Big Data, Platforms, Tools, and Advanced Analytics Applications and Capabilities – What do you Need

Steven Hagan, Vice President, Server Technologies
Global Digital Data Growth
Growing leaps and bounds by 40+% YoY!

2009 = .8 Zetabytes
= .08 ZB Structured Data
= .72 ZB Unstructured Data

2020 = 35 Zetabytes
= 3.5 ZB Structured Data
= 31.5 ZB Unstructured Data*

(1 Zetabyte = 1 Trillion Gigabytes)

- Chart conservatively assumes a constant 9:1 ratio of unstructured data vs. structured data (based upon IDC’s estimate that 90% of all digital data is unstructured).
- Chart does not reflect IDC’s projection that unstructured data is currently growing twice as fast as structured data at the rate of 63.7% vs. 32.3% CAGR.

Source: IDC Digital Universe Study, A Digital Universe Decade – Are You Ready?, 2010
Global Drivers for Analyzing Place-Events-People

- **BIG Data** – Terabytes, Petabytes, Exabytes, Zettabytes, Yottabytes
  - Sensors, RFID, VIDEO, LIDAR, Raster, 3D, Terrain and City Models
  - Tagged Data, Social Media, Semantics, Ontologies
  - SDIs, INSPIRE, Linked Open Data — Persistent Relationships

- **BIG Hardware:**
  - CLOUD Platforms – Public and Private
  - Cheaper, more powerful – Clusters of Commodity Servers, Virtualization: = Greener
  - Massively parallel database machines (Software Enablement) – e.g. Hadoop, Oracle

- **BIG Software** –
  - Real Time Analytics – Biggest value from fastest response – Streams and Events — Internet of Things; Spatially Aware
  - CyberSecurity
  - Location Enable All Applications: ERP, CRM, Business Intelligence, Public Sectors
  - Engineered Systems – Fully installed and tested (Labor Cost is now Dominant Factor)
  - Support Standards for Interoperability – W3C, ISO, OGC, Wide Range
Data Volume & Variety Explosion Continues - Terabytes, Petabytes, Exabytes, Zettabytes

- Sensors, RFID, LIDAR, Raster, 3D, Terrain and City Models, SDIs
- New data products for consumers, mobility, defense, intelligence, land and water mgmt, transportation, environment, agriculture, and constituent services
- Terrain Models and 3D for planning, maintenance, emergency response, tourism
- Tagged Data, Semantics, Ontologies - Location is a Powerful Organizing Principle
- Integrate Social Media (Video, Audio, Text, Wikis, Facebook, Imagery) with Spatial
Data Velocity: Spatially-aware Real-Time Streams / Events / “Internet of Things”

Track Moving Objects – Cars, UAVs

- Ultra-high throughput (1 million/sec++) and microsecond latency
- Detect patterns in the flow of events and message payloads, Complex Event Processing-CEP
- Filtering, correlation, and aggregation across event sources
- Business Intelligence in Real Time
Definition: What is Big Data Anyway?

‘The 4 Vs’ of Big Data?

**Volume** – It’s about 100s TeraBytes and Petabytes... BUT could be smaller volumes – size is in the eye of beholder

**Variety** – Highlights the new types of data from the ‘internet of things’ that can be stored and analysed

**Velocity** – Relates to streams of high frequency data arriving e.g. From a sensor, network, Imagery, UAV, or high volume event

**Value** – At volume low value data can highlight high value patterns, trends and insights of significant commercial value
Detecting Signal in Noise = New Insight

What’s the effectiveness of a re-employment campaign? How has it effected the economy?

What do citizens think about policies & services?

Can I track certain suspicious individuals?

How can sensor data improve maintenance efforts for key assets?

Can we detect issues and trends before liabilities increase?

What do citizens think about policies & services?
Big Data in Public Sector
Examples of Different Program Areas

1. Fraud Prevention
2. Maintenance & Utilities
3. Constituent Sentiment
4. Threat Identification
5. Economic Analysis
6. Healthcare
7. Regulatory Compliance, Licensing & Enforcement
8. Open Government
9. Tax Collections

Use Cases
Big Data Poses Big Challenge For Military Intelligence

The amount of data generated is reaching epic proportions

• Number of sensors deployed on the ground or aloft on unmanned aerial vehicles (UAVs) and satellites continues to soar, as does the ability to keep this asset in place for longer periods
• Volume of data being gathered for intelligence, surveillance and reconnaissance (ISR) is already huge. In Afghanistan, ISR acquisition systems gather more than 53TB of data every day.

