SWE 307 BIG DATA PROJECT - 2

Redis for RDB caching and HDFS as Data Storage

Due date: 6.11.2025 Thursday, in class.

This project aims to build a real-time information system that combines the speed of in-memory data processing with the reliability and scalability of distributed storage (See Figure 1.). The system uses **Redis** to handle fast, real-time ingestion and temporary caching of incoming data, while **Hadoop HDFS** provides a distributed and fault-tolerant file system to store documents that are uploaded or downloaded. The Hadoop HDFS is expected to improve response time as well.

Key Objectives:

- Use scott database (https://github.com/rsp/pg-scott) for RDBMS(MySQL). At the beginning, import data from given site to your database.
- Use **Redis** to temporarily store and process incoming data for fast access.Periodically flush the data from Redis to **RDBMS** for durable storage and future processing. Show that the caching mechanism works as expected. (See example: https://www.youtube.com/watch?v=0a-RlJx09rg)
- Use **Hadoop-HDFS** for document storage.

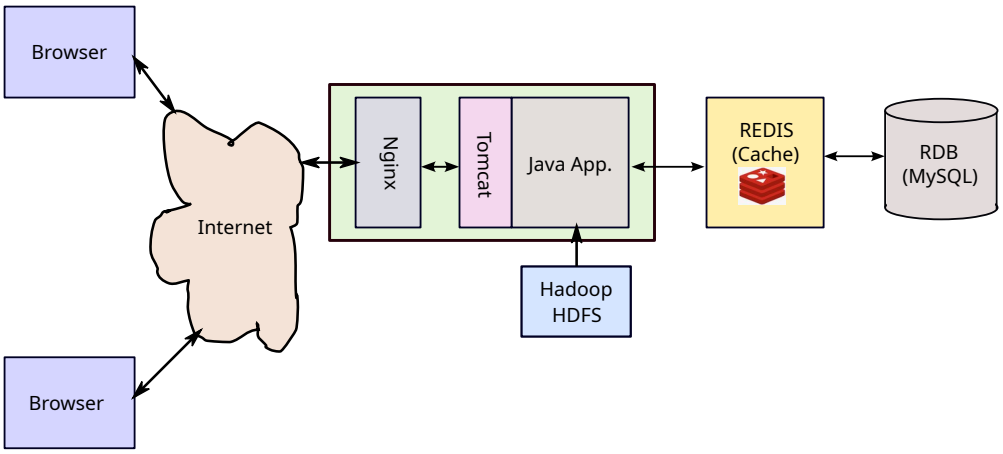

Figure 1. Architectural block diagram of project 2.

What is required from you is as follows:

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

2) Redis and MySQL must be installed on 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/updated from the database via Redis.
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.

Notes:

1) Some groups may want to implement the project using G-Drive or AWS-S3 Object storage, these implementations will also be accepted. This will be discussed in the class.
2) Example image files and data files will be provided on Github repository, you can clone/download everything provided.
Link: https://github.com/ozmen54/SWE307-2025.git