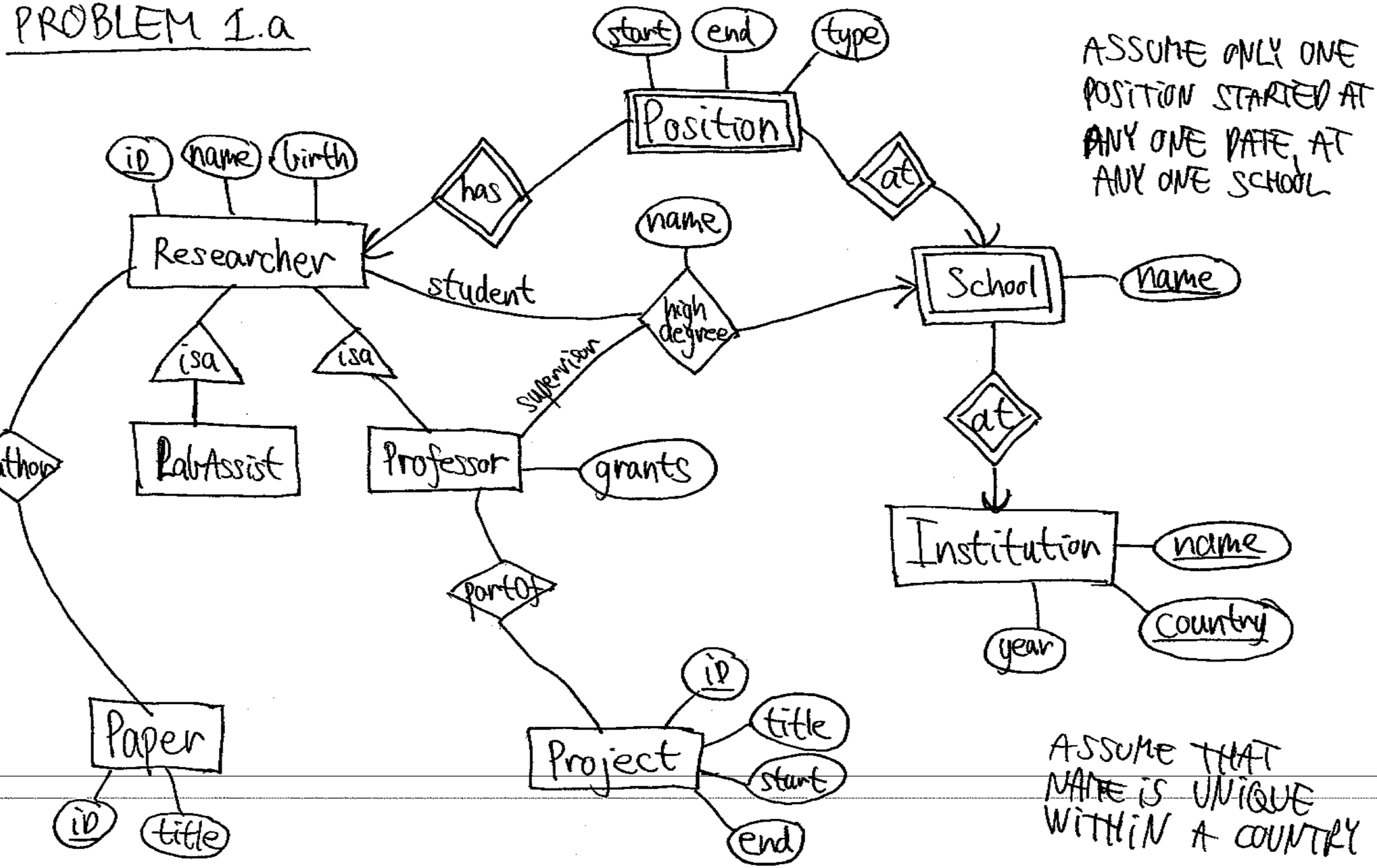
Case Study: Lufthansa Cargo Database

Carsten Schürmann

Today's lecture

- More on data modelling
- Introduction to Lufthansa Cargo Database
- Entity Relationship diagram
- Boyce-Codd normal form

From Lecture 2: Data Modeling



From Lecture 3: Normalization

- **So far:** Start from scratch mentatility

- **ER diagram**

  `Movies(title,year,length,filmType,studioName,starName)`

  where the length of a movie is repeated several times (once for each `starName`).