“There’s a gap between our growing ability to collect data and our limited ability to process that data. We are collecting 1,500% more data than we did 5 years ago. At the same time, our head count has barely risen.”

Gen. Robert Kehler, commander of the U.S. Strategic Command, late 2011

Throughout the defense industry, there’s a concerted effort to deal with so-called big data. There is a recognition that it is simpler to resolve the problems with technology than to boost the staff of skilled analysts.

Source: Terry Coslow, Defense Systems, 29 March 2012
What Can Federal Agencies Do With Big Data?

Example: Weapons of Mass Destruction Tracking

Combining Information on:

- People
- Goods
- Money
- Materials
- Transportation

To give analysts awareness never before possible
Example: DHS Risk Analysis

External Data Sources
- Transactional & Operational Systems
- Contents Repository
- Databases
- Web resources
- Blogs, Mails, news

Real-time Data Streams

Search, Presentation, Report, Visualization, Query

Enterprise Data Management Infrastructure
- GeoSpatial
- POIs
- Documents
  - Historical Records
  - Demographics
  - Customer Data
  - Call Records

Automatic Responses and Publishing
- SMS
- Console Alerts
- EV Grid Management
- Workflow Initiation
- Real-time Dashboards

historical records
phone records
internet traffic
financial data
telephone records
internet traffic

SMS
Console Alerts
EV Grid Management
Workflow Initiation
Real-time Dashboards

Oracle
Condition Based Maintenance (CBM+) relies on sensors to collect information about the vehicle condition. The condition data is combined with other data such as operational use and cost data. Fleet-wide data sets need to be analyzed for trends that lead to increased costs, but analysts do not always know what they are looking for. Big Data can help direct the analysts to the areas of most “bang-for-the-buck”. Utilities can use these to spot trends in the massive time series data from millions of smart meter readings and lower costs or prevent blackouts.
Non-Defense & Security Uses

Use Cases

National Institutes of Health
Disease outbreak detection & tracking

USDA
Food Stamp Fraud detection/tracking

Housing and Urban Development
Development and use of affordable housing
Big Data, Analytics: For Your Domain: What Tools & Platforms are Most Useful to You?

• SCALE: Terabyte, Petabyte, Exabyte Databases
  – Hadoop, Map Reduce, Many Open Source tools
  – NoSQL, Graph, Relational, Semantics, Ontologies,
  – Cyber Security Attacks, Multi-Level Security
  – Versioning, Compression, Archiving
  – Analytics Software as part of Platforms

• HOW do you identify all the technologies you need

• HOW to get ALL those technologies in an COTS Platform
  – Engineered Solution for you – Without Extra Labor Costs!
Hadoop, MapReduce and Related Technologies

Hadoop, HDFS, Hive, MapReduce, HBase
Mahout, Sqoop, Zookeeper, Flume,
Oozie, Pig, Whirr
NoSQL, Key Value Stores

What are these? When and why are they important?
For what Use Cases?

What Problems are YOU trying to solve?

And Oracle really supports all of these Open Source Technologies?
Oracle Integrated Solution Stack for Big Data

ACQUIRE
- HDFS
- Oracle NoSQL Database
- Enterprise Applications

ORGANIZE
- Hadoop (MapReduce)
- Oracle Big Data Connectors
- Oracle Data Integrator

ANALYZE
- Data Warehouse
- In-Database Analytics

DECIDE
- Analytic Applications

InDatabase
Analytic
Applications
Big Data Processing & Query Characteristics

<table>
<thead>
<tr>
<th>Batch-Oriented</th>
<th>Real-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process data to use</td>
<td>Deliver a service</td>
</tr>
<tr>
<td>Bulk storage</td>
<td>Fast access to specific record</td>
</tr>
<tr>
<td>Write once, read all</td>
<td>Read, write, delete update</td>
</tr>
</tbody>
</table>
## Big Data Storage Options: Batch vs. Real-Time

