DOMANDE E RISPOSTE DI DATA MINING

Nella `parte 1` si trovano le domande senza highlighting delle risposte corrette, mentre nella `parte 2` si trovano le domande comprendenti delle risposte corrette in **grassetto**.

PARTE 1: Domande senza risposte corrette

1) In which part of the CRISP methodology we perform the test design activity?

- Data Understanding
- Evaluation
- Business Understanding
- Modeling

2) Talking about the general idea of database, what is the purpose of the "Schema on write" strategy?

Scegli una o più alternative:

- Optimisation for specific types of queries
- Clean design of the data structure
- Avoid preprocessing of data before writing
- Flexibility and efficiency for any kind of query

3) Talking about the general idea of database, what is the meaning of "Schema on write"?

Scegli un'alternativa:

- Create a schema for data after writing into the database
- Create a schema for data while writing into the database
- Change the schema of the data after each write
- Create a schema for data before writing into the database

4) Talking about the general idea of database, what is the purpose of the "Schema on read" strategy?

Scegli una o più alternative:

- Avoid preprocessing of data before writing
- Optimisation for various types of queries
- Possibility to extract data in various shapes
- Flexibility for any kind of query

5) Talking about the general idea of database, what is the meaning of "Schema on read"?

Scegli un'alternativa:

- Write the raw data in the database, then at reading time shape then according to the reader's needs
- Create a schema for data after reading them from the database
- Read the schema of the data before each data read
- Change the schema Of the data before each read

6) Which of the definition below describes the OLAP operation Roll-Up?

Scegli un'alternativa:

- Reduces data aggregation and adds a detail level to a hierarchy
- Changes the layout, in order to analyse a group of data from a different viewpoint
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value
- Creates a link between concepts in interrelated cubes, to compare them
- Causes an increase in data aggregation and removes a detail level in a hierarchy

7) Which of the definition below describes the OLAP operation Pivot?

Scegli un'alternativa:

- Causes an increase in data aggregation and removes a detail level in a hierarchy
- Reduces data aggregation and adds a detail level to a hierarchy
- Changes the layout, in order to analyse a group of data from a different viewpoint
- Creates a link between concepts in interrelated cubes, to compare them
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value

8) Which of the definition below describes the OLAP operation Drill-Down?

Scegli un'alternativa:

- Reduces data aggregation and adds a detail level to a hierarchy
- Creates a link between concepts in interrelated cubes, to compare them
- Changes the layout, in order to analyse a group of data from a different viewpoint
- Causes an increase in data aggregation and removes a detail level in a hierarchy
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value

9) Talking about ETL, which of the following activities is related to the Transformation step?

Scegli una o più alternative:

- Usage of dictionaries to solve inconsistencies
- Snapshot of the operational data
- Denormalisation
- Calculation of derived data

10) Talking about ETL, which of the following activities is related to the Cleansing step?

Scegli una o più alternative:

- Elimination of duplicates
- Association of a timestamp to the operational data
- Usage of dictionaries to solve inconsistencies
- Snapshot of the operational data

11) Talking about ETL, which of the following activities is related to the Extraction step?

Scegli una o più alternative:

