XRootD Popularity on Hadoop Clusters

- Dataset (DS) popularity is very important to CMS operations

- Current implementation in Oracle has scalability issue

- Migration to Hadoop, in harmony with CERN IT strategy

- DS popularity on Hadoop scales with data volume

  - Oracle vs Hadoop performance

  - Pig and Spark implementation

Prediction of Dataset Popularity

- Reactive vs Proactive approach

- First attempt on Big Data architecture

  - Training of several Classifiers

  - Compare performance

  - Calculate Predictor's accuracy