

# DATA MANAGEMENT



ITC / DEPARTMENT OF GEO-INFORMATION PROCESSING  
Presented by Lucas De Olo - September 2023




1

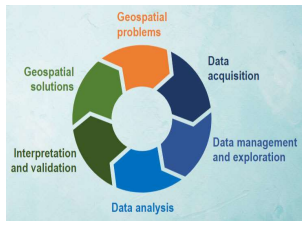

## OVERVIEW

- Databases and DBMS
  - Definitions and basic terminology
- Relational data model
  - Concepts and terminology
  - Domain, Attribute, Relation, Schema, SQL
  - Constraints
- GIS and DBMS
  - Spatial databases
  - GDB design

3

## OVERVIEW

2

## DATABASE

DATA, DATASET, DATABASE, DBMS AND DB SYSTEM

**Data**


- Is a **resource held on paper or in digital format** that serves to record or administer some **facts and descriptions of phenomena of interest**.

**Data set (or dataset):**

- A homogeneous **collection of data** normally describing a single kind of phenomenon

**Database (DB)**

- A **collection of interrelated data sets properly structured** by means of, and stored through a DBMS (Data Base Management System)



4

1

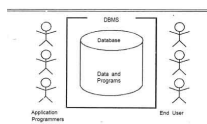
2

## DATABASE

DATA, DATASET, DATABASE, DBMS AND DB SYSTEM


**Database management system (DBMS)**

- A **software package** that is designed for the purpose of managing databases. This means, DBMS allows to set-up, maintain and explore one or more databases.



**Database system**

- Combination of a database and its DBMS.



5

## TERMINOLOGY

### DATA MODEL



A DBMS works based on a certain **data model**.

A data model **organizes elements of data and standardizes how they relate to one another**

A data model is an integrated collection of:

- Data structuring primitives,**
- Rules of how to structure, and**
- Mechanisms to handle the data**

In other words, a data model is a **toolbox** that allows us to **create/define** a database structure and **manipulate** the data stored in it.





7

## DATABASE MANAGEMENT SYSTEM (DBMS)

WHAT DOES A DBMS OFFER TO USERS?

- Supports proper storage and manipulation of **very large data sets**.
- Can be instructed to guard over **data correctness**.
- Allows the **control of data redundancy**.
- Supports the **concurrent use** of the same data set by many users.
- Includes data **backup and recovery functions** to ensure data protection and availability at all times.



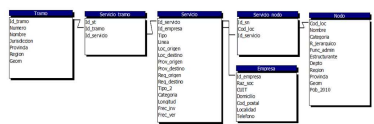

6

## RELATIONAL DATA MODEL

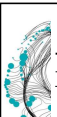
WHY STUDY RELATIONAL DATA MODEL?

There are different data models: hierarchical, network, object-oriented, etc. → DBMS linked to most GIS packages make use of the **relational data model**

The **relational data model** structures a DB as a **collection of inter-related tables (or relations)** → we model a piece of reality as linked tables

8




# TERMINOLOGY

THE LANGUAGE USED IN RELATIONAL DATA MODEL

---


**Structured Query Language (SQL)** – the relational database language:

- SQL is a DSL (*domain-specific language*) used in programming and designed for managing data held in a relational DBMS.
- **Natural language** → its syntax is easy to read
- Developed in the 70's → initially called SEQUEL (Structured English QUERY Language)
- **Particularly useful in handling structured data where there are relations between different entities/variables of the data.**



UNIVERSITY OF TWENTE

9




# TERMINOLOGY

## RELATIONAL DATA MODEL

---


The relational data model is an integrated collection of:

1. Data structuring primitives = **attributes, tuples, and relations**
2. Rules of how to structure = **data definition language**
3. Mechanisms to handle the data in a database = **data manipulation language**.



