Big Data: Next Frontier for Innovation

Selim Burduroğlu
Global Innovation Evangelist & Architect
Education & Research Industries
03.26.14

@theselimb
Big Things!
“The implementation of a **state-of-the-art Data Analytics** program is the foundational platform required for the three Discovery Themes to have a transformative impact on the advancement of science and technology research, education, and outreach across the university.”

—**David Cole**, Professor and Ohio Research Scholar, School of Earth Sciences, College of Arts and Sciences
“The notion of “Big Data” has created a stir and promises to revolutionize all walks of human endeavor ranging from the financial industry to the health-care industry, from the social sciences (e.g., study of personalized and social interactions) to engineering (e.g., analyzing sensory data from the manufacturing process), and from scientific discovery to cyber-security.”

—David Tomasko, Professor of Chemical and Biomolecular Engineering, Associate Dean for Undergraduate Education and Student Services, College of Engineering
“Data Analytics includes the set of skills necessary to identify how to manage, represent and manipulate large and complex data stores, how to abstract, model, and effectively analyze such data to facilitate the identification and evaluation of appropriate hypothesis and actionable patterns, and how to create quantitative visual cues to help explain and take informed decisions from those data.”

—David Tomasko, Professor of Chemical and Biomolecular Engineering, Associate Dean for Undergraduate Education and Student Services, College of Engineering
Big Data Defined
hype
Data Velocity

Real Time
Near Real Time
Periodic
Batch
Reports
Table
Database
KB
MB
GB
TB
PB

Data Volume

Data Variety

Marketing Automation
CMS
Audio
Video
Web
Social Tech
SMS
Automated Demand Generation
Photo
App
Mobile
Web
Periodic
Batch
Reports
Table
Database
KB
MB
GB
TB
PB

Source - http://jeffhurtblog.com
understanding the data deluge: comparison of scale with physical objects

1 megabyte
(A large novel)

1 gigabyte
(Information in the human genome)

1 terabyte
(Annual world literature production)

1 petabyte
(All US academic research libraries)

1 exabyte
(Two thirds of annual production of information)

A tiny ant
Height of a short person
Length of the Auckland Harbour Bridge
Length of New Zealand
Diameter of the Sun

Source: Julian Carver
Where does Big Data come from?
#Genlike Generation
College Scorecards in the U.S. Department of Education’s College Affordability and Transparency Center make it easier for you to search for a college that is a good fit for you. You can use the College Scorecard to find out more about a college’s affordability and value so you can make more informed decisions about which college to attend.

To start, enter the name of a college of interest to you or select factors that are important in your college search. You can find scorecards for colleges based on factors such as programs or majors offered, location, and enrollment size.

[Search for a college by name...](http://www.whitehouse.gov/issues/education/higher-education/college-score-card)
Massive
Open
Online
Course
The Rise of the MOOCs
90% of the data in the world was created in the last two years

235 terabytes data collected by the US Library of Congress by April 2011

Walmart handles more than 1 million customer transactions every hour

Source: McKinsey

Source: The Economist
Google announces Android software for wearable tech

Launching a new software development kit, Google is preparing for a new market of smartwatches and other wearable technologies - starting with a Motorola ‘moto 360’ watch in summer 2014
50% of campus leaders agree that President Obama's plan to collect and report data on college outcomes is appropriate - but just 13 percent are confident that the data will be accurately collected and reported.
The Internet of Things
Machine to Machine Era
Infinite Possibilities – By 2020

8 BILLION
People in the WORLD

200 BILLION
CONNECTED Devices
BigData

Enterprise Data

Transactions

Public Data

Social Media

Sensor Data
History of Big Data
Records Department of FBI
Implementing analytics and applying it to make **data driven decisions** is a major differentiator between high performing and low performing organizations.

*Big Data: The Next Frontier for Innovation, Competition and Productivity*
What Makes it Big Data?

VOLUME

VELOCITY

VARIETY

VALUE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.
BIG DATA

“No other book offers such an accessible and balanced tour of the many benefits and downsides of our continuing infatuation with data.”

WALL STREET JOURNAL

Viktor Mayer-Schönberger
and
Kenneth Cukier
COLUMBUS, Ohio – A surgeon at The Ohio State University Wexner Medical Center is the first in the United States to consult with a distant colleague using live, point-of-view video from the operating room via Google Glass, a head-mounted computer and camera device.

“It’s a privilege to be a part of this project as we explore how this exciting new technology might be incorporated into the everyday care of our patients,” said Dr. Christopher Kaeding, the physician who performed the surgery and director of sports medicine at Ohio State. “To be honest, once we got into the surgery, I often forgot the device was there. It just seemed very intuitive and fit seamlessly.”

