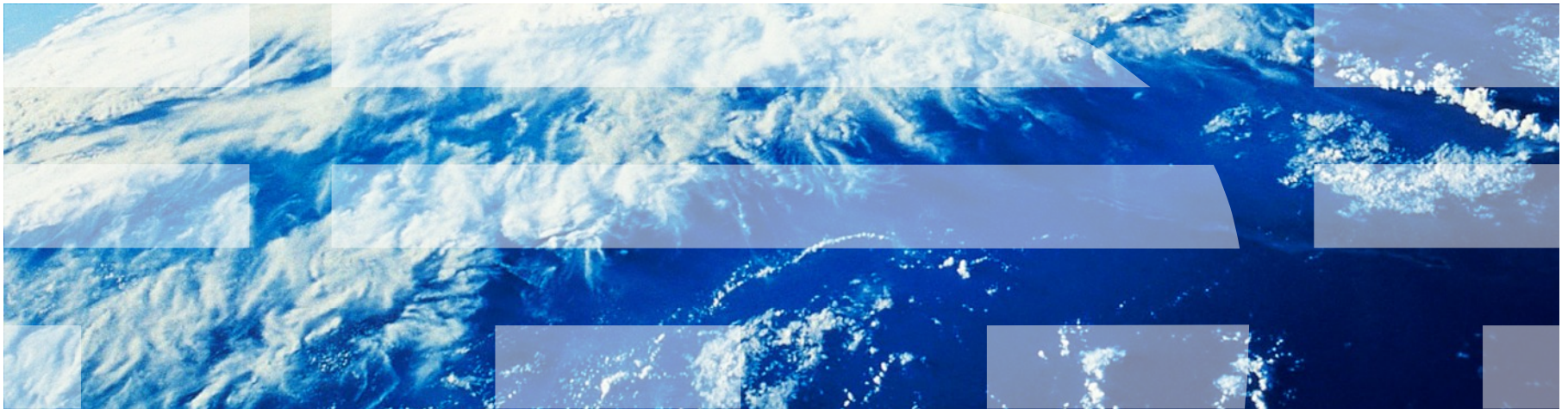
# > Featured Speakers

### Leon Katsnelson

Program Director,
IM Cloud Computing and Emerging Technologies

IBM

### **Uri Budnik**

**Director,
ISV Partner Program**

**RightScale**

# Agenda

- DB2 and Cloud

- The Big Data Challenge

- IBM's Approach to Big Data

- Big Data on Cloud

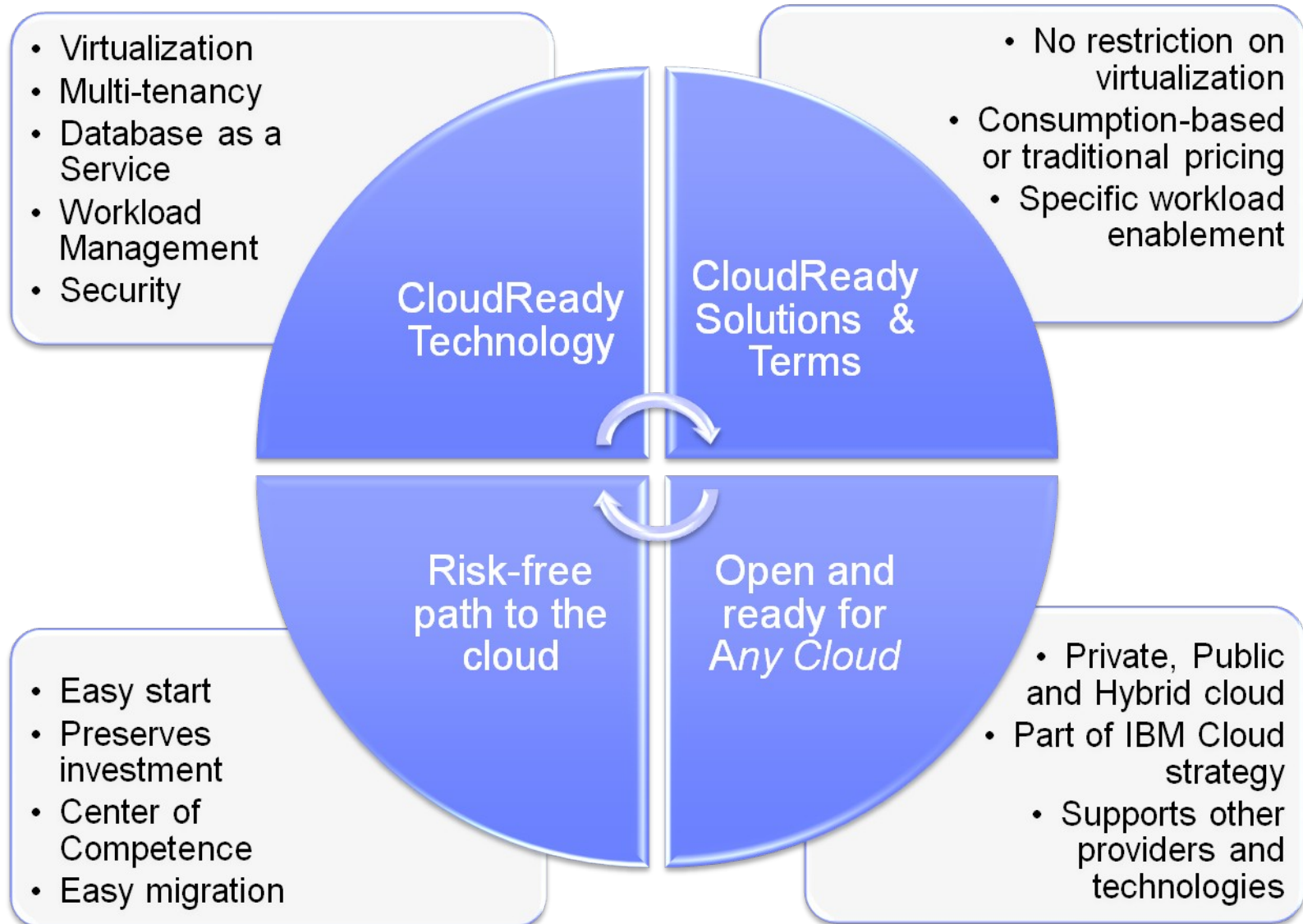- RightScale Cloud Management Platform

# DB2 ON CLOUD

# … many reasons clients consider Cloud Computing

We need to reduce our IT costs

We are constrained on space, energy and cooling in our data center

Need to support remote teams

We are always constrained on resources for proper Q&A of our solutions

We need to improve business agility

We need to reduce our CAPEX

We need to integrate web born data

We need to deliver better resiliency

**DB2 is CloudReady**

# DB2 Strategy for Cloud Computing

- Virtualization
- Multi-tenancy
- Database as a Service
- Workload Management
- Security

**CloudReady Technology**

**CloudReady Solutions & Terms**

- No restriction on virtualization
- Consumption-based or traditional pricing
- Specific workload enablement

**Risk-free path to the cloud**

**Open and ready for *Any Cloud***

- Easy start
- Preserves investment
- Center of Competence
- Easy migration

- Private, Public and Hybrid cloud
- Part of IBM Cloud strategy
- Supports other providers and technologies

# DB2: Ready for Any Cloud

IBMSmart**Cloud** Enterprise

**amazon** web services™

G**O**GRID

**NEW**

**IBM Workload Deployer**

**RiGHT SCaLe**®

**cloud**.com
now a part of Citrix Systems

**rackspace**® **HOSTING**

**EUCALYPTUS**

# Free DB2-based Website for a year

- All-in-one instance (server):
  - DB2 Express-C
  - Web application server (Nginx, Ruby on Rails)
  - Content Management System (RadiantCMS)

- Runs on Amazon EC2 Micro instance:
  - 613MB of memory
  - Up to 2 EC2 Compute Units (for short periodic bursts)
  - EBS storage
  - 32-bit or 64-bit platform

- Free for a year for new AWS customers

- Tutorial http://programmingzen.com/2011/04/04/running-radiant-cms-on-db2-in-the-cloud/

# Dogfooding

@your place, @your pace DB2 skills acquisition

Community operated property

10,500+ Registered Students

Runs on the cloud, Managed by RightSale

Leverages DB2 technologies e.g. HADR

Continuous availability since launch in January 2011

**IBM**

# DB2 Galileo Early Experience Program

Clients get hands-on experience with the upcoming version of DB2 in 30 minutes

• No need to purchase, rack and cable servers

• No operating system patching

• No need to install DB2

• Always use the latest drop i.e. no wasted time dealing with resolved issues

How:

• Apply for DB2 Galileo Early Experience at http://freedb2.com/db2-early-experience-program-on-the-cloud/

