Information Systems (Informationssysteme)

Jens Teubner, TU Dortmund
jens.teubner@cs.tu-dortmund.de

Summer 2016

Part III

A Very Brief Introduction to SQL

SQL—Structured Query Language

By far the most important query language today is **SQL**.

- Structured Query Language
- Originally meant to be used by end users ☺
- Today supported by virtually any database system

SQL operates on **relational data**:

Ingredients			
Name	Alcohol	InStock	Price
Orange Juice	0.0	12	2.99
Campari	25.0	5	12.95
Mineral Water	0.0	10	1.49
Bacardi	37.5	3	16.98

- Real databases may contain 100s or 1000s of **tables**, sometimes with billions of **rows** (also: **tuples**).

Our First SQL Query

The key construct of SQL is the `SELECT-FROM-WHERE` clause:

```
SELECT Name, Price
  FROM Ingredients
 WHERE Alcohol = 0
```

`SELECT` Choose a **set of columns** to be reported in the query result.

We'll later call this **projection**, **not** selection.

`FROM` Choose a **table** where rows should be taken from.

`WHERE` Additional **conditions** that rows must satisfy in order to appear in the result (the `WHERE` clause is optional).

→ **This** is what we call a **selection**.

Ingredients			
Name	Alcohol	InStock	Price
Orange Juice	0.0	12	2.99
Campari	25.0	5	12.95
Mineral Water	0.0	10	1.49
Bacardi	37.5	3	16.98

+

```
SELECT Name, Price
  FROM Ingredients
 WHERE Alcohol = 0
```

=

Name	Price
Mineral Water	1.49
Orange Juice	2.99

Data in Multiple Tables

Cocktail ingredients are sold by various suppliers (for a certain price), which could be represented as

SoldBy			
Ingredient	Supplier	DelTim[3]	Price
Orange Juice	A&P Supermarket	1	2.49
Orange Juice	Shop Rite	3	2.79
Campari	Joe's Liquor Store	2	14.95
Bacardi	Liquor's & More	5	13.99
Mineral Water	Shop Rite	3	1.89
Bacardi	Joe's Liquor Store	2	14.99

[3] Delivery time in days.

When multiple tables are reference in the `FROM` clause, this is interpreted as the **Cartesian product** of the referenced tables:[4]

```
SELECT *
  FROM Ingredients, SoldBy
```

Ingredients				SoldBy			
Name	Alcohol	InStock	Price	Ingredient	Supplier	DelTim	Price
Orange Juice	0.0	12	2.99	Orange Juice	A&P Supermarket	1	2.49
Orange Juice	0.0	12	2.99	Orange Juice	Shop Rite	3	2.79
Orange Juice	0.0	12	2.99	Campari	Joe's Liquor Store	2	14.95
Orange Juice	0.0	12	2.99	Bacardi	Liquors & More	5	13.99
Orange Juice	0.0	12	2.99	Mineral Water	Shop Rite	3	1.89
Orange Juice	0.0	12	2.99	Bacardi	Joe's Liquor Store	2	14.99
Campari	25.0	5	12.95	Orange Juice	A&P Supermarket	1	2.49
Campari	25.0	5	12.95	Orange Juice	Shop Rite	3	2.79
Campari	25.0	5	12.95	Campari	Joe's Liquor Store	2	14.95
Campari	25.0	5	12.95	Bacardi	Liquors & More	5	13.99
Campari	25.0	5	12.95	Mineral Water	Shop Rite	3	1.89
Campari	25.0	5	12.95	Bacardi	Joe's Liquor Store	2	14.99
Mineral Water	0.0	10	1.49	Orange Juice	A&P Supermarket	1	2.49
Mineral Water	0.0	10	1.49	Orange Juice	Shop Rite	3	2.79
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

[4] Use `*` in the `SELECT` clause when you simply want to choose all columns.

Queries over Multiple Tables

In practice, you rarely want to see this Cartesian product in the final result.