- Obvious problem:
  Uses more memory than is necessary.

From Lecture 3: Normalization

- Principled approach to avoiding (or at least being aware of) anomalies in a database design.
- Captures situations where unrelated facts are placed in a single relation.
- Decompose (split) to avoid anomalies:

```
Movies(title,year,length,filmType,studioName,starName)
```

becomes

```
Movies1(title,year,length,filmType,studioName)
Movies2(title,year,starName)
```

Recall: Keys

- A *candidate key* for a relation is a set K of its attributes that satisfy:
  - **Uniqueness:** The values of the attribute(s) in K uniquely identify a tuple.
  - **Minimality:** The uniqueness property goes away if we remove any attribute from K.
- If uniqueness is satisfied the attributes are said to form a *superkey*.
- **Example**: For `Movies`,
  - `{Title,year,starName}` is a candidate key.
  - `{Title,year,starName,length}` is a superkey.
  - `{Title, year}` is not a key.

Recall: Functional dependency (FD)

- We say that A (functionally) determines B, written A→B, if the value of B is ***always*** determined by the value(s) of A (for ***any*** possible relation).
- **Examples**:
  - `cpr → name` in `Person(cpr,name)`
  - `title year → length` in `Movie`
- **Non-example**:
  - `title year → starName` does not hold for `Movie`

Normalization

- First normal form
  - All data value are atomic
- Second normal form
  - 1NF
  - No non-key attribute is partially dependend on a candidate key
- Third normal form
  - 2NF
  - No non-key attribute depends transitively on a candidate key

Recall: Boyce-Codd Normal Form (BCNF)

- A relation R is in BCNF:

  $A_1A_2A_3\ldots A_n$ -> B is a FD

  iff $\{A_1, A_2, A_3, \ldots, A_n\}$ is a superkey of R
- **Example**: `Movies` has the FD

  `title year → length`

  where `{title,year}` is not a superkey.
  - This means that `Movies` is not in BCNF.
- The anomalies we saw in `Movies` are in fact *caused* by the above FD!
  - requires us to store the same movie length again and again.

Recall: Decomposing into BCNF

- Suppose relation R is not in BCNF. Then there is an FD $A_1A_2...A_n \rightarrow B_1B_2...B_m$ that is not *unavoidable*.
- To eliminate the FD we split R into two relations:
  - R1 with all attributes of R except $B_1B_2...B_m$.
  - R2 with attributes $A_1A_2...A_n \rightarrow B_1B_2...B_m$. Note that $A_1A_2...A_n$ is a superkey of R2, so a join recovers the original relation R.
- This process is repeated until all relations are in BCNF.

BCNF (Summary)

- A relation is in Boyce-Codd normal form if for every FD $A \rightarrow B$ either
  - B is contained in A (the FD is trivial), or
  - A contains a candidate key of the relation,
- In other words: every determinant in a non-trivial dependency is a (super) key
- The same as 3NF except in 3NF we only worry about non-key Bs
- If there is only one candidate key then 3NF and BCNF are the same

Example

Decomposing Courses

- Schema is Courses(Number, DepartmentName, CourseName, Classroom, Enrollment, StudentName, Address).
- BCNF-violating FD is
  Number DepartmentName → CourseName Classroom Enrollment.
  - What is $\{$Number, DepartmentName$\}^+$?

