

Abschlussveranstaltung

Praktikum: Data Warehousing und Mining

Agenda für heute

- DMC-Rückblick
- ETL-Rückblick
- Praktikumsevaluation
- Scheinvergabe

DMC: Ergebnisse

Platz	Team	Punkte
1	Inst_KIT_2	12545,5
2	TU_Dortmund_2	12445,5
3	Inst_KIT_1	12377,5
4	RWTH_Aachen_1	12119,5
5	FH_Trier_1	12072,5
6	RWTH_Aachen_2	12072,0
7	Uni_Marburg_1	12006,5
8	HS_Anhalt_2	11961,5
9	HS_Anhalt_1	11941,5
10	Uni_UPC Barcelona_1	11566,0
11	Uni_Hamburg_2	11532,5
12	FH_Köln_1	11444,0

DMC: Interne Ergebnisse

Team	DMC-Punkte	virt. Platz	DWM-Punkte
1	11866,0	10	6,5
2	11566,0	10	6,5
3	11399,5	15	6
4	11707,5	10	6,5

- Für Abgabe 1 Punkt
- Bis zu 6 Punkte je nach Platzierung:

$$6 - \frac{\lfloor \frac{VP}{67} * 12 \rfloor}{2}$$

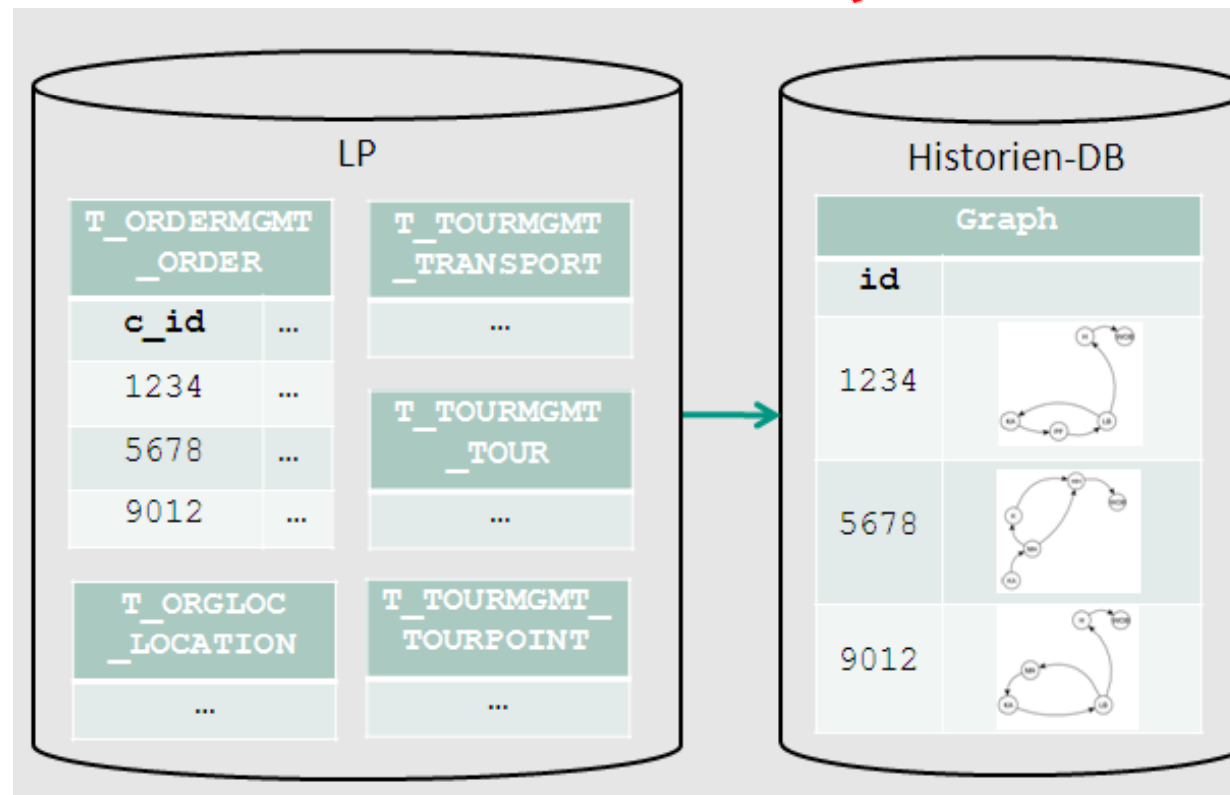
...Lust auf den nächsten Wettbewerb? 😊

- 2010 UC San Diego Data Mining Contest
- 01.07. – 31.08.2010
- Aufgabe: „Identifying potential new customers from a population of consumers“
- Trainingsdaten: 130.475 Kundendatensätze, 334 Variablen, Ratio Kunden/Nichtkunden 1:10
- Testdaten: 86.691 Datensätze
- Evaluationsmetrik: Area under the Curve (AUC)
- Live-Leaderboard

