

zubin.dowlaty@mu-sigma.com

Twitter: @crunchdata



Mu Sigma

HPC User Forum

Trends Discussion

Do The Math

Chicago, IL

Bangalore, India

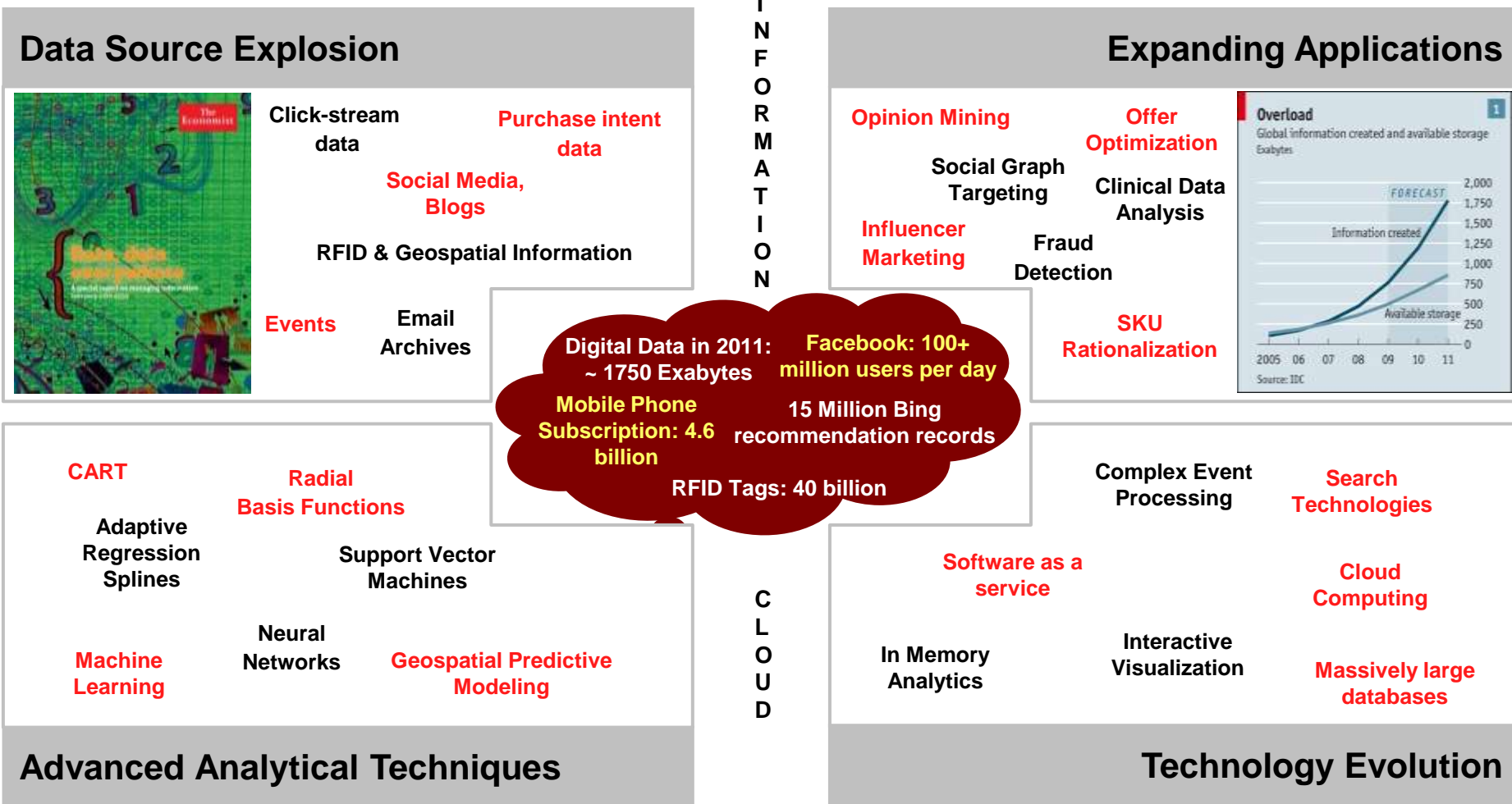
www.mu-sigma.com

2013

Proprietary Information

"This document and its attachments are confidential. Any unauthorized copying, disclosure or distribution of the material is strictly forbidden"

The information explosion is leading to an unprecedented ability to store, manage and analyze data; we need to run ahead..



