**Faculty of Computing, Engineering & Media (CEM)**

**Coursework Brief 2022/23**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Module name:** | Data Warehouse Design and OLAP | | | | | | | | |
| **Module code:** | IMAT 5167 | | | | | | | | |
| **Title of the Assessment:** | Data Mart Design | | | | | | | | |
| **This coursework item is:** ***(delete as appropriate)*** | | | | Summative | | |  | | |
| **This summative coursework will be marked anonymously: *(delete as appropriate)*** | | | | | | Yes | | |  |
| **The learning outcomes that are assessed by this coursework are:**   1. A comprehensive understanding of the concepts, purposes, architectures, evolution and benefits of DW. 2. A systematic knowledge of how to apply ERD and Star Schema to design databases, DW or data mart. 3. The ability to design appropriate data extraction, transformation and loading strategy and create reasonable queries. | | | | | | | | | |
| **This coursework is**: ***(delete as appropriate)*** | | | Individual | | | | |  | |
| If other or mixed ... explain here: | | | | | | | | | |
| **This coursework constitutes** 50 % **of the overall module mark.** | | | | | | | | | |
| **Date Set:** 4th November, 2022 | |  | | | | | | | |
| **Date & Time Due (the deadline):** | | **9/12/2022 at 12.00 noon** | | | | | | | |
| **In accordance with the University Assessment and Feedback Policy (**<https://www.dmu.ac.uk/about-dmu/quality-management-and-policy/academic-quality/learning-teaching-assessment/assessment-feedback-policy.aspx>), **your marked coursework and feedback will be available to you on:** | | | | | **9/01/2023** | | | | |
| You should normally receive feedback on your coursework by **no later than 15 university working days after the formal hand-in date,** provided that you have met the submission deadline  If for any reason this is not forthcoming by the due date your module leader will let you know why and when it can be expected. The Associate Professor Student Experience ([CEMstudentexperience@dmu.ac.uk](mailto:CEMstudentexperience@dmu.ac.uk)) should be informed of any issues relating to the return of marked coursework and feedback. | | | | | | | | | |
| **When completed you are required to submit your coursework via:**   1. As a Word Document via Turnitin (via the appropriate Turnitin facility set up within the Turnitin folder of IMAT 5167 module’s BB shell)   If you need any support or advice on completing this coursework please visit the Student Matters tab on the CEM Blackboard shell. | | | | | | | | | |
| **Tasks to be undertaken:** See attached. | | | | | | | | | |
| **Deliverables to be submitted for assessment:** See attached. | | | | | | | | | |
| **How the work will be marked:** See page 7 | | | | | | | | | |
| **Module leader/tutor name:** | Yingjie Yang | | | | | | | | |
| **Contact details:** | yyang@dmu.ac.uk | | | | | | | | |

Should you need any further information or advice please email [cemadvicecentre@dmu.ac.uk](mailto:cemadvicecentre@dmu.ac.uk)

**Midlands Theatre (MT) Company**

**Scenario**

***Company Overview:***

Midlands Theatre (MT) is a chain of small theatres that are found within the suburbs of several cities and towns within the Midlands, including Leicester and Birmingham. It specialises in high quality theatre productions that may be seen to be insufficiently “mainstream” for the general population. The company purchased its first theatre in January 1982, and since then has seen substantial increases in the number of clients that want to visit a local theatre to see a production that is more alternative to those typically found within the city centre theatres and/or large entertainment complexes. In response to the growing client base, MT sometimes offers a matinee (i.e., afternoon) performance of a popular production in addition to the traditional evening performance time.

Alongside the increase in number of theatre goers, comes the issue of managing the additional bookings and ticketing requirements. The Managing Director of MT, Ms. Heritage, developed a computerised booking system for MT. Although the ultimate aim is to enable the on-line booking of tickets for any MT productions at any MT theatres, Ms. Heritage has created a simple booking system first. This first version booking system has automated the existing manual booking processes, which are detailed below.

***Information regarding current booking processes at Midlands Theatre (MT):***

Each of the theatres in MT’s portfolio has a schedule of productions for a given month in a particular calendar year. For instance, a schedule for a production entitled “Wind Blows”, which ran for one week at the Cropston Theatre during March 2019, is provided in Table 1.

**Table 1. Example of one week performance at the Cropston Theatre**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Production** | **Day and Date** | **Afternoon** | **Evening** |
| **“Wind Blows”**  ***Scriptwriters***: William  ***Director:*** Henry | Monday 4/3/2019 | No performance | 7.30pm |
|  | Tuesday 6/3/2019 | No performance | 7.30pm |
|  | Wednesday 6/3/2019 | 2.30pm | 7.30pm |
|  | Thursday 7/3/2019 | No performance | 7.30pm |
|  | Friday 8/3/2019 | 2.30pm | 7.30pm |
|  | Saturday 9/3/2019 | 2.30pm | 7.30pm |
|  | Sunday 10/3/2019 | 2pm – note earlier start | 6.45pm – note earlier start |

**ER Diagram for MT’s first version, computerised booking System**



1..1

offers

Theatre

makes

has

has

0..\*

1..\*

1..\*

1..1

1..1

0..\*

Client

TicketPurchase

Performance

Client#

Purchase#

Per#

1..1

Production

Only one production is shown at a theatre at a time, and some of the more popular productions run both afternoon (matinee) and evening performances.

