Dr Greg Wadley



INFO90002 Database Systems & Information Modelling

Week 02
Data Modelling & SQL (1)



MELBOURNE The first half of semester

Week	Lecture 1	Lecture 2
2	Modelling 1 conventions, 1-M relationships	SQL 1 group by, having, joins
3	Modelling 2 M-M, 1-1, unary relationships	SQL 2 subqueries, functions
4	Modelling 3 ternary relationships, views	SQL 3 set operations, formatting
5	Normalization 1st, 2nd, 3rd normal forms	Physical design data types, indexes, denormalising
6	Data Dictionaries	Wrapup and Q&A

end of Week 6: Assignment One on data modelling is due

Data modelling

- ER modelling conventions
- Identifying entities and business rules
- one-to-many (1-M) relationships

SQL

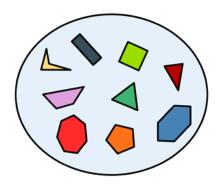
- Overview and history
- Create tables
- Insert data into tables
- Read data from tables



Databases and ER models

- A database can be thought of as a representation of
 - a collection of entity sets, and
 - relationships between the entities
- An entity is an object or abstract concept or event which can be distinguished from other entities
 - example: product, order, sale, person, movie, tweet
- Entities have attributes that describe the entity and distinguish it from other entities in the same entity set
 - example attributes: EmployeeName, Address

- (reminder what are "sets"?
 - union, intersection, Cartesian product)





Valid and invalid entities

An Entity

- Will have many instances in the database
- Has several attributes
- Is necessary for the system to work

Examples

Person: EMPLOYEE, STUDENT, PATIENT

Place: STORE, WAREHOUSE, STATE, CITY

Object: PRODUCT, MACHINE, BUILDING, VEHICLE

Event: SALE, REGISTRATION, BROADCAST

Abstract: ACCOUNT, UNI SUBJECT, ROLE

Entities do not usually include:

- An output of the system (i.e. a report)
- The system itself
- The company that owns the system

- Entities
 - singular nouns
 - Employee, Customer, Sale (without or without capital letter)
- Attributes
 - usually a noun
 - itemColour, quantitySold, id
- Relationships
 - verbs or verb phrases
 - has, wants, manages, performs work for
- Use names meaningful to the domain
 - try not to abbreviate names
 - except num or nbr for number, ID for identifier
 - conventions on UPPER and lower case

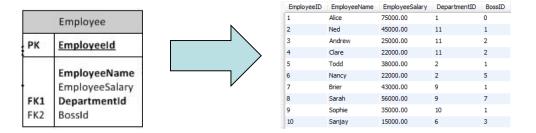
Modelling a Single Entity

- By searching for nouns in the case we can identify entities (for example – Customer)
- What things would we need to record about the Customer
 - these become the customer's Attributes

- How can we identify individual Customers?
 - by name?
 - by address?
- Now we can draw it as an <u>entity</u> in the ER diagram



Mapping ER diagram to database tables



- Entity set
 - Often corresponds to a table in the database
- Entity instance
 - Often corresponds to a row in a table
- Attribute
 - Often corresponds to a column in a table
- Relationship set (link between entity sets)
 - Often corresponds to a Foreign Key in a table
- Relationship instance (link between entity instances)
 - Foreign Key value = Primary Key value

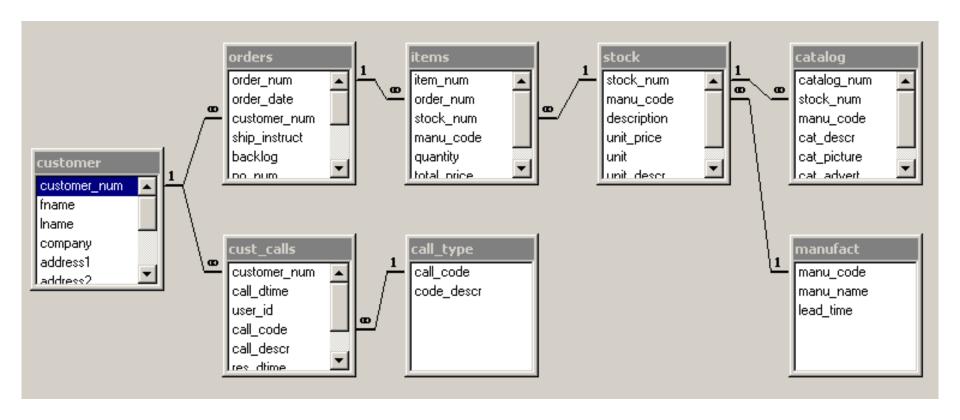


How to draw ER diagrams?

