SWE 307 BIG DATA PROJECT - 2 DESCRIPTION

HDFS as Data Storage

Due date: 8.11.2024 Thursday, in class.

In this project study, you are asked to create the big data-based information system shown in Figure 1. In the project work, the employee information of a company is stored in the "emp.csv" file. The company's department information is stored in the "dept.csv" file.

What is required from you is as follows:

1) Hadoop-HDFS cluster must be installed as single node in your computer.

2) A simple Java Spring-Boot application will be developed to perform the following tasks:

    a) Personnel and department data will be read from the database (can be MySQL or PostgreSQL).

    b) Personnel images will be stored to/read from HDFS.

    c) There will be a single web page, on this page the information will be displayed in a table using the JOIN operation on database. Information to display: employee name, manager name, salary, commission, department.
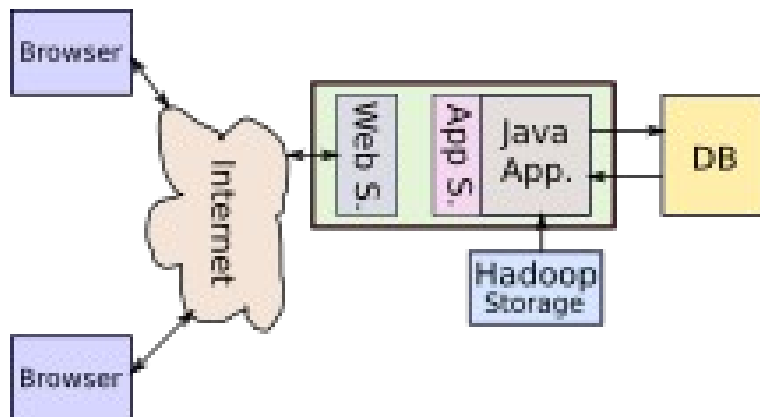

Figure 1. Big data based web information system.

PS: 1) Some groups may want to implement the project on G-Drive or AWS-S3 as file storage, it is also accepted. This will be discussed in the class.
2) Example image files and csv files will be provided on Github repository, you can clone/download everything provided.
Link: https://github.com/ozmen54/SWE307-2024.git

Here you are text data as well:
emp.csv
empno,ename,job,mgr,hiredate,sal,comm,deptno,img
7369,SMITH,CLERK,7902,17-DEC-1980,800,,20,smith.jpg
7499,ALLEN,SALESMAN,7698,20-FEB-1981,1600,300,30,allen.jpg
7521,WARD,SALESMAN,7698,22-FEB-1981,1250,500,30,ward.jpg
7566,JONES,MANAGER,7839,2-APR-1981,2975,,20,jones.jpg
7654,MARTIN,SALESMAN,7698,28-SEP-1981,1250,1400,30,martin.jpg

7698,BLAKE,MANAGER,7839,1-MAY-1981,2850,,30,blake.jpg
7782,CLARK,MANAGER,7839,9-JUN-1981,2450,,10,clark.jpg
7788,SCOTT,ANALYST,7566,09-DEC-1982,3000,,20,scott.jpg
7839,KING,PRESIDENT,\000,17-NOV-1981,5000,,10,king.jpg
7844,TURNER,SALESMAN,7698,8-SEP-1981,1500,0,30,turner.jpg
7876,ADAMS,CLERK,7788,12-JAN-1983,1100,,20,adams.jpg
7900,JAMES,CLERK,7698,3-DEC-1981,950,,30,james.jpg
7902,FORD,ANALYST,7566,3-DEC-1981,3000,,20,ford.jpg
7934,MILLER,CLERK,7782,23-JAN-1982,1300,,10,miller.jpg

dept.csv
deptno,dname,loc
10,ACCOUNTING,NEW YORK
20,RESEARCH,DALLAS
30,SALES,CHICAGO
40,OPERATIONS,BOSTON