# DS 206 Business Intelligence: Final Exam

|  |  |  |  |
| --- | --- | --- | --- |
| Instructor: | Arman Asryan | Name: |  |
| TA | Yelena Hakhumyan | Semester: | Spring, 2022 |
| Course | Business Intelligence | Date: | May 16, 2022, 12:30-14:30 |

**PRELIM: C**lick **Enable Content/Editing** when opening the file (if the pop-up appears).

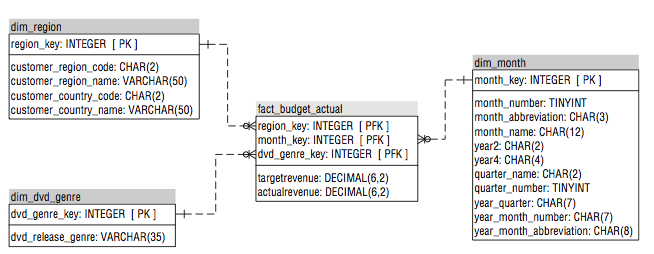
**PRELIM:** Make sure to insert Your full name in the **Name** field of the table above.

**PRELIM:** Make sure to rename this file to **Name\_Surname\_DS206\_FinalExam.docx**.

**DESCRIPTION:** The exam is composed ofquestions. There are multiple-choice, short-answer, fill-in-the-blank, drag-and-drop (copy-and-paste) questions included in the exam. The last two questions are about dashboards and data ingestion respectively. If

**PRELIM:** You encounter technical difficulties when typing in the **[Type an answer here]** boxes, feel free to remove them and just insert Your answer instead.

1. **(3 points)** Identify the type of the fact table in the star schema below:



Consolidated fact table

1. **(3 points)** Which measure is used to assess whether there are too many visuals and/or information in a BI solution?

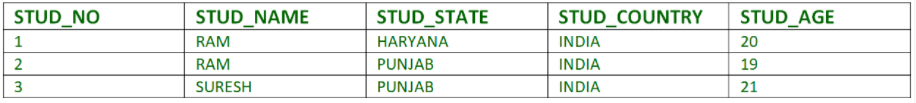
Data-ink ratio

1. **(6 points)** The BI team has found out they would not be able to fulfill all the requirements under the planned budget within the required time frame. Suggest workarounds for the situation (write as many as You can).

Lecture 1 [Page 50]

* Proceed with a set of must-haves that can be accomplished within the existing plan and budget.
* Reduce the scope of some of the must-have requirements so that they can be accomplished within the existing plan and budget.
* Increase the budget and lengthen the schedule based on a revised list of requirements for the BI project.

1. **(3 points )** Which normal form does the table depicted below violate ? If you find that it does not violate any normal forms, write *None*  instead.



3NF (No transitive dependency)

1. **(6 points)** In order to eliminate the existing data shadow systems and to implement a new BI infrastructure, the company hires an experienced BI team. During the first technical meeting, one of the BI team members kept insisting that the existing data shadow systems should be used when designing the newer alternative. Explain how the existing data shadow systems can contribute to the new BI infrastructure.

Lecture 1 [Page 48]

* Contribution 1: Although business users may complain about their existing reports or data shadow system, they have been using them for analysis for some period of time, so these systems are now the benchmark for comparison.
* Contribution 2: Even when the old systems are being discarded and the business is providing ”new” requirements, it is fairly common for many of the business rules and data used by the business in the old systems to be relevant. Reverse engineering fills in the inevitable gaps in detailed requirements, keeping the new BI solution from being derailed when business users start using it.

1. **(5 points)** What are the dimensions upon which overlapping, redundant, outdated, and conflicting requirements may be aggregated (highlight all that apply).
   1. Business value
   2. Metadata
   3. Business roles
   4. Business processes
   5. Business group
   6. Data sources
   7. Analytical functionality
   8. None of the above.
2. **(5 points)** What are the limitations of using EDW (enterprise data warehouse) without data marts (highlight all that apply)

* not supporting robust BI,
* not having a correct data served to users,
* being inflexible with the inevitable changes in data requirements,
* not being able to correctly aggregate data across the desired dimensions.

