

Recap

Friday, August 12, 2022

10:23 PM

1. Case Class
2. Pattern Matching
3. Inheritance
4. Traits
5. Layered Traits
6. Higher Order Functions
7. Spark Components

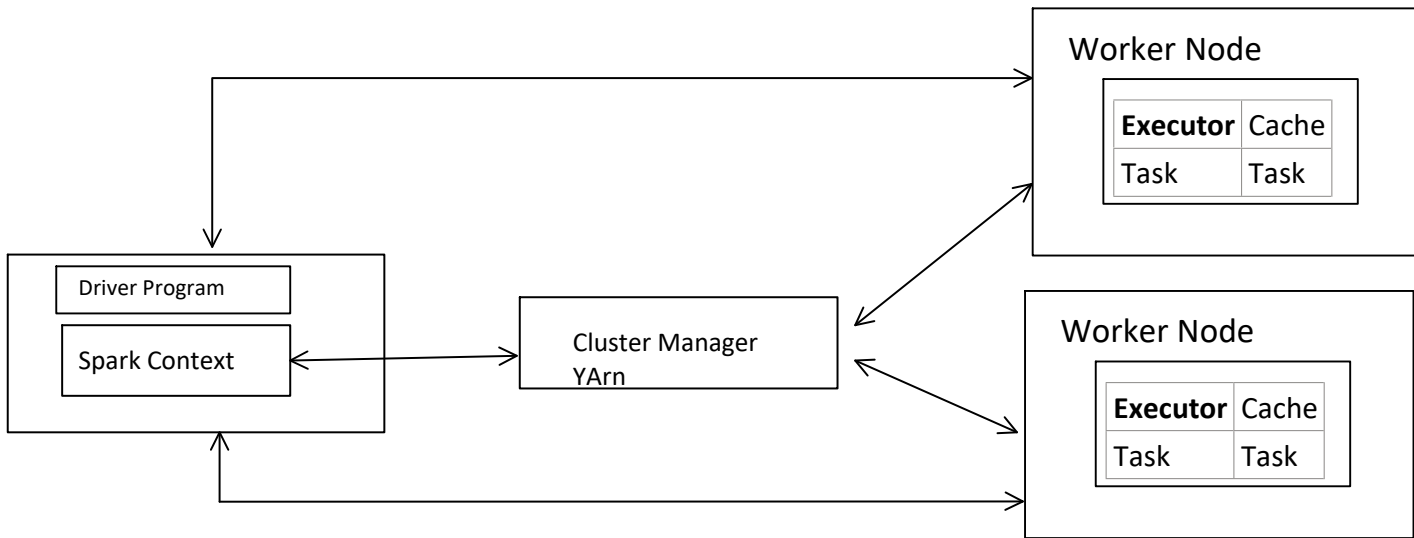
Topics for Today

Friday, August 12, 2022 10:25 PM

1. Spark Architecture
2. Deployment Mode
3. Building a First Spark Application
4. Sqoop
5. Sqoop Commands
6. Sqoop Incremental

Spark Architecture

09 July 2020 14:19



I need to open a locked door

Key

I don't have a key

Key Maker

I create a key

Spark is Door

To Access Spark - Spark Context

Driver Program Creates the Spark Context

Deployment Mode

09 July 2020 23:14

Client Mode - Spark-shell
Cluster Mode -

Go To slide 33

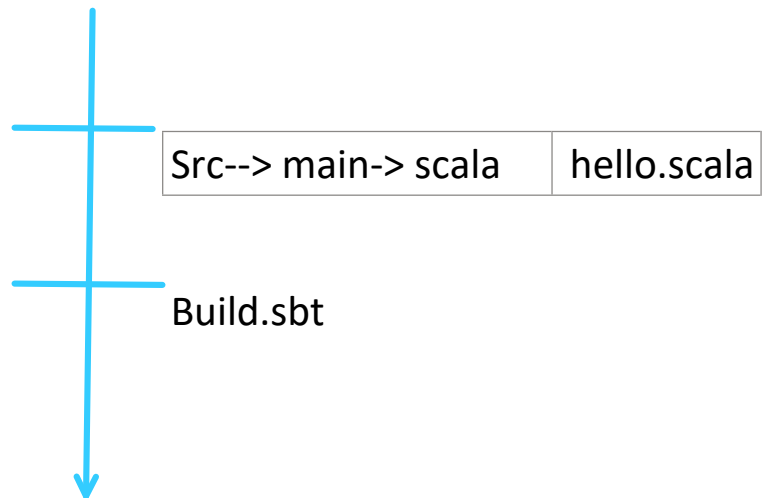
Building a First Spark Application

09 July 2020 23:16

Build- Convert scala code into Byte Code class files

Combine all class files to create one single ,which is called a jar file

SBT - Simple Build Tool



hello.scala

```
object hello {  
    def main(args:Array[String])  
        = {println("Hello World")}  
}
```

```
name := "Hello World"
```

```
version := "1.0"
```

```
scalaVersion := "2.11.8"
```

```
sbt package
```

```
spark2-submit \  
--class "hello" \  
--master yarn \  
/mnt/home/edureka_967855/fordownload/sbtproject/target/scala-2.11/hello-world_2.11-1.0.jar
```

