IT471 Data Warehousing Fall 2020 – Final Project
Hotel Heaven

Hotel Heaven is a well-known chain of hotels that is held with a degree of regard proportional to the signs they have in front of their hotels. In recent years, the company's growth has become stagnant and profits been dormant. This is largely attributed to the fact that, unlike other hotel chains where "the lights are always on", when it comes to understanding the potential of business analytics Hotel Heaven's lights have yet to be switched on. In fact, the company has not even invested in a data warehouse.

This is about to change, and this case is about Hotel Heaven's efforts to build a data warehouse to begin addressing those needs.

While Hotel Heaven's vacancy rate for analytics is 100%, they are not completely vacuous because they do have data from a reservations system. To summarize that system, customers make reservations through the Hotel Heaven's website. Customers request the type of room they want (room with a king size bed, double beds, view of the "courtyard", etc.) at one of the hotels, along with the check-in and check-out date and any applicable discount code. The system checks if such a room is available, and if so, quotes a price. If the customer accepts, they proceed to make the reservation and the system issues a confirmation number. The information is then stored in a back-end database.

Keep in mind that operational systems (OLTP's) are used to run a business. They enable the business to function by handling business processes, which in turn consist of steps and procedures - as you see in the paragraph right above. As you know, operational systems are ill-suited for analytics.

The purpose of data warehouses is to enable managers to get the information they need quickly and without the need for technical knowledge on their part. Data warehouses are neither intended nor used to run business operations (such as shipping products, issuing paychecks, ordering from vendors, etc.). Data warehouses do not directly incorporate steps and procedures from operational systems, instead they focus on being able to satisfy the information needs of the business.

Hotel Heaven's management is most eager to analyze the reservations data, and not just because lately people seem to be checking out faster than they are checking in. Therefore, they have pointed one single business process – that of a customer making a reservation at one of their hotels – as the most important incorporate data from first in their data warehouse.

Your job for Hotel Heaven and this project is two-fold, first to design the data warehouse and then implement it with a data cube to show how well it meets the target. Each of these two parts is described next.

1. Design a data warehouse to waken Hotel Heaven from nightmares

The exercise discussing how to approach designing a data warehouse is available with this project. While you do not have to turn in the items it describes, such as a bus matrix, following the steps in that document will assuredly produce far better results.

The first step in that approach is to identify requirements, in other words, what the data warehouse is intended to be used for. This step can be difficult, especially for people doing this for the first time. Therefore, to give everyone a starting point here are the requirements to use and what Hotel Heaven dreams of having in their data warehouse:

Operational measures (O)

Hotel Heaven needs metrics for basic operations including the following:

- O1) Booked revenue – Total revenue based on the room rate that was confirmed at time of reservation,
- O2) Days booked – This can be the average number of days rooms are booked (reserved),
- O3) Room rate – This can be the average rate per day on reservations, and
- O4) Number of rooms booked – Hotel Heaven needs to look at the number of rooms reserved, at each hotel and by type of room (King-size bed, Luxury Suite, etc.).

Marketing (M)

Hotel Heaven wants to track reservations so they can market to each guest relentlessly until they threaten to sue, and at that point offer them a generous rewards program as a way of apologizing. To that end, they want to know:

- M1) Value – This is the total value of bookings that have been made by each guest (or customer).
- M2) Frequency - The number of customers who have made a single reservation and never come back, the number who have made two reservations, three reservations, and so on.

Pricing (P)

Hotel Heaven offers various discounts, such as corporate rates, AAA discounts, and group rates. In addition to evaluating the effectiveness of these discounts, Hotel Heaven would like to be able to set prices that will maximize occupancy rates, with lower prices to fill their hotels during slack time and increase them for surges in occupancy rates. As part of that effort Hotel Heaven needs to look at their occupancy rates:

- P1) For our purposes we can define this rate as the percent of rooms in a hotel that have been reserved (or booked).

Repeat Business (R)

Hotel Heaven knows that repeat business is key to their success and wants to be able to predict which customers are likely (or willing) make a future reservation. They want to start testing and evaluating different ideas aimed at increasing their chances of getting customers to come back, something which customers are generally loath to do.

As keen as they are to be able to do this, predictive analytics is new to Hotel Heaven. They are unsure of what data they will need but are sure that the data warehouse will be useful for that purpose so look to you for guidance.

R1) Hotel Heaven does not know what data they will need for this purpose, and so R1 is not defined as the items above are. Therefore, this requirement gives you the latitude to define what you believe would be useful and build it into your design.

While the other requirements all need to be based on the reservations database for this project, you may add columns to the tables in the database for R1. If you do, then add a few rows of data to them for the implementation part of this project.

Additional

Hotel Heaven knows there are many other metrics and analyses they need, and that you might have some good ideas for designing and implementing them in their data warehouse.