Google Glass has a frame similar to traditional glasses, but instead of lenses, there is a small glass block that sits above the right eye. On that glass is a computer screen that, with a simple voice command, allows users to pull up information as they would on any other computer. Attached to the front of the device is a camera that offers a point-of-view image and the ability to take both photos and videos while the device is worn.
CHALLENGES

- Stale Data
- Islands of Data
- Drowning in Spreadsheets
- Lack of Data Skill
- Gather and Organize All Data
Executives who feel they understand the impact data will have on their organizations
Oracle’s Strategy for Big Data

Produce
Data

Use
Data

How do you shrink this gap?

- Capture massive data volumes
- Analyze all data
- Secure and unify the data platform

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.
Big Data Connectors and Data Integrator
Technically, **Hadoop** consists of two key services: Reliable data storage using the Hadoop Distributed File System (HDFS) and high-performance parallel data processing using a technique called **MapReduce**.
**Hadoop** is a large-scale distributed batch processing infrastructure. While it can be used on a single machine, its true power lies in its ability to scale to hundreds or thousands of computers, each with several processor cores. **Hadoop** is also designed to efficiently distribute large amounts of work across a set of machines.
Oracle’s Strategy for Big Data

Produce
Data

Use
Data

How do you shrink this gap?

✔ Capture massive data volumes

Analyze all data

Secure and unify the data platform
Analyze All Your Data In-Place

Big Data Appliance + Hadoop / NoSQL

Advanced Analytics

Exadata + Oracle Database
Expose All Data to End Users

Big Data Appliance + Hadoop / NoSQL

Endeca

OBI EE

Exadata + Oracle Database
Oracle’s Strategy for Big Data

Produce Data

Use Data

How do you shrink this gap?

- Capture massive data volumes
- Analyze all data
- Secure and unify the data platform
Large, Research Intensive University

Big Data Appliance
+ Hadoop

Data Reservoir

Exadata
+ Oracle Database
Securing Big Data

- Increasingly, Big Data solutions are capturing sensitive information must be protected and audited
- This is no different than critical data stored in an RDBMS
Enhanced Big Data Security

Authenticate
users with secure Kerberos protocol

Authorize
access to data with fine grained controls

Audit
activity and access with Oracle Audit Vault and Database Firewall

Encrypt
data as it flows thru the system
Oracle Strategy for Big Data

Produce Data

How do you shrink this gap?

- Capture massive data volumes
- Analyze all data
- Secure and unify the data platform

Use Data
Advanced Query & Analysis
Full Power of SQL and Advanced Analytics

Transparent to Applications
No Changes to Application Code

Single View of All Data
Unified Metadata Across RDBMS & Hadoop

Fastest Performance
Utilize SQL Processing Across the Platform

Leverage Existing Skills
Lower Cost & Complexity of Big Data Adoption
Skills Gap
190,000
Projected Shortage
In Data Scientists

"Data Scientist….the sexiest job
in the 21st century"
Big Data...
It's unbelievable how much you don't know about the game you've been playing all your life.

-Mickey Mantle
“Every day I wake up and ask, How can I flow data better, manage data better, analyze data better?”

Rollin Ford, CIO, WalMart
Source: The Economist, “Data, Data Everywhere”
Analytics 3.0 – The Era of Impact
Tom Davenport – IIA

1.0 Traditional Analytics
- Primarily descriptive analytics and reporting
- Internally sourced, relatively small, structured data
- "Back room" teams of analysts
- Internal decision support

2.0 Big Data
- Complex, large, unstructured data sources
- New analytical and computational capabilities
- "Data Scientists" emerge
- Online firms create database-based products and services

3.0 Rapid Insights Providing Business Impact
- Analytics integral to running the business; considered strategic competitive asset
- Rapid and agile insight delivery
- Analytical tools available at point of decision
- Cultural evolution embeds analytics into decision and operational processes
- All businesses can create database-based products and services

http://iianalytics.com/a3/
“... In the coming years, 75 to 80 percent of new cloud applications will be data-intensive, and data is becoming much more strategic than applications.”

