

Applying Big Data

**Presented by John Dougherty
4/25/2013**

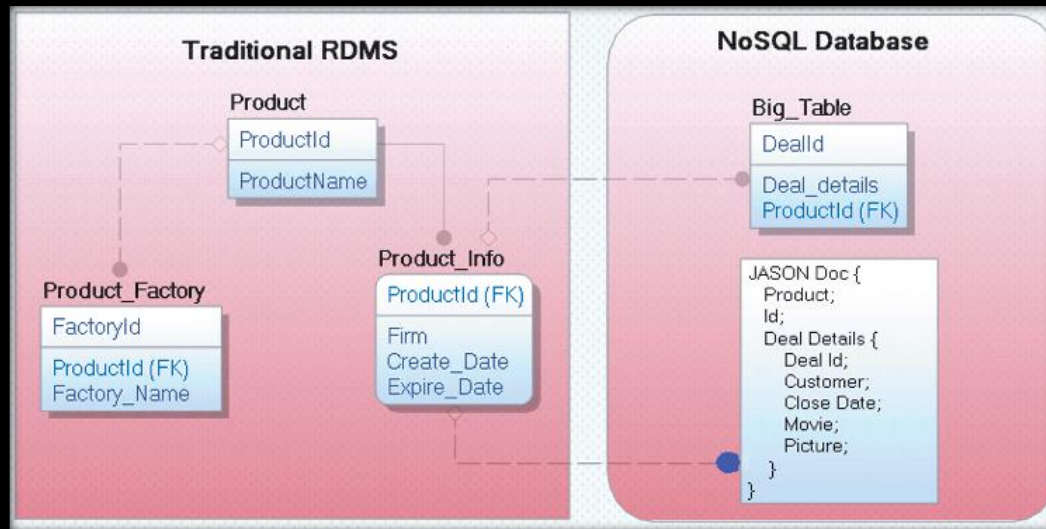
Big Data Buzzwords

- **Volume, Velocity, and Variety**
The 3V's originated in the early 2000's. META (Gartner, now)
Volume...self contained. Velocity = Speed of transaction
Variety = Data profiling from multiple data sources
Veracity has been added, but has not yet been fully adopted
- **Agility/Agile Development**
The Agile Manifesto, created February 2001 (Remember Scrum?)
Incorporation into Big Data software becoming mandatory
Adaptive and Predictive approaches are hotly contested
- **Modeling Data**
Data Modeling is paramount, given Big or Small datasets
Design must be confronted at ingress and egress
Hybrid data modeling and remodeling existing models

Big Data Buzzwords – Agile Dev.

