

Paper Reading

Database

Read an overview of the relational model e.g. from <http://www.utexas.edu/its-archive/windows/database/datamodeling/rm/overview.html> which is much easier to read than Codd's paper. You probably want to read up to the normalization part. As you read Codd's paper, observe how many of the ideas had their roots in the paper and were further developed and refined.

A Relational Model of Data for Large Shared Data Banks

by E. F. Codd

The paper introduces two problems, *data independence* and *data inconsistency*, and proposed the relational model to handle these issues.

1. What does data independence mean?
2. What does data inconsistency mean?

At the time the paper was written, the dominant data models are the tree structured models.

3. How do the ordering, indexing and access path dependences affect the independence of application programs from growth and changes in data representations?
4. Do the competing models handle data independence well? Does the relational model handle data independence well?

The relational model consists of a set of tables

5. What is the definition of the term relation?
6. What is the definition of the term domain?
7. What are the properties of a relation?
8. What are the advantages of using unordered rather than ordered relations?
9. What is a primary key?
10. What is a foreign key?

Given a tree structured model, it is possible to convert it into a relational model.

11. What is the (first) normal form?
12. How do you normalize a hierarchical model into the normal form?

Languages based on first order predicate calculus can be used to provide logical inference on databases. Codd describes a more restricted set of operators that became known as relational algebra or relational calculus. Given that this is a more restricted set of operators, what are natural questions you may want to ask about it?

13. What do the operations on relations, permutation, projection, join, composition and restriction do?
14. What is the natural join?
15. What is a point of ambiguity?
16. What condition would ensure that there is no point of ambiguity? Why is this important?

The relationship between redundancy and consistency was further developed by Codd and other into the theory of normal forms. See e.g. <http://www.utexas.edu/its-archive/windows/database/datamodeling/rm/overview.html>. You may find it interesting to compare later development with Codd's suggestions in this paper.

17. What is strong redundancy?
18. How can the system utilize redundancy?
19. What strategies does Codd suggest for checking consistency?

Why did the paper (and follow-on ideas) have impact? What are the fundamental advantages of the relational model that motivated Codd to formulate the database problem this way? Are the advantages obvious?