Databases

1. Introduction

Arni Magnusson

United Nations University
Fisheries Training Programme

18–20 Nov 2014

Introduction
Database systems
Create database

Database
Design
Data types

Outline

What is a database

  purpose, design, data types

Outline

What is a database

purpose, design, data types

Create database

software, import data

Introduction
Database systems
Create database

Database
Design
Data types

Outline

What is a database

purpose, design, data types

Create database

software, import data

Query

get data, join tables, SQL language

Introduction
Database systems
Create database

Database
Design
Data types

Outline

What is a database

purpose, design, data types

Create database

software, import data

Query

get data, join tables, SQL language

Interface

connect to database from other program

Introduction
Database systems
Create database

Database
Design
Data types

Goals

After this three-day module, you should:

1. **Understand** what a database is, and how it works

Goals

After this three-day module, you should:

1. **Understand** what a database is, and how it works

2. Be able to **create** a simple database

Introduction
Database systems
Create database

Database
Design
Data types

Goals

After this three-day module, you should:

1. **Understand** what a database is, and how it works

2. Be able to **create** a simple database

3. Be able to **get data** from any database

Introduction
Database systems
Create database

Database
Design
Data types

Database

What is a database?

Introduction
Database systems
Create database

Database
Design
Data types

Database

Introduction
Database systems
Create database

Database
Design
Data types

Database

Fisheries data are expensive and important, for general research and to give management advice

Datasets live much longer than computers

Introduction
Database systems
Create database

Database
Design
Data types

Database

Fisheries data are expensive and important, for general research and to give management advice

Datasets live much longer than computers

What we DON'T want

- Data only on a laptop: can be lost by accident
- Other people cannot access data
- Difficult to relate with other datasets
- Difficult to manipulate (aggregate, subset, calculate)

Introduction
Database systems
Create database

Database
Design
Data types

Database

Fisheries data are expensive and important, for general research and to give management advice

Datasets live much longer than computers

What we DO want

- Data are safe: can undo mistakes, automatic backups
- Everyone can access data
- Combine different datasets
- Efficient data manipulation (aggregate, subset, calculate)

Databases also

- Handle massive amounts of data
- Compute very fast

Database

Good data management

  Archived and made available in a database

Good programming

  Archived and made available in a repository

Good science

  Archived and made available in a journal

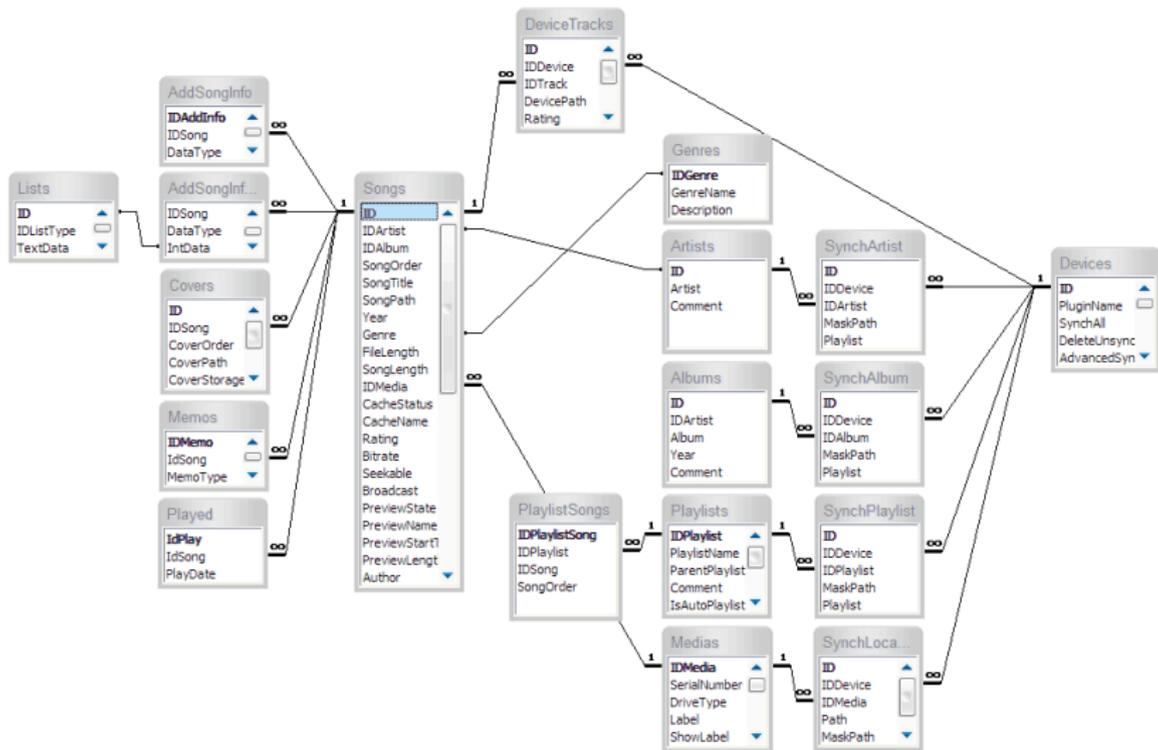$\Rightarrow$ Takes extra effort, but is worthwhile in the long run

Introduction
Database systems
Create database

Database
Design
Data types

Design

What is a database made of?

