

Hi



Hadoop Meets Exadata

Presented by: Kerry Osborne

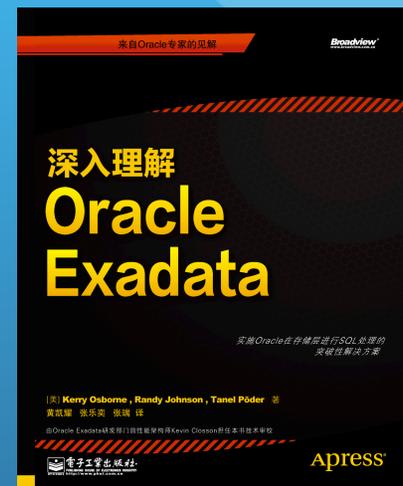
DW Global Leaders Program

Decemeber, 2012



whoami -

Never Worked for Oracle
Worked with Oracle DB Since 1982 (V2)
Working with Exadata since early 2010
Work for Enkitec (www.enkitec.com)
(Enkitec owns an Exadata Half Rack – V2/X2)
(Enkitec owns an Oracle Big Data Appliance)
Exadata Book (recently translated to Chinese)
Hadoop Aficionado



Blog: kerryosborne.oracle-guy.com
Twitter: @KerryOracleGuy



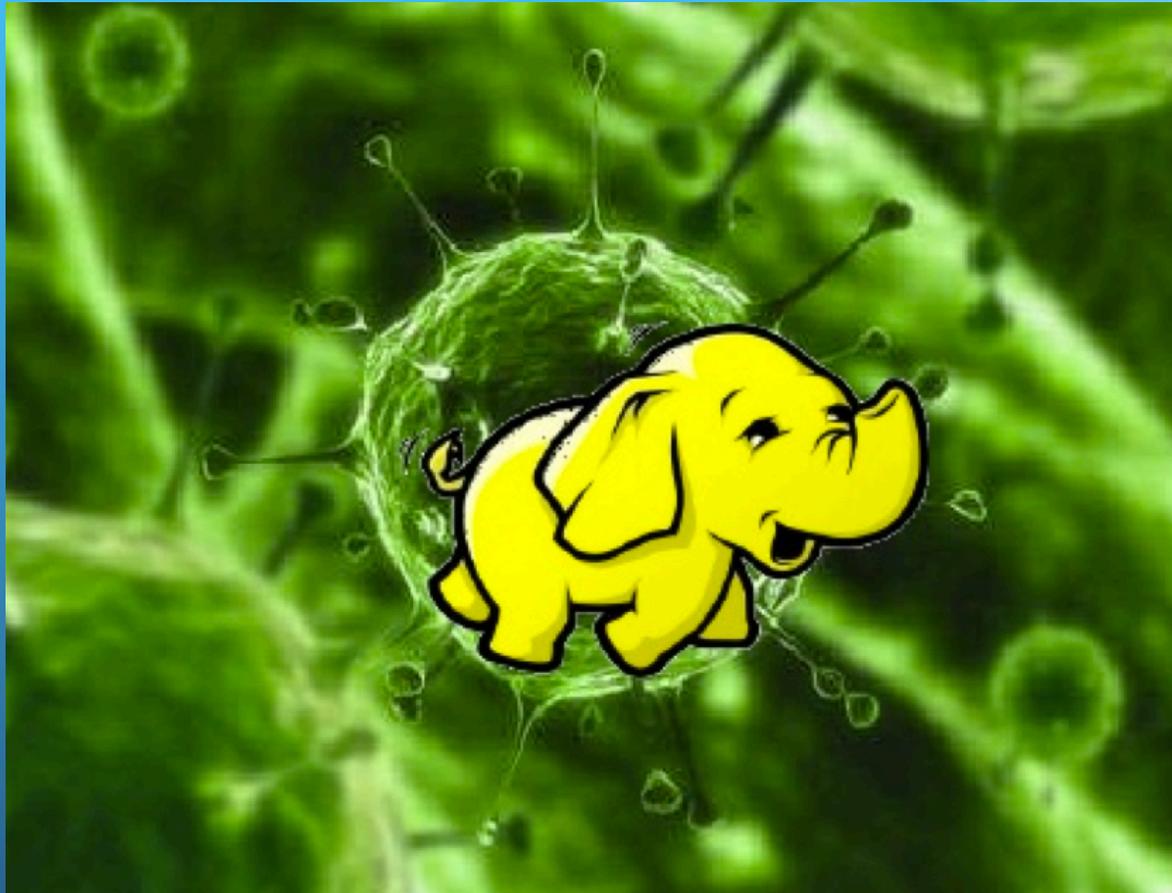
What's the Point?



Data Volumes are Increasing Rapidly
Cost of Processing / Storing is High
Scalability is Big Concern

And ...