<table>
<thead>
<tr>
<th>Hadoop Distributed File System (HDFS)</th>
<th>NoSQL Database (Oracle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>File System</td>
<td>Database</td>
</tr>
<tr>
<td>Parallel scanning</td>
<td>Indexed storage / Key Value Pairs</td>
</tr>
<tr>
<td>No inherent structure</td>
<td>Simple data structure</td>
</tr>
<tr>
<td>High volume writes</td>
<td>High volume random reads and writes</td>
</tr>
<tr>
<td>Batch Oriented</td>
<td>Real-Time</td>
</tr>
</tbody>
</table>
NoSQL & RDBMS – Used Alongside Each Other

• NoSQL architectures
  – Low value, low density, simple data
  – Very simple relationships
  – Schema-free, unstructured or semi-structured data
  – Distributed storage and processing
  ➢ Lower overhead ($ per operation)

• RDBMS
  – High value, high density, complex data
  – Complex data relationships
  – Schema-centric
  – Designed to scale up & out
  – Lots of general purpose features/functionality
  ➢ High overhead ($ per operation)
Oracle NoSQL Database Enterprise Edition
Scalable, Highly Available, Key-Value Database

Features

- Flexible Key-Value Data Model
- ACID transactions
- Horizontally Scalable
- Highly Available
- Elastic Configuration
- Simple administration
- Intelligent Driver
- Commercial grade software and support

Java SE 6 (JDK 1.6.0 u25)+; Solaris or Linux
Why Have Graph Databases Become Very Popular

Graph Database

- Model data in terms of relationships
- Flexible schema evolves easily by adding new relationships
- Supports querying and discovery by graph patterns and traversal
- Enables graph analytics such as reachability, connectivity, transitivity, same as, proximity, centrality…

- Oracle Has Been Shipping Graph Databases
- For over 10 YEARS

Query:  SELECT ?x ?y
FROM  ...
WHERE {  ?x :partOf ?y }
Buzzwords For Apps using Graph Technology
What terms to look for:

- Semantic Web
- W3C RDF/OWL/SPARQL
- RDF – Resource Description Framework
- Graph Data Management
- Knowledge Discovery
- Knowledge Mining
- Inferencing / Reasoning
- Social Network Analysis (SNA)

- Ontologies
- Taxonomy/Terminology Mgmt
- Property Graphs
- Sentiment Analysis
- Faceted Search
- Text Mining
- NoSQL Database
- Big Data
Using Graphs for Enterprise Applications

- Enterprise metadata framework for enterprise data & public cloud
- Semantic terms link instance data with other resources and apps
- Linked resources enable interoperability between apps
RDF / OWL for Enterprise Integration

Access & Presentation Layer

RDF metadata layer
(integrated graph metadata)

Data Servers

Data Sources / Types

Machine Generated Data
Social Media
Human Sourced Information
Subscription Services
Transaction Systems
Supporting Breadth of Enterprise Data

Semantic Metadata Layer

End-user and Developer Environments

Developers
Data Integration
JDeveloper

Data Scientists
Discovery
Statistics
Mining

Business Users
Business Intelligence
Dashboards

Big Data Sources
Structured Data
Unstructured Data
Social Media

Streaming Services
Event Processing

Data Management
NoSQL
Hadoop
Relational

Data Services
Statistics
Text Analytics
Graph Analytics
Spatial

Data Mining
Natural Lang. Processing
Sound and Video
Images

App Services
Web-log Sessionization and Enrichment
Sentiment Analysis
Reference Architecture

ODBC
JDBC

Vertical Applications
Horizontal Applications

Compression
Security & Encryption

Sound and Video
Images

ORACLE
Oracle: RDF, OWL Semantic Graph

Key Features

- Proven scalability for billions of relationships, terabytes of data
- Standards-based native reasoning
- Supports open source 3rd party & Oracle tools
- Graph querying with SPARQL & SQL languages
- Compression, parallelism, and partitioning to enhance query, load, and inference performance
- DoD strength security: top secret, compartmented

Querying is based on graphs

Ex: Find sub-prime mortgage exposure for “Wells Fargo” bank...
Graph Alignment of Complex Workflows

Data Sources
- Contents Repository
- Databases
- Web resources
- Blogs, Mails, news, RSS feeds