1. **(5 points)** Name the six principles upon which the data integration approach(es) must be architected:
2. Iterative
3. Holistic
4. Incremental
5. Documented
6. Reusable
7. Auditable
8. **(5 points)** Match the following concepts with their definitions:

|  |  |  |  |
| --- | --- | --- | --- |
| **Concept** | **Insert the matching number** | **Number** | **Definition** |
| Information architecture | 4 |  | defines the data along with the schemas, integration, transformations, storage, and workflow required to enable the analytical requirements of the information architecture. |
| Technical architecture | 3 |  | the products, their configurations, and how they are interconnected to implement the technology requirements of the BI framework. |
| Data architecture | 1 |  | the technologies that are used to implement and  support a BI solution that fulfills the information and data architecture requirements |
| Product architecture | 2 |  | the business context necessary for successful BI solutions to be built on a sustaining basis. |
| Data preparation | 6 |  | the description of the data as it created, transformed, stored, accessed, and consumed in the enterprise |
| Metadata | 5 |  | the core set of processes for data integration that gather data from diverse source systems, transform it according to business and technical rules, and stage it for later steps in its life cycle |

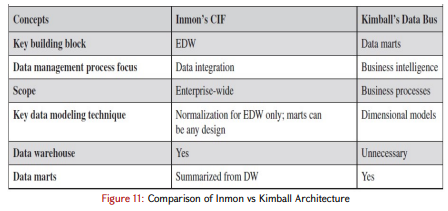
1. **(3 points)** The company has a BI environment that can be characterized by the following statements:

* There are multiple BI and performance management (PM) tools or applications scattered throughout the given enterprise.
* The company continuously engages new vendors to bring in new BI solutions to make up for the current environment’s shortcomings.
* There are many data shadow systems in the company. Moreover, there is a continuous need to build more in order to successfully cope with reporting and analysis tasks.

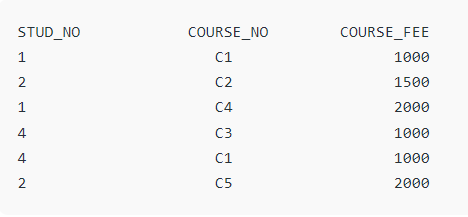
Summarize the situation (problem) the company might be facing using a single expression/sentence.

Accidental Architecture

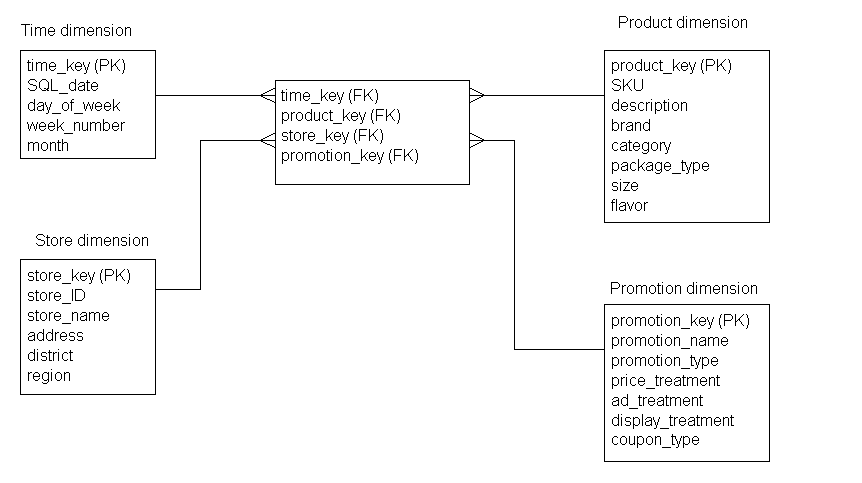
1. **(5 points)** Select (underline) whether each of the following statements is true or false.
   1. Data franchising is responsible for designing conforming dimensions, managing slowly changing dimensions, updating dimensional hierarchies. (True / False)
   2. Data preparation is the data cleansing processing necessary to accumulate data from numerous sources. (True / False)
   3. Operational BI is responsible for shifting the business users to a specific application reporting silo. (True / False)
   4. Data franchising enables self-service BI. (True / False)
   5. Business transformations and metric calculations are performed at the data franchising stage. (True / False)
2. **(6 points)** Explain the primary differences between the *corporate information factory* and *enterprise data bus* architectures.



