Big Analytics: A Next Generation Roadmap

Cloud Developers Summit & Expo: October 1, 2014

Neil Fox, CTO: SoftServe, Inc.

© 2014 SoftServe, Inc.
Remember Life Before The Web?

1994
Even Revolutions Take Time

- Developed world
- World globally
- Developing world

* Estimate
Big Analytics?
Big Analytics also describes the technology solutions used by giants such as Google®, Amazon®, and Facebook® to process enormous data to provide you the best in internet, real-time services.
Value Proposition

Public Safety:
Durham, NC Police use Predictive Analytics to reduce crime rate by 50%

Fraud Prevention:
Insurance company uses Predictive Analytics to save $12 m annually

Supply Chain Optimization
Retailers using Predictive Analytics to forecast product demand, price, promotion and inventory management

Healthcare
Impacting every aspect of the healthcare system – giving more personalized data to patients, providers and payers. Areas of focus include lifestyle, diet, exercise, research, and clinical trials

Retail:
Macy’s boosts same store sales by 10% using 10s of millions of data points from twitter, social media, in-store and on-line.
...And just how powerful is this stuff?

Marketing (e.g. Target):

"... computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a “pregnancy prediction” score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy." [1]

Government (via SAP survey):

"87% of federal and 75% of state IT officials believe that real-time Big Data has the potential to save a significant number of lives." [2]


A Success Story

The Durkheim Project

• Developed linguistics-driven prediction models to estimate the risk of suicide. & Reached 70% accuracy in cohort distinction [3]

• Deployed a real-time big data framework to capture opt-in mobile and social media data from veterans

• Fast Company said “This may be the most vital use of big data we’ve ever seen.” [4]


So What is Slowing Adoption Down? ...

1. Resistance to change

“... there's the drag exerted by relational database administrators who badly want to stick to what they know.” [5]

2. Volume, Variety, and Velocity

“... big data problems have just as much to do with changing how you do data querying and processing as they do with handling the oft-cited "three V's" -- the big data parameters of volume, variety, and velocity.”

- Volume (of data under management) - Data is growing from the Terabytes to the Petabytes... for everyone
- Velocity (of transactions)- NoSQL simply lets you access your data differently
- Variety (of data) – Unstructured?, structured?, semi-structured?

Big Analytics Engineering Challenges

Real Time Intelligence

How to achieve **Low Latency** for personalized customer experience in real-time?

How to improve **System Performance** for Data Science/Analytics team?

How to implement **Self-Service** with high **Data Quality** over terabytes and petabytes?

Data Scientists/Analysts

Consumers

Intelligent Agents

Business Users

Data Discovery

Business Reporting
Complex Environment

Big Data Landscape

Vertical Apps
- Predictive Policing
  - bloomreach
  - myráx

Ad/Media Apps
- rocketfuel
  - collective

Business Intelligence
- ORACLE Hyperion
  - SAP BusinessObjects
  - IBM Cognos

Analytics and Visualization
- Tableau
  - Palantir

Log Data Apps
- splunk
  - loggly
  - sumologic

Media Science
- TURN
  - DataXu

Data as a Service
- factual
  - GNIP

INRIX
  - LexisNexis

Data as a Service
- Loggie
  - knoema

LOGATE

Analytics Infrastructure
- Hortonworks
  - cloudera

MapR
  - Vertica

Informatica

Operational Infrastructure
- Couchbase
  - 10gen

HADOOP

Teradata
  - Hadoop

Hadapt

Infrastructure as a Service
- Windows Azure
  - Google BigQuery

Structured Databases
- ORACLE
  - SQL Server

IBM DB2

SoftServe
Empowering your Business through Software Development

Copyright © 2012 Dave Feinleib

dave@vcdave.com

blogs.forbes.com/davefeinleib
Sample Technologies

- Hadoop
- R
- Vertica
- Spark
- Tableau
- Qlik
- Apache HBase
- Cassandra

SoftServe Empowering your Business through Software Development
But now Hadoop...

**Hadoop In Action**

How is Hadoop used, or how will it be used, at your organization?