UNIVERSITY OF TWENTE

11



# TERMINOLOGY

## THE LANGUAGE USED IN RELATIONAL DATA MODEL

Find all cities (codes and names) that belong to the department called Tlaxcala


**A**

Code	Name	Dep
38077020	Ciudad de Palcos	Santa Catarina
38077030	Palcos	Santa Catarina
38084000	Corumbá	Sergipe
38084010	Do Isidro de Delfino	Sergipe
38084020	Praia do	Sergipe
38084030	Maracá	Sergipe
38084040	Praia do Chico	Sergipe
38084050	Chico Chico	Sergipe
38084060	Mata Pretendida	Sergipe
38084070	El Temo	Sergipe
38084080	Chaparral	Sergipe
38084090	Jamua	Sergipe
38084100	Chico	Sergipe
38084110	Tlaxcala	Tlaxcala
38084120	Manzanillo	Tlaxcala
38084130	Colima San José	Tlaxcala
38084140	Juárez	Tlaxcala
38084150	Manzanillo	Tlaxcala
38084160	Benito	Tlaxcala
38084170	Veracruz	Tlaxcala

SELECT Code, Name  
FROM A  
WHERE Dep = Tlaxcala

UNIVERSITY OF TWENTE

10




## TERMINOLOGY

### DATA STRUCTURING PRIMITIVES

---


- **Attribute:** properties that describe an entity
- **Tuple:** the entire row of attribute's values corresponding to a particular entity
- **Relation (or table):** a collection of tuples that are similarly shaped



12

5

6



# TERMINOLOGY


## ATTRIBUTE

### Attribute

- In the relational data model, we represent **real-world objects** by **tuples** stored in **relations** (*tables*).
- These real-world objects have particular **properties** that describe them. These properties are called **attributes**.

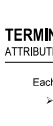
The relation *Productions* has these attributes: *cid*, *crop*, *annum*, *score*, *quality*

Productions				
cid	crop	annum	score	quality



UNIVERSITY OF TWENTE

13




## TERMINOLOGY

### ATTRIBUTE & DOMAIN

Each attribute has an associated domain:

- In the relation *Productions*:
  - ✓ *cid*: integer
  - ✓ *crop*: varchar(255)
  - ✓ *annum*: integer
  - ✓ *scores*: integer
  - ✓ *quality*: varchar(255)

15




# TERMINOLOGY

## DOMAIN

---


### Domain

- A set of acceptable values that an attribute is allowed to contain
- Original set of **atomic values** used to model data.
- By atomic value, we mean that each value in the domain is **indivisible** as far as the relational model is concerned; e.g. the domain of the attribute *Day* has the set of all possible days: {Mon, Tue, Wed...}.
- An attribute domain belongs to a certain **data type (also called domain)**
- There are many system-defined data types/domain: Numeric (Integer, Floating-point); Character (Char, Varchar); Temporal (Date, Time); Boolean (true/false, y/n)



UNIVERSITY OF TWENTE

14



# TERMINOLOGY

## NULL VALUES


Some attributes may have **missing values**. An attribute value may be:

- 'unknown'
- 'not applicable'

} In SQL we use **NULL** in both cases

### NULL (or Null)

- A special marker used in SQL to indicate that a data value does not exist in the DB.
- It enables the representation of missing and inapplicable information
- Not to be confused with a value of 0



UNIVERSITY OF TWENTE

16

### TERMINOLOGY

#### TUPLE

**Tuple**

- A tuple is a **record**
- In the relational data model, tuples represent **real-world objects/phenomena** that have certain attributes
- A tuple can be defined as a **list of attribute values**

One tuple from the relation *Production* is: (cid=3, crop="Rice\_paddy", annum=2000, score=300, quality="F")

cid	crop	annum	score	quality
3	Rice, paddy	2000	300	F

UNIVERSITY OF TWENTE.

17

### TERMINOLOGY

#### RELATION SCHEMA AND RELATION INSTANCE

**Relation schema**

- Basic information describing a table or relation:
  - Name of the relation;
  - List of attributes;
  - Domain of each attribute

**Relation instance**

- Set of tuples that adheres to all the requirements that are formulated by the relation schema
- The set of tuples in a relation at some point in time

UNIVERSITY OF TWENTE.

19

### TERMINOLOGY

#### RELATION (or TABLE)

**Relation (Table)**

- A collection of tuples that are similarly shaped: all tuples have the same attributes

The relation *Productions* in *FAOcrops.mdb*

Relation

Tuples

Attributes

UNIVERSITY OF TWENTE.

