DATA DISRUPTION
AND INNOVATION

Closing Keynote
DBTA Data Summit – May 13, 2015

John O’Brien | Radiant Advisors
@obrienjw  john.obrien@radiantadvisors.com
The world around us has changed to consumer-centric and data-centric economy where companies must transform to remain competitive …and survive.
Data Disruption and Innovation

3 STEPS TO FOLLOW

Business Disruption  Capability by Design  Platform of Speed
**Business Disruption**

**THE POWER OF DISRUPTION**

- Understand digital life
- Understand new consumers
- Experiment and move fast
- Think like a startup
- Competitors are moving
- Innovate then optimize
- Practice Creative Destruction
Business Disruption

FAMILIAR STORIES?

- Amazon.com
- Best Buy
- Blockbuster
- Barnes & Noble
- Circuit City
- Netflix
- Borders
- Starbucks
- HBO Now
Business Disruption
DIFFERENT MINDSET NEEDED

“Advantages”
• Ready to compete on analytic ability
• Deeper understanding of your customers
• Deeper understanding of your products usage

“Data Centric”
• Treat data as an asset – Treat data as oxygen
• Willing to give services for free
• Willing to buy data
Business Disruption

CUSTOMER LIFETIME VALUE
Business Disruption

BRAND-TO-CUSTOMER
Business Disruption

PRODUCT RELATIONSHIPS
Business Disruption

BRAND-TO-CUSTOMER FOCUS

“Brand Analytics”
• Sentiment analysis
• Partnerships
• Influence

“Customer Analytics”
• Identify potential valuable customers
• Nurture to customers
• Engage lifelong customers

Brand Loyalty Becomes Customer Loyalty
Business Disruption

CONNECTED CUSTOMER

“Mobile”
- Individually centered – Me, Here, Now
- Generation “C” Connected
- Opportunity for more data

“Engaged”
- Engaging experiences
- Throughout their day
- Platform invitation
Business Disruption
ALWAYS ON THE GO
Business Disruption

OPPORTUNITY TO INFLUENCE

RT Gr8 #Breakfast at @FavCasino #GoBuffet #Reno

Gr8 #datasummit #keynote w/ @obrienjw #GoBigData #Empowered
Business Disruption

EXPERIENCES CAPTURED
Business Disruption
BE A PART OF SHARING
Business Disruption
WHAT’S NEXT?

• What would it take for you to have wearable technology?
Data Disruption and Innovation
CAPABILITY BY DESIGN

- Discovery is iterative
- Experiment / Fail Fast
- Exploratory; can be unclear
- New paths; new ideas
- Driven by unknowns
- Foundation for discovery
Capability by Design

FOUR ELEMENTS OF DESIGN

Self-Sufficiency

Collaboration

Visualization

Mobile
Capability by Design
DESIGNED FOR SELF-SERVICE

- Self-sufficiency
- From access to ability
- Partners in discovery
- Greater autonomy
- Remove barriers
Capability by Design
DESIGNED FOR COLLABORATION

• Investment strategy
• Fosters collective innovation
• Closes the business/IT gap
• Establishes transparency
• Data governance and security
Capability by Design

DESIGNED FOR VISUALIZATION

- Cognitive hardwiring
- Makes analytics approachable
- Understanding to insight
- Curated images and color
- Visual discovery
Capability by Design

DESIGNED FOR MOBILITY

- More than consumption
- Visually interact with data
- Portable discovery
- Personalized experience
BI Casual Users - Consumers

- Represent 80% of BI Users
- Reporting and Monitoring generations of BI

Power Users and Business Analysts – Data Provisioners

- Represent 20% of BI Users
- Analysis and Prediction generations of BI

Data Scientists and Analytic Modelers

- A few professionals in the organization
- Creators of statistical models
Capability by Design
ITERATIVE “FRICIONLESS” DISCOVERY

“Get Data Access”

“Collaborate”

“Explore Data”

“Agilely Blend Data”

“Discover”

99x for each Discovery

“Get Data Access”

“I need to work with new data”
“I want to check something”

“What do you think of this?”
“How can we use this?”

