CAS Information Engineering

## Leistungsnachweis für Modul DB & DWH

**Gruppenmitglieder**:

Andreas Fischer

Bernd Novotny

Tobias Schieferdecker

Inhaltsverzeichnis:

Leistungsnachweis für Modul DB & DWH 1

1. Übersetzung des gegeben ERMs in den uns bekannten Dialekt 1

2. Auslagern der relevanten Daten in csv-Dateien 1

3. Erstellung der staging-area 2

4. Reparieren der korrupten Daten 2

5. Konzipierung des Sternschemas für das DWH 2

6. Erstellung und Befüllen des DWH 2

7. Auswertungen auf Basis des DWH 2

8. Fazit und Schlusswort 2

# **Übersetzung des gegeben ERMs in den uns bekannten Dialekt**

Es soll die gegebenen Datenbank «Northwind» genommen und gemäss einem neu anzufertigen Schema migriert werden. Das Schema soll im ERM-Dialekt von H.-W. Buff ausgeführt werden.

## Ausgangsdaten

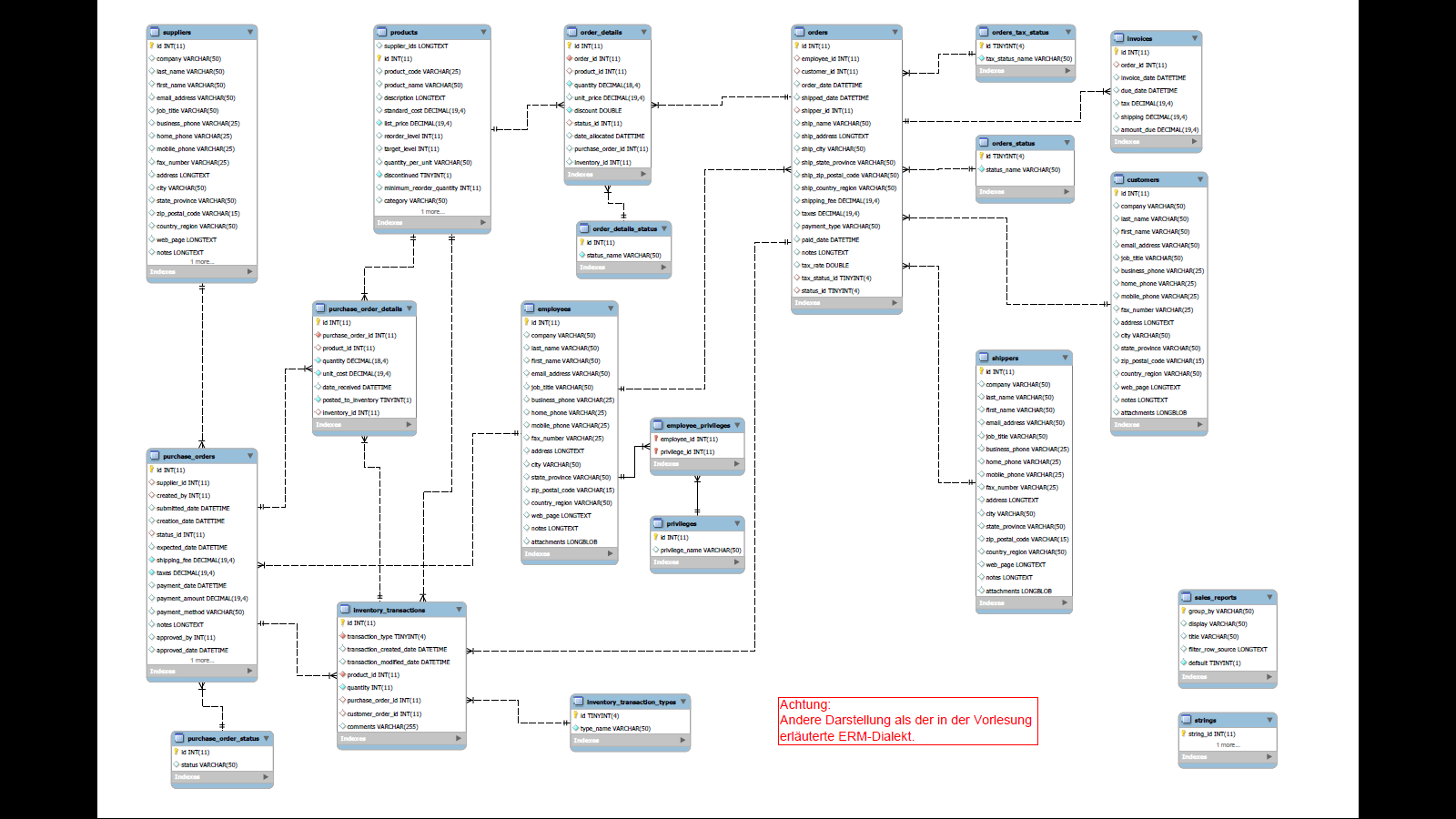
Zur Erzeugung der Datenbank liegen die beiden SQL-Skripte

