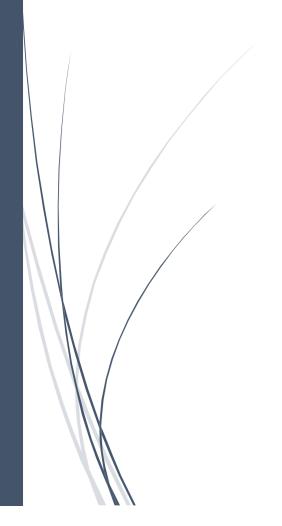Big Data Lab

PIG Installation

1. Pig Installation

1.1 From Source :-

- Open terminal and cd into Downloads folder.

- Download Pig source code tarball using the following command in the terminal.

  ```
  wget https://downloads.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz
  ```

  Or download the source code tarball directly from the website.

- Unzip the file

  ```
  tar -xvf pig-0.17.0.tar.gz
  ```

  Note: pig_path is the path of where the extracted PIG folder is present.

  It can be found by using the following commands.

  ```
  cd pig-0.17.0
  ```

  ```
  pwd
  ```

- Set PIG globally by updating the system bash file.

  ```
  gedit ~/.bashrc
  ```

- Paste the following code at the end of the file

  ```
  export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
  ```

  ```
  export PIG_HOME=pig_path

  export PATH=$PATH:$PIG_PATH/bin

  export PIG_CLASSPATH=$HADOOP_HOME/conf
  ```

- Save and close the bashrc file and run the source command to load and save the new variables globally.

```
source ~/.bashrc
```

- To verify installation run the following command, it will return pig version number.

```
pig -version
```

1.2. Installation using a shell script (easier method) :-

Running this shell script automates the complete installation process.

- Open the terminal and create a new file named install-pig.sh

- Copy paste the below code into the new file and save it.

```sh
#!/bin/sh

cd ~/Downloads
echo ""
wget "https://downloads.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz"
echo Downloaded pig-0.17.0 successfully
echo ""
tar -xvf pig-0.17.0.tar.gz
echo Unzipped pig-0.17.0 successfully
echo ""
rm pig-0.17.0.tar.gz
cd pig-0.17.0
PIG_PATH=`pwd`
cd ..
echo export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64 >> ~/.bashrc
source ~/.bashrc
echo ""
echo JAVA_HOME set to /usr/lib/jvm/java-8-openjdk-amd64
echo ""
echo export PIG_HOME=$PIG_PATH >> ~/.bashrc
source ~/.bashrc
echo ""
echo PIG_HOME set to $PIG_PATH
echo ""
echo export PATH=$PATH:$PIG_PATH/bin >> ~/.bashrc
source ~/.bashrc
echo ""
echo PATH set to $PATH:$PIG_PATH/bin
echo ""
echo export PIG_CLASSPATH=$HADOOP_HOME/conf >> ~/.bashrc
source ~/.bashrc
echo ""
echo PIG_CLASSPATH set to $HADOOP_HOME/conf
echo ""
echo "Installation completed successfully"
```

After saving the file run the following command.

```
bash install-pig.sh
```

2. Steps to run Pig programs

- Create a pig script and write the business logic in this file

```
gedit file_name.pig
```

- To run the code

```
pig -x local file_name.pig
```