Assignment	**03**
Name	**Murtaza Jamali**
Course	**Big Data Analytics**
Instructor	**Moeed Tariq**
Submission date	**2*/11/2023**

What Is PSCP Utility

It provides a secure alternative to traditional file transfer methods and is commonly employed for secure file copying in networked environments. PSCP is particularly useful for transferring files to or from systems that support SSH, offering encrypted data transmission.

Block size vs Split size in Hadoop.

In Hadoop, block size refers to the size of data blocks in HDFS (typically 128 MB), while split size relates to the logical division of data for processing in MapReduce. While HDFS uses fixed size blocks, MapReduce split size is configurable, offering flexibility based on processing needs. Optimal tuning of these sizes is essential for efficient data storage and processing in a Hadoop cluster.

Serdes in the hive (e.g JSON, CSV) (Load a sample JSON file to hive table)

Json file:

```
{"name":"ravi","age":30,"details":{"a":100,"b":200},"city": "hyd"}
{"name":"hari","age":20,"details":{"a": 200, "b":300), "city":"usa"}
```

Solution:

```
use bd15;

Create Table:
create table jsontab(col1 string);

Load Data:
load data local inpath 'json.txt' into table jsontab;


Create another table:

create table jsontab2(name string,age int,details string,city string);


insert value jsontab to jsontab2:
insert overwrite table jsontab2 select
get_json_object(col1,'$.name'),get_json_object(col1,'$.age'),get_json_o
bject(col1,'$.details'),get_json_object(col1,'$.city') from jsontab;
```

Small file problem in Hadoop

The small file problem in Hadoop refers to challenges arising from managing numerous small files in HDFS. It leads to increased metadata overhead, inefficient block storage, and slower data processing.

Certificate ID Number: **8c4b7e59a3ae4a3ba42fd253431ecbf7** **November 27, 2023**