* Northwind-Database.sql
* Northwind-Daten.sql

vor. Sie können im Workbench geöffnet und ausgeführt werden.

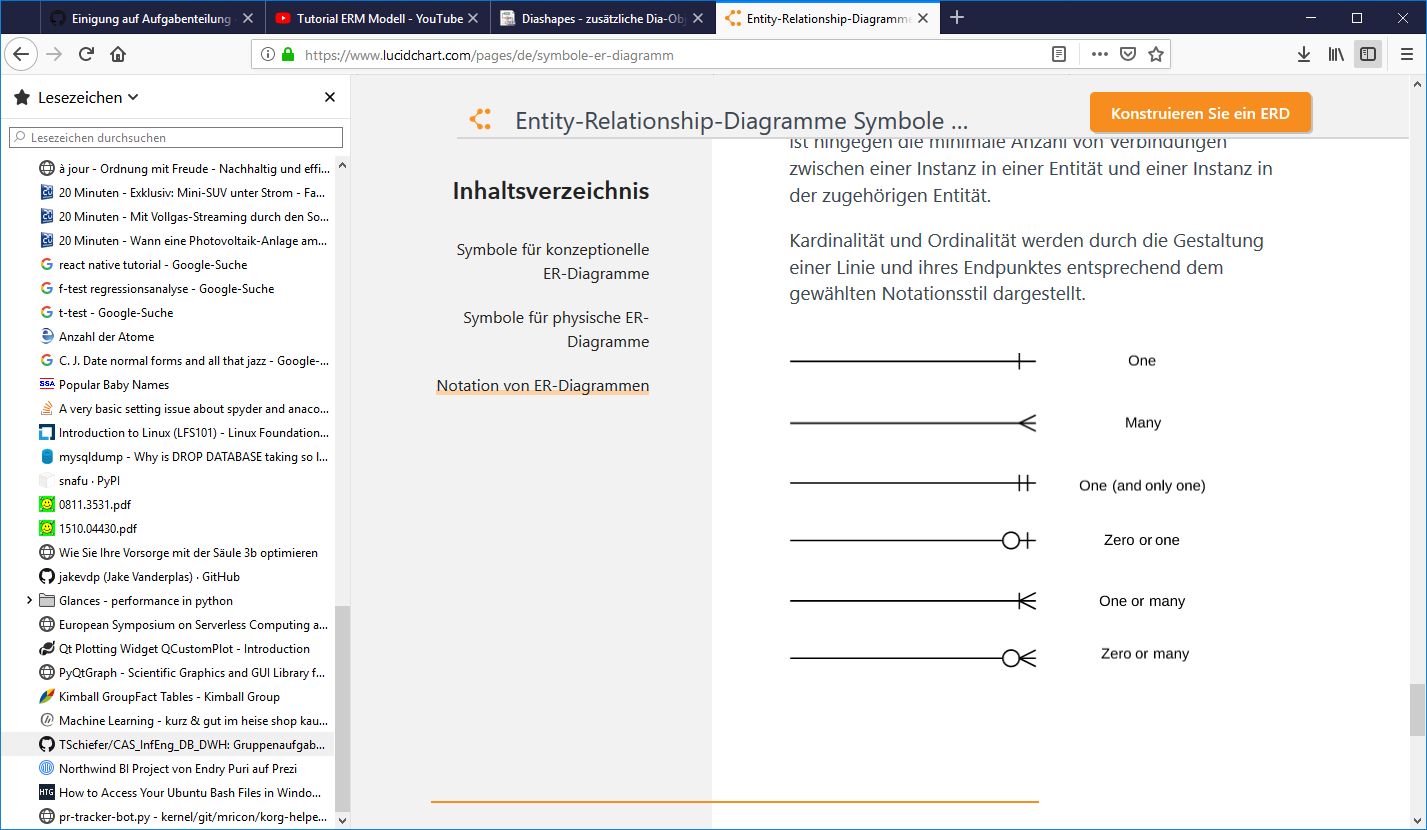
Nach erfolgreichem Durchlauf liegt die Datenbank mit 20 Entitäten und einigen Testdaten eingelesen vor.

