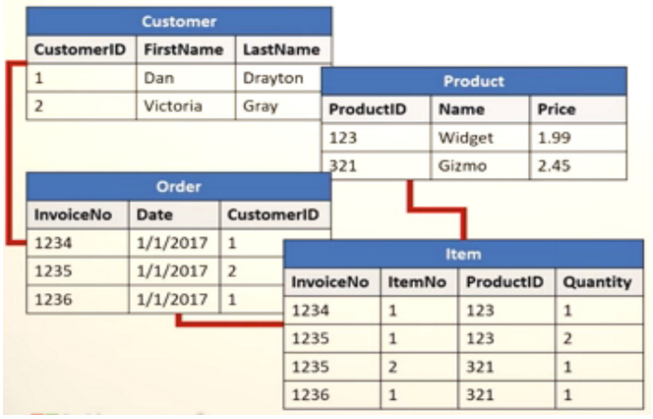
DAT229x Introduction to Big Data

Wednesday, 6 June 2018 10:49 AM

Lab 2. Introduction to Relational Databases

Lab 2.1 Introduction to Databases

- Database is a special system for storing data, and there are sorts or different types of databases.
- Relational database is a digital database based on the relational model of data, as proposed by E. F. Codd in 1970.[1] Virtually all relational database systems use SQL (Structured Query Language) for querying and maintaining the database.[2]
- Example. All of tables which we call relations in a relational database, have discrete information about individual elements of business entities. Linking these tables together by using common e-field. CustomerID from Table1 is related to the customer that's mentioned in my order Table2. Within my order to my item Table3, my items refer to the invoice number, and within those items, i am referencing the products that's been actually ordered by customer. By having those primary keys we are able to put together all of those different tables.



Customer

CustomerID	FirstName	LastName
1	Dan	Drayton
2	Victoria	Gray

Product

ProductID	Name	Price
123	Widget	1.99
321	Gizmo	2.45

Order

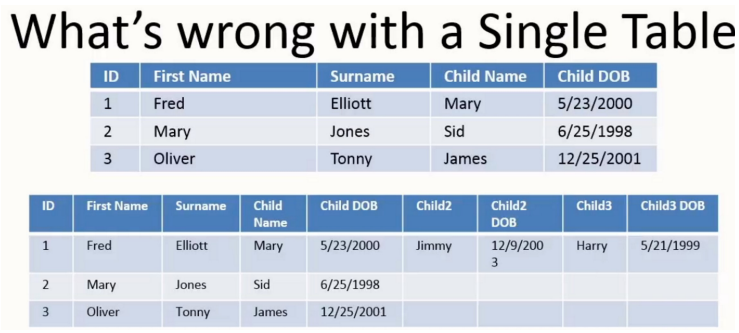
InvoiceNo	Date	CustomerID
1234	1/1/2017	1
1235	1/1/2017	2
1236	1/1/2017	1

Item

InvoiceNo	ItemNo	ProductID	Quantity
1234	1	123	1
1235	1	123	2
1235	2	321	1
1236	1	321	1

Lab2.2 Relational Databases

- Tables - contain data
- Field/Column - broken into fields or columns; each column should be a single type of data (i.e., first-name or last-name)
- Row - each row is a record of information
- Data type - important sequel so we know how to write (i.e., dates of birth should go into dates fields; first-name or surnames can go into text fields; Ids need to be number fields)
- Primary key - unique reference anyone record in a given table (i.e., numbers friends 1/2/3; parents number 1/2/3); numeric or alphanumeric reference
- Foreign key - link that a primary key links to (i.e., Parents Table. primary key is the parent number; foreign key is the parent ID in the Children Table because it links to the primary key in the parents table)
- Joins - all table joins in some way
- SQL language used to manipulate our databases accessing the information (i.e., SELECT Product ID, Name FROM Product Where Price > 2.00;)

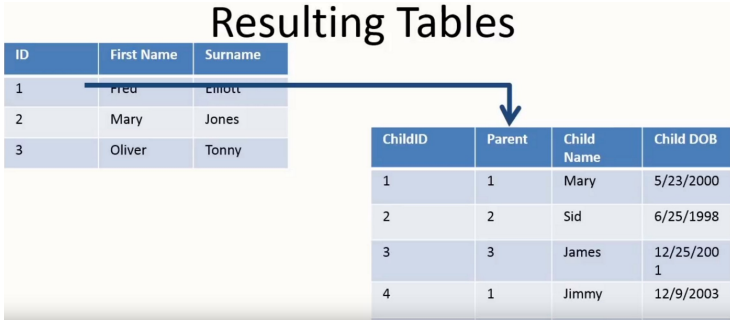


What's wrong with a Single Table

ID	First Name	Surname	Child Name	Child DOB
1	Fred	Elliott	Mary	5/23/2000
2	Mary	Jones	Sid	6/25/1998
3	Oliver	Tonny	James	12/25/2001

