188.429 VU Business Intelligence WS2016

Test 2

1. **Types of analytics** (6 points)

   What is the difference between *descriptive, predictive,* and *prescriptive* analytics?

2. **Data Warehouse Architecture Components** (5 points)

   Which statements about components in the data warehouse reference architecture are correct?

   - ☐ The *landing area* is a database that stores a single data extract of a subset of one source database.
   - ☐ The *staging area* is a database that supports one or more types of business transactions.
   - ☐ The *master data store factors* out information that establishes the context of data collected in business transactions.
   - ☐ The *metadata component* factors out of information that establishes the context of data collected in business transactions.

3. **Data Warehouse** (5 points)

   Data Warehouse...

   - ☐ typically uses a schema that is still more or less in 3NF.
   - ☐ stores data by operational applications rather than by business subjects.
   - ☐ streams continously updated data.
   - ☐ is a copy of transaction data specifically structured for query and analysis.

4. **OLTP VS OLAP** (5 points)

   Which of the statements about OLTP and OLAP are correct?

   - ☐ OLTP aims to turn raw data into strategic information.
   - ☐ OLAP optimizes for many short and "small" transactions.
   - ☐ OLTP systems store up-to-date data.
   - ☐ OLTP systems tend to use normalized.

5. **Fact or Dimension** (5 points)

In which star schema table would you most likely expect to find the following fields?

Order Line Quantity	◯ Fact	◯ Dimension
Street	◯ Fact	◯ Dimension
Revenue	◯ Fact	◯ Dimension
Sales amount	◯ Fact	◯ Dimension
Day of week	◯ Fact	◯ Dimension

6. **OLAP Operations: cube operator** (10 points)

What does the cube Operator do and what is it useful for?

Illustrate how it can be used by means of a minimal example (simple SQL query using the operator and example query result).

7. **Category vs. measure attributes** (5 points)

Which of the following statements about category and measure attributes are correct?

- ☐ A measure attribute is an independent attribute that primarily serves to group or segment business data.
- ☐ Measure attributes are used to express quantitative properties of objects.
- ☐ Measure attributes assign objects to one of a relatively small number of discrete categories, based on the value of an attribute.
- ☐ A measure attribute is an independent attribute that primarily serves to group or segment businnes data.

8. **Vertical Partitioning** (5 points)

Which of the following statements about vertical partitioning are correct?

- ☐ Vertical partitioning is particularly efficient when only a few attributes of a table are accessed.
- ☐ Vertical partitioning is based on the idea of splitting a table into disjoint parts with the same schema.
- ☐ When executing a filter operation, vertical partitioning allows to skip certain partitions based on query predicates.

☐ Horizontal partitioning is based on the idea of dividing a table into multiple tables that contain fewer columns.

9. **Time Dimension** (5 points)

Which of the following statements about time dimensions in Data Warehouse applications are correct?

☐ The historization of dimension tables in relative straightforward, but the historization of fact tables poses a significant conceptual challenge.

☐ Rows in dimension tables are tipically directly associated with time.

☐ Fact tables are tipically associated with a specific time by the foreign key reference to the time dimension.

☐ Historization tracks changes in attribute values, relations and entities accoss time in order to facilitate analysis.

10. **ETL** (5 points)

Which of the following statements about ETL are correct?

☐ Data auditing aims at judging the quality of data.

☐ The loading component transfers the conformed data from the staging area into the DWH.

☐ The extraction component transfers data from the source systems into the staging area.

☐ Both file and DBMS technologies may be used in the implementation of an ETL process.

11. **Agile BI** (5 points)

Which of the following statements about Agile BI are correct?

☐ Agile BI advocates a rigid, highly architectural top-down design approach.

☐ Agile BI aims to address evolving user requirements and business needs.

☐ An agile development approach in BI is necessarily associated with lower costs than a waterfall-style development approach.

☐ Agile development approaches aim for short time-to-value.

12. **Data Warehouse Development** (8 points)

Describe the Inmon (top-down) approach for data warehouse development and list its advantages and disadvantages.

13. **Hadoop** (5 points)

Which of the following statements about Hadop are correct?

- ☐ Hadoop provides fault-tolerance.
- ☐ Hadoop is a NoSQL database.
- ☐ Hadoop provides scalable large-scale data processing.
- ☐ Hadoop is ideally suited for low-latency applications.

14. **Hadoop use cases** (6 points)

List three typical use cases for Hadoop.

15. **HDFS** (5 points)

- ☐ Files stored in HDFS can be changed.
- ☐ HDFS is optimized for large, sreaming
- ☐ HDFS sits on top of the native filesystem.
- ☐ HDFS performs best with a very large number of small files.

16. **MapReduce** (5 points)

- ☐ is a distributed programming model for parallel data processing.
- ☐ MapReduce is a distributed file system.
- ☐ MapReduce is a streaming data processing tool.
- ☐ In the typical map reduce job, the number of mappers is much larger than the number of reducers.

17. **HBase** (5 points)

- ☐ falicitates parallelized queries over massive data sets.
- ☐ is a distributed relational database.
- ☐ is a graph database.
- ☐ is based on a column-family oriented data model.

18. **Hive** (5 points)

- ☐ imposes a relational structure on unstructured big data.
- ☐ provides low-latency and supports real-time queries.
- ☐ Schema and data are integrated and stored in the same location.
- ☐ supports typed columns.