- You can use paper or the software of your choice
- e.g. Visio, https://www.lucidchart.com/ ...
- Assignment 1 model should be made in MySQL Workbench
- In the exam you'll need to model on paper
- Diagrams in lecture slides are made in Workbench and Visio
- My suggestion is:
 - Use pen-and-paper or whiteboard for early Conceptual modelling
 - Use paper, Visio or Workbench for subsequent Logical modelling
 - Use Workbench for the final Physical model

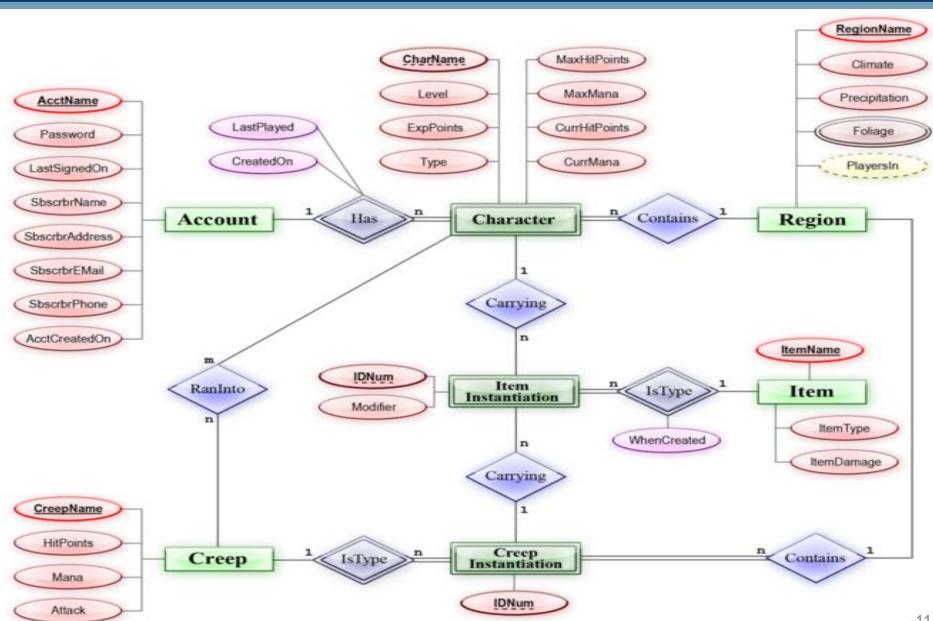


MELBOURNE Variation in ER diagram standards





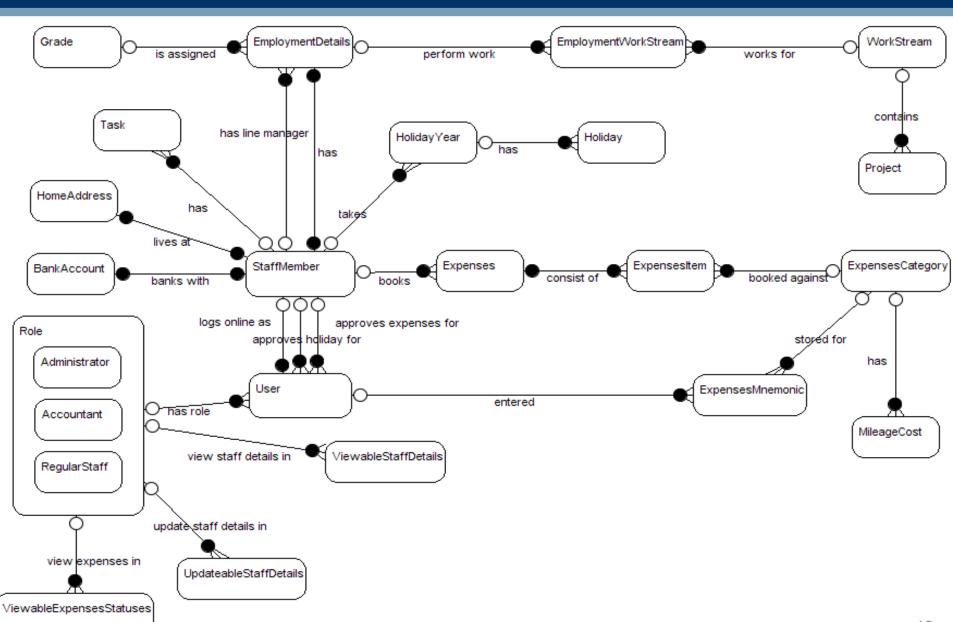
Variation in ER diagram standards



-11-



MELBOURNE Variation in ER diagram standards





Representing entities and attributes

Entity

Entity1				
PK <u>Identifier</u>				
	Ent1Attribute1 Ent1Attribute2			

Attributes

EntityAttributeExample

- 💡 PartialIdentifier
- PartialIdentifier2
- Mandatory
- Optional
- Item1
- Item2

- Key (or Identifier)
 - Fully identifies an instance
- Partial Key
 - Partially identifies an instance
- Attributes
 - Mandatory
 - Optional
 - Derived
 - [YearsEmployed]
 - Multivalued
 - {Skill}
 - Composite
 - Name (First, Middle, Last)



Conceptual design for single entity

(drawn using Visio software)

Customer1				
PK	CustomerID			
	CustFirstName CustMiddleName CustLastName BusinessName CustType CustAddress(Line1, Line 2, Suburb, Postcode, Country)			

- underline = primary key
- bold = not null
- () = composite attribute



Convert to *logical* design

Customer1				
PK	CustomerID			
	CustFirstName CustMiddleName CustLastName BusinessName CustType CustAddress(Line1, Line 2, Suburb, Postcode, Country)			

