BLUE WATERS

SUSTAINED PETASCALE COMPUTING

3/1/19

Machine Learning with Blue Waters Monitoring Data

Status as of 3/1/19

NCSA

GREAT LAKES CONSORTIUM
FOR PETASCALE COMPUTATION

CRAY

Main Components

- Monitoring Data
- Datastore
- ML Models for Analysis

Monitoring Data

- Starting with OVIS metric data
- Doing analysis on all (most?) data simultaneously

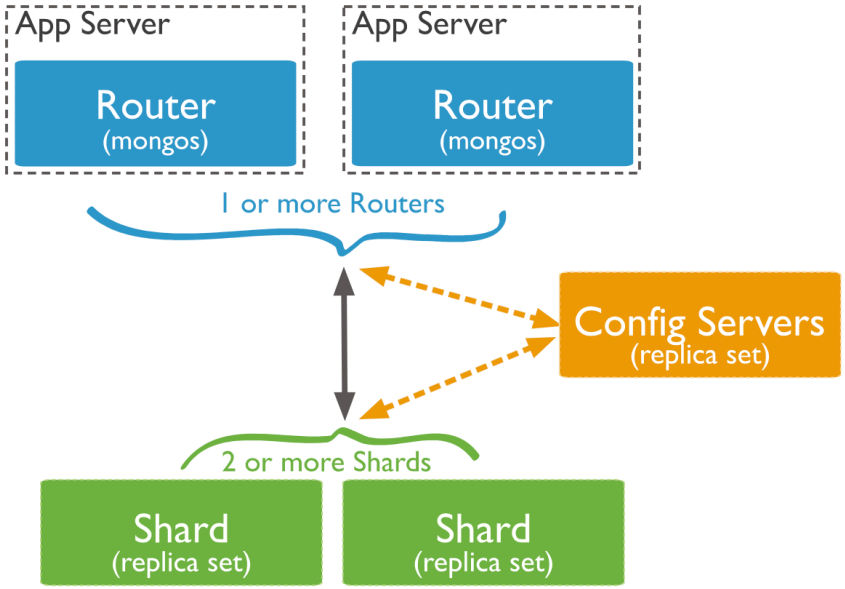
- BW Team has studied the monitoring data extensively
  - Frequency of collection
  - Transport mechanisms
  - Important metrics

Datastore

- Data science pipeline
  - Descriptive, Prescriptive, Predictive
- Flexible Query Engine to support pipeline
- Distributed datastore for scalability
- MongoDB deployed as a batch queued HPC job

MongoDB on BW Hosting Metric Data

- MongoDB Components
  - Config Server
  - Router – Hosts Query
  - Shards – Hosts Data
- One component per processing element
- Studying component balance to optimize ingest and queries



ML Model

- Sequence Modeling
  - LSTMs, Novel CNNs, Autocorrelation
- Unsupervised Auto Encoding Schemes
  - Clustering
- Job Classification, Anomaly Detection

Status Summary

- Monitoring data has been established as the prototype testbed
- Version 1 of MongoDB is on BW is built and being studied
- Prototype model created and briefly studied on “adjacent problem”