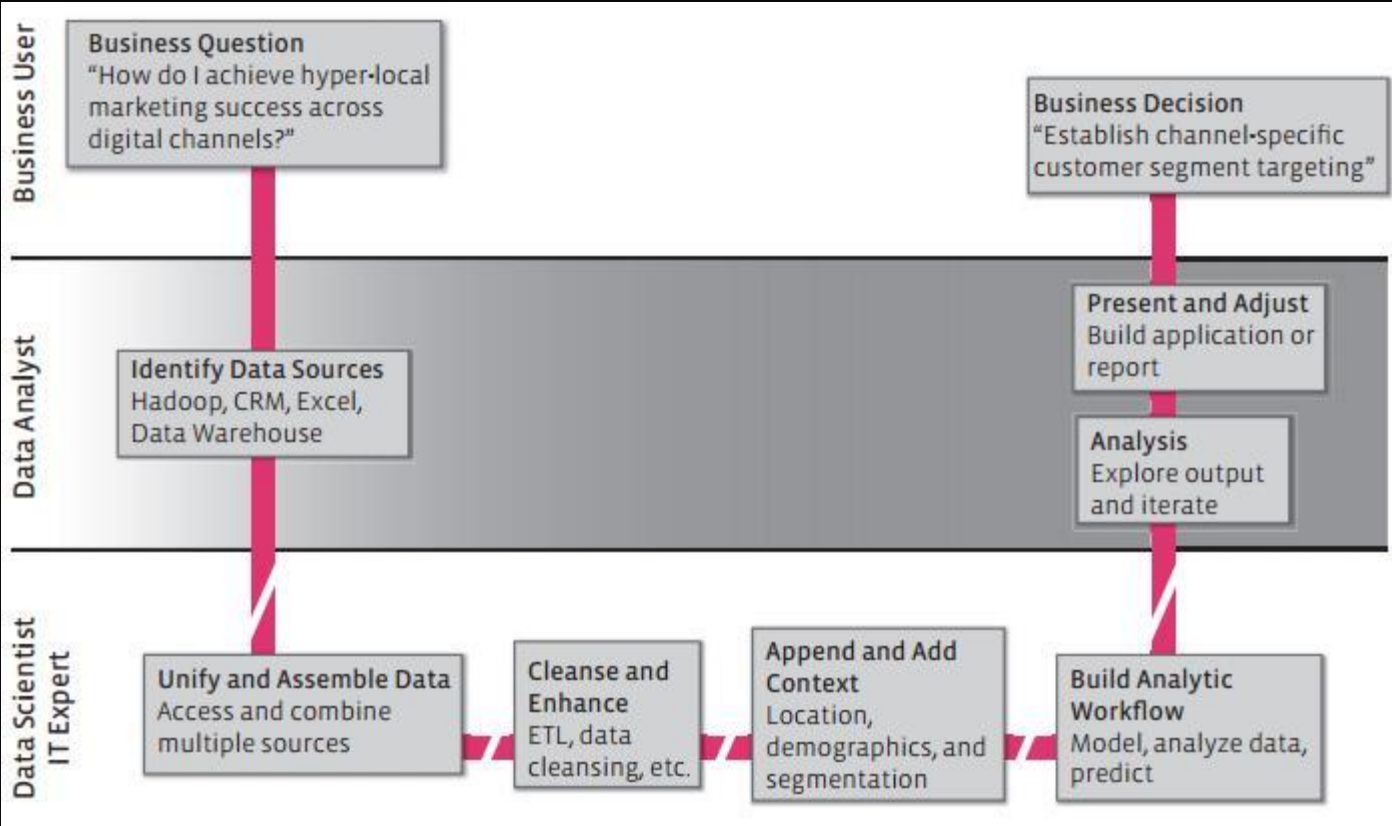
- **Informatics**
 - Informaticists are leveraged across multiple disciplines
 - There is no strict definition for a data scientist/informaticist
 - Greatest likelihood to adopt an agile/adaptive model
- **Daily Batch**
 - Development should be incorporated into existing process workflows. Seamlessness should be the goal.
 - Utilizing an agile approach to finding new uses to existing data
- **Classic Dept.**
 - Least likely to need/adopt new development approaches
 - Relevant data must still be filtered through
 - Staff should not be re-learning the wheel with deployment

Big Data Buzzwords – Data Models



- Example of Hybrid Modeling
- Every project/objective must have properly defined models to reach maximum efficacy
- Data silos are losing their complicit positioning
- Transitioning modeling to enumeration

Big Data Buzzwords – Question Inception



Connecting these lines is a great example of the work that lies ahead in identifying the objectives and goals of the business environment

Big Picture

There is a lot of data

As of 2009, Google generates at least >2 EB per year, >2TB indexed URLs, >9B page views per day

The Internet of Things
(connected devices and data)

YouTube houses >1EB of data, >72 hours of video per minute, >4B views per day

Facebook houses one billion users; utilizing >500TB per day, housing 35% or more of the world's photos

Twitter >125B tweets per year, >390M per day, approximately 4500 per second

~2.3B people use the internet today, of which, 90% of the world's data has been generated within the last two years

What will you be aggregating?

In 2002, recorded media and electronic information flows generated about 22 exabytes (10^{18}) of information
In 2006, the amount of digital information created, captured, and replicated was 161 EB

Big Data for Business

Use Cases

IBM's 5 High Value Use Cases



Big Data Exploration

Find, visualize, understand all big data to improve decision making. Big data exploration addresses the challenge that every large organization faces: information is stored in many different systems and silos and people need access to that data to do their day-to-day work and make important decisions.



Enhanced 360° View of the Customer

Extend existing customer views by incorporating additional internal and external information sources. Gain a full understanding of customers—what makes them tick, why they buy, how they prefer to shop, why they switch, what they'll buy next, and what factors lead them to recommend a company to others.



