

Lecture #3 –
Introduction to SQL Server &
Database Design

GEOG/ECOL/ENR 5050

Fall 2016

Instructor: Dr. Shannon E. Albeke

What is a Database?

- noun: a comprehensive collection of related data organized for convenient access, generally in a computer.
(<http://dictionary.reference.com/>)
- It is a container with associated software and protocols to help organize, manage, manipulate, share and visualize related data

Available Database Platforms

- For this class we are going to use MS SQL Server Express (<http://www.microsoft.com/en-us/download/details.aspx?id=52679> ; ENU\x64\SQLEXPADV_x64_ENU.exe)
 - Free, up to 10 GB per database, supports spatial data types.
 - Unlimited size for the Licensed Version
- Other options
 - MS Access (not Enterprise scale, just desktop)
 - Oracle (proprietary)
 - MySQL (open source, multiple operating systems)
 - PostgreSQL with PostGIS and pgAdmin III add-ons (open source, multiple operating systems)**

SQL Management Studio

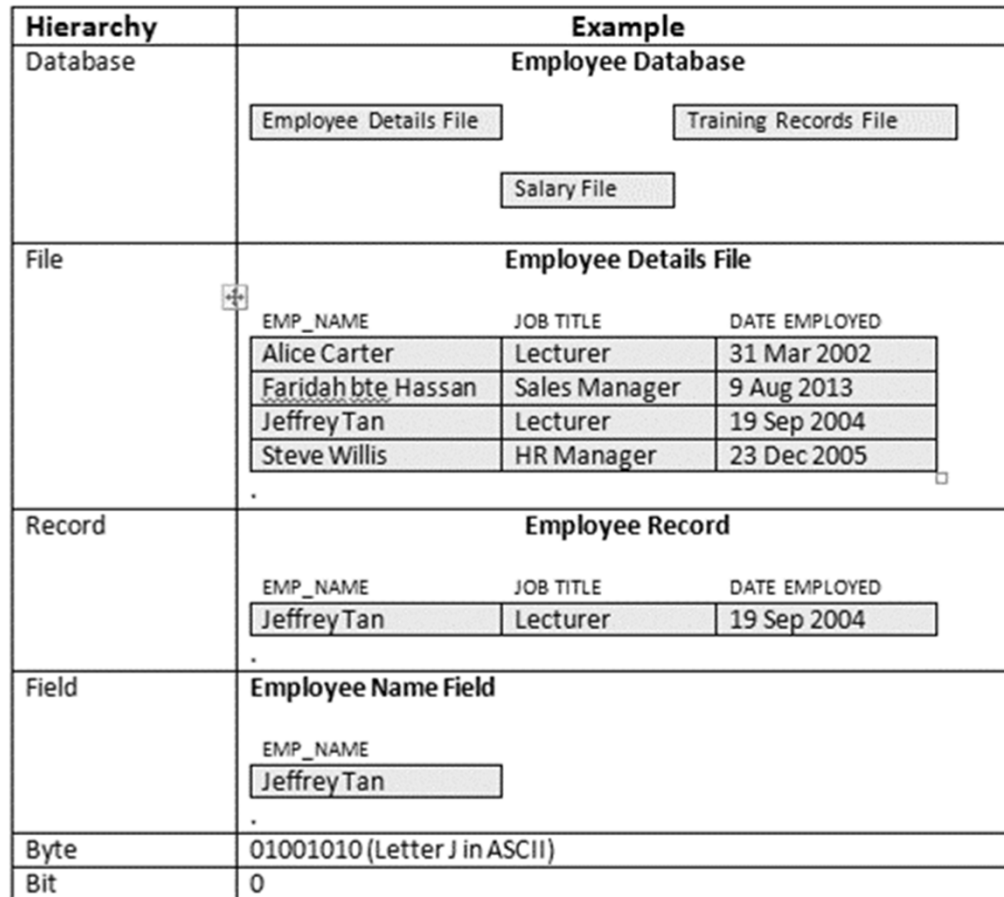
- It is an Integrated Development Environment (IDE) to assist in creating and managing SQL Server Databases
- Allows you to connect to an **Instance** of SQL Server
 - E.g. ECOINFO,6004\ECODEV
(server_name\instance_name)
 - You can have more than one Instance on a computer/server

Within An SQL Instance

- You can have one or many **Databases** within an Instance
- A Database is a container for:
 - Tables (the primary component of a database)
 - Store your hierarchical, structured data here
 - Tables consist of Rows (data records) and Columns
 - A Row represents a single entity
 - Each Column represents an attribute describing the entity
 - » Each column can be one data-type only (character, numeric, etc.)
 - Best to define a Primary Key for each table
 - PK guarantees each row of data are unique
 - PK can consist of one or more columns

Within An SQL Instance

- You can ... an Instance
- A Data ...
 - Tables
 - Sto
 - Tab
 -
 -
- Bes



Note: EMP = employee

Source: Jeffrey TL Tan Wikipedia original contributor for Data Hierarchy, 9 Aug 2013
 Permission is given to freely use this diagram in its entirety & unedited.

