week 2 reading relational databases

wbg231

January 2023

1 Introduction

-

- Hector Garcia-Molina, Jeffrey D Ullman, and Jennifer Widom. Database systems: the complete book - Chapter 2. Pearson Education, 2009.
- i am going to keep this kind of high level

overview of data models

- a Database model has three parts
  - structure of the data that is conceptually how is the data organized
  - operations on the data ie a limited set of things that can be done to the data (searching or modification) the limitations of this data allow the system to be easy to work with but still fast
  - constraints on the data, we only allow the data to be format in a certain way or take on certain values
- the two most important data modes today are the relational and semi-structured model

the semi structured model

- semi structured data resembles trees or graphs rater than tables
- the most common semi structured model is XML
- the data structure are tags that describe the data.
- operations in semi structured data usually involve following paths down a tree or graph

- constraints in this model are about the types of data that can be associated with a certain tag
- this model has a lot more freedom than the relational model
- however the speed of the relational model and the easy of SQL make it preferable in most cases

the relational mode

- the relational model lest us represent data as two dimensional tables called relations where each row represents and individual and each col represents a dimension
- the columns of a table are named by attributes ie the col names
- the schema is called a has the name of the relation and the set of attributes of the form table(A_1 ... A_d)
- attributes are a set not a list (that is there are no repeated and order does not matter)
- Database in the relational Database model a database is one more relations
- the schema of a database is called a database schema
- each row of a database is called a tuple the elements in a tuple must be ordered
- each feature needs to have an associated fixed data type such as STR
- within tuple order does not matter
- a set of tuples for any schema is called an instance these instances can be updated as new data is added
- a relations key is a set of attributes for which no two tuples (ie rows) can have the same value for all attributes in key

defining a schema in sql

- sql is used for database manipulation
- sql has two attributes
    1. the data definition sub language for declaring schema
    2. the data manipulation sub language for querying and modification of a database

- sql has three types of relations
    1. sorted relations which are called tables, these are what we normally deal with
    2. views which are relations defined by computation and are not stored after seen
    3. temporary tables which are constructed by SQL as intermediate steps in computation but not stored.
- the types within sql are char, string, int, date, float, bool
- tables are declared like this CREATE TABLE name(title, char(100), year int, length int)