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real time. Augment and enhance cyber security and intelligence analysis platforms with big data technologies to process and analyze new types (e.g. social media, emails, sensors, Telco) and sources of under-leveraged data to significantly improve intelligence, security and law enforcement insight



Operations Analysis

Analyze a variety of machine and operational data for improved business results. The abundance and growth of machine data, which can include anything from IT machines to sensors and meters and GPS devices requires complex analysis and correlation across different types of data sets. By using big data for operations analysis, organizations can gain real-time visibility into operations, customer experience, transactions and behavior.



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency. Optimize your data warehouse to enable new types of analysis. Use big data technologies to set up a staging area or landing zone for your new data before determining what data should be moved to the data warehouse. Offload infrequently accessed or aged data from warehouse and application databases using information integration software and tools.

Use Cases

Visual Analytics

Science: Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

- Applied since data science began, 1970's
- Many different products available, augmenting existing solutions, and providing all-in-one
 - SAS
 - SAP (though called predictive analytics, still fits)
- Same problems incur with extensibility as do with design/deployment

Sensor Analytics

Internet of Things: The first speaking of the gargantuan brontobyte

(1 Bit = Binary Digit · 8 Bits = 1 Byte · 1024 Bytes = 1 Kilobyte · 1024 Kilobytes = 1 Megabyte · 1024 Megabytes = 1 Gigabyte · 1024 Gigabytes = 1 Terabyte · 1024 Terabytes = 1 Petabyte · 1024 Petabytes = 1 Exabyte · 1024 Exabytes = 1 Zettabyte · 1024 Zettabytes = 1 Yottabyte · 1024 Yottabytes = 1 Brontobyte · 1024 Brontobytes = 1 Geopbyte)

ROI?

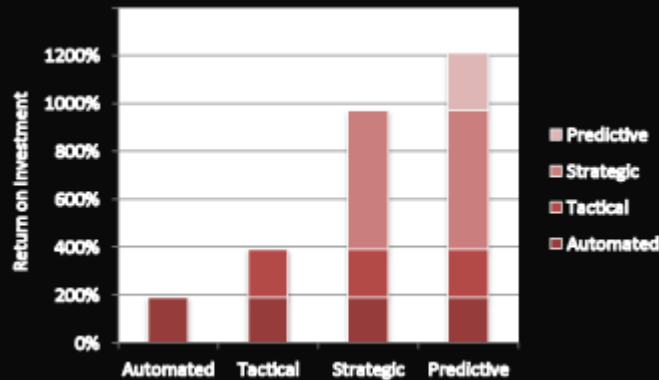
- ROI Metrics are difficult to predict, but follow a trend of double and triple digits
- What keeps the CEO up at night, decision journeys
- An anecdotal report (questionnaire) shows 44% of CMO's can measure their ROI
- Design and development will continue to be tantamount to a successful return



ROI

Nucleus Research

Four Stages of an Analytic Enterprise



- Becoming an analytic enterprise requires Big Data
- Average ROI of 241%
- Increased productivity
 - A major metropolitan police department achieved an 863 percent ROI when it combined its criminal records database with a national crime database created by a major university.
- Reduced labor costs
 - A major resort earned an ROI of 1,822 percent when it integrated shift scheduling processes with data from a national weather service, enabling managers to avoid unnecessary shift assignments and increase staff utilization.

Telco reduces costs associated to CO management and circuit deployment by 230%

QoS data expected to expand well into Petabytes for the Telco industry

Moving Forward

How to formulate the right questions

- Communication between C-Suite and VP isn't enough
- Considering old data, holistic approaches work best
- Objectives and goals begin with dialogue at the highest levels
- What are the questions we should be asking?

Should we start now? Yes.

Brought to you by:



Viriton

Big Data for Business

BIBLIOGRAPHY

- <http://blogs.starcio.com/2012/12/what-is-big-data-real-challenges-beyond.html> - Big Data for All Businesses
- http://www.nytimes.com/2013/03/24/nyregion/mayor-bloombergs-geek-squad.html?pagewanted=all&_r=2& - NYC Mayor's use case
- <http://www.312analytics.com/what-is-machine-learning-big-data-modeling/> - Data Modeling, the big challenges
- <http://goo.gl/wH3qG> - Origin of VVV
- <http://www.businessinsider.com/cia-presentation-on-big-data-2013-3?op=1> - CIA CTO Presentation
- <http://www.rosebt.com/1/post/2013/03/data-science-and-analytics-workflow.html> - Rose Business Technologies, Workflow
- <http://nucleusresearch.com/research/notes-and-reports/the-big-returns-from-big-data/> - Nucleus Research