


Report 2

Project Name	Spark Backend Support for Gora (GORA-386)
Project URL	
Report #	2
Report compiled by	Furkan KAMACI
Report date	1 June 2015

Project Description

Apache Gora open source framework provides an in-memory data model and persistence for big data. Gora supports persisting to column stores, key value stores, document stores and RDBMSs, and analyzing the data with extensive Apache Hadoop MapReduce support. Even Spark is so powerful compared to Map/Reduce which Gora currently supports; there is no Spark backend for Gora. This proposal aims to develop a solution for Gora to support Spark.

Checklist

Berkeley's first paper for Spark is analyzed (http://www.cs.berkeley.edu/~matei/papers/2010/hotcloud_spark.pdf).

Berkeley's RDD paper is analyzed (https://www.cs.berkeley.edu/~matei/papers/2012/nsdi_spark.pdf).

Spark documentation and programming guide is followed.

Examples are developed to analyze RDD.

Apache Gora and Spark integration is analyzed.

Next Steps

Analyze Hbase and Spark integration.

Supervisors Comments