- PK guarantees each row of data are unique
- PK can consist of one or more columns

_within

abase)

e

columns

the entity

aracter,

Table Relationship Example

Table Design

StateCode (PK)	Text (2)
StateName	Text (50)

Table Data

StateCode	StateName
CO	Colorado
WY	Wyoming

Table Relationship Example

Table Design

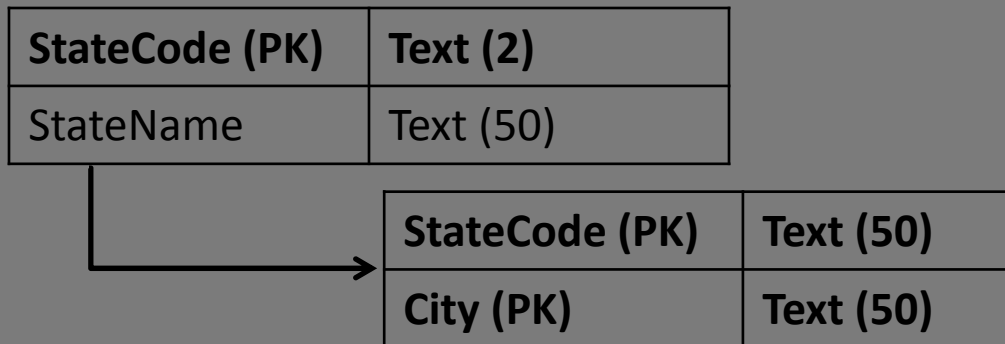
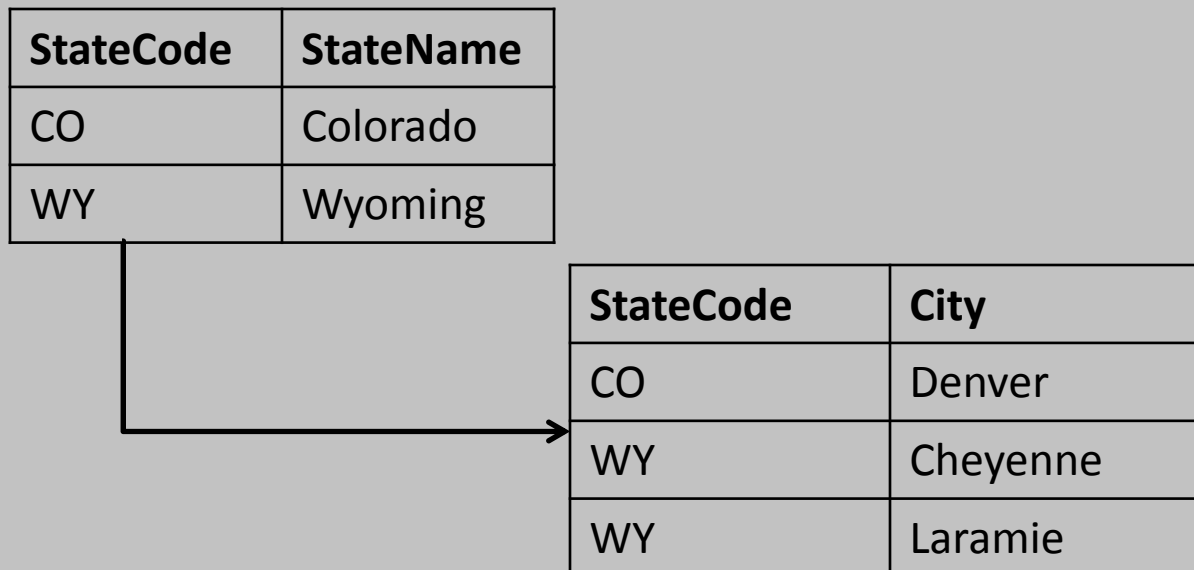


Table Data



Within An SQL Instance

- You can have one or many **Databases** within an Instance
- A Database is a container for:
 - Tables (the primary component of a database)
 - Views (also known as stored queries)
 - Views allow you to Select data from one or more related tables. Can be All data or a Subset.
 - Primary tool for asking questions of our data
 - Views can be used just like a Table, but you can't design the Data Types directly, they are obtained from the Parent Tables

Within An SQL Instance, cont'd

- A Database is a container for:
 - Relationships
 - Set rules for how two tables are related to each other
 - Stored Procedures
 - Saved set of Functions that can be called for 'action' upon your data or a given user input
 - Security
 - Methods for specifying which users have access (read vs. write/edit) to your data

Database Design Considerations

Space and Time

- For Natural Resource Data, more often than not, you collect information at a location in space multiple times
 - Sample site
 - Take measurements one or more times
 - This is a natural hierarchy and lends itself to a simple database format/structure/schema/design

Basic Design Considerations

- THINK about how you might store your data
 - Especially if you haven't collected it yet
 - This may become a part of your legacy, think big!
 - Requires some imagination
 - Study design = Database design?
 - Include Spatial information?
 - Points, lines, polygons or raster
 - Must maintain metadata!
- Data types (again!)
 - Text, Long Integer, Double/Float, Date/Time...etc.

Basic Design Considerations

- Normalization – set of methods to break down tables to their constituent parts (i.e. each table represents 1 ‘thing’ only)
- Primary Key – a column containing a value that makes the row of data unique (can be multiple columns...called a Composite key)
 - Best to use Natural Key vs Surrogate Key (auto #)
- Child Key is found in the related table, linking the two tables together through a common value

Horizontal Table Structure

- Also known as Flat File
- Positives
 - Most intuitive
 - 1 record contains all of the attributes describing the ‘thing’
 - Matches format for most Reports, Stats packages, GIS
- Negatives
 - Not flexible when adding new fields
 - Null values (what do they mean?)
 - Dummy values
 - Require many lookup tables

Vertical Table Structure

- Stores data going down instead of across
 - Can use Crosstab Query to make flat
- Positives
 - Flexible, can easily add in new Attributes (field)
 - No Null values stored in the table
 - Fewer lookup tables
 - I find it easier to calculate statistics on the data
- Negatives
 - Loss of strong data typing (general field type (text) to store all information (Number, Text, DateTime))
 - Less control over data

SQL Server Express

- <http://www.microsoft.com/en-us/download/details.aspx?id=52679>
- Class SQL Server
 - ServerName\InstanceName
 - ECOINFO, 6004\ECODEV
 - In R you would use ECOINFO, 6004\\ECODEV