

DAILY ONLINE ACTIVITIES SUMMARY

| | | | |
|--|------------------------|-------------------|------------|
| Date: | 02-05-2020 | Name: | PRASANNA |
| Sem & Sec | 8 th ,B | USN: | 4AL16CS068 |
| Online Test Summary | | | |
| Subject | ----- | | |
| Max. Marks | ----- | Score | ----- |
| Certification Course Summary | | | |
| Course | Introduction to Hadoop | | |
| Certificate Provider | Great learner academy | Duration | 6 Hrs |
| Coding Challenges | | | |
| Problem Statement: prob1- <i>To print number is even or odd</i> | | | |
| Status: Solved | | | |
| Uploaded the report in Github | | Yes | |
| If yes Repository name | | prasanna_p | |
| Uploaded the report in slack | | Yes | |

Online Test Details: (Attach the snapshot and briefly write the report for the same)

Certification Course Details: (Attach the snapshot and briefly write the report for the same)

Coding Challenges Details: (Attach the snapshot and briefly write the report for the same)

1. Certification Course Details:



The screenshot shows the Great Learning Learning Management System (LMS) interface. At the top, there's a navigation bar with the Great Learning logo, 'Home', 'Live Sessions', and 'Certificates' links. On the right, there's a 'My Courses' button and a user profile icon. Below the navigation bar, a breadcrumb trail reads 'Courses / Introduction to Hadoop / What is ETL'. A left-hand sidebar titled 'Content' lists various learning videos. The video 'What is ETL' is currently selected and highlighted. The main area of the page displays a video player for the selected video. The video shows a man standing in front of a whiteboard with the text 'Polyglot Persistence' written on it. The video player includes standard controls like play/pause, volume, and a progress bar. Below the video player, there's a rating section asking 'How would you rate this video' with five stars. At the bottom right, there's a 'Activate Windows' watermark.

Introduction to Hadoop :

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed *storage* and *computation* across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

MapReduce

MapReduce is a parallel programming model for writing distributed applications devised at Google for efficient processing of large amounts of data (multi-terabyte data-sets), on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. The MapReduce program runs on Hadoop which is an Apache open-source framework.

Hadoop Distributed File System

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. It is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides high throughput access to application data and is suitable for applications having large datasets.

How Does Hadoop Work?

It is quite expensive to build bigger servers with heavy configurations that handle large scale processing, but as an alternative, you can tie together many commodity computers with single-CPU, as a single functional distributed system and practically, the clustered machines can read the dataset in parallel and provide a much higher throughput. Moreover, it is cheaper than one high-end server. So this is the first motivational factor behind using Hadoop that it runs across clustered and low-cost machines.

Hadoop runs code across a cluster of computers. This process includes the following core tasks that Hadoop performs –

- Data is initially divided into directories and files. Files are divided into uniform sized blocks of 128M and 64M (preferably 128M).
- These files are then distributed across various cluster nodes for further processing.
- HDFS, being on top of the local file system, supervises the processing.
- Blocks are replicated for handling hardware failure.
- Checking that the code was executed successfully.
- Performing the sort that takes place between the map and reduce stages.
- Sending the sorted data to a certain computer.
- Writing the debugging logs for each job.

Advantages of Hadoop

Advantages

- Hadoop framework allows the user to quickly write and test distributed systems. It is efficient, and it automatically distributes the data and work across the machines and in turn, utilizes the underlying parallelism of the CPU cores.
- Hadoop does not rely on hardware to provide fault-tolerance and high availability (FTHA), rather Hadoop library itself has been designed to detect and handle failures at the application layer.
- Servers can be added or removed from the cluster dynamically and Hadoop continues to operate without interruption.
- Another big advantage of Hadoop is that apart from being open source, it is compatible on all the platforms since it is Java based.

2) Coding Challenges:

1. To check whether given number is even or odd

Pgrm1:

```
num = int(input("Enter a number: "))  
  
if (num % 2) == 0:  
  
    print("{0} is Even".format(num))
```

else:

```
print("{0} is Odd".format(num))
```