Extract, Transform, Load
- Feature/term Extraction
- Semantic categorization
- Transformation to RDF triple

Extracted Entities & Relationships

Intelligence Ontology

Search, Presentation, Report, Visualization, Query

SQL/SPARQL

National Intelligence Scenario

Person: Abduwali Abdukhadir Muse
- Nationality: Somalian
- Country: UK
- Group: Al-Shabab
- Ideology: Islamist

Person: Chehab Abdouljamid Boulay
- Nationality: Pakistani
- Country: Pakistan
- Group: Al Qaeda
- Currently resides
- Supports

Person: ?
- Nationality: Pakistani
- Country: Pakistan
- Group: ?
- Currently resides
- Member of
- Has

Link ?
## Allied Nation Intelligence Service

**Oracle Spatial and Graph: Social Analysis**

### Objectives
- Profile suspects through telephone, email and social network communications
- Produce “data products” for analysts

### Solution
- RDF Graph modeling of the social network: people, groups and places of interest
- Graph analytics discover relationships among individuals & meaning of pseudonyms, aliases, codes, terminology

### Benefits
- Find & label “same-as” relationships
  - Discover of ~100 million relationships / month
- Tagging for 600 TB / 10b triples graph
- Top-secret, compartmented security for data
- Standards-based tools: W3C RDF & SPARQL
Core Inference Features and Semantic Indexing

• Forward chaining based inference engine in the database
  • Various native rulebases provided
    • E.g., RDFS, OWL 2 RL, SNOMED (subset of OWL 2 EL), SKOS
  • Validation of inferred data
  • User-defined rules, Proof generation
  • Performance enhancements
    • Parallel Inference, Incremental Inference, Compact data structures
    • Optimized owl:sameAs handling

• Semantic document indexing
  – Programmable API to plug-in 3rd party extractors (for any data) into the database.
  – If you have software that “understands” text (NLP), image, audio, or video, plug it in
More functions in Oracle Database Release

- **Inference**
  - Native OWL 2 EL inference support
  - User defined inferencing
    - Allows generation of new RDF resources
    - Temporal reasoning, Spatial reasoning
    - Web service callouts
  - Ladder Based Inference
    - Fine grained security for inference graph
  - Performance optimization for user defined rules
  - Integration with TrOWL*, an external OWL 2 reasoner
    - TrOWL is a transformation based, tractable reasoner for OWL 2

* http://trowl.eu/
Oracle Spatial and Graph: Graph Use Cases

**Data Integration Metadata Layer**
- Unified content metadata for federated resources
- Validate semantic and structural consistency

Apps:
- Metadata warehouse
- Master Data Management
- Taxonomy/vocabulary store
- Media metadata repository

**Text Mining & Entity Analytics**
- Find related content & relations by navigating connected entities
- "Reason" across entities

Apps:
- Documents summary repository
- Documents catalogue

**Social Media Analysis**
- Analyze social relations using curated metadata
- Blogs, wikis, forums, IM, calendars, video, voice

Apps:
- Social graphs
- Recommender systems
- HR skills database

**Network Analysis**
- Graph analysis of road, utility and multimodal networks, drive-time polygons

Apps:
- Trade area management
- Service delivery optimization
- Water, Gas, Electric Utility
## Oracle Spatial and Graph

### Graph Use Cases

<table>
<thead>
<tr>
<th>Intelligence, Law Enforcement</th>
<th>Web and Social Network Solutions</th>
<th>Health Care and Life Sciences</th>
<th>Transportation</th>
<th>Network Management</th>
<th>Financial Services</th>
<th>Media, Content Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Build graphs in social networks for Threat Analysis</td>
<td>Model a graph of relationships for Social Network Analysis (SNA)</td>
<td>Terminology management of EHR systems</td>
<td>Fleet management</td>
<td>Facilities mgt. Assets monitoring &amp; maintenance</td>
<td>Model “same as” relationships for Fraud detection</td>
<td>Query graph patterns in Media Metadata to find media of interest</td>
</tr>
<tr>
<td>Semantically index results of Text Mining</td>
<td>Query graph patterns for Activity Analysis</td>
<td>Semantic concept alignment of clinical, research and operational data</td>
<td>Multimodal transportation planning Routing and Tracking</td>
<td>Outage management</td>
<td>Semantically link data sets for complete results in Compliance Management</td>
<td>Semantically relate metadata to Repurpose Content</td>
</tr>
<tr>
<td>Semantic ally link data sets for Integrated Justice system</td>
<td>Attribute Clusters</td>
<td>Information discovery using inferencing rules</td>
<td>Network planning</td>
<td>Coverage area analysis</td>
<td>Link relevant metadata for Risk Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Oracle Spatial and Graph - Types