1. **(5 points)** Select (underline/highlight) whether each of the following statements is true or false.
   1. The foreign keys in fact table can be null. (True / False)
   2. The source system’s primary key is usually maintained in the dimension table. (True / False)
   3. Transaction fact tables store a record for the entire lifetime of the event. (True / False)
   4. Periodic fact table has the lowest level of granularity. (True / False)
   5. You can consolidate facts that do not have the same grain into a consolidated fact chart. (True / False)
2. **(3 points )** Which normal form does the table depicted below violate ? If you find that it does not violate any normal forms, write *None*  instead.



2NF (No partial dependency)

1. **(5 points)** In ADA (analytical data architecture), the data model of the EDW (enterprise data warehouse) integration schema must be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Meanwhile, the distribution schema of the EDW should adhere to the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Select the most inclusive alternative solution.
   1. dimensional, hybrid dimensional-normalized
   2. 3NF or hybrid dimensional-normalized, 3NF
   3. same as the source system, 3NF
   4. 3NF or hybrid dimensional-normalized, hybrid dimensional-normalized
2.  **(3 points)** Identify the type of the fact table in the star schema below:

Event (Factless) fact table

1. **(5 points)** Copy and paste the following characteristics to their respective groups.
   1. Minimal redundancy,
   2. Increased index use,
   3. Eliminate inconsistent data,
   4. Increased maintenance concerns
   5. Efficient use of storage space
   6. Increased redundancy

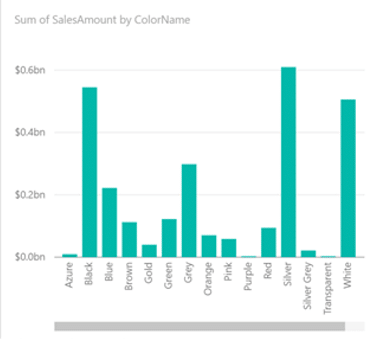
|  |  |
| --- | --- |
| **CHARACTERISTICS SUPPORTING THE USE OF NORMALIZED MODELS** | **CHARACTERISTICS SUPPORTING THE USE OF DIMENSIONAL MODELS** |
| Minimal redundancy | Increased redundancy |
| Efficient use of storage space | Increased index use |
| Eliminate inconsistent data | Increased maintenance concerns |

1. **(6 points)** Explain the differences between the **SCD Type 6** and **SCD Type 3**.

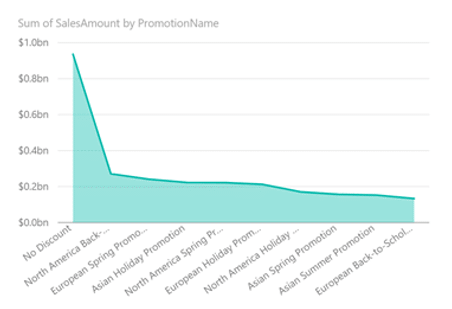
With the SCD type 3 technique, you continue to overwrite the rows as in type 1, but you keep track of the new or current value of that particular attribute as well as the original or the previous value in separate fields at the time of change.

The SCD type 6 technique blends portions of SCD types 2, 1, and 3. Similar to type 2, a type 6 schema creates a new record with every change in attribute values that are being tracked. The current value (type 1) of a specific attribute is added as a column (type 3) onto every row. This allows rows to be filtered or grouped by either the current or historical value of the type 3 column. The type 2 portion of this technique enables ”as is” and ”as was” analysis.

