Evaluation Criteria for Reference Architecture
Criteria for Evaluating Reference Architecture

- These are my personal suggestions
- Evaluation of Reference Architectures should be based on pre-defined criteria
- Without criteria, it will be hard to determine best Reference Architectures
- These criteria should include possible mappings to other subgroups deliverables
- Another criterion should be relative harmony with existing Reference Architectures
- The draft brainstorming Reference Architecture meets these criteria
Brainstorming Big Data Reference Architecture

- **Applications and User Interfaces**
  - **Real-time Analytics and Interfaces**
  - **Interactive Analytics and Interfaces**
  - **Batch Analytics and Interfaces**

- **Infrastructure**
  - **Data Sources**
  - **Stream Processing and ETL**

- **Data Processing**
  - **Foundation Data Store**

- **Analytics Database**

- **Operational Databases**
  - **In Memory or SSD Operational Databases**

- **Interactive Analytics and Interfaces**
  - **Real-time Analytics and Interfaces**

- **Security**
  - **Process Management**
  - **Data Resource Management**

- **Design, Develop, and Deploy Tools**

- **Systems Management**

*Wednesday, July 10, 2013*
Criteria Examples

- Taxonomy subgroup should be able to define and extend Architecture Components
- Requirements subgroup output should be mappable to Architecture Components
- Use Case Actors should be mappable to Reference Architecture
- Roadmap subgroup gap analysis should be mappable to Architecture Components
- Big Data technology stacks (e.g. Apache) should be mappable to Architecture
- The final slide shows how Apache components fit into brainstorming Architecture
Apache Big Data Framework in Reference Architecture

Real-time Analytics and Interfaces
- Camel RX

Interactive Analytics and Interfaces
- Drill, Solr, Tez, MRQL

Batch Analytics and Interfaces
- Pig, Hive, Mahout, Giraph

In Memory Databases
- Gora

Operational Databases
- Accumulo, CouchDB, HBase, Cassandra,

Stream Processing and ETL
- S4

Data Sources
- Kafka

Flume, Sqoop

Infrastructure
- CloudStack

Data Processing
- Map-Reduce, Yarn, Hama

Foundation Data Store
- HDFS

Applications and User Interfaces
- Thrift, Pivot

Design, Develop, and Deploy Tools
- Cascading

Security
- Knox

Process Management
- Zookeeper, Oozie, Mesos

Data Resource Management
- HCatalog, Falcon

Systems Management
- Ambari