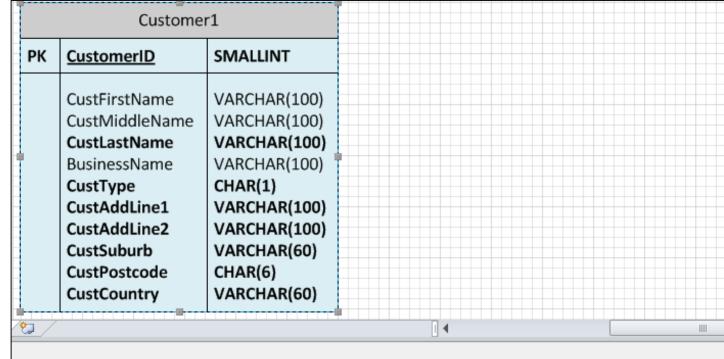
Customer1				
PK	CustomerID			
	CustFirstName CustMiddleName CustLastName BusinessName CustType CustAddLine1 CustAddLine2 CustSuburb CustPostcode CustCountry			

- Composite attributes become individual attributes
- Multi-valued attributes become a new table
- Resolve many-many relationships via a new table
- Add foreign keys at crows foot end of relationships



Convert to physical design

 Determine data types for each attribute



Key, nullable

	Physical Name	Data Type	Req'd	PK	Notes
	CustomerID	SMALLINT	~	~	CustomerID identifies Customer1
	CustFirstName	VARCHAR(100)			CustFirstName is of Customer 1
	CustMiddleName	VARCHAR(100)			CustMiddleName is of Customer 1
	CustLastName	VARCHAR(100)	~		CustLastName is of Customer 1
	BusinessName	VARCHAR(100)			BusinessName is of Customer 1
\blacktriangleright	CustType	CHAR(1)	~		NOTE: This will be implemented as an ENUM type in MySQL
	CustAddLine1	VARCHAR(100)	~		CustAddLine1 is of Customer1
	CustAddLine2	VARCHAR(100)	~		CustAddLine2 is of Customer 1
	CustSuburb	VARCHAR(60)	~		CustSuburb is of Customer 1
	CustPostcode	CHAR(6)	~		CustPostcode is of Customer 1
	CustCountry	VARCHAR(60)	~		CustCountry is of Customer 1



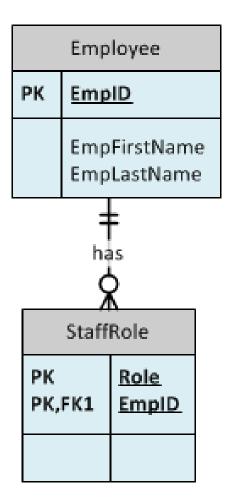
Dealing with multi-valued attributes

Conceptual Design

Employee				
PK EmpID				
	EmpFirstName EmpLastName {Role}			

StaffRole is an example of a weak entity

Logical Design





Business Rules and Relationships

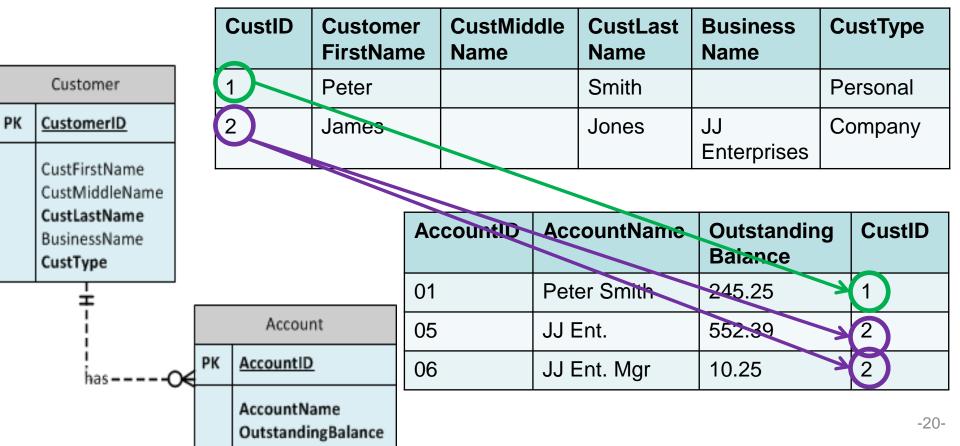
- Business rules are assertions that constrain entities
- Can impact structure and behaviour of the database
- Business rules can be assertions about attributes
 - "Quantity bought must be between 1 and 200." (assertion)
 - quantity bought (attribute)
- Or business rules can be assertions about entities
 - "A customer sets up at least one account." (assertion)
 - customer (entity)
 - account (entity)
- The latter kind of business rules are represented in our data models as relationships between entities.

