Unleash the Power of Big Data

Presented by Matt Ziegler
Key Trends

- **Device Explosion**: >5.5 billion (70+% of global population)
- **Ubiquitous Connection**: Web traffic
  - 2010: 130 Exabyte (10 E18)
  - 2015: 1.6 ZettaByte (10 E21)
- **Social Networks**: >2 Billion users
- **Sensor Networks**: >10 Billion
- **Cheap Storage**: $100 gets you 3 million times more storage in 30 years
- **Inexpensive Computing**: 1980: 10 MIPS/$
  - 2005: 10M MIPS/$
Insight

Time Series

Relational

Unstructured
Real-time Data isn’t perfect

The Truth about Real-time Data

- Naturally incomplete
- Doesn’t look like SQL (unevenly spaced, no transactions)
- Subject to errors in measurement
- Varies in fidelity
Decision-Ready Data
Big Data and the PI System

Data Warehousing

Statistical Analytics

Visual Analytics

Bring calculations to the data

Systems of engagement vs systems of record

Identify the conversation
Data Warehousing
PI Connectors and Integrators

In Context

Auto-configured w/ Events w/ Assets
Visual Analytics
PI Infrastructure for Visual Analytics

Connect via Internet or PI Interface

Visual Analytics – BI Tools
- Excel Power BI
- Spotfire
- Tableau
- Qlikview

In Memory DB

CAST
- Aggregate / Augment
- Transmit
- Cleanse
- Shape

Data

PI System: System of Record / Real-time Analytics

PI Data Archive
- Assets / Events
- Asset Based Analytics

Business User

Operational User

© Copyright 2014 OSIsoft, LLC.
Enabling the Smart Grid

Conservation Voltage Regulation (CVR)

ANSI C-84.1 → 114 – 226V

- Utilities operate at the high end of range
- Potential 3% continuous energy savings
- 6,500 MW*Years (56.9 MM MW*hrs)

Violation defined as 5 consecutive reads under 115V

Grand Coulee Dam: #1 US Producer
Direct Visual Analytics
Statistical Analytics
PI Infrastructure for Statistical Analytics

Information → Insights: Complex Analytics

- Cleansing
- Mining Modeling Learning

Data CAST

- Aggregate / Augment
- Transmit
- Cleanse
- Shape

PI System: System of Record / Real-time Analytics

- PI Data Archive
- Assets / Events
- Asset Based Analytics

Connect via Internet or PI Interface

DATA

Business User

Operational User

DATA
Steam Cycle Statistical Analytics

File dependencies to include with job:
[Auto-detected] MapReduce.exe
[Auto-detected] Microsoft.HadoopMapReduce.dll
14/03/20 22:11:27 INFO streaming.StreamJob: C:\bin\hadoop job -Dmapred.job.tracker=jobtrack_0038
14/03/20 22:11:28 map 0% reduce 0%
14/03/20 22:12:10 map 10% reduce 0%
14/03/20 22:13:59 map 50% reduce 0%
14/03/20 22:15:47 map 90% reduce 0%
14/03/20 22:16:11 map 99% reduce 0%
14/03/20 22:16:17 map 100% reduce 0%
14/03/20 22:17:48 map 100% reduce 17%
14/03/20 22:17:54 map 100% reduce 33%
14/03/20 22:17:57 map 100% reduce 67%
14/03/20 22:21:01 map 100% reduce 90%
14/03/20 22:22:25 map 100% reduce 100%
14/03/20 22:22:33 Job complete: job_2014031006;
14/03/20 22:22:33 Output: asv://energy@blobtes;

Job Completed [0] in 701.993 sec.
Reading Results...
Reading Results... Done.
Calculating.....
Calculating...... Done
Cluster Analysis
Conclusions

- Need Higher Fidelity Data
- Change my model
- Add more data
- Add facets (time of day, temperature, coal quality)

High Pressure Steam does more Work
Call to Action

• Whisper Suite (SAP HANA, HDFS, Power BI, Office 365)

• Power BI at Product Expo

• MVP or Beta registration
THANK YOU