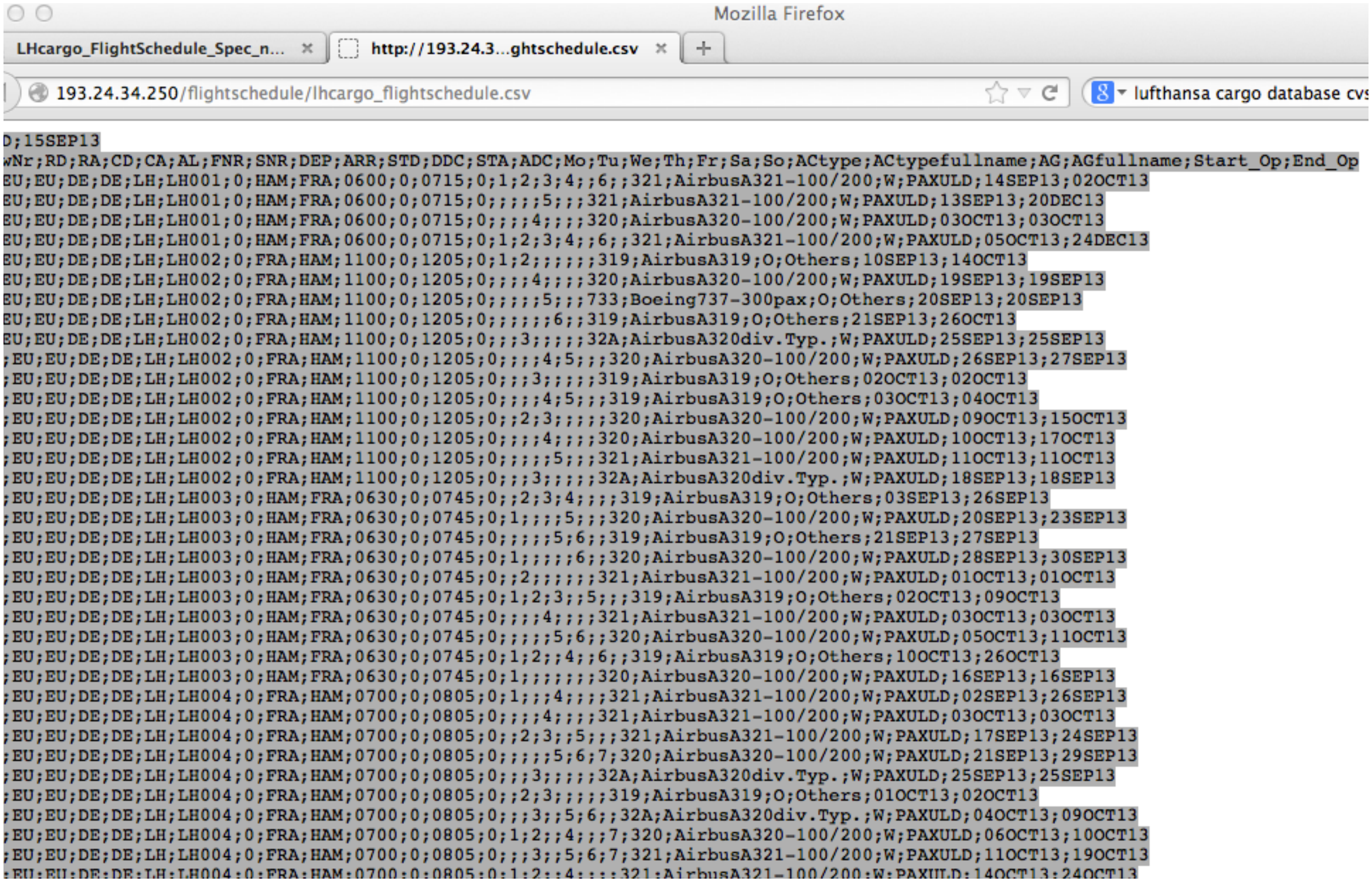
{Number, DepartmentName, Coursename, Classroom, Enrollment

- Decompose Courses into
  Courses1(Number, DepartmentName, CourseName, Classroom Enrollment) and
  Courses2(Number, DepartmentName, StudentName, Address).

