**ACCT / MIS 6309 Business Data Warehousing**

**Instructor: Kevin R. Crook**

**Data Warehouse Design Project**

**General Instructions**

- This is an individual project, not a group project (due to the course objectives required to be met for accreditation). While more than one student could independently select the same subject matter, each student must turn in a unique project
- Students may ask other students to review each other's completed work for suggestions, but the substantial body of work must be original from the student
- Please use the provided word document template
- Please save the completed word document into PDF format before uploading
- Please submit the PDF file electronically through eLearning before the due date and time
- ERDs should be completed using ERWin using the same notation settings used in class. ERDs should be copied and pasted into the word document template

Select a business problem for a fictitious company that requires a data warehouse to solve:

- Write a 5 to 7 page summary of the business problem including the following:
    - A description of the fictitious company
    - A description of the industry the company is in and an analysis of where the company fits in the industry
    - A description of various business problems the company has that can be solved by designing and implementing a data warehouse
    - Select a data warehouse architecture and perform an analysis
        - Which method: Inmon, Kimball, or Stand-alone?
        - Analyze why you selected that architecture
        - Analyze why did you not select the other architectures

- Narrow down the various business problems to a subset of at least 5 specific business questions the executives want the data warehouse to be able to answer (at least 2 of the questions must be financial related)

- Using ERWin, design a subset of a data warehouse that will answer the 5 specific business questions (not the entire data warehouse which would be hundreds or thousands of star schemas)

    - Include the following in the design as a minimum:

        - Multiple stars with at least 3 fact tables
            - Following as a minimum
                - At least 1 factless fact table
                - At least 1 stovepipe
        - All dimensions should be as rich as possible
        - Following as minimum
          (ok if a dimension meets more than one requirement)
            - At least 1 highly browsable dimension
            - At least 3 affinity dimension
            - At least 1 junk dimension
            - At least 1 degenerate dimension
            - At least 1 time-stamped dimension
            - At least 3 conformed dimensions
                - At least 1 shared dimension
                - At least 1 conformed rollups
                - At least 1 overlapping dimension
        - At least 1 snowflake
            - At least 2 dimensions with outriggers
                - At least 1 of those dimension with 2 levels of outriggers
        - At least 1 bridge
            - Be sure as use the method shown in class to ensure the dimension can be conformed to other schemas

- Provided a written detail description of your above data warehouse design:

    - General description of each fact table

        - Short written description at a high level of what data it holds
        - Describe the grain
        - Which facts are additive, which are non-additive
        - What type of business problems can be answered from a single query to this star (can use SQL although not required)

- For each dimension
    - Short written description of the dimension
    - What type of dimension: affinity, junk, degenerate, etc.
    - What type of change: 1, 2, time-stamped, or 3
    - Browsability of the dimension including what business problems can be answered by browsing this dimension (can use SQL although not required)
    - List rich attributes and how are they derived
    - For a time-stamped dimension, how you you convert it to type 3 if space becomes an issue
    - What does it conform with and describe how (shared, rollup, overlap, etc)
- For each snowflake
    - Short written description of the snowflake
    - Describe the hierarchy it represents
- For each bridge
    - Short written description of the bridge
    - Describe the relationship that caused the bridge to be added
- Finally, for each of the 5 specific business questions
    - Describe how you would design a query to answer the question
        - Can use SQL if it's easier for you, but not required
    - Any drill across needed
- Give an example of a business question that cannot be answered due to a stovepipe
    - Short written description