ASSIGNMENT

TASK 1:

1. NoSQLDATABASE:

NoSQL is an approach to database design that can accomodate a wide variety of data models, including key-value, document, columnar and graph formats. NoSQL is an alternative to traditional relational databases in which data is placed in tables and data schema is carefully designed before the database is built.

Features:

- Generic data model
- Data structure can be of any form
- No primary and foreign key concept
- Data stored in single table

2. Types of NoSQL DATABASE

- Wide-Column Store
- Document Store
- Key-Value Data Store
- Graph Store
- Multi-Model

3. CAP THEOREM:
   C-Consistency: data remains consistence after execution of operations
   A-Avalability: data should be available to all users
   P-Partition tolerance: system continues to function even if the communication among the severis unreliable.

4. Hbase vs RDBMS

HBASE	RDBMS
Column-oriented	Row oriented
Flexible schema	Fixed schema
Good with sparse tables	Not optimized for sparse tables
Joins using MR –not optimized	Optimized for joins
Tight integration with MR	Not tight integration
Horizontal Scalability	Hard to shard and scale
Good for semi and unstructured data	Good for structured data

Task 2:

1:Table created in hbase using create command

```
hbase(main):001:0> create 'bulktable','cf1','cf2'
0 row(s) in 3.8570 seconds

=> Hbase::Table - bulktable
hbase(main):002:0>
```

2:inserting values in the table on terminal widow.

File   Edit   View   Search   Terminal   Help
```
1 amit 3
2 girija 4
3 jatin 5
4 swati 3
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
```

3:put command: This command puts data from local filesystem to hdfs.

Cat command :This command shows us the contents of the file in hdfs.

```
[acadgild@localhost Hbase]$ hadoop fs -put bulk_data.tsv /hbase/
19/01/14 23:41:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
asses where applicable
[acadgild@localhost Hbase]$ hadoop fs -cat /hbase/bulk_data.tsv

19/01/14 23:41:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
asses where applicable
1 amit 3
2 girija 4
3 jatin 5
4 swati 3
You have new mail in /var/spool/mail/acadgild
```

4:Here we have used the command to copy data from hdfs to Hbase .

```
[acadgild@localhost Hbase]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.columns=HBASE_ROW_KEY,cf1:name,cf2:
exp bulktable /hbase/bulk_data.tsv
2019-01-14 23:45:45,146 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using b
uiltin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/Static
LoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2019-01-14 23:45:47,808 INFO  [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x6025e1b6 connecting to
ZooKeeper ensemble=localhost:2181
2019-01-14 23:45:47,854 INFO  [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-1569965, built on 02/20/
2014 09:09 GMT
```

5:Mapping done 100% .Job completed.

```
465047_0002/
2019-01-14 23:46:06,289 INFO  [main] mapreduce.Job: Running job: job_1547488465047_0002
2019-01-14 23:46:45,379 INFO  [main] mapreduce.Job: Job job_1547488465047_0002 running in uber mode : false
2019-01-14 23:46:45,387 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2019-01-14 23:47:09,281 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2019-01-14 23:47:11,719 INFO  [main] mapreduce.Job: Job job_1547488465047_0002 completed successfully
```

6:scan command :This command is used to view contents of files in Hbase.

```
hbase(main):020:0> scan 'bulktable'
ROW                            COLUMN+CELL
 1                             column=cf1:name, timestamp=1547490319068, value=amit
 1                             column=cf2:exp, timestamp=1547490324523, value=3
 2                             column=cf1:name, timestamp=1547490336338, value=girija
 2                             column=cf2:exp, timestamp=1547490485682, value=4
 3                             column=cf1:name, timestamp=1547490382001, value=jatin
 3                             column=cf2:exp, timestamp=1547490474287, value=5
 4                             column=cf1:name, timestamp=1547490405674, value=swati
 4                             column=cf2:exp, timestamp=1547490416376, value=3
4 row(s) in 0.3930 seconds
```