SWEN-220
Mathematical Models of Software

Normalization
Well Structured Relations

• Contain minimal *redundancy* (data duplication).

• Allow users to insert, delete and modify data without *anomalies* – errors or inconsistencies that result when updating a relation that contains redundant data.
  
  – Insertion Anomaly
  – Deletion Anomaly
  – Modification Anomaly
Maintaining Consistency

Consider the following relation:

$$\text{SIS}(\text{dept, num, sec, ms, days, time, room, cap, inst})$$

And this example table:

**SIS**

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

$\text{ms} = \text{major (M), service (S) or both (MS)}$

$\text{cap} = \text{capacity}$

$\text{inst} = \text{instructor}$
Consider the following relation:

\[ \text{SIS}(\text{dept}, \text{num}, \text{sec}, \text{ms}, \text{days}, \text{time}, \text{room}, \text{cap}, \text{inst}) \]

And this example table:

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

ms = major (M), service (S) or both (MS)
cap = capacity
inst = instructor

What's wrong with this picture?
Maintaining Consistency

Consider the following relation:

\[
\text{SIS(dept, num, sec, ms, days, time, room, cap, inst)}
\]

And this example table:

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

What's wrong with this picture?

**REDUNDANCY!**

Multiple copies of:

- major / service (ms)
- room capacity

Think about changing:

- capacity of room 1550
- 220 from M to MS

Violates the DRY principle
(Don’t Repeat Yourself)

Let’s see why.

\[ ms = \text{major (M), service (S) or both (MS)} \]
\[ cap = \text{capacity} \]
\[ inst = \text{instructor} \]
Maintaining Consistency

SIS(dept, num, sec, ms, days, time, room, cap, inst)

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

ms = major (M), service (S) or both (MS)
cap = capacity
inst = instructor

What is/are candidate key(s)?
Maintaining Consistency

SIS$(\text{dept, num, sec, ms, days, time, room, cap, inst})$

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

ms = major (M), service (S) or both (MS)
cap = capacity
inst = instructor

What is/are candidate key(s)?
dept + num + sec
Maintaining Consistency

SIS(dept, num, sec, ms, days, time, room, cap, inst)

What is/are candidate key(s)?

dept + num + sec

Do we need sec to determine major/service?

Do we need dept + num + sec to determine room capacity?

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

ms = major (M), service (S) or both (MS)
cap = capacity
inst = instructor
Maintaining Consistency

SIS(\text{dept, num, sec, ms, days, time, room, cap, inst})

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

\(\text{ms} = \text{major (M), service (S) or both (MS)}\)
\(\text{cap} = \text{capacity}\)
\(\text{inst} = \text{instructor}\)

What is/are candidate key(s)?

\text{dept + num + sec}

Do we need \text{sec} to determine \text{major/service}?  
Do we need \text{dept + num + sec} to determine \text{room capacity}?  

That's the source of our problem.  
Our relation is \textit{unnormalized}.  

Maintaining Consistency

What is/are candidate key(s)?

depth + num + sec

Do we need sec to determine major/service?

Do we need dept + num + sec to determine room capacity?

That's the source of our problem.

Our relation is unnormalized.

How to we fix this?

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>ms</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>M</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>M</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>S</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>S</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>S</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

ms = major (M), service (S) or both (MS)
cap = capacity
inst = instructor
Normalization

• Normalization is a set of transformations whose goal is reducing data duplication in relations.

• Doing so reduces the risk of insert, update & delete anomalies.

• Normalization introduces new relations to "factor out" duplicates.

• Our focus: The first three normalizations (normal forms).
  – 1NF, 2NF, 3NF
  – A relation normalized at level N is also normalized at all levels less than N.
  – There are levels beyond 3NF, but they are:
    • Rarely used in practice.
    • Address unusual and esoteric redundancies.

• To normalize, we have to know about functional dependencies.
Functional Dependency

Consider a relation \( R( a_1, a_2, \ldots a_N ) \)

Assume that if we know attributes \( a_i, a_j \) then we also know \( a_k, a_m, \) and \( a_n \) uniquely.