Hadoop Is A Virus



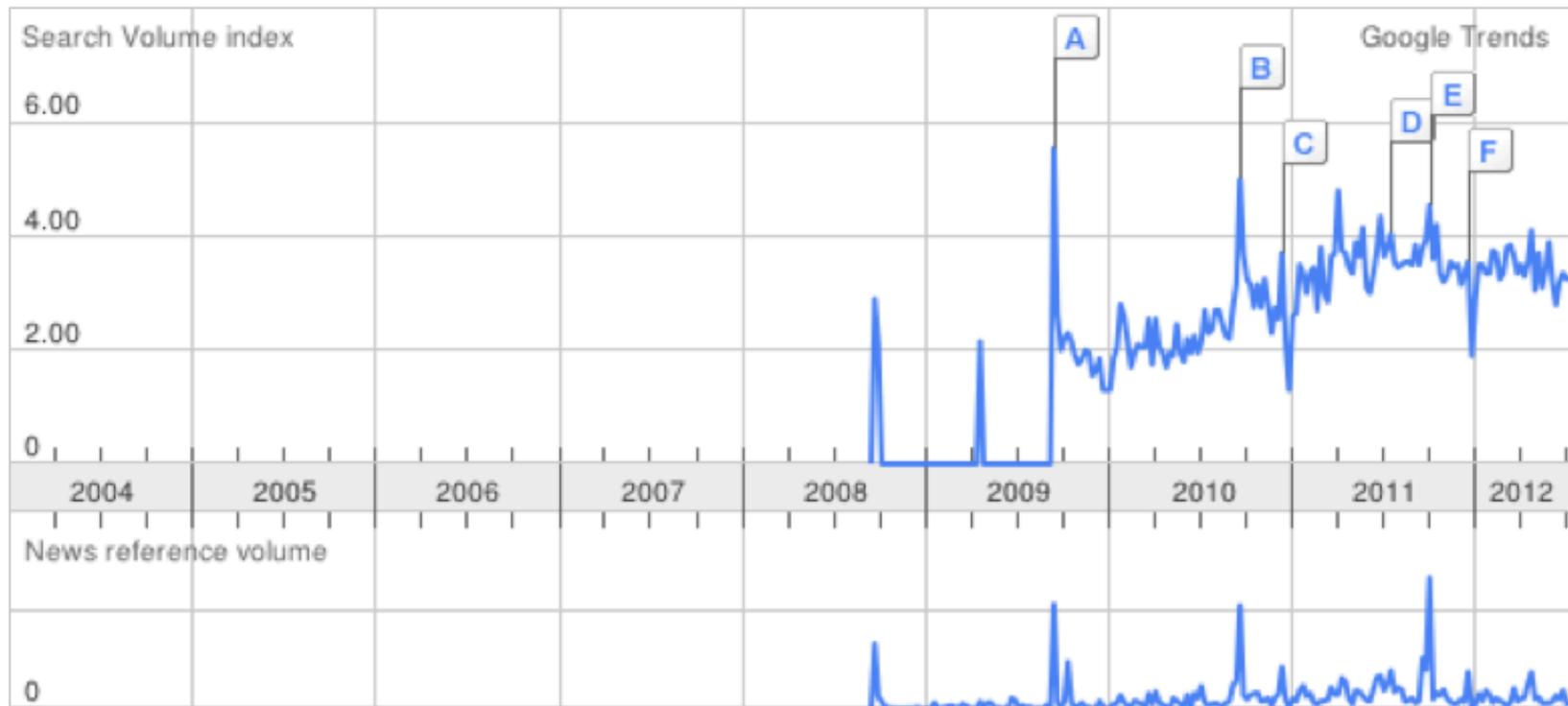
* Stolen from Orbitz

enkitec

Google Trends

exadata

1.00



Rank by

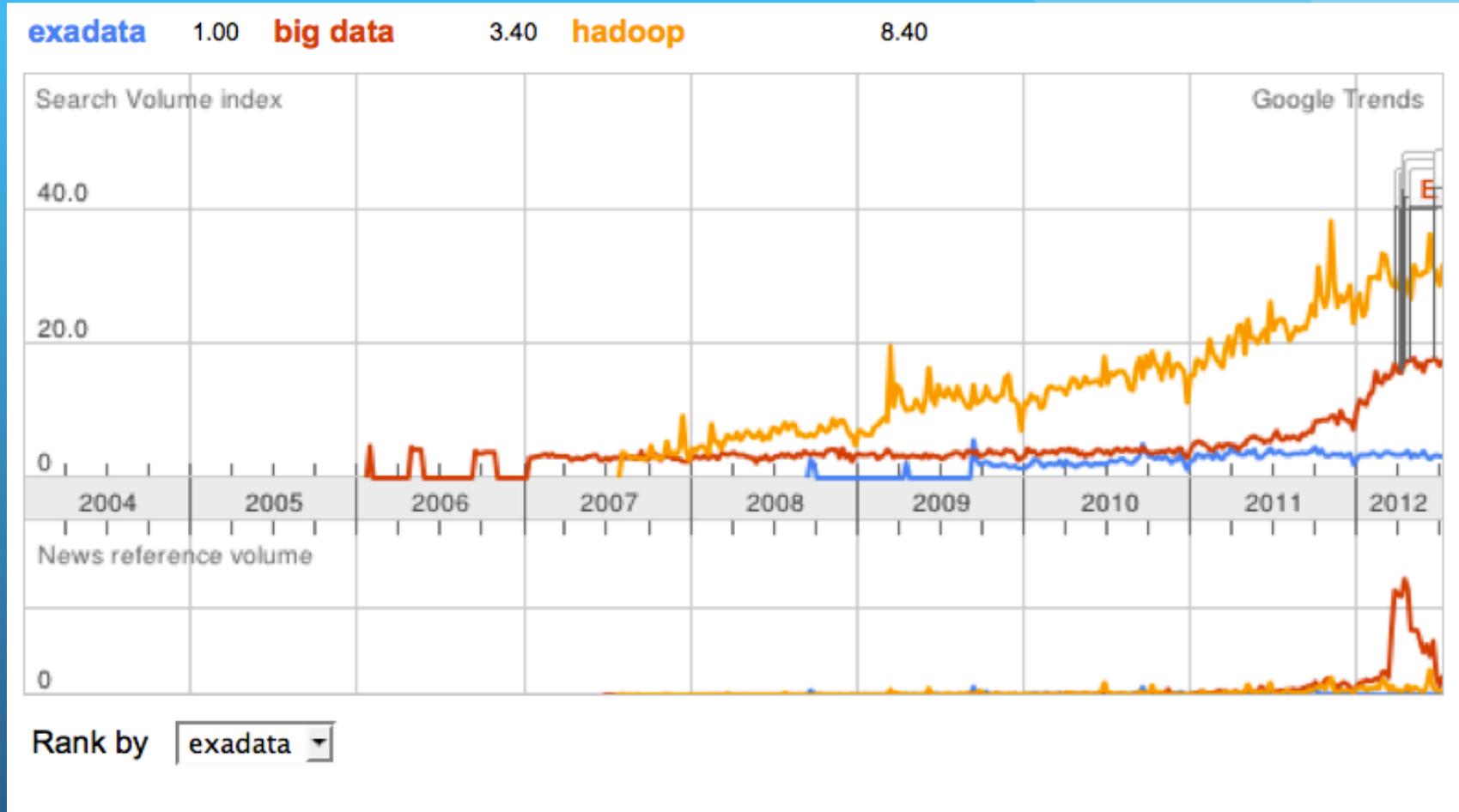
Google Trends

exadata 1.00 big data 3.40



Rank by exadata

Google Trends

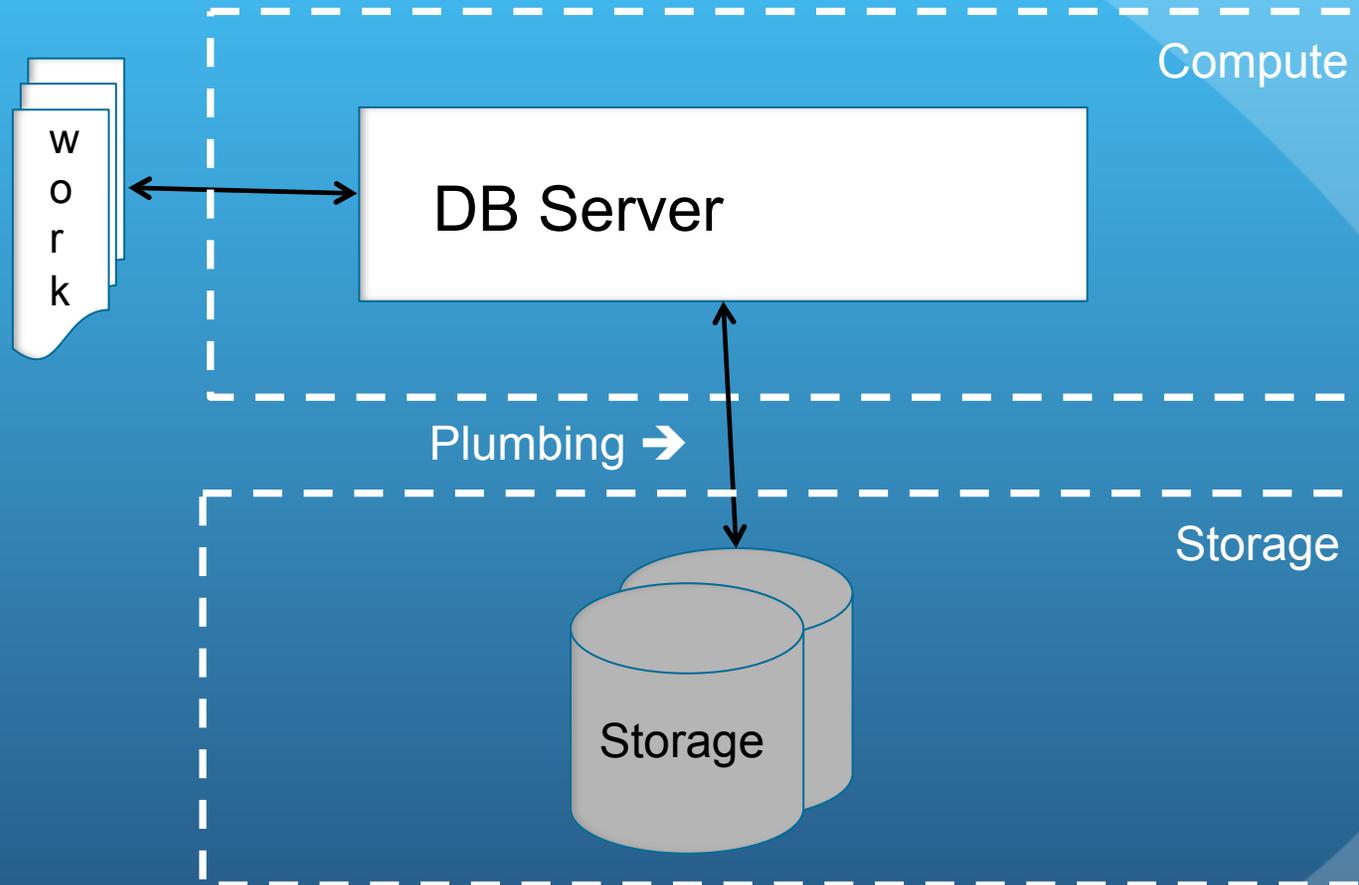


Disjointed Presentation ???