Three Key Macro Analytics Trends Targeting Consumption

Analytics Trends

▶ Computational Enablement

- Scalable Analytics, Big Data & NoSQL technologies, GP-GPU, In – Memory & High Performance Computing
- Master Data Management, Cloud, and Federated Data – the end of the central EDW?
- Analytics as a Service and the Cloud
- Open Source Maturity
- Rise of Agile Self Service
- Data Model Disruption: ETL vs. ELT

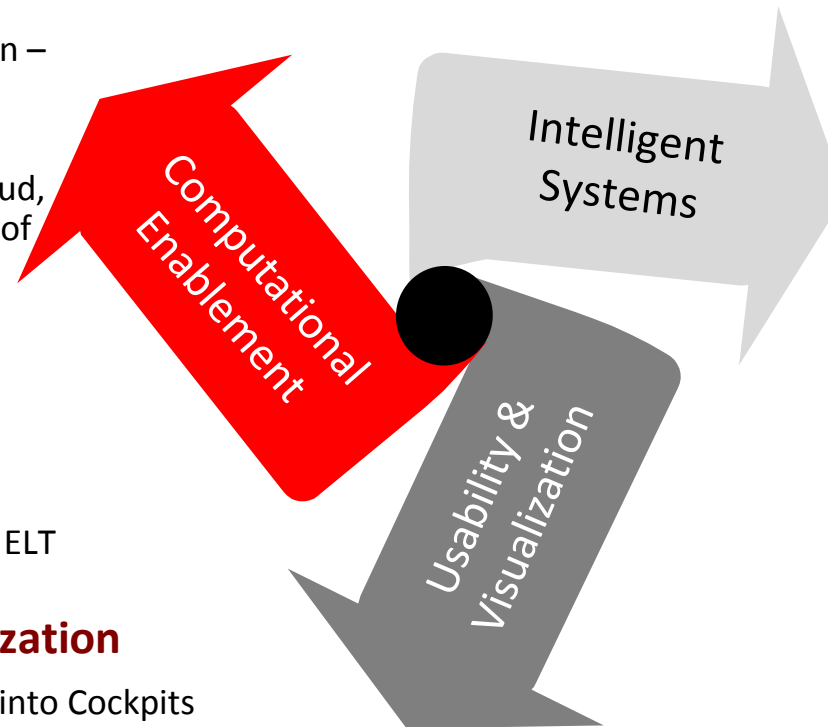
▶ Usability and Visualization

- Dashboards will Evolve into Cockpits
- Improved Interactive Data Visualization & Aesthetics
- Augmented / Virtual Reality / Natural User Interfaces
- Graph Analytics sparked by Social Data
- Discovery with Computational Geometry
- Mobile BI and the Integration of Consumer & Enterprise Devices
- Gamification: Do you want to play a game?

▶ Intelligent Systems

(“Anticipation Denotes Intelligence”)

- Operational Smart Systems are Back propagating – Pervasive Analytics
- Real Time Analytics and Event Streams
- Artificial Intelligence (AI) is not what we expected
- Don’t forget the 4th tier it’s Juicy
- Metaheuristics and the Ants
- Operational Analytics, You are the Exception
- Experiments in the Enterprise and the Quasi
- Just Say No to OLS
- Energy Informatics if Radiating
- Intelligent Search





Big Data's recent enterprise popularity can be attributed to two key factors

▶ Technology Trigger

- Robust Open Source Technology for Parallel Computation on Commodity Hardware (Hadoop / NoSQL)

▶ Frustration Trigger..

- Tyranny of the Data Model, Cost & Complexity for data integration, Agility, Impact, Bureaucracy, Difficulty & Costly to scale of existing EDW technology

Google Trigger: Releases "Map Reduce" & "GFS" paper

Faced with storing massive data sets from indexing the web, Google searches for & finds a way to store & retrieve the data in commodity hardware



Apache: Hadoop project



Teradata buys Aster (map-reduce DB)



Cloudera named most promising startup funded in 2009



IBM Acquires



Hortonworks



Yahoo: 10K core cluster



EMC Acquires Greenplum

HP Acquires Vertica

Oracle MSOFT SAS



A Mindset for Enabling Decision Sciences At Scale

Dataset -> Toolset -> Skillset -> Mindset

▶ Skillset

- Map-Reduce, HDFS, HIVE, R, PIG, NoSQL, Unstructured Data
- Parallel computation & Stream Processing
- Data Scientist and Data Explorer
- “...big data job postings on Dice more than tripled year-over-year – Jan 2013”

▶ Mindset

- Automation and Scale
- Higher utilization of machines for decisions
- Agile
- Learning vs. Knowing
- Consumption
- Algorithmic and Heuristic (Man & Machine)



Trend: Dashboards will Evolve to be More Actionable and Forward Looking

Figure 1: Business Intelligence Maturity Levels

Maturity	Analogy	Additional Capability
Pre-BI	Report	Report
I	Investigation	Interactive Reporting
II	Dashboard	Real-Time Reporting
III	Cockpit	Actionable Reporting
IV	Flight Simulator	Business Simulation
V	Autopilot	Business Optimization

Dashboredom?

