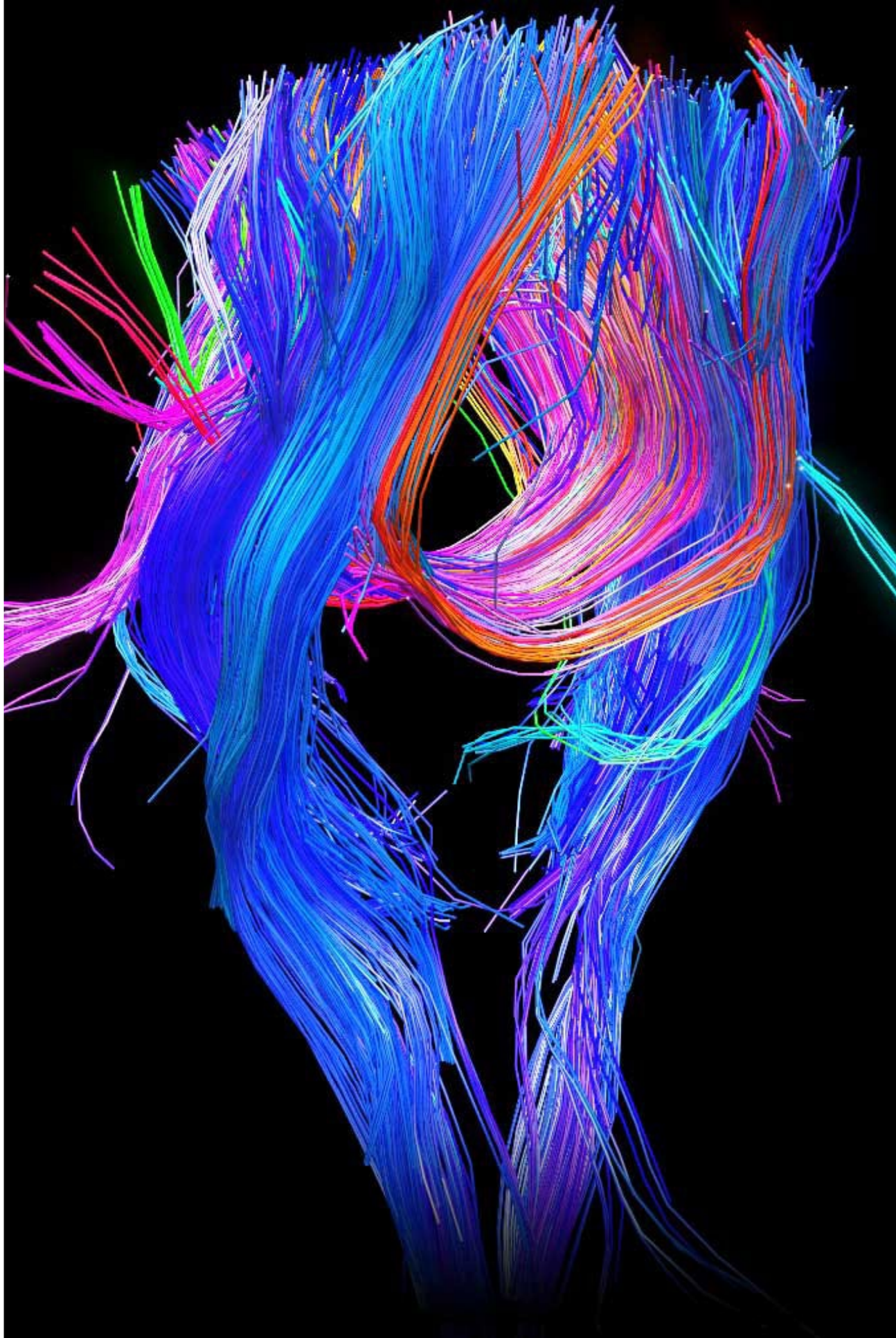
Big Data Approach to fMRI Data Analysis with Intel DAAL and Full Correlation Matrix Analysis

Haoran Shu (CUHK)
Yin Lok Wong (HKU)