→ Use a `WHERE` clause to select only semantically related data.

```
SELECT Name, InStock, Supplier
  FROM Ingredients, SoldBy
 WHERE Name = Ingredient
```

↓

Name	InStock	Supplier
Orange Juice	12	A&P Supermarket
Orange Juice	12	Shop Rite
Campari	5	Joe's Liquor Store
Mineral Water	10	Shop Rite
Bacardi	3	Liquors & More
Bacardi	3	Joe's Liquor Store

Queries over Multiple Tables

Resolve ambiguities by prepending column names with their table name:

```
SELECT Name, InStock, Supplier, SoldBy.Price
  FROM Ingredients, SoldBy
 WHERE Name = Ingredient
   AND SoldBy.Price < Ingredients.Price
```

Name	InStock	Supplier	Price
Orange Juice	12	A&P Supermarket	2.49
Orange Juice	12	Shop Rite	2.79
Bacardi	3	Liquors & More	13.99
Bacardi	3	Joe's Liquor Store	14.99

Tuple Variables

. . . or introduce **tuple variables** for easier reference:

```
SELECT Name, InStock, Supplier, s.Price
  FROM Ingredients AS i, SoldBy AS s
 WHERE Name = Ingredient
   AND s.Price < i.Price
```

Name	InStock	Supplier	Price
Orange Juice	12	A&P Supermarket	2.49
Orange Juice	12	Shop Rite	2.79
Bacardi	3	Liquors & More	13.99
Bacardi	3	Joe's Liquor Store	14.99

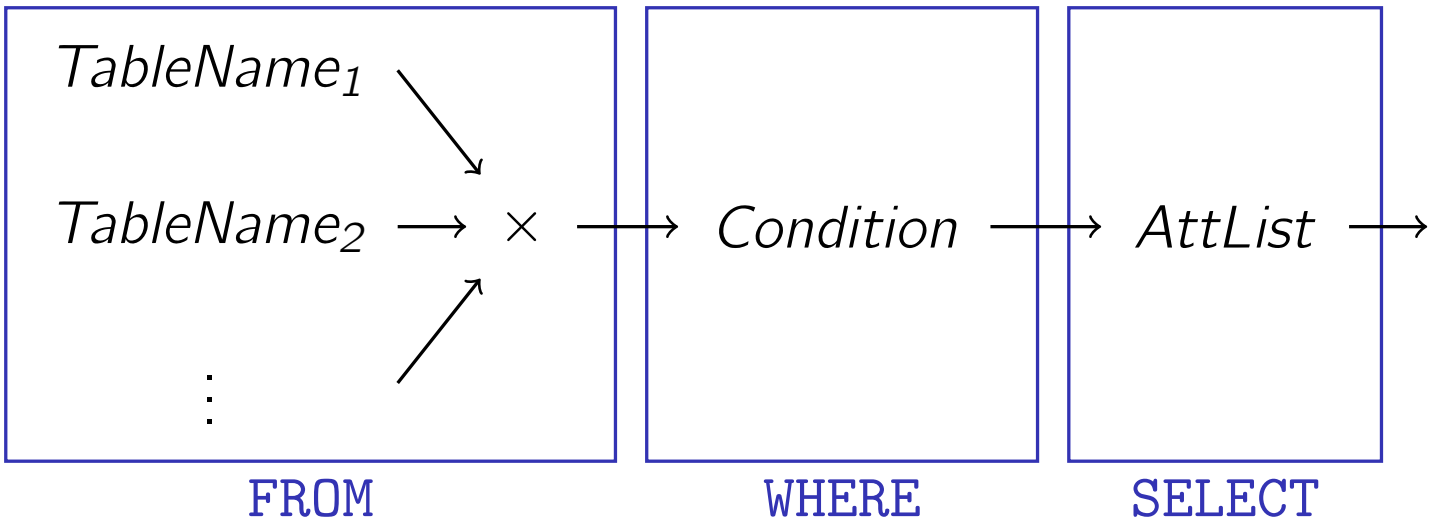
(The keyword `AS` is optional; '`SoldBy s`' would mean just the same.)

Semantics of SQL SELECT-FROM-WHERE Expressions

Conceptually, the query

```
SELECT AttList
  FROM TableName1 , TableName2 , ...
 WHERE Condition
```

does the following:



(But most likely, the database system will choose a better strategy to actually execute the query.)

Concluding Remarks

- SQL is **case insensitive**; use ’ as a **string delimiter**.
- It is okay to reference the **same table multiple times** in a `FROM` clause ($\rightarrow$ “self-join”). Use **tuple variables** then to tell things apart.

Never, **never ever**, write queries where the correctness depends on the current table contents.

E.g., the correct answer to “give me names and prices of all non-alcoholic ingredients” is **not**

```
SELECT Name, Price
  FROM Ingredients
 WHERE Name = 'Orange Juice' OR Name = 'Mineral Water'
```