|  |  |
| --- | --- |
| **Assignment Case** |  |
| COMP6579  Big Data Processing |
| **Computer Science** | **E203-COMP6579-WS02-01**  **E203-COMP6579-SL04-01** |
| ***Valid on*** *Even Semester Year 2019/2020* | **Revision 00** |

## Soal

*Case*

**GameZ Ztore**

**GameZ Ztore** is a new game store developed by **Software Laboratory**. As GameZ Ztore become more famous in the society, they wanted to do some analytics on their data.

From the sales business process, the data can be analyzed to gain sales insight. The data is stored in **MySQL** dump file and the data schema is drawn using **Entity Relationship Diagram (ERD)** below:

A screenshot of a cell phone

Description automatically generated

Figure 1. GameZ Ztore ERD

You were given the task to gain some insight from the **sales** data using **Hadoop** tools. Below is the task you must do:

# **Load data from MySQL to Hive**

Given the file “**GameZ.sql**” that consists of the data about **sales**, **product**, **developer**, **publisher**. You need to load the data to **MySQL** database, then **ingest** the data from **MySQL** database to **Hive** for data integration.

# **Query Analysis**

From the data in **Hive**, you need to gain some sales insight in **GameZ Ztore**, below are some statements you need to answer using **Hive** / **Impala** query:

* 1. Show **games** which **quantity** have been sold **greater than 25**
  2. Show **top 10** **games** which is developed by **Konami** and is **sold the most between 2017 and 2019**
  3. Show **top 5 publisher** that published **more than 100** games and the game price is **below 500000**
  4. Show **games** which **occurred in multiple transaction** and **sold greater than 10 times** where the **rating is 'T'**
  5. Show **games** and **location** where it is **mostly sold (either NA, EU, and JP)** for every game where the **rating is 'E10+'** and have been **sold greater than 6**. For example:

|  |  |
| --- | --- |
| NA Sales | 0.05 |
| EU Sales | 0.8 |
| JP Sales | 0.14 |
| Then the game is mostly sold in EU | |

**BlueBG Store**

**BlueBG Store** is a **boardgame** retail store under **PT**. **Software Laboratory**. As **BlueBG Store** has grown rapidly and planned to “go public” there is a need to improve the store sales more quickly. To do that they intended to do some **analysis** from **different** **kinds** of data they have.

From the sales business process, some data can be analyzed to gain sales insight in **BlueBG Store**. The data is stored in **MySQL** dump file and the data schema can be drawn using **Entity Relationship Diagram** (**ERD**) below:

A screenshot of a social media post

Description automatically generated

Figure 1. BlueBG Store ERD

You were given the task to gain some insight from the **sales** data using **Hadoop** tools. Below is the task you must do:

# **Load data from MySQL to Hive**

Given the file “**bluebgstore.sql**” that consists of the data about **sales**, **boardgame, staff, and customer**. You need to load the data to **MySQL** database, then **ingest** the data from **MySQL** database to **Hive** for data integration.

# **Query Analysis**

From the data in **Hive**, you need to gain some sales insight in **Bluejack Store**, below are some statements you need to answer using **Hive** / **Impala** query:

* 1. Show **top 10** total **boardgame** copies sold that is **released in 2017**
  2. Show the **average price per category** where the **total of boardgame** in the category is **more than 50**
  3. Show **5 most bought** category in **January**
  4. Show all staff who has **handled transaction** **more than the average**

|  |  |  |
| --- | --- | --- |
| **Maximum Player** | **Average Time** | **Recommendation Message** |
| >= 5 | <= 120 | Recommended for Beginners |
| >= 5 | > 120 | Recommended for Parties |
| < 5 | > 120 | Not Recommended for Beginners |

* 1. Show boardgames **recommendation message**. All boardgame with **minimum** player of 2 or more and **maximum** player of 4 or more will get a recommendation message with terms and condition as below:

**Files to be collected**:

[NIM].txt that consist of:

* Command to Load data from MySQL to Hive
* Hive query for analysis

**Note:**

1. Remember to clearly separate the answers since there are two different themes in this quiz.
2. For any question regarding this quiz, feel free to contact your assistant through the forum in Binusmaya or personal contact.

**Good Luck 😊**