Mentors:
Pragnesh Kumar
Kwai Wong
Junqi Yin



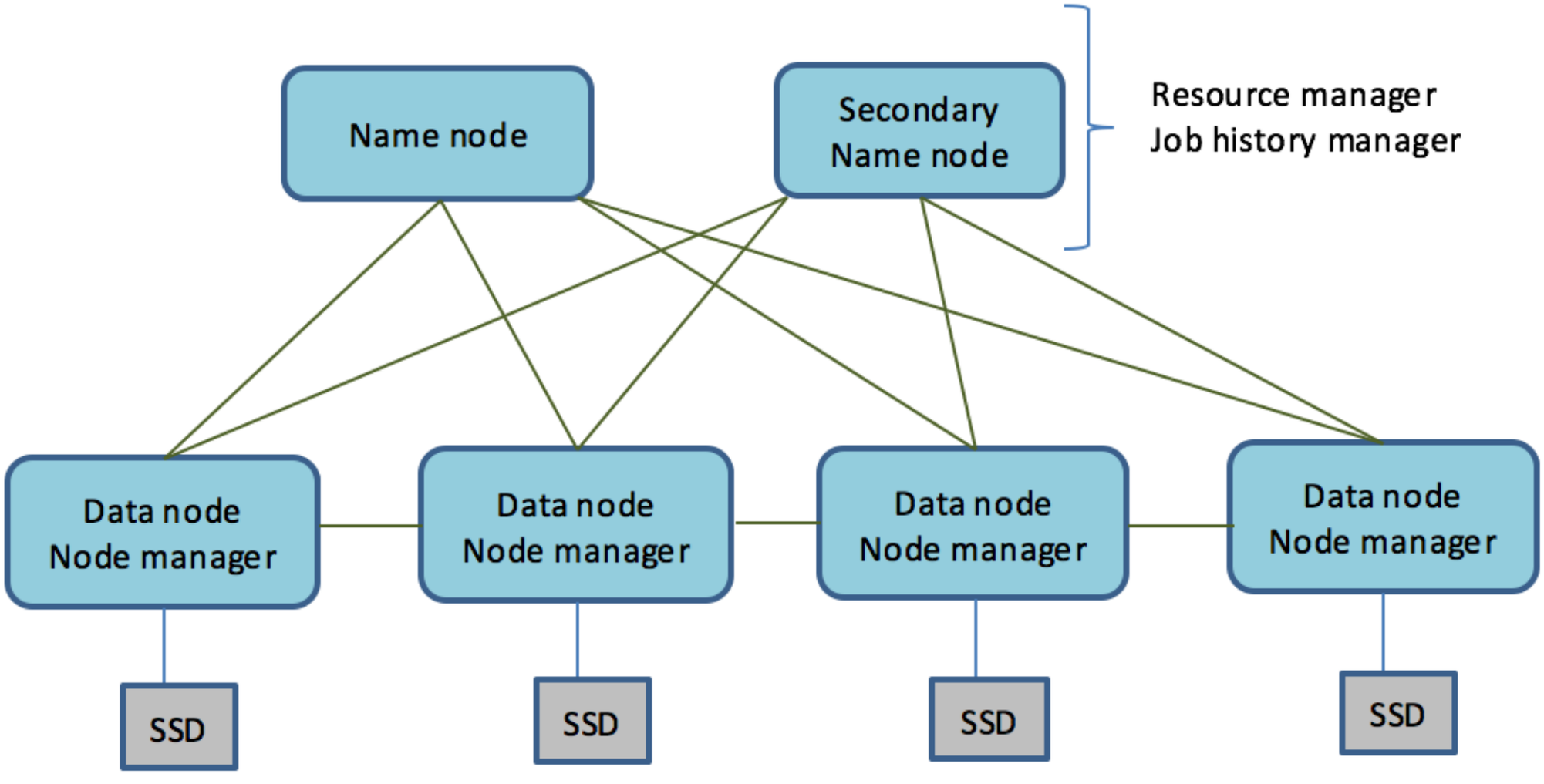
Big Data

fMRI data analysis deals with data represented in large scale matrices and operations numbered easily in Giga magnitude, which breeds an ideal scenario for the use of Big Data framework such as Apache Spark.