“Is this data usable?”
“What is this data?”

“Is this what I’m looking for?”
“This is interesting…”

“That helped, need more.”
“Nope, keep looking…”

“Nope, keep looking…”

“That helped, need more.”

“Is this what I’m looking for?”
“This is interesting…”

“How can we use this?”

“Let’s mix with other data”
“Does this help my data set?”

“Let’s mix with other data”

“Does this help my data set?”
Capability by Design
“FRICITIONLESS” DATA INGESTION

1 hour steps
x 99
= 495 hours
= 62 days

1 minute steps
x 99
= 495 minutes
= 8.25 hours
1. COLLECT DATA FROM EVERYWHERE

Start with business goal, then imagine the data needed

Get external public and purchased data – Big Data
• Facebook, Pintrest, Twitter, Weather, demographics

Create new ways to collect the data you want
• Mobile apps, wearables, sensors, “Internet of Things”

Current systems are the feedback loop
2. ANALYTICS CULTURE

Capability by Design

Customers, Their Network
Their Influence

Interactions
Relationships
Usage

Business
Functions &
Processes

New and
Disruptive
Business

Customer Insights

Product Insights

Optimize

Innovate
3. ACT ON ANALYTICS

- Ongoing A/B experiments
- Continuously improving analytic models
- Tracking prediction and recommendation engines
- Lean product development and interaction
Design for Capability

GOVERNED DATA DISCOVERY

Governing People, Roles, Responsibilities
Define roles and responsibilities for accessing and working with data
Define ownership and delegation for semantic context discovered

Governing Discovery Results
Governing semantic context after discovery
Operationalizing and monitoring analytic models

Modern Data Platforms
Managing data abstraction and mixed workloads
Defining discovery or analytic sandboxes as dev, prod or both
Data Disruption and Innovation

PLATFORM OF SPEED

• Built to Change
• Polyglot Persistence approach
• Data unification is key
• Platform oriented and not a bunch of database silos
• Requires data governance
• Enabling platform for analytics and discovery
Platform for Speed

POLYGLOT PERSISTENCE PRINCIPLE

Polyglot Persistence Principle

Transactions  Analytics  Behavioral  Sensors

RDBMS  MPP  Doc  Hadoop  Graph

Optimized Columnar  In-Memory  MPP  OLAP

Relational SQL  NoSQL

Specialized Key-Value  Document  Distributed  Graph

UX  UX  UX

UX  UX  UX

UX  UX  UX

UX  UX  UX

UX  UX  UX
Platform for Speed

DATA TECHNOLOGIES

Key Value Store (Hadoop)
Discovery Oriented

Highest Scalability
Lowest Cost
Schema-less
Without Context

Analytic Database Technologies

EDW RDBMS

Accessibility:
Programming -> SQL

Workload:
Flexible, Scalable

Maturity:
Emerging

SQL, MDX, UDF
Analytic Optimized
Accepted

SQL Access
Reference Data Mgmt
Mature
## Platform for Speed

### MODERN DATA PLATFORM

<table>
<thead>
<tr>
<th>Flexible Class</th>
<th>Analytic Optimized Class</th>
<th>Reference Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hortonworks</td>
<td>Highly Optimized for BI</td>
<td>Data Warehouses</td>
</tr>
<tr>
<td>Cloudera</td>
<td>Highly Specialized for BI</td>
<td>Master Reference Data</td>
</tr>
<tr>
<td>MapR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassandra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MongoDB</td>
<td></td>
<td>Stable, Context, SQL</td>
</tr>
</tbody>
</table>

- Discovery, Scalable, Programmable
- Discovery and Analytics Oriented
- Stable, Context, SQL
Platform for Speed
ARCHITECTED FOR DISCOVERY

Modern Data Platform

Data Acquisition
- Data Unification (SQL/Data Services)
- STG EDW
- Analytics Optimized Tier
- Analytics DB
- MPP Col
- Replication Technology
- Governed Acquisitions