18

### TERMINOLOGY

#### RELATION SCHEMA AND RELATION INSTANCE

**Relation schema**

*Productions* (cid: integer, crop: varchar(255), annum: integer, score: integer, quality: varchar(5))

**Relation instance**

cid	crop	annum	score	quality
1	Maize	2000	120000	
1	Maize	2001	100000	
1	Maize	2002	200000	
1	Maize	2003	250000	
1	Maize	2004	400000	
1	Maize	2005	350000	
1	Maize	2006	250000	
1	Maize	2007	300000	
1	Maize	2008	200000	
1	Maize	2009	200000	sf
2	Potatoes	2000	220000	

on 18/09/2021 – 15:47 relation *Productions* had five attributes and 6930 tuples

UNIVERSITY OF TWENTE.

20

9

10

### TERMINOLOGY

#### RELATIONAL DATA MODEL

The relational data model is an integrated collection of:

- Data structuring primitives = **attributes, tuples, and relations**
- Rules of how to structure = **data definition language**,
- Mechanisms to handle the data in a database = **data manipulation language**.

UNIVERSITY OF TWENTE.

21

### TERMINOLOGY

#### RELATIONAL DATA MODEL

The relational data model is an integrated collection of:

- Data structuring primitives = **attributes, tuples, and relations**
- Rules of how to structure = **data definition language**,
- Mechanisms to handle the data in a database = **data manipulation language**.

UNIVERSITY OF TWENTE.

23

### TERMINOLOGY

#### DATA DEFINITION LANGUAGE

**Data Definition Language (DDL)**

- A (subset of a) computer language used to create and modify the **structure** of database objects in a database:
- Set of commands that can be used to define the database **schema**:  
CREATE, ALTER, DROP

```
CREATE TABLE Productions(
  cid single,
  crop varchar(255), annum integer,
  score integer, quality varchar(255)
)
```

cid	crop	annum	score	quality

Productions

UNIVERSITY OF TWENTE.

22

### TERMINOLOGY

#### DATA MANIPULATION LANGUAGE

**Data Manipulation Language (DML)**

- A (subset of a) computer language used for **adding (inserting), deleting, and modifying (updating)** data in a DB
- Set of commands that deals with the **manipulation of data**:  
SELECT, INSERT, UPDATE, DELETE

```
SELECT *
FROM Productions AS p, Countries AS c
WHERE c.ID=p.cid AND p.crop="Potatoes" AND
c.CHANGE="slovakia"
```

cid	crop	annum	score	quality
1	Maize	2000	120000	
1	Maize	2001	100000	
1	Maize	2002	200000	
1	Maize	2003	250000	
1	Maize	2004	400000	
1	Maize	2005	350000	
1	Maize	2006	250000	
1	Maize	2007	300000	
1	Maize	2008	200000	
1	Maize	2009	200000	sf
2	Potatoes	2000	220000	

UNIVERSITY OF TWENTE.

24

## TERMINOLOGY

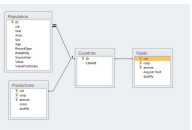
### RELATIONAL DATABASE

**Relational database**

- A DB based on the **relational data model**
- A collection of relations (*tables*) **structured to recognize relations** (*links/associations*) between stored items of information.

**Database schema**

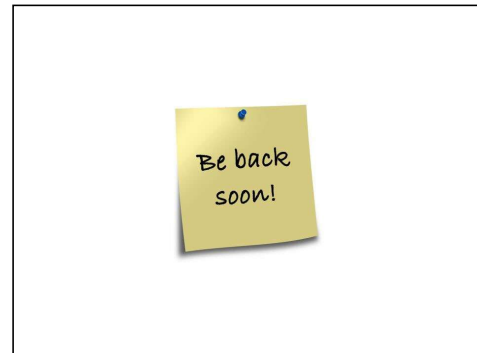
- The DB **structure** described in a formal language
- A **collection** of relation schemas and the **associations** amongst them
- The **skeleton structure** that represents the **logical view** of the entire database



The GDB *FAOcrops.mdb* contains 4 relations: (*Productions*, *Yields*, *Countries*, *Population*)

UNIVERSITY OF TWENTE

25



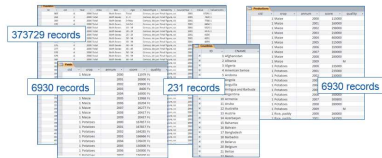
27

## TERMINOLOGY

### DATABASE INSTANCE

**Database instance**

- A **collection of relation instances**, one for each relation in the database schema