Das Schema der ERP-Datenbank «Northwind» liegt in der «Bachman's crow-foot notation» vor. Sie kann z.B. erzeugt werden durch das MySQL-Tool Workbench.



Die Relationen sind hier durch Verbindungen mit speziellen Symbolen dargestellt, die die Kardinalitäten angeben.

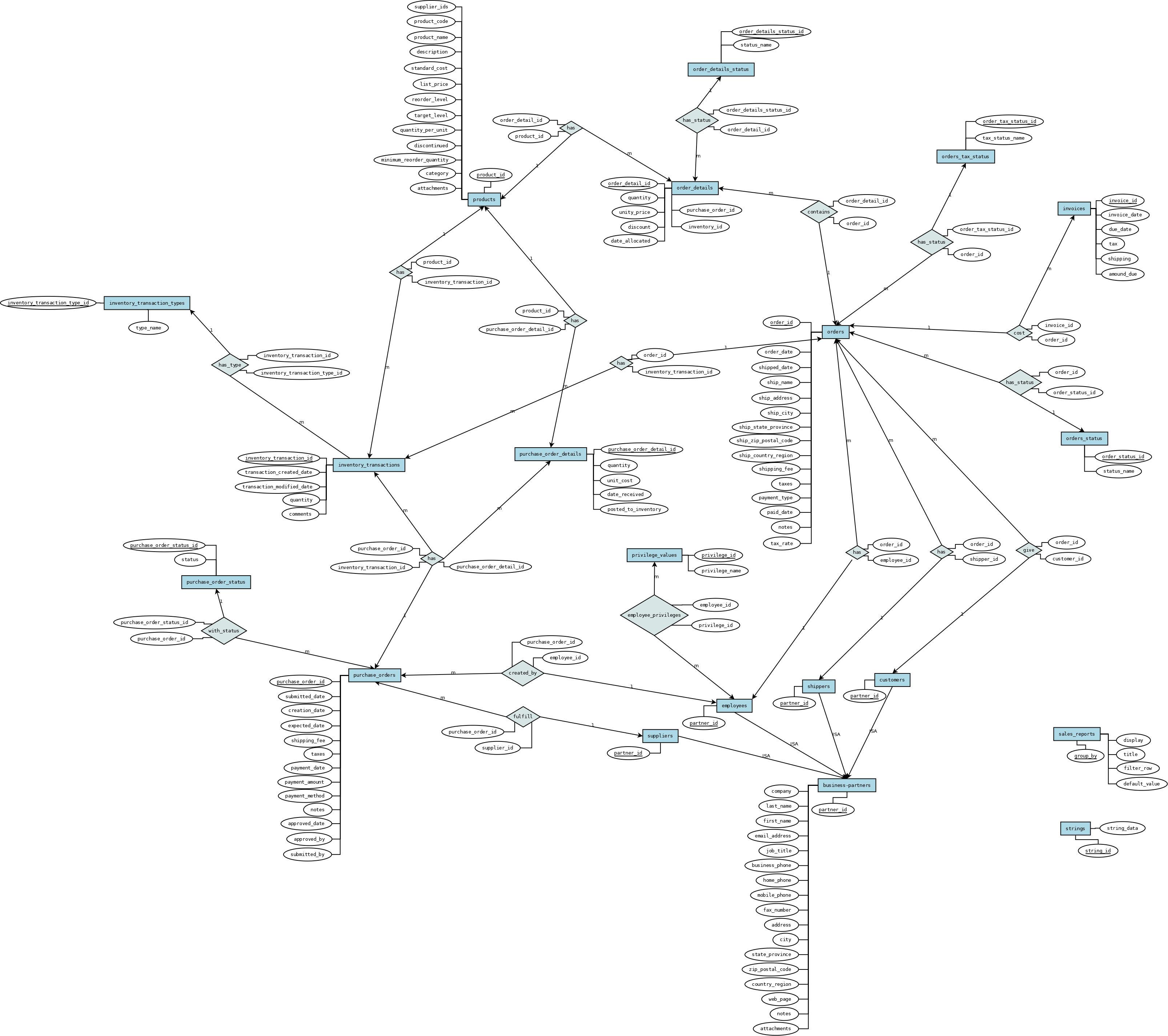
Die folgende Figur listet die verschiedenen Varianten auf.