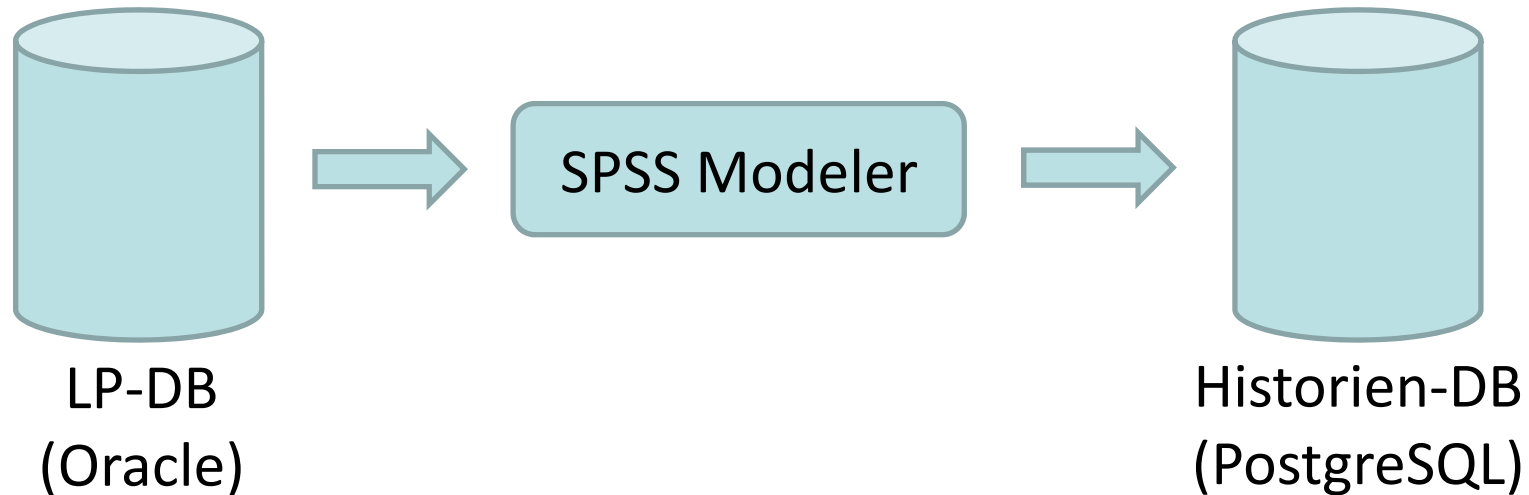
Transformationsprozess

- ETL-Aufgabe

**Folie aus
Veranstaltung 5**

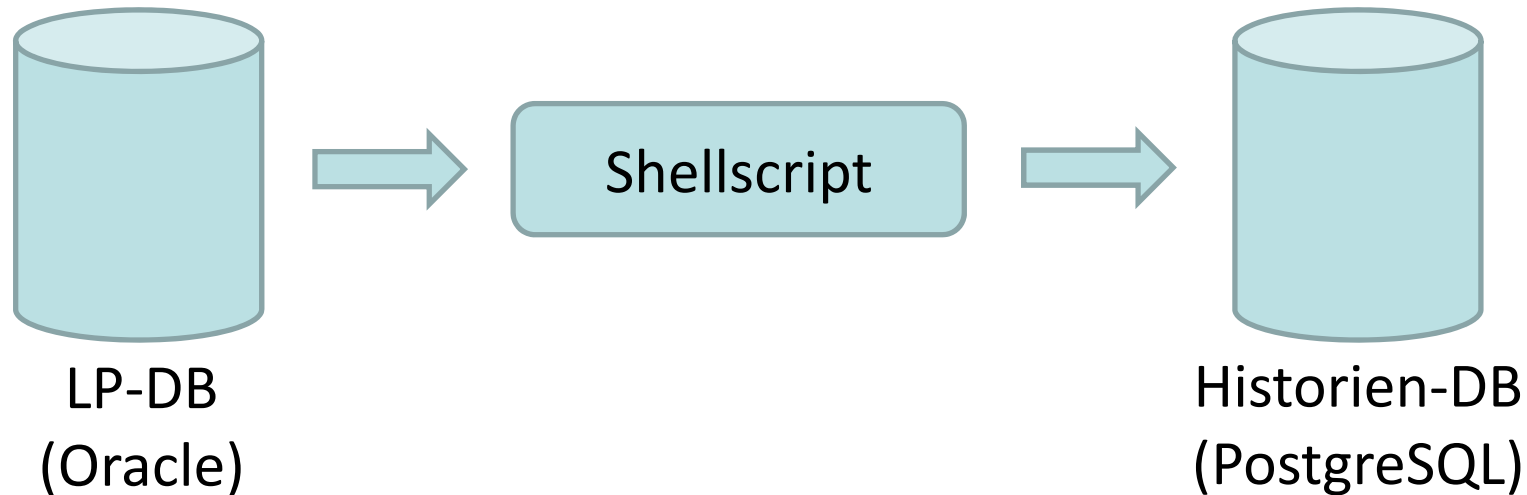


ETL-Rückblick: Häufigste Lösungsvariante