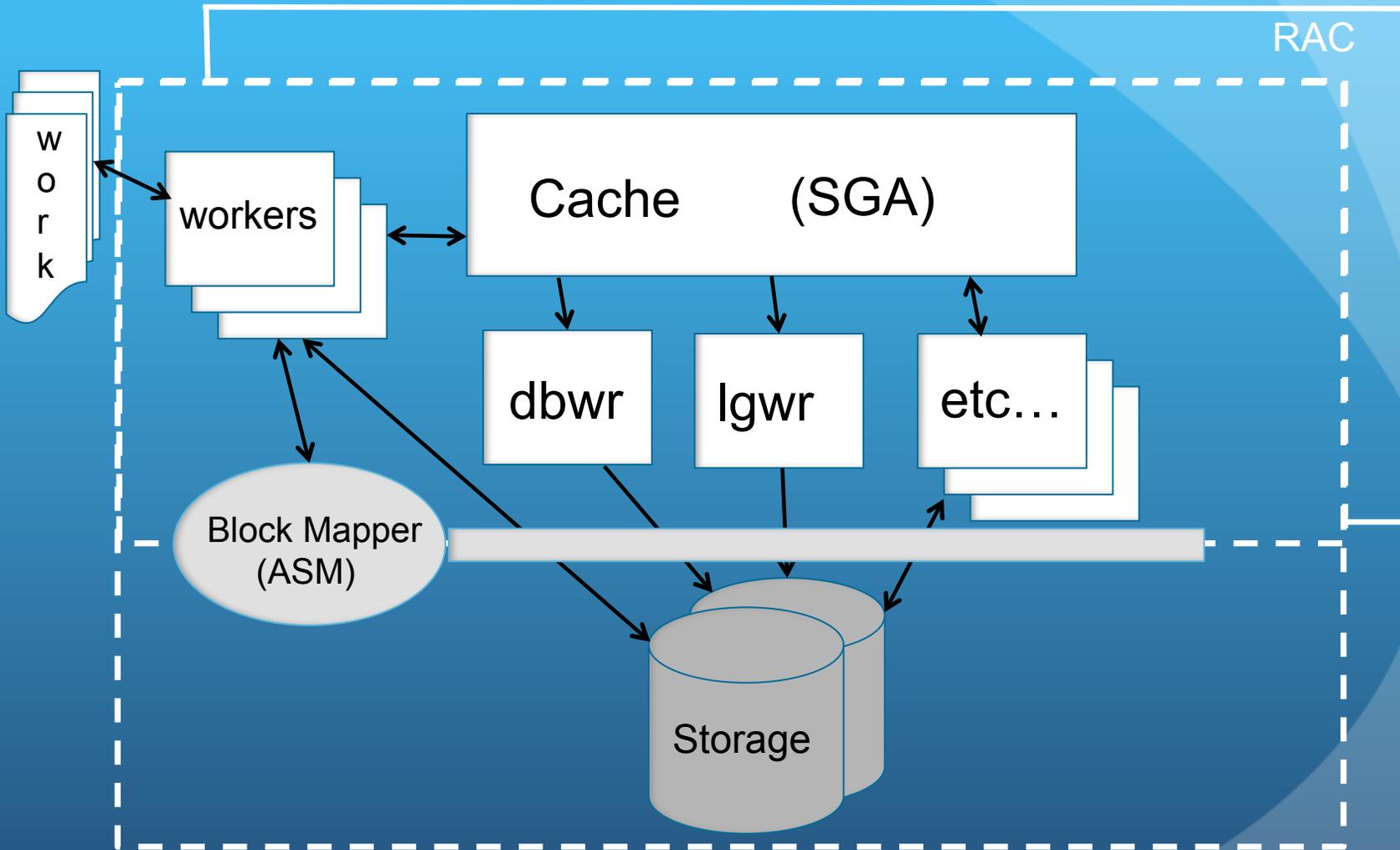


Architectures
Integration Approaches
Oracle Products
Exadoop Case Study

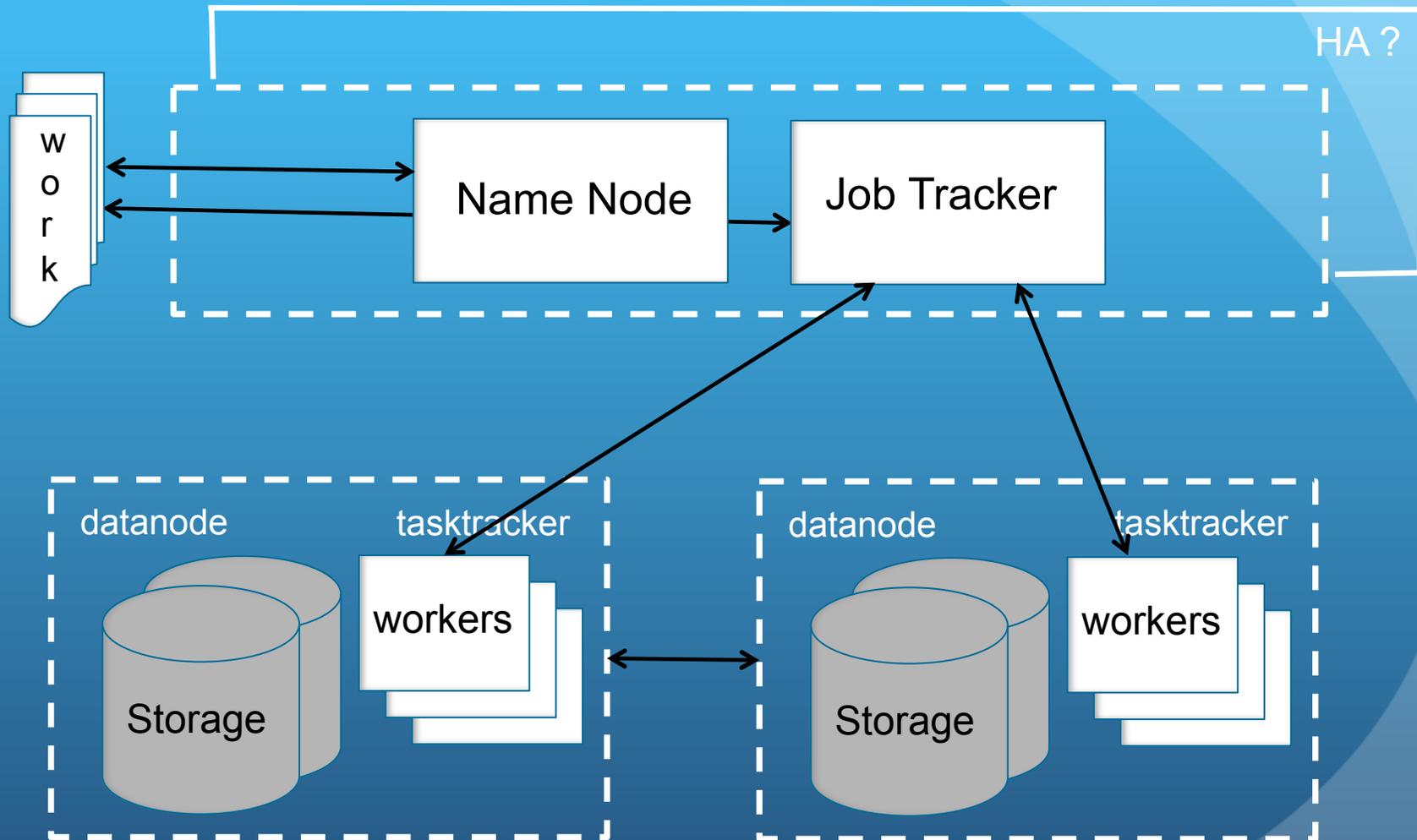
Traditional RDBMS Architecture



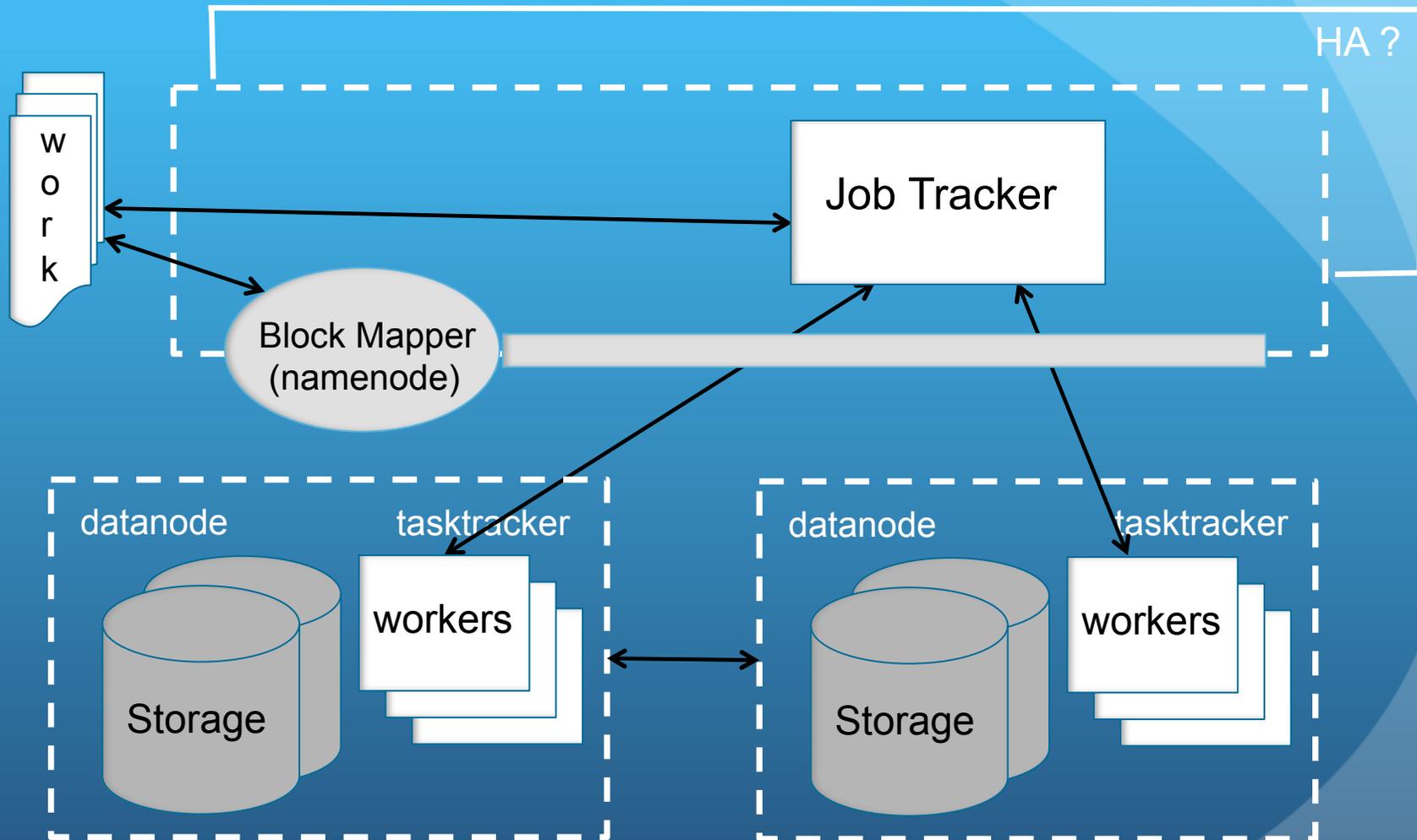
Traditional Oracle Architecture