• Email IMcloud@ca.ibm.com to request $500 credit

• Enjoy your own private virtual DB2 server in one of the data centers WW

From zero to DB2 Galileo in about 30 minutes
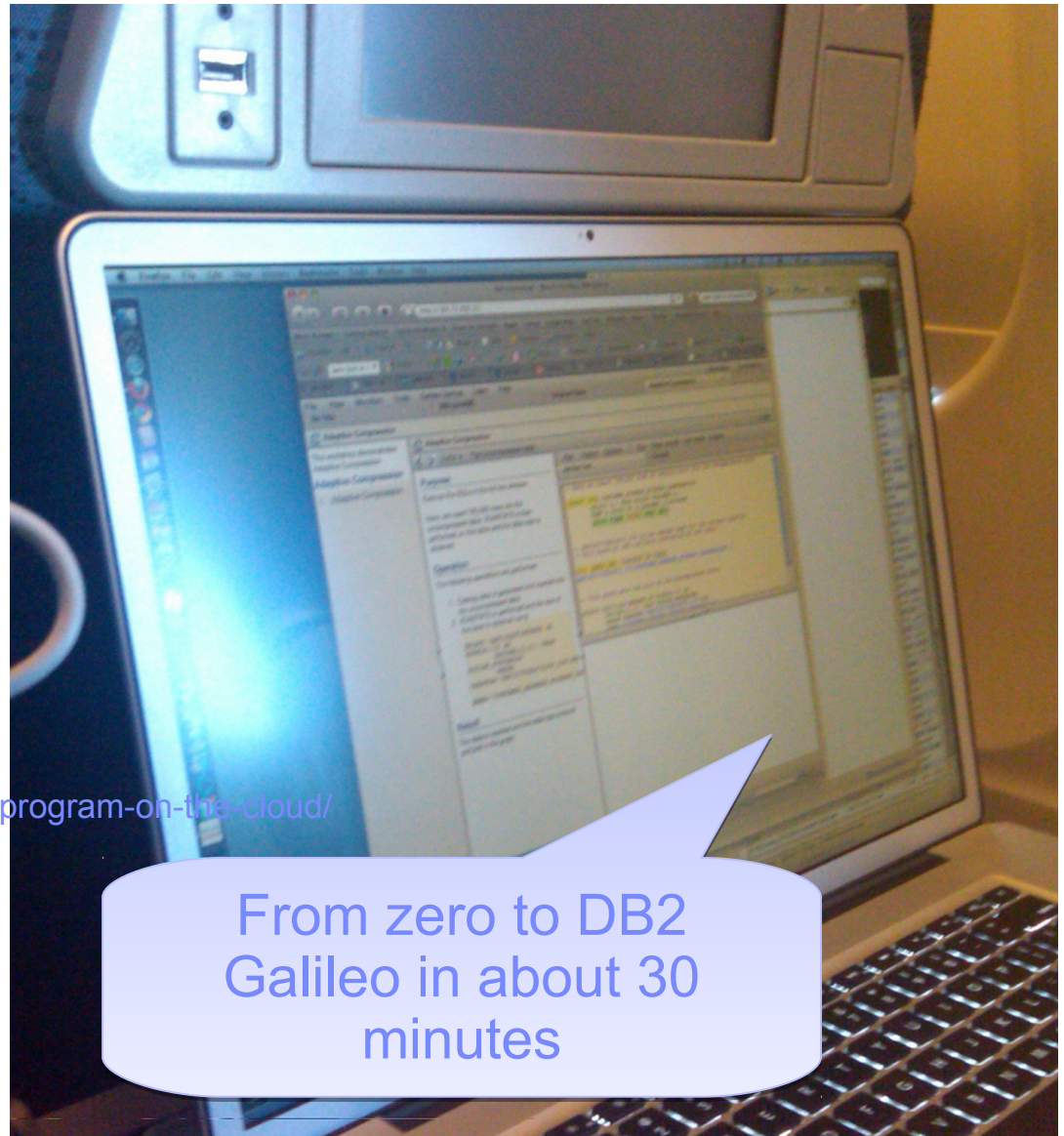
# DB2 on Cloud Resources

- IM Cloud Computing Center of Competence
  - IMcloud@ca.ibm.com

- DB2 on Cloud homepage
  - www.ibm.com/db2/cloud

# THE "BIG DATA" CHALLENGE

# What we hear from customers . . . .

- Lots of potentially valuable data is dormant or discarded due to size/performance considerations

- Large volume of unstructured or semi-structured data is not worth integrating fully (e.g. Tweets, logs, . . .)

- Not clear what should be analyzed (exploratory, iterative)

- Information distributed across multiple systems and/or Internet

- Some information has a short useful lifespan

- Volumes can be extremely high

- Analysis needed in the context of existing information (not stand alone)

# Big Data Presents Big Opportunities

*Extract insight from a high volume, variety and velocity of data in a timely and cost-effective manner*



**Variety:** Manage and benefit from diverse data types and data structures

**Velocity:** Analyze streaming data and large volumes of persistent data

**Volume:** Scale from terabytes to zettabytes

# Big Data Scenarios Span Many Industries



Multi-channel customer sentiment and experience a analysis

Detect life-threatening conditions at hospitals in time to intervene

Predict weather patterns to plan optimal wind turbine usage, and optimize capital expenditure on asset placement

Make risk decisions based on real-time transactional data

Identify criminals and threats from disparate video, audio, and data feeds

# Customer Engagements

## Use patterns

- Customer sentiment analysis (cross-sell, up-sell, campaign management)

- Integrated retail and web customer behavior modeling

- Predictive modeling (credit card fraud)

- System log analytics (reduce operational risk)

## Common requirements

- Extract business insight from large volumes of raw data (often outside operational systems)

- Integrate with other existing software

- Ready for enterprise use

Text, Blog, Weblog

Click streams

Log & transactions

Biological Sequences

Operational system & streams data sources

Consumer Insight

Multi-channel sales

Next Gen Fraud Models

New Business Development

ext Analytics

Statistical Model Building
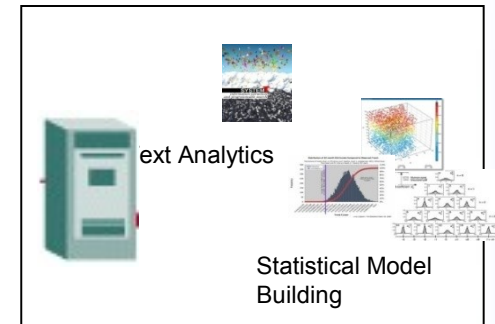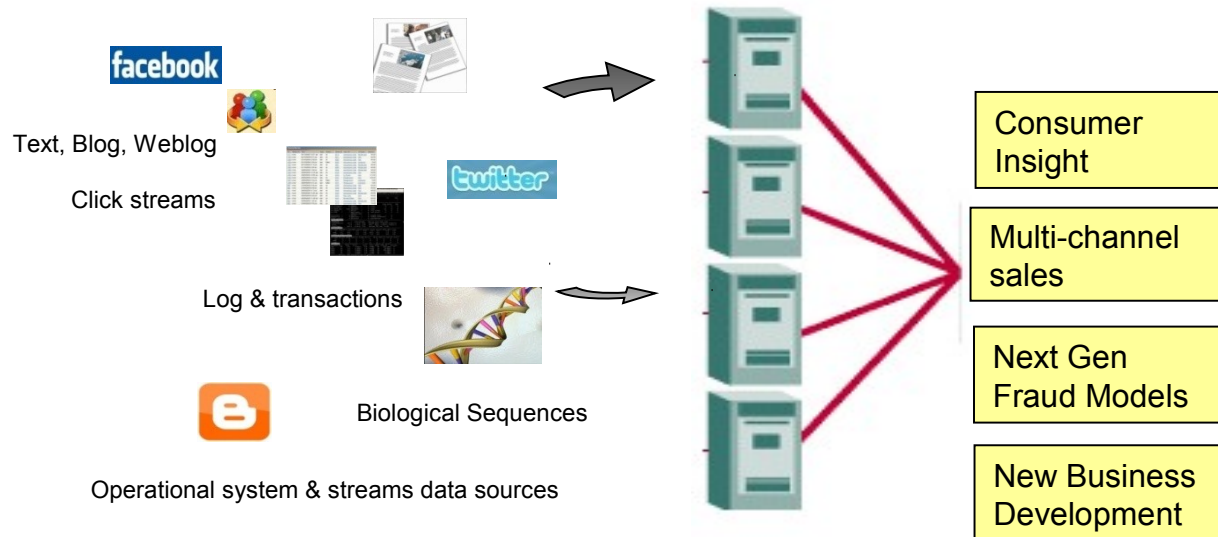
Big Data

# IBM'S APPROACH

# Big Data: an integral part of an enterprise data platform

- Manage Big Data from the instant it enters the enterprise
- High fidelity – no changes to original format
- Available for new uses, analyses, and integrations.