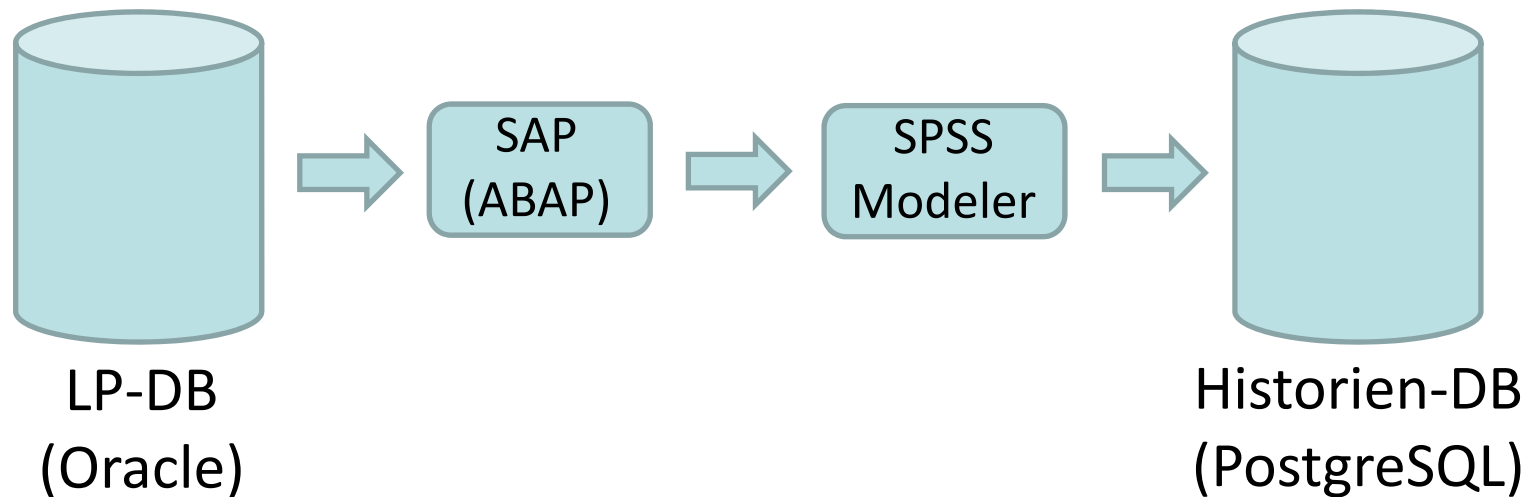
- Prozesssteuerung durch SPSS Modeler
- Transformation:
 - meist im Modeler
 - gelegentlich unterstützt durch SQL-Statements in den Datenbanken