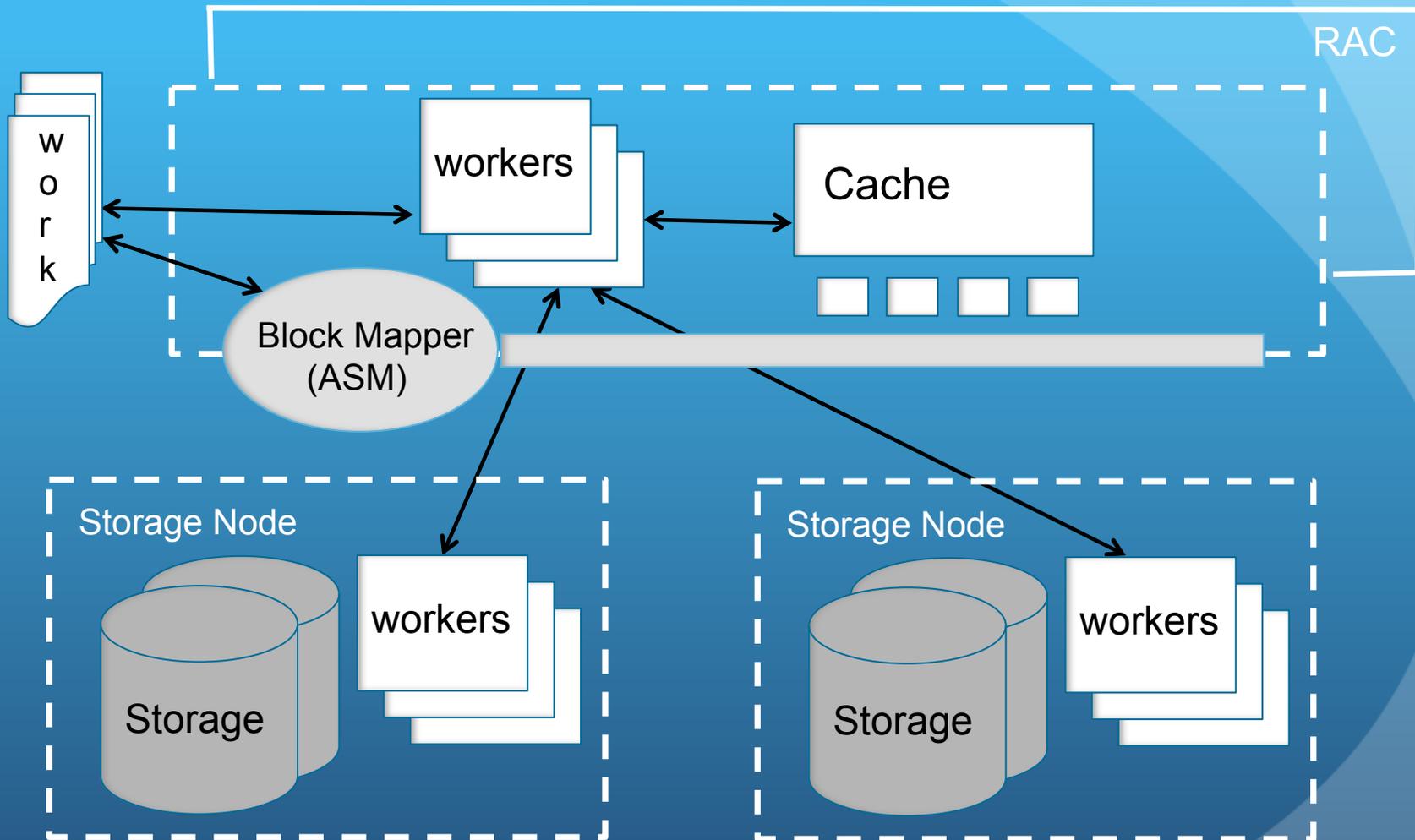
HDFS/Hadoop Architecture



HDFS/Hadoop Architecture

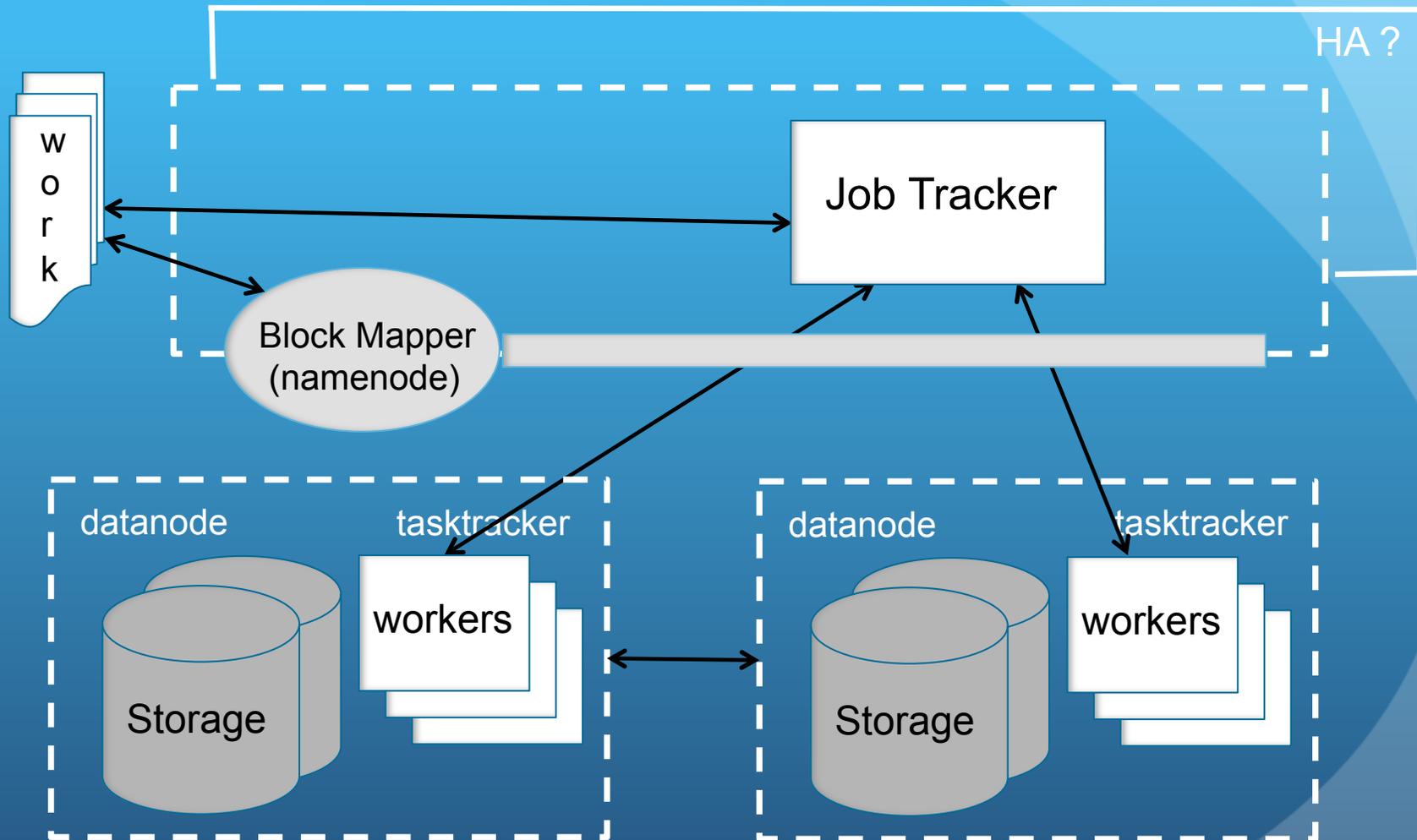


Exadata Architecture



enkitec

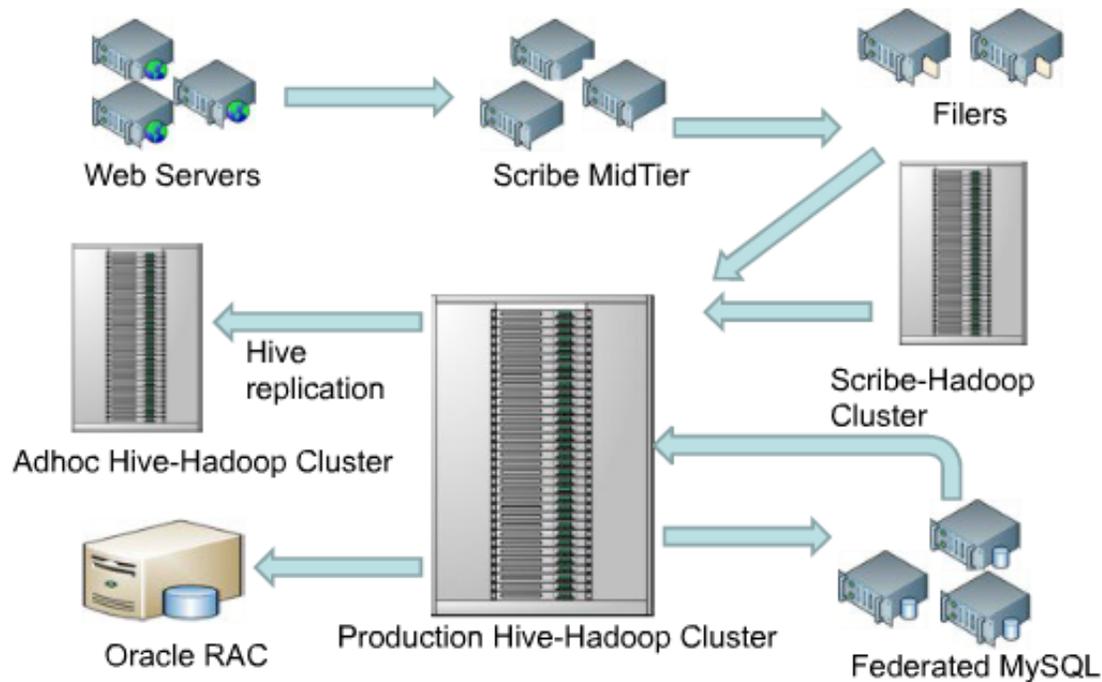
HDFS/Hadoop Architecture



