# Project: Student Competency Analysis

The competence of the students is the key to a college's success. To ensure competency level of student performance in various courses he participated and trained. The college placement officer is interested in getting insights about each student and recommending core companies based on his core strengths.

**Task 1: Defining Tables**

1. Create two internal Hive tables for the provided dataset –
   1. **StudentCourseCompletionStatus** -
      1. Distribute **StudentCourseCompletionStatus** data randomly into 5 buckets based on **marks** column
      2. Table Schema - studentsid string, courseid string, examdate date, attendedstatus string, marks int , result string
   2. **CourseDetails** -
      1. Split the **CourseDetails** table into two partitions using dynamic partitioning based on the **CourseType** column
      2. Table schema - coursetype string, courseid string, title string, competency string, complexity string

**Task 2: Data Analysis using Hive**

1. List the count of students qualified for various courses
2. List the courses available.
3. Extract the following fields from **StudentCourseCompletionStatus** file from HDFS directory and save the following result set to HDFS output directory -
   1. StudentID
   2. CourseTitle
   3. Result

**Task 3: PySpark**

1. Create PySpark dataframes and read data from the two Hive tables.
2. Find the following using the PySpark dataframes -
   1. Find the total number of students per result category
   2. Find the total number of students absent
   3. Find the maximum, minimum and average marks scored by students

**Task 4: Version control**

1. Create a Github account and create a new repository.
2. Create a new branch in the repository.
3. Upload the PySpark notebook to the new branch.
4. Merge the code to the master branch.

**Steps to connect to Hive -**

1. Hive connection
   1. open terminal and type following -
      1. **beeline**
   2. then connect to Itversity Hive warehouse -
      1. **!connect jdbc:hive2://m02.itversity.com:10000/;auth=noSasl**
   3. Enter Itversity username and password (from dashboard)
   4. Change warehouse to individual warehouse -
      1. **set hive.metastore.warehouse.dir = /user/ana002669/warehouse;**
   5. Create and new DB and use it for the project