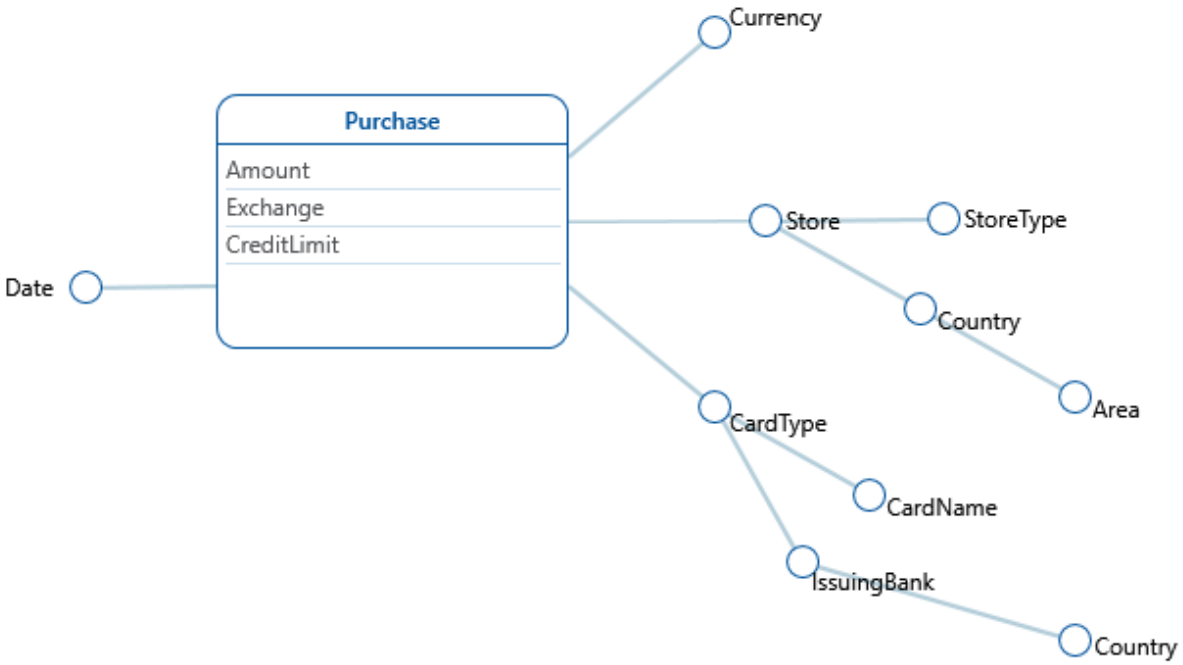
- Elimination Of duplicates
- Snapshot of the operational data
- Association of a timestamp to the operational data
- Usage Of dictionaries to solve inconsistencies

12) Order the ETL operations in order to obtain the correct sequence

1. Cleansing
2. Transformation
3. Extraction
4. Loading

Order: ___

13) Which of the texts describes the DFM schema below?



Scegli un'alternativa:

- We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type; stores have also a country and an area. Purchases have a card type (e.g. credit or debit) and a card name (e.g. MyCard or JeansCard). Card types have also an issuing bank and a country.
- We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type and country; stores have also an area. Purchases have a card name (e.g. MyCard or JeansCard)t card_names have a card type (e.g. credit or debit). Card types have also an issuing bank and a country.
- We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type; stores types have also a country and an area. Purchases have a card type (e.g. credit or debit) and a card name (e.g. MyCard or JeansCard). Card names have also an issuing bank and a country.

14) Which of the following are typical data warehouse queries?

Scegli una o più alternative:

- Which products maximize the profit?
- Which is the stock of product "XXX" in the "YYY" warehouse
- What is the total revenue of yesterday in shop "B003" ?
- What is the total revenue per product category and state?
- What is the relationship between profits gained by products "WWW" and "ZZZ"?

15) Select the sentences describing a characteristic of a Data Warehouse

Scegli una o più alternative:

- Solves the inconsistencies
- Non volatile
- Includes the time dimension
- Constantly updated as soon as the base data are updated
- Includes all the base data in their native format

16) Associate the various Data Warehouse Architectures to the respective advantages

Architectures:

1. Single Layer
2. Two Layers
3. Three Layers

Advantages:

- The workloads for analysis and daily operations are separated
- An intermediate level of consolidated data is available
- The occupation of space is minimised

17) Which of the activities below is part of "Business Understanding" in the CRISP methodology?

- Which data are available?
- Which machine learning functions are necessary for my problem?
- Which data must be collected with a specific campaign?
- Which are the resources available (manpower. hardware, software,...)

