Running Mobius Application

Introduction

Mobius provides C# language binding to Apache Spark enabling the implementation of Spark driver program and data processing operations in the languages supported in the .NET framework like C# or F#.

Architecture

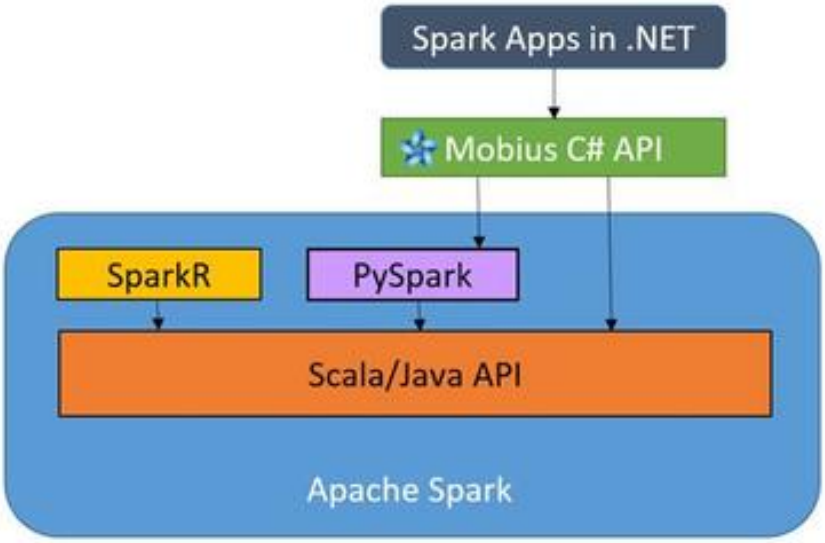


Figure 1: Shows the dependency between the .Net application, Mobius C# API and Spark API

- Mobius application can be developed in any IDE that supports .Net environment (like Visual Studio).
- Mobius applications can be compiled to an executable file and from this Spark job can be submitted. Similarly, job submission can be made in different ways like, Local mode, Debug mode and Yarn Cluster-client mode. There are also other ways available, like Yarn Cluster and Standalone mode, but this hasn't been tested.
- "Microsoft.Spark.CSharp.Adapter" library holds the Mobius API to submit/execute Spark job in .Net application.
- CsharpRunner is an agent come along with the Mobius to bridge communication between Spark and "Microsoft.Spark.CSharp.Adapter" library.

Steps to configure and run Mobius application/shell using Big Data Studio

Step 1: Ensure Pre-Requisites

- Windows Server 2012 or above; or, 64-bit Windows 7 or above.
- Developer Command Prompt for Visual Studio 2013 or above, which comes with .NET Framework 4.5 or above.
- 64-bit JDK 7u85 or above; or, 64-bit JDK 8u60 or above.

Step 2: Download the Mobius

Download the spark-clr_2.11-X.X.XXX.zip file (latest is suggested) available in below location: https://github.com/Microsoft/Mobius/releases

Step 3: Configure Mobius in Big Data Studio

Use attached configured files

Place the following Mobius files/folders in the mentioned (Big Data Studio) location. You can find these files that have been configured already from the attachment (Mobius.zip - *Mobius/Mobius Configuration Files*).

Batch file	**sparkclr-submit.cmd**	*%InstalledDrive%Syncfusion\BigData\X.X.X.XX\BigDataSDK\SDK\Spark\bin*
Batch file	**sparkclr-shell.cmd**	*%InstalledDrive%Syncfusion\BigData\X.X.X.XX\BigDataSDK\SDK\Spark\bin*
Folder	**repl**	*%InstalledDrive%Syncfusion\BigData\X.X.X.XX\BigDataSDK\SDK\Spark*
JAR file	**spark-clr_2.11-2.0.000-PREVIEW-2**	*%InstalledDrive%Syncfusion\BigData\X.X.X.XX\BigDataSDK\SDK\Spark\jars*

Eg. C:\Syncfusion\BigData\3.1.0.59\BigDataSDK\SDK\Spark\bin

Manual configuration changes

If the attached configure files are not helpful, you can manually configure the Mobius. Kindly make the following changes in the "sparkclr-submit.cmd" file to run the Mobius using our Big Data Studio environment.