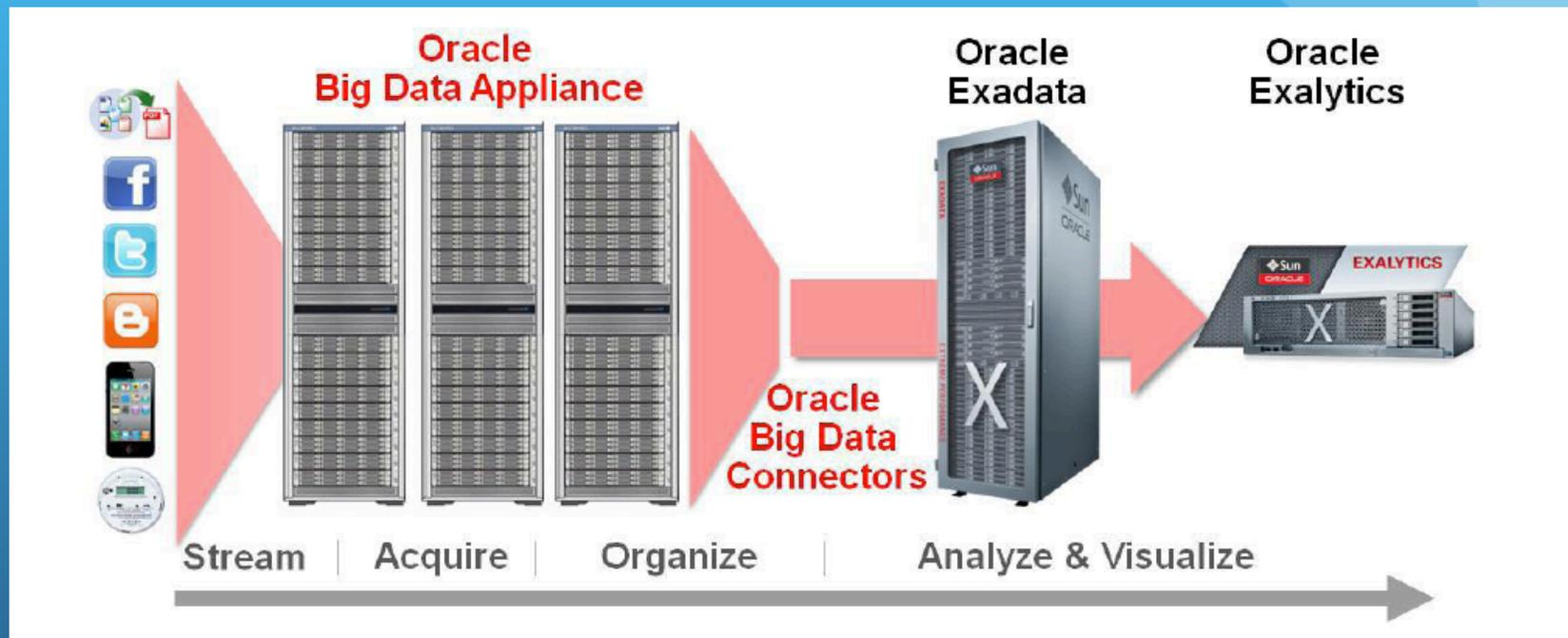
Oracle + Hadoop Integration

facebook

Data Flow Architecture at Facebook



Obligatory Marketing Slide



Oracle Big Data Appliance

Prebuilt Hadoop Stack in a Rack
Engineered System
Open Source Software
Includes Cloudera Distribution