Frank Gens – International Data Corporation (IDC) Analyst
Analytics 3.0 Reference Architecture
Oracle Integrated Solution for Big Data

ACQUIRE

ORGANIZE

ANALYZE

DECIDE
### Oracle Big Data Architecture Capabilities

<table>
<thead>
<tr>
<th>Data</th>
<th>Acquire</th>
<th>Organize</th>
<th>Analyze</th>
<th>Decide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master &amp; Reference</td>
<td>DBMS (OLTP)</td>
<td>ETL/ELT</td>
<td>ODS</td>
<td>Scorecards Dashboards Query Reporting</td>
</tr>
<tr>
<td>Transactions</td>
<td>Files</td>
<td>ChangeDC</td>
<td>Warehouse</td>
<td>Real-Time Decisions</td>
</tr>
<tr>
<td>Machine Generated</td>
<td>NoSQL</td>
<td>Real-Time</td>
<td>Graph Model</td>
<td>EPM, BI, Social Applications</td>
</tr>
<tr>
<td>Social Media</td>
<td>HDFS</td>
<td>Streaming Event Processing</td>
<td>In-Database Analytics</td>
<td>Text Analytics and Search</td>
</tr>
<tr>
<td>Text, Image Video, Audio</td>
<td>HDFS</td>
<td>Hadoop (MapReduce)</td>
<td>In Memory Analytics</td>
<td>Advanced Analytics</td>
</tr>
</tbody>
</table>

**Unstructured**
- NoSQL
- HDFS

**Structured**
- DBMS (OLTP)

**Semi-structured**
- Files

**Specialized Hardware**
- Big Data Cluster
- High Speed Network
- RDBMS Cluster
- In Memory Analytics

**Management Security, Governance**
- Oracle Big Data Architecture Capabilities

---

*Copyright © 2013, Oracle and/or its affiliates. All rights reserved.*
Why Now?

Big – Fast – Smart – Hyper Data

• Demand for Accountability and Transparency
  • Fueled by Funding Shortages
  • Increased Consumer Like Services

• To be Competitive in Global Knowledge Economy
  • 7 Billion People in the World only 317 Millions are Americans
  • To Lead – Everyone needs to be educated and equipped with the skills of the future

• Innovates use of disruptive technologies such as Cloud, Social, Analytics, and Mobile...can make Education
  • More Accessible, Affordable, and Relevant
21\textsuperscript{st} Century Learning
Essential Ingredients

✓ Proficiency
✓ Personalized
✓ Data-Driven
✓ Science of Learning
✓ Exploit Technology to Enhance Learning Experience
Big Idea About Big Data for Education & Research

Why?

• Increase the **quality** of **decisions** made by:
  • Students
  • Faculty
  • Advisors/Counselors
  • Chief Business Officers
  • Colleges
  • Principal Investigators (PIs)
Big Data in Education...

The Future is Now!

- Help predict student success
- Improve graduation rates and student retention
- Determine what learner does know and does not know
- Monitor student’s behavior and level of engagement
- Trigger alerts when the learner is getting off track, bored, or frustrated
- Increase engagement with game mechanics
- And more…
Closed-Loop Student Success & Experience

1. PREDICT
2. IDENTIFY
3. ENGAGE
4. INTERVENED
5. SUCCEED
6. ENHANCE
7. GROW
8. ACQUIRE

EMPOWER

ACQUIRE Knowledge & Skills

GROW Professional & Personal

INTERVENE

ANALYZE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.
When we understand the differences in our student needs, and drive our services based on that fact based decision making, we will give them the soil they need to flourish.

Anonymous
Hardware and Software

Engineered to Work Together
Oracle Big Data Appliance
Cost Effective High Performance Hadoop Appliance

- Fastest, Pre-Configured Hadoop Appliance
  - Best-in-Class Pre-Integrated Software
- Superfast connectivity to Exadata: 15TB/Hour
  - Partitioned & Non-partitioned; Online & Offline Loads
- Pre-integrated Software
  - Linux, Java VM, Cloudera Distribution of Hadoop
  - Oracle R Distribution & NoSQL Database
- Start Small, Add Servers, Add Storage
Oracle Exadata
Cost Effective Fastest Database Machine

- **Speed Up Query Performance 10-100X**
  - Smart-Scan eliminates IO Bottlenecks
  - Flash Cache enable In-Memory Data Access
  - Parallel Query provides Scale out Processing

- **Fast, Cheap Storage**
  - Columnar Compression provides 10X Savings
  - Capacity of Disk; IOs of Flash; Speed of D-RAM

- **Start Small, Add Servers, Add Storage**
Oracle In-Memory DBMS

Fastest Query Performance

>100X Faster

- Scales Up or Scales out for very large data sets
- Scans use super fast SIMD vector instructions
- **Billions of rows/sec** scan rate per CPU core
- Joins up to **10X Faster**
Oracle DBMS SQL & R
Analyze Data across all your Systems

- Expand the data and users who can analyze on Hadoop
- Use Full Power of Oracle SQL and R on all data
- Stream Hadoop Resident data through Fast Big Data Connectors
National Cancer Institute
Identifying Relationship between Gene to Cancer Interaction

17,000 Genes

5 Major Cancer Types

60M Patients

20M Medical Publications
Turkcell
Location-based Mobile Billboard on Oracle Event Processing

1.5B Events Per Day

800K Events per Second

50 Simultaneous Campaigns