[Zaki Malik'08]

Case Study

Raw data as CVS File




```
D;15SEP13
wNr;RD;RA;CD;CA;AL;FNR;SNR;DEP;ARR;STD;DDC;STA;ADC;Mo;Tu;We;Th;Fr;Sa;So;ACtype;ACtypefullname;AG;AGfullname;Start_Op;End_Op
EU;EU;DE;DE;LH;LH001;0;HAM;FRA;0600;0;0715;0;1;2;3;4;;6;;321;AirbusA321-100/200;W;PAXULD;14SEP13;02OCT13
EU;EU;DE;DE;LH;LH001;0;HAM;FRA;0600;0;0715;0;;;;5;;;321;AirbusA321-100/200;W;PAXULD;13SEP13;20DEC13
EU;EU;DE;DE;LH;LH001;0;HAM;FRA;0600;0;0715;0;;;4;;;;320;AirbusA320-100/200;W;PAXULD;03OCT13;03OCT13
EU;EU;DE;DE;LH;LH001;0;HAM;FRA;0600;0;0715;0;1;2;3;4;;6;;321;AirbusA321-100/200;W;PAXULD;05OCT13;24DEC13
EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;1;2;;;;;319;AirbusA319;O;Others;10SEP13;14OCT13
EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;4;;;;320;AirbusA320-100/200;W;PAXULD;19SEP13;19SEP13
EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;;5;;;733;Boeing737-300pax;O;Others;20SEP13;20SEP13
EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;;;6;;319;AirbusA319;O;Others;21SEP13;26OCT13
EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;3;;;;;32A;AirbusA320div.Typ.;W;PAXULD;25SEP13;25SEP13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;4;5;;;320;AirbusA320-100/200;W;PAXULD;26SEP13;27SEP13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;3;;;;;319;AirbusA319;O;Others;02OCT13;02OCT13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;4;5;;;319;AirbusA319;O;Others;03OCT13;04OCT13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;2;3;;;;;320;AirbusA320-100/200;W;PAXULD;09OCT13;15OCT13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;4;;;;320;AirbusA320-100/200;W;PAXULD;10OCT13;17OCT13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;;;5;;;321;AirbusA321-100/200;W;PAXULD;11OCT13;11OCT13
;EU;EU;DE;DE;LH;LH002;0;FRA;HAM;1100;0;1205;0;;3;;;;;32A;AirbusA320div.Typ.;W;PAXULD;18SEP13;18SEP13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;;2;3;4;;;;319;AirbusA319;O;Others;03SEP13;26SEP13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;1;;;;5;;;320;AirbusA320-100/200;W;PAXULD;20SEP13;23SEP13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;;;;5;6;;319;AirbusA319;O;Others;21SEP13;27SEP13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;1;;;;;6;;320;AirbusA320-100/200;W;PAXULD;28SEP13;30SEP13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;;2;;;;;;321;AirbusA321-100/200;W;PAXULD;01OCT13;01OCT13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;1;2;3;;5;;;319;AirbusA319;O;Others;02OCT13;09OCT13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;;;;4;;;;321;AirbusA321-100/200;W;PAXULD;03OCT13;03OCT13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;;;;;5;6;;320;AirbusA320-100/200;W;PAXULD;05OCT13;11OCT13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;1;2;;4;;6;;319;AirbusA319;O;Others;10OCT13;26OCT13
;EU;EU;DE;DE;LH;LH003;0;HAM;FRA;0630;0;0745;0;1;;;;;;;320;AirbusA320-100/200;W;PAXULD;16SEP13;16SEP13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;1;;;4;;;;321;AirbusA321-100/200;W;PAXULD;02SEP13;26SEP13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;;;4;;;;321;AirbusA321-100/200;W;PAXULD;03OCT13;03OCT13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;2;3;;5;;;321;AirbusA321-100/200;W;PAXULD;17SEP13;24SEP13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;;;;5;6;7;320;AirbusA320-100/200;W;PAXULD;21SEP13;29SEP13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;3;;;;;32A;AirbusA320div.Typ.;W;PAXULD;25SEP13;25SEP13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;2;3;;;;;319;AirbusA319;O;Others;01OCT13;02OCT13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;;3;;5;6;;32A;AirbusA320div.Typ.;W;PAXULD;04OCT13;09OCT13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;1;2;;4;;;7;320;AirbusA320-100/200;W;PAXULD;06OCT13;10OCT13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;;;3;;5;6;7;321;AirbusA321-100/200;W;PAXULD;11OCT13;19OCT13
;EU;EU;DE;DE;LH;LH004;0;FRA;HAM;0700;0;0805;0;1;2;;4;;;;321;AirbusA321-100/200;W;PAXULD;14OCT13;24OCT13
```

