Integrity Constraints in the Relational Model

**Integrity constraints** are logical statements that restrict the set of allowable relations in a database.

**Example.**
- database schema, \( R = \{ \text{EMP, DEPT} \} \), with
- \( \text{schema}(\text{EMP}) = \{ \text{ENAME, DNAME, ADDRESS, POSTCODE, LOC} \} \)
- \( \text{schema}(\text{DEPT}) = \{ \text{DNAME, MNAME, NO_EMPS, LOC} \} \)
- database, \( d = \{ r_1, r_2 \} \) over \( R \),
- \( r_1 \) is a relation over \( \text{schema}(\text{EMP}) \), and
- \( r_2 \) is a relation over \( \text{schema}(\text{DEPT}) \).

Keys and Functional Dependencies (FDs)

- Stating that \( \text{ENAME} \) is a key of \( \text{EMP} \), means that no two distinct tuples in \( r_1 \) have the same \( \text{ENAME} \).
- Stating that \( \text{DNAME} \) is a key of \( \text{DEPT} \), means that no two distinct tuples in \( r_2 \) have the same \( \text{DNAME} \).
- Keys are special cases of Functional Dependencies (FDs).
- An example of an FD which is not the result of a key, is the constraint that an \( \text{ADDRESS} \) has a unique \( \text{POSTCODE} \).

Inclusion Dependencies (INDs)

- Stating that \( \text{DNAME} \) in \( \text{EMP} \) is a foreign key referencing the key \( \text{DNAME} \) in \( \text{DEPT} \), means that whenever there is a tuple in \( r_1 \) with a nonnull \( \text{DNAME} \)-value, say Sales, then there is a corresponding tuple in \( r_2 \) whose \( \text{DNAME} \)-value is also Sales.
- Foreign keys are special cases of Inclusion Dependencies (INDs).
- An example of an IND which is not the result of a foreign key is the constraint that the \( \text{LOC} \)ation an employee works in is included in the \( \text{LOC} \)ations of the departments.

An example database

Relation schemas \( R \) and \( S \), and relations \( r \) and \( s \)

\( R \) is a relation schema, with \( \text{schema}(R) = \{ \text{ENAME, DNAME, MNAME} \} \)

Relation \( r \) over \( R \) is given by

\[
\begin{array}{ccc}
\text{ENAME} & \text{DNAME} & \text{MNAME} \\
\text{Mark} & \text{Computing} & \text{Steve} \\
\text{Angela} & \text{Computing} & \text{Steve} \\
\text{Graham} & \text{Computing} & \text{Steve} \\
\text{Paul} & \text{Math} & \text{Donald} \\
\text{George} & \text{Math} & \text{Donald} \\
\end{array}
\]
Definition. An FD over $R$ is a statement of the form $R : X \rightarrow Y$ (or simply $X \rightarrow Y$) where $X$ and $Y$ are subsets of schema$(R)$.

We say that $X$ functionally determines $Y$.

- $R : \{ENAME\} \rightarrow \{DNAME, MNAME\}$, each employee works in one department and has one manager.
- $S : \{ENAME\} \rightarrow \{SAL\}$, each employee has one salary.

Definition. An FD $X \rightarrow Y$ is satisfied in a relation $r$, if whenever two rows in $r$ have the same $X$-value they also have the same $Y$-value.

Alternative definition. An FD $X \rightarrow Y$ is satisfied in a relation $r$, if for each $X$-value of $r$ there is at most one $Y$-value.

$NAME \rightarrow AGE$ is satisfied in the following relation:

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>CHILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack</td>
<td>20</td>
<td>John</td>
</tr>
<tr>
<td>Jack</td>
<td>20</td>
<td>Jane</td>
</tr>
</tbody>
</table>

$NAME \rightarrow AGE$ is violated in the following relation:

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>CHILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack</td>
<td>20</td>
<td>John</td>
</tr>
<tr>
<td>Jack</td>
<td>30</td>
<td>Jane</td>
</tr>
</tbody>
</table>
An example of one FD satisfied and the other violated.  

NAME $\rightarrow$ AGE is satisfied in the following relation, but AGE $\rightarrow$ CHILD and NAME $\rightarrow$ CHILD are violated:

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>CHILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack</td>
<td>20</td>
<td>John</td>
</tr>
<tr>
<td>Jack</td>
<td>20</td>
<td>Jane</td>
</tr>
</tbody>
</table>

Definition. An FD $X \rightarrow Y$ holds on relation schema $R$ if every allowable (legal) relation $r$ over $R$ satisfies $X \rightarrow Y$.

Definition. A set $F$ of FDs holds on relation schema $R$ if every allowable (legal) relation $r$ over $R$ satisfies each FD in $F$.

Here "allowable" and "legal" mean that the relations would correctly model some part of the real world.