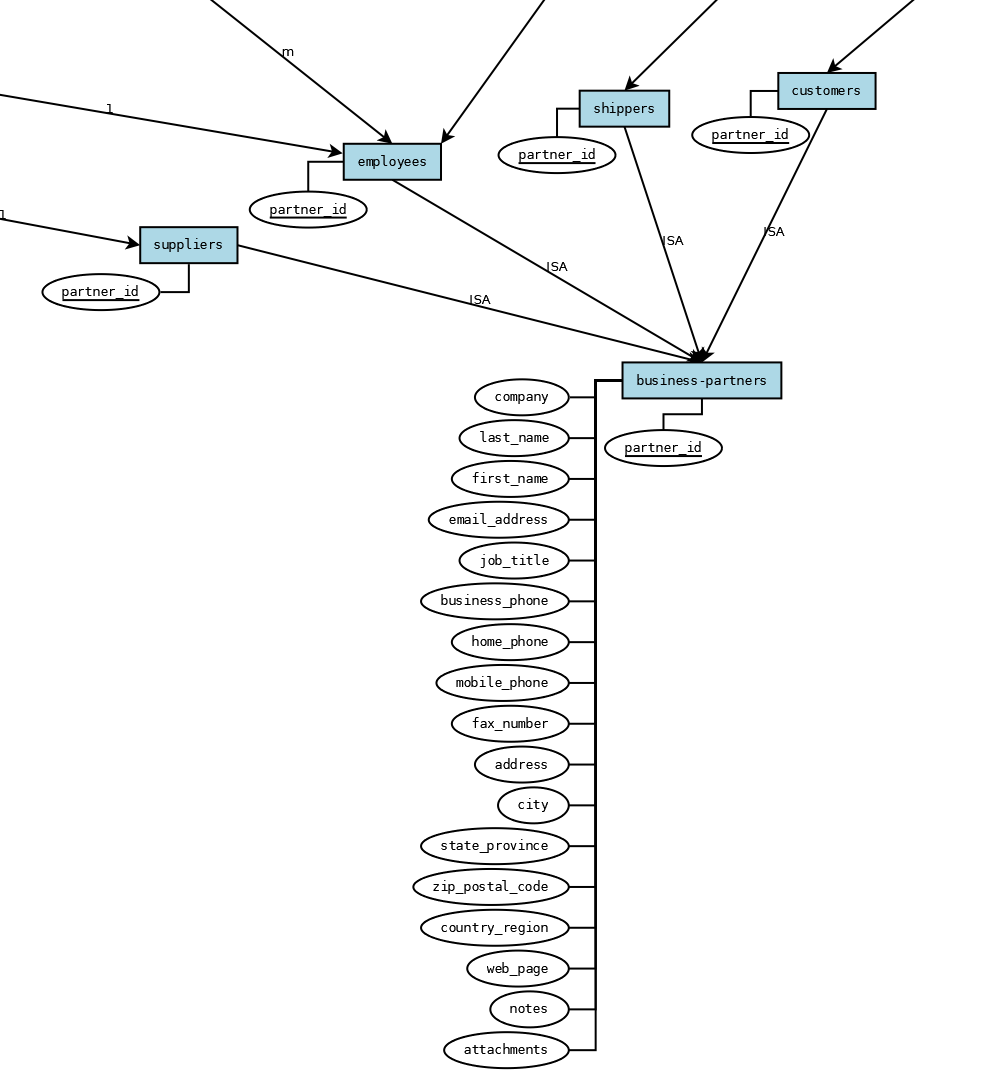
Um das neue ERM-Diagramm zu zeichnen wird das Tool Dia verwendet. Siehe dazu <https://de.wikipedia.org/wiki/Dia_(Software>) oder <http://dia-installer.de/index.html.de>.

