Spark Shared RDD

Cluster Setup
Spark provides three types of cluster set-up: standalone configuration, Mesos integration and YARN. Ignite should be able to start alongside spark using all three ways.

- Standalone cluster setup. Spark provides a script spark-class that starts WorkerProcess. As a part of the integration we should provide a script that will start Ignite node together with spark workers.
- Mesos. TBD
- YARN. Spark provides internal Hadoop Job that starts Spark application processes on all YARN nodes instead of spark workers. We should provide a similar Hadoop job that will be able to start Ignite nodes on task nodes.

Reading Data From Ignite
Ignite should provide the following RDDs:

Cache iterator RDD
This RDD can be properly partitioned and collocated with Ignite nodes. Cache name or optional cache configuration should be passed to construct an RDD so that user has an ability to create caches on the fly. User also may specify a predicate that is passed to ignite scan query.

Cache SQL/Fields query iterator
This RDD is not partitioned and should be parallelized by Spark if necessary. Cache name and SQL clause should be passed to construct an RDD.

Saving Data To Ignite
Utility object that takes any Spark RDD and stores it to Ignite using Streamer.