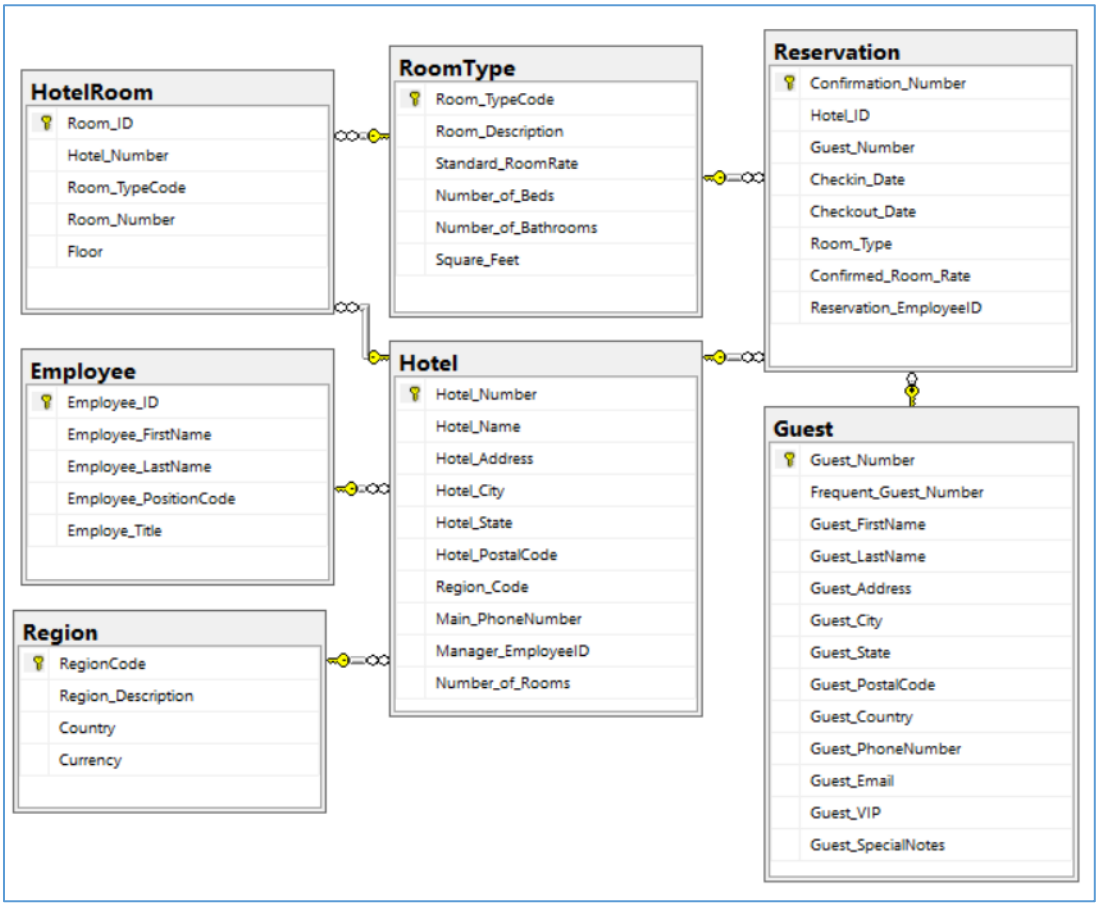
A1) This is for any other requirements beyond those lister above that you believe would be useful for Hotel Heaven and that you are able to define and incorporate in your design.

One way to approach this would be to look up metrics and measures used by hotels on the web, and then make sure those you choose can be calculated using only data you have in the reservations database for this project.

The requirements for Operational Measures (O) and Marketing (M) are important; Pricing (P) is also important but may be a bit more involved so may be near same category as the remaining two.

The last two requirements are open-ended and not as clearly defined as the others. This is to give you latitude to define them as you see fit, but because they might also be more difficult you should only do them if you can. In other words, it is ok if you do not fulfill some of the requirements, particularly R1 and A1, but if you do you will receive extra credit on your score.

As you would expect, the star schema you design needs to be based on the database Hotel Heaven for their transactional system for reservations. As you know, once a data warehouse has been designed and built, ETL processes are used to load data from transactional systems into the data warehouse. The schema from Hotel Heaven's reservation database is as follows:



Database schema from Hotel Heaven's OLTP reservation system

A backup file of a database with this schema is included with this project. After restoring it you should carefully review the design of the reservation system and open the tables to get a sense of Hotel Heaven's system and data. Several views are included in that database to help you explore and understand the company's systems and data.

As the transaction system will be the source of data for your data warehouse, the schema for reservations provided here will be the source of the attributes for all dimensions, attributes, facts, and measures in your design.

It is important that all attributes in your design be based on fields in the source database. In other words, the attributes in your design need to be either computed values, transformations, or direct usage of fields that can be found in that database.

It is equally important for your design to include all the attributes, dimensions, and measures that the end user would want to see and use. Remember that dimensions need to be "wide not deep".

Most importantly, your design needs to be to address the information needs of Hotel Heaven's management stated above.

