1

Introduction

Review Questions

1.1 Why are traditional databases inappropriate for data analysis?
1.2 Discuss four main characteristics of data warehouses.
1.3 Describe the different components of a multidimensional model, i.e., facts, measures, dimensions, and hierarchies.
1.4 What is the purpose of online analytical processing (OLAP) systems and how are they related to data warehouses?
1.5 What is spatial data? Give an example of an application that manipulates spatial data.
1.6 Explain the differences between spatial databases and spatial data warehouses.
1.7 What is temporal data? Describe different types of temporal support needed by applications.
1.8 Do current data warehouses manipulate temporal data?
1.9 Describe the different levels of schemas that are used for designing databases.
1.10 Explain the advantages of using a conceptual model when designing a data warehouse.
1.11 Specify the different steps used for designing a database.
1.12 Why do we need a method for data warehouse design?
Introduction to Databases and Data Warehouses

Review Questions

2.1 What is a database? What is a database management system?
2.2 Describe the four phases used when designing a database.
2.3 Define the following terms: entity type, entity, relationship type, relationship, role, cardinality, population.
2.4 Illustrate with one example each of the following kinds of relationship types: binary, n-ary, one-to-one, one-to-many, many-to-many, and recursive.
2.5 Discuss different kinds of attributes according to their cardinality and their composition. What are derived attributes?
2.6 What is an identifier? What is the difference between a strong and a weak entity type? Does a weak entity type always have an identifying relationship? What is an owner entity type?
2.7 Discuss the different characteristics of the generalization relationship.
2.8 Define the following terms: relation (or table), attribute (or column), tuple (or line), domain.
2.9 Explain the various integrity constraints that can be described in the relational model.
2.10 Discuss the basic rules for translating an ER schema into a relational schema. Give an example of a concept of the ER model that can be translated into the relational model in different ways.
2.11 Illustrate with examples different types of redundancy that may occur in a relation. Why redundancy in a relation can induce problems in the presence of insertions, updates, and deletions?
2.12 What is the purpose of functional and multivalued dependencies? What is the difference between them?
2.13 What are normal forms? Specify several normal forms that can be defined on relations. For each one of these normal forms, give an example of a relation that does not satisfy the particular normal form.
2.14 Why was the object-relational model proposed? How does this model differ from the traditional relational model?

2.15 Describe the composite types and the user-defined types supported by SQL:2003.

2.16 What is the difference between final and not final types? What is the difference between instantiable and noninstantiable types?

2.17 Explain the differences between a table and a type in SQL:2003. Are there several types of tables in SQL:2003?

2.18 Discuss the basic rules for translating an ER schema into an object-relational schema.

2.19 What is the objective of physical database design? Explain some factors that can be used to measure the performance of database applications and the trade-offs that have to be resolved.

2.20 Explain different types of file organization. Discuss their respective advantages and disadvantages.

2.21 What is an index? Why are indexes needed? Explain various types of indexes.

2.22 What is clustering? What is it used for?

2.23 What is an operational database system? What is a data warehouse system? Explain several aspects that differentiate these systems.

2.24 What is the meaning of the acronyms OLAP and OLTP?

2.25 Using an example of an application domain that you are familiar with, describe the various components of a multidimensional model, i.e., facts, measures, dimensions, and hierarchies.

2.26 Why are hierarchies important in data warehouses? Illustrate with examples various kinds of hierarchies.

2.27 Discuss the role of measure aggregation in a data warehouse. How can measures be characterized?

2.28 Describe the various OLAP operations using the example you defined in Question 2.25.

2.29 Describe the differences between the following concepts:

   (a) Relational OLAP (ROLAP), multidimensional OLAP (MOLAP), hybrid OLAP (HOLAP).

   (b) Star schema, snowflake schema, starflake schema, constellation schema.

2.30 What is the objective of physical data warehouse design? Specify different techniques that are used to achieve such objective.

