Mining and Monetizing the Content: Big Data, Data Virtualization and Analytics
Lopez Cisco Day

Ceejay Dideles
Services Solutions Specialist

July 10, 2015
Moving from Opinions to Intelligence

“If we have data, let’s look at data. If all we have are opinions, let’s go with mine.”

- Jim Barksdale, CEO, Netscape
The Challenge – 100’s of Big Data Use Cases

<table>
<thead>
<tr>
<th>AUTOMOTIVE</th>
<th>COMMUNICATIONS</th>
<th>CONSUMER PACKAGED GOODS</th>
<th>FINANCIAL SERVICES</th>
<th>EDUCATION &amp; RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto sensors reporting location, problems</td>
<td>Location-based advertising</td>
<td>Sentiment analysis of what’s hot, problems</td>
<td>Risk &amp; portfolio analysis</td>
<td>Experiment sensor analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIGH TECHNOLOGY / INDUSTRIAL MFG.</th>
<th>LIFE SCIENCES</th>
<th>MEDIA/ENTERTAINMENT</th>
<th>ON-LINE SERVICES / SOCIAL MEDIA</th>
<th>HEALTH CARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mfg quality Warranty analysis</td>
<td>Clinical trials Genomics</td>
<td>Viewers / advertising effectiveness</td>
<td>People &amp; career matching Web-site optimization UTILITIES</td>
<td>Patient sensors, monitoring, EHRs Quality of care</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OIL &amp; GAS</th>
<th>RETAIL</th>
<th>TRAVEL &amp; TRANSPORTATION</th>
<th>UTILITIES</th>
<th>LAW ENFORCEMENT &amp; DEFENSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling exploration sensor analysis</td>
<td>Consumer sentiment Optimized marketing</td>
<td>Sensor analysis for optimal traffic flows Customer sentiment</td>
<td>Smart Meter analysis for network capacity,</td>
<td>Threat analysis - social media monitoring, photo analysis</td>
</tr>
</tbody>
</table>

Crosses all Industries and Verticals
PLDT to invest over $30 million in 'Big Data' services

'In the long run, these 'Big Data' solutions will help improve PLDT’s bottom line and increase the subscriber base of the group,' PLDT chief strategy officer Winston Damarillo says.
Components of a Big Data and Analytics Solution
Big Data vs. Data Analytics

- **Big Data:**
  - Save money:
    - Cost effective scale
    - Consolidate many types of data
    - Operational efficiencies

- **Data Analytics:**
  - is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making
  - New ways to make money:
    - Predict, understand, and monetize customer behavior
    - Fact-based decision making
    - Enable real-time tactical decisions
Why a Different Approach for Storing Big Data?
Common “Database” Method

Constrained by Processing, Storage, and Networking Speed

Data In

Network

Store Data in Predefined Tables

SAN

Network

Data Processing

SAN

Network

Data Out
Traditional Enterprise Data Management

Data Sources
- Relational, Mainframe
- Documents and Emails
- Social Media, Web Logs
- Machine Device, Cloud

Data Warehouse (DW)

ETL Process

BI/Analytics
Data Warehouses Cannot Cost-Effectively Support Data Growth

Today, growth is accommodated by additional investment in your data warehouse.

100TB 100% DATA GROWTH 100TB

Data Warehouse $20,000 – $100,000/TB

To Add Capacity to the Data Warehouse Incremental spend of $2M – $10M
Data Warehouses Cannot Cost-Effectively Support Data Growth

Today, growth is accommodated by additional investment in your data warehouse

- 100TB Data Warehouse $20,000 – $100,000/TB
- 100% DATA GROWTH
- To Add Capacity to the Data Warehouse Incremental spend of $2M – $10M

Big Data complements your data warehouse, offloading data to defer/avoid more costly spend

- 100TB Lower Value Data
- 50TB Higher Value Data

Keeps the Right Data in the Data Warehouse:
- Operational Analytics
- Reporting
- Business Analytics

Offloading Everything Else to Big Data:
Saves $1.85M – $9.8M
- Historical Data
- Data Processing
- Data Hub/Ad Hoc Exploratory
- Transformation/Batch

Hadoop Cluster Cost $1000 – $2000/TB Incremental Cost $240K – $300K

© 2013-2015 Cisco and/or its affiliates. All rights reserved.
Data Optimization
How does it work?

