Suggested Strategy for Exploring Big Data Standards Opportunities using NIST Use Cases

Bob Marcus
robert.marcus@et-strategies.com
Suggested Strategy Steps

• Map collected NIST Big Data Use Cases to the NIST Reference Architecture. Abstraction may be necessary to provide value beyond a specific application domain

• Use these mappings to drive drill-downs in the Reference Architecture and describe interfaces between components

• Determine opportunities for standardization of interfaces based on analysis of Use Cases

• Solicit feedback on the expanded Reference Architecture and possible standardizations from people with experience implementing Big Data projects (e.g. Use Cases)

• Create exploratory implementations to evaluate open issues involving interfaces and possible standardization

• Modify deliverables based on feedback and exploratory implementations
Example of Strategy using 10 Simple Generic Use Cases

See NIST Big Data Document M0297
http://bigdatawg.nist.gov/_uploadfiles/M0297_v2_8205832946.doc
Ten Simple Generic Use Cases

- Multiple users performing interactive queries and updates on databases
- Perform real-time analytics on data streams and notify users when specified events occur
- Extract data from external data sources and load and transform in horizontally scalable data store,
- Perform batch analytics on the data in a highly horizontally scalable data store
- Perform interactive analytics on data in analytics-optimized database
- Visualize data extracted from horizontally scalable Big Data store
- Move data from a horizontally scalable data store into a Data Warehouse
- Extract, process, and move data from data stores to archives
- Combine data from Cloud and on premise data stores for analytics
- Orchestrate multiple sequential and parallel data transformations and/or analytic processing using a workflow manager
Possible Standards Suggested by Mappings

- Interfaces to databases (e.g. SQL queries to Hadoop)
- Representation of events
- Standard descriptions of transformations (e.g. JSON and/or XML to tables)
- Interface between analytics and data processing. Metadata to support data access and data flow for diverse analytic interfaces.
- Standard descriptions of database schemas and configurations to support analytics.
- Query language for visualization data caches
- Transformations from unstructured data to Data Warehouse relational schema.
- Standards for describing archival data store schemas and possibly generating ETL, curation, and packaging code
- Standard representations for data sets (e.g. schemas, data types, semantics)
- Declarative format for workflow across multiple operations.
General Areas for Possible Big Data Standards

- Standard interfaces to data stores for end-user queries, analytics applications, and visualization tools e.g. (SQL to HDFS, JSON, Arrays)
- Standard metadata descriptions of data for extraction, transformation, and/or loading
- Standard pattern descriptions for event processing and declarative queries
- Standard descriptions of workflows for system orchestration
Possible Next Steps using Geoffrey Fox’s Work

- Use Geoffrey Fox’s HPC-ABDS as a drill-down from the current NIST Reference Architecture (See M0297)

- Map Ogre Facets presented by Geoffrey Fox to the drilled-down NIST Reference Architecture

- Note at a high level, the Ogre Facets map to the Reference Architecture:
  - Problem Architecture -> System Orchestration
  - Data Sources -> Platform Provider and/or Data Sources
  - Core Analytic Kernels/Mini-applications -> Application Provider

- Use the drilled-down mappings to define interfaces and explore possible standards in a similar fashion as the ten Generic Use Cases