ANALYTICS DATA ARCHITECTURE FOR HIGH PERFORMACE ANALYTICS

Marc Smith
Principal Solution Architect, SAS
Presenting to Alberta Data Architects, Wednesday, May 14
marc.smith@sas.com or @canmjs
GOOGLE TRENDS

“ANALYTICS”

Internet of Things
Big Data
Hadoop
Analytics
“BIG DATA”
"HADOOP"
"INTERNET OF THINGS"
INTERNET OF THINGS
"Big data is what happened when the cost of storing information became less than the cost of making the decision to throw it away."

George Dyson
TWO ERAS . . . TWO MINDSETS

SCARCITY MINDSET
- Technology constrained
- Process-centric
- Focus on cost control
- Everything is forbidden unless it is permitted

ABUNDANT MINDSET
- Technology empowered
- Discovery-centric
- Focus on value
- Everything is permitted unless it is forbidden
CUSTOMER USE CASES . . . SHOW ME THE MONEY

- Leverage all sensor data for improved asset reliability
- Evaluate Risk & Fraud in Real-time
- Automate network configurations dynamically
- Drive evidence-based patient care
- Personalize product & content recommendations
CAN YOU SCALE?

DO YOU GROW A CULTURE OF INNOVATION?

WILL DATA BE GOVERNED?

CAN YOU IMPROVE INFRASTRUCTURE?

WILL YOU REDUCE COST?
SAS ANALYTICS

THE ANALYTICS LIFE CYCLE

BUSINESS MANAGER
Domain Expert
Makes Decisions
Evaluates Processes and ROI

IT SYSTEMS / MANAGEMENT
Model Validation
Model Deployment
Model Monitoring
Data Preparation

IDENTIFY / FORMULATE PROBLEM
DATA PREPARATION
DATA EXPLORATION
TRANSFORM & SELECT
BUILD MODEL
VALIDATE MODEL
DEPLOY MODEL
EVALUATE / MONITOR RESULTS

BUSINESS ANALYST
Data Exploration
Data Visualization
Report Creation

DATA MINER / STATISTICIAN
Exploratory Analysis
Descriptive Segmentation
Predictive Modeling

Copyright © 2012, SAS Institute Inc. All rights reserved.
REQUIRES ENTERPRISE ARCHITECTURE APPROACH TO ANALYTICS

BIG DATA ANALYTICS

GRID COMPUTING

IN-MEMORY

IN-DATABASE

EDW

ADW

LINE OF BUSINESS

PROGRAM MANAGER

FINANCE

SUPPLY CHAIN

I.T.

HR
HADOOP WHY SHOULD WE CARE?
Analytic Data Architecture

- **Raw Data**
- **Summarized, Shared Data**
- **Analytic Base Tables**

Business Glossary (Governance)

---

Analytic Technical Architecture

- **Data Management Cluster / Appliance**
- **SAS Client Application**

10Gb Ethernet

- **High Performance Analytics Cluster / Appliance**
- **Shared Disk**