2.31 Discuss advantages and disadvantages of using materialized views.

2.32 What are bitmap and join indexes? Why are they important?

2.33 Describe the different components of a typical data warehouse architecture. Identify variants of this architecture and specify in what situations they are used.

2.34 Briefly describe the multidimensional model implemented in Analysis Services and in Oracle OLAP.
Conventional Data Warehouses

Review Questions

3.1 Discuss the following concepts: dimension, level, attribute, key, fact relationship, role, measure, hierarchy, parent-child relationship, cardinalities, root level, leaf level, and analysis criterion.

3.2 Explain the difference, at the schema and at the instance level, between balanced and unbalanced hierarchies.

3.3 Give an example of a recursive hierarchy. Explain how to represent an unbalanced hierarchy with a recursive one.

3.4 Explain the usefulness of generalized hierarchies. To which concept of the entity-relationship model these hierarchies are related?

3.5 What is a splitting level? What is a joining level? Does a generalized hierarchy always have a joining level?

3.6 Explain why noncovering (or ragged) hierarchies are a particular case of generalized hierarchies.

3.7 Describe using an example how the type of a generalization in the ER model (total vs. partial, exclusive vs. overlapping) influences the corresponding generalized hierarchy in the multidimensional model.

3.8 What is the difference between strict and nonstrict hierarchies?

3.9 Illustrate with one example the problem of double counting of measures for non-strict hierarchies. Describe different solutions to this problem.

3.10 What is a distributing factor? Explain the importance of choosing an appropriate distributing factor.

3.11 Explain in what situations alternative hierarchies are used.

3.12 Describe the difference between parallel dependent and parallel independent hierarchies.

3.13 Illustrate with examples the difference between generalized, alternative, and parallel hierarchies.

3.14 What are role-playing dimensions?

3.15 Why some dimensions are called fact or degenerate?
3.16 Relate the problem of double counting to the functional and multivalued dependencies that hold in a fact relationship.

3.17 Why a fact relationship must be decomposed in the presence of dependencies? Show an example of a fact relationship that can be decomposed differently according to the dependencies that hold on it.

3.18 Discuss the mapping rules for translating a MultiDim schema into a relational schema. Are these rules similar to those used for translating an ER schema into a relational schema?

3.19 Explain how a balanced hierarchy can be mapped into either normalized or denormalized tables. Discuss the advantages and disadvantages of these alternative mappings.

3.20 How to transform at the logical level an unbalanced hierarchy into a balanced one?

3.21 Describe different approaches for representing generalized hierarchies at the logical level.

3.22 Explain how a nonstrict hierarchy can be represented in the relational model and in the object-relational model.

3.23 Is it possible to distinguish between generalized, alternative, and parallel dependent hierarchies at the logical level?

3.24 Identify the kind of hierarchies that can be directly represented in Analysis Services and in Oracle OLAP.
Review Questions

4.1 What are spatial databases? Describe two complementary ways of modeling spatial data in database applications.
4.2 Describe the various spatial data types, giving for each one of them an example of its use.
4.3 Define the various topological relationships in terms of the boundary, interior, and exterior of spatial values.
4.4 What is the difference between the raster and vector data models for representing spatial data?
4.5 What is the difference between the spaghetti, network, and topological models for storing collections of spatial objects?
4.6 Discuss the following concepts: spatial dimension, spatial level, spatial attribute, spatial fact relationship, spatial measure, spatial hierarchy, and topological relationship.
4.7 What are the differences between a spatial level, a spatial level with spatial attributes, and a conventional level with spatial attributes?
4.8 Give an example of each of the following hierarchies: balanced, unbalanced, and generalized spatial hierarchies.
4.9 Illustrate with an example how a distributing factor allows one to solve the double counting problem in spatial nonstrict hierarchies.
4.10 What is the difference between alternative and parallel spatial hierarchies?
4.11 Classify the topological relationships according to the procedures required for measure aggregation. Give an example of a situation that requires a specific aggregation procedure.
4.12 Why n-ary topological relationships are needed in spatial fact relationships? Does such relationships are usual in spatial databases?
4.13 How does a spatial measure differ from a conventional measure computed with spatial operators?
4.14 Does a spatial measure requires to be related to spatial dimensions?
4.15 Give an example of a multidimensional schema containing a spatial measure. Transform the spatial measure into a spatial dimension. Compare the two schemas with respect to the various queries that can be addressed to them.