Data Sources
- Relational, Mainframe
- Documents and Emails
- Social Media, Web Logs
- Machine Device, Cloud

Extract Transform Load (ETL)

Offload ETL jobs to Hadoop

Offload infrequently used data to Hadoop

HOT Data
Data Warehouse (DW)
~30% of the data

WARM/Cold Data
Data Warehouse Optimization (Hadoop)
~70% of the data

Cisco UCS Servers

BI/Analytics

Deliver Enriched BI/Analytics
Virtualize data from DW and Hadoop
Deliver richer & deeper data for Analytics

Optimize Storage / Costs
Migrate infrequently used data to Hadoop
Process Unstructured, New Sources in Hadoop

Process Unstructured, New Sources in Hadoop
Data Virtualization as Feed to Analytics
Data Virtualization

...is an approach to data management that allows an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted or where it is physically located.
Logical Data Warehouse

• One *logical* place to go for data
• Business view of data
<table>
<thead>
<tr>
<th>Key Analytic and Business Intelligence Partners</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>splunk</strong></td>
</tr>
<tr>
<td>- A scalable analytic platform primarily for event and machine data analysis.</td>
</tr>
<tr>
<td><strong>tableau</strong></td>
</tr>
<tr>
<td>- Highly intuitive, visualization data discovery, dash boarding, and data mash up capabilities</td>
</tr>
<tr>
<td>- Enables business users to do analytics without extensive skills or training with a BI platform</td>
</tr>
<tr>
<td><strong>Actian</strong></td>
</tr>
<tr>
<td>- Actian Analytics platform – Enterprise-grade SQL in Hadoop</td>
</tr>
<tr>
<td>- Analytics platform based on Open Source</td>
</tr>
<tr>
<td><strong>platfora</strong></td>
</tr>
<tr>
<td>- #1 Big Data analytics platform native on Hadoop</td>
</tr>
<tr>
<td>- End to End platform built for Multi-Structured data</td>
</tr>
<tr>
<td>- Self-service, iterative, interactive and fast</td>
</tr>
<tr>
<td><strong>Sas</strong></td>
</tr>
<tr>
<td>- Industry leader with portfolio spanning many disciplines including BI, data warehousing, and so on.</td>
</tr>
<tr>
<td>- Advanced analytics is a core strength. Also offers industry-specific analytic applications built on its product portfolio.</td>
</tr>
<tr>
<td><strong>MicroStrategy</strong></td>
</tr>
<tr>
<td>- An enterprise-grade end-to-end platform for large and complex enterprise requirements.</td>
</tr>
<tr>
<td>- Big focus on mobility</td>
</tr>
<tr>
<td><strong>informatica</strong></td>
</tr>
<tr>
<td>- Incumbent in ETL data management operations</td>
</tr>
<tr>
<td>- Focused on Data Integration</td>
</tr>
</tbody>
</table>
Case Study
Cisco Data Virtualization Use Case
Sony Entertainment Network

Customer Challenge
- Want to analyze integrated collection of data for PlayStation 3
- Analysis frequency is extremely limited
- Calculations can only be run once / month or once / week; need this data daily
- Combine different data sources and make it look like unified source; limit replication

Service Solutions
- A flexible data model/abstraction layer across all source systems
- Bridge dimensional data from Exadata with clickstream data and system usage data from Hadoop
- Enhanced agility through a flexible data delivery infrastructure

Impact on Customer
- 9M projected annual revenue increase due to more complete and timely analysis
- IT staff savings of $415,000 on first project
- Infrastructure cost avoidance of $304,000 on first project
- 7.2x ROI in first year
# Cisco Data Virtualization Use Case

## Comcast

<table>
<thead>
<tr>
<th>Customer Challenge</th>
<th>Service Solutions</th>
<th>Impact on Customer</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Analyze data pre and post deployment to identify changes</td>
<td>• A flexible virtual data layer across all source systems</td>
<td>• $21M in increased revenue in first year</td>
</tr>
<tr>
<td>• Understanding customers’ needs and ways to influence subscriber base</td>
<td>• Able to translate back and forth with different granularity on different platforms</td>
<td>• Decreased time to solution</td>
</tr>
<tr>
<td>• Analyze impact of proactive email campaigns</td>
<td>• Run predictive analysis to make better business decisions and improve the customer experience</td>
<td>• 4 projects in 6 months, a single project would have taken 6 months using ETL</td>
</tr>
<tr>
<td>• Integrate 3 different data sets; Customer Survey Data, Operational Data, Demographic Data</td>
<td></td>
<td>• Manpower reduced from 5 FTEs to 1 FTE</td>
</tr>
</tbody>
</table>

- Now 5-6x faster customer adoption of products
What’s next for Lopez Group?

ABS-CBN Broadcasting And iWanTV

ABS-CBN Mobile

Sky Broadband and SKY Cable

Head Office

Energy

What You Need:
- Data Optimization
- Data Virtualization

SKYBiz’ Retail and Real Estate
Thank You