

Tools for Big Data Analytics Focus Group Report Out

Smart Grid Big Data Workshop

Users of Analytics and Scenarios

- Multiple beneficiaries for analytical scenarios
 - Trade off decisions should involve grid, building, business process, and human aspects
- Extreme and rare event prediction and detection
 - Economics related to these events
- Scenarios to integrate consumer choices and human behavior in real-time into decision-support systems
- An end-to-end scenario involving all subsystems does not exist
 - Instead multi-level analytical systems that can interoperate
 - Distribution grid analysis in near real-time
 - Scada systems for monitoring and stability testing
 - Control systems for event detection and resilience
 - Timescale and extent to precompute/real-time compute on different analytical subsystems

Challenges and BD Technology Needs

- Advancing analytic capabilities in operations
 - How to utilize machine intelligence, integrate exploratory research products
 - Nimble systems for dynamic data-driven operational insights to action loop
- Grid analysis and control in near real time, latency of communication
 - Data compression from the edge to distributed processing systems for analytics and grid management
 - Streaming and in memory computing capabilities
- Extreme event characterization and prediction
- Interoperability of the interaction between subsystems
 - Ability to combine the legacy systems with big data systems
 - Spatiotemporal resolution and connectivity issues
 - Common information models needed
 - Standards to send data out of subsystem boundaries via reusable and secure services
- Smart data capture
 - With a purpose and model in mind
 - Need to collect and preserve long-term data for understanding usage patterns over the years
 - Provenance of aggregate, cleaned and approximated data
 - Data preparation needs to be automated

What does it take for an exploratory research product to be integrated into smart grid workflows?

- Lessons learned bulletins (from ?) - a recent one is on a shut down notice to generators, but many of them should not be sent out. We don't have the tools now to avoid wrong signals before showing them to operators as actionable insights in real-time.
- Some of the data that can be useful for others over time does not get out as reusable services.
- Finding out analytics products that can enrich data systems in the smart grid. This comes down to finding the tools to collect and integrate such products.
- What we learn from the dirty data and the systems related to quality-evaluation/stamping of the data should be made a part of the analytics and learning ecosystem.
- Finding ways to send out data from the usual system boundaries. Designing systems around purposeful data capture.