1. Add these commands before the line 15, as that will throw SPARK_HOME and JAVA_HOME non-set error

```
set SPARK_HOME=%CMDHOME%\..
set JAVA_HOME=%CMDHOME%\..\..\..\Java\jdk1.7.0_51
```

2. Modify the sparkclr*.jar path in the lines 19 and 47 as below.

```
19: if not exist %CMDHOME%\..\jars\spark-clr*.jar (goto :sparkclrhomeerror)

47: set SPARKCLR_CLASSPATH=%SPARKCLR_HOME%\jars%SPARKCLR_JAR%
```

3. Modify the classpath to support YARN-Cluster in the lines 62 and 81.

```
62: "%JAVA_HOME%\bin\java" -Xmx512m -cp
"%HADOOP_CONF_DIR%;%LAUNCH_CLASSPATH%"
org.apache.spark.launcher.SparkCLRSubmitArguments %* > %LAUNCHER_OUTPUT%

81: "%JAVA_HOME%\bin\java" -Xmx512m -XX:MaxPermSize=300m -cp
"%HADOOP_CONF_DIR%;%LAUNCH_CLASSPATH%;%SPARK_HOME%\conf"
org.apache.spark.deploy.csharp.CSharpRunner debug
```

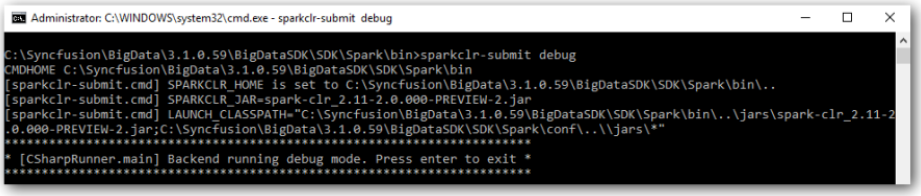
Note: *The line number that have been mentioned are listed based on the original batch file. (i.e.) sparkclr-submit.cmd file without any changes.*

Step 4a: Running Mobius C# Application

Starting Mobius C# Runner (Debug)

To execute the Mobius C# application with debugger, we should start the "CSharpRunner" in debug mode. Debug mode is used to step through the C# code in Visual Studio during a debugging session. The command to start the CSharpRunner in debug mode is as follows:

> *sparkclr-submit debug*



There are also other options available to run the mobius application, please refer to the following

link:
https://github.com/Microsoft/Mobius/blob/master/notes/running-mobius-app.md

Prerequisites

- Ensure whether the Mobius CSharpRunner is started (in debug mode) before the execution of the application.
- Ensure the files listed in the sample is available in the location mentioned / you can change it.
- For YARN Cluster – Client mode, make sure to set the Hadoop configuration environment variable "**HADOOP_CONF_DIR**" before starting the CSharpRunner.

YARN – Local mode

Set the master for the application as YARN and in default the job gets submitted in the local machine.

```
sparkConf.SetMaster("yarn");
```

