

ORACLE

Big Data – Are You Ready?

Thomas Kyte http://asktom.oracle.com The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

What's New Since Oracle OpenWorld 2010?

Dec 2010 Oracle Exadata Database Machine X2-8 shipping



- Feb 2011 Oracle Database Firewall
- Apr 2011 Oracle Exadata Automatic Service Requests
- May 2011 Oracle Exadata Solaris support
- Jun 2011 Oracle Exadata certification of SAP
- Jun 2011 Oracle Exadata Storage Expansion Racks
- Sep 2011 Oracle Database 11.2.0.3 Patch Set released
- Sep 2011 Oracle Database Appliance shipping



Big Data Buzz

"Why big data is a big deal"

InfoWorld – 9/1/11

"Keeping Afloat in a Sea of 'Big Data"

ITBusinessEdge – 9/6/11

"The challengeand opportunityof big data"

McKinsey Quarterly—5/11

"Getting a Handle on Big Data with Hadoop"

Businessweek-9/7/11

"Ten reasons why Big Data will change the travel industry"

Tnooz -8/15/11

"The promise of Big Data"

Intelligent Utility-8/28/11

Big Data Use Cases

Today's Challenge	oday's Challenge New Data			
Healthcare Expensive office visits	Remote patient monitoring	Preventive care, reduced hospitalization		
Manufacturing In-person support	Product sensors	Automated diagnosis, support		
Location-Based Services Based on home zip code	Real time location data	Geo-advertising, traffic, local search		
Public Sector Standardized services	Citizen surveys	Tailored services, cost reductions		
Retail One size fits all marketing	Social media	Sentiment analysis segmentation		

What Makes it Big Data?



Why Is Big Data Important?



ORACLE

Source: * McKinsey Global Institute: Big Data - The next frontier for innovation, competition and productivity (May 2011)





Acquire all available data



Acquiring Big Data Challenge

Need to process high volume, lowdensity information Application will need to change frequently

Must scale out to meet aggressive roll out plan



Key value pair database Dynamic data model Highly scalable, available Transparent load balancing Built using BerkeleyDB



Key value pair database Dynamic data model Highly scalable, available Transparent load balancing Built using BerkeleyDB



Oracle NoSQL: Practically ACID

. . .

The serious part of Oracle NoSQL is a practical approximation of ACID compliance, the standard that SQL databases like to offer. ACID means "Atomic, Consistent, Isolated, Durable transactions," and there's a robust debate about just what this translates to in excruciating detail. Most NoSQL systems promise a different acronym, BASE, which stands for "Basically Available, Soft State, and Eventually Consistent." In other words, you'll probably get the right answer except when you don't. Key value pair database

Dynamic data model

Highly scalable, available

Transparent load balancing

Built using BerkeleyDB

In all, Oracle NoSQL was a pleasure to try because it offered so many serious features developed by a company with a deep history of serious data management. There are dozens of small ways in which the tool is more thorough and sophisticated than the simpler NoSQL projects. You get a number of different options for increasing the durability in the face of a node crash or trading that durability for speed. The documentation is solid and written by working engineers with deep experience in storing data for enterprise customers.

