**Components of NSF 144305**

- Design and build SPIDAL – a Scalable Parallel Interoperable Data Analytics Library
- Domain specific libraries – mainly from project
- Core Machine Learning Libraries
- High Performance for Java and MIDs
- **NIST Big Data Analysis** Analysis: Features of data intensive Applications deriving 50 Ogres and 64 Convergence Diamonds
- HPC-ABDS: Cloud-HPC Interoperable software with performance of HPC and rich functionality of commodity Apache Stack
- Implementations: HPC and Clouds with DevOps

**HPC-ABDS Apache Big Data Stack**

- Programming model and runtime language: Java
- Distributed computing framework: Apache Hadoop
- Big Data storage: HBase, Cassandra, MongoDB, HBase, Cassandra, MongoDB
- Processing: Apache Spark, Apache Flink, Apache Storm, Apache Beetle
- Machine Learning: Mahout, Weka, H2O, R
- Cloud computing: Amazon Web Services, Google Cloud Platform, Microsoft Azure
- Security: Apache Knox, Apache Atlas

**Applications in action with MIDAS/SPIDAL**

- Multi-Scale Imaging and Spatial Data Analytics
- MIDAS and Biomolecule Simulations
- Polar Remote Sensing Algorithms
- WebPlotViz - Browser Visualization of High Dimensional Data

**Midas and Spiadal Java High Performance Middleware and Language**


**The Concept of Harp Plug-in**

- **Application**
- **Framework**
- **Resource Manager**
- **MapReduce Model**
- **MapCollective Applications**
- **MapReduce**
- **MapReduce V2**

**Multidimensional Scaling with Flink**

- **Applications**
- **Performance**
- **Comparison**

**DA-MDS speedup for 200K with Different Optimization Techniques**

- **Best MPI**
- **Intel**
- **Intel intra node**

**Multidimensional Scaling with Flink**

- **Performance**
- **Comparison of Flink and MPI**
- **Flink vs MPI**
- **DA-MDS**
- **Benchmark of Flink and MPI**
- **Best MPI**
- **Intel**
- **Intel intra node**