ID	First Name	Surname	Child Name	Child DOB	Child2	Child2 DOB	Child3	Child3 DOB
1	Fred	Elliott	Mary	5/23/2000	Jimmy	12/9/2003	Harry	5/21/1999
2	Mary	Jones	Sid	6/25/1998				
3	Oliver	Tonny	James	12/25/2001				

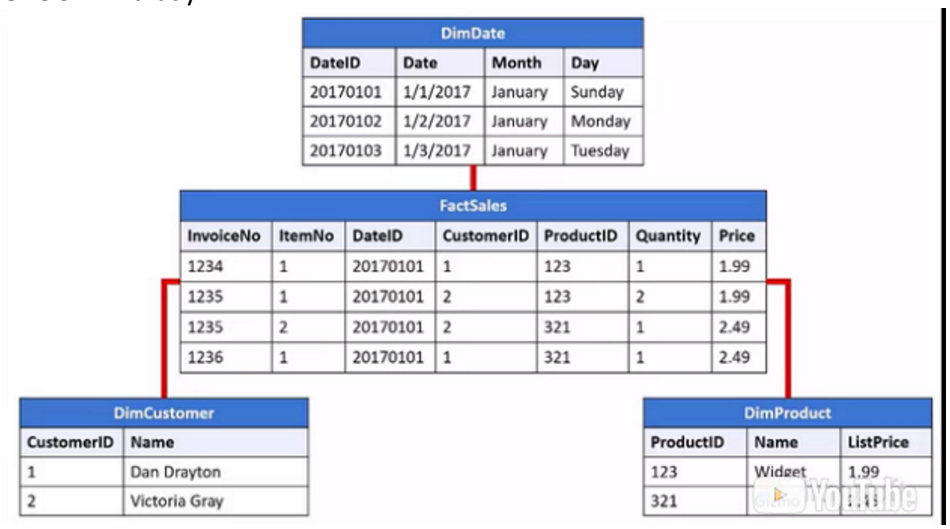
Larger and larger; blanks and empty.

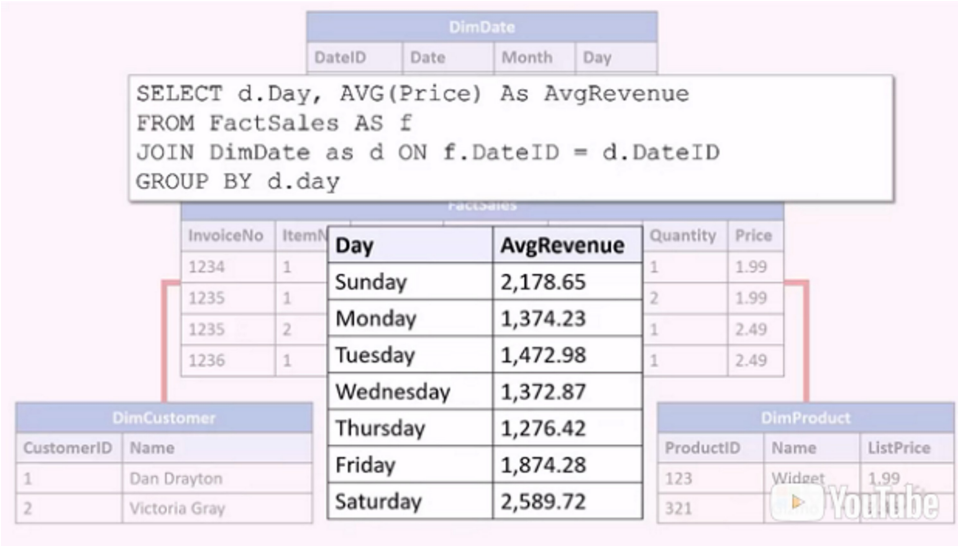


Broking into parents table and children table.
ID in table1 is the primary key and Parent in table2
Is the foreign key that links to table1.

Lab2.4 Data Warehouses

- A special form of relational database
- Fact table (i.e., my sales) and dimension tables that are related to fact table that have all things that I would aggregate the data by (i.e., aggregate sales by date or by customer or by product)
- Denormalized relational database (i.e., allowing duplication; merged name together for customer into one field; allowing the idea that my date is stored in a separate fields)
- Rearrange the structure of database, to optimal design for the type of query that i want to do when I am working with historical data and get some reporting from it.
- SELECT d.Day, AVG (Price) As AvgRevenue
  From FactSales AS f
  JOIN DimData as d ON f.DateID = d.DateID
  GROUP BY d.day





Asking the day and price from FactSales table, joining the DimDate table, based on common key by date. Grouping and aggregating the data so I get the average revenue per day of the week.

Further Reading
https://mva.microsoft.com/en-US/training-courses/sql-database-fundamentals-16944?l=w7qq6nAID_6805121157

https://docs.microsoft.com/en-us/azure/sql-database/