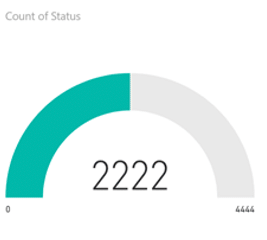
1. **(7 points)** What are the problems with the following dashboard (bullet points)?
2. Filters should be on the left side of the page.
3. The following bar chart should be horizontal (given the number of categories) and sorted:



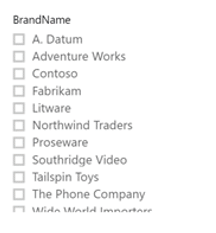
1. Data-ink ratio is too high (too many visuals on a single page).
2. Injudicious use of colors (comparison and contribution plots don’t have color differentiation at all),
3. Wrong chart type chosen (should have been a Treemap or a bar chart)



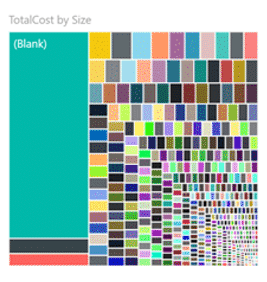
1. Wrong chart (should have been a card visual)



1. The following slicer is scrollable:
   1. Too many categories (worth grouping),
   2. Too small dedicated area

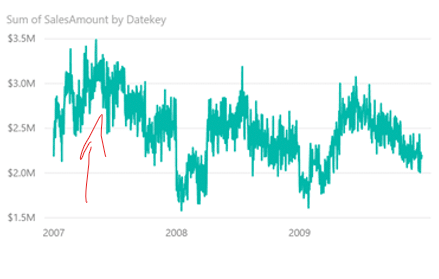


1. Wrong visual chosen:



* 1. Treemap is good but the number of categories is too large (worth grouping).
  2. There is no legend for connecting various colors with different sizes.
  3. (Blank) – most likely a very big chunk where Size = NULL (should have been labeled properly).

1. No date hierarchy (the date filter only has years, while there are details pointing to the fact the data is at least monthly):



1. The titles of visuals are not embedded judiciously (they are the default ones set by Power BI Desktop).

**View online:**

[**https://drive.google.com/file/d/1kYHYr9fNnXAzFatTDLzmT\_AqJguqA2AU/view?usp=sharing**](https://drive.google.com/file/d/1kYHYr9fNnXAzFatTDLzmT_AqJguqA2AU/view?usp=sharing)

***Download (in case You cannot view the PNG online):***

[**https://drive.google.com/uc?export=download&id=1kYHYr9fNnXAzFatTDLzmT\_AqJguqA2AU**](https://drive.google.com/uc?export=download&id=1kYHYr9fNnXAzFatTDLzmT_AqJguqA2AU)

1. **(11 points)** The client has given the BI team a task to implement a Type 1 slowly changing dimension table for its customers. The BI specialists were given the source and destination table design scripts, as well as the initial stored procedure (SQL).
   1. Debug the stored procedure script and make it work. **(5 points)**
   2. Add the DELETE operation to the stored procedure **in a new textbox**. **(6 points)**

Designing the source and the destination tables

-- CREATING THE SOURCE TABLE

DROP TABLE IF EXISTS [dbo].[Customer];

GO

CREATE TABLE [dbo].[Customer](

[id] [int] IDENTITY(1,1) NOT NULL,

[customer\_name] [varchar](150) NULL,

[country] [varchar](50) NULL,

CONSTRAINT [PK\_Customer] PRIMARY KEY CLUSTERED

([ID] ASC) WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]) ON [PRIMARY];

GO

-- POPULATING THE SOURCE TABLE

INSERT [dbo].[Customer] ([customer\_name], [country]) VALUES ('John Smith', 'UK')

