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 2

2. Auslagern der relevanten Daten in csv-Dateien 2

3. Erstellung der staging-area 2

4. Reparieren der korrupten Daten 3

5. Konzipierung des Sternschemas für das DWH 3

6. Erstellung und Befüllen des DWH 3

7. Auswertungen auf Basis des DWH 3

8. Fazit und Schlusswort 3

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

# **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 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). Die Tabellen, die uns für spätere Auswertungen interessant schienen, waren:

* “customers”
* “products”
* “orders”
* “order\_details

Hierbei wurden an dieser Stelle keine Einschränkung der Daten vorgenommen, sondern das dem SQL-Befehl

“SELECT \* FROM ‘TABLE’ ;” (mit ‘TABLE’ gleich einer der vier von uns ausgesuchten Tabellen) entsprechenden Resultat in das csv-Format exportiert. Die entsprechenden csv-Files sind dem zip-File “skripte\_und\_daten\_CAS\_DB\_DWH.zip”(?) beigefügt, das im Email-Anhang an Herrn Aebi verschickt wurde.

# **Erstellung der staging-area**

Wir waren uns bezüglich der staging-area nicht sicher, wie hoch der geforderte Anspruch an deren „physisches Vorhandensein“ ist. Sprich, soll eine zur Northwind-Datenbank parallele Datenbank mittels MySQL-Workbench erstellt werden, die dann eine eingedampfte Version der Daten enthält? Wir haben uns letztendlich dafür entschieden, eine staging-area-Datenbank zu kreieren und zu befüllen, haben aber letztendlich für die Reparatur-Transformationen die Daten aus den csv-Files gelesen und nach den Transformationen die Daten auch wieder in csv-Files geschrieben. Genauso wird bei unserer Lösung die DWH-Datenbank („NORTHWIND\_DWH“) mit Daten aus den csv-Dateien befüllt und nicht mit solchen aus der staging-area.

Nichtsdestotrotz hier eine kurze Erläuterung, wie die staging-area-Datenbank erzeugt und befüllt wird.

1. Das Skript zur Erzeugung der leeren Datenbank ist in MySQL-Workbench auszuführen und lautet „create\_staging\_area\_schema.sql“.
2. ... usw.

# **Reparieren der korrupten Daten**

# **Konzipierung des Sternschemas für das DWH**

# **Erstellung und Befüllen des DWH**

# **Auswertungen auf Basis des DWH**

# **Fazit und Schlusswort**