Guidelines

Understand Possible Queries!

- Can I ship my goods every Thursday?
- Is it possible to ship twice a week?
- Is enough capacity available?
- Can my goods reach a destination directly?
- How many flights go to Asia?

Understand Data!

lufthansa-cargo.com/fileadmin/user_upload/corporate/pdf/01_eSe

lufthansa cargo dat

Page: 1 of 5 Automatic Zoom

Example of the Data Record:

Ref.	***Record Content***
Line 1	SCD;01MAR05
Line 2	RowNr;RD;RA;CD;CA;AL;FNR;SNR;DEP;ARR;STD;DDC;STA;ADC;Mo;Tu;We;Th;Fr;Sa;So;ACtype;ACtypefullname;AG;AGfullname;Start_Op;End_Op
Line 3	1;NA;DE;US;DE;LH;LH445;0;ATL;FRA;1610;0;0630;1;1;2;3;4;5;6;7;34Z;Airbus A330-400;W;PAXwidebody;25NOV04;31MAR05
Line 4	END#

Ref. No.	***Usage***	***Data Element***	***Character Format***	***Note***
1	**M**	**Schedule File Header**		
1.1	M	Schedule File Header Identifier	aaa	SCD
1.2	M	Separator	Semicolon	
1.3	M	Date of File Creation		
1.3.1	M	Day of the month	nn	e.g.: 01
1.3.2	M	Month	aaa	e.g.: MAR (IATA data element 201)
1.3.3	M	Year	nn	e.g.: 05

Ref. No.	***Usage***	***Data Element***	***Character Format***	***Note***
2	**M**	**Flight Schedule Information Header**		
2.1	M	Header Row Number	RowNr;	
2.2	M	Header Region Code of Departure Airport	RD;	
2.3	M	Header Region Code of Arrival Airport	RA;	
2.4	M	Header Country Code of DepartureAirport	CD;	
2.5	M	Header Country Code of Arrival Airport	CA;	
2.6	M	Header Airline Code	AL;	
2.7	M	Header Flight Number	FNR;	
2.8	M	Header Segment Number	SNR;	

Ref. No.	*Usage*	*Data Element*	*Character Format*	*Note*
2	**M**	**Flight Schedule Information Header**		
2.1	M	Header Row Number	RowNr;	
2.2	M	Header Region Code of Departure Airport	RD;	
2.3	M	Header Region Code of Arrival Airport	RA;	
2.4	M	Header Country Code of DepartureAirport	CD;	
2.5	M	Header Country Code of Arrival Airport	CA;	
2.6	M	Header Airline Code	AL;	
2.7	M	Header Flight Number	FNR;	
2.8	M	Header Segment Number	SNR;	
2.9	M	Header Airport Code of Departure	DEP;	
2.10	M	Header Airport Code of Arrival	ARR;	
2.11	M	Header Schedule Time of Departure	STD;	
2.12	M	Header Date Variation of Departure	DDC;	
2.13	M	Header Schedule Time of Arrival	STA;	
2.14	M	Header Date Variation of Arrival	ADC;	
2.15	M	Header Operating Days	Mo;Tu;We;Th;Fr;Sa; So;	
2.16	M	Header Aicraft Type Code	ACtype;	
2.17		Header Aircraft Type Full Name	ACtypefullname;	
2.18	M	Header Aircraft Group Code	AG;	
2.19	M	Header Aircraft Group Full Name	AGfullname;	
2.20	M	Header Start Date of Flight Operations	Start_Op;	
2.21	M	Header End Date of Flight Operations	End_Op	