18) Rank the technologies for increasing level of abstraction: 1 = generate the most specific information, ... 3 = higher level knowledge

- SQL queries on operational databases → ___
- Data Mining on operational databases → ___
- OLAP analysis on Data Mart → ___

19) Which of the following sentences describes an advantage of a Data Warehouse with respect to a standard DBMS

Scegli una o più alternative:

- Allows efficient execution of key-based queries
- Manages efficiently data updates
- Has tools for helping to solve inconsistencies
- Allows analysis along the time dimension
- Allows efficient execution of multi-dimensional queries

20) Link the names of the OLAP operations below to their definitions

names:

- Slice dice → ___
- Drill across → ___
- Pivot → ___
- Drill down → ___
- Roll-up → ___

definitions:

1. Creates a link between concepts in interrelated cubes, in order to compare them
2. Reduces the number of cube dimensions after setting one of the dimensions to a specific value
3. Causes an increase in data aggregation and removes a detail level in a hierarchy
4. Reduces data aggregation and adds a detail level to a hierarchy
5. Changes the layout, in order to analyse a group of data from a different viewpoint

21) What is Data Ingestion?

- A process that copies data from sources to a repository, taking care of possible differences in speed between the generation and the storing process
- A process that copies data from sources to a repository, ensuring high data quality
- A process that copies data from sources to a repository, making the transformation required by the users
- A process that copies data from sources to a Data Warehouse guaranteeing the correctness of data with respect to the schema

22) Check the features of Data Lakes in comparison with Data Warehouses, Multiple answers allowed. Selecting a wrong option will be penalised.

- Can run on less expensive HW/SW
- Do not enforce data quality
- More complex management
- Completely defined use cases

23) Check the main objectives pursued when choosing a Data Lake architecture

- Store the data directly from the sources, as they are; the most appropriate structure will be decided later, according to the user needs
- The data must be accurately structured, in order to provide fast answers to complex queries
- The storage must be scalable and cheap, at the expenses of latency
- Ensure high quality of data through an accurate cleaning and transformation step

PARTE 2: Domande con risposte corrette (in grassetto)

1) In which part of the CRISP methodology we perform the test design activity?

- Data Understanding
- Evaluation
- Business Understanding
- **Modeling**

2) Talking about the general idea of database, what is the purpose of the "Schema on write" strategy?

Scegli una o più alternative:

- **Optimisation for specific types of queries**
- **Clean design of the data structure**
- Avoid preprocessing of data before writing
- Flexibility and efficiency for any kind of query

3) Talking about the general idea of database, what is the meaning of "Schema on write"?

Scegli un'alternativa:

- Create a schema for data after writing into the database
- Create a schema for data while writing into the database
- Change the schema of the data after each write
- **Create a schema for data before writing into the database**

4) Talking about the general idea of database, what is the purpose of the "Schema on read" strategy?

Scegli una o più alternative:

- **Avoid preprocessing of data before writing**
- Optimisation for various types of queries
- **Possibility to extract data in various shapes**
- **Flexibility for any kind of query**

5) Talking about the general idea of database, what is the meaning of "Schema on read"?

Scegli un'alternativa:

- **Write the raw data in the database, then at reading time shape then according to the reader's needs**
- **Create a schema for data after reading them from the database**
- Read the schema of the data before each data read
- Change the schema Of the data before each read

6) Which of the definition below describes the OLAP operation Roll-Up?

Scegli un'alternativa:

- Reduces data aggregation and adds a detail level to a hierarchy
- Changes the layout, in order to analyse a group of data from a different viewpoint
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value
- Creates a link between concepts in interrelated cubes, to compare them
- **Causes an increase in data aggregation and removes a detail level in a hierarchy**

7) Which of the definition below describes the OLAP operation Pivot?

Scegli un'alternativa:

- Causes an increase in data aggregation and removes a detail level in a hierarchy
- Reduces data aggregation and adds a detail level to a hierarchy
- **Changes the layout, in order to analyse a group of data from a different viewpoint**
- Creates a link between concepts in interrelated cubes, to compare them
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value

8) Which of the definition below describes the OLAP operation Drill-Down?

