LET THE TRANSFORMATION BEGIN
EMC IT Big Data Analytics Journey

Mahmoud Ghanem
Sr. Systems Engineer
Agenda

1. Introduction To Big Data
2. EMC IT Big Data Journey
3. Marketing Science Lab Use Case
4. Technical Benefits
5. Lessons Learned
What is Big Data

How to manage very large amounts of data and extract value and knowledge from them

No single standard definition…

“Big Data” is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.

Three V’s

<table>
<thead>
<tr>
<th>Volume</th>
<th>Velocity</th>
<th>Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data at Rest</td>
<td>Data in Motion</td>
<td>Data in Many Forms</td>
</tr>
<tr>
<td>Terabytes to exabytes of existing data to process</td>
<td>Streaming data, milliseconds to seconds to respond</td>
<td>Structured, unstructured, text, multimedia</td>
</tr>
</tbody>
</table>
EMC IT Big Data Journey
A Journey Of Maturity

1. AGGREGATE
   - Consolidation
   - Master Data
   - Common BI Tools
   - BI Governance
   FOUNDATION

2. LIBERATE
   - BA-as-a-Service
   - Data Scientist Services
   - Collaborative Analytic Tools
   - Unified Analytical Platform
   ANALYTICS ENABLEMENT

3. INNOVATE/ITERATE
   - Flexible / Scalable
   - Mission Critical
   - Real Time Capable
   - Collaborative Delivery
   DATA LAKE

4. HARNESS
   - Analytics -based decision making
   - Leveraging data to predict future models
   - Transforming operations by applying analytics
   ANALYTICS ENTERPRISE
Building The Data Lake

PROCESS EXECUTION

MONITOR THE BUSINESS

MEASURE BUSINESS PERFORMANCE

IMPROVE THE BUSINESS

DATA LAKE

01001010 1010010101 10101000 10101010 10100101

EMBED INTO BUSINESS APPS “MAKE THEM SMARTER”

GOVERNANCE

INNOVATE ITERATE REFINE
EMC IT Data Lake Architecture

ANALYTICS TOOLBOX APPLICATIONS

APPLICATIONS

MICRO-BATCH

DATA PLATFORM

EXECUTION

CASSANDRA

POSTGRESQL

MEMORY DB

GEMFIRE

SPRING XD

PIVOTAL HD

GREENPLUM DB

BATCH

REAL-TIME

INGESTION

UNSTRUCTURED

HDFS ON ISILON

HADOOP ON SCALEIO

STRUCTURED

BIG DATA FOUNDATIONS | BIG DATA SYSTEMS

NETWORK

WEB

SUPPLIER

MARKET

ERP

CRM

PLM
Big Data Analytics

Security
- Predict failure
- User behavior analysis

IT
- Predict failure of exchange servers

Manufacturing
- Proactively identify disk drives
- Better yield rates

Services
- Proactive and predictive services

Sales
- Service contracts renewals ~100M

Marketing
- Cross sell products ~10M
- Blending data target marketing & Segmentation to drive incremental revenue
Agenda

1. Introduction To Big Data
2. EMC IT Big Data Journey
3. Marketing Science Lab Use Case
4. Technical Benefits
5. Lessons Learned
Thought Leadership

Leadership’s Big Data marketing vision.

“...the best data may reside outside the organization’s database. Combining this data with your CRM and making sense of it requires robust algorithms. However, the benefits are tremendous—you’ll get better understanding of your customer profiles to create more targeted solutions and messaging.” (2013)

“To get a 360 degree view of your customer, you must find a way to pull all of this information into your grid and merge it with the data you do own” (2015)

The EMC Marketing Science Lab funnels CRM data into EMC’s data lake solution for predictive modeling … by marrying our internal data with external unstructured data … we are able to find individuals” (2015)

“The marketing science lab had the goal to build a lean of data science guys that could deliver insight en masse to the marketing team to make it more productive. Less communications, but more targeted, in a way that we think will be more attractive to that target audience.” (2013)