4.16 Briefly describe the spatial model implemented in Oracle Spatial.

4.17 How are a spatial level, a spatial level with a spatial attribute, and a conventional level with a spatial attribute represented in the object-relational model?

4.18 How is a topological relationship between spatial levels represented in a logical schema?

4.19 How can one check in a logical schema the topological relationship of a fact relationship?

4.20 How is a spatial measure represented in the object-relational model?
Temporal Data Warehouses

Review Questions

5.1 Describe with an example the three implementation types of slowly changing dimensions. Discuss their potential problems.

5.2 What are temporal databases? Discuss different ways of interpreting the time frame associated with the facts contained in a temporal database.

5.3 Describe the various temporal data types, giving for each one of them an example of its use.

5.4 Define the various synchronization relationships in terms of the boundary, interior, and exterior of temporal values.

5.5 Does current DBMSs, and SQL in particular, provide support for dealing with time-varying data?

5.6 Explain the various temporality types supported by the MultiDim model.

5.7 Discuss the following concepts: temporal dimension, temporal level, temporal attribute, temporal fact relationship, temporal measure, temporal hierarchy, temporal parent-child relationship, instant cardinality, lifespan cardinality, and synchronization relationship.

5.8 Does a temporal MultiDim schema include a time dimension?

5.9 What are the differences between a temporal level, a temporal level with temporal attributes, and a conventional level with temporal attributes?

5.10 Given a parent-child relationship between two levels, analyze the various situations that result when the level, the relationship, or both the level and the relationship are temporal. Identify potential problems that may arise in each situation.

5.11 In the case of temporal relationships between temporal levels, how the lifespan of a relationship instance must be constrained in order to ensure correctness of roll-up and drill-down operations?

5.12 Are both instant and lifespan cardinalities allowed for non-temporal parent-child relationships? Is it possible to have both types of cardinalities when the related levels are nontemporal?
5.13 State the constraints implied by the instant cardinalities in the case of temporal relationships between temporal levels. What can be implied by combining these constraints with the constraint expressed in Question 5.11?

5.14 Explain the role of synchronization relationships in temporal fact relationships.

5.15 Illustrate with examples the usefulness of including in a multidimensional schema various combinations of valid time, loading time, and transaction time. Analyze this according to the temporal support provided by source systems.

5.16 How is the issue of measure aggregation in the presence of temporal relationships related to the problem of managing different temporal granularities? Explain various solutions for these problems.

5.17 Why is it useful to first translate a temporal MultiDim schema into the ER model prior to its translation at the logical level?

5.18 How are the various temporality types represented in the ER and object-relational models?

5.19 Illustrate with examples how to map into the object-relational model a temporal attribute, a temporal level with a temporal attribute, and a conventional level with a temporal attribute.

5.20 Discuss different options that can be used for mapping a temporal relationship into the object-relational model.

5.21 How are measures mapped into the object-relational model? How does the mapping differ when measures represent states or events?

5.22 What are the advantages of implementing a temporal multidimensional schema using the object-relational model instead of the relational model?

5.23 How can the various temporal constraints included in a conceptual multidimensional schema be represented in a logical schema?

5.24 Illustrate with an example different approaches to measure aggregation in the presence of temporal relationships.
6

Designing Conventional Data Warehouses

Review Questions

6.1 What are the similarities and the differences between designing a database and designing a data warehouse?

6.2 Compare the top-down and the bottom-up approaches for data warehouse design. Which of the two approaches is more often used? How does the design of a data warehouse differ from the design of a data mart?