Oracle Big Data Appliance

Big Data Appliance	
Hardware Specification and Details	
18 Compute and Storage nodes	<p>Per Node:</p> <ul style="list-style-type: none"> • 2 x Six-Core Intel® Xeon® 5675 Processors (3.06 GHz) • 48 GB Memory (expandable to maximum 144GB) • Disk Controller HBA with 512MB Battery backed write cache • 12 x 3TB 7,200 RPM High Capacity SAS Disks • 2 x QDR (40Gb/s) Ports • 4 x 1 Gb Ethernet Ports • 1 x ILOM Ethernet Port
2 x 32 Port QDR InfiniBand Switch	<ul style="list-style-type: none"> • 32 x InfiniBand ports • 8 x 10GigE ports
1 x 36 Port QDR InfiniBand Switch	<ul style="list-style-type: none"> • 36 InfiniBand Ports
Additional Hardware included:	<ul style="list-style-type: none"> • Ethernet switch for administration of the Appliance • Keyboard, Video or Visual Display Unit, Mouse (KVM) hardware for local administration • 2 x Redundant Power Distributions Units (PDUs) • 42U rack packaging
Spares Kit Included:	<ul style="list-style-type: none"> • 2 x 3 TB High Capacity SAS disk • InfiniBand cables

BDA Software

Big Data Appliance

Integrated Software

Oracle Enterprise Linux 5.6

Oracle Hotspot Java Virtual Machine

Cloudera's Distribution including Apache Hadoop

Cloudera Manager

Open Source Distribution of R

Oracle NoSQL Database Community Edition

Top Secret Feature of BDA



enkitec

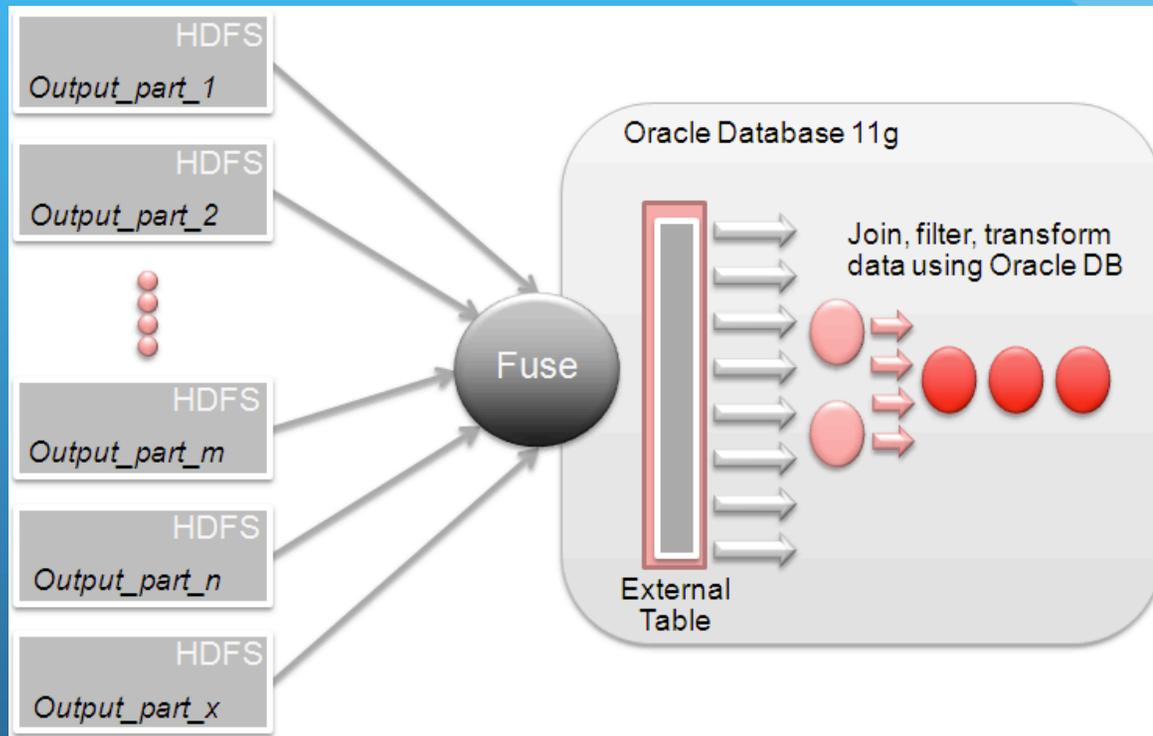
Integration Options

Many Ways to Skin the Cat

- Fuse
- Sqoop
- Oracle Big Data Connectors



Fuse - External Tables



Sqoop (SQL-to-Hadoop)



- Graduated from Incubator Status in March 2012
- Slower (no direct path?)
- Quest has a plug-in (oraoop)
- Bi-Directional

Oracle Big Data Connectors

Oracle Loader for Hadoop - OLH

Oracle Direct Connector for HDFS - ODCH

Oracle R Connector for Hadoop – ORHC

Oracle Data Integrator Application Adapter for Hadoop

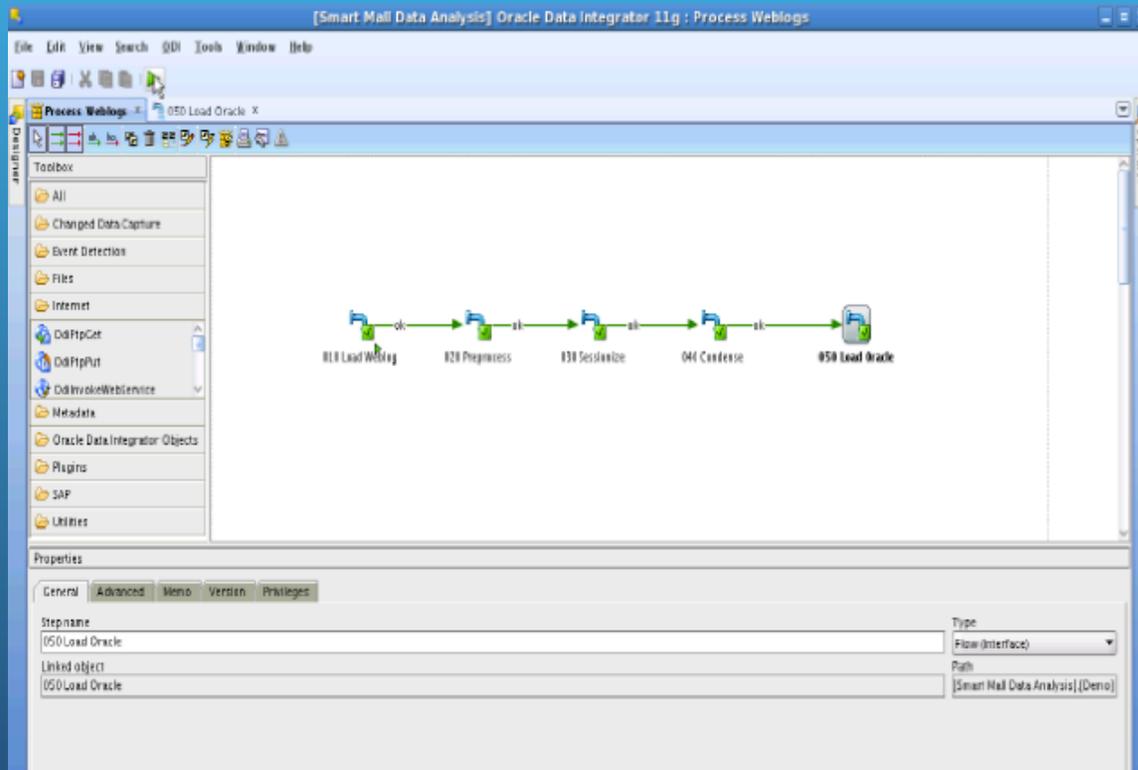
Note:

All Connectors are One Way



Oracle Data Integrator Application Adapter for Hadoop

ODIAAH ?



