Big Data for Government Symposium
http://www.ttcus.com

Linkedin/Groups:
Technology Training Corporation

@TECHTrain
Operational Challenges and Considerations in Large Scale Data Analytics

Hemanth Setty, Chief Technology Officer
1962, a computer pioneer said, "There is no reason to suppose the average boy or girl cannot be master of a personal computer." Soon after, PCs arrived and have kept changing and changing.
As Humans We Need More and More Data

Worldwide internet traffic, 1990-2020 PB/month

- WWW is born
- Digital decade
- 17x Growth 2008-2020
- +27% 2008-2020 CAGR

Forecast Model

MACHINE TO MACHINE?

+32% Video
+20% Non-video
The Hype Circle – Data is the CORE

Cloud, Mobile, and Information Assurance are Enablers of Data.
Explore Data In It’s Simplest Form

Structured Data

Data Residing in Databases

Unstructured Data

Documents, Websites, Emails, etc.
Quiz...

• Today should we care if data is structured or unstructured?
• Can we ask intelligent questions across structured and unstructured data?
Quiz...Answers

• Today should we care if data is structured or unstructured?
  – Answer: No

• Can we ask intelligent questions across structured and unstructured data?
  – Answer: Yes
Operational Challenges to Unlocking the Power of Big Data

“The Big Data Gap,” reveals that Federal IT professionals believe big data can improve government but that the promise of big data is locked away in unused or inaccessible data. *Meritalk 2012*

- **People**
- **Technology**
- **Process**

**Fresh Minds with New Approaches**
- New Skills

**Business Drives Technology**
- Define the Questions to Ask of the Data

**Business Intelligence Software**
- Cloud
- Big Data Solutions
- Mobile
RATB Business Intelligence Portal Project
Five Step Agile Approach to Kick Start
Business Intelligence / Analytics

(1) Business Requirements Analysis
- Who are your customers?
- What questions need answers?
- Are there privacy concerns with answering the questions?

(2) Data Discovery
- What data exists?
- Who are the data owners?
- What is the refresh cycle?
- What is the unverified data quality?
- What are the data retrieval alternatives?

(3) Backend Development
- Develop queries
- Validate data quality
- Develop logic to increase data quality

(4) UI Requirements
- Conduct Focus Group Analysis
- UI Solution Alternatives
- Build UI working with Focus Groups

(5) BI Development
- Build UI working with Focus Groups

Three Dimensions to Data Analytics

**Portal**

- Pre-Defined Queries Delivered via BI / BA
- "Quick Answers to Complex Questions"
- "Web Interface" – With Access Controls
- Minimal Ad-hoc Business Intelligence

**Self Service Business Analytics**

- Remove Specialist Tools From Equation
- "Comprehensive, Light Weight, BI / BA Solution"
- Rapid Access to Enterprise Data
- Ad-hoc Pivot Table, Chart, Geospatial

**Detailed Analysis**

- High End Analytic Tools
- Rapid Access to Enterprise Data
- Specialized Experience Required
Three Dimensions to Data Analytics

Quick Hits

Self Service
Business Analytics

Advanced Data Management System

Detailed Analysis

Office

ESRI

Microsoft Office SharePoint

Office PerformancePoint Server 2007

SQL Server 2014

Highcharts JS

ESRI

Palantir

uReveal

Oracle

ENDECA
High Level Logical Architecture
Advanced Data Management System

TIER 1
Display Analytic Layer / Data Federation

TIER 2
Data Layer

Data Automation
"SAP DATA AUTOMATION"

Unstructured Data Automation

Data Types
Resource Data – Entities Doing Business with the Federal Government
Risk Data – Supporting Bad Actors Risk Framework
Operational Challenges and Considerations in Large Scale Data Analytics

Questions?

Hemanth Setty, Chief Technology Officer