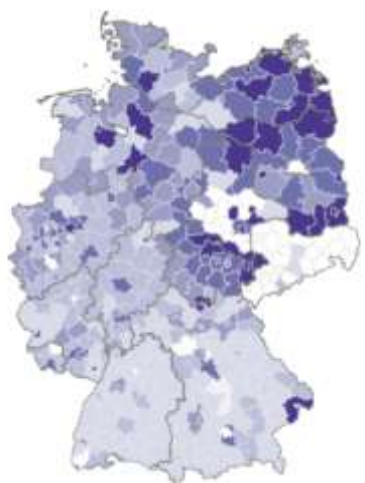
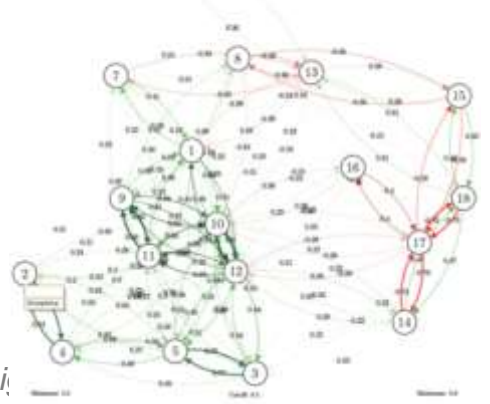
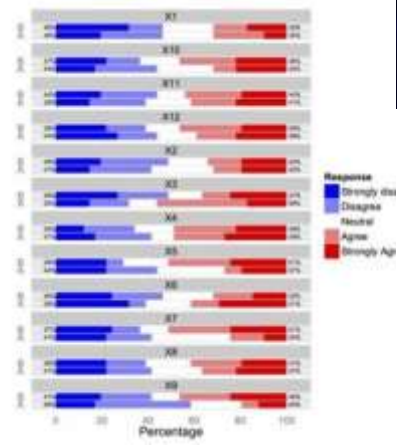
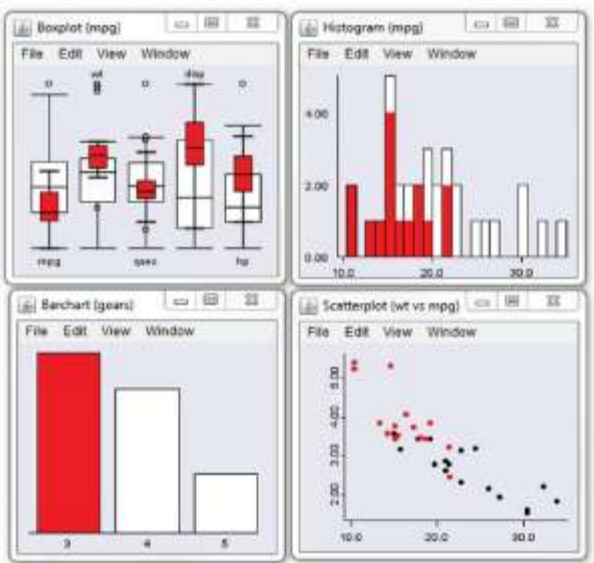
http://www.information-management.com/issues/20_7/dashboards_data_management_quality_business_intelligence-10019101-1.html

“As we move from the information age into the intelligence age”..