Cognos

Operational Data Store

Warehouse Applications

Big Data Applications

Warehouse

http://ww

**Big Data Platform**

**IBM Big Data Solutions**　　**Client and Partner Solutions**

Big Data User Environment

**Developers**　　**End Users**　　**Admin.**

Traditional data sources (ERP, CRM, databases, etc.)

Big Data Enterprise Engine

Streaming analytics　　Internet-scale analytics

Source data (Web, sensors, logs, media, etc. )

# IBM's Platform Addresses Key Requirements

1. Platform for $V^3$ – Variety, Velocity, Volume
   - Variety - manage data & content "As Is"
   - Handle any velocity - low-latency streams and large volume batch
   - Volume - huge volumes of at-rest or streaming data

2. Analytics for $V^3$
   - Analyze Sources in their native format - text, data, rich content
   - Analyze all of the data - not just a subset
   - Dynamic analytics - automatic adjustments and actions

3. Ease of Use for Developers and Users
   - Developer UIs, common languages & automatic optimization
   - End-user UIs & visualization

4. Enterprise Class
   - Failure tolerance, Security and Privacy
   - Scale Economically

5. Extensive Integration Capabilities
   - Integrate wide variety of sources
   - Leverage enterprise integration technologies

## Big Data Platform

# Platform Vision

**IBM**

**IBM Big Data Solutions**

**Client and Partner Solutions**

## Big Data Accelerators

**Text**   **Statistics**   **Financial**   **Geospatial**   **Acoustic**

**Image/Video**   **Mining**   **Times Series**   **Mathematical**

| Connectors | Applications | Blue Prints |

## Big Data Enterprise Engines

**InfoSphere Streams**

**InfoSphere BigInsights**

## Productivity Tools & Optimization

| Workload Management & Optimization | Provisioning | Workflow | Job Scheduling | Job Tracking | Data Ingestion |
|---|---|---|---|---|---|
| Management | Admin Tools | Configuration Manager | Activity Monitor | Identity & Access Mgmt | Data Protection |

**INTEGRATION**

Information Server

| **Rules / BPM** |
| iLog & Lombardi |

| **Data Warehouse** |
| InfoSphere Warehouse |

| **Warehouse Appliances** |
| IBM & non-IBM |

| **Master Data Mgmt** |
| InfoSphere MDM |

| **Database** |
| DB2 & non-IBM |

| **Content Analytics** |
| ECM |

| **Business Analytics** |
| Cognos & SPSS |

| **Marketing** |
| Unica |

| **Data Growth Management** |
| InfoSphere Optim |

http://ww

twitter

# BigInsights Summary

- **BigInsights = analytical platform for persistent "Big Data"**
  - Based on open source & IBM technologies
  - Managed like a start-up . . . . Emphasis on deep customer engagements, product plan flexibility

- **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

# InfoSphere BigInsights

Platform for volume, variety, velocity -- $V^3$

- Enhanced Hadoop foundation

Analytics for $V^3$

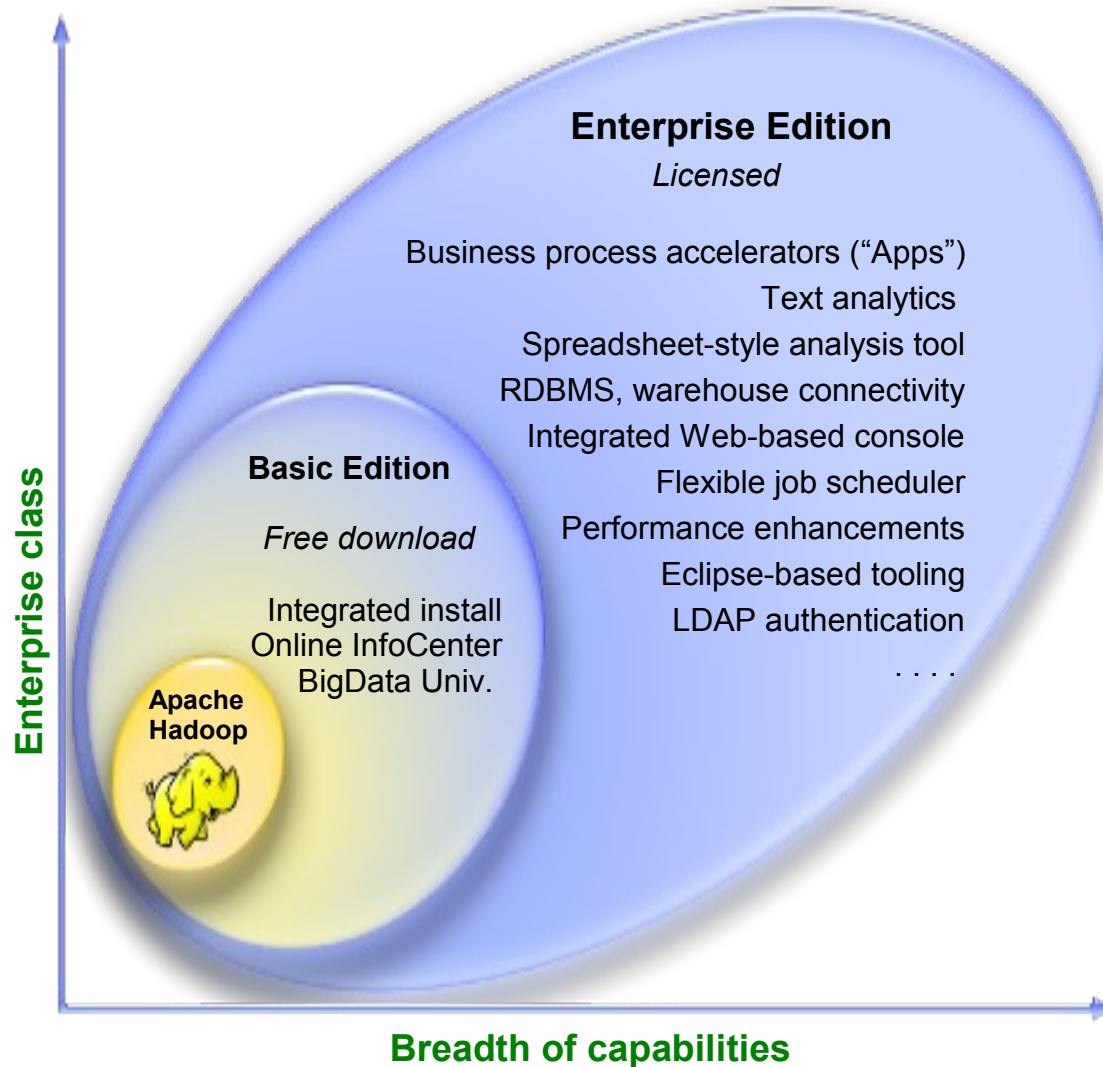- Text analytics & tooling

Usability

- Web console
- Integrated install
- Spreadsheet-style tool
- Ready-made "apps"

Enterprise Class

- Storage, security, cluster management

Integration

- Connectivity to DB2, Netezza, JDBC databases

**Enterprise Edition**

*Licensed*

Business process accelerators ("Apps")

Text analytics

Spreadsheet-style analysis tool

RDBMS, warehouse connectivity

Integrated Web-based console

Flexible job scheduler

Performance enhancements

Eclipse-based tooling

LDAP authentication

. . . .

**Basic Edition**

*Free download*

Integrated install
Online InfoCenter
BigData Univ.

**Apache Hadoop**

**Enterprise class**

**Breadth of capabilities**

Big Data
# ON CLOUD

# … Hadoop is great at getting the value out of large volumes of data

"94% of Hadoop users perform analytics on large volumes of data not possible before; 88% analyze data in greater detail; while 82% can now retain more of their data"