- Keys or Identifiers are used to identify individual entity instances
 - Primary Key
 - (set of) columns, the values in which uniquely identify each instance
 - no column can be removed from the key without losing uniqueness
 - Candidate Key
 - the set of possible primary keys (choose one to be the PK)
 - Surrogate Key
 - system-assigned serial number (used if natural PK is unavailable or unsuitable)
 - Composite Key
 - a key which is made up of more than one attribute
 - e.g. for the entity "airline flight" we might use the composite key
 - » FlightNumber + FlightDate
 - Foreign Key
 - the key used to link to a primary key in another table
 - helps us to join tables in a Select statement
- Primary Keys are
 - unique
 - never null
 - do not change their value



Two entities with 1-M relationship

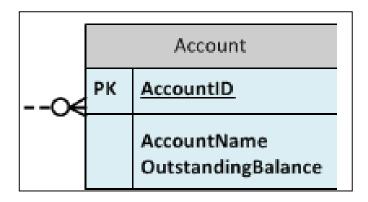
- Example: "A customer can set up several accounts."
 - The tables get linked through a foreign key





Conceptual to logical design – with FK

Conceptual Design



Logical Design

		Account
	PK	<u>AccountID</u>
·Œ	FK1	CustomerID AccountName
		OutstandingBalance

- Add foreign keys at crows feet end of relationships
 - FK1 CustomerID
 - This is the link to the customer table
 - Every CustomerID in Account must be present in Customer
 - » Referential integrity



Physical design - with FK

Attribute data types

FK must have the same data type as the PK it refers to.

Customer						
PK	PK <u>CustomerID</u> SMALLINT					
CustFirstName VARCHAR(100) CustMiddleName VARCHAR(100) CustLastName VARCHAR(100) BusinessName VARCHAR(100) CustType CHAR(1)						
H has						
	Accou	nt				
PK	PK AccountID INTEGER					
FK1	K1 CustomerID AccountName OutstandingBalance		SMALLINT VARCHAR(100) DECIMAL(10,2)			



Now we can create the tables

```
CREATE TABLE Customer (
  CustomerID
                   smallint
                                               auto increment,
 CustFirstName
                 varchar(100),
 CustMiddleName
                   varchar(100),
  CustLastName
                   varchar(100)
                                               NOT NULL.
  BusinessName
                   varchar(200),
                   enum('Personal','Company') NOT NULL.
 CustType
  PRIMARY KEY (CustomerID)
 ENGINE=InnoDB:
```

```
□CREATE TABLE Account (
                         smallint
   AccountID
                                         auto increment,
   AccountName
                                         NOT NULL,
                         varchar(100)
                         DECIMAL(10,2)
   OutstandingBalance
                                         NOT NULL,
   CustomerID
                         smallint
                                         NOT NULL,
   PRIMARY KEY (AccountID),
   FOREIGN KEY (CustomerID) REFERENCES Customer(CustomerID)
         ON DELETE RESTRICT
         ON UPDATE CASCADE
                                Referential Actions
 ) ENGINE=InnoDB;
                                how foreign keys
                                guarantee referential
                                integrity.
```



Insert - with FK

Current Database

CustID	CustomerFirstName	CustMiddle Name	CustLastName	BusinessName	CustType
1	Peter		Smith		Personal
2	James		Jones	JJ Enterprises	Company

AccountID AccountName Outstand	lingBalance CustID
------------------------------------	--------------------

Insert a row...

INSERT INTO ACCOUNT VALUES (DEFAULT, 'My New Account', 0, 5);

What happens?

INSERT INTO ACCOUNT VALUES (DEFAULT, ... Error Code: 1452. Cannot add or update a child row: a for

Error Code: 1452. Cannot add or update a child row: a foreign key constraint fails (`db_seanbm/account`, CONSTRAINT `account_ibfk_1` FOREIGN KEY (`CustomerID`) REFERENCES `customer` (`CustomerID`))

Run the Inserts...

```
INSERT INTO ACCOUNT VALUES (DEFAULT, 'Peter Smith', 245.25, 1);
INSERT INTO ACCOUNT VALUES (DEFAULT, 'JJ ENt.', 552.39, 2);
INSERT INTO ACCOUNT VALUES (DEFAULT, 'JJ ENt. Mgr', 10.25, 2);
```

CustID	CustomerFirstName	CustMiddle Name	CustLastName	BusinessName	CustType
1	Peter		Smith		Personal
2	James		Jones	JJ Enterprises	Company
3	Akin		Smithies	Bay Wart	Company

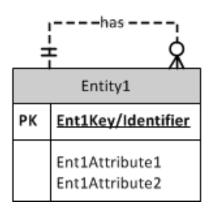
AccountID	AccountName	OutstandingBalance	CustID
01	Peter Smith	245.25	1
02	JJ Ent.	552.39	2
03	JJ Ent. Mgr	10.25	2



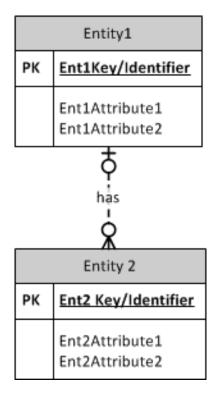
Relationship Degree

How many entities take part in the relationship?

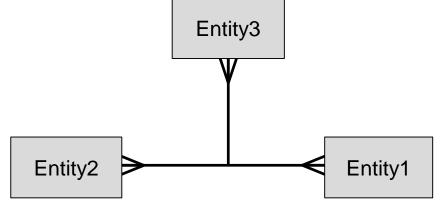








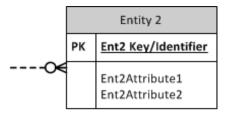
Ternary (3)



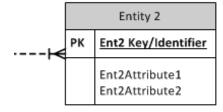


Relationship Cardinality

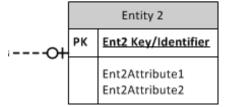
- One to One
 - Each entity in one set is related to 0 or 1 in the other.
- One to Many
 - Each entity in one set is related to many in the other.
- Many to Many
 - Each entity in either set can be related to many in the other set
 - These require an extra step to implement in a relational database.



