

Hadoop Streaming

Where do we use Hadoop Streaming?

While running Hadoop Map Reduce Jobs, we come across Hadoop Streaming. This is implemented using a Jar file whose path we need to initialize in Hadoop versions below 2.6 and is initialized automatically in versions higher.

What is Hadoop Streaming?

Most of us know the Hadoop's primary language is Java since Hadoop has been programmed in Java. But developers who do not know Java can use Hadoop as well. This extension of functionality is because of Hadoop Streaming Jar files. Loading these Jar files basically enables Java based Hadoop to read input, process mappers and reducers and display output on standard IO devices as defined in languages other than Java. Hadoop streaming supports Python, PHP, Ruby, Perl, bash etc. The Input and output are written into Java Objects and the mappers/reducers have limited functionality or options to customize with different languages other than Java. However, for beginners and simple tasks, Hadoop streaming is a better option than Java based definitions since the functionality provided by Hadoop streaming covers basic necessities sufficiently, especially if the user is more comfortable with the other languages.

How to use Hadoop streaming?

Loading Hadoop Streaming Using Python(with MR job)

```
python PyFileContainingMapperAndReducerDefinitions.py -r Hadoop -hadoop-streaming-jar  
/locationOfHadoopStreamingJar/hadoop-streaming.jar MapperReducerData.data
```

Loading Python Using Hadoop Streaming(without MR job)

```
./locationOfHadoopStreamingJar/hadoop-streaming.jar -input myinputfile -output myoutputfile -  
mapper /locationOfMapperInPy/mapper.py -reducer /locationOfReducerInPy/reducer.py
```

https://www.tutorialspoint.com/hadoop/hadoop_streaming.htm

Other interesting articles-

<https://hadoop.apache.org/docs/r1.2.1/streaming.html>

<https://www.kdnuggets.com/2018/12/different-conventional-programming-machine-learning.html>

<http://hadooptutorial.info/introduction-to-hadoop-streaming/>

https://www.youtube.com/watch?v=H_8CMENqEZ4

http://mlwhiz.com/blog/2015/05/09/Hadoop_Mapreduce_Streaming_Tricks_and_Techniques/