Then \( a_k, a_m, \) and \( a_n \) are *functionally dependent* on \( a_i \) and \( a_j \), written:

\[
a_i, a_j \rightarrow a_k, a_m, a_n
\]
Using Our Example

\[ SIS(\text{dept}, \text{num}, \text{sec}, \text{ms}, \text{days}, \text{time}, \text{room}, \text{cap}, \text{inst}) \]

By definition, *non-candidate key attributes are dependent on any candidate key:*
Using Our Example

\[
\text{SIS(\text{dept}, \text{num}, \text{sec}, \text{ms}, \text{days}, \text{time}, \text{room}, \text{cap}, \text{inst})}
\]

By definition, \textit{non-candidate key attributes are dependent on any candidate key}:

\[
\text{dept, num, sec} \rightarrow \text{ms, days, time, room, cap, inst}
\]
Using Our Example

SIS(\texttt{dept, num, sec, ms, days, time, room, cap, inst})

By definition, non-candidate key attributes are dependent on any candidate key:
\texttt{dept, num, sec \rightarrow ms, days, time, room, cap, inst}

But \texttt{ms} only depends on the department and course number:
\texttt{dept, num \rightarrow ms}
Using Our Example

\[
\text{SIS}(\text{dept, num, sec, ms, days, time, room, cap, inst})
\]

By definition, *non-candidate key attributes are dependent on any candidate key*:

\[
\text{dept, num, sec} \rightarrow \text{ms, days, time, room, cap, inst}
\]

But ms only depends on the department and course number:

\[
\text{dept, num} \rightarrow \text{ms}
\]

This is a *partial key dependency*. 
Using Our Example

SIS(dept, num, sec, ms, days, time, room, cap, inst)

By definition, *non-candidate key attributes are dependent on any candidate key*:  
\[ \text{dept, num, sec} \rightarrow \text{ms, days, time, room, cap, inst} \]

But ms only depends on the department and course number:  
\[ \text{dept, num} \rightarrow \text{ms} \]

This is a *partial key dependency*.

And capacity only depends on the room (not part of the key):  
\[ \text{room} \rightarrow \text{cap} \]
Using Our Example

SIS(dept, num, sec, ms, days, time, room, cap, inst)

By definition, non-candidate key attributes are dependent on any candidate key:

\[ \text{dept, num, sec} \rightarrow \text{ms, days, time, room, cap, inst} \]

But ms only depends on the department and course number:

\[ \text{dept, num} \rightarrow \text{ms} \]

This is a \textit{partial key dependency} - we eliminate this in 2NF.

And capacity only depends on the room (not part of the key):

\[ \text{room} \rightarrow \text{cap} \]

This is a \textit{transitive dependency}. 
The Three Main Normal Forms Defined

1NF - Dependent on the key.
   Given a candidate key (simple or composite), every other attribute is dependent on the key.

2NF - The whole key.
   Given a composite candidate key, every non-key attribute is dependent on the *all* the attributes in the candidate key. That is, there are no *partial key* dependencies.

3NF - And nothing but the key.
   Given a candidate key, no non-key attribute is dependent on any attributes that are not in the candidate key. That is, no *transitive* dependencies.
1NF - The base

• Every relation we develop from a well-formed ERD is technically in 1NF.
• This is true even if we use a sub-optimal approach to multi-valued attributes.
• However, "factoring out" multi-value attributes is usually considered the prerequisite to 1NF.
• As we know how to do this, we'll simply assume we have 1NF relations.
• $\text{SIS}(\text{dept, num, sec, ms, days, time, room, cap, inst})$ is in 1NF
2NF - No partial key dependencies

Every non-key attribute depends on the whole key. This is only a problem with composite keys (why?)
2NF - No partial key dependencies

Every non-key attribute depends on the whole key. This is only a problem with composite keys (why?)

