Aligning Your Strategic Initiatives with a Realistic Big Data Analytics Roadmap

3 key strategic advantages, and a realistic roadmap for what you really need, and when
Topics to be discussed

1. Effective exploration of your data to identify your next Analytics opportunity for innovation, and competitive advantage?

2. Key tools and partners to align your strategic goals to big data KPI's & metrics?

3. Leverage successful patterns for big data insight and avoid re-invention of technology

4. Optimized HW/SW Platforms to reduce time, cost, risk, and waste

5. Use Cloud, SAAS, and external data platforms to enrich and extend my on-premises IT Investments

The Answer?
Why is Big Data Multiplying the Effectiveness of the BI / Analytics Investment?

**VOLUME**

Desire to run analytics on entire data sets, not sample sizes—better accuracy

—“By 2014, 85% of the data warehouses currently deployed will fail to scale to meet new information volume and complexity demands.” Source: Gartner

**VARIETY**

Rapid data generation coming from diverse sources. New sources of data is unstructured / semi-structured.

—“Through 2015, enterprises integrating high-value and diverse new information types and sources into a coherent information management infrastructure, will financially outperform their industry peers by more than 20%.” Source: Gartner

**VELOCITY**

Demand for near real-time processing of new data (for customer targeting, loyalty, fraud, etc.)

---

©2012, Cognizant
What Problems Are Driving You to Look at New Data Management Solutions?

- Poor query response: 45%
- Can’t support advanced analysis: 40%
- Inadequate data load speed: 39%
- Can’t scale up to large data volumes: 37%
- Cost of scaling up is too expensive: 33%
- Poorly suited to real-time or on-demand workloads: 29%
- Can’t support data modeling we need: 23%
- Current platform is a legacy we must phase out: 23%
- We need platform that supports mixed workloads: 21%
- Can’t support large concurrent user count: 20%
- Inadequate high availability: 19%
- Inadequate support for web services and SOA: 16%
- Inadequate support for in-memory processing: 16%
- Current platform is 32-bit, and we need 64-bit: 15%
- Current platform is SMP, and we need MPP: 14%
- We need platform better suited to cloud or virtualization: 13%
- Can’t secure data properly: 11%
- Other: 4%
- No Problem: 3%
How can I ensure Big Data has Value to the Business?
The Difference Between BI & Analytics
What Investments Does Big Data Touch?
WHO: Market Leaders are Using Analytics… but Now Want a Big Data Advantage

THE ANALYTICS HABITS OF TOP PERFORMERS
Top-performing organizations were twice as likely to use analytics to guide day-to-day operations and future strategies as lower performers.

Percent of respondents whose organizations do these activities well or very well.

- Use insights to guide future strategies:
  - Top Performers: 45%
  - Lower Performers: 20%

- Use insights to guide day-to-day operations:
  - Top Performers: 53%
  - Lower Performers: 27%

ANALYTICSTRUMPS INTUITION
The tendency for top-performing organizations to apply analytics to particular activities across the organization compared with lower performers. A likelihood of 1.0 indicates an equal likelihood that the organizations will use either analytics or intuition.

- Financial management and budgeting: 22.1
- Operations and production
- Strategy and business development
- Sales and marketing
- Customer service
- Product research and development
- General management
- Risk management
- Customer experience management
- Brand or market management
- Workforce planning and allocation

Overall Average
Deep Analytics and Visualization Will Supplant Traditional BI

Extreme analytics Allow Us to Find:

• Patterns
• Trends
• Predictions Never Before Feasible
Analytic "Coaches" are Key to Market Performers

1. How to Use analytics to Improve the Business

2. Lack of Management Bandwidth / Priorities
   - Lack of Analytics Skills at LOB
Every Level Can Benefit from an Analytics Advisor:

- I don't know what I want
- I don't know how
- I know, but need help
- I know, and I can do, but want to optimize
Where: Industry Analytic Advisors Provide a Critical Relationship Between Executives & Solution Makers...

Use Case: focusing on a compelling business opportunity

Big Data Solution Architect

Industry Analytic Advisor

Most Companies Try to use IT to Hand off Key Insights...
Where are you in your Big Data Analytics Journey?
I Know I Need an Analytical Advantage...

But What’s Really in Reach?

1. Enterprise Agility & Market Interaction
2. Strategic & Predictive Planning Clarity
3. Decision Accuracy & Insight Innovation
4. Performance Measurement

Cost to Acquire Customer, Lifetime Customer Value, RFM Analysis
3 Key Advantages: Now Within Reach

What do I Really need?

1. Enterprise Agility & Market Interaction
2. Strategic & Predictive Planning Clarity
3. Decision Accuracy & Insight Innovation
4. Performance Measurement
Not Usually a Strategic Advantage: But Big Cost Advantages

Performance Management is most often used in an Operational Fashion, for Optimization.

Data Cache or “Data Lake” allows for a Massive Data SandBox

(At 1/10 the cost of In-DW Storage?)
Free up your DW to focus on what it does best:

Analyze Structured Data, and Build Repeatable metrics

A Data Staging Area at Pennies on the Dollar

Use the Horsepower of Hadoop to Transform Your Data

Scale from 4 to 4000 Nodes - Incrementally, with Fairly Predictable Results
Scale your Metrics & Algorithms to Broader Parameters

Data Mining Algorithms Haven't Changed That Much in 20 Years - But Experts Agree That More Data Increases Accuracy & Insight

Analytic Leaders are Running Their Analytics Directly Within Hadoop or MPP

Use the Strengths of Hadoop to Make Broader & Deeper Analysis Than DW Alone

Manage Diverse Semi & Unstructured Data, Correlate & Analyze Within Hadoop
Just In Time: Worth More Than Decades of Hindsight

Much of Traditional BI and Some Analytics are Obsolete Upon Completion.