ETL-Rückblick: Lösungsvarianten (2)



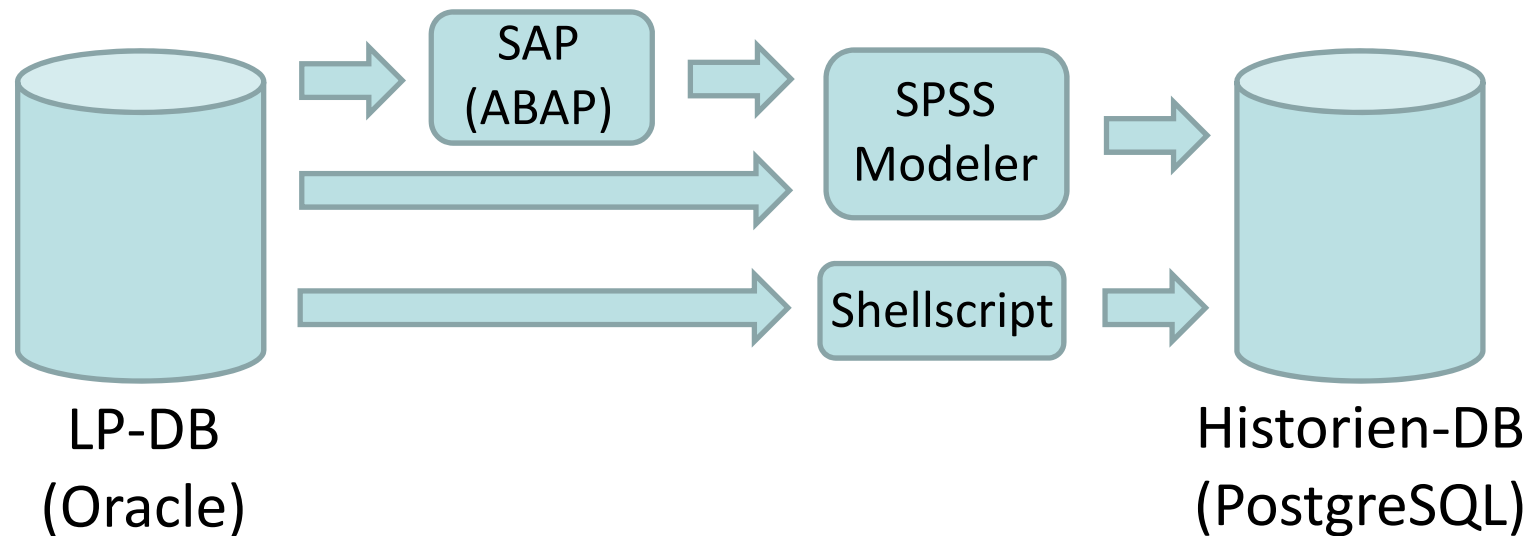
- Prozesssteuerung durch Shellscript
 - Aufrufe von sqlplus, psql
 - Daten zwischenzeitlich als csv-Dateien gespeichert
- Transformation mit PL/pgSQL

ETL-Rückblick: Lösungsvarianten (3)



- Transformation in ABAP
- SPSS Modeler nur zum Laden verwendet

ETL-Rückblick: Lösungsübersicht



Fazit:

- Die Transformation ist sowohl innerhalb als auch außerhalb der Datenbanken möglich.
- Der SPSS Modeler ist (problemlos?) als ETL-Tool verwendbar.

