**Task1**

1) CREATE TABLE CourseDetails\_tmp(coursetype string, courseid string, title string, competency string, complexity string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

2) LOAD DATA INPATH '/user/ana002675/StudentCourseDetails.csv' INTO TABLE CourseDetails\_tmp;

3) create table CourseDetails (courseid string, title string, competency string, complexity string)

partitioned by(coursetype string)

row format delimited

fields terminated by ',';

3) INSERT OVERWRITE TABLE CourseDetails PARTITION(coursetype) SELECT courseid,title,competency,complexity,coursetype from CourseDetails\_tmp;

Text

Description automatically generated

CREATE TABLE StudentCourseCompletionStatus(

studentsid string, courseid string, examdate string, attendedstatus string, marks int , result string)

CLUSTERED BY (marks) INTO 5 BUCKETS

ROW FORMAT DELIMITED

fields terminated by ',';

LOAD DATA INPATH '/user/ana002675/StudentCourseCompletionStatus.csv' INTO TABLE StudentCourseCompletionStatus;

Text

Description automatically generated

HDFS location details

Text

Description automatically generated

**Task 2**

1.

select title as courseTitle,result,count(\*) cnt from (select cd.courseid ,title,competency,complexity ,coursetype, studentsid,examdate,attendedstatus,marks, result from CourseDetails cd join StudentCourseCompletionStatus sc ON cd.courseid=sc.courseid) f where result="Qualified" group by title,result;

Graphical user interface, text

Description automatically generated

2.

select distinct courseid,title from CourseDetails;

Text

Description automatically generated with low confidence

3.

INSERT OVERWRITE DIRECTORY "/user/ana002675/task3\_extract" ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT studentsid,title,result FROM StudentCourseCompletionStatus left outer join CourseDetails ON StudentCourseCompletionStatus.courseid=CourseDetails.courseid;

A screenshot of a computer

Description automatically generated