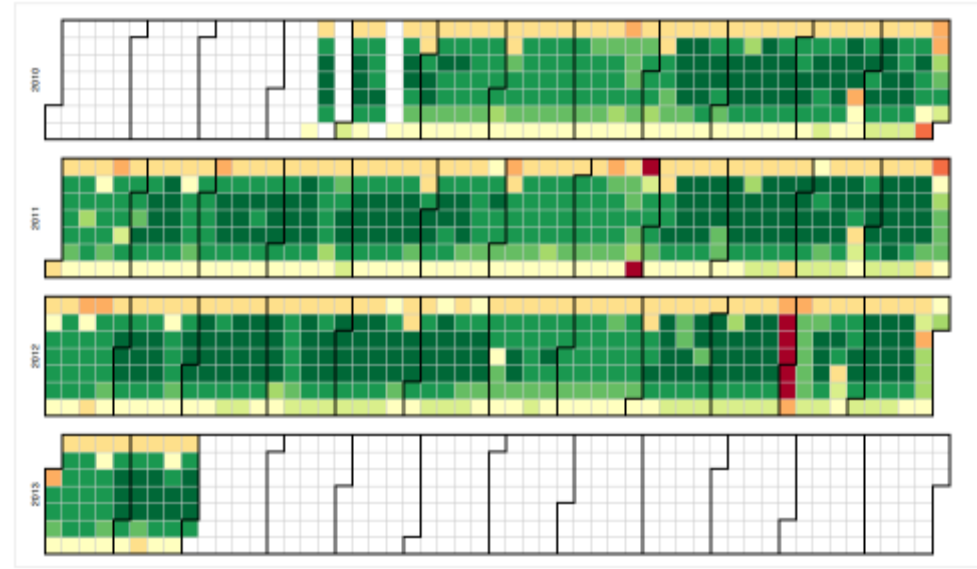
Open Source:

- ▶ R is the premier language for data scientists
 - Strength in data visualizations..
 - <http://www.r-project.org/>

