INFO8115 - Group Project Presentation

Prepare all the following and place in a PowerPoint presentation to be delivered during week 14 of the course.

Part 1: Normalization/Star and Snowflake Schema:

* Using the Excel spreadsheet 'INFO8115 - Data Source for Normalizing' showing all the details of each of the steps to normalize from 1NF, 2NF, 3NF and a Star schema. When creating the slides from these, ensure you show the row contents as well
* Create a Snowflake of the Star Schema on the Date Dimension so that the database holds the last 12 rolling Moving Quarters (MQTR1=last month+2prior months, MQTR2=2 months ago+2prior months to that, etc.)   Example: if Jun/2022 just closed MQTR1=Apr/2022->Jun/2022, MQTR2=Mar/2022-<May/2022, etc.)

Part 2: Case Study - Big Data Architecting for new company

* The new Company has the following datasets that need to be integrated into a solution for Big Data reporting:
  + Operational/Transactional systems:
    - Finance
    - Marketing
    - Sales
  + In-house solutions
    - Finance Excel spreadsheets
  + Web Based
    - Product Reviews (unstructured and millions of entries)
* A Data Governance team will be formed so require Master Data Management
* The company will be significant is size so will require an ETL and related solutions to move data
* The company wants to have standard reports, dashboards, predictive analytics including an area for Data Scientists to analyze without requiring IT resources.
* Don’t worry about cost of the tools. You can as well if you want provide options, i.e., one that you know is free and another one that has a cost associated with it.
* Try to minimize the number of tools required to be purchased/supported overall.

Part 3: Prepare the shell of the PowerPoint presentation to contain:

* Cover page
  + includes eye catching picture, course name (INFO8115-{X99}-SEC{9}) team number, team name and team members
* Table of Contents
* Document history chart
  + most recent rows first
  + columns of Version, Date, Name, Activity
* Page Headers, Footers and Numbering
* RACI
  + Dividing the deliverables amongst team members
* Question/Answer slide
* References slide
  + To capture sites you research, etc.

Part 4: PowerPoint deliverables:

* For the ‘Normalization/Star and Snowflake Schema’
  + Create slides for the 1NF, 2NF, 3NF and 4NF, Star schema AND Snowflake schema.
  + Describe in detail how you moved from one normalization level to the next level.
* For the ‘Case Study - Big Data Architecting for new company’
  + Slides required:
    - Case study requirements as a slide
    - A slide for the attached ‘topology baseline’ diagram as a starting point.
    - A slide showing all the components you’re recommending aren’t required and are blanked out of the ‘topology baseline’ diagram.
    - A slide describing why you believe those components you recommended aren’t required by component or by group of components (i.e., new company-not now, too complex, not needed, etc.)
    - For all remaining components:
      * Create a slide for each tool that you would recommend taking care of each component with reasons why (i.e., highly recommended by Gartner, will take care of multiple categories on the ‘topology baseline’ diagram). Ensure when you prepare the slides that you show which component(s) you are specifically talking about for that slide.
      * Create as many slides as possible that are required to complete all the remaining components on the ‘topology baseline’ diagram
    - Summary slide of all tools that you’re recommending would accommodate all the components required. If possible, highlight the topology diagram to show the tools in some way. (i.e., different color circle for different tools, etc.)