**DATA AGGREGATIONS**

**Dataset**: Video Game Sales With Ratings

**Date**: July 31, 2022

**Tools Used:**

* Azure HDInsight Cluster
* Azure file storage
  + Resource group: SNeudorf
  + Location: Canada Central
  + Storage account: sneudorfhdistorage
* Apache Hive 2.0

**Data Aggregations Performed:**

Within Apache Hive 2.0:

1. Shows the average critic score for Sports games. Store this information in a column called “sports\_critic\_score”
2. Shows the average critic score for Shooter games. Store this information in a column called “shooter\_critic\_score”
3. Created three statistics tables (average, min, max) for the global\_sales for: all games, Sports games, Shooter games. For each table, label the columns as: “average\_global\_sales”, “min\_global\_sales”, and “max\_global\_sales”
4. Created two tables containing the global\_sales and the count of games with that global\_sales for: Sports games and Shooter games. For each table, label the columns as: “global\_sales” and “count”.

**Scripts Used:**

Below were the scripts used for the data aggregations listed in the ‘Data Aggregations Performed’ section above:

1.

*SELECT AVG(Critic\_Score) AS sports\_critic\_score*

*FROM sales\_genre*

*WHERE genre = "Sports";*

2.

*SELECT AVG(Critic\_Score) AS shooter\_critic\_score*

*FROM sales\_genre*

*WHERE genre = "Shooter";*

3. a) **All Games**

*SELECT AVG(global\_sales) AS average\_global\_sales,*

*MIN(global\_sales) AS min\_global\_sales,*

*MAX(global\_sales) AS max\_global\_sales*

*FROM sales\_genre;*

3. b) **All Sports Games**

*SELECT AVG(global\_sales) AS average\_global\_sales,*

*MIN(global\_sales) AS min\_global\_sales,*

*MAX(global\_sales) AS max\_global\_sales*

*FROM sales\_genre*

*WHERE genre = "Sports";*

3. c) **All Shooter Games**

*SELECT AVG(global\_sales) AS average\_global\_sales,*

*MIN(global\_sales) AS min\_global\_sales,*

*MAX(global\_sales) AS max\_global\_sales*

*FROM sales\_genre*

*WHERE genre = "Shooter";*

4. **TABLE 1: Sports Games – Global Sales**

*SELECT COUNT(global\_sales) AS count,*

*SUM(global\_sales) AS global\_sales*

*FROM sales\_genre*

*WHERE genre="Sports";*

**TABLE 2: Shooter Games – Global Sales**

*SELECT COUNT(global\_sales) AS count,*

*SUM(global\_sales) AS global\_sales*

*FROM sales\_genre*

*WHERE genre="Shooter";*