6.3 Discuss the various phases in data warehouse design, emphasizing the objective of each phase.

6.4 Summarize the main characteristics of the analysis-driven, source-driven, and analysis/source-driven approaches for requirements specification. How do they differ from each other? What are their respective advantages and disadvantages? Identify in which situations one approach would be preferred over the others.

6.5 Using an application domain that you are familiar with, illustrate the various steps in the analysis-driven approach for requirements specification. Identify at least two different users, each one with a particular business goal.

6.6 Using the application domain of Question 6.5, illustrate the various steps in the source-driven approach for requirements specification. Define an excerpt of an ER schema from which can be derived some multidimensional elements.

6.7 Compare the steps for conceptual design in the analysis-driven, source-driven, and analysis/source-driven approaches.

6.8 Develop a conceptual multidimensional schema for the application domain of Question 6.5 using among the three approaches the one that you know best.

6.9 Illustrate the different aspects of the logical design phase by translating the conceptual schema developed in Question 6.8 into the object-relational model.
6.10 Describe several aspects that are important to consider in the physical design phase of data warehouses.

6.11 Describe how the ETL process is taken into account in the different phases of data warehouse design.
Designing Spatial and Temporal Data Warehouses

Review Questions

7.1 Describe the current approaches to the design of spatial and temporal databases.

7.2 How do the analysis-driven, source-driven, and analysis/source-driven approaches for designing data warehouses must be modified to take into account the spatial and temporal aspects?

7.3 Compare the early and the late inclusion of spatial and temporal support when designing data warehouses. What are the factors that determine the choice between the two options?

7.4 In the analysis-driven approach, describe the steps for requirements specification and conceptual design of spatial and temporal data warehouses.

7.5 In the source-driven approach, describe the steps for requirements specification and conceptual design of spatial and temporal data warehouses.

7.6 Justify the usefulness of the analysis/source-driven approach for designing spatial and temporal data warehouses.

7.7 Using an application domain that you are familiar with, develop a spatial multidimensional schema following the analysis-driven approach. Choose between early or late inclusion of spatial support.

7.8 Illustrate how the spatial aspects are taken into account in the logical design phase by translating the conceptual schema developed in Question 7.7 into the object-relational model.

7.9 Using an application domain that you are familiar with, develop a temporal multidimensional schema following the source-driven approach. Choose between early or late inclusion of temporal support.

7.10 Illustrate how the temporal aspects are taken into account in the logical design phase by translating the conceptual schema developed in Question 7.9 into the object-relational model.
Conclusions and Future Work

Review Questions

8.1 Provide your own assessment of the advantages and disadvantages of using the MultiDim model for designing conventional, spatial, and temporal data warehouses.

8.2 Why the operations for querying and aggregating data should be considered in the context of a conceptual multidimensional model?

8.3 List several constructs of the ER model that are not considered in the MultiDim model. Why would these constructs be useful in a conceptual multidimensional model?

8.4 Discuss an actual application domain that deals with three-dimensional (3D) spatial data. Do you think that spatial data warehouses would benefit from the support of 3D data?

8.5 Give examples of spatial data for which a continuous view is more appropriate than a discrete view. Would the combination of these views enhance the analysis possibilities in spatial data warehouses?

8.6 What is meant by multiple representations of spatial data? Identify issues that may arise when dealing with multiple representations in spatial data warehouses.

8.7 Describe the problem of dealing with multiple granularities in temporal data warehouses. What kind of support for this is provided by current data warehouse systems?

8.8 What are the problems raised by measure aggregation in temporal data warehouses?

8.9 What is schema versioning? How is this related to temporal data warehouses?

8.10 Give examples of actual applications that manipulate spatio-temporal data. Why a data warehouse should support such data?

8.11 Why is it important to assess a data warehouse design method? How could such an assessment be realized? How to include spatio-temporal aspects in the assessment?