Productions could be at a theatre for several weeks, although the majority of productions run only for a few days over one week. The run of the next production cannot start until the previous production has ended (in other words there is no interleaving of the performance runs of different productions). It is common for the same production to have runs in all of MT’s theatres - one theatre after another - so that the widest number of potential clients for a production is reached.

When a potential theatre goer wants to purchase a ticket for a particular performance of a production at a MT theatre in advance, s/he contacts the central booking service via the website, telephone or an App. Whenever a ticket is purchased, a unique purchase number is allocated, the performance to be attended, the theatre, the client details (such as name, postcode and house number, and telephone), the total amount paid and the method of payment (cash, Paypal, debit card or credit card) and the method of ticket delivery (either by mail/email or by client pick up from the local theatre box office) are recorded. If several tickets are bought at the same time by the same person, each ticket is recorded as if it were bought separately.

**Table 2. Further Information about each Entity (Type) on the ER Diagram**

|  |  |
| --- | --- |
| ***Entity Type*** | ***Entity Instance*** |
| Theatre | A particular MT theatre e.g., Cropston |
| Production | A particular production e.g., How the Wind Blows |
| Performance | A particular performance of a production run, e.g., a performance of Wind Blows at Cropston Theatre on Monday 4/3/19 at 7.30pm |
| Client | A particular client of a theatre e.g., Mr Eric Charles |
| TicketPurchase | A particular ticket bought for a particular performance e.g., Ticket Purchase No. 29, which is for Cropston Theatre for the 9/3/19 performance at 7.30pm of Wind Blows and is bought by Ms Helen w2544. |

Any remaining tickets for a performance are available to buy on the day from the local theatre box office. The same details are recorded, although clearly in this case the ticket will always be collected from the local theatre box office!

A relational database has been created according to the ERD. The tables

are as follows:

* Theatre(**Theatre#**, Name, Address, MainTel);
* Production(**P#**, Title, ProductionDirector, PlayAuthor);
* Performance(**Per#**, *P#, Theatre#,* pDate, pHour, pMinute, Comments);
* Client(**Client#**, Title, Name, Street, Town, County, telNo, e-mail);
* TicketPurchase(**Purchase#**, *Client#, Per#,* PaymentMethod, DeliveryMethod, TotalAmount).

Having established the online transaction processing database, Ms. Heritage wants more intelligence information from the available data and she is looking for a potential data warehouse for MT. As the first step in this process, she expects to setup a data mart for ticket sales as the first step. The data mart should satisfy the following analysis requirements:

* The total sale value of each production.
* Monthly sale value of each theatre.
* The theatre name (for each theatre) and the names of clients who have the highest spending in that theatre.

Here, by ticket sale value/spending, we refer to the value of money rather than number of tickets. For this data mart, only the data involved in ticket sales are stored, and any data not involved in at least one ticket sale should not be included.

You can access the OLTP database in my database schema: ops$yyang00. For example, you can access the data for Theatres using the following SQL statement:

select \* from ops$yyang00.theatre;

**Your Tasks**

You are to develop a prototype of an ORACLE data mart for ticket sales as a part of a potential data warehouse for MT. For this assignment you are required to work individually.

You have been given the MT company scenario, together with an Entity Relationship (ER) Diagram and corresponding tables for the existing relational database.

* Task 1: Analyse the given database design and the requirement for a data mart, and identify the dimensions and fact for your data mart. Here for this assignment, only the essential (minimum) dimensions and attributes for the required queries need to be included.
* Task 2: Design the star schema for the data mart, and determine the relevant attributes and suitable granularity.
* Task 3: Map your star schema to logical relations and identify the corresponding PKs and FKs.
* Task 4: Create the corresponding tables in Oracle using SQL
* Task 5: Identify your source data from the OLTP database and design your data extraction rules. You need to give a detailed mapping and transformation list from the source to the destination.
* Task 6: Implement your data extraction, transformation and loading through Oracle SQL. The number of rows extracted into each dimension or fact table in your data mart should be printed from Oracle query.
* Task 7: Comment on how your data mart satisfies the requirements of MT. Implement the required data analysis requirements for both the data mart and the original OLTP database. Compare your queries and comment on the advantages of a data mart in analysis operations. Test results from Oracle query should be included.

Evidence of the successful execution of your queries is required. For large outputs, you need to provide only the last page of outputs with the number of rows in the end.

