Hadoop Assignment 1

Tanzeel-ur Rehman

648882

<https://github.com/htanzeel01/Hadoop-/tree/main>

A screenshot of a computer

Description automatically generated with medium confidence

Pig is a high-level programming language useful for analysing large data sets. Pig Hadoop was developed by Yahoo! and is generally used with Hadoop to perform a lot of data administration operations.

For writing data analysis programs, Pig renders a high-level programming language called Pig Latin. Several operators are provided by Pig Latin using which personalized functions for writing, reading, and processing of data can be developed by programmers.

For analyzing data through pig, we need to write scripts using Pig Latin. Then, these scripts need to be transformed into MapReduce tasks. This is achieved with the help of Pig Engine.

MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster. MapReduce facilitates concurrent processing by splitting petabytes of data into smaller chunks and processing them in parallel on Hadoop commodity servers. In the end, it aggregates all the data from multiple servers to return a consolidated output back to the application.

Tez is an Apache Software Foundation project that provides a framework for building high-performance batch and interactive data processing applications on top of Hadoop. Tez is designed to improve the execution speed and efficiency of complex data processing tasks by enabling optimized data flow and task scheduling.

# Code

**Cleanup**

**A picture containing text, font, screenshot

Description automatically generated**

Reading the data and filtering out any missing/null values

**Count games according to release date**

**A screenshot of a computer code

Description automatically generated with low confidence**

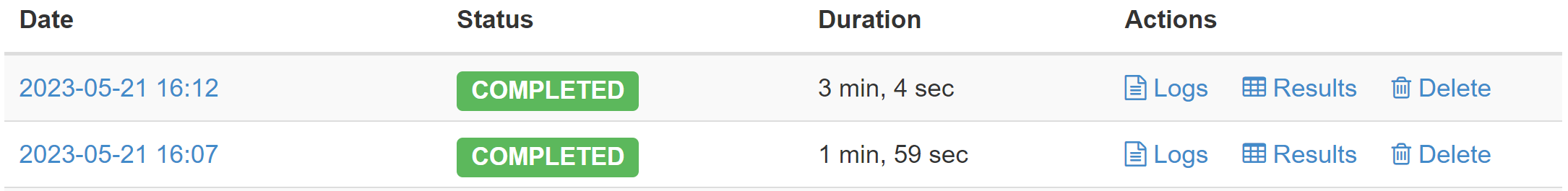
**Results**

**A picture containing text, font, screenshot, typography

Description automatically generated**

**Which System was faster?**

I tested it on both tez and mapreduce and noticed that the code executed fasted on tez.



The difference is almost more than a minute which shows that the tez executed code faster than mapreduce.