Ventana Research Benchmark Study "Hadoop and Information Management" 2011

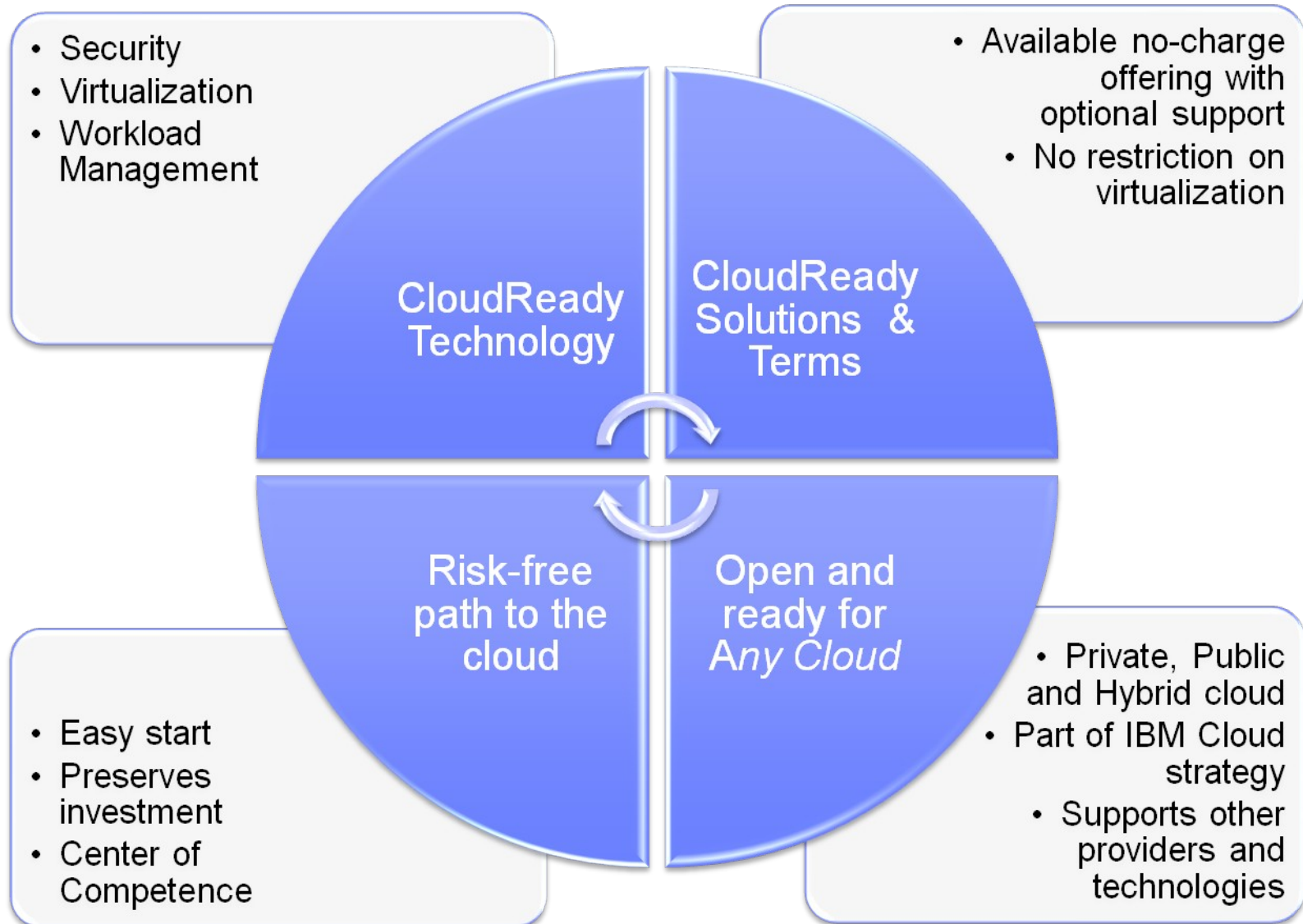# … but it requires large scale compute clusters and lots of storage

- Yahoo! Has over 40000 nodes running Hadoop managing 180-200 petabytes of data.

- Facebook has 2000 compute nodes with 20 Petabytes of data.

- Average cluster size is 200 servers but most clusters around 30 servers

- Most companies can't make upfront capital investment in compute, storage and networking resources:
  - Makes it hard to evaluate and pilot
  - Difficult to develop real hands on skill
  - Impractical for non continuous needs

# VALUE: The Fourth **V** of the Big Data Story

What insight could you gain if you had full use of a 100-node Hadoop cluster for an hour?

What if one hour of this 100-node cluster would cost $34?

# BigInsights Strategy for Cloud Computing

- Security
- Virtualization
- Workload Management

- Available no-charge offering with optional support
- No restriction on virtualization

**CloudReady Technology**

**CloudReady Solutions & Terms**

**Risk-free path to the cloud**

**Open and ready for** *Any Cloud*

- Easy start
- Preserves investment
- Center of Competence

- Private, Public and Hybrid cloud
- Part of IBM Cloud strategy
- Supports other providers and technologies

# IBM BigInsights: Ready for Any Cloud
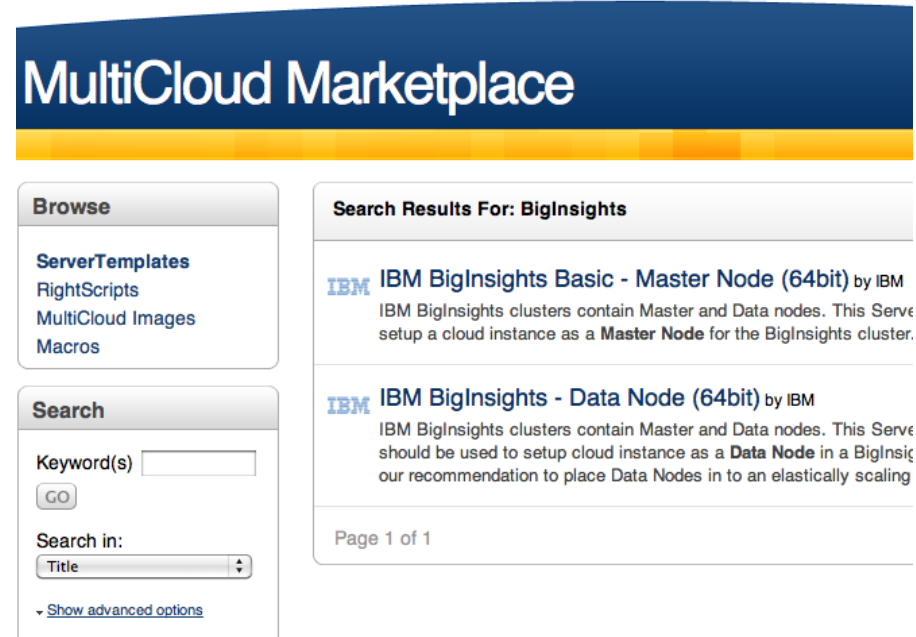
## IBM BigInsights on Cloud
## Hadoop for everyone … no upfront investment

- Your own Hadoop cluster on the cloud in less than 30 minutes

- No need to buy install, patch, maintain hardware

- Deploy on Amazon, IBM, Rackspace or your private cloud infrastructure

- Pay as you go for your infrastructure starting at $0.34/node/hour and only pay for what you actually use

- Use BigInsights Basic Edition at no charge, available low cost support. Seamlessly transition to BigInsights Enterprise Edition when ready

# IBM BigInsights and RightScale.com
# Hadoop on Amazon, Rackspace, Private & Hybrid Cloud Infrastructure

- Your own Hadoop cluster on the cloud in less than 30 minutes

- Deploy to Amazon, Rackspace cloud DCs or on your private cloud

- Pay as you go for your infrastructure starting at $0.34/node/hour

- Sophisticated cloud resource management by RightScale

- Use BigInsights Basic Edition … seamlessly transition to BigInsights Enterprise Edition when ready

- Take a free course on BigDataUniversity.com

# Available Now!
## Deploy BigInsights on IBM SmartCloud Enterprise

- Your own Hadoop cluster on the IBM cloud in less than 30 minutes

- No need for hardware, install, patch, maintain

- Locate your Hadoop cluster in one of IBM Cloud data centers WW

- Low hourly charges starting at $0.30/cluster/hour.

- Explore using BigInsights Basic seamlessly transition to BigInsights Enterprise when ready

- Take a free course on BigDataUniversity.com

# Evaluating Big Data Technology
# IBM IM Demo Cloud (in limited beta)

- IBM representative initiates the project, invites customer to participate.

- Types of projects: **Demos**, **Proof of Technology**, **Proof of Concept**:

- Also great for Beta, Product Introductions, Technology Previews

- Skip construction and go straight to evaluation. No need to get hardware, install/patch OS, BigInsights software, write MapReduce jobs, data sets

- BigInsights sandbox:
  - Dedicated cluster of virtual systems. Not a shared sandbox
  - Customer operates the systems as if local
  - Ability to bring large data sets
  - Data centers in NA, EU and AP

# BigInsights on the Cloud
# Making Learning Hadoop Easy and Fun

- Flexible on-line delivery allows learning @your place and @your pace

- Free courses, free study materials.

- **Cloud-based sandbox for exercises – zero setup**

- 10500+ registered students.

- Hadoop Programming Challenge - 3 students sent to IOD 2011 in Las Vegas, all expenses paid!

# … in Summary

- Big Data analytics using technologies such as Hadoop can deliver great value but its appeal is somewhat diminished by the extensive capital requirements

- Cloud is a great way to sidestep the CAPEX challenges

- IBM InfoSphere BigInsights, enterprise-ready distribution of Hadoop is available for cloud deployment on Amazon, IBM SmartCloud Enterprise, Rackspace and on private clouds

- IBM helps clients leverage the power of Hadoop and ease of deployment of the cloud. Contact IM Cloud Computing Center of Competence IMCloud@ca.ibm.com.