Our SIS example has a partial key dependency:

\[ \text{SIS}(\text{dept}, \text{num}, \text{sec}, \text{ms}, \text{days}, \text{time}, \text{room}, \text{cap}, \text{inst}) \]

- \text{dept, num, sec} \rightarrow \text{ms, days, time, room, cap, inst}
- \text{dept, num} \rightarrow \text{ms}

\text{dept, num} \rightarrow \text{ms}
Normalizing 1NF to 2NF

SIS( dept, num, sec, ms, days, time, room, cap, inst )

department, course → offerings, instructors

department, course → ms

department, course → ms

1. Extract the partial dependency into its own relation:

   CourseCategory( department, course, offering )
Normalizing 1NF to 2NF

SIS( dept, num, sec, ms, days, time, room, cap, inst )
  dept, num, sec → ms, days, time, room, cap, inst
  dept, num → ms

1. Extract the partial dependency into its own relation:
   CourseCategory( dept, num, ms )

2. Make the partial dependency key into the primary key:
   CourseCategory( dept, num, ms )
Normalizing 1NF to 2NF

\[
SIS( \text{dept}, \text{num}, \text{sec}, \text{ms}, \text{days}, \text{time}, \text{room}, \text{cap}, \text{inst} ) \\
\text{dept, num, sec } \rightarrow \text{ ms, days, time, room, cap, inst } \\
\text{dept, num } \rightarrow \text{ ms }
\]

1. Extract the partial dependency into its own relation:
   \[
   \text{CourseCategory( dept, num, ms )}
   \]

2. Make the partial dependency key into the primary key:
   \[
   \text{CourseCategory( dept, num, ms )}
   \]

3. Remove dependent attribute(s) from the original relation - here ms:
   \[
   SIS( \text{dept, num, sec, days, time, room, cap, inst} )
   \]
Normalizing 1NF to 2NF

\begin{itemize}
\item SIS( \texttt{dept, num, sec, ms, days, time, room, cap, inst } )
\item dept, num, sec → ms, days, time, room, cap, inst 
\item dept, num → ms
\item 1. Extract the partial dependency into its own relation:
   CourseCategory( \texttt{dept, num, ms } )
\item 2. Make the partial dependency key into the primary key:
   CourseCategory( \texttt{dept, num, ms } )
\item 3. Remove dependent attribute(s) from the original relation - here \texttt{ms}:
   SIS( \texttt{dept, num, sec, days, time, room, cap, inst } )
\item 4. Turn partial key in original relation into a foreign key to the new relation:
   SIS( \texttt{dept, num, sec, days, time, room, cap, inst } )
   (\texttt{dept, num}) refers to (\texttt{dept, num}) in CourseCategory.
\end{itemize}
## Tables After Normalization

The SIS table is as follows:

\[
\text{SIS(} \text{dept, num, sec, days, time, room, cap, inst} \text{)}
\]

(dept, num) refers to (dept, num) in CourseCategory

### CourseCategory(dept, num, ms)

#### SIS

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>cap</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>40</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>40</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>20</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>20</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>20</td>
<td>JH</td>
</tr>
</tbody>
</table>

#### CourseCategory

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>M</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>S</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>S</td>
</tr>
</tbody>
</table>
3NF - No transitive dependencies

Every non-key attribute depends on nothing but the whole key.
3NF - No transitive dependencies

Every non-key attribute depends on nothing but the whole key. Our 2NF SIS example has a *transitive dependency*:

SIS(\text{dept, num, sec, days, time, room, cap, inst})

\text{dept, num, sec} \rightarrow \text{days, time, room, cap, inst}

\text{room} \rightarrow \text{cap}
3NF - No transitive dependencies

Every non-key attribute depends on nothing but the whole key.

Our 2NF SIS example has a *transitive dependency*:

\[
\text{SIS(} \text{dept, num, sec, days, time, room, cap, inst)} \\
\text{dept, num, sec } \rightarrow \text{ days, time, room, cap, inst} \\
\text{room } \rightarrow \text{ cap}
\]

This is *transitive* because:

- If I know dept, num, sec then I know the room.
- If I know the room then I know the capacity.
Normalizing 2NF to 3NF

\[ \text{SIS( dept, num, sec, days, time, room, cap, inst )} \]
\[ \text{dept, num, sec } \rightarrow \text{ days, time, room, cap, inst} \]
\[ \text{room } \rightarrow \text{ cap} \]

1. Extract the transitive dependency into its own relation:

\[ \text{Capacity( room, cap )} \]
Normalizing 2NF to 3NF

SIS( dept, num, sec, days, time, room, cap, inst )
  dept, num, sec → days, time, room, cap, inst
  room → cap