YARN Cluster – Client mode

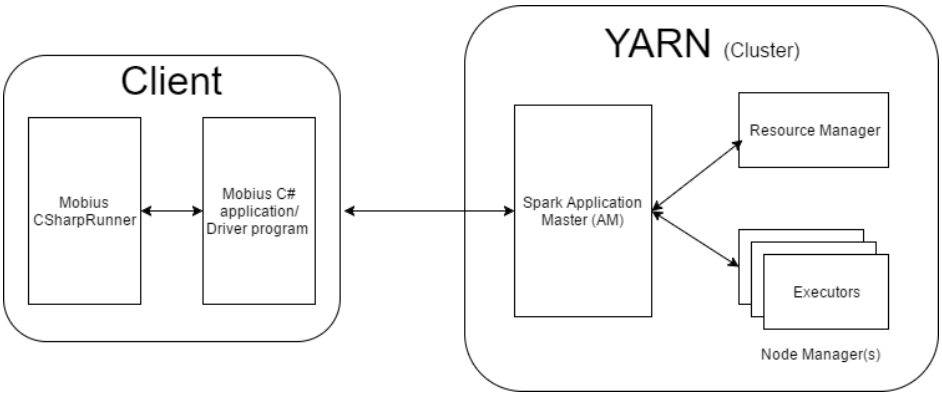


Figure 2: Mobius YARN Cluster – Client mode communication

For submitting job to the cluster, set the deploy mode property to the spark configuration and that will do the action. In Yarn – Client mode, the Mobius CSharpRunner, driver program and Spark driver will be running the client machine and the executors will be running on the cluster nodes.

```
sparkConf.Set("spark.submit.deployMode", "client");
```

Note: *If you've missed setting environment variable for Hadoop configuration*

(HADOOP_CONF_DIR) before starting the CSharpRunner, then the job will be submitted in the local machine.

C# sample application

Description

The attached (Mobius.zip - *Mobius/Mobius C# Application*) sample application helps you run/debug the Mobius APIs in C# and contains the code snippets of few Spark SQL – DataFrame APIs in Mobius C#. The listed APIs are as follows:

- Read dataframe
- Get columns in the dataframe
- Describe dataframe – Mean, Min, Max etc.
- Register as temporary table
- Read table using SQL query.

Similarly, we can use the other APIs too. For more Mobius C# APIs, please refer to the following link:
https://github.com/Microsoft/Mobius/blob/master/csharp/Adapter/documentation/Mobius_API_Documentation.md

Step 4b: Running Mobius Shell (REPL)

Mobius Shell supports the interactive mode that enables the analytics to express computation/analysis as a set of C# code statements that gets executed in the REPL (Read, Evaluate, Print, Loop) way.

Prerequisites

- Ensure the files listed in the sample is available in the location mentioned / you can change it.
- Ensure the folder "repl" is available in the location mentioned above.
- For YARN Cluster – Client mode, make sure to set the Hadoop configuration environment variable "**HADOOP_CONF_DIR**" before starting the mobius shell.

YARN – Local mode

Execute the following command to run the mobius shell in YARN - Local mode.

```
...SDK\Spark\bin> sparkclr-shell.cmd --master yarn
```

YARN Cluster – Client mode

Execute the following command to run the mobius shell in YARN Cluster – Client mode.

```
...SDK\Spark\bin> sparkclr-shell.cmd --master yarn-client
```

Note: *If you've missed setting environment variable for Hadoop configuration (HADOOP_CONF_DIR) before running the mobius shell, then the job will be submitted in the local machine.*

Built-in directives in shell

Command	Syntax/Directive	Description
Help	**:help**	Display help on available commands.
Load	**:load**	Load extra library to current execution context, e.g. :load "myLib.dll".
Quit	**:quit**	Exit REPL.

Sample script

```
//Load required namespace
> using System.Linq;

//Read dataframe from CSV file
> var dataFrame = sqlContext.Read().Format("com.databricks.spark.csv")
.Option("header", "true")
.Option("inferSchema", "true")
.Load("/Data/Customers/Customers.csv");

//Display the top 20 rows in the dataframe
> dataFrame.Show();

//Dataframe columns
> var columns = dataFrame.Columns().ToList();
```

Reference links

Mobius: https://github.com/Microsoft/Mobius
Mobius build: https://github.com/Microsoft/Mobius/releases
Mobius Architecture: https://databricks.com/blog/2016/08/03/developing-apache-spark-applications-in-net-using-mobius.html
Mobius examples: https://github.com/Microsoft/Mobius/tree/master/examples
Running Mobius: https://github.com/Microsoft/Mobius/blob/master/notes/running-mobius-app.md
Mobius C# API documentation: https://github.com/Microsoft/Mobius/blob/master/csharp/Adapter/documentation/Mobius_API_Documentation.md
Running Mobius in YARN cluster: https://github.com/Microsoft/Mobius/blob/master/notes/running-mobius-app.md#yarn-cluster
Mobius Shell: https://github.com/Microsoft/Mobius/blob/master/notes/mobius-shell.md#run-shell