Big Data

Manuel Sevilla, VP, Global CTO | BIM

Jeff Hunter, VP, Business Technology Innovation
Business Information Management TLI Overview

**Skills**

- Strategy & Performance Mgmt.: 400
- BI & Analytics: 3000
- Data warehousing: 2000
- Content management: 1500
- Information management: 3000
- Big Data: 500

**Major Wins 2013**

- The Coca-Cola Company
- Johnson & Johnson
- ERDF
- T-Mobile
- Boehringer Ingelheim
- ABN-AMRO
- Bayer
- Posten
- Southwest Airlines
- Smiths
Global Partnership: 24th of October 2013

Cloudera proposes the best Hadoop distribution especially in term of:
- **Security** (Manager, Sentry...)
- **Performance** (Impala, Spark/ & Shark)
- **Search** (SolR native integration with Hadoop)

Recent investments (esp. Intel) will help making significant improvements (technical + sales)

250+ trained people on Cloudera

40% of Big Data pipeline is with Cloudera (NA, FS, Europe)
Global Partnership: 4th of December 2013

Pivotal proposes a unique complete Big Data software infrastructure stack with:
- Pivotal HD (Hadoop)
- Gemfire (event driven + in memory)
- HAWQ & Greenplum (data warehouse with or w/o Hadoop)
- Spring & Cloud Foundry (custom dev)
All this using commodity/VM hardware

Gold Member of Cloud Foundry

200+ trained people on Pivotal

40% of Big Data pipeline is with Pivotal (NA, FS, Europe)
In order to plan and justify inter-continental fiber deployment

- Broadband connectivity, and voice, satellite, corporate, IP interconnection and mobility services, operating in 30+ countries
- Analyzing granular network data with 5 years of history
- Identify new opportunities by taking into account socio-economic indicators into the predictive analysis of network usage
- Using socio-thermodynamic models for analyzing and generating predictions on network usage on internet traffic in underwater cables
- Vibrating models has revealed links between countries and population

- Cloudera CDH platform on Amazon Web Services Cloud (IaaS) using Impala for SQL querying
- Tableau for reporting
- R for Analytics
Being aware of the real level of risk exposure in real-time

- Increased regulatory demand and market forces demand retention of large scale granular data and quick access to it
- They desire to move from the current ‘brittle’ DW architecture to a Data Reservoir, with a ‘minimalistic’ model
- Load data relating to trades, positions, valuations, etc – and classify hot, warm and cold data according to latency of access desired (e.g., hot data represents most recent 5 days data and will reside in memory)
- Real Time Risk assessment
- Warm and cold data will be accessed from the Reservoir via a SQL-like interface.

### Investment Banking

#### Data Reservoir

**Stream Ingestion**
- **S1**: Constant data load
- **S2**: Continuous data load and aging (partial)

**In-Memory Processing**
- **S3**: On-demand cache load of reservoir data
- **S4**: SQL access to warm data
- **S5**: SQL access to warm data
- **S6**: SQL access to data in the reservoir
- **S7**: Direct access to reservoir data
- **S8**: SQL access under peak data load conditions
- **S9**: Cache server node failure
- **S10**: Backup store node failure
- **S11**: Data reservoir node failure

**Hadoop Feed**

**Pivotal HD**

**GemFire XD**

**Spring Batch**

**SQL Fire Transformation**

**Data Reservoir**
- **HDFS**
Helping our customers in identifying Big Data opportunities and Key initiatives to be taken

<table>
<thead>
<tr>
<th>Weekly project meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
</tr>
<tr>
<td>Week 1</td>
</tr>
<tr>
<td>△</td>
</tr>
</tbody>
</table>

### Business Stream 1
**Core Business**
- Reach a new level of operational excellence

### Business Stream 2
**User Centric**
- Foster personalization & recommendation

### Business Stream 2
**B2B & Market Data**
- New revenue streams (selling data, cross data)

### Transverse
**IT/Infrastructure**
- Technical expertise: Hadoop, ...

---

**Preparation of workshops**
- WS#1
- WS#2
- WS#3
- WS#4
- WS#5
- WS#6
- WS#7
- WS#8
- WS#9
- WS#10

---

**Customer Experience Management**
**Social CRM**
**Predictive Maintenance**
**Social Listening**
**Data API offer**
**Open Data**
**B2B2C services**
**Data selling**
**Data reports selling**
**Hadoop Expertise**
**Analytics services**
**Target Big Data reference architecture for Telco**
**Privacy Management**