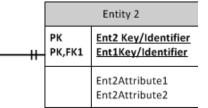
Optional Many



Mandatory Many



Optional One



Mandatory One



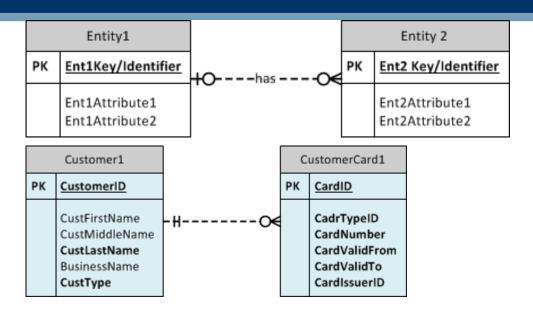
Strong and Weak entities

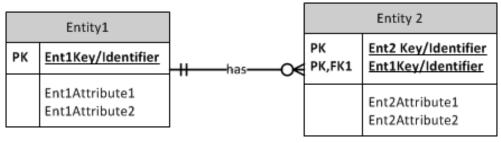
Strong Entity

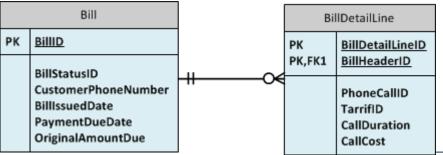
entity 2's PK is independent of the PKs of other entities

Weak Entity

entity 2's PKdepends on(includes)the PK of entity 1









Structured Query Language (SQL)

- SQL or SEQUEL is a language used to create, access and maintain relational databases
- Based on relational algebra and relational calculus
- SQL (DML) supports CRUD (Create, Read, Update, Delete)
 - Insert, Select, Update, Delete commands
- You can see the latest SQL 2011 standard at
 - http://www.jtc1sc32.org/doc/N2151-2200/32N2153T-text_for_ballot-FDIS_9075-1.pdf
- SQL is a widely used standard and there are resources online:
 - http://en.wikipedia.org/wiki/SQL
 - http://en.wikipedia.org/wiki/Category:SQL_keywords
 - https://dev.mysql.com/doc/refman/5.7/en/sql-syntax.html
 - https://www.w3schools.com/sql/

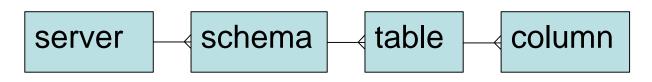


Brief history

1974	IBM develops SEQUEL (renamed to SQL) based on Codd research
1979	Oracle, IBM etc release RDBMS with SQL language
1986	1 st SQL Standard (ANSI)
1989	2 nd SQL Standard (ANSI) – includes referential integrity
1992	3 rd SQL Standard (ISO) – most widely conformed to by vendors
1997	dynamic websites enabled by SQL
1999	SQL-1999 – 4 th SQL Standard (ISO) – Object support, recursion,
	procedures and flow control
2003	SQL-2003 – 5 th SQL Standard (ISO) – XML support, auto number
2006	SQL-2006 – 6 th SQL Standard (ISO) – Defines SQL use with XML
2008	SQL-2008 – 7 th SQL Standard (ISO) – FETCH command added
2008	HTML 5 with SQLite built in
2011	SQL-2011 – 8 th SQL Standard (ISO) – temporal databases

- during Implementation of the database
 - Implement tables from physical design using DDL Create Table
- during Production
 - use Select commands to read the data from the tables
 - use DML Insert, Delete, Update commands to update data
 - use DML Alter, Drop commands to update the database structure

- We are using the MySQL implementation of SQL
 - If you are using other DBMS (such as ORACLE or SQLServer) you will need to check their implementation of SQL.
 - differences can range from valid keywords to data types
- The university's MySQL server = version 5.7.9
- You can get the latest version of MySQL (5.7) from
 - http://dev.mysql.com/downloads/
 - Community edition = FOSS
 - Get syntax help for MySQL SQL statements at
 - http://dev.mysql.com/doc/refman/5.7/en/sql-syntax.html
- Explore your server with these commands:
 - show schemas; (alternatively, 'show databases')
 - show tables;
 - describe table;



Consists of:

- Data Definition Language (DDL)
 - to define and set up the database
 - CREATE, ALTER, DROP
 - also TRUNCATE, RENAME
- Data Manipulation Language (DML)
 - to manipulate and read data in tables
 - SELECT, INSERT, DELETE, UPDATE
 - MySQL also provides others.... eg REPLACE
- Data Control Language (DCL)
 - to control access to the database
 - GRANT, REVOKE
- Other Commands
 - administer the database
 - transaction control