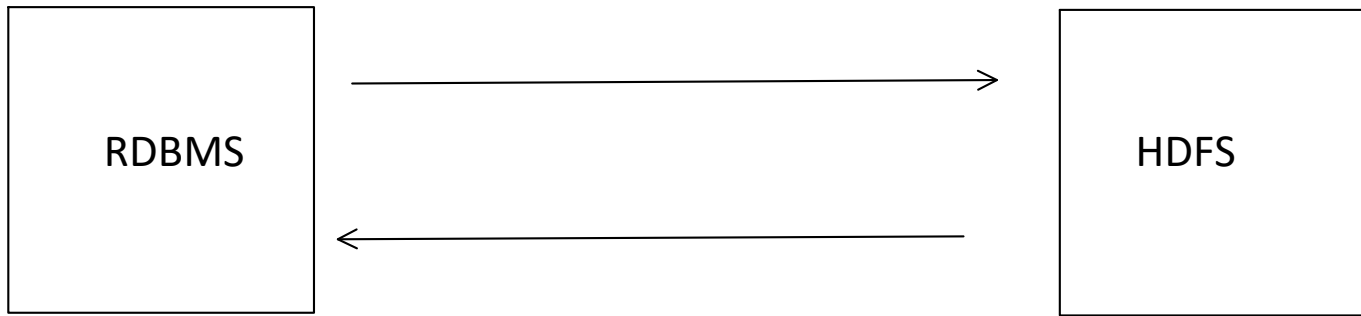
```
spark-submit --class "hello" --master yarn
```

```
/home/deepakpec296edu/FirstSparkApplication/target/scala-2.11/hello-world_  
2.11-1.0.jar
```

To check the list of classes in Jar Files
`jar -tvf hello-world_2.11-1.0.jar`

Sqoop

10 July 2020 23:08



Sqoop Commands

10 July 2020 23:37

Connect to MySQL Database

```
mysql -u deepakpec296edu-p
```

BlueMongoose84\$

```
mysql -h sqoopdb.edu.cloudlab.com -u labuser -p
```

```
CREATE TABLE CUSTOMERS (  
ID      INT          NOT NULL,  
NAME  VARCHAR(30) NOT NULL,  
AGE    INT          NOT NULL,  
ADDRESS CHAR(25) ,  
SALARY DECIMAL(18,2),  
PRIMARY KEY (ID)  
);  
SHOW TABLES;
```

```
Sqoop import \  
--connect jdbc:mysql://<ip address>/<database name>  
--table <mysql_table name>  
--username <username_for_mysql_user>  
--password <password>  
-m <number of mappers to run>  
--target-dir <target directory where data needs to be imported>  
--driver < Manually specify JDBC driver class to use>  
sqoop import --connect jdbc:mysql://ip-10-1-1-204.ap-  
south-1.compute.internal/deepakpec253edu --username
```

```
deepakpec253edu --password SilverKangaroo82$ --table CUSTOMERS -m
3 --target-dir '/user/deepakpec253edu/sqoop_sql/' --driver
'com.mysql.jdbc.Driver' --delete-target-dir
```

```
USE deepakpec253edu;
```

```
SHOW TABLES;
```

```
INSERT INTO CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)
VALUES(1,'Sehwag',23,'Delhi', 1000.00);
INSERT INTO CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)
VALUES(2,'Sachin',25,'Mumbai', 1200.00);
INSERT INTO CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)
VALUES(3,'Saurav',24,'Kolkatta', 1100.00);
```

```
SELECT * FROM CUSTOMERS;
```

```
sqoop import --connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/deepakpec296edu --driver com.mysql.jdbc.Driver --username
deepakpec296edu --password BlueMongoose84$ --table CUSTOMERS -m 3 --target-dir '/user/deepakpec296edu/sqoop_sql/' --delete-target-dir
```

```
sqoop import \
--connect jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_967855 \
--username labuser \
--password edureka \
--table CUSTOMERS \
-m 3 \
--target-dir '/user/edureka_967855/sqoop_sql/' \
--delete-target-dir
```

```
delete from CUSTOMERS;  
select * from CUSTOMERS;
```

Sqoop Exports;

```
sqoop export \  
--connect jdbc:mysql://<ip address>/<database name>  
--table <mysql_table name>  
--username <username_for_mysql_user>  
--password <password>  
-m <number of mappers to run>  
--export-dir <directory from where data needs to be exported>
```

```
sqoop export --connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/deepakpec296edu --table CUSTOMERS --username  
deepakpec296edu --password BlueMongoose84$ -m 3 --export-dir '/user/deepakpec296edu/sqoop_sql/' --driver 'com.mysql.jdbc.Driver'
```

```
select * from customers;
```

