This assessment addresses the following Unit Learning Outcomes:

1. Describe the common data sources that exist in organisations and their use in BI
2. Demonstrate practical skills in the processes associated with extraction, transformation and loading (ETL) of organisational data
3. Design and implement a simple data warehouse environment

# Submission Instructions:

Please submit a single word document (or similar, e.g., pdf is acceptable) using the link in Moodle. The due date is: **Sunday 1st March, 2020.**

From the Unit Information Guide:

It is your responsibility to **keep a copy of any work handed in for assessment purposes**. It is recommended that you keep both a hard copy and an electronic copy.

Assessments that are not received by the due date will be regarded as late unless an extension has been granted by the unit coordinator. Applications for extensions should be made as soon as a problem is experienced. Late submission will be penalised at a rate of five percent of the total marks available for the assignment per day (or part thereof). **No submission will be accepted more than two weeks after the due date, as assignment return will have begun**

# The Scenario

Silverback University has asked you to help them to design their student performance data mart. Your requirements elicitation with the clients has determined that they are interested in answering questions such as:

* Number of students in each unit in each year
* Number of students who have come to Silverback from various other institutions (e.g., how many students from IT Polytechnic, Downtown University etc) and which Silverback courses they have enrolled in.
* Number of students in each unit offering (i.e., ICT394 in Semester 1 2020, is a different unit offering to ICT394 in Semester 2 2020) by year.
* Grade distributions of students in units by year
* Grade distributions of students in units by offering and year
* Grade distributions by Lecturer by unit and offering
* Grade distributions by Course
* Grade distributions by School
* Grade distributions by Previous Institution (e.g., Results for students in IT courses from SP, etc).

There are a number of data sources from which you will need to draw your data. Below are the sources and the tables within them that are/may be of interest to you. These are discussed below (**Primary Key**, Foreign Key):

# Source 1: Course Handbook

The Course Handbook is a FileMaker Pro Database. It contains data regarding all courses, units and offerings of units that are offered by Silverback University. A course is made up of units, and a unit will have at least one offering each year.: COURSE (**CourseCode**, **Version**, CourseName, SchoolName)

UNIT (**UnitCode**, CourseCode, Version, UnitName)

UNIT\_OFFERING (**OfferingNumber**, UnitCode, Year, TeachingPeriod) OFFERING\_COORDINATOR (**StaffID**, **UnitOfferingNumber**)

# Source 2: Student Information System

The Student Information System has its data stored in a relational DBMS (Oracle) at present.

STUDENT (**StudentID**, StudentName, DateOfBirth)

ENROLMENT (**EnrolNumber**, StudentID, UnitOfferingNumber, Grade)

# Source 3: Human Resources System

The HR System is a proprietary system that is owned by the HR Department. STAFF\_MEMBER(**StaffNumber**, StaffName, SchoolCode) SCHOOL(**SchoolCode**, SchoolTitle)

# Source 4: Credit Database

The Credit Database records where students have studied prior to attending Silverback, and the amount of credit they receive toward their Silverback degree. The data are stored in a relational DBMS (MySQL) that was developed as a student project at Silverback.

STUDENT (**StudentNumber**, Student Name)

PREVIOUS\_INSTITUTION (**Institution\_Code**, Institution\_Name, Country) COURSE (**Institution\_Code**, **Course\_Code**, Course\_Name) PREVIOUS\_STUDIES (**StudentNumber**, **Institution\_Code**, **Course\_Code**, Credit\_Points)

# What you have to do:

**TASK 1 (30%)**: Discuss two (2) issues that may be problematic in the creation of the data warehouse that are apparent from the description above. For both, explain what you see as being the issue, why it is problematic, and what you will suggest needs to be done. This should take no more than two (2) pages in total.

The issue noticed from above are :

1. The table **STUDENT** appears in two different sources respectively (student information and credit database) with different primary keys. This may suggest that the student table in student information carries different records from the student table in the credit database. Therefore it poses a challenge when creating one dimension for student information as now we have details of students in Silverback indexed by **StudentID** and also we have credit details of students indexed by **StudentNumber** without any foreign key referenced from **studentID** to show that they are the same students.

**Suggestion of a solution**

Create one dimension for student details with primary key as **StudentID** and another dimension name it Previous with primary key **StudentNumber** to store details of students’ previous institution.

1. In the credit database there is Previous table which contains details of students’ previous institution credit points , the problem is that the table’s primary key is same as the student’s table in the credit database (StudentNumber), this poses a challenge when determining which table has the granular data between the two (which one looks up studentNumber from the other or which one will be used as the dimension to lookup student number).

**Suggestion of a solution**

**TASK 2 (25%)**: Based on the list of questions the client wants answered (see above), discuss what you see as being the most appropriate level of **granularity** for your data warehouse. Your discussion will need to **explain why** you have made this choice, and why the alternatives have been discarded. This should take not more than one (1) page.

**TASK 3 (30%)**: Assuming that the issues you have raised in TASK 1 have been addressed to your satisfaction, design a Star Schema that will support the analyses as listed above.

**TASK 4 (15%)**: Provide the SQL statements you would use to create the tables if you were to be implementing this design using Oracle.

# To Submit:

You will need to submit a single word-processed document that includes

* The written answers for Tasks 1 and 2
* The SQL statements for Task 4, and
* Screen shot of the design you have created for Task 3 (please do not submit your visio file, or whatever file you use, just a screen shot, copied and pasted into a word document please).

# What if I have questions?

It is most likely that you will have questions about this assignment. There will be a Discussion Forum on Moodle. Please use this, firstly, to see if you query has already been answered and, then, if it hasn’t, to ask your question.

# A final word…

This is an INDIVIDUAL ASSIGNMENT. While it is OK to discuss the assignment with other students, it is NOT OK to submit the same work. As final year students, the onus is on you to ensure that you do not overstep the mark as regards Academic Integrity.