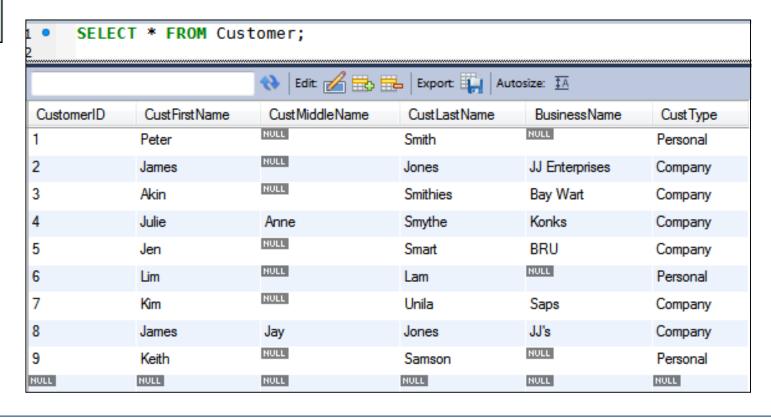
* MELBOURNE The SELECT Statement in detail

```
SELECT [ALL | DISTINCT] select_expr [, select_expr ...]
   List the columns (and expressions) that are returned from the query
[FROM table references
   Indicate the table(s) or view(s) from where the data is obtained
[WHERE where condition]
   Indicate the conditions on whether a particular row will be in the result
[GROUP BY {col_name | expr } [ASC | DESC], ...]
   Indicate categorisation of results
[HAVING where condition]
   Indicate the conditions under which a particular category (group) is included
   in the result
[ORDER BY {col_name | expr | position} [ASC | DESC], ...]
   Sort the result based on the criteria
[LIMIT {[offset,] row_count | row_count OFFSET offset}]
   Limit which rows are returned by their return order (ie 5 rows, 5 rows from
   row 2)
```



Select entire contents of table

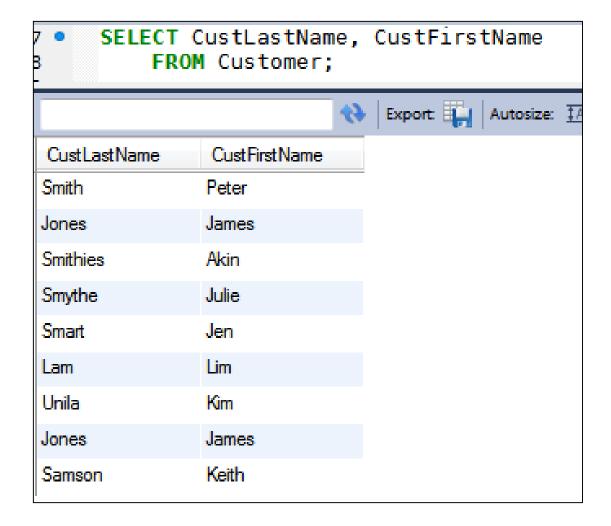
Customer		
PK	CustomerID	
	CustFirstName CustMiddleName CustLastName BusinessName CustType	





Select specific columns

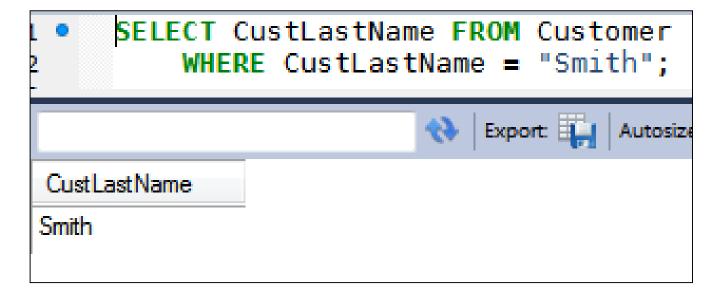
Customer	
PK	CustomerID
	CustFirstName CustMiddleName CustLastName BusinessName CustType





MELBOURNE WHERE clause: select specific rows

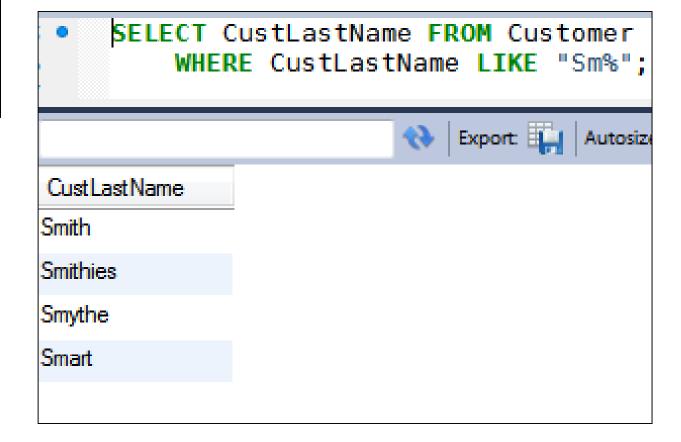
Customer		
PK	CustomerID	
	CustFirstName CustMiddleName CustLastName BusinessName CustType	





MELBOURNE WHERE clause with LIKE

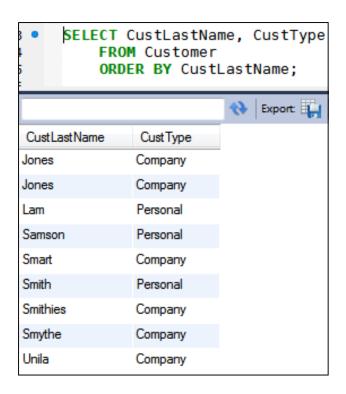
Customer		
PK	CustomerID	
	CustFirstName CustMiddleName CustLastName BusinessName CustType	