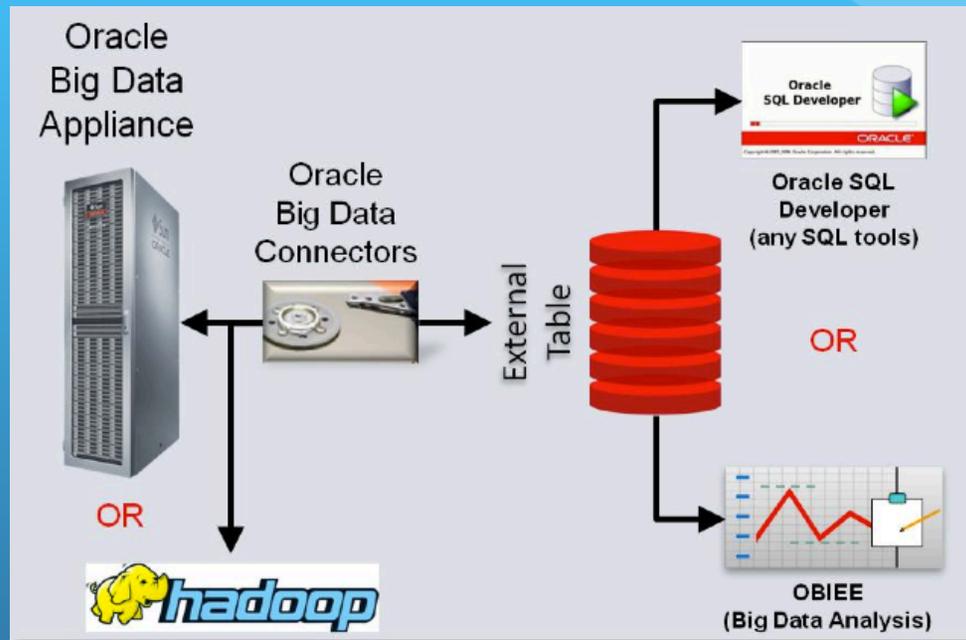
Oracle R Connector for Hadoop (ORHC)

- Provides ability to pull data from Oracle RDBMS
- Provides ability to pull data from HDFS
- Provides access to local file system
- Not really a loader tool
- Most useful for analysts

Oracle Loader for Hadoop (OLH)

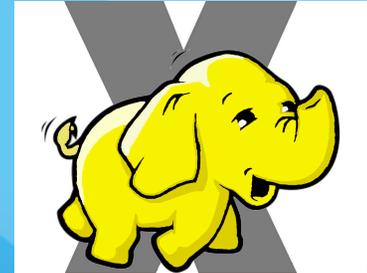
- Implemented as a MapReduce job (oraloader.jar)
- Saves CPU on DB Server
- Can convert to Oracle datatypes
- Can partition data and optionally sort it
- Online – direct into Oracle tables
 - Can load into Oracle via JDBC or OCI Direct Path
- Offline – generate preprocessed files in HDFS (DP format)

Oracle Direct Connector for HDFS (ODCH)



- Uses External Tables
- Fastest - 12T per hour
- Can load DP files preprocessed by OLH
- Allows Oracle SQL to query HDFS data
- Doesn't require loading into Oracle
- Downside – uses DB CPU's

Exadoop

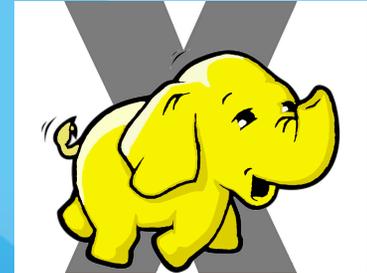


* Mad Scientist Project



enkitec

Exadoop



Unusual Situation!

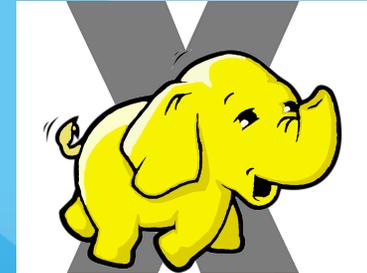
Exadata Half Rack with 4 Spare Storage Servers

Company Playing with “Big Data” Technology

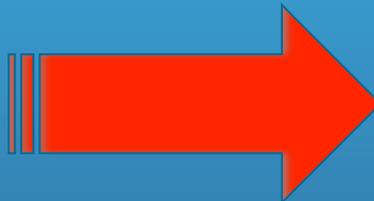
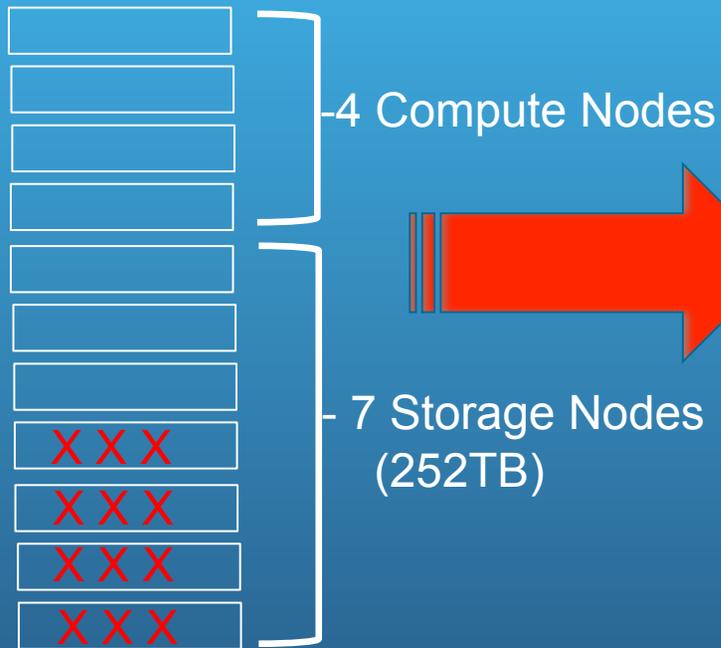
Exadata Cells Very Similar to BDA Servers