Operational Reporting Tier
- STG EDW
- Analytics DB
- SSE
- Tableau Serv
- Tableau Mobile
- Tableau

Data Unification (SQL/Data Services)
- BI Tool
- SAS
- Excel
- Business Reporting
- Business Discovery
- Executive
- Analysts
- Data Science
- Social Response

Internal Operational Systems
- Users
- Report User
- User data

External Data
- Managed File Transfers
- Partner
- 3rd Party
- Website Activity
- Mobile Activity
- IoT Activity
- Social Media Aggregator

Behaviorsal Data
- Social Networks
- Activity

Proxy Server

Governed Acquisitions
- Proxy Server

User data

User data

User data

User data

User data
Platform for Speed

KEY SQL CONSIDERATIONS

**Scalability**
- How many nodes max?
- All nodes in cluster?
- Subset of cluster?
- Data duplication?

**Speed**
- Response time
- Ad-hoc workloads
- Without caching
- Concurrency

**SQL Capability**
- Tools Compatibility
- ANSI SQL
- Analytic SQL
- User Defined Functions

**Architecture**
- YARN compatible
- Data file formats
- Data Lake strategy
- Semantic Layer

Evaluation Criteria
Platform for Speed

**EVOLVING ARCHITECTURE FOR SQL**

**Hadoop v1**
- Hive-QL
- PIG
- MapReduce
- Hadoop HDFS

Batch-oriented SQL

**Hadoop v2**
- Hive 0.13
- Tez
- PIG
- MapReduce
- YARN
- Hadoop HDFS

Interactive SQL

**Hadoop v2 with more SQL options**
- Hive
- PIG
- MapReduce
- YARN
- YARN
- Hadoop HDFS

Architectural SQL
Platform for Speed

MODERN DATA PLATFORM UNIFIED SQL

Extending SQL Access to Big Data and Hadoop via Hive and other HDFS SQL engines

Highly Optimized for Analytics
- MPP
- Columnar
- In-memory
- MOLAP

Highly Specialized for Analytics
- Document Stores
- Graphs
- Text Analytics

Enterprise Data Warehouses
Master Reference Data

Operational Systems, Big Data, Streams
- Apache Hadoop
- Map Reduce
- Hive SQL
- Generate

Flexibility Class
- Discovery, Scalable, Programs

Optimized Class
- Discovery & Analytics Oriented

Reference Class
- Stable, Context, SQL
<table>
<thead>
<tr>
<th>Business Demands</th>
<th>Hadoop Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet and e-Commerce companies needed to cost-effectively work with big data (both operationally and analytically)</td>
<td>Hadoop v1 inspired by Google papers and from Apache projects delivered affordable scalability and a parallel programming framework</td>
</tr>
<tr>
<td>Business needed a more interactive environment for users and apps with improved resource management and multi-tenancy</td>
<td>Hadoop v2 improved the two-tier architecture with YARN for managing resources and reorganized for YARN applications</td>
</tr>
</tbody>
</table>
Platform for Speed

HADOOP EVOLVES FOR BUSINESS

<table>
<thead>
<tr>
<th>Business Demand</th>
<th>Hadoop Generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiently integrating big data, BI and operational data apps in a singular data architecture meeting data management requirements</td>
<td>Hadoop v2x focus on improvements in operational capabilities: SQL, performance, security, governance while Data Lake provides a strategy</td>
</tr>
<tr>
<td>Long term “data operating system” architecture for efficiently and effectively manage data workloads, reusability, and leveraging cloud</td>
<td>Future Hadoop will focus on IT ecosystem integration, more Hadoop applications with improved distributed file systems</td>
</tr>
</tbody>
</table>
Data Disruption and Innovation

3 STEPS TO FOLLOW

Business Disruption  Capability by Design  Platform of Speed
For more information

www.RadiantAdvisors.com

Twitter: @RadiantAdvisors

Email: info@RadiantAdvisors.com

LinkedIn: www.linkedin.com/company/radiant-advisors

THANK YOU!