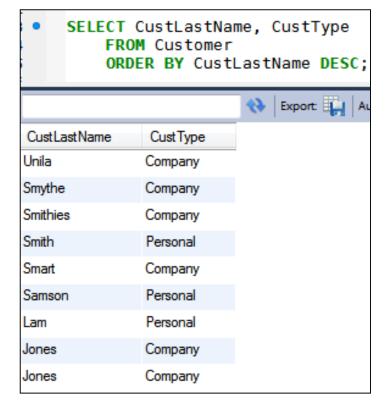




Select with ORDER BY

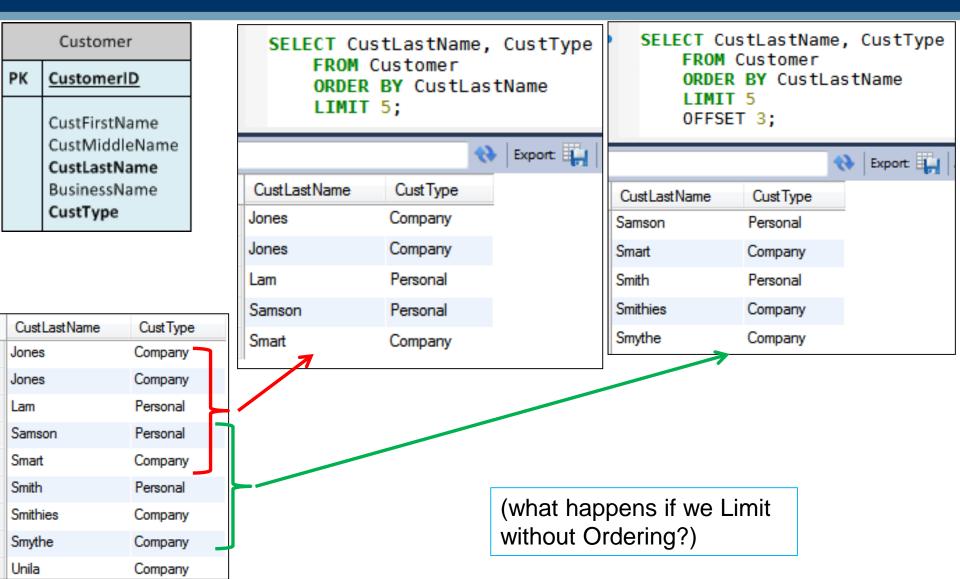
Customer		
PK	CustomerID	
	CustFirstName CustMiddleName CustLastName BusinessName CustType	







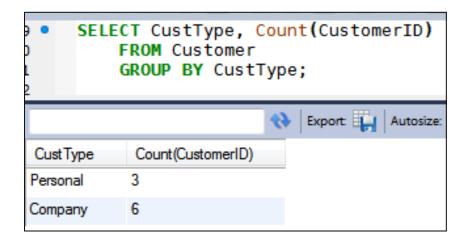
Select with LIMIT

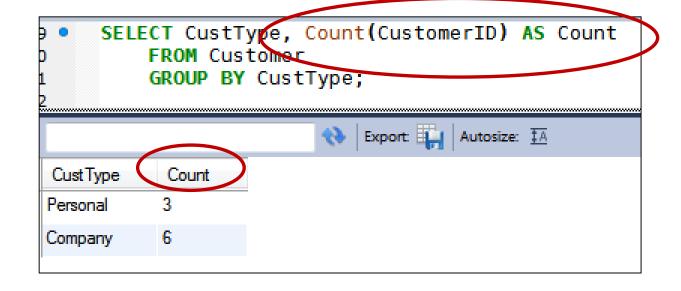




Select with GROUP BY

Customer		
PK	CustomerID	
	CustFirstName CustMiddleName CustLastName BusinessName CustType	

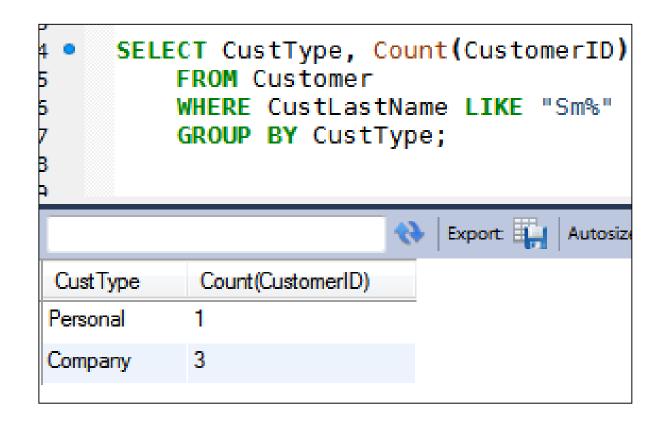






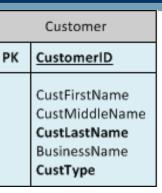
Select with WHERE and GROUP BY

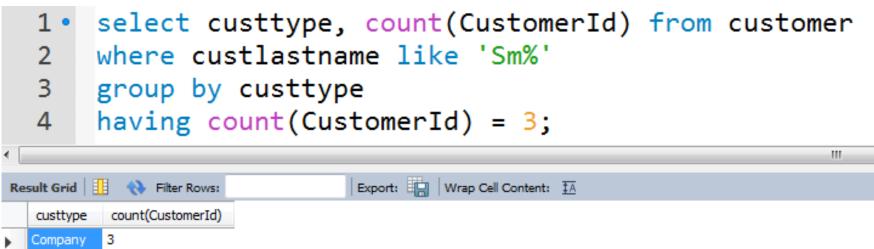
Customer	
PK	CustomerID
	CustFirstName CustMiddleName CustLastName BusinessName CustType