## Entity-Relationship-Modell (ERM)

Die folgende Figur zeigt das neu erstellt ERM-Diagramm in Buff-Notation.



Speziell wurde versucht ISA-Beziehungen einzubauen, da die Entitäten «suppliers», «employees», «shippers» und «customers» identische Attribute aufweisen. Dazu wurde neu die Entität «business-partners» eingefügt.



Beobachtungen:

* Das Diagramm ist grafisch anspruchsvoller und grösser als das ursprüngliche. Vor allem wenn viele Attribute vorhanden sind.
* Es sind mehr Elemente nötig, denen gute Namen gegeben werden müssen. Dies ist nicht immer einfach.
* Das gewählte Tool ist nicht ideal. Das Handling und die Darstellung könnten verbessert werden. Vor allem das Zeichnen der Relationen und deren Beschriftung sollte besser sein.
* Den Schlüsseln der Entitäten müssen eindeutige Bezeichnungen gegeben werden, da sie nicht nur in den Entitäten sondern auch in den Relationen verwendet werden.

## SQL Scripte

Zur Umsetzung des ERM-Diagramms in eine Datenbank werden SQL-Scripte benötigt. Als Basis werden die ursprünglichen Scripte genommen und angepasst.

Zur Erzeugung der Datenbank muss als erstes immer das Skript **Northwind2-Database.sql** ausgeführt werden. Es löscht eine schon vorhandene Datenbank und baut sie dann neue wieder auf.

Danach können die Test-Daten eingelesen werden durch das Skript **Northwind2-Data.sql.**

Beobachtungen:

* Die ursprünglichen Scripte wurden scheinbar aus der Datenbank im MySQL-Workbench exportiert. Darin sind die Bezeichner alle «gequoted» mit backticks. Dies wurde in den neuen Scripten nicht mehr beibehalten.
* Die Schlüssel wurden wie im Diagramm konsequent mit Entitätsnamen und angehängtem «\_id» bezeichnet.
* Die Reihenfolge beim Einlesen der Daten ist wichtig. Daten auf die später verwiesen wird, müssen schon in der Datenbank sein. Die Constraints werden sonst nicht erfüllt.
* In den ursprünglichen Daten sind einzelne Records vorhanden, die auf Entitäten verweisen sollten aber mit NULL gefüllt. Diese konnten erst eingelesen werden, als die NULL-Werte durch vorhandene Fremdschlüssel ersetzt wurden.
* Da die Entitäten «suppliers», «employees», «shippers» und «customers» nur noch aus Schlüsseln bestehen und eine ISA-Beziehung zu der Entität «business\_partners» haben, mussten die Schlüssel angepasst werden.

Die beiden Scripte laufen ohne Fehlermeldungen oder Warnungen durch und erzeugen bzw. füllen die Datenbank.

# **Auslagern der relevanten Daten in csv-Dateien**

Nachdem die Northwind-Datenbank mit der uns zur Verfügung gestellten Datei “Northwind-Database.sql” mittels MySQL-Workbench erzeugt und im Folgenden im selben Tool mittels der Datei “Northwind-Data.sql” befüllt wurde, haben wir uns der Aufgabe gewidmet, die für unsere späteren Auswertungen relevanten Daten in csv-Dateien auszulagern.



Bild 1 Der Knopf für den Export einer Datentabelle aus MySQL-Workbench ist rot umrandet.

Für diesen Schritt haben wir kein Skript verwendet. Wir haben vielmehr einfach die Funktionalität von MySQL-Workbench verwendet haben, auf Knopfdruck ein csv-File aus der gerade betrachteten Tabelle zu erzeugen (siehe Bild 1)

# **Erstellung der staging-area**

# **Reparieren der korrupten Daten**

# **Konzipierung des Sternschemas für das DWH**

# **Erstellung und Befüllen des DWH**

# **Auswertungen auf Basis des DWH**

# **Fazit und Schlusswort**