1. Extract the transitive dependency into its own relation:
   Capacity( room, cap )

2. Make the transitive dependency key into the primary key:
   Capacity( room, cap )
Normalizing 2NF to 3NF

SIS( dept, num, sec, days, time, room, cap, inst )
  dept, num, sec \rightarrow days, time, room, cap, inst
  room \rightarrow cap

1. Extract the transitive dependency into its own relation:
   Capacity( room, cap )

2. Make the transitive dependency key into the primary key:
   Capacity( room, cap )

3. Remove dependent attribute(s) from the original relation - here cap:
   SIS( dept, num, sec, days, time, room, inst )
Normalizing 2NF to 3NF

SIS( dept, num, sec, days, time, room, cap, inst )
  dept, num, sec → days, time, room, cap, inst
  room → cap
1. Extract the transitive dependency into its own relation:
   Capacity( room, cap )
2. Make the transitive dependency key into the primary key:
   Capacity( room, cap )
3. Remove dependent attribute(s) from the original relation - here cap:
   SIS( dept, num, sec, days, time, room, inst )
4. Make original relation transitive key a foreign key to the new relation:
   SIS( dept, num, sec, days, time, room, inst )
   room refers to room in Capacity.
## Tables After Normalization

SIS(\( \text{dept, num, sec, days, time, room, inst} \) )

(\( \text{dept, num} \)) refers to (\( \text{dept, num} \)) in CourseCategory
room refers to room in Capacity

CourseCategory(\( \text{dept, num, ms} \))
Capacity(\( \text{room, cap} \))

### SIS

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>sec</th>
<th>days</th>
<th>time</th>
<th>room</th>
<th>inst</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>01</td>
<td>MWF</td>
<td>11:00</td>
<td>1550</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>220</td>
<td>02</td>
<td>TR</td>
<td>9:30</td>
<td>1550</td>
<td>ML</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>01</td>
<td>TR</td>
<td>11:00</td>
<td>1550</td>
<td>TR</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>02</td>
<td>MWF</td>
<td>9:00</td>
<td>1530</td>
<td>KM</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>01</td>
<td>MWF</td>
<td>9:00</td>
<td>2200</td>
<td>SS</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>02</td>
<td>TR</td>
<td>9:20</td>
<td>2200</td>
<td>JH</td>
</tr>
</tbody>
</table>

### Capacity

<table>
<thead>
<tr>
<th>room</th>
<th>cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1530</td>
<td>20</td>
</tr>
<tr>
<td>1550</td>
<td>40</td>
</tr>
<tr>
<td>2200</td>
<td>20</td>
</tr>
</tbody>
</table>

### CourseCategory

<table>
<thead>
<tr>
<th>dept</th>
<th>num</th>
<th>ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>se</td>
<td>220</td>
<td>M</td>
</tr>
<tr>
<td>se</td>
<td>383</td>
<td>S</td>
</tr>
<tr>
<td>cs</td>
<td>220</td>
<td>S</td>
</tr>
</tbody>
</table>
RSN to SQL

SIS( dept, num, sec, days, time, room, inst )
(dept, num) refers to (dept, num) in CourseCategory
room refers to room in Capacity

CourseCategory(dept, num, ms)

Capacity(room, cap)

CREATE TABLE Capacity (  
    room STRING PRIMARY KEY,  
    cap INTEGER,  
) ;

CREATE TABLE CourseCategory (  
    dept STRING,  
    num INTEGER,  
    ms STRING,  
    PRIMARY KEY ( dept, cnum )  
) ;

CREATE TABLE SIS(  
    dept STRING,  
    num INTEGER,  
    sec INTEGER,  
    days STRING,  
    time STRING,  
    inst STRING,  
    room STRING REFERENCES Capacity( room ),  
    PRIMARY KEY ( dept, num, sec ),  
    FOREIGN KEY ( dept, num ) REFERENCES CourseCategory( dept, cnum )  
) ;
How Do Unnormalized Relations Arise?

1. Sometimes as a result of mediocre ER models.
2. Sometimes from increased information about the domain.
3. Sometimes because of database refactoring.
4. Sometimes as a performance compromise (joins ain't cheap!).
   - Normalization means more tables.
   - SQL Queries require that tables be joined to perform the query – means more time & space required.
5. Sometimes when converting a "spreadsheet" database into a "real" database.