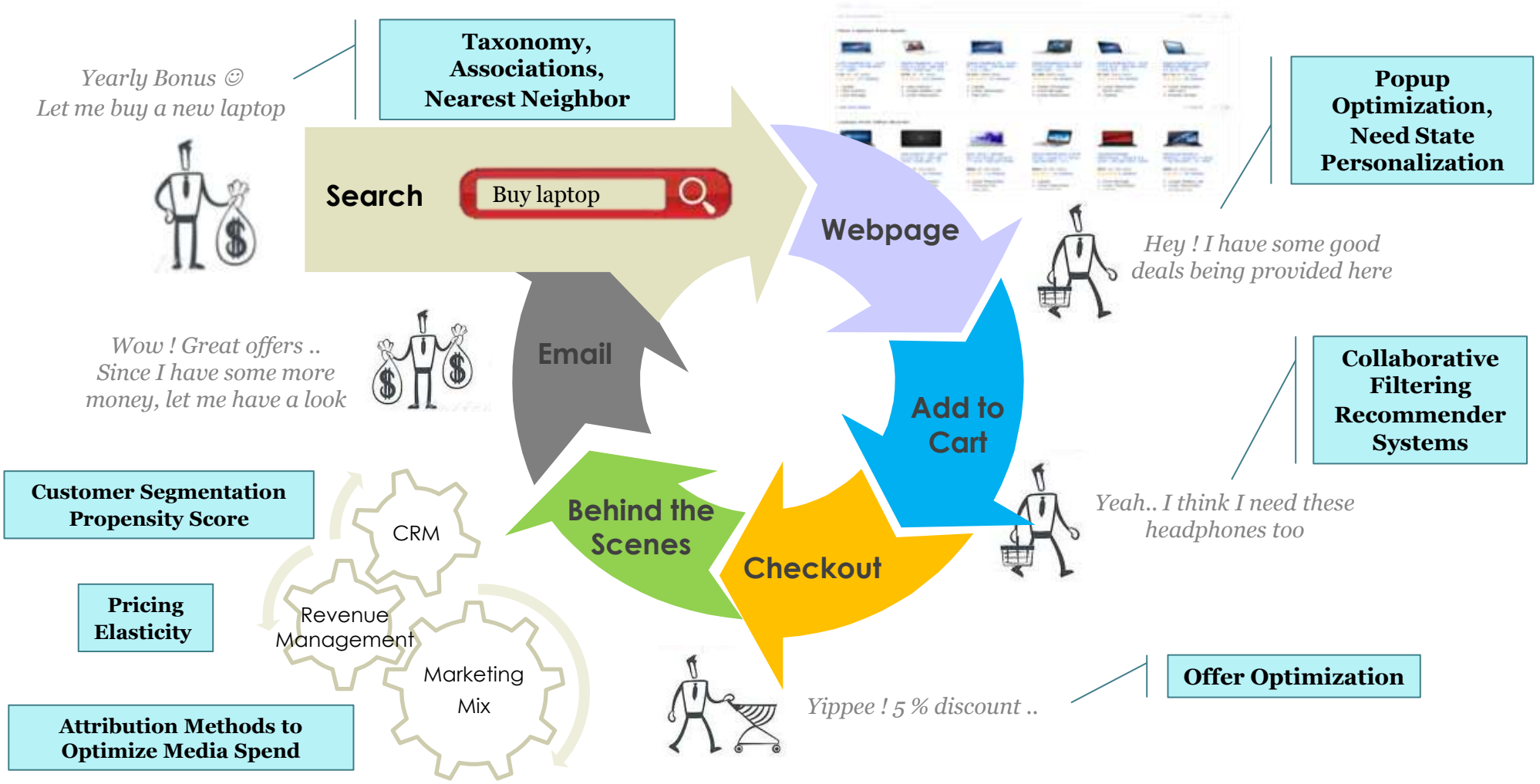


NYC Subway Usage

March 14, 2013



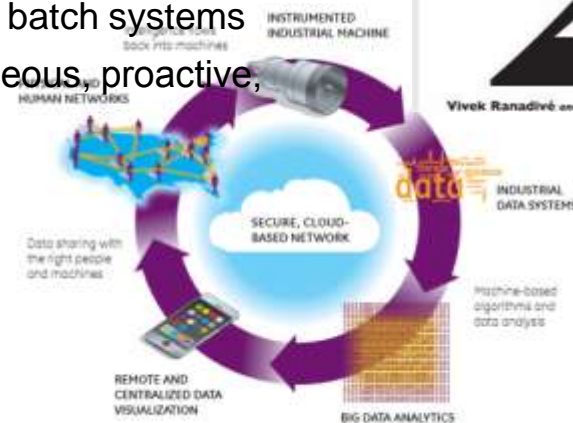
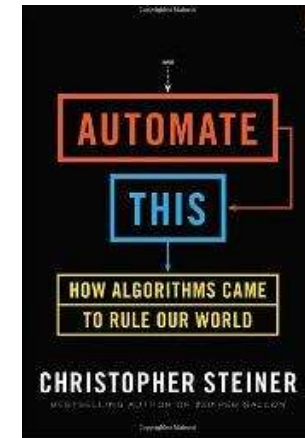
The current business scenario demands automating and injecting intelligence into various enterprise touch points