*FAOcrops.mdb* instance on 18,09,2021 at 15:47

UNIVERSITY OF TWENTE


26

## TERMINOLOGY

### INTEGRITY CONSTRAINTS

**Integrity constraints**

- A **set of rules** that are used to **maintain the quality of information** in a DB
- **Ensure** that the data manipulation is performed in such a way that **data integrity is not affected**.
- Used to **guard against accidental damage/errors** to/in the DB
- Integrity constraints are **specified** when the schema is defined
- Also used to **establish links between relations/tables**



UNIVERSITY OF TWENTE

28

13

14

## TERMINOLOGY

### INTEGRITY CONSTRAINTS

**Integrity constraints**

- There are various types of integrity constraints that a DB provides

We will discuss 3 types:

1. **Domain** integrity constraint
2. Entity integrity constraint: **Primary key**
3. Referential integrity constraint: **Foreign key**

UNIVERSITY OF TWENTE

29


## INTEGRITY CONSTRAINTS

### PRIMARY KEY

**Primary key**

- The primary key *K* of a relation *R* is one of (or a combination of) *R*'s attributes, such that:

1. It is **unique** – there are no two distinct tuples of *R* that have the same attribute value.
2. It is **minimal** – there is no proper subset of *K* that is unique.



UNIVERSITY OF TWENTE

31

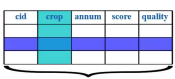
## INTEGRITY CONSTRAINTS

### DOMAIN CONSTRAINTS

**Domain integrity constraints**

- Refers to the **definition of a valid set of values for an attribute**
- In order to add a new value, it needs to meet the criteria defined for that attribute. Otherwise, we get an error

```
CREATE TABLE Productions(
    cid single,
    crop varchar(255), annum integer,
    score integer, quality varchar(255)
)
```



UNIVERSITY OF TWENTE

30


## INTEGRITY CONSTRAINTS

### PRIMARY KEY

Example – For the relation *Countries* in *FAOcrops.mdb* we can think of three candidate keys:

- a. *ID*
- b. *CNAME*
- c. *ID+CNAME*

Which one would you choose as the primary key? Why?



UNIVERSITY OF TWENTE

32

## INTEGRITY CONSTRAINTS

### PRIMARY KEY

**Golden rule of keys:**

- Check suitability carefully → Whether one (or a set) of the attributes can become a primary key or not can never be judged from looking at a relation instance.
- Each stored relation in a DB **must** have a defined primary key.
- Primary keys are conceived and defined (or even created!) when designing the database schema

UNIVERSITY OF TWENTE

33

## INTEGRITY CONSTRAINTS

### REFERENTIAL INTEGRITY

**Primary key**

**Foreign key**

The attribute *oid* is a foreign key in the table *Productions* pointing at the attribute *ID* of the table *Countries*. It only allows:

- Existing values in *ID*
- NULL values

UNIVERSITY OF TWENTE

35

## INTEGRITY CONSTRAINTS

### REFERENTIAL INTEGRITY

**Foreign key**

- A set of attributes that is used to **refer** to a tuple in another relation.
- It must correspond with a primary key value in the second relation.
- A foreign key behaves like a 'logical pointer'.
- Links between **primary** and **foreign** keys determine the relationships amongst tables in a DB

UNIVERSITY OF TWENTE

34

## DBMS AND GIS

A GIS is **not** a DBMS  
Main GIS packages **can be linked** to a DBMS to store and manage data

SQL Server 2008 R2, Oracle, MySQL, PostgreSQL, Informix, esri, PostgresSQL

UNIVERSITY OF TWENTE

36

17

18

24/09/2023

24/09/2023

## GIS AND DATABASES

### LINKING SPATIAL AND NON-SPATIAL DATA

**Spatial data**

**Non-spatial data**

Code	Name	City	Type	Sex	Men	Women	Total
0001701	San Felipe	Colombia	Urban		4733	6066	10799
0001702	San Felipe	Colombia	Urban		217	261	478
0001703	San Felipe	Colombia	Urban		463	481	944
0001704	San Felipe	Colombia	Urban		1267	1260	2527
0001705	San Felipe	Colombia	Urban		35	34	69
0001706	San Felipe	Colombia	Urban		456	471	927
0001707	San Felipe	Colombia	Urban		154	164	318
0001708	San Felipe	Colombia	Urban		245	241	486
0001709	San Felipe	Colombia	Urban		653	663	1316
0001710	San Felipe	Colombia	Urban		362	361	723
0001711	San Felipe	Colombia	Urban		161	164	325
0001712	San Felipe	Colombia	Urban		121	121	242
0001713	San Felipe	Colombia	Urban		75	84	159
0001714	San Felipe	Colombia	Urban		2203	2264	4467
0001715	San Felipe	Colombia	Urban		3013	2864	5877
0001716	San Felipe	Colombia	Urban		37	64	101

