HKBU Department of Computer Science

COMP 7810/4096 Business Intelligence

Individual Project (2019-20)

Please submit your project to BUelearning website

16 April 2020 (Thu) 6 pm

Consider a database of NBA and ABA statistics derived from [1]. The database contains 6 relations stored in excel formats:

1. Players – List of all players and their information
2. Regular Season – Regular season statistics for all players
3. Playoffs – Playoff statistics for all players
4. Teams – List of all teams and their information
5. Team Season – Regular season team offense and defense statistics, won and lost statistics
6. State – Teams' state code and state name

These 6 relations are not structured for efficient online analytical processing (OLAP). However, analysts of NBA and ABA prefer to analyze the teams' and players' statistics of NBA and ABA by using Microsoft SQL Server and SSDT.

Produce a report (in docx format) which contains the answers of tasks 1-4:

- Task 1: Import the Excel files to the database using SSMS. Combine **Teams** and **State** tables to form one new table [**Teams_State**] using a Transformation function in SSIS. **Capture the screenshots** for the transformation flow and describe the details for the ETL process. Perform **normalization** or **denormalization** if necessary. Design a **schema** of a data mart. Perform proper **data selection** and **stability analysis**. Briefly explain your schema design. Specify your assumptions.

- Task 2: **Implement the data warehouse** using **SSDT**, you should contain evidences of the implementation (such as screenshots of data source view, cube, deployment messages, and details in Solution Explorer). Add dimension intelligence if necessary. In addition, indicate the following clearly:
  1. Fact table(s)
  2. Dimension table(s)
  3. Primary key(s)
  4. Foreign key(s)
  5. Attributes and user hierarchies (if any) in each dimension

- Task 3: Create two **calculated fields** in your schema. The **offensive_index** is the first **calculated field** defined by the sum of all attributes of offense (i.e., the attributes with name o_xxx). Similarly, create the second calculated field called **defense_index** (i.e., sum of the attributes with name d_xxx)

offense_index = [sum of all attributes of offense (i.e., the attributes with name o_xxx)]
defense_index = [sum of all attributes of defense (i.e., the attributes with name d_xxx)]

- Task 4: Use **MDX** to get the results of the following questions. Copy the MDX and the screenshots of the results; paste them in your report. You need to remove NULL values for all questions (if any) and create a chart to show the results using SQL server report service (SSRS) for each question.

1. The `offensive index` and `defensive index` for the teams `Chicago Stags, Floridians` and `Orlando Magic`.
2. The `won` and `lost` index for each teams (with team code and team name) in the state of `Florida`.
3. The `top 10` teams (with team code and team name) having the best `offensive index` for all years.
4. The teams (showing team name) with `defensive index greater than 1300000`, sort the index in `descending` order.
5. The `top 5 players` (showing IDs and names) with the best `Pts` in `Playoffs` and `weight between 135 and 150` (using where clause).
6. The `won` and `lost` indexes for `BOS` team in `1950` and NYK team in `1951`.
7. The `bottom 5 teams` (showing team name) with the `worst defensive index` for all years.
8. The `won` and `lost` index for each team (showing team name) during `1950 to 1952`.
9. **Use ORDER ()** function with order specification **BDESC** to sort the `lost` index from highest to lowest for different teams during `1949 to 1950`. (Hint: You should NOT sort the performance within each year but for all 2 years)
10. The team (showing team code) with the best `offensive index` in each year.
11. The `turnover` index in Playoffs (with their IDs and names) where player's `position` is `G` and `turnover` index `greater than 300`
12. The top 5 players who played the longest `minutes` in `Regular Season` with position `F`.
13. Average `minutes` for different positions in `Playoffs`. (Format the average value to 1 d.p.) [Hint: you may need to use the measure **Playoffs Count]**
14. Compare the `offensive index` for each team (showing the team code) in `1951 and 1952` using **prevMember** or **nextMember**, and calculate the differences.

Deliverables

Please submit

1) your codes i.e. **visual studio project files** (in zip format), and
2) your report (in pdf format) [please covert the docx to pdf]

to the site http://buelearning.hkbu.edu.hk/

It is preferred to be concise, which should avoid excessive printouts of result data.

References:

[1] A Moore. Datasets and project suggestions. Available at http://www.cs.cmu.edu/~awm/10701/project/data.html. Last accessed Mar 2013.

[2] http://stackoverflow.com/questions/8321014/performance-tips-for-top-5-bs-for-each-a-mdx-queries-especially-ssas

[3] http://bimentalist.com/2011/08/29/msbi-28-ssas-9-modifying-measures-attributes-and-hierarchies-in-ssas/

Notes:

Penalty will be imposed on the act of academic dishonesty such as plagiarism, or submission of materials for assessment which is not your own work. A student found to have committed an act of plagiarism shall receive an "F" grade for the course. You are strongly advised to read the handbook on Avoiding Plagiarism, especially the section on "The Cost of Plagiarism". The handbook is available on the Academic Registry website at http://buar.hkbu.edu.hk/index.php/current_students_and_alumni/academic_guidelines/avoiding_plagiarism

Hints:

You may need to look up MDX reference books or online manual to look for "build-in" functions to complete your tasks, which is a common practice in real world.

Table 1 Meaning of Attributes (Non-exhaustive) of the Dataset

Attribute name	Meaning
fgm	Field Goals Made
fga	Field Goals Attempted
ftm	Free Throws Made
fta	Free Throws Attempted
3pm	Three Points Made (Team)
3pa	Three Points Attempted (Team)
tpm	Three Points Made (Player)
tpa	Three Points Attempted (Player)
oreb	Offensive Rebound
dreb	Defensive Rebound
reb	Total Rebound (oreb + dreb)
asts	Assists
pf	Personal Fouls
stl	Steals
to	Turnovers
blk	Blocks
pts	Points
gp	Game Played
o * or d *	Offensive Stats or Defensive Stats for Team