- "Points"
- "Lines"
- "Polygons"

Web Services (OGC)
- SPARQL End Point

Rasters

Network Graphs

Oracle Spatial and Graph

Geocoding
Routing
Inferencing

3D

Topologies

RDF Semantic Graphs
Ontology-driven Geospatial Applications - Connect Actionable Knowledge

- Simple Features
- GeoRaster
- Topology
- Networks
- Gazetteers

- Data Integration
- National Map schemas
- Geographic names
- Temporal
- Naïve Geography

Application Ontologies

RDF & OWL Metadata

Environmental Monitoring

Famine Relief

Disaster Response
It’s Not Just about “Linking”

- Integrate domains of knowledge through common vocabularies (ie SKOS)
- Manage relationships between collections of images and associated metadata
- RDF as flexible and extensible data model supports powerful search and end-user discovery of related content
- Rich platform for data integration, data repurposing, and better quality control and classification
Automatic Data Optimization
ILM: Hot/Cold Data Classification

Enhanced Insight into Data Usage: “heat map”

Recently inserted, actively updated

Infrequently updated, Frequently Queried

Retained for long term analytics and compliance with corporate policies and regulations

- Block and Segment level statistics on last Read and last Update
ALTER TABLE orders
ILM ADD CompressionPolicy
COMPRESS Partitions for Query
AFTER 90 days from creation;

ALTER TABLE sales
ILM ADD MovePolicy
TIER Partitions TO ‘Archive_TBS’
ON OrdersClosedPolicy;
Data Compression
Reduce storage footprint, read compressed data faster

- **Hot Data**
  - Advanced Row Compression: 3X
  - Columnar Query Compression: 10X
  - Columnar Archive Compression: 15X
Connecting: CYBERSECURITY is Major Challenge

Information Security and Privacy

Oracle Database

Monitoring
- Configuration Management
- Audit Vault
- Total Recall

Access Control
- Database Vault
- Label Security

Encryption & Masking
- Advanced Security
- Secure Backup
- Data Masking
Oracle Database Security Solutions
Comprehensive Security Platform

PREVENTIVE
- Encryption
- Redaction and Masking
- Privileged User Controls

DETECTIVE
- Database Firewall and Network Monitoring
- Audit Consolidation
- Reporting and Alerting

ADMINISTRATIVE
- Privilege Analysis
- Sensitive Data Discovery
- Configuration Scanning

*Oracle*
Redacting Data Challenges

- Secure sensitive personal information
- Redact data in applications, queries and reports
- Avoid changing applications, queries and reports
Redacting Sensitive Data
Mask Application Data Dynamically

Policy enforced redaction of sensitive data

Call Center Operator

Payroll Processing

<table>
<thead>
<tr>
<th>Soc. Sec. #</th>
<th>DOB</th>
<th>PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>115-69-3428</td>
<td>11/06/71</td>
<td>5623</td>
</tr>
</tbody>
</table>

Redacting Sensitive Data
Mask Application Data Dynamically

<table>
<thead>
<tr>
<th>CLIENT CONTACT</th>
<th>SOC. SEC. NO.</th>
<th>DOB</th>
<th>PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARA JONES</td>
<td>*<strong>-</strong>-3428</td>
<td>01/01/01</td>
<td>****</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CLIENT CONTACT</th>
<th>SOC. SEC. NO.</th>
<th>DOB</th>
<th>PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAMES SMITH</td>
<td>*<strong>-</strong>-3256</td>
<td>01/01/01</td>
<td>****</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMPLOYEE NAME</th>
<th>SOC. SEC.#</th>
<th>DOB</th>
<th>PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARA JONES</td>
<td>115-69-3428</td>
<td>11/06/71</td>
<td>5623</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMPLOYEE NAME</th>
<th>SOC. SEC.#</th>
<th>DOB</th>
<th>PIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAMES SMITH</td>
<td>114-59-3256</td>
<td>01/12/68</td>
<td>3763</td>
</tr>
</tbody>
</table>
## Connecting: Tools to Find Connections: Discovery & Predictive Analysis