List databases

```
sqoop list-databases \  
--connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/ \  
--username deepakpec253edu \  
--password SilverKangaroo82$
```

```
sqoop list-databases --connect jdbc:mysql://ip-10-1-1-204.ap-  
south-1.compute.internal/ --username deepakpec253edu --password  
SilverKangaroo82$
```

List Tables

```
sqoop list-tables \
```

```
--connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/deepakpec253edu \
```

```
--username deepakpec253edu \
```

```
--password SilverKangaroo82$
```

```
sqoop list-tables --connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/deepakpec253edu --username deepakpec253edu --password SilverKangaroo82$
```

```
hdfs dfs -rm -r -f -skipTrash /user/edureka_967855/sqoop_sql/
```

```
sqoop import \
```

```
--connect jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_967855 \
```

```
--table CUSTOMERS \
```

```
--username labuser \
```

```
--password edureka \
```

```
-m 1 \
```

```
--where "Age > 24" \
```

```
--target-dir '/user/edureka_967855/sqoop_sql/' \
```

```
--delete-target-dir
```

```
INSERT INTO CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)  
VALUES(4,'Rahul',23,'Delhi', 1000.00);
```

```
sqoop eval \
```

```
--connect jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_967855 \  
--username labuser \  
--password edureka \  
--query "SELECT * FROM CUSTOMERS LIMIT 10"
```

```
delete from CUSTOMERS;  
DROP TABLE CUSTOMERS;  
DROP DATABASE edureka_967855;
```

What is Incremental Data?

Append Mode - Only New Records

Mode

Column Name ID

Last Value that has been imported - 3

Last Modified Mode - Allow New as well as Updated Records

Source table in Database should have a timestamp column

Mode

Column Name

Last Value of Timestamp

Retrieve newer rows than previously imported set

of rows

Append - New Rows without updates to existing Rows

There is some id

LastModified - New Rows + Updates to existing rows

Column - Which Timestamp Value

```
mysql -h sqoopdb.edu.cloudlab.com -u labuser -p
```

Append Example

```
INSERT INTO
```

```
CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)
```

```
VALUES(4,'Vishnu',23,'Delhi', 1000.00);
```

```
INSERT INTO
```

```
CUSTOMERS(ID,NAME,AGE,ADDRESS,SALARY)
```

```
VALUES(5,'Rajat',23,'Delhi', 1000.00);
```

```
sqoop import \
```

```
--connect
```

```
jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_  
967855 \
```

```
--table CUSTOMERS \  
--username labuser \  
--password edureka \  
-m 1 \  
--target-dir '/user/edureka_967855/sqoop_sql/' \  
--incremental append \  
--check-column ID \  
--last-value 3
```

```
sqoop import --connect jdbc:mysql://ip-10-1-1-204.ap-south-1.compute.internal/deepakpec296edu --  
table CUSTOMERS --username deepakpec296edu --password BlueMongoose84$ -m 1 --target-dir  
'/user/deepakpec296edu/sqoop_sql/' --driver 'com.mysql.jdbc.Driver' --incremental append --check-  
column ID --last-value 3
```

```
sqoop eval \  
--connect  
jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_  
967855 \  
--username labuser \  
--password edureka \  
--query "SELECT * FROM CUSTOMERS LIMIT 10"
```

Job

```
sqoop job --create empupdatejob -- import --connect jdbc:mysql://ip-10-1-1-204.ap-  
south-1.compute.internal/deepakpec296edu --username deepakpec296edu --password  
BlueMongoose84$ --table CUSTOMERS --m 1 --incremental append --check-column ID --last-value 3 --
```



```
driver 'com.mysql.jdbc.Driver' --target-dir '/user/deepakpec296edu/sqoop_sql/'  
sqoop job --list
```

```
sqoop job --exec empupdatejob  
sqoop job --delete empupdatejob
```

```
sqoop import --connect jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_967855 --username labuser --  
password-file '/user/edureka_967855/sqoop/sqoop.password' --table CUSTOMERS --columns  
'NAME,ID' -m 1 --target-dir '/user/edureka_967855/sqoop_sql/' --delete-target-dir
```

```
sqoop import --connect jdbc:mysql://sqoopdb.edu.cloudlab.com/edureka_967855 --username labuser --  
password-file '/user/edureka_967855/sqoop/sqoop.password' --query 'SELECT NAME,ID FROM  
CUSTOMERS WHERE $CONDITIONS AND ID > 3' -m 1 --target-dir '/user/edureka_967855/sqoop_sql/' --  
delete-target-dir
```

Challenges in Existing Computing Methods

12 July 2020 14:23

Slide 7 - 10

Iterative Distributed Computing



Problem - MapReduce - Lots of i/o operations

Rdd store the Lineage

IntelliJSet up

Friday, March 3, 2023

10:38 PM

[How to setup Scala on IntelliJ IDEA IDE + First SBT/Scala project](#)