Key value pair database

Dynamic data model

Highly scalable, available

Transparent load balancing

Built using BerkeleyDB





Organize and distill big data using massive parallelism

Organizing Big Data Challenge

Have existing Oracle data warehouse Also want to perform analysis on big data

Can't negatively impact data warehouse SLAs

Analysis Sandbox



Provides analysis workspace

Controlled access to resources and data

Doesn't impact production system

Sandboxing with Oracle Enterprise Manager

RACLE				Hab ≁	soc. 🎎	Log Out 🔮
tabase Cloud Self Service Portal				Page Rafreshed Sep 25, 2	011 5:33:571	рм рот 🕻
Home Chargeback Ny Preference	CRE				Servers	Databaran
Notifications	Hy Databases					_
🦄 Detabases Due to Expire in Next 7 Days 2	Veve 💌 Request Database 💥	Delete				
-	Service Name 🛆 🗸	Type $\Delta \nabla$	Status	Zorse Name	Start Date	End Dat
Your Usage You have permission to use these cursulative note allowerner when onlying distributes	D6353bac.adc2101112.us.oracle.com	Database Instance	Ŷ	11202_5106_11M0032	9/22/2011	10/7/20
	DB+79444.adc2101112.us.orade.com	Database Instance	÷	1.1202_SIDB_LINU#32	9/22/2011	10/7/20:
	D6681ea5.adc2101112.us.orade.com	Database Instance	Ŷ	11202_505_1040032	9/14/2011	9(25)(20)
equeste.	D669dtaf.sta00138.us.orade.com	Database Instance		11202 SIDE 20NE 64 BIT	9(13)(2011	9/30/20
femory: 3.42 GB	My Requests					
50 50	Remark Name	Sabu	Creat	ion Date Start Date	End	Date
tonege: 1.72 GB	JDOE - Tue Sep 13 18/21/36 PDT 2011	Success	903	9/2011 9/13/2011	9,130	12011
	JDOE - Wed Sep 14 12:20:50 PDT 2011	Success	9(14	4/2011 9/14/2011	9,128,	2011
	300E - Thu Sep 22 15:07:27 PDT 2011	Success	9/23	2/2011 9/22/2011	10,17,	2011
0 100	JDOE - Thu Sep 22 15:25:30 PDT 2011	Success	9/2	2/2011 9/22/2011	10,17,	2011
	<					2

Simple to set up

Efficient server utilization

Secure and scalable

Accountable via charge back

Ideal for Oracle Exadata



Organizing and Distilling Big Data Challenge

Must transform big data into something easily analyzed Want to avoid writing lots of Hadoop code Need to load data quickly into Oracle Data Warehouse

Hadoop Architecture



Distributed file system with redundant storage

Map/Reduce programming paradigm

Highly scalable data processing

Cost-effective model for high volume, low density data

A Map/Reduce Pipeline



Oracle Data Integrator



Reduces Hadoop complexities through graphical tooling

Oracle Loader for Hadoop





Big Data in Action

Oracle NoSQL Database

Oracle Enterprise Manager

Oracle Data Integrator

Oracle Loader for Hadoop



Analyzing Big Data Challenge



Want to perform statistical analysis using R

Doing analysis on a laptop is slow and not secure

R Statistical Programming Language



Open source language and environment

Used for statistical computing and graphics

Strength in easily producing publication-quality graphs

Highly extensible

Why R Wasn't Ready for the Enterprise



Small data models only are stored and run on user's laptop



Oracle R Enterprise Approach





Oracle Enterprise Manager

Oracle Data Integrator

Oracle Loader for Hadoop

Oracle R Enterprise





Making Decisions Based on Big Data Challenge

Big data has been transformed into actionable insight Want to add new insights into BI dashboard

How do we quickly integrate R analytics into dashboard?

Dashboard Analytics

- Oracle Business Intelligence Enterprise Edition
 - -Advanced dashboard visualization
 - -Runs BI and EPM applications
- Integrating R Analytics
 - -Embed R script's web interface in BI dashboard
 - -Graphics will stream to BI dashboard





Oracle Exalytics Hardware

Engineered for extreme analytics

- •40 Intel processor cores
- •1 Terabyte main memory





Oracle Exalytics Software

Oracle TimesTen In-Memory Database
–Adaptive in-memory caching of analytics
–In-memory columnar compression
–Tightly integrated with Oracle Exadata
–Enables speed-of-thought visualization



•Oracle Business Intelligence Foundation Suite



Oracle Integrated Solution Stack for Big Data



Oracle Big Data Appliance Hardware

- •18 Sun X4270 M2 Servers
 - -48 GB memory per node = 864 GB memory
 - -12 Intel cores per node = 216 cores
 - -24 TB storage per node = 432 TB storage
- •40 Gb p/sec InfiniBand
- •10 Gb p/sec Ethernet





Oracle Big Data Appliance Software

- Oracle Linux
- Java Hotspot VM
- Apache Hadoop Distribution
- R Distribution
- Oracle NoSQL Database
- Oracle Data Integrator for Hadoop
- Oracle Loader for Hadoop

No. 10



The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Hardware and Software

ORACLE

Engineered to Work Together





Maximizing the Value of Enterprise Big Data

- •Hardware and software for Big Data
- Integrates all enterprise data
 - -Structured and unstructured
 - -SQL and NoSQL
- •Fastest time-to-market
- •Single vendor support