**Deliverables:**

The deliverable is a report that summarises your work and justifies your design decisions. In your DW assignment, please follow this format to organise your report. Please name all your entities, tables (logical relations) with a prefix of your student ID. For example, if your student id is P12345678, and your fact is ‘sp’, then you should name it as ‘P12345678\_sp’. Failure to do so may reduce your marks. Everything should be put into one MSWord document for your submission.

1. **Star schema design**
   1. Star schema

Please draw a star schema here

* 1. Explanation on your star schema (no more than half page)

1. What are the fact and dimensions and why they are chosen
2. Why their attributes are chosen
3. What is the granularity

1. **Derivation of logical relations**
   1. A list of logical relations (it has to follow the formal representation as our lecture notes)
   2. Explanation on the mapping from your star schema to the logical relations (no more than half page)
2. How is the logical relations mapped from your star schema (which relations is corresponding to which entity in your star schema)
3. What are the Primary Keys (PK) and Foreign Keys (FK) in the fact and dimensions. For each FK, what does it refer to ?
4. **Creation of tables with integrity rules**

For each table:

* 1. SQL code for each table creation
  2. Evidence of the code working and table created (small screenshots)

1. **Data source identification, data extraction, transformation and loading**
   1. Data source mapping

Please draw the map between data source and their destination in your Data Mart

For each ETL operation:

* 1. ETL code
  2. Evidence of ETL code works

Please provide the screenshot to show your code works and give the correct result. If the output is long, just show the last screen with the returned number of rows inserted.

1. **Justification of the data mart design and comparison of data mart and OLTP**

For each required analysis query

* 1. SQL code for the required queries (both Data Mart and the relational model)
  2. Evidence of the SQL code working and their results

Please provide the screenshots for your execution of your code and their results. If the result is too long, provide only the last screen.

For the comparison

* 1. Please compare your two sets of queries and comment on how they satisfy the requirements of the data analysis and the comparison between the Data Mart and the relational model (no more than one page).

**IMAT5167 DW&OLAP Criteria Marking Grid for DW Design Assignment**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TASK | *No work*  *0-10%* | *Little attempt*  *10-19%* | *Poor attempt*  20*-29%* | *Fail attempt**30-39%* | *Marginal fail**40-49%* | *Pass 50-59%* | *Merit 60-69%* | *Distinction*  *70-79%* | *Excellent*  *80-89%* | *Outstanding 90-100%* |
| Star schema design | No work | Attempted, but No idea on star schema | Poor attempt, inappropriate choice of dimensions and fact | Some effort insufficient granularity, dimensions and fact quality | Serious errors, but some evidence of understanding star schema | Satisfactory but with some significant errors/ limitations | As for Distinction level but with some minor errors/ limitations | All dimensions and the fact identified, correct granularity | Same as previous, no unjustified redundancy | As excellent. Minimum set of dimensions and facts in fact table. Perfect granularity |
| Derivation of logical relations | No work | Attempted but No idea on logical relations | Poor attempt, inappropriate choice of tables & attributes | Some effort but insufficient table/PK/FK quality overall. | Serious errors, but some evidence of understanding logical relations | Satisfactory but with some significant errors/ limitations | As for Distinction level but with some minor errors/ limitations | Correct choice of logical relations, and attributes, primary & foreign keys correct. | Excellent choice of logical relations, and attributes, primary & foreign keys | Same as excellent. Clear explanation on the map from star schema to logical relations. |
| Creation of tables with integrity rules | No work | Attempted, but No idea on how to create a table with integrity | Poor tables and attempt at integrity rules | Some effort but tables and integrity is not as reasonable quality | Some tables and associated integrity rules created but serious errors exist | Satisfactory  but with some significant errors/ limitations | As for Distinction level but with some minor errors/ limitations | Tables created correctly. Appropriate implementation of integrity rules | All tables created correctly. Appropriate implementation of integrity rules | Same as excellent.  Optimised table creation with minimum code and full set of constraints |
| Data source identification, data extraction, transformation and loading | No work | Attempted, but No idea on data source and ETL | Poor data sources, wrong or no transformation | Some effort but data source is not sufficient and transformation has significant errors | Some data sources are identified, and attempts are made for transformation | Satisfactory  but with some significant errors/ limitations | As for Distinction level but with some minor errors/ limitations | Data sources are correctly identified, good transformation rules are set and data are loaded correctly | All data sources are correctly identified, efficient transformation rules are set and all data are loaded correctly | Same as excellent. Minimum data sources and optimised ETL code produced, any potential problems discussed |
| Justification of the data mart design and comparison of data mart and OLTP | No work | Attempted, but No idea on data sources and queries | Little or no evidence of justification | Some effort insufficient understanding of data mart and weak justifications provided. | Some incomplete explanation are provided. | Satisfactory  but with some significant errors/ limitations | As for Distinction level but with some minor errors/ limitations | Correct justification for requirements with a good comparison between OLTP and DM queries | Excellent justification for all requirements with excellent comparison between most OLTP and DM queries | Same as excellent. More than one solutions are presented and difference discussed. |

**Overall Mark:**

**Overall Comments:**