Scegli un'alternativa:

- **Reduces data aggregation and adds a detail level to a hierarchy**
- Creates a link between concepts in interrelated cubes, to compare them
- Changes the layout, in order to analyse a group of data from a different viewpoint
- Causes an increase in data aggregation and removes a detail level in a hierarchy
- Reduces the number of cube dimensions after setting one of the dimensions to a specific value

9) Talking about ETL, which of the following activities is related to the Transformation step?

Scegli una o più alternative:

- Usage of dictionaries to solve inconsistencies
- Snapshot of the operational data
- **Denormalisation**
- **Calculation of derived data**

10) Talking about ETL, which of the following activities is related to the Cleansing step?

Scegli una o più alternative:

- **Elimination of duplicates**
- Association of a timestamp to the operational data
- **Usage of dictionaries to solve inconsistencies**
- Snapshot of the operational data

11) Talking about ETL, which of the following activities is related to the Extraction step?

Scegli una o più alternative:

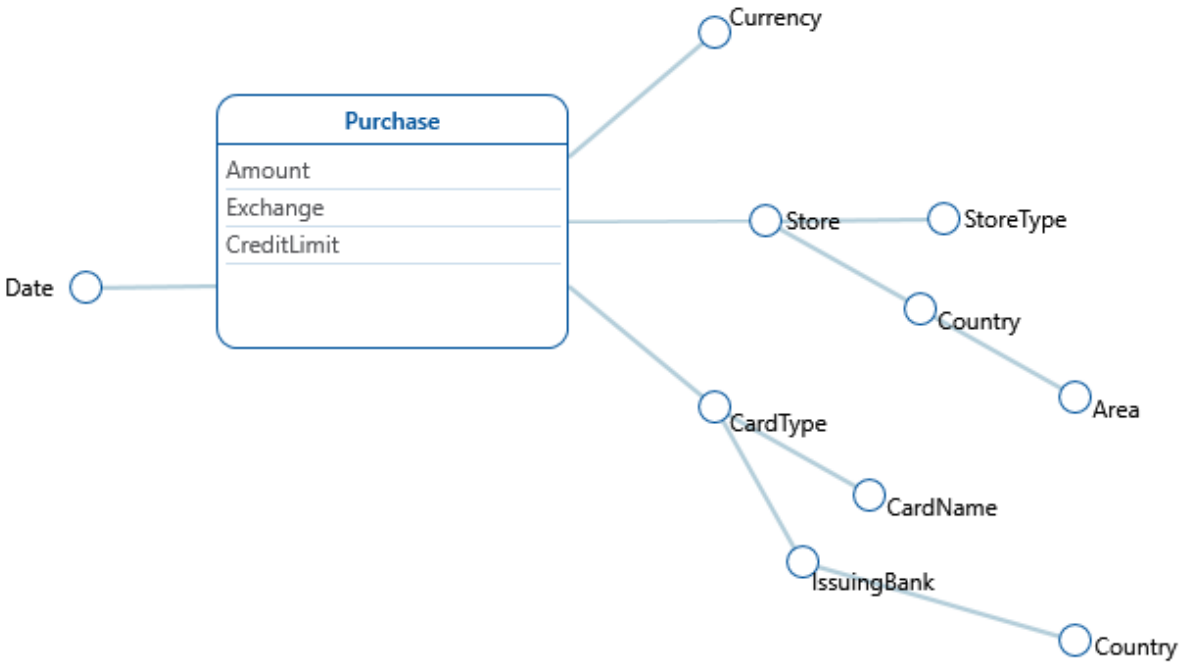
- Elimination Of duplicates
- **Snapshot of the operational data**
- **Association of a timestamp to the operational data**
- Usage Of dictionaries to solve inconsistencies

12) Order the ETL operations in order to obtain the correct sequence

1. Cleansing
2. Transformation
3. Extraction
4. Loading

Order: 3-1-2-4

13) Which of the texts describes the DFM schema below?



