## **Part 1 – Business Rules**

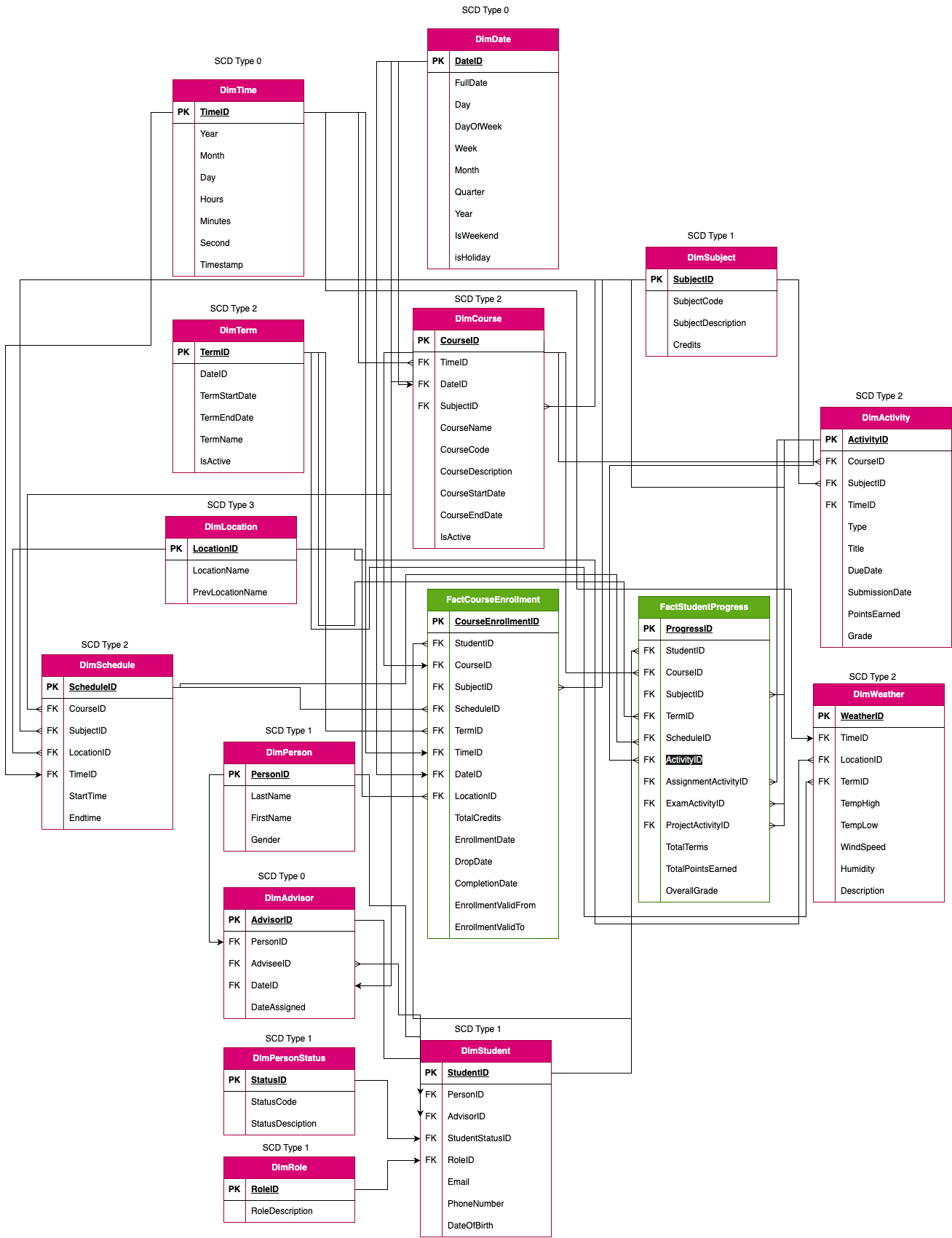
**1. Determine four business questions your data warehouse design will answer. Keep these questions in mind as you move on to the rest of the assignment. One of the questions needs to consider some sort of correlation with external weather data.**

* What percentage of students are receiving low grades, and which advisors are assigned to these students? This information can help determine if advisors should reach out for wellness check-ins. On the other hand, the students receiving good grades can get an update from their advisors that they are doing well in those respective terms.
* What is the average number of credits that students enroll in each semester?This information can help us understand student workload and performance trends.
* What types of course materials (assignments, paper discussions, projects, quizzes) are assigned to students, and how do these materials influence the points earned? Analyzing this relationship can provide insights into effective teaching methods and student engagement.
* How do weather conditions influence the submission dates of course assignments and the corresponding scores? Understanding this relationship can help identify external factors impacting student performance and engagement.

## Part 2 – Design a constellation schema warehouse

**Create and insert an ERD showing the constellation ERD schema below. Requirements are as follows:**

1. **Determine four to five (non-date/time) SCDs - make sure to include at least one type 2 and one type 3**
2. **Determine two to three date dimensions of different grains, consider a role-playing or bitemporal date dimensions. You will use table in question 2b to outline the facts and explain your design choice for the temporal dimensions**
3. **Determine two to three fact tables, your design should include at least one Snapshot (transactional) fact, and at least one cumulative fact.**