UNIVERSITY OF TWENTE

37

## GIS AND DATABASES

### LINKING SPATIAL AND NON-SPATIAL DATA

Spatial and non-spatial data in a GIS environment are stored separately (.shp, .dbf) but dynamically linked

Code	Name	City
0001701	San Felipe	Colombia
0001702	San Felipe	Colombia
0001703	San Felipe	Colombia
0001704	San Felipe	Colombia
0001705	San Felipe	Colombia
0001706	San Felipe	Colombia
0001707	San Felipe	Colombia
0001708	San Felipe	Colombia
0001709	San Felipe	Colombia
0001710	San Felipe	Colombia
0001711	San Felipe	Colombia
0001712	San Felipe	Colombia
0001713	San Felipe	Colombia
0001714	San Felipe	Colombia
0001715	San Felipe	Colombia
0001716	San Felipe	Colombia

UNIVERSITY OF TWENTE

39

## GIS AND DATABASES

### LINKING SPATIAL AND NON-SPATIAL DATA

GIS support both!  
This is what makes GIS a very powerful tool

UNIVERSITY OF TWENTE

38

## GIS AND DATABASES

### LINKING SPATIAL AND NON-SPATIAL DATA

1 Raster, Cell value  
2 Vector, Object ID

LandUseClass, Id, Description, Perc

LandUseClass	Id	Description	Perc
A	Primary forest	11.8	
B	Secondary vegetation	25.8	
C	Pasture	31.2	
D	Built-up area	26.5	
E	Rivers, lakes	4.1	

Parcel, Pld, Location, OwnerID

Parcel	Pld	Location	OwnerID
3421	2001	435	
5871	1462	550	
2109	2623	1040	
1515	2023	245	
3434	2020	488	
6371	1802	950	
2209	2542	1040	
1505	2020	145	

UNIVERSITY OF TWENTE

40

19

20

### SPATIAL DATABASES

Databases have evolved over the last 20 years towards supporting more complex data, such as **spatial data**

A **spatial database** allows users to store, query and manipulate collections of **spatial data by means of a new data type: geometry**

Parcel	Pld	Geometry	OwnerID
1421	MULTIPOLYGON((257462 704978333 464768 750851561 257463 86786 ...))	435	
1519	MULTIPOLYGON((257462 815525058 464768 815525058 257463 86786 ...))	220	
2105	MULTIPOLYGON((257765 714911912 464768 838972187 257762 58794 ...))	1045	
1515	MULTIPOLYGON((257765 672100448 464807 13792555 257765 838978 ...))	245	
1545	MULTIPOLYGON((257432 672100448 464807 13792555 257432 838978 ...))	475	
1534	MULTIPOLYGON((257432 476097654 464819 84850292 257433 147915 ...))	650	
2209	MULTIPOLYGON((257444 88627332 464828 555046319 257446 43201 ...))	1840	
1555	MULTIPOLYGON((257559 750107457 464935 203946059 257561 ...))	145	

UNIVERSITY OF TWENTE

41

### SPATIAL DATABASE MANAGEMENT SYSTEM (SDBMS)

A S(patial)DBMS is a software package that:

- offers **all benefits of a DBMS** for data storage/management
- supports the **storage of spatial data models** – e.g. **point, line, or polygon**,
- can **manage coordinate systems** and transformations
- extends querying** and manipulating capabilities of traditional SQL by adding special commands **aimed at spatial data** – e.g. **SELECT p.\* FROM parcels AS p WHERE Area(p.geom)>1000**
- provide storage of the relationships between features, including the **creation and storage of topological relationships**.

UNIVERSITY OF TWENTE

42

### GIS & SDBMS

A GOOD TEAM

**GIS:**

- Built-in 'understanding' of geographic space,
- Functions for spatial analysis of (almost) any kind, and
- Equipment for efficient map production
- BUT** lack of fully developed query language to operate on tabular data.