- BigDataUniversity.com is a great place to start skills acquisition when budget, time or travel is an issue.

# Resources

- **Use IBM BigInsights on IBM SmartCloud Enterprise**
  - http://www.db2university.com/courses/course/view.php?id=310

- **Use IBM BigInsights on Amazon, Rackspace or your private cloud:**
  - http://www.db2university.com/courses/course/view.php?id=309

- **Learn Hadoop and other Big Data technologies**
  - http://BigDataUniversity.com

- **Get free help from IM Cloud Computing Center of Competence:**
  - IMcloud@ca.ibm.com

- **Download Free IBM BigInsights Basic Edition**
  - http://www-01.ibm.com/software/data/infosphere/biginsights/basic.html

- **Visit home page for IBM BigInsights on ibm.com:**
  - http://www-01.ibm.com/software/data/infosphere/biginsights/

Cloud Management Platform
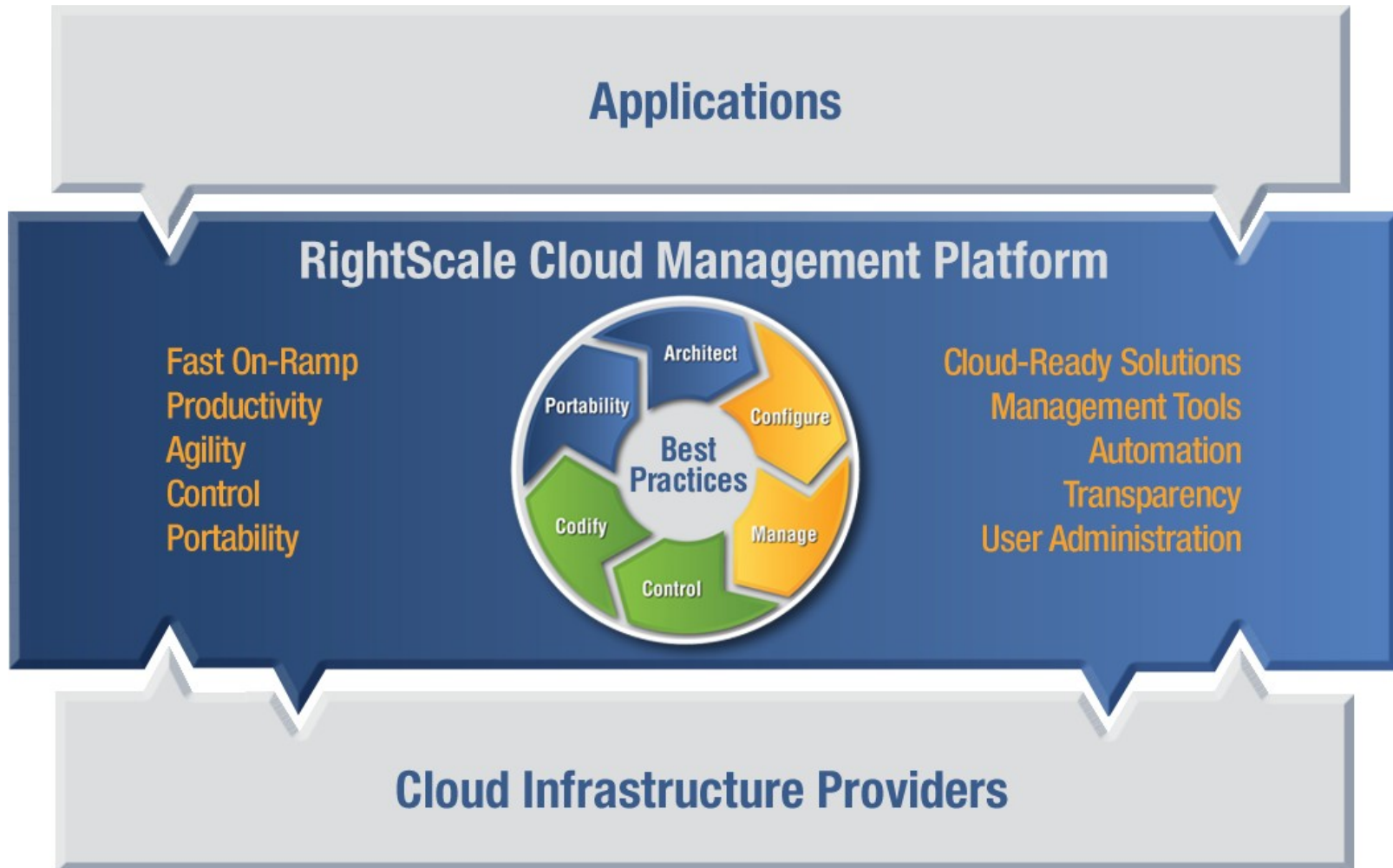
# RIGHTSCALE

## RightScale
## Real Customers, Real Deployments, Real Benefits

- Managed Cloud Deployments for 4 Years — globally

- More than 45,000 users; launched more than 3MM servers!

- Powering the largest production deployments on the cloud

# Complete Systems Management

# What do we Mean by Cloud Computing?
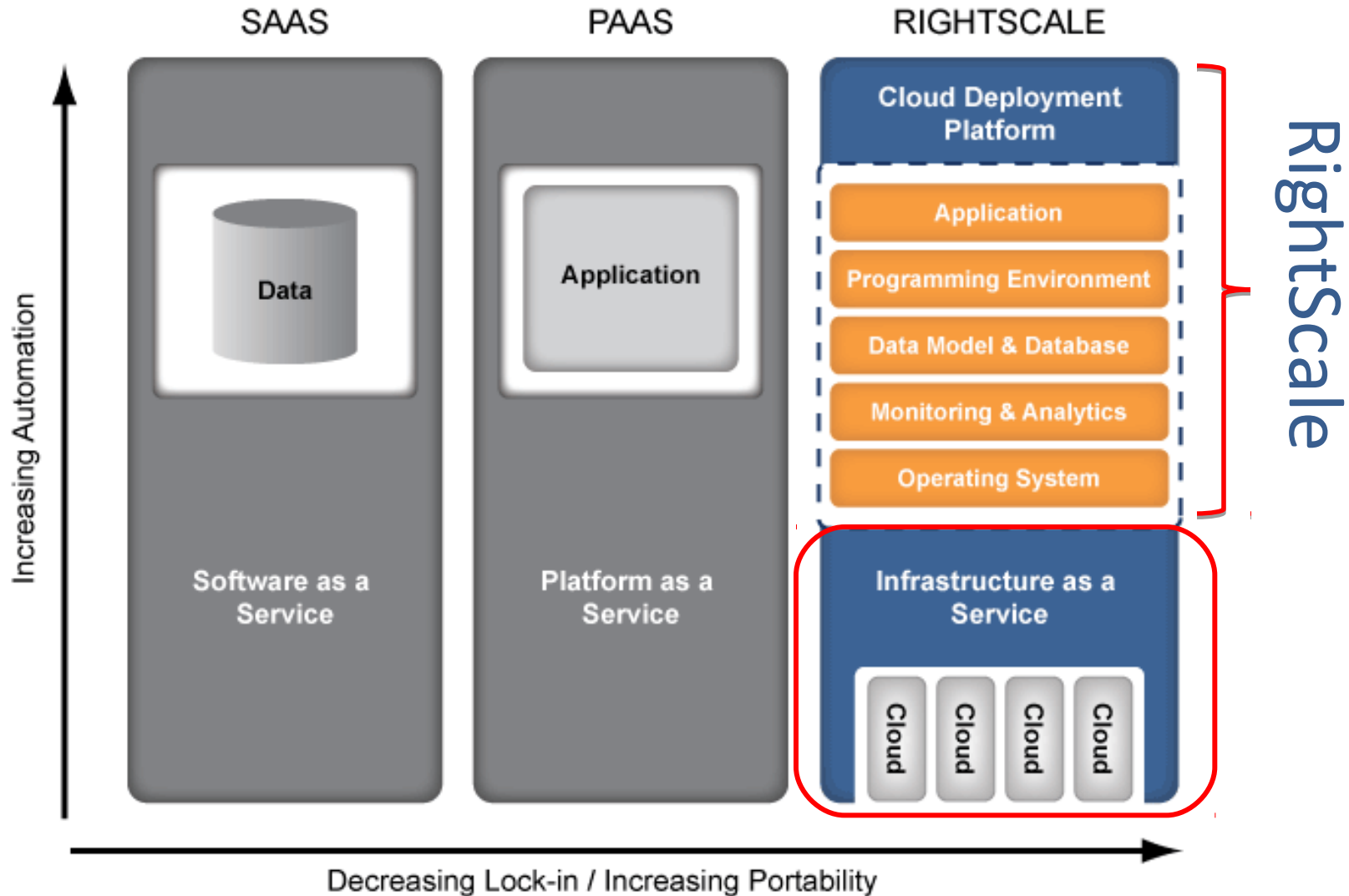
# RightScale Manages IaaS Clouds

# Take advantage of many resource pools



**Private Clouds**

**Public & Managed Clouds**

| | |
|---|---|
| Amazon Web Services | Rackspace |
| Datapipe | SoftLayer |
| Yahoo! Japan / IDCF | Tata |
| Korea Telecom | UnGeo |
| Logicworks | |