<table>
<thead>
<tr>
<th><strong>Problem Classification</strong></th>
<th><strong>Sample Problem</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anomaly Detection</strong></td>
<td>Given demographic data about a set of customers, identify customer purchasing behavior that is significantly different from the norm (Fraud?) . For Sensors, Events – problem? Good News?</td>
</tr>
<tr>
<td><strong>Association Rules</strong></td>
<td>Find the items that tend to be purchased together and specify their relationship – market basket analysis</td>
</tr>
<tr>
<td><strong>Clustering</strong></td>
<td>Segment demographic data into clusters and rank the probability that an individual will belong to a given cluster. For customers / govt constituents, what do they want to know about?</td>
</tr>
<tr>
<td><strong>Feature Extraction</strong></td>
<td>Given demographic data about a set of customers, group the attributes into general characteristics of the customers</td>
</tr>
</tbody>
</table>
Analysis with R

R workspace console

Oracle Database 11g statistics engine

OBIEE, Web Services

Function push-down
– data transformation & statistics

Transparencyly leverage Hadoop for High Performance Analytics to Oracle Big Data Appliance (Oracle R Connector for Hadoop)
Predictive Analytics with R, BI, Data Mining

- Function push-down – data transformation & statistics
- No changes to the user experience
- Scale to large data sets
- Embed in operational systems

R workspace console

Oracle statistics engine

OBIEE, Web Services

Development → Production → Consumption
Information Discovery and Hadoop

Complementary Big Data Capabilities

- Massive scalability
- Batch execution
- Technical experts develop analyses

- Data Variety - structured and unstructured data
- No fixed pre-defined model required

- In-memory analytics
- Interactive query response times
- Targets business users

Oracle Endeca
Information Discovery

Deep Large-Scale Processing

Populates

Fast Self-service
Information Discovery

Informs
Big Data in Healthcare

Find relationship between gene to cancer interaction

- Cross-referenced the relationships between 17000 genes and five major cancer types across 20 million medical publication abstracts.

- Cross-referenced genes from 60 Million patients and miRNA for a simulated 900 Million population.

- Understanding additional layers of the pathways these genes operate in and the drugs that target them is expected to help researchers in their work.
Oracle In-Database Analytics Platform

Parallel Processing Engine

Oracle R Enterprise
Oracle Data Mining
Spatial Analytics
Text and Search
RDF Graph Analytics
SQL Analytics

Data Layer
XML
Relational
OLAP
Spatial
RDF
Media
Big Data Apparent Disconnect
Two Separate Developer Worlds

Unstructured Data
- Hadoop, HDFS, Hive, MapReduce, HBase
- Mahout, Sqoop, Zookeeper, Flume, Oozie, Pig, Whirr

Structured Data
- DBMS (OLTP)
- Extract Transform and Load (ETL)

Big Data Environments
- MapReduce Solutions
  - Distributed File Systems
  - Transaction (Key-Value Stores)

Acquire | Organize | Analyze

Traditional Environments
- DBMS (OLTP)
- DBMS (DW)
- Advanced Analytics

Acquire | Organize | Analyze

NoSQL
- Flexible
- Specialized
- Developer Centric

SQL
- Trusted
- Secure
- Administered
Oracle’s Approach To Big Data

Provide all Technologies, Engineered Together

Oracle Exadata
Database Machine

Oracle Big Data Appliance

Unstructured Data

Structured Data

Oracle

MapReduce Solutions

Distributed File Systems

NoSQL
Flexible Specialized Developer Centric

SQL
Trusted Secure Administered

Hadoop, HDFS, Hive, MapReduce, HBase
Mahout, Sqoop, Zookeeper, Flume, Oozie, Pig, Whirr

Extract, Transform, and Load (ETL)

Trusted Administered

Enhanced Transform and Load (ETL)

NoSQL

Flexible Specialized Developer Centric

DBMS (OLTP)

