GettingStarted

Getting Started with MRQL

Installation Instructions

Apache MRQL can run in 4 modes: in Map-Reduce mode using Apache Hadoop, in BSP mode (Bulk Synchronous Parallel mode) using Apache Hama, in Spark mode using Apache Spark, and in Flink mode using Apache Flink.

The latest stable MRQL version, MRQL-0.9.6-incubating, is compatible with the following Apache releases:
The MRQL MapReduce mode is compatible with Apache Hadoop releases 1.x and 2.x (Yarn). You can download the Hadoop tarball from Apache Hadoop. The BSP and Spark modes are optional. The BSP mode is compatible with Apache Hama 0.6.2, 0.6.3, 0.6.4, and 0.7.0. You can download the latest tarball from Apache Hama. The Spark mode is compatible with Apache Spark 1.0.0 through 1.6.0. You can download the latest tarball prebuilt for Hadoop1 or Hadoop2 from Apache Spark. The Flink mode is compatible with Apache Flink 0.10.1 and 0.10.2 in local and Yarn modes. You can download the latest tarball prebuilt for Hadoop2 from Apache Flink.

The following instructions assume that you have already installed Hadoop MapReduce and you have deployed it on your cluster successfully.

How to install MRQL

Download the latest stable MRQL binary release from http://www.apache.org/dyn/closer.cgi/incubator/mrql and extract the files. The MRQL 0.9.6 binary release uses Hadoop 2.7.1 (Yarn), Hama 0.7.0, Spark 1.6.0, and Flink 0.10.2 The scripts `bin/mrql`, `bin/mrql.bsp`, `bin/mrql.spark`, and `bin /mrql.flink` evaluate MRQL queries in Hadoop, Hama, Spark, and Flink modes, respectively.

How to run MRQL on a Hadoop **MapReduce** cluster:

Change the configuration file `conf/mrql-env.sh` to match your Hadoop installation. For a test, run the PageRank example or the k-means clustering example on a small Hadoop MapReduce cluster.

How to run MRQL on a Hama cluster:

* Follow the instructions in Getting Started with Hama to set up and start Hama.
* Change the configuration file `conf/mrql-env.sh` to match your Hama installation.

For a test, run the PageRank example or the k-means clustering example on a Hama cluster.

How to run MRQL on a Spark standalone cluster:

* Follow the instructions in Spark Standalone Mode to set up and start Spark in standalone deploy mode.
* Change the configuration file `conf/mrql-env.sh` to match your Spark installation.

For a test, run the PageRank example or the k-means clustering example on a Spark cluster.

How to run MRQL in Spark mode on a Yarn cluster:

Set `SPARK_MASTER=yarn-client` in `conf/mrql-env.sh` (see Running Spark on YARN).

How to run MRQL in Flink mode on a Yarn cluster:

Change the configuration file `conf/mrql-env.sh` to match your Flink installation. Then run the PageRank example or the k-means clustering example using the `bin/mrql.flink` script.

How to Recompile MRQL

Download the latest stable MRQL source release from http://www.apache.org/dyn/closer.cgi/incubator/mrql and extract the files. You can get the latest source code using:

```
git clone https://git-wip-us.apache.org/repos/asf/incubator-mrql.git
```

To build MRQL using maven, use

```
mvn clean install
```

To validate the installation, use

```
mvn -DskipTests=false clean install
```

which runs the queries in `tests/queries` in memory, local Hadoop mode, local Hama mode, local Spark mode, and local Flink mode.

Currently, the "mvn install" in MRQL 0.9.6 builds MRQL using Hadoop 2.7.1 (Yarn), Hama 0.7.0, Spark 1.6.0, and Flink 0.10.2. To build MRQL on Hadoop 1.x, such as 1.0.3, use:

```
mvn -Dhadoop1 -Dhadoop.version=1.0.3 clean install
```

To build MRQL on another Hadoop 2.x (yarn) use:

```
mvn -Dyarn.version=2.2.0 clean install
```

To build MRQL on cloudera cdh use:

```
mvn -Dcdh -P-yarn -Dhadoop.version=2.5.0-cdh5.3.0 clean install
```