**基于Hadoop可视化大数据分析显示**

|  |  |
| --- | --- |
| 班级 | 软件2202 |
| 姓名 | 刘扬 |
| 学号 | 2230200063 |
| 教师 |  |
| 日期 | 20241208 |

软件与通信学院编制

2024年2月

|  |  |
| --- | --- |
| 任务1 | 建立数据库 |
| 代码截图 | Create database tjnuworkmovie; |
| 运行结果截图 | 01 |
| 任务2 | Hive建表(以genome\_tags为例) |
| 代码截图 | create external table genome\_tags(  tagId string,  tag srting  )  row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'  with serdeproperties  ('separatorChar' = ',') |
| 运行结果截图 | 0201 |
| 任务3 | 数据导入hdfs（以movies为例） |
| 代码截图 | hadoop fs -put /export/data/tjnu/movies.csv /tjnu/demo01/tjnu/ |
| 运行结果截图 | 03 |
| 任务4 | 数据导入hive表 |
| 代码截图 | load data inpath ‘/tjnu/demo01/tjnu/movies.csv’ into table movies |
| 运行结果截图 | 0301030203030304 |
| 任务5 | 数据处理：统计出前10中标签 |
| 代码截图 | CREATE TABLE IF NOT EXISTS tag\_movie\_count AS  SELECT tag, COUNT(DISTINCT movieId) AS movie\_count  FROM tags  GROUP BY tag;  CREATE TABLE IF NOT EXISTS top\_10\_tags AS  SELECT tag, movie\_count  FROM tag\_movie\_count  ORDER BY movie\_count DESC  LIMIT 10 |
| 运行结果截图 | top10tags |
| 任务6 | 数据处理：评级最高前10个电影的名字 |
| 代码截图 | CREATE TABLE IF NOT EXISTS top10\_movies AS   SELECT m.title, AVG(r.rating) AS avg\_rating  FROM movies m JOIN ratings r ON m.movieId = r.movieId  GROUP BY m.title  ORDER BY avg\_rating DESC  LIMIT 10 |
| 运行结果截图 | top10movies |
| 任务7 | 数据处理：前10个用户 |
| 代码截图 | CREATE TABLE IF NOT EXISTS top10\_users AS  SELECT userId, COUNT(movieId) AS total\_movies  FROM ratings  GROUP BY userId  ORDER BY total\_movies DESC  LIMIT 10 |
| 运行结果截图 | top10_users |
| 任务8 | 数据处理：评级最高前10条的电影信息 |
| 代码截图 | CREATE TABLE IF NOT EXISTS top10\_rated\_movies AS  SELECT m.movieId, m.title, m.genres, MAX(r.rating) AS max\_rating  FROM movies m JOIN ratings r ON m.movieId = r.movieId  GROUP BY m.movieId, m.title, m.genres  ORDER BY max\_rating DESC LIMIT 10 |
| 运行结果截图 | top10_rated_movies |
| 任务9 | 数据处理：每个标签下的电影总数量 |
| 代码截图 | CREATE TABLE IF NOT EXISTS tag\_movie\_count AS  SELECT tag, COUNT(DISTINCT movieId) AS movie\_count  FROM tags  GROUP BY tag |
| 运行结果截图 | tag_movie_count |
| 任务10 | 数据传输：从hive到mysql |
| 代码截图 | 12 |
| 运行结果截图 | 0101010201030104 |
| 任务11 |  |
| 代码截图 |  |
| 运行结果截图 |  |
| 任务12 |  |
| 代码截图 |  |
| 运行结果截图 |  |
| 任务13 |  |
| 代码截图 |  |
| 运行结果截图 |  |