Oracle Exalytics
In-Memory Machine

DBMS (DW)

Acquire

Acquire

Analyze

Analyze

Organize

Organize

Oracle’s Approach To Big Data

Provide all Technologies, Engineered Together

Oracle Exadata
Database Machine

Oracle Big Data Appliance

Unstructured Data

Structured Data

Oracle

MapReduce Solutions

Distributed File Systems

NoSQL
Flexible Specialized Developer Centric

SQL
Trusted Secure Administered

Hadoop, HDFS, Hive, MapReduce, HBase
Mahout, Sqoop, Zookeeper, Flume, Oozie, Pig, Whirr

Extract, Transform, and Load (ETL)

Trusted Administered

Enhanced Transform and Load (ETL)

NoSQL

Flexible Specialized Developer Centric

SQL
Trusted Secure Administered
Big Data and Cloud Computing
Next Generation Enterprise Data Platform

- Big Data
  - Drive Business Value
  - Deliver Top Line Growth

- Private Cloud
  - Drive Business Efficiency and Agility
  - Deliver Bottom Line Savings

- Secure

CIO

Business Value
Public Clouds and Private Clouds

- Used by multiple tenants on a shared basis
- Hosted and managed by cloud service provider

- Exclusively used by a single organization
- Controlled and managed by in-house IT

Trade-offs

- Lower _upfront_ costs ↔ Lower _total_ costs
- Outsourced management ↔ Greater control over security, compliance, QoS
- OpEx ↔ CapEx & OpEx

Oracle Technology Supplies both Public and Private clouds
Oracle Complete Big Data Platform

**Big Data Appliance**
- Cloudera’s Distribution including Apache Hadoop
- Oracle NoSQL Database
- Open Source R
- Applications

**Exadata**
- Oracle Big Data Connectors
- Oracle Advanced Analytics
- Data Warehouse
- Oracle Database
- In-Database Analytics

**Exalytics**
- Analytic Applications
- Alerts, Dashboards, MD-Analysis, Reports, Query
- Web Services
- BI Abstraction

**Steps**
- Acquire
- Organize
- Analyze
- Decide
Why Build A Hadoop Appliance?

Time to Build
Optimizations
Maintenance
Required Skills for MapReduce Development

- Java
- Hadoop Framework
- Parallel Algorithms

& Manage Large Clusters of Machines
Big Data: Batch-Oriented Processing using Hadoop: Map – Reduce – *Simple* Example
HW/SW Efficiencies: But Labor Costs Growing - Complete, Engineered System Needed
Oracle Big Data Appliance

• Optimized and Complete
  – All you need to store and integrate big data

• Integrated with Oracle Exadata
  – Analyze all your data

• Easy to Deploy
  – Risk Free, Quick Installation and Setup

• Single Vendor Support
  – Full Oracle support for all hardware and software
Oracle Exadata Database Machine

Data Warehousing, Transaction Processing, Consolidation

- **Fastest** Data Warehouse & OLTP
- **Best Cost/Performance** Data Warehouse & OLTP
- Optimized Hardware (per rack)
  - Processor: up to 128 Intel Cores and 2 TB DRAM
  - Network: 880 GB/Sec Throughput
  - Storage: 5 TB Flash and up to 336 TB Disk
- Software Breakthroughs
  - Exadata Smart Storage Grid
  - Smart Flash Cache
  - Hybrid Columnar Compression
  - Parallel Scale-Out Database and Storage
- Scales from ¼ Rack to 8 Full Racks
Oracle Exalytics In-Memory Machine

First engineered system for analytics

Visual Analysis without limits

Smarter analytic applications
Oracle Complete Big Data Platform

**Big Data Appliance**
- Cloudera’s Distribution including Apache Hadoop
- Oracle NoSQL Database
- Open Source R
- Applications

**Exadata**
- Oracle Big Data Connectors
- Oracle Advanced Analytics
- Data Warehouse
- Oracle Database
- In-Database Analytics

**Exalytics**
- Analytic Applications
- Alerts, Dashboards, MD-Analysis, Reports, Query
- Web Services
- BI Abstraction

**ACQUIRE**
**ORGANIZE**
**ANALYZE**
**DECIDE**