4 Cells \approx Mini BDA! (happy face)

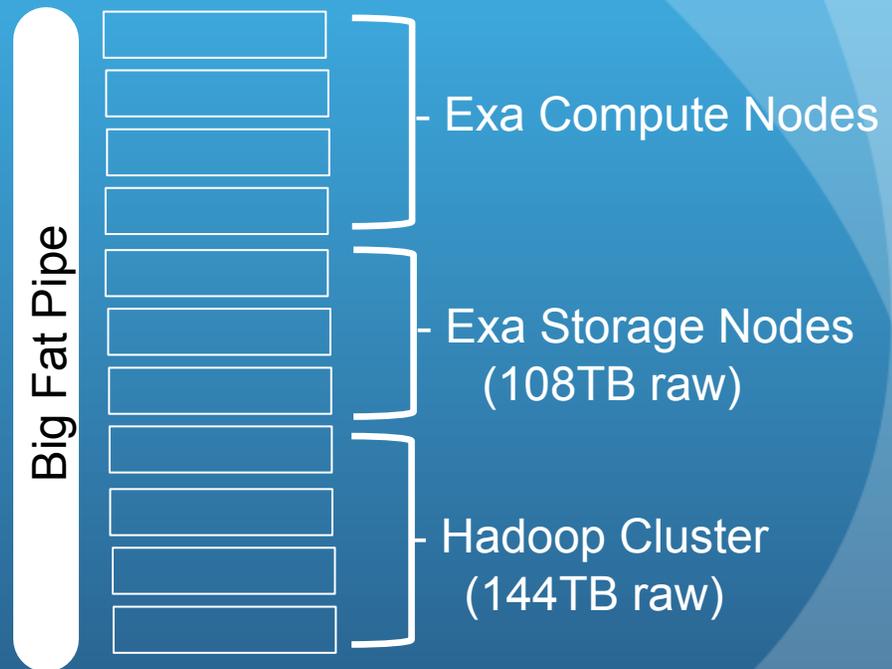
Exadoop Layout



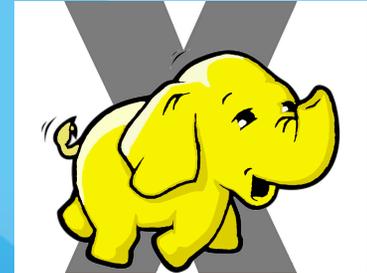
Exa Half Rack



Exadoop

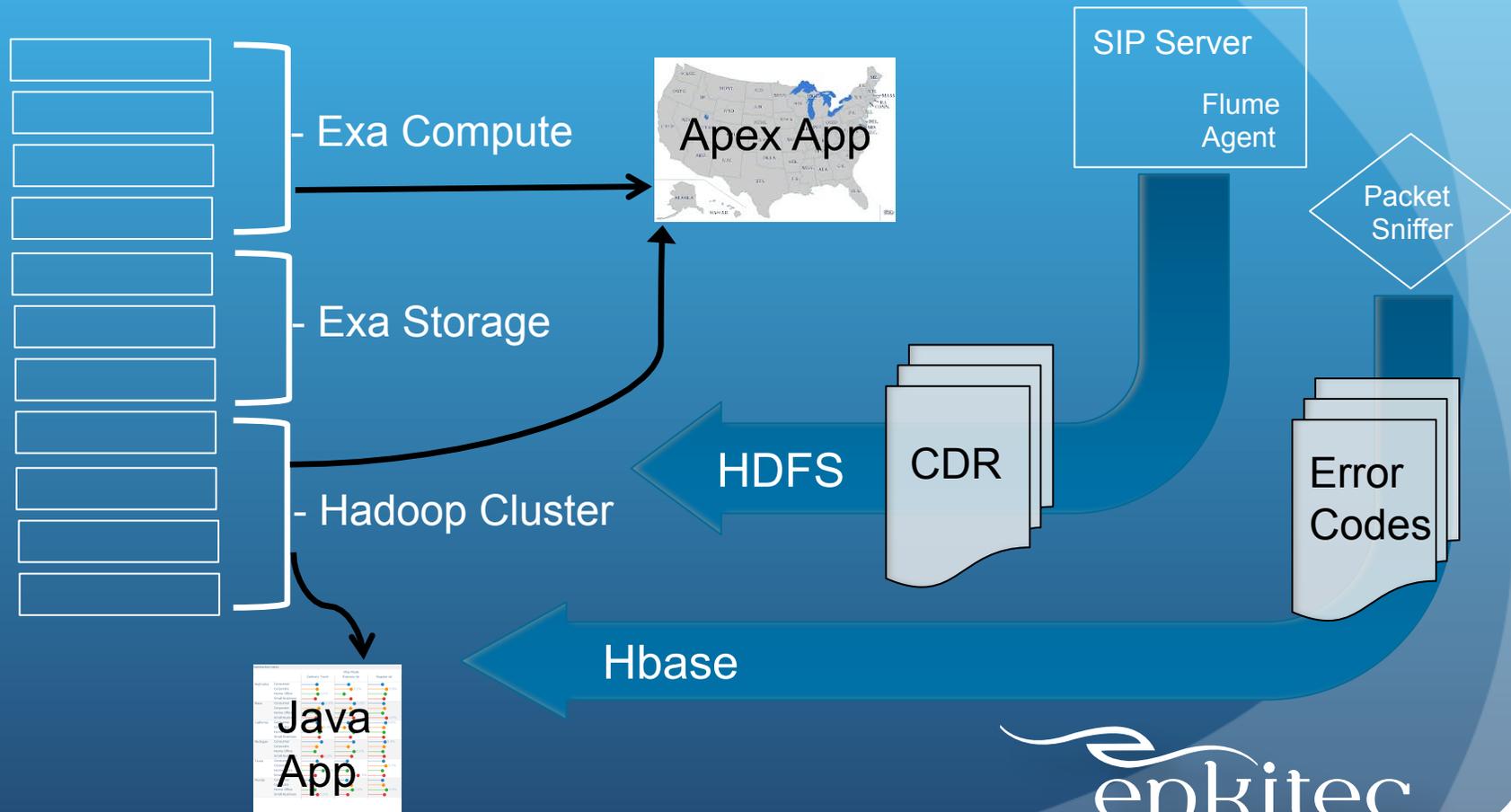
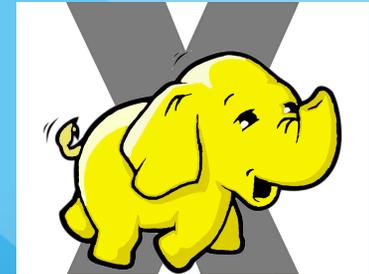


Exadood Applications



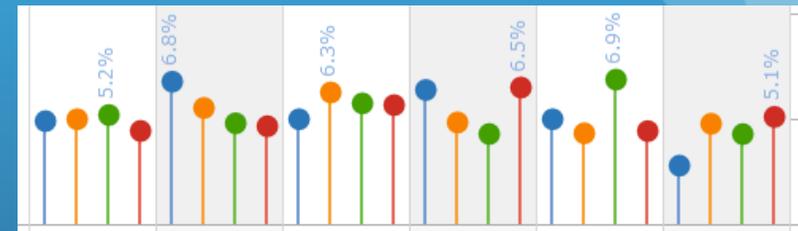
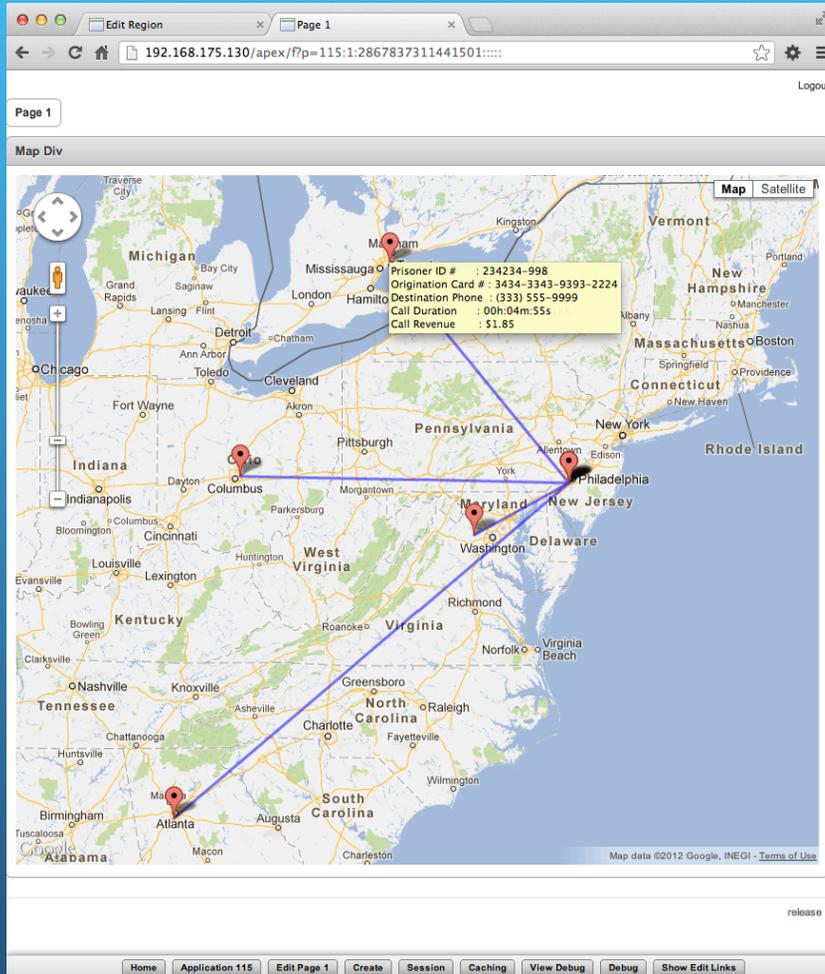
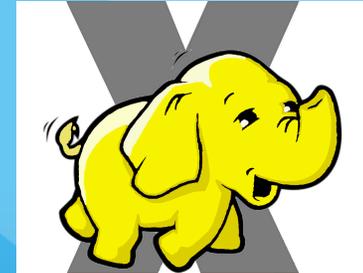
Telecom Company
Call Detail Records Dumped by Switches
Loaded into HDFS via Flume

Exadoop - Proposed Architecture



enkitec

Exadoop Applications



Wrap Up



Is Hadoop the right tool for the job?

Maybe

All the Cool Kids Are Doing It!



Questions?

Contact Information : Kerry Osborne
kerry.osborne@enkitec.com
kerryosborne.oracle-guy.com
www.enkitec.com