Streaming Data is Approaching Tera-Petabytes / Day in Telecom, Web, Media, etc.

Your Competitors Are Investigating It Now

Data Does Not Have to Be Retained, But Business Rules Can Take Action or Trickle Feed Key Insights to DashBoards or DW
Why You Need a Partner with Credibility and Methodology for Big Data Analytics

Key Considerations

Physical security of data center
• Operational & Application security
• Cluster security and user authorization
• Effective disaster recovery plan

Integration with existing IT Systems
• Integration strategy with Hadoop cluster
  • Control access groups

Information Security & Governance

Hardware Requirements

Software Requirements

Monitoring and alerting mechanism
• Library set up for various projects
• Data access frameworks to be set up (HBASE, PIG, HIVE)
• Hadoop Core setup (Common Java runtime for HDFS)

Requirement based configuration
• Light Processing config
• Balanced Compute config
• Storage heavy config
• Compute intensive config
## Cognizant’s integrated Social Media Analysis & Reporting Tool (iSMART)

<table>
<thead>
<tr>
<th><strong>Listener</strong></th>
<th>Calls social media web services (Facebook, Twitter, Blogs, E-Mails, Foursquare, Youtube etc) and crawl through social media websites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Integrate</strong></td>
<td>Houses social media data and associated BI/analytics structures. Integrates with the warehouse and data mart to bring the 360 degree view</td>
</tr>
<tr>
<td><strong>Analyze</strong></td>
<td>Measures brand’s sentiment on Social media sites using Cognizant’s proprietary Sentiment Analysis algorithm</td>
</tr>
</tbody>
</table>
| **Track**    | Mobile BI ready dashboards and reports  
Brand Sentiment Analytics  
Competitive Intelligence & Customer Insights  
Social Media ROI Calculation  
Marketing Campaign Effectiveness |
| **Act**      | Make informed decision  
Take action, and  
Track the results |

### Predictable Big Data Frameworks for Your Vertical Innovation
Offerings for your Big Data Analytics Journey Must Be Aligned & Integrated
Backup: Why You Need Big Data Solution Frameworks
Hadoop Creates New Possibility and Efficiency in the New Analytic Platform... But **MAJOR** Assembly is Required
Big Data Analytics-as-a-Service

Science
Medical imaging, sensor data, genome sequencing, weather data, satellite feeds, etc.

Industry
Financial, pharmaceutical, manufacturing, insurance, airline, energy & retail data

Docs / Legacy
Sales data, customer behavior, product databases, accounting data, etc.

Social & Machine
Log files, health & status feeds, activity streams, network messages, Web analytics, intrusion, spam list

High Volume Data Flows
Map Reduce Process
Consume results

"Integration is often 80% of the total cost in advanced analytics." - Data Mining Expert Ronnie Kohavi, PHD
Hadoop as DW Augmentation

Stage 1: Add unstructured Data

Application Request

OLTP Database, Structured Enterprise Data

ETL

EDW or Analytic Data Mart

Hadoop

Business intelligence

Analytics

OLTP Database, Structured Enterprise Data

©2012, Cognizant
Hadoop as an ETL “Pre Processing” Perform

Stage 2  Structure and Store

Application Request

OLTP Database, Structured Enterprise Data → Hadoop → Data Warehouse

Business Intelligence

Analytics

©2012, Cognizant
Hadoop as an ETL Perform

Stage 3  Ad-hoc query support

Application Request

OLTP Database, Structured Enterprise Data

Hadoop + Hive

Data Warehouse

Business intelligence

Analytics

Analytics
## Optimized Hadoop & Big Data Environments: Use Proven Building Blocks for Success

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Description</th>
<th>Reasoning</th>
<th>Known Customers &amp; Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Storage</strong> (Hadoop + DBMS)</td>
<td>Integrated in the data processing pipeline Hadoop is used to pre-process unstructured data (i.e. Apache web server logs) before uploading the data in the Data Warehouse for advanced analytics</td>
<td>Hadoop licensing and support is 10x cheaper than DW</td>
<td>Sears (retail) Monsanto (life science) CBS (media) Home Depot (retail)</td>
</tr>
<tr>
<td>2</td>
<td><strong>Compute</strong> (standalone Hadoop HDFS+MR)</td>
<td>Massively-parallel compute engine Hadoop is used to run complex compute-intensive analysis</td>
<td>In-house MR software skills Strong bias towards open-source</td>
<td>GE (broad variety) CODONiS (life science)</td>
</tr>
<tr>
<td>3</td>
<td><strong>Database</strong> (Hadoop Hbase)</td>
<td>Massively-parallel NoSQL database Hadoop is used to run SQL queries against extremely large volumes of data</td>
<td>Leverage old HW In-house MR/SQL skills Strong bias towards open-source</td>
<td>Ricoh (content mgmt)</td>
</tr>
</tbody>
</table>
Thank you