2. **Questions on Dimensions**

a. For EACH non-date/time SCD use the following table explain your SCD design

| **Table Name and Screenshot** | **SCD Type** | **Design explanation** |
| --- | --- | --- |
|  | 1 | Contains First and last name of student or advisor. Other fields can be added to this. This will be a FK for the Student and Advisor dimensions. The SCD Type is 1 because we do not need to track the history of previous names or other identifiable data, those can be overwritten. |
|  | 1 | Contains Status code and description. If we need to track the status code history across time, we may choose type 2. Since that is not needed in our case , type 1 is sufficient |
|  | 1 | This table contains student data, except the name is linked through the PersonID, along with the Advisor details, status and role. Since we do not need to track the history of previous email addresses, phone numbers or the date of birth, the SCD type for this dimension is type 1 |
|  | 2 | To track any likely changes that might occur for term dates, we are using SCD Type 2 |
|  | 1 | Since there is no history to track for the fields in this dimension, we are using SCD Type 1. If we need to track the history of changes across time so we can track how grades are affected by the amount of credits for that subject, we can use SCD Type 2 |
|  | 1 | This dimension has details about the advisor and their advisees, who will be the students. It also has a DateAssigned field which indicates when the student was assigned to the advisor. We can use SCD Type 1 since only the current advisees are relevant. This dimension can also serve as a transactional snapshot as it records assignments as they occur. |
|  | 2 | Since different subjects will have different schedules and we would want to track how these schedules would affect over time (according to weather conditions), we are using SCD Type 2 |
|  | 3 | Used SCD Type 3 for this so that we can keep track of where a class was scheduled in the past. |
|  | 2 | This is similar to Blackboard, where there is a track of the history of any changes made to assignments. Due Dates can be changed by the professor, and they can have a track of the old due dates for assignments. One activity can have multiple submissions, as we have in Blackboard. Hence, we have used SCD Type 2. |
|  | 2 | We need to maintain a comprehensive history of any changes made to courses, which is why we are using SCD Type 2. We can maybe add two more columns called StartDate and EndDate, which will track the duration of the old courses which are not active. |
|  | 2 | Contains weather data which we need a history of to track how the weather across time affects student grades. Hence, we are using Type 2 |

b. Outline the fact tables which contain role-playing or bitemporal design

| **Table Name and Screenshot** | **Temporal Type** | **Design explanation** |
| --- | --- | --- |
|  | Bitemporal | EnrollmentDate is when the student enrolled for the subject and EnrollmentValidFrom is the actual date from when the enrollment begins. Similar for CompletionDate and EnrollmentValidTo. So this is tracking both the transaction time and the valid time. |
|  | Role-playing | Using the Activity dimension, we can maintain a structure that tracks different types of activities for courses |

**3. Questions on Fact tables**

**a. For each of your business questions in part 1, how are the measures tied to your questions? Highlight the dimensions, facts and measures involved.**

| **Question 1:** What percentage of students are receiving low grades, and which advisors are assigned to these students? This information can help determine if advisors should reach out for wellness check-ins. On the other hand, the students receiving good grades can get an update from their advisors that they are doing well in those respective terms. | |
| --- | --- |
| **Screenshot of tables answering question** | **How measures are tied to your question** |
|  | Grades below a certain threshold can be filtered out from the fact which are derived from DimActivity and DimSubject. They can then be categorized according to courses using DimCourse. That can be linked to the relevant advisors using DimAdvisor, which will list out the advisors for students who have low grades. |
| **Question 2:**What is the average number of credits that students enroll in each semester? | |
| **Screenshot of tables answering question** | **How measures are tied to your question** |
|  | We can link FactCourseEnrollment to find the aggregate of credits taken up by students and we can then group them by the term. Since the fact is already referring DimStudent via the StudentID, we would be able to find the credits for each subject they have taken(via SubjectID) |
| **Question 3:**What types of course materials (assignments, paper discussions, projects, quizzes) are assigned to students, and how do these materials influence the points earned? | |
| **Screenshot of tables answering question** | **How measures are tied to your question** |
|  | We can join FactStudentProgress with DimActivity to see how different types(DimActivity.Type) of activities contribute to TotalPointsEarned. We can maybe include DimCourse to distribute them according to the courses. These will be selected according to the ActivityID,CourseID and StudentID. Finally, they will be grouped by ActivityType |
| **Question 4:**How do weather conditions influence the submission dates of course assignments and the corresponding scores? | |
| **Screenshot of tables answering question** | **How measures are tied to your question** |
|  | We can first join FactStudentProgress with DimActivity to link scores with submission dates, join with DimStudent to list out the activities of the students, join with DimDate to link submission dates with weather data and finally join with DimWeather to link those dates with the weather conditions. The average scores can be calculated via TotalPointsEarned based on submission date. |

b. **Question:** Outline the fact tables and explain why they are cumulative or snapshot

| **Table Name and Screenshot** | **Fact Type** | **Design explanation** |
| --- | --- | --- |
| FactCourseEnrollment | Cumulative | This table continuously updates with each enrollment or drop event across courses and terms. It provides a historical view showing how many students are enrolled in each course over time |
| FactStudentProgress | Snapshot | This table contains a students performance at a specific point in time, for each course they are enrolled in. It does not continuously update past records for activities or grades, aka data at discrete points in time |

c. **Question:** Which attributes in the OLTP schema will transform to measures and what measures can be derived/calculated that should be included?

| **Table and Attribute name from OLTP** | **Fact and Measure name** | **How is it transformed** |
| --- | --- | --- |
| Term, TermID | FactStudentprogress,TotalTerms | Calculate total number of terms based on termID for each student |
| DimActivity,PointsEarned | FactStudentProgress,TotalPointsEarned | Calculate sum of all points earned from all assignments in a term for a subject |
| DimActivity,Grade | FactStudentProgress,OverallGrade | Calculate Overall Grade using the grades from each activity from grade field of DimActivity |