Introduction
Database systems
Create database

Database
Design
Data types

Design

Tables

- A database is a collection of tables

- Related tables are joined using key columns

Introduction
Database systems
Create database

Database
Design
Data types

Design

Introduction
Database systems
Create database

Database
Design
Data types

Design

Tables

- A database is a collection of tables

- Related tables are joined using key columns

- Each table column has one data type

Data types

	Bytes	Example
Text		
VARCHAR*(length)*	*length*	Some text
Date		
DATE	4	1999-12-31
Number		
TINYINT	1	any number from -128 to 127
SMALLINT	2	$-32\,768$ to $32\,767$
INTEGER	4	$-2\,147\,483\,648$ to $2\,147\,483\,647$
DECIMAL*(signif, round)*	*signif*	123.45

brown: standard SQL
gray: LibreOffice SQL

Introduction

Database systems

Create database

Database systems

Software

LibreOffice

Database systems

Where can I find a database?

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Database systems

Local database

- Notes that are not research data
  example: literature database (articles, books, etc.)

- Copy of a remote database
  example: global fisheries data (FishStat)

Introduction
Database systems
Create database
Database systems
Software
LibreOffice

Database systems

Local database

- Notes that are not research data
  example: literature database (articles, books, etc.)

- Copy of a remote database
  example: global fisheries data (FishStat)

Remote database

Web interface
  example: FishBase, Google Scholar, Web of Science

Direct access
  example: typical fisheries data (catch, surveys, tags, biology)

Introduction
Database systems
Create database
Database systems
Software
LibreOffice

Database systems

The database system we will focus on is a remote database that you have direct access to

The database is running on a powerful server that is available 24/7, servicing multiple users

The data are always safely backed up, even in the case of power failure, flooding, fire, etc.

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Database systems

When you work for an institute, they already have a database

Building and maintaining a large database is a complicated job, requiring years of training

Important foundation of all research and management advice

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Database systems

The database administrator gives you read access to parts of the
database that you need for your work

Data that you work with should be in a central database for the
institute, not in a personal database on your computer

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Software to run a database system

Powerful

IBM DB2

MariaDB

Microsoft SQL Server

MySQL

Oracle

PostgreSQL

brown: free software

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Software to run a database system

Powerful

IBM DB2

MariaDB

Microsoft SQL Server

MySQL

Oracle

PostgreSQL

Simple

Firebird

LibreOffice Base

Microsoft Access

SQLite

brown: free software

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Interface

The program we use to communicate with the database is called a client, or interface

For example, here at Hafro I'm using R as an interface to the Oracle database system

An R script for data analysis often starts with a few lines of code to get a dataset from the database

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Interface

Usually, you don't need to think about what software the database system is running on

The database administrator will help you to connect your preferred interface to the database

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

LibreOffice

We will use LibreOffice Base to learn how a database system works

It is very limited, but good enough to learn from

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

LibreOffice

We will use LibreOffice Base to learn how a database system works

It is very limited, but good enough to learn from

Tomorrow, we will use R as an interface to query data from our LibreOffice database

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Install LibreOffice

Open in a web browser

`libreoffice.org`

and select: Download

Introduction
Database systems
Create database

Database systems
Software
LibreOffice

Terminology

Different database systems use different words to describe things

In LibreOffice Base:

- Field = table column

- Record = table row

Introduction
Database systems
Create database
Prepare
Create database
Create table

Create database

Now we will create our first database

Introduction
Database systems
Create database
Prepare
Create database
Create table

Data

Let's make up some data

[on the projector]

Introduction
Database systems
Create database

Prepare
Create database
Create table

Data

Name	Country	Capital	Siblings	Cars	Movie
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

Name	what is your first name?
Country	where do you come from?
Capital	what is the capital of that country?
Siblings	how many siblings do you have (incl. yourself)?
Cars	how many cars have you owned?
Movie	have you ever watched an Icelandic movie (yes/no)?

Introduction
Database systems
Create database

Prepare
Create database
Create table

Data

Name	Country	Capital	Siblings	Cars	Movie
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

What is the data type of each column?

Text	VARCHAR			
Date	DATE			
Number	TINYINT,	SMALLINT,	INTEGER,	DECIMAL
(max)	127	32 767	2 147 483 647	999999.999...

Introduction
Database systems
Create database

Prepare
Create database
Create table

Create database

- Create directory c:/database

- Start LibreOffice Base

- Create a new database   [next]

- No, do not register   [finish]

- Navigate to c:/database and save as TeamDB.odb

Introduction
Database systems
Create database

Prepare
Create database
Create table

Create table structure

- Database – Tables

- Create table in design view

- Fill in column names and types

- File – Save as PEOPLE

- Create primary key: Yes

- Window – Close window

(Table names in LibreOffice Base should be UPPERCASE)

Introduction
Database systems
Create database

Prepare
Create database
Create table

Import data

- Open data in LibreOffice Calc

- Copy area containing data

- Paste into database

(Pasting from LibreOffice Calc is much faster than pasting from other programs)

Introduction
Database systems
Create database

Prepare
Create database
Create table

Primary key

A primary key is the backbone of a database table

Every value in this column must be unique

Usually the first column of a table

Introduction
Database systems
Create database

Prepare
Create database
Create table

Real data

Logbook data from Icelandic fisheries

- Take a look at catch.csv

- Find out how many columns (and rows) there are in the data

- Find out what the column data types are

- Import into a new database called onetable.odb

- Call the table CATCH and set a two-column primary key

(A two-column primary key means that every row has a unique combination of these two columns)