“The best data may reside outside the organization’s database. Combining this data with your CRM and making sense of it requires robust algorithms. However, the benefits are tremendous—you’ll get better understanding of your customer profiles to create more targeted solutions and messaging.” (2014)
# Marketing Science Lab

## Waterfall of Capabilities

### SKILLS

<table>
<thead>
<tr>
<th>Applied Statistics</th>
<th>Programming</th>
</tr>
</thead>
</table>

### TOP METHODOLOGIES

<table>
<thead>
<tr>
<th>Clustering/Classification</th>
<th>Next Likely Purchase</th>
<th>Propensity-To-Buy</th>
<th>Text Analytics</th>
</tr>
</thead>
</table>

### BIG DATA ANALYTICS

<table>
<thead>
<tr>
<th>Response</th>
<th>Purchase Rules</th>
<th>Segmentation</th>
<th>Journeys</th>
<th>Full ROI</th>
<th>Marketing Mix</th>
<th>Lead Score</th>
</tr>
</thead>
</table>
EMC’S Business Data Lake

Blending more data than ever before for targeting and segmentation!

Media Sources
- TechTarget
- Bombora

Social Data
- Sysomos
- Sprinklr
- WordPress
- Disqus
- Twitter

Competitive
- Aberdeen (Harte Hanks)
- CITDB
- Fairfax
- Fortune 500
- SFDC
- HG Data

Marketing Response
- 3rd Party Events
- Aprimo Marketing Contacts
- Adobe Analytics (Omniture)
- EBC Attendance
- Product Training Data

Enterprise
- Installed Base
- Opportunities
- Telemetric Data (SYR)
- Contracts
- Professional Services
Analytical Process

Customer segmentation

A wide range of attributes (giving the rich customer profile shown below) were used to categorize ABM accounts into three distinct segments based on purchase readiness – Undecided, Motivated and Ready —using a statistical technique called cluster analysis.

1. EMC Business Data Lake
2. Develop Target
3. Statistical Analysis for Segmentation
4. Tableau and Aprimo Integration
5. Deploy
Segmentation Using Classification Techniques and Business Rules

Use case #1

Business Data Lake

Enterprise Hybrid Cloud

EHC BUYER’S JOURNEY CONSISTS OF THREE STAGES

CTA IT TRANSFORMATION WORKSHOP

CTA CLOUD ADVISORY SERVICE

READY

MOTIVATED

UNDecided

© Copyright 2015 EMC Corporation, All rights reserved.
The EMC Marketing Science Lab was created to deliver incredibly deep, **360 degree views** of customer behaviors, needs, sentiments, and drive unprecedented marketing **effectiveness** for EMC.
Agenda

1. Introduction To Big Data
2. EMC IT Big Data Journey
3. Marketing Science Lab Use Case
4. Technical Benefits
5. Lessons Learned
Technical Benefits

**FASTER ANALYTICS**
Data Lake enables faster data analytics & modelling

**SIMPLIFIED PROVISIONING**
Reduced number of tables/report required

**TOOL AGNOSTIC**
Business logic in the DB not the tool provides increased flexibility

**FASTER DATA INTEGRATION**
Marketing Automation Data into BDL reduced from 7-10 days to <24 hours

**DATA GOVERNANCE**
Provides one version of the truth

**SHARE WORKSPACE**
Ability to share data analysis with others
Lessons Learned

The right people, aligned in the right way, working with the right technology

DEDICATED TEAM

ESTABLISH INDIVIDUAL / TEAM TO LEAD THE EFFORT

PARTNERSHIP

DETERMINE WHICH FUNCTIONS & TECHNOLOGIES TO INVOLVE

GAIN EXECUTIVE SPONSORSHIP

COMMUNICATE, EDUCATE, DRIVE ADOPTION

IT’S A JOURNEY
It’s a journey worth …
Thank You