### Problem for you to work on

Consider the following relation about films:

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
<th>Genre</th>
<th>StarName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>SciFi</td>
<td>Carrie Fisher</td>
</tr>
<tr>
<td>Star Wars</td>
<td>1977</td>
<td>SciFi</td>
<td>Harrison Ford</td>
</tr>
<tr>
<td>Raiders ...</td>
<td>1981</td>
<td>Action</td>
<td>Harrison Ford</td>
</tr>
<tr>
<td>Raiders ...</td>
<td>1981</td>
<td>Adventure</td>
<td>Harrison Ford</td>
</tr>
<tr>
<td>When Harry ...</td>
<td>1989</td>
<td>Comedy</td>
<td>Carrie Fisher</td>
</tr>
</tbody>
</table>

1. What FDs are satisfied by the above relation?
2. What FDs are violated by the above relation?
3. What FDs would you expect to hold on the above relation schema?

### Keys and Superkeys

Definition. A set of attributes $X$ contained in schema$(R)$ is a superkey for a relation schema $R$ if $X \rightarrow$ schema$(R)$ holds on $R$.

Definition. A set of attributes $X$ contained in schema$(R)$ is a key for $R$ if

1. $X$ is a superkey for $R$, and
2. for no proper subset $Y$ of $X$, is $Y$ a superkey for $R$.

* What are the superkeys and keys for our earlier $R$ and $S$ about employees?
Let $schema(R) = SPJ$ (we often abbreviate the set notation $\{S, P, J\}$):

- $S$ stands for student,
- $J$ stands for subject and
- $P$ stands for position.

Let $F$ be the following set of FDs over $R$:

- $SJ \rightarrow P$, i.e. every student has one position in each subject.
- $PJ \rightarrow S$, i.e. every position has one student in each subject.

What are the keys of relations over $schema(R)$ that satisfy $F$?

$F$ is the following a set of FDs over $U$:

- $C \rightarrow T$, i.e. a course has one teacher.
- $HR \rightarrow C$, i.e. a room can only only have one course at any time.
- $HT \rightarrow R$, i.e. a teacher can only be in one room at any time.
- $CS \rightarrow G$, i.e. a student has one grade per course.
- $HS \rightarrow R$, i.e. a student can only be in one room at any time.

What are the keys for this example?

A comprehensive example.

Let $U$ be a relation schema with $schema(U) = CTHRSG$:

- $C$ stands for a course,
- $T$ stands for a teacher,
- $H$ stands for hour,
- $R$ stands for room,
- $S$ stands for student and
- $G$ stands for grade.

Closure of a set of attributes

The closure of set of attributes $X$ with respect to $F$, denoted by $CLOSURE(X, F)$, is the set $Y$ of all attributes such that $X$ functionally determines $Y$.

The closure of $X$ is computed by the following algorithm:

1. $CI := X$;
2. Done := false;
3. while not Done do
4.   Done := true;
5.   for each $W \rightarrow Z$ in $F$ do
6.     if $W$ is a subset of $CI$ and $Z$ is not a subset of $CI$ then
7.       $CI := CI \cup Z$;
8.     Done := false;
9.   end if
10. end for
11. end while
12. return $CI$;
For our example we have:

- \( \text{CLOSURE}(C, F) = CT \)
- \( \text{CLOSURE}(HR, F) = HRCT \)
- \( \text{CLOSURE}(RSG, F) = RSG \)
- \( \text{CLOSURE}(HRSG, F) = HRSGCT \)
- ...

**Superkeys**

**Alternative Definition.** A set of attributes \( X \) contained in \( \text{schema}(R) \) is a superkey for \( R \) with respect to \( F \) if \( \text{CLOSURE}(X, F) = \text{schema}(R) \).  

* What are the keys of our comprehensive example?

**Problem for you to work on**

Consider relation schema \( R \) where \( \text{schema}(R) = ABCD \).

Let the set \( F \) of FDs which hold on \( R \) be \( \{AB \rightarrow C, C \rightarrow D, D \rightarrow A\} \).

1. Compute
   1.1 \( \text{CLOSURE}(A, F) \)
   1.2 \( \text{CLOSURE}(B, F) \)
   1.3 \( \text{CLOSURE}(C, F) \)
   1.4 \( \text{CLOSURE}(D, F) \).
2. What are all the keys of \( R \)?
3. What are all the superkeys for \( R \) which are not keys?

**Problem for you to work on**

Let \( R \) be a relation schema with \( \text{schema}(R) = \{A_1, A_2, A_3, B_1, B_2, B_3, C\} \).

Let \( F = \{A_1 \rightarrow B_1, A_2 \rightarrow B_2, A_3 \rightarrow B_3, B_1 \rightarrow A_1, B_2 \rightarrow A_2, B_3 \rightarrow A_3, \{B_1, B_2, B_3\} \rightarrow C\} \).

How many keys does \( R \) have with respect to \( F \) and what are they?