Derive Functional Dependencies

FNR, SNR -> DEP

FNR, SNR -> ARR

FNR, SN

RD-> CD

RA -> CA

DEP -> RD

ARR -> RA

…

Group work 10 minutes!

BCNF normalization

- Take the big table and decompose!
- Formulate each decomposition
- Group work 10 minutes
- Draw ER diagram

History Preservation

– Databases are persistant and have state

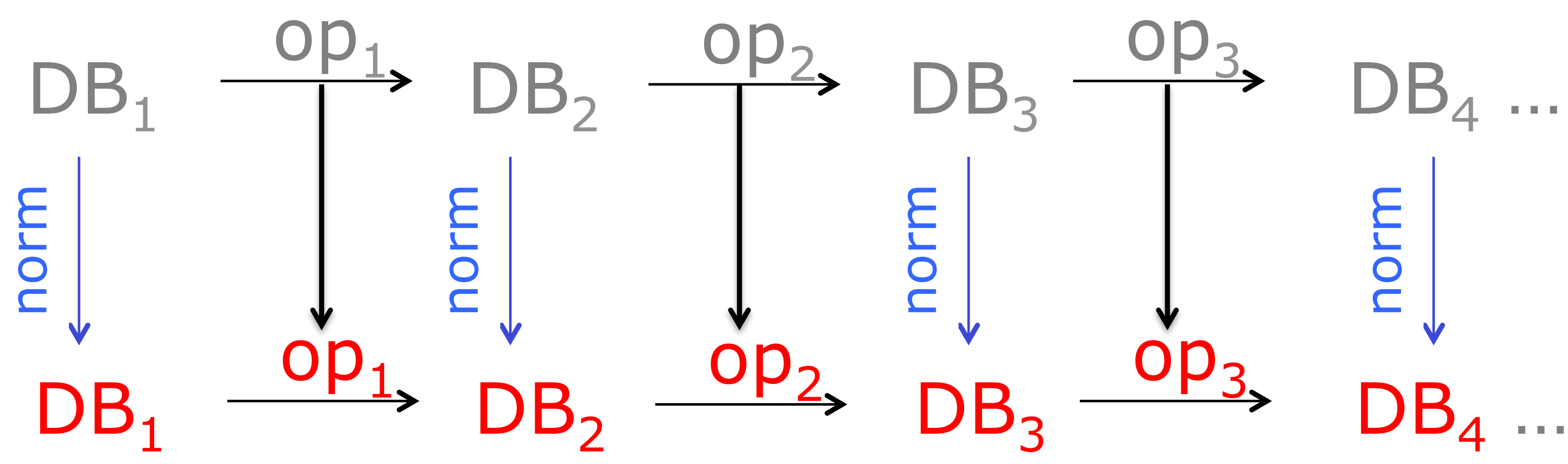
DB_n

– Oerations (such as search, update, delete)

$$\xrightarrow{op_1}$$

History Preservation

- Standard evolution of the database



- Simulated evolution of the normalized database

Semantic Transparency

- Very difficult to get right
  - Simple case:
    - Insertion of one record in the original DB
    - May translate into several insertions
  - Complex case:
    - Complex transaction exist on the original DB
    - How are these transactions translated?

- Yet necessary
  - Normalization should be semantically transparent

Course goal

After the course the students should be able to:

- Use tools so far to transform a live database into a new one, while reducing redundancy and preserving history

Next steps...

- Exercises from 10:00 as usual.
- Will start by a TA presentation of some exercises from last week (<15 min.)
- Work on the first hand-in
- Next week: SQL