2. Implement the data warehouse to bring Hotel Heaven sweet dreams

This part is for you to show how your design meets the requirements by building a data cube to implement it. It is recommended that you enter your design for Part 1 in the database provided for your project because you will have the tables ready to go for this part. In any case, you will need to start this part with the tables in your star schema design and then:

a) Enter data in the tables with your star schema design

Enter data in your star schema tables to represent about ten reservations.

It is best to start with your dimension tables, and for this you should use values that are from the database for this project. For example, if you were entering a field for a customer's name you would use one of the first and last names from the customer table in the transaction database (the database given for this project). If you are sure there are no corresponding values in the transaction database, you can use values of your choosing.

You do not have to enter all the rows from the tables in the database you were given or anything close to it. In fact, you do not have to enter a great deal of data for your dimensions. For example, if you had a dimension for SKUs you could decide that only three SKUs would be used in the fact table and so enter only three rows of data for that dimension. Also, you may have noticed that there is already a dimDate table in the database for you.

With your dimension tables complete you can enter fact table data because you will have the associated surrogate keys.

b) Build your cube

Use Data Tools in Visual Studio .NET and the data from your star schema tables.

Make sure you have only additive measures in your design and use calculated measures as appropriate. Use dimension hierarchies as appropriate and remember not to show key fields and to have formatted values in your cube.

c) Create an offline cube file

As we have done in previous projects, once you have your cube in an Excel pivot table it is done by using OLAP Tools – Offline OLAP mention options.

With this you can create the deliverables for this project as described next.

Deliverables

The deliverables for this project are two files. One is the offline cube file (**.cub** file) from the steps above; the other is a single pdf file with your name at the top and three sections:

1 - Design

- A screen shot of your star schema design.
- Add any comments you have, including assumptions you made, caveats about the design, or anything you want to point out about it.

2 – Requirements

- List the requirements that you have met with your design and how it meets them. More specifically, for each requirement show the calculation(s) it would use using column names in your design. For example, if requirement M1 were for a metric based on average price and you were working with a different database, you might have:

  M1) Average price

  SUM(factOrders.GrossSales) / COUNT(factOrders.GrossSales)

3 – Implementation

- For each requirement you met, indicate the requirement with at least one screenshot of a pivot table from your cube showing demonstrating how it was met. To continue the example of from above for a different database, for this section you might have something along the lines of:

  M1) Average price metric - for each product (SKU) and over time

SKU	Metric for Average Price
100200 Std. Scuba Tank, Magenta	$116.06
101100 Dive Mask, Small Clear	$106.00
101200 Dive Mask, Med Clear	$105.05
201000 Half-dome Tent	$111.95
202000 Half-dome Tent Vestibule	$110.00
204000 Toasty Sleeping Bag	$156.92
204010 Fry 'em Cooking Stove	$94.23
Overall	**$110.81**

Quarter	Metric for Average Price
2018	
2018-Q1	$110.54
2018-Q2	$107.14
2018-Q3	$111.27
2018-Q4	$119.93
Overall	**$110.81**

As always, please add any comments, clarifications, or suggestions throughout your document.

- - - - - - - -

Lastly, management is counting on you to bring Hotel Heaven to Data Warehouse Heaven. As you can see from the sign hanging on their flagship hotel, they really need it.

Grading

The grading for this project will be based the following factors.

First, for the pdf document:

1) Organization and format

   Is a .**pdf** file type submitted? Is it structured as listed for the deliverables? Is it easy to follow (for example, making sure the requirements - O1, O2, etc. - that apply to each pivot table screenshot are shown)? Are all screenshots readable?

2) Standards and guidelines

   How well does the design meet the guidelines and standards for star schema designs? These are the same as previous projects and are set out in the document **Checklist for Star Schema designs.pdf** on Blackboard? These specifically include guidelines for formatting, naming standards, and diagramming conventions.

3) Star schema structure

   Is the star schema in the design appropriately structured? Do keys match dimensions, are surrogate and natural keys present, are measures mixed with dimensions, are all measures additive?

4) Granularity

   In the design, does the granularity of the dimension tables match that of the associated fact table? Is each measure at the same level of detail (i.e., one-to-one relationship) with respect to each dimension connected to their fact table as well as the other measures in it?

5) Requirements

   How many and which of the requirements are listed as being met? For those listed as being met, how well do they actually meet those requirements? Do the requirements listed as being met match with the star schema in the design? Were any of the optional requirements addressed (and therefore extra credit)?

6) Implementation

   How well do the screenshots of pivot tables match the requirements being met? How effectively do they demonstrate that the requirement was met?

For the offline cube file:

1) File type

   Was a **.cub** file uploaded? (The file you upload needs to be an offline cube file that you create using Excel using the way as we have done previously. Keep in mind this needs to be a **.cub** file and not a VS.NET project, zipped file, .cube file, or any other file type).

2) Guidelines

   How well formatted and complete is your cube? Are all key fields hidden? Are calculated measures and dimension hierarchies used appropriately?

3) Completeness

   Is the cube complete? Does it match the star schema? Does the cube match the screenshots presented and capable of the requirements listed as being met?