# ServerTemplates: Reproducible servers on demand

- Dynamic configuration
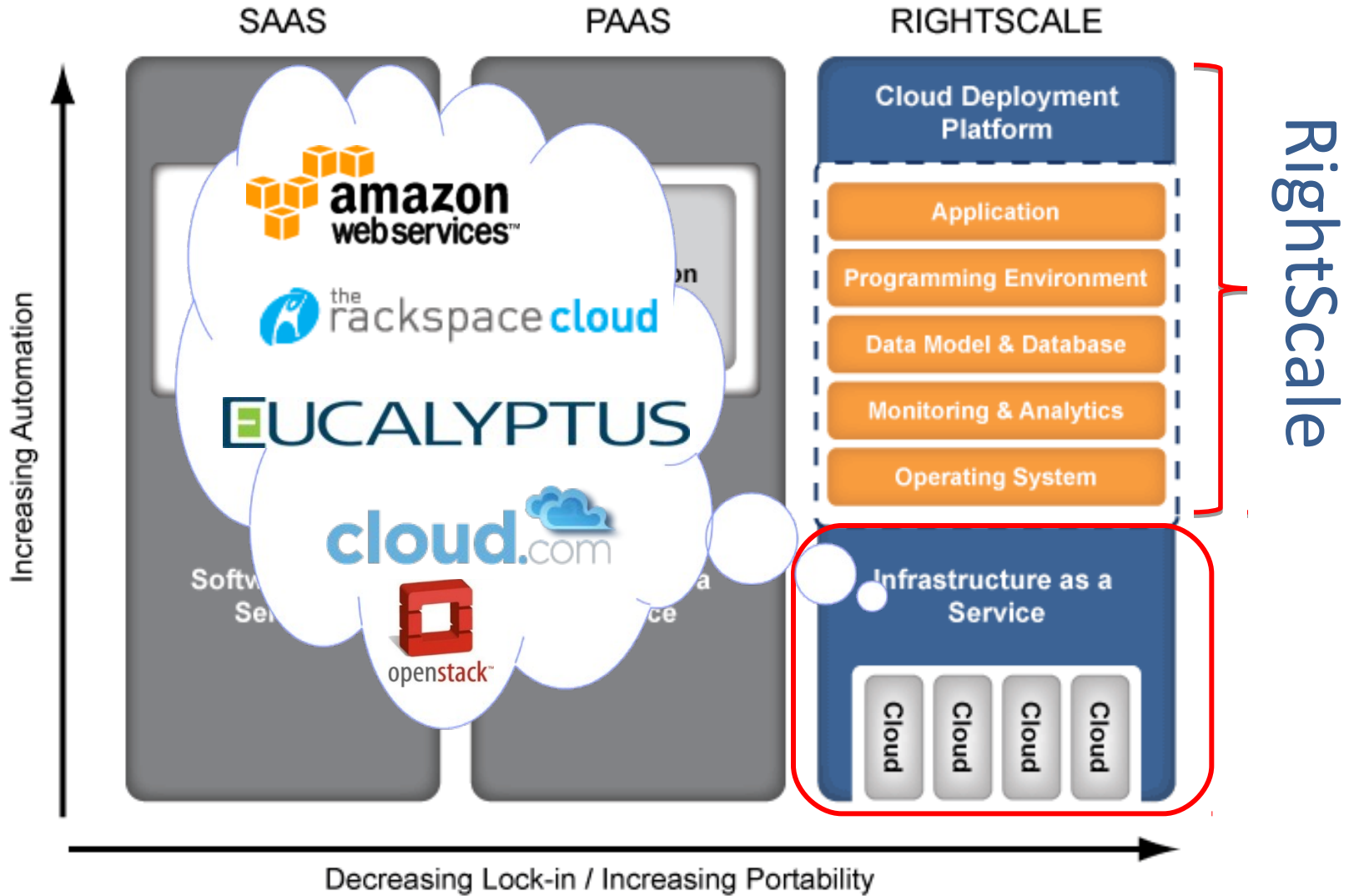
- Abstract role and behavior from cloud infrastructure

- Predictable deployment

- Cloud agnostic / portable

- *Object-oriented programming for sysadmins*

**Server Template**

**Scripts**

**Boot-Run at Startup**

Install App Server Software

Install Database Software

**Operational - Run on Demand**

Backup Database

**Decommisson - Run prior to Terminate**

Shutdown Database

**Base Machine Image**

**Load on Startup**

Cloud Machine Image with RightLink
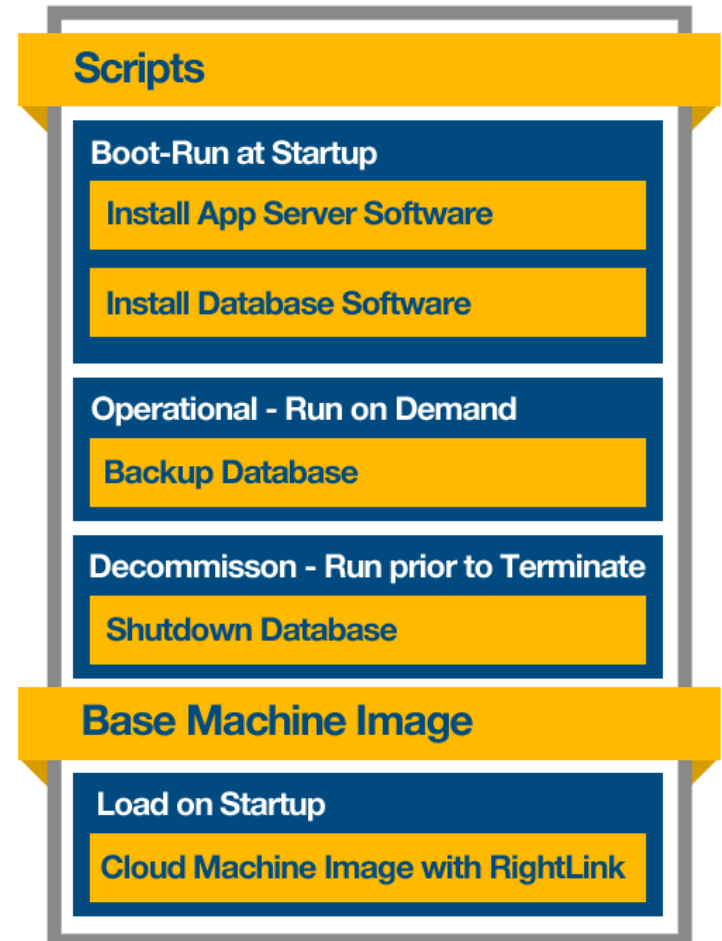
# ServerTemplates: Reproducible servers on demand

## Configuring servers through bundling images:

DB2 Express-C 9.7.4 (CentOS 5.2)

DB2 Express-C 9.7.4 (CentOS 5.4)

BigInsights (CentOS 5.4)

BigInsights (Ubuntu 8.10)

DB2 Express-C 9.7.4 (Ubuntu 8.10)

Frontend Apache 1.3 (Ubuntu 8.10)

Frontend Apache 2.0 (Ubuntu 9.10) -patched

CMS v1.0 (CentOS 5.4)

CMS v1.1 (CentOS 5.4)

My ASP appserver  (windows 2008)