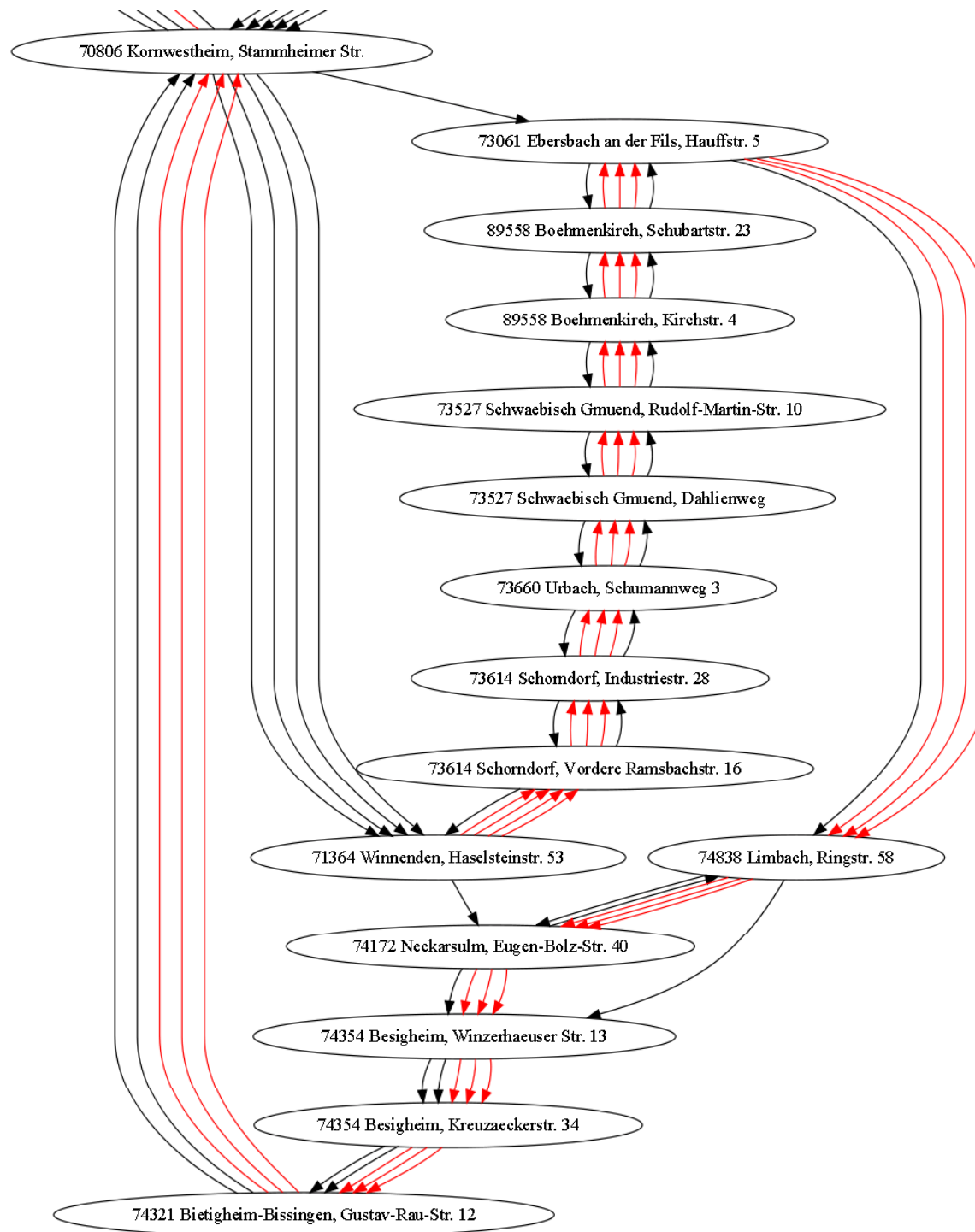
Pilotbetrieb LogoTakt

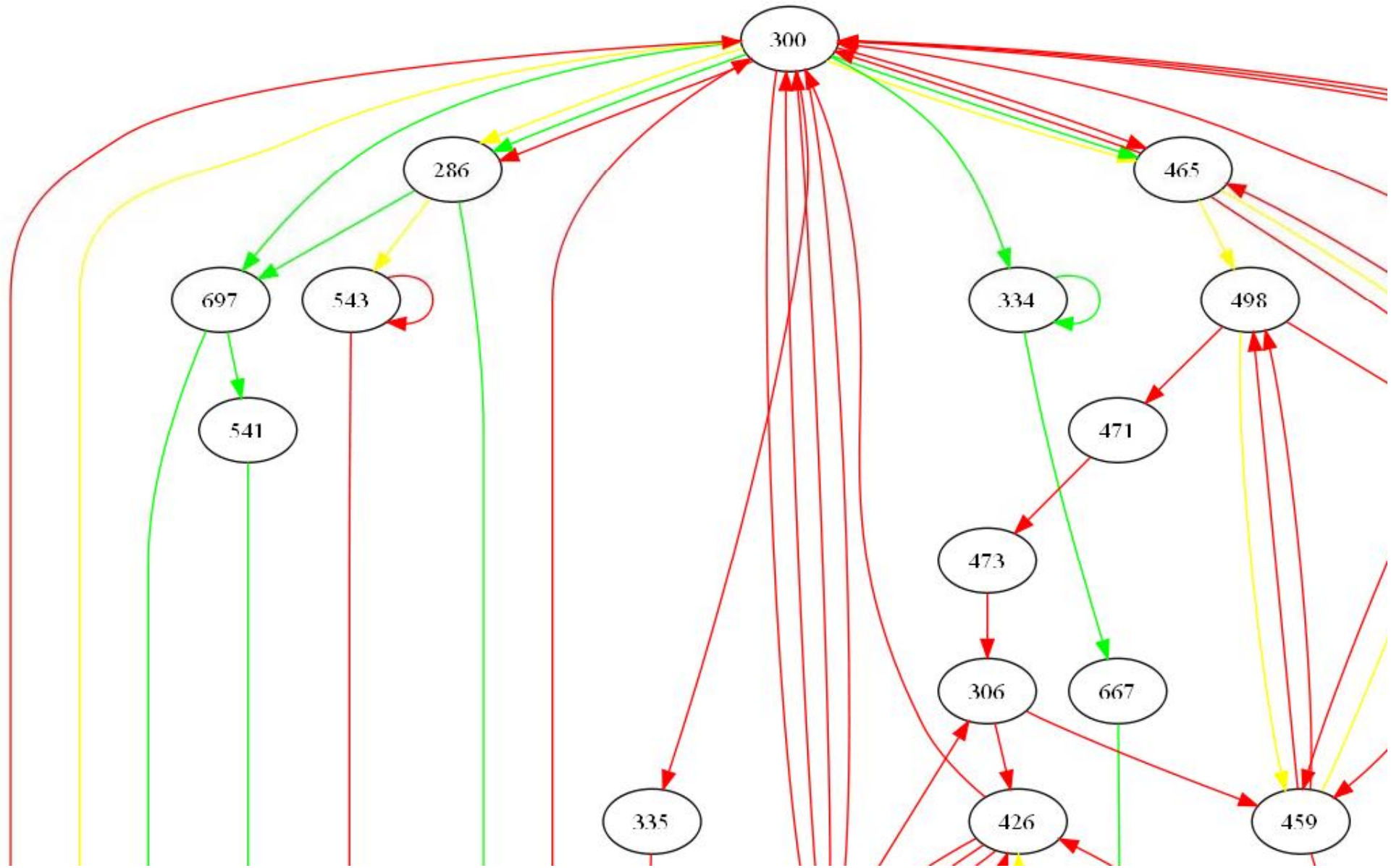


- Erste Realdaten werden seit diesem Monat gesammelt.
- Anschließend sind erste Realwelt-Erkenntnisse mit vorgestellten Analysetechniken (Graph Mining, Graphanfragen) zu erwarten.
- Ausgewählte ETL-Prozesse werden je nach Analysebedarf zugehörige Datenbasis erzeugen.

