Big Data - Project 1: Best Books API

Project Description:

Project 1 is a Scala, Spark Hive command line application that connects to a CSV file that is from amazon book reviews. Through qruns queries showing the stats of books using user ratings, author, title, price, year released, and genre.

Technologies Used

\* Scala - 2.12.10

\* Spark - version 3.1.2

\* MySQL Workbench - version 8.0.28

\* IntelliJ IDEA 2021.3.2 (Community Edition)

\* OpenJDK 8

\* Hadoop - 3.3.2

\* Hive - 3.1.2

\* ubuntu - 20.04.4

Features

\* User Account Management

\* Administrative user controls

\* Different Menus for users and Admins

To-do list:

\* Add password encryption feature

\* add a Master user method or class that supersedes all users, even admins

Getting Started

(include git clone command) (include all environment setup steps)

First setup steps:

Step 1: clone git hub repository using this link: https://github.com/AarSchoSkIg/Project1-AarSchoSkIg.git

Step 2: install all the above technologies on to your machine

Step 3: Go to Linux cd to Hadoop 3.3.2. Then start data nodes and yarn

Step 4: open project in Intellij

Usage

* Start MySQL Workbench and run the script called: bestbooksapiproject1bdscript
* Run code in IntelliJ
* Select option 2 to Sign In
* Enter username: Codingboss00 ; password: Zelda
* Select option 3 for books menu
* Select which query to run by entering in corresponding number
* See results of the ran queries

Text

Description automatically generated

A computer screen capture

Description automatically generated with medium confidence

Text

Description automatically generated

Text

Description automatically generated

License

This project uses the following license: no license.