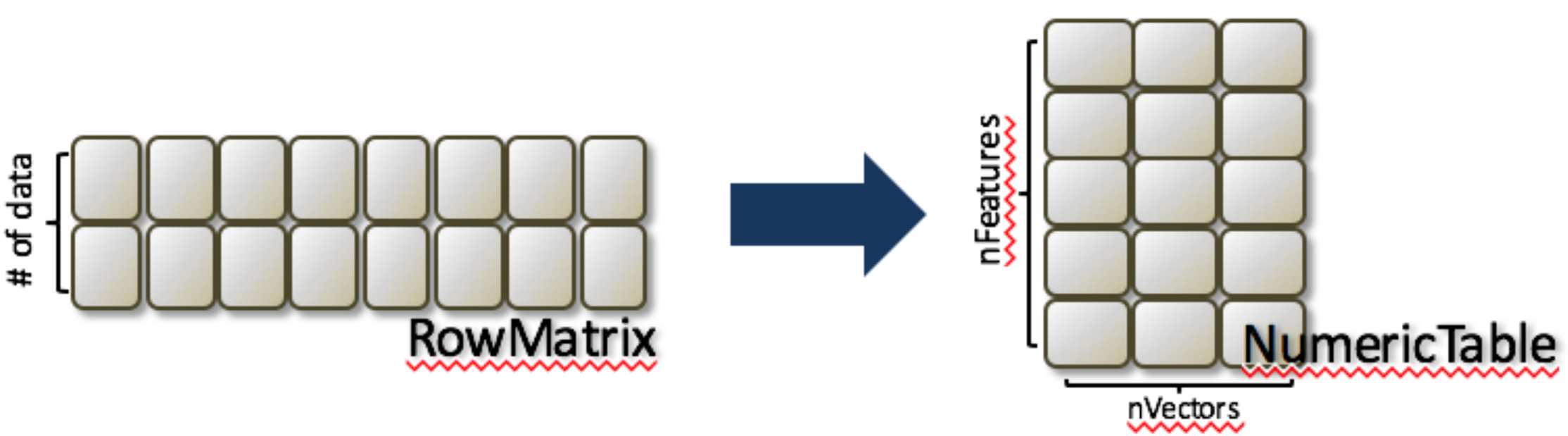
The test data used contains the matrix with 480115 elements.