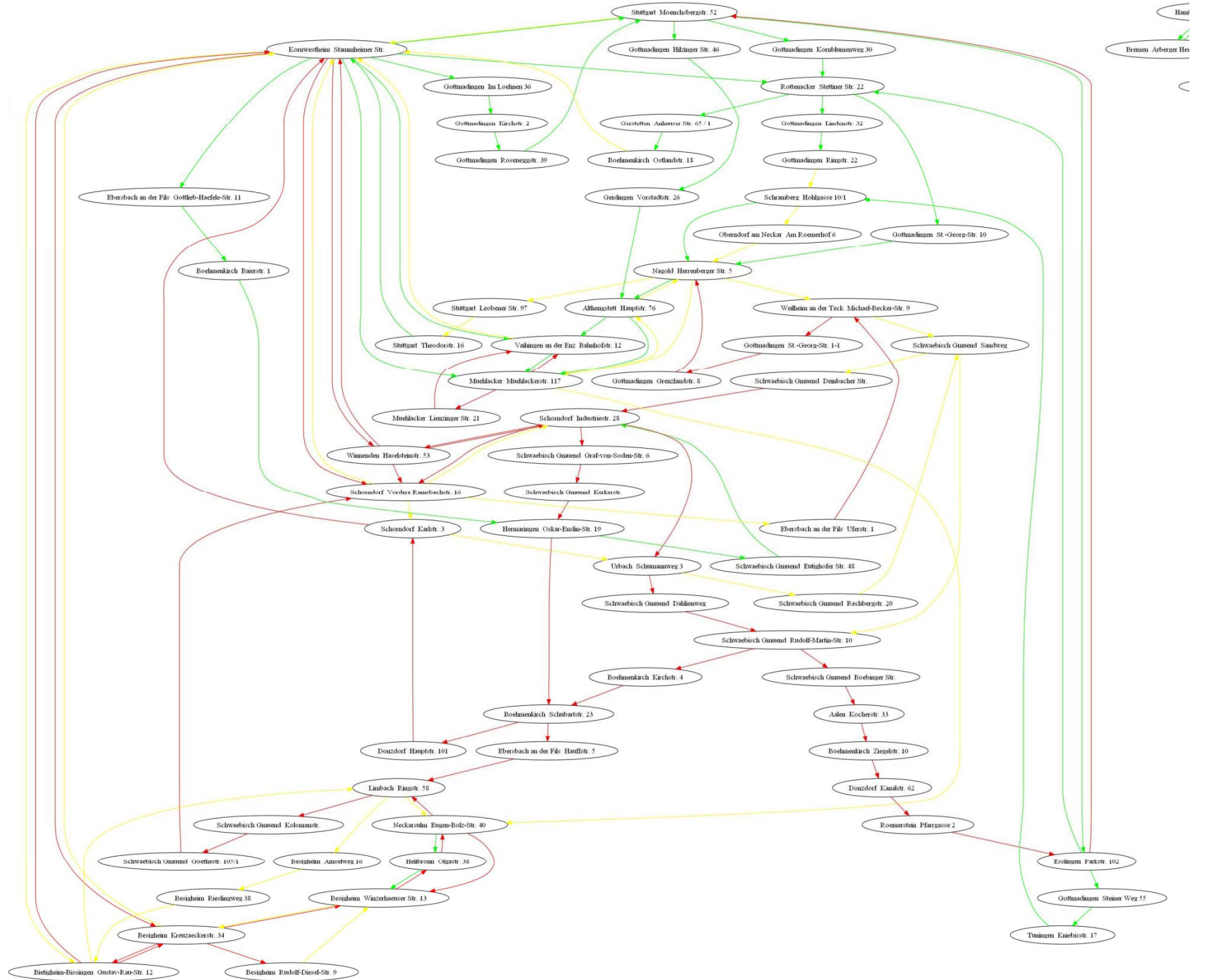
Rückblick: Graphvisualisierung

- (Nicht gefordert, aber wir haben uns trotzdem darüber gefreut:)
- Automatisierte Lösungen bis hin zu einem fertig konfigurierten Eclipse-Projekt
- <Demo>