INSERT [dbo].[Customer] ([customer\_name], [country]) VALUES ('Bauhaus Motors', 'UK')

INSERT [dbo].[Customer] ([customer\_name], [country]) VALUES ('Honest Fred', 'UK')

INSERT [dbo].[Customer] ([customer\_name], [country]) VALUES ('Fast Eddie', 'Wales')

INSERT [dbo].[Customer] ([customer\_name], [country]) VALUES ('Slow Sid', 'France')

GO

-- CREATING THE DESTINATION TABLE

DROP TABLE IF EXISTS [dbo].[Customer\_SCD1];

GO

CREATE TABLE [dbo].[Customer\_SCD1](

[client\_id\_sk] [int] IDENTITY(1,1) NOT NULL,

[business\_key\_nk] [int] NOT NULL,

[customer\_name] [varchar](150) NULL,

[country] [varchar](50) NULL

) ;

GO

Designing the stored procedure for ingesting data into the dimension table

DROP PROCEDURE IF EXISTS dbo.Customer\_SCD1\_ETL

GO

CREATE PROCEDURE dbo.Customer\_SCD1\_ETL

AS

BEGIN TRY

MERGE dbo.customer\_scd1 AS DST

USING dbo.customer AS SRC

ON ( SRC.id = DST.business\_key\_nk )

WHEN NOT MATCHED THEN

INSERT (client\_id\_sk,business\_key\_nk,customer\_name,country)

VALUES (SRC.id,SRC.id,SRC.customer\_name,SRC.country)

WHEN MATCHED AND (Isnull(DST.customer\_name, '') <> Isnull(SRC.customer\_name, '') OR

Isnull(DST.country, '') <> Isnull(SRC.country, ''))

THEN

UPDATE SET DST.customer\_name = SRC.customer\_name,

DST.country = SRC.country;

END TRY

BEGIN CATCH

THROW

END CATCH

GO

--a

DROP PROCEDURE IF EXISTS dbo.Customer\_SCD1\_ETL

GO

CREATE PROCEDURE dbo.Customer\_SCD1\_ETL

AS

BEGIN TRY

MERGE dbo.customer\_scd1 AS DST

USING dbo.customer AS SRC

ON ( SRC.id = DST.business\_key\_nk )

WHEN NOT MATCHED THEN

INSERT (business\_key\_nk,customer\_name,country)

VALUES (SRC.id,SRC.customer\_name,SRC.country)

WHEN MATCHED AND (Isnull(DST.customer\_name, '') <> Isnull(SRC.customer\_name, '') OR

Isnull(DST.country, '') <> Isnull(SRC.country, ''))

THEN

UPDATE SET DST.customer\_name = SRC.customer\_name,

DST.country = SRC.country;

END TRY

BEGIN CATCH

THROW

END CATCH

GO

-- b

GO

CREATE PROCEDURE dbo.Customer\_SCD1\_ETL

AS

BEGIN TRY

MERGE dbo.customer\_scd1 AS DST

USING dbo.customer AS SRC

ON ( SRC.id = DST.business\_key\_nk )

WHEN NOT MATCHED THEN

INSERT (business\_key\_nk,customer\_name,country)

VALUES (SRC.id,SRC.customer\_name,SRC.country)

WHEN MATCHED AND (Isnull(DST.customer\_name, '') <> Isnull(SRC.customer\_name, '') OR

Isnull(DST.country, '') <> Isnull(SRC.country, ''))

THEN

UPDATE SET DST.customer\_name = SRC.customer\_name,

DST.country = SRC.country

WHEN NOT MATCHED BY Source THEN

DELETE

OUTPUT $action,

DELETED.business\_key\_nk AS business\_key\_nk\_d ,

DELETED.customer\_name AS customer\_name\_d,

INSERTED.business\_key\_nk AS business\_key\_nk\_s,

INSERTED.customer\_name AS customer\_name\_s;

END TRY

BEGIN CATCH

THROW

END CATCH

GO