Biananes - Spark Nifti Reader Library

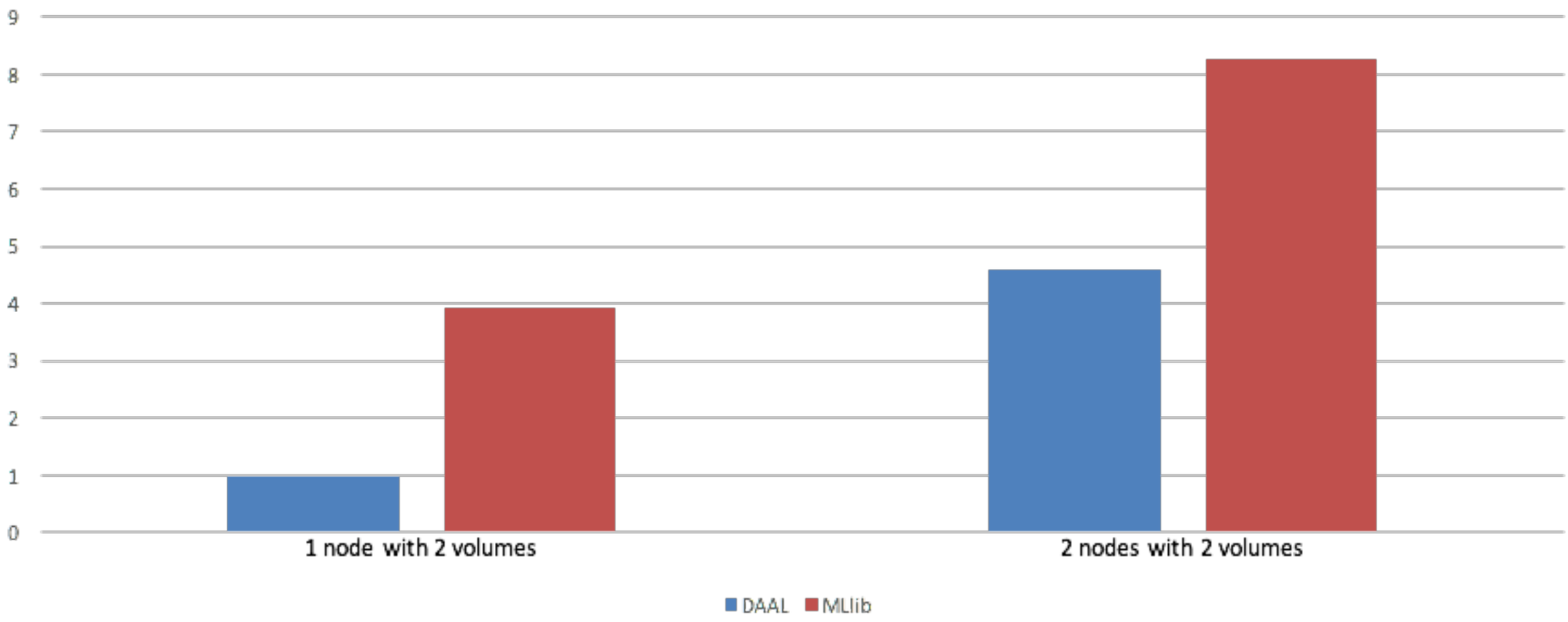
Read Nifti(.nii) to RowMatrix RDD



Incorporate with Intel DAAL



Benchmarking



SVD computation with Nifti input under Spark

Benchmarking

Currently working on further testings with QRDecomposition and expanding tests with more computing nodes and larger data sets.

Computation attempts on Principal Component Analysis and Covariance were taken but failed to draw comparison as with input matrix size larger than 65535, these computations in MLlib are not supported

FCMA

Parallelism

Correlation Computation

- Optimized to a matrix multiplication by normalizing the data
- Paralleled with MPI

Classifier Analysis

- LibSVM used
- GotoBlas to further speed up

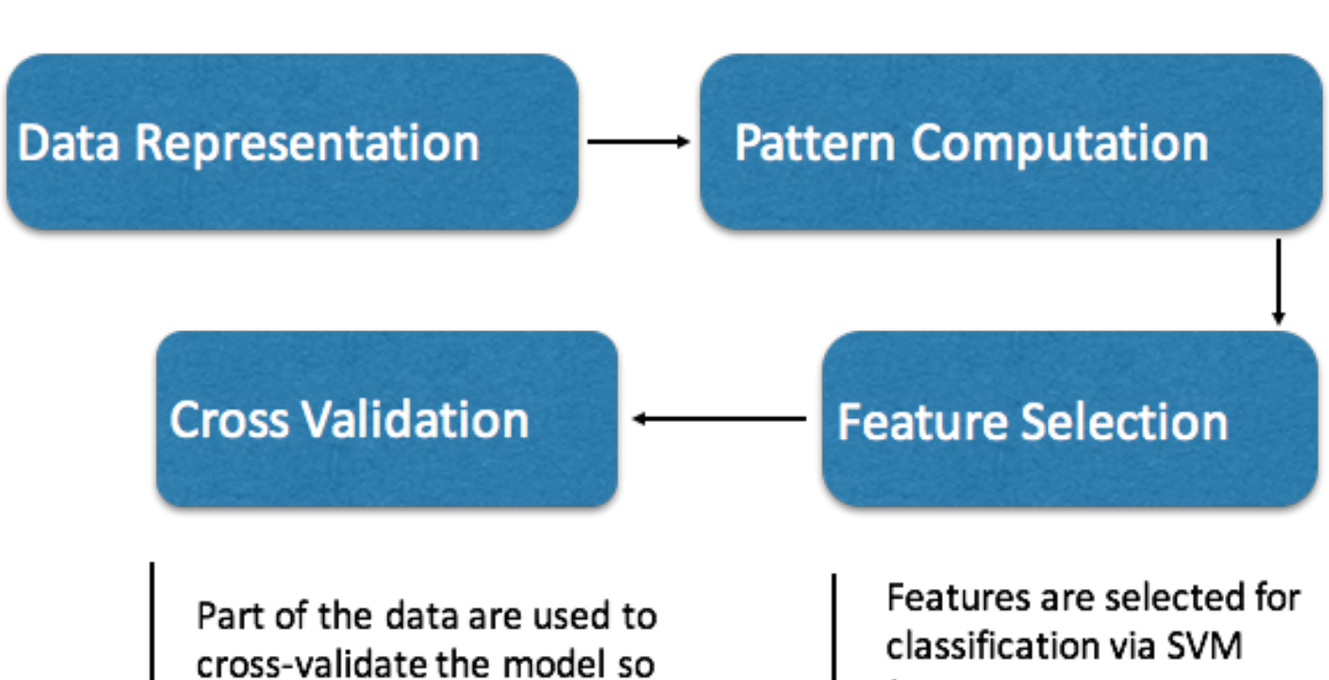
MVPA

- Activity Patterns
- Correlation Patterns

Classification Procedure

fMRI images are represented by matrices for each voxel at different epochs

Activities or *Correlations* are computed according to the MVPA pattern used



Part of the data are used to cross-validate the model so as to pinpoint the best parameters

Features are selected for classification via SVM (Supporter Vector Machine)