**DBMS**

- Specifically designed for handling attribute data (i.e., administrative, non-spatial, tabular, thematic)
- Long tradition in multi-user concurrent management of large amounts of data
- BUT** not support provided for spatial data storage and representation

UNIVERSITY OF TWENTE

43

### GIS & SDBMS

A GOOD TEAM

**SDBMS:**

- Focuses on storage, querying and sharing **large spatial data sets**,
- Understand geometry and topology
- Provide **fully-fledged querying language** (i.e. SQL)
- Spatial querying functions
- Basic visualization

PostgreSQL

UNIVERSITY OF TWENTE

44

21

22

### GIS & SDBMS

SAME DATA DIFFERENT VIEWS

GIS environment  
e.g. ArcMap

SDBMS environment  
e.g. Postgres + PostGIS

UNIVERSITY OF TWENTE

45

### GIS & SDBMS

SAME DATA DIFFERENT VIEWS

GIS environment  
e.g. ArcMap

SDBMS environment  
e.g. Postgres + PostGIS

UNIVERSITY OF TWENTE

47

### GIS & SDBMS

SAME DATA DIFFERENT VIEWS

DBMS environment - e.g. MS Access

UNIVERSITY OF TWENTE

46

### TERMINOLOGY

UNIVERSE OF DISCOURSE

We always aim at representing **only a part of the real world**.

**Universe of discourse**

- A part of the real world that is of interest.

We use data models for representing the universe of discourse and storing that representation in a database.

A.K.A. = the **database miniworld**

*It must be well understood by the designers of the DB in order to be able to properly represent it!*

UNIVERSITY OF TWENTE

48



### TERMINOLOGY

UNIVERSE OF DISCOURSE – EXAMPLE

Example:

- We are interested in understanding *if and how crop production relates to population size and structure in different countries of the world.*

Part of the real world that constitutes the *universe of discourse*:

- Crop types
- Crop production
- Population (size and structure)
- Country

UNIVERSITY OF TWENTE.

49

### DATABASE CREATION

STAGES

Relational database creation involves 4 stages:

- 1. Data investigation**
  - Type, quantity and quality of the data to be included in the DB
  - Entities and attributes are decided
- 2. Data modelling**
  - Create a conceptual model of the data based on relationships between entities
- 3. Database design**
  - Create a practical design for the DB
- 4. Database implementation**
  - Create DB structure and populate it

UNIVERSITY OF TWENTE.

51

### DATABASE CREATION

PROCESS OF DATABASE DESIGN

Relational database design involves:

- Deciding which **relations** will be present in the database, and
- Defining **their schemas**, which in turn implies:
  - Deciding which **attributes** relations will have, and
  - What are the **domains** of these attributes.

Database design also includes:

- Defining **relationships** between relations by **integrity constraints**
- Defining desirable **queries**

UNIVERSITY OF TWENTE.

50

### DATA MODELLING

STAGES

Data modelling or entity relationship modelling involves 4 steps:

1. Identification of entities
2. Identification of relationships
  - One to one
  - One to many
  - Many to many
3. Identification of attributes of entities
4. Design of tables

The process should be led by:

- Purpose of the DB
- Questions that the DB should be able to answer

UNIVERSITY OF TWENTE.

52

25

26

### DATA MODELLING

RELATIONSHIPS

Relationships:

- many visitors stay at one hotel (M:1);
- one travel company organizes holidays for many visitors (1:M);
- one ski school teachers many visitors (1:M); and
- several different travel companies may use more than one ski school (M:N).

UNIVERSITY OF TWENTE.

53

### Thanks for joining!

Coming up

Practical: Today 11:00 – 15:30

Wrap up session: Tomorrow 16:00 – 17:30

Questions?

Discussions > Lesson 9: Data Management

UNIVERSITY OF TWENTE.

55

### DATA MODELLING

TABLES

Hotel ID	Name	Other attributes...
001	Mountain View	...
002	Palace Deluxe	...
003	Ski Lodge	...

Travel Co. ID	Travel Co. Name	Other attributes...
T01	Sea Tours	...
T02	Green Shores	...

Visitor ID	Visitor Name	Hotel ID	Travel Co. ID	Ski School ID	Other attributes...
V001	Smith J.	002	T01	002	...
V002	Schmidt S.	001	T02	003	...

Ski School ID	Ski School Name	Other attributes...
S01	Snow Fun	...
S02	Blue's Ski School	...
S03	Ski School Professional	...

UNIVERSITY OF TWENTE.

54

27

28