1- HDFS Read

2- Map reduce analysis

3- HDFS write

donwload txt from gutenberg

create a folder in HDFS root- gutenberg

move file to this folder

 hadoop fs -put /home/ankit/Downloads/petrified.txt /gutenberg

Create maven java project-- hadoopmr4

How to do MapReduce on hadoop

      0- look at input file to select appropriate input format class-- inputformat class

      1- create a mapper class

      2- create a reducer class

      3- create a driver class with a main method to create a jar

      a- create a driver method with a man method to create a job
      b- decide what to write to hdfs using and output format class

      4- build this project to create a jar file
      5- run the jar file as follows

from hadoop.org-commands-jar

hadoop jar <jar> [mainClass] inputPathOnHDFS outputPathonHDFS

import hadoop core dependency

extends mapper class for Mapper

use most appropriate class and compact datatype

last 2 of mapper should match 1st 2 of reducer- output of mapper- input of reducer

just like in mongoDB values is iterable

hadoop jar <jar> [mainClass] inputPathOnHDFS outputPathonHDFS

these args will go to set input and output path

java 9 docs api

build jar file

com.ankit.mr.hadooplab4.WordCountMR

<mark>Note Very Imp--</mark>

There are following 2 ways to run MapR on hadoop

1- hadoop jar <jar> [mainClass] inputPathOnHDFS outputPathonHDFS

use this command is the main class is not declared in the pom.xml file

hadoop jar /home/ankit/Downloads/mrwordcount-2.jar com.ankit.mr.hadooplab4.WordCountMR /gutenberg/petrified.txt /gutenbergOutput2

2- hadoop jar <jar> inputPathOnHDFS outputPathonHDFS

use this one if main class is declared in pom.xml

hadoop jar /home/ankit/Downloads/mrwordcount.jar /gutenberg/petrified.txt /gutenbergOutput2

it can be done as follows:

```
    <build>
        <plugins>
            <plugin>
                <groupId>org.apache.maven.plugins</groupId>
                <artifactId>maven-jar-plugin</artifactId>
                <version>2.4</version>
                <configuration>
                    <archive>
                        <manifest>
                            <addClasspath>true</addClasspath>
<mainClass>com.ankit.mr.hadooplab4.WordCountMR</mainClass>
                        </manifest>
                    </archive>
                </configuration>
            </plugin>
        </plugins>
    </build>
```