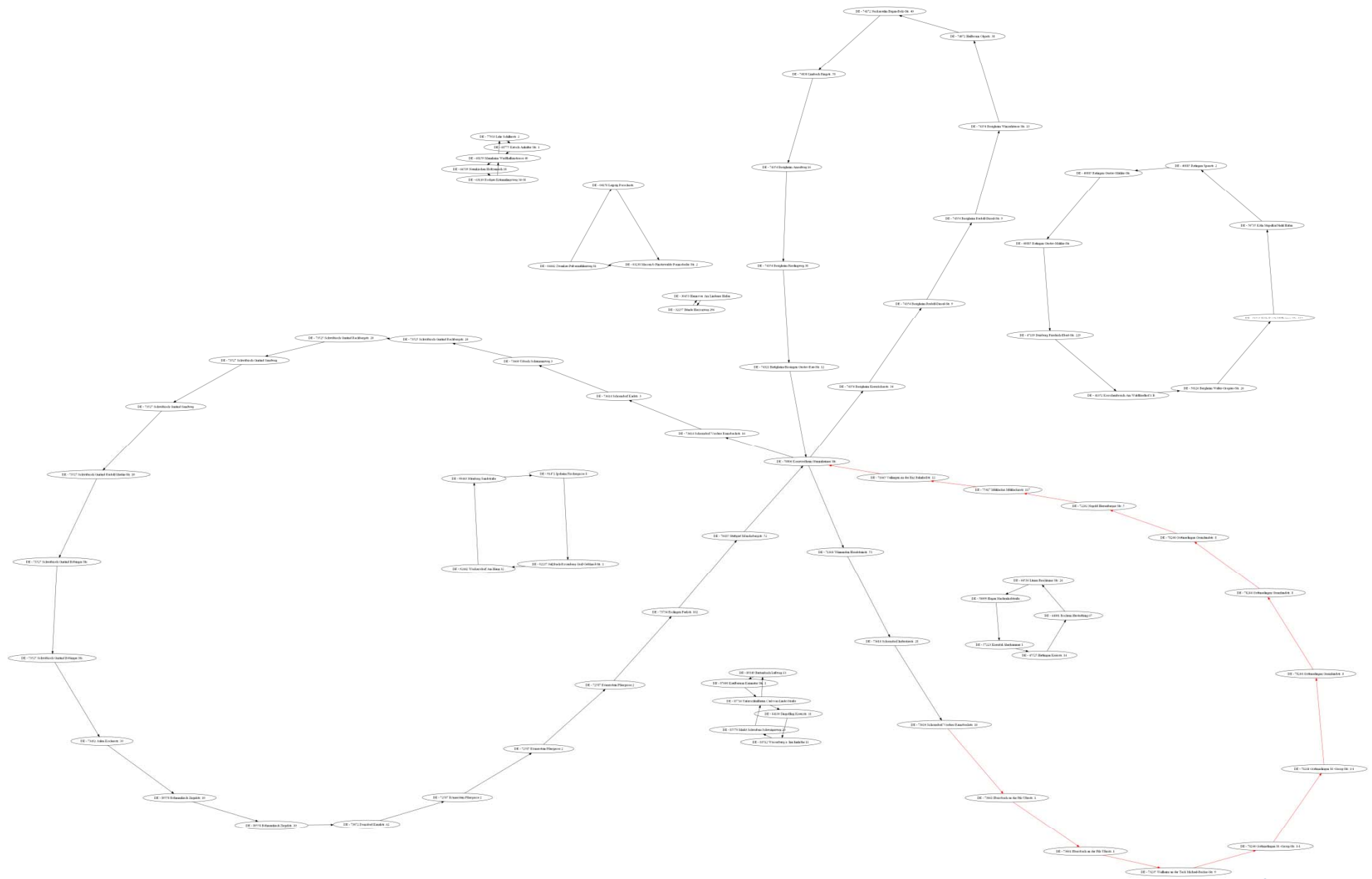




15



Haus
Bremen Arberger Heer



Praktikumsevaluation

- „Vorlesungsteil“: Montag, 9:45 Uhr
- „Praktikumsteil“: Aufgaben (DMC, ETL, Cognos) sowie deren Bearbeitung in den Tutorien
- Uns helfen am meisten: Freitextantworten (für Abschnitt 2 auch in 3.12/3.13)
- Gesamtbewertung: 2.12 und 3.11
- Sonstige Kommentare auch gerne per E-Mail

Arbeiten am Institut

- Studien-/Diplomarbeiten
bzw. Bachelor-/Masterarbeiten
- Hiwi-Jobs

- Bei Fragen bezüglich verfügbarer Themen
und Aufgaben an Frank oder Matthias
wenden

Scheinvergabe

Vielen Dank für eure Teilnahme
und viel Erfolg in eurem weiteren Studium!