Scegli un'alternativa:

- **We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type; stores have also a country and an area. Purchases have a card type (e.g. credit or debit) and a card name (e.g. MyCard or JeansCard). Card types have also an issuing bank and a country.**
- We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type and country; stores have also an area. Purchases have a card name (e.g. MyCard or JeansCard)t card_names have a card type (e.g. credit or debit). Card types have also an issuing bank and a country.
- We need to track the purchases with credit cards. Each purchase has an amount, an exchange and a credit limit. The purchases are done in a date and in a store of a given type; stores types have also a country and an area. Purchases have a card type (e.g. credit or debit) and a card name (e.g. MyCard or JeansCard). Card names have also an issuing bank and a country.

14) Which of the following are typical data warehouse queries?

Scegli una o più alternative:

- **Which products maximize the profit?**
- Which is the stock of product "XXX" in the "YYY" warehouse
- What is the total revenue of yesterday in shop "B003" ?
- **What is the total revenue per product category and state?**
- **What is the relationship between profits gained by products "WWW" and "ZZZ"?**

15) Select the sentences describing a characteristic of a Data Warehouse

Scegli una o più alternative:

- **Solves the inconsistencies**
- **Non volatile**
- **Includes the time dimension**
- Constantly updated as soon as the base data are updated
- Includes all the base data in their native format

16) Associate the various Data Warehouse Architectures to the respective advantages

1. Single Layer: c. The occupation of space is minimised
2. Two Layers: a. The workloads for analysis and daily operations are separated
3. Three Layers: b. An intermediate level of consolidated data is available

17) Which of the activities below is part of "Business Understanding" in the CRISP methodology?

- Which data are available?
- Which machine learning functions are necessary for my problem?
- Which data must be collected with a specific campaign?
- **Which are the resources available (manpower. hardware, software,...)**

18) Rank the technologies for increasing level of abstraction: 1 = generate the most specific information, ... 3 = higher level knowledge

- SQL queries on operational databases → 1
- Data Mining on operational databases → 3
- OLAP analysis on Data Mart → 2

19) Which of the following sentences describes an advantage of a Data Warehouse with respect to a standard DBMS

Scegli una o più alternative:

- Allows efficient execution of key-based queries
- Manages efficiently data updates
- **Has tools for helping to solve inconsistencies**
- **Allows analysis along the time dimension**
- **Allows efficient execution of multi-dimensional queries**

20) Link the names of the OLAP operations below to their definitions

- Slice dice → 2. Reduces the number of cube dimensions after setting one of the dimensions to a specific value
- Drill across → 1. Creates a link between concepts in interrelated cubes, in order to compare them
- Pivot → 5. Changes the layout, in order to analyse a group of data from a different viewpoint
- Drill down → 4. Reduces data aggregation and adds a detail level to a hierarchy
- Roll-up → 3. Causes an increase in data aggregation and removes a detail level in a hierarchy

21) What is Data Ingestion?

- **A process that copies data from sources to a repository, taking care of possible differences in speed between the generation and the storing process**
- A process that copies data from sources to a repository, ensuring high data quality
- A process that copies data from sources to a repository, making the transformation required by the users
- A process that copies data from sources to a Data Warehouse guaranteeing the correctness of data with respect to the schema

22) Check the features of Data Lakes in comparison with Data Warehouses, Multiple answers allowed. Selecting a wrong option will be penalised.

- **Can run on less expensive HW/SW**
- **Do not enforce data quality**
- More complex management
- Completely defined use cases

23) Check the main objectives pursued when choosing a Data Lake architecture

- **Store the data directly from the sources, as they are; the most appropriate structure will be decided later, according to the user needs**
- The data must be accurately structured, in order to provide fast answers to complex queries
- **The storage must be scalable and cheap, at the expenses of latency**
- Ensure high quality of data through an accurate cleaning and transformation step