My ASP.net  (windows 2008) – security update 1

My ASP.net  (windows 2008) – security update 8

SharePoint v4 (windows 2003) – 32bit

SharePoint v4 (windows 2003) –64bit

SharePoint v4.5 (windows 2003) –64bit

…

## Configuring servers with ServerTemplates:

boot sequence

Setup DNS and IPs

Restore last backup

Configure DB2

Install DB2

Install monitoring

MultiCloudImage

| CentOS 5.2 | Ubuntu 8.10 | Win 2003 |
| CentOS 5.4 | Ubuntu 9.10 | Win 2007 |
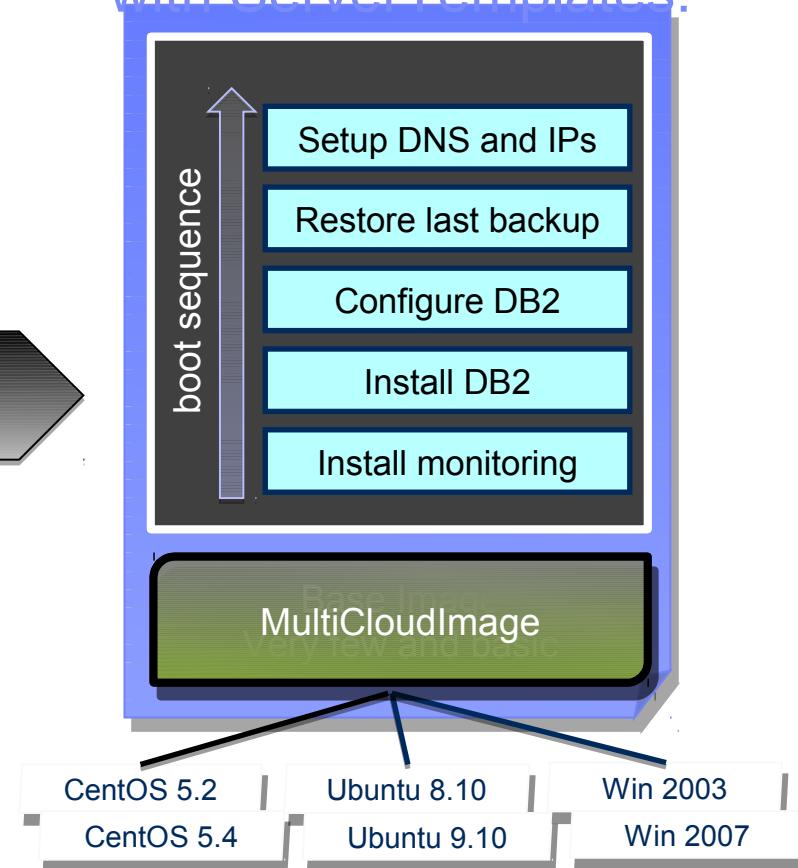
# ServerTemplates – Reproducible servers on demand
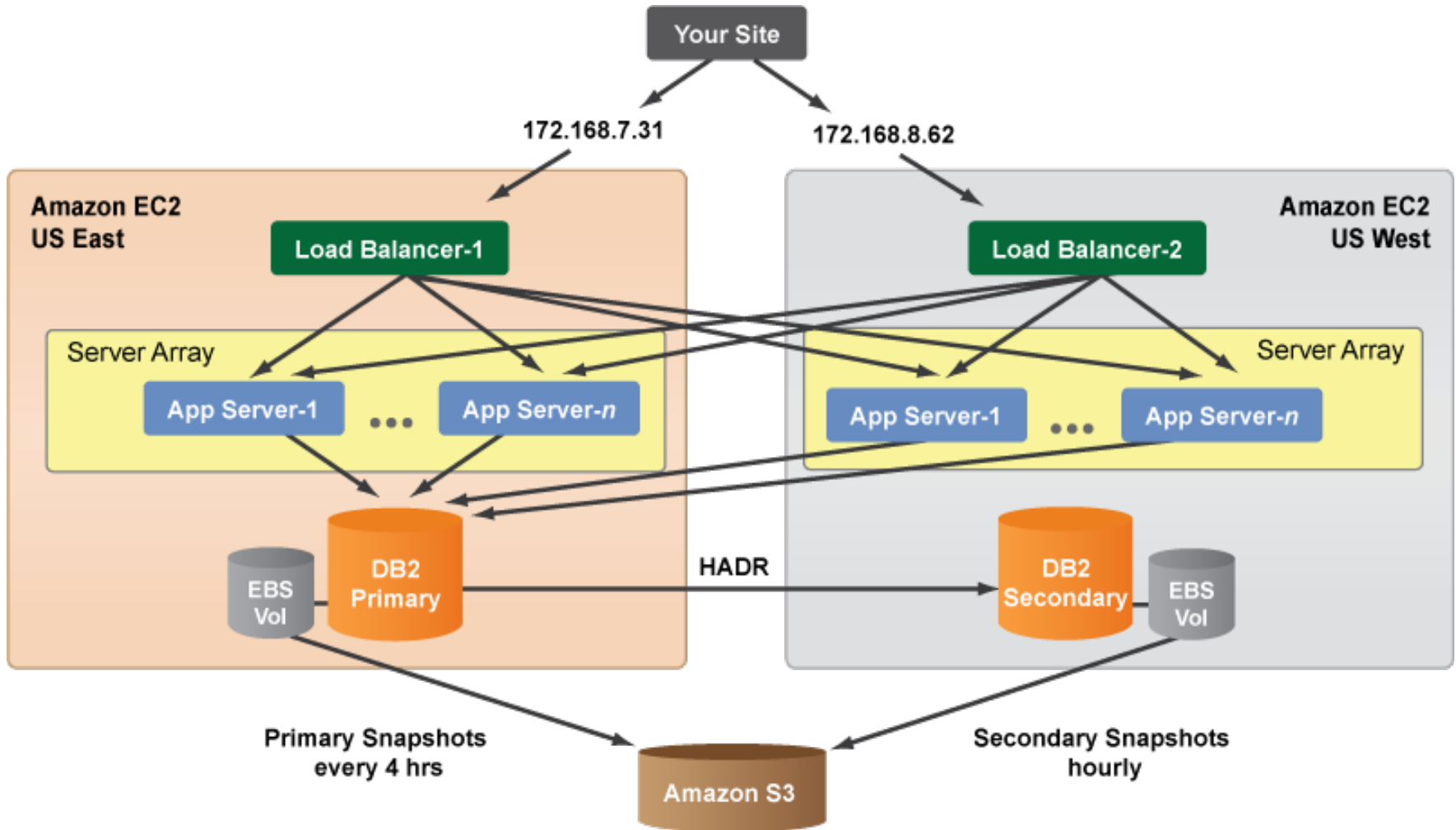
**ServerTemplates are like Playlists**

VS.

**VMs and Standard Cloud Images are like burned CDs**

Integrated approach that puts together all the parts needed to architect single & multi-server deployments

# Not just single servers — complete environments

# RightScale Runs in Your Web Browser

# Organize Servers in Deployments

# Organize Servers in Deployments

Reports ▾    Settings ▾                    Search 🔍    Production ▾    Support ▾    Logout

Help ⊗

Monitoring    Alerts

...e | October 7th ServerTemplate Release | Customize the widgets below!

**Credential Status**

🟢 Online

🟢 cloud controller
🟢 cloud account

erver Statistics

**Deployment Budget Estimate**

| Deployment | MTD* | Servers |
|---|---|---|
| App Test | – | 🟥🟥🟥...10x |
| Default | – | 🟢🟢 🟥🟥🟥...56x |
| DotNet IIS | – | 🟢🟢🟢...6x + [ 🟢 in arrays] |
| DotNet IIS–PERF TEST | – | 🟥🟥🟥🟥 |
| Elasticity Web Demo | – | 🟢🟢🟢🟢🟢 |
| IBM BigInsights | – | 🟢 🟡🟡🟡🟡🟡 |
| Load Test Group | – | 🟢 🟥🟥🟥 |
| Nokia Siemens Networks | – | |

# Manage Systems, not Servers

IBM

**RIGHT SCALE®**   Manage ▾   Design ▾   Clouds ▾   Reports ▾   Settings ▾   Se

**MultiCloud Marketplace ServerTemplates**

### Browse

**ServerTemplates**
RightScripts
MultiCloud Images
Macros

### Search

Keyword(s) [        ]

[GO]

Search in:
[Title        ▾]

▾ Show advanced options

### Categories

**All**

Shared

11H1 Compatible

All-In-One

Application Server

---

**Search Results For: ibm**          Sort By [Newest First ▾]

IBM  **Hadoop - IBM BigInsights Basic Edition (Data Node)**
by IBM                                                                No Ratings

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired      Nov 22, 2011
by, and is compatible with, open source Apache Hadoop and used to store, manage, and
gain insights from Internet-scale data at rest.

IBM  **Hadoop - IBM BigInsights Basic Edition (Master Node)**
by IBM                                                                No Ratings

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired      Nov 18, 2011
by, and is compatible with, open source Apache Hadoop and used to store, manage, and
gain insights from Internet-scale data at rest.

IBM  **IBM DB2 Express-C 9.7.4** by IBM                                ★★★★
                                                                      2 Ratings
IBM DB2 Express-C is the "no-limits" free edition of the enterprise-class IBM database       Nov 15, 2011
server. It features advanced technologies such as pureXML, a hybrid relational/XML
database engine. It's easy-to-use, includes self-managing capabilities, and offers
enterprise-proven performance and scalability.

IBM  **BigData University - IBM BigInsights Basic Edition
(Master Node)** by IBM                                                No Ratings

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired      Sep 22, 2011
by, and is compatible with, open source Apache Hadoop and used to store, manage, and
gain insights from Internet-scale data at rest.