- Running analytics: 66%
- Business intelligence: 63%
- For extract, transform, and load functions: 35%
- Archiving: 23%

Data: InformationWeek 2014 State of 321 Database Technology Survey of 956 business technology professionals using or considering using Hadoop, January 2014
...though the adoption is still behind
Big Analytics Reference Architectures and Practices

- ROI-driven Big Analytics systems design based on proven Architectures and Technologies
- Maximum efficiency with the lowest Cost per Terabyte
So What to Do?

To Properly Frame a Big Data + Analytics Project

1) What is the business goal?
2) What data can we get in support of this?
3) How can we display business intelligence intuitively?
4) Iterate, iterate, iterate

“Start with a focused, business-driven project, make sure the data is consistent with your vision and then apply advanced analytics without moving beyond human-understandable decisions.” [6]

SoftServe, Inc. is a unique software development partner, offering specialized outsource technical staffing... and integrative, and professional services to some of the world’s premier technology companies.
Our Unique Integrated Approach

Abiliton™
Abiliton Big Analytics

SoftServe’s adaptive best practice framework for Big Data/Business Analytics transformation and optimization

**People**
- Competence Development
  - Knowledge Model
  - Performance Management Practices
  - Training Catalog
- Organization Structure
  - Optimal Team Structure
  - Roles and Balance

**Process**
- SDLC Optimization
  - Governance
  - Project Management
  - Business Analysis
  - Software Engineering
  - Quality Control
  - DevOps
  - Metrics for Continuous Improvement
  - Project Status Dashboard

**Technology**
- Data Science & Analysis
  - Predictive Modeling
  - Statistical Analysis
  - Standard and Ad-Hoc Reports
- Data Engineering
  - Solution Architecture
  - System Modernization
  - System Optimization
  - Operations Automation
Big Analytics: Technology Framework

A modern integrated approach for solving Big Data/Business Analytics needs across multiple verticals and domains

All Data
- Application data
- Media data: images, video, etc
- Social data
- Enterprise content data
- Machine, sensor, log data
- Docs and archives data

Data Acquisition and Storing
- Data Integration
- Data Lake (Landing, Exploration and Archiving)
- Enterprise Data Warehousing

Analytics
- Reporting and Analysis
- Predictive Modeling
- Data Mining
- UX and Visualization

Applications
- Customer Analytics
- Marketing Analytics
- Web/Mobile/Social Analytics
- IT Operational Analytics
- Fraud and Risk Analytics

Real-time Data Processing
- Complex Event Processing
- Real-time Query and Search

Data Management
(Governance, Security, Quality, MDM)
Big Analytics Case Study: Network Security

**Business Goals:**
1. Provide reporting platform in the cloud for services & applications usage analysis
   Charge customers based on the platform they are using, number of consumers’ applications etc.

**Technical Specs:**
- Machine generated data
- Big Data: 7.5BLN log records per day
- Near real-time reporting
- Reports which “touch” billions of rows

**Solution:**
- ETL - Talend
- DW – HP Vertica/ InfoBright EE
- OLAP – Pentaho
- BI – JasperServer Pro
Big Analytics Case Study: Online Analytics

**Business Goals:**
- Insights and optimization of all web, mobile, and social channels
- Optimization of recommendations for each visitor
- High return on online marketing investments

**Technical specs:**
- Big Data > 1PB
- 10+ GB per customer/day
- 10+ Hadoop Clusters
- 15+ Aster Data Clusters

**Solution:**
- Hadoop/HBase/Hive
- Aster Data
- Oracle
- Java/Flex
Takeaway: Metrics for Success

• **Qualitative Performance**: For example, compelling visualizations that make tasks easier (i.e. less complex)

• **Quantitative Performance**: For example, maximizing systems ROI, reducing TCO, or even saving lives...

• **Compliance and Data Governance**: For example, privacy concerns & jurisdictional issues
2025?

- Applications were generic
- Doctors (and patients) had one data point per year
- Doctors had to rely on their own ability to research
- Shoppers had to search for deals
- Companies had exabytes of data they were not using
Thank you!

US OFFICES
Austin, TX
Fort Myers, FL
Boston, MA
Newport Beach, CA
Salt Lake City, UT

EUROPE OFFICES
United Kingdom
Germany
Netherlands
Ukraine
Bulgaria

EMAIL
info@softserveinc.com

WEBSITE:
www.softserveinc.com

USA TELEPHONE
Toll-Free: 866.687.3588
Office: 239.690.3111