Add Intelligent Analytics Everywhere You Can to Unlock The Value In Information

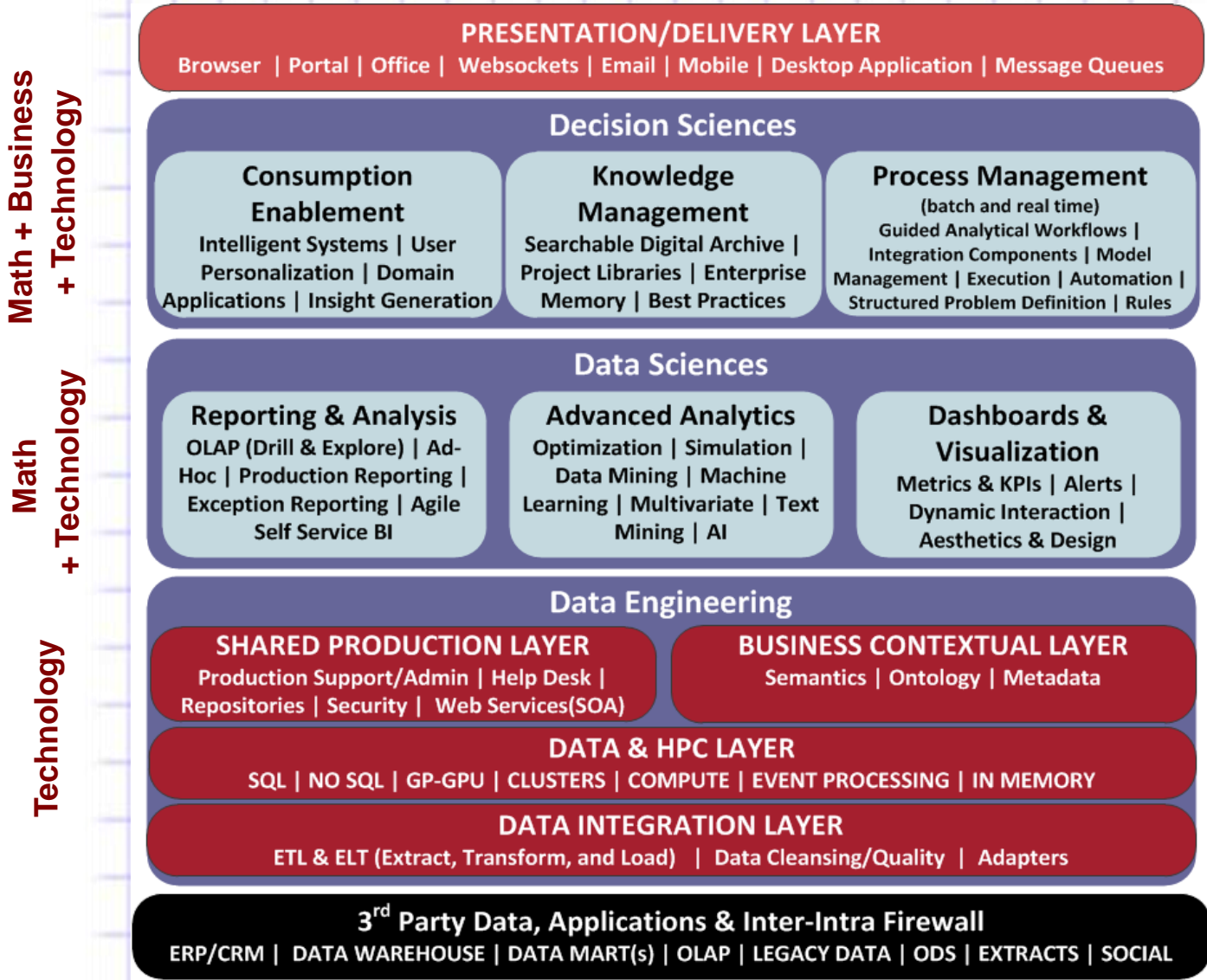
Current News on Intelligent Systems

- ▶ **CEP Vendors Acquired** June 2013: Streambase (Tibco) and Progress Apama (Software AG)
- ▶ USA Today, September 16, 2012
 - <http://www.usatoday.com/money/business/story/2012/09/16/review-automate-this-probes-math-in-global-trading/57782600/1>
- ▶ **Two Second Advantage:** Published 2011
 - Excellent portrayal of Intelligent Systems, patterns from the human brains – mental models
 - Wayne Gretzky: Edge: He predicted a little better than anyone else, where the hockey puck would be.
 - **Enterprise 2.0: Every transaction became a bit of digital data**
 - **Enterprise 3.0: Every EVENT can become a bit of digital data**
 - There will always be value in making projections, days, months, years, predictive analytics, data mining, and analytics systems on historical data – batch systems
 - But in today's world, enterprises need something more, instantaneous, proactive, predictive capability – real time systems
- ▶ **GE Mind an Machines:** November 2012
 - » Industrial Internet, Intelligent Systems, and Intelligent Machines
 - » http://www.ge.com/docs/chapters/Industrial_Internet.pdf
 - » <http://www.youtube.com/watch?v=Svl3Pmv-DhE>





Decision Support Stack for Operationalizing Analytics



Platforms	Assets/ Products
Decision Sciences	muESP™ muBuzz™ muFlow™ muMix™ muPDNA™ muRx™
Data Sciences	muXo™ muHPC™ muText™

Mu Sigma Future State Reference Analytics Design

Mu Sigma's I&D Big Data team has developed MapReduce implementations of frequently utilized algorithms in R

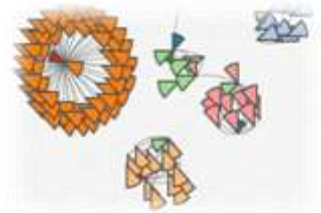
muHPC™ Packages - Map Reduce implementations of frequently utilized statistical algorithms in R

muEDA – Built on **RHIVE** (A package for executing HIVE queries through R)



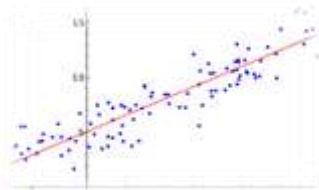
Consists of functions to perform frequency analysis, univariate analysis and other EDA techniques

muKMeans – Built on **RMR** (A package for writing MapReduce codes in R)



Consists of functions to perform data clustering using K-Means

muGLM – Built on **RMR** (A package for writing MapReduce codes in R)



Consists of functions to fit linear and generalized linear models

zubin.dowlaty@mu-sigma.com

Twitter: @crunchdata



Thank You

**Chicago, IL
Bangalore, India
2013
www.mu-sigma.com**

Proprietary Information

"This document and its attachments are confidential. Any unauthorized copying, disclosure or distribution of the material is strictly prohibited"