---


Copyright © Capgemini 2014. All Rights Reserved
Data Driven Customer Experience

- Data Driven Customer Experience in Automotive

- Social profiling
- Micro-targeting
- PROFILING & TARGETING

- Demand Prediction
- Price Prediction
- PRICING & DEMAND PREDICTION

- Repurchase Impulse
- Franchise Dealer Loyalty
- LOYALTY

- Analytical Solutions
- OEM functions
- Importers
- Dealers

- SERVICE DELIVERY
- INNOVATION

- Analytics Service Unit

- Retail
- Vehicle
- Connected Car

- Data integration
- Online sales
- Social
- External data providers

- Intensity of relationship along customer lifecycle
- Time
- Prospect
- New customer
- Regular customer
- Loyal customer
- Churning customer
- Win-back
- Retention
- Won-back customer
- Closing the loop
- Keeping the dialogue
- Starting the dialogue
- Acquisition
- Consolidation
- Intensification
- Churn prevention
- Conquest
- Loyalisation
- Lost customer
- After Sales

- Analytical Solutions
- Retail
- Online sales
- Social
- External data providers

- Connected Car
- OEM functions
- Importers
- Dealers
Predictive Modeling for Corporate Planning and Forecasting

Predictive Sales Forecast Model Based on External Market Indicators

**Situation**
- Client had significant issues with the accuracy of sales forecasts
- This led to credibility concerns when providing guidance to equity analysts and in business decision making

**Objective**
- Client desired a demand forecast model based on external market indicators
- The goal was to improve ability to derive insights and provide longer-term visibility into strategic direction of key markets

**Capgemini Approach**

1. Interviewed Business Stakeholders to Define Demand Drivers

   - End Market Applications
     - Water Heaters/Tankless
     - Barbecues & Grills
     - Oven Interiors
   - Potential Indicators
     - Per capita disposable income
     - Demand for semi-durable goods
     - Consumer Confidence Index
   - Sample Indices
     - IFR Microwaves
     - IFR Non-Durable Goods

   Matched demand drivers to market indices

2. Assessed Data Providers and Recommended Data Vendor

   - Compared data providers to select vendor

3. Ran Multivariable Regressions

   - Synthesis of historical sales and economic data
   - $Y = f(x)$

4. Developed Predictive Forecast Formulae and Values

   - Generated granular and consolidated forecast views
Capgemini’s Data Science & Analytics as a Service provides many options for identifying and realizing value from your enterprise data. These services range from providing seasoned data scientist with deep data exploration and visualization experience within specific industries, to providing management consultants capable of defining effective data monetization strategies and value realization roadmaps.

Since enterprise information management is critical to enabling an efficient data monetization strategy, Data Scientist Insights can also assist you in growing your existing enterprise architecture and information services to be inclusive of those capabilities needed for enterprise-level, and data science-based value generation. Due to our deep expertise in extracting value from disparate data sources (enterprise, IT, and social), proven enterprise architectural capabilities, and ability to program manage the integration of new complex capabilities into existing organizations, we are able to provide an end-to-end source for realizing tomorrow’s value from today’s data.

---

**Define Objectives & Levers**
- Define Business Objectives
- Assess Current Analysis Situation
- Identify Casual Levers
- Produce Modeling Plan

**Data Inventory Collection & Understanding**
- Collect Initial Data
- Summarize and Describe Data
- Explore Data and Produce Report
- Verify Data Quality

**Data Prep Selection & Cleansing**
- Define Data Selection Rational
- Select Data
- Determine Data Causality
- Clean Data
- Integrate and Format Data

**Modeling Design & Development**
- Select Modeling Techniques
- Generate Test & Validation Designs
- Build Model Parameters

**Simulation, Optimization, & Evaluation**
- Integrate Individual Models
- Simulate and Optimize Ensemble Space
- Evaluate Results
- Approve Models

**Deployment Solution within the System**
- Identify Deployment Environment
- Create Map-Reduce Plan or Model Integration Strategy
- Plan Model Deployment, Monitoring, & Maintenance

---

"Data Science is the organic and systematic practice of transforming hypotheses and data into actionable predictions"
About Capgemini

With more than 130,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2013 global revenues of EUR 10.1 billion. Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.