Select with GROUP BY and HAVING



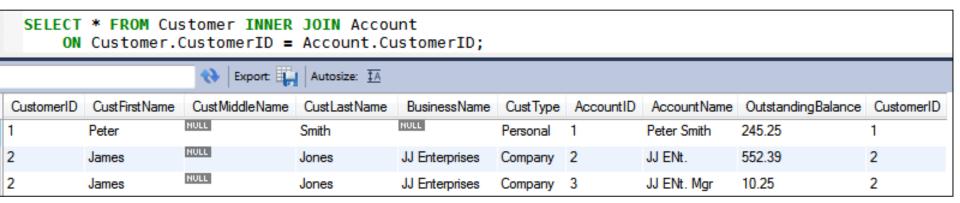


"Having" works on groups the way "Where" works on individual rows

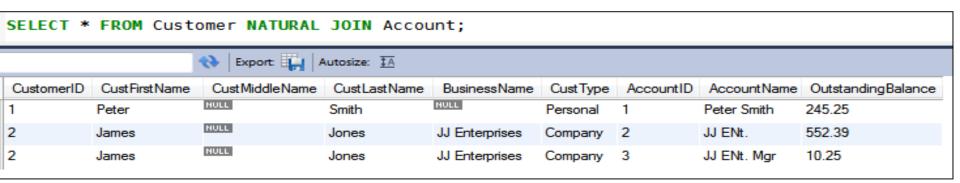


Inner join, Natural join

- Data about an entity is spread across 2 tables so join them
- Inner/Equi join Join rows where FK value = PK value

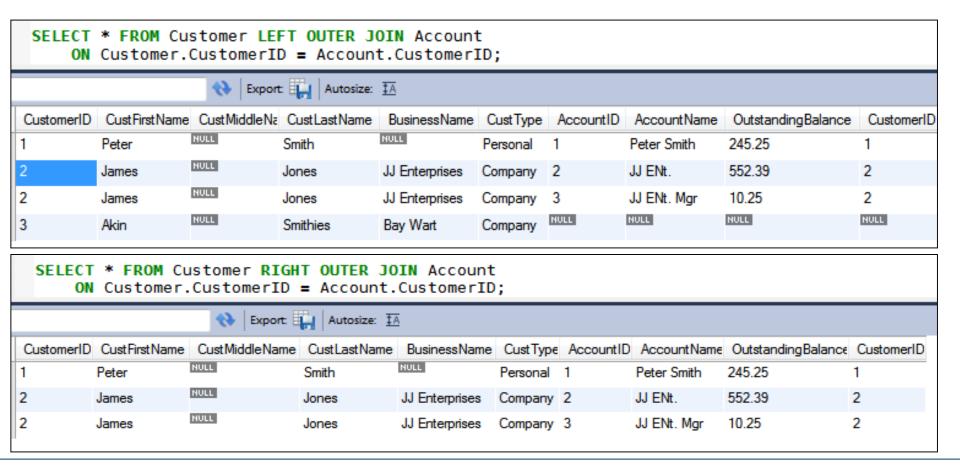


- Natural Join gives the same result as Inner Join
 - requires PK and FK columns to have the same name



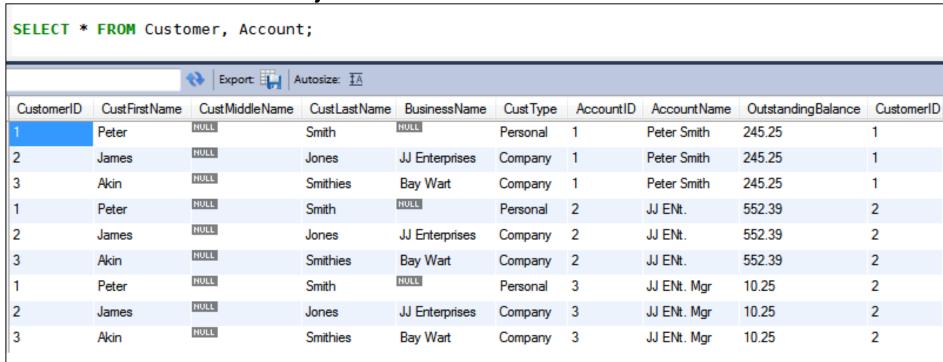
Outer join

- Can be left or right (see difference below)
- Includes records from left/right table that don't have a matching row





What if there is no join condition?



NOT CORRECT: lack of join conditions -> Cartesian product

(every row in Customer combined with every record in Account)