IBM

## RiGHTSCALE®

Manage ▾   Design ▾   Clouds ▾   Reports ▾   Settings ▾

# MultiCloud Marketplace Macros

### Browse

ServerTemplates
RightScripts
MultiCloud Images
**Macros**

### Search

Keyword(s) [        ]

[ GO ]

Search in:
[ Publisher      ▾ ]

▾ Show advanced options

### Categories

**All**

**Shared**

---

**Search Results For: ibm**

Sort By [ Newest First ▾ ]

IBM  **Create BigInsights Cluster Deployment** by IBM

Oct 14, 2011  ⊙

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired by, and is compatible with, open source Apache Hadoop and used to store, manage, and gain insights from Internet-scale data at rest.

IBM  **Create BigInsights Deployment - BigData University**
by IBM

Sep 22, 2011  ⊙

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired by, and is compatible with, open source Apache Hadoop and used to store, manage, and gain insights from Internet-scale data at rest.

InfoSphere BigInsights Basic Edition, IBM entry-level offering, is availab

IBM  **Create BigInsights (Hadoop) Cluster Deployment - Simple two node setup** by IBM

Jul 14, 2011  ⊙

A core component of IBM's platform for Big Data, IBM InfoSphere BigInsights is inspired by, and is compatible with, open source Apache Hadoop and used to store, manage, and gain insights from Internet-scale data at rest.

IBM  **Setup Highly Available DB2 database cluster (BETA 1)**
by IBM

Apr 12, 2011  ⊙

Create a cluster of 2 DB2 servers in a high availability configuration (HADR).

Page 1 of 1

# Free Training at Your Own Pace

- RightScale is a Sponsor of Big Data University

- Free Course on Hadoop in the Cloud Includes:

  - RightScale

  - IBM BigInsights

  - Amazon EC2 $25 credits (limited quantity)

- Go to: BigDataUniversity.com

- Course Code: BD005EN

- Direct Link: http://www.bigdatauniversity.com/courses/course/view.php?id=309

# Resources – Learn More

- Web Resources:
  - RightScale.com/webinars
  - RightScale.com/whitepapers
  - Vimeo.com/RightScale
  - IBM.com/smartercomputing
  - BigDataUniversity.com

- Follow presenters on Twitter
  - @uribudnik
  - @katsnelson

### Contact Us

RightScale
sales@rightscale.com
1-866-720-0208

IBM IM Cloud Team
IMcloud@ca.ibm.com

IBM

> Questions

# Additional / Backup Slides

# Two Really "Hot" Areas: Cloud and Data

- "Big Data will earn its place as the next 'must have' competency in 2012" (IDC)

- "Through 2015, more than 85 percent of Fortune 500 organizations will fail to effectively exploit big data for competitive advantage" (Gartner).

- "2012 is likely to be a busy year for Big Data-driven mergers and acquisitions" (IDC)

- "[In 2012,] 80% of new commercial enterprise apps will be deployed on cloud platforms" (IDC)

- IP traffic over data center networks will reach 4.8 zettabytes a year by 2015, and cloud computing will account for one-third of it, or 1.6 zettabytes. (CISCO)

- "Amazon Web Services [will] exceed $1 billion in cloud services business in 2012 with Google's Enterprise business to follow within 18 months" (IDC).

# Merging the Traditional and Big Data Approaches

## Traditional Approach
*Structured & Repeatable Analysis*

## Big Data Approach
*Iterative & Exploratory Analysis*

**Business Users**

Determine what question to ask

**IT**

Delivers a platform to enable creative discovery

**IT**

Structures the data to answer that question
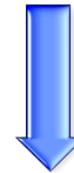
**Business**

Explores what questions could be asked

Monthly sales reports
Profitability analysis
Customer surveys

Brand sentiment
Product strategy
Maximum asset utilization

# BigInsights Content

| Function | Version | Basic Edition | Enterprise Edition |
|---|---|---|---|
| Integrated Install* | | Inc | Inc |
| Hadoop (including common utilities, HDFS, MapReduce framework) | 0.20.2 | Inc | Inc |
| Jaql (programming / query language) | 0.5.2 | Inc | Inc |
| Pig (programming / query language)* | 0.8.1 | Inc | Inc |
| Flume (data collection/aggregation) | 0.9.1 | Inc | Inc |
| Hive (data summarization/querying)* | 0.7.1 | Inc | Inc |
| Lucene (text search)* | 3.3.0 | Inc | Inc |
| Zookeeper (process coordination) | 3.3.3 | Inc | Inc |
| Avro (data serialization) | 1.5.1 | Inc | Inc |
| HBase (real time read/write)* | 0.90.4 | 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 |

*New or upgraded*

# BigInsights Content (cont'd)

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

*New or upgraded*

# BigInsights: Value Beyond Open Source
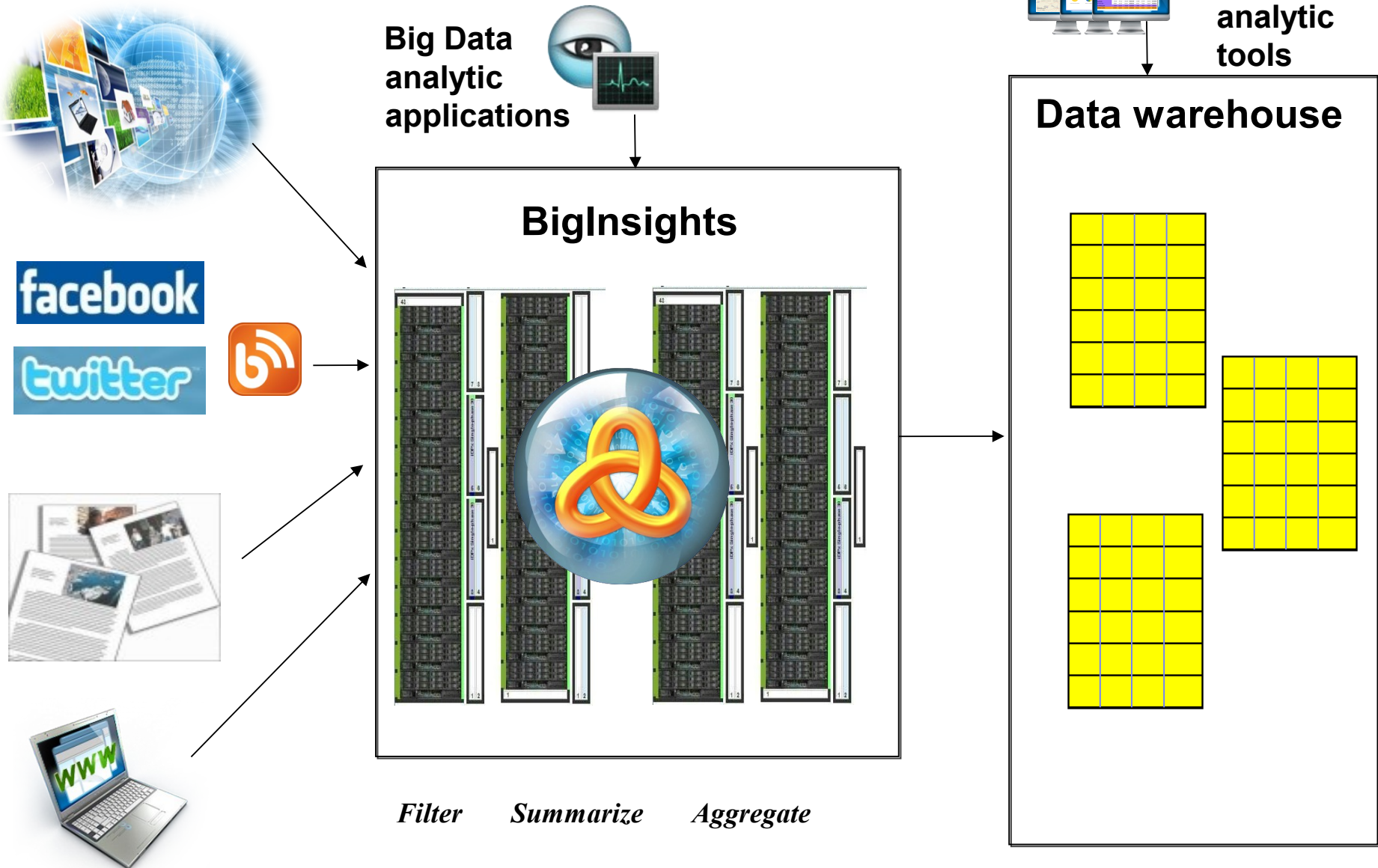
- **Technical differentiators**
  - Built-in analytics
    - Text processing engine, annotators, Eclipse tooling
    - Interface to project R (statistical platform)
  - Enterprise software integration (DBMS, warehouse)
  - Spreadsheet-style analytical tool for analysts
  - Ready-made business process 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**

## BigInsights

### Data warehouse

*Filter      Summarize      Aggregate*

# BigInsights and the data warehouse

**Traditional analytic tools**

**Big Data analytic applications**

**Data Warehouse**

**BigInsights**

- *Query-ready archive for "cold" warehouse data*