



# Konzeption und Evaluation eines Data Warehouse

zur Analyse von Daten im  
Verkehrsbereich

eine Studienarbeit von Jan Sandberger

Betreuer: Heiko Schepperle



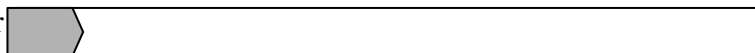
# Data Warehouse System

- System zur Auswertung von großen Datenmengen (evtl. basierend auf relationalen Datenbanken)
- Multidimensionale Speicherung
- Data Warehouse Schema enthält eine Faktentabelle und mehrere (möglichst orthogonale) Dimensionstabellen
- Redundante Datenverwaltung zur Beschleunigung von Anfragen



# Einführung

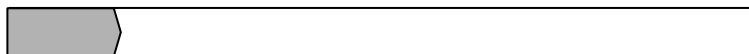
- Ein Ziel des Projekts OVID: intelligente Verkehrsleitsysteme
- Bisher nur simulierte Messdaten:
- Simulationsprogramm VISSIM:
  - Verkehrswegenetze
  - Messstationen
  - verschiedene Fahrzeuge





# Roadmap

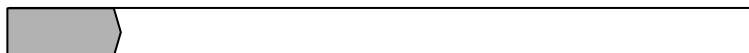
- Das Nutzdatenschema
- Das Füllen der Nutzdatenbank
- Das Data Warehouse Schema
- Anfragen an das Data Warehouse





# Roadmap

- Das Nutzdatenchema
  - VISSIM und seine Logdaten
  - Besonderheiten einer Simulation
  - Umsetzung in ein relationales Schema
- Das Füllen der Nutzdatenbank
- Das Data Warehouse Schema
- Anfragen an das Data Warehouse

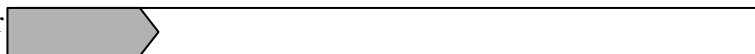




# VISSIM und die Logdaten

Messung 111: Strecke 1 Spur 1 bei 137.509 m, Länge0.000 m.  
Messung 112: Strecke 1 Spur 2 bei 136.416 m, Länge0.000 m.  
Messung 1311: Strecke 13 Spur 1 bei 109.251 m, Länge0.000 m.  
Messung 1312: Strecke 13 Spur 2 bei 109.412 m, Länge0.000 m.  
Messung 1611: Strecke 16 Spur 1 bei 17.832 m, Länge0.000 m.  
Messung 2811: Strecke 28 Spur 1 bei 25.557 m, Länge0.000 m.

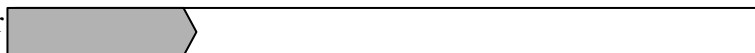
Messung	t(Einf)	t(Ausf)	Fz-Nr	Fztyp	Linie	v[m/s]	b[m/s <sup>2</sup> ]	Bel	Pers	tStau	FzLaenge
2811	600.62	-1.00	960	100	0	10.8	-1.69	0.08	2	0.0	4.40
2111	-1.00	600.62	954	100	0	11.9	-1.50	0.02	2	0.0	4.55
3211	600.99	-1.00	912	100	0	12.9	0.58	0.01	2	0.0	4.11
2811	-1.00	601.04	960	100	0	10.0	-1.68	0.04	2	0.0	4.40
3411	601.07	-1.00	952	100	0	17.1	-0.10	0.03	1	0.0	4.11
6912	601.24	-1.00	817	150	0	18.0	0.14	0.06	1	56.4	6.31
3211	-1.00	601.31	912	100	0	13.0	0.18	0.01	2	0.0	4.11
3411	-1.00	601.31	952	100	0	17.1	-0.11	0.01	1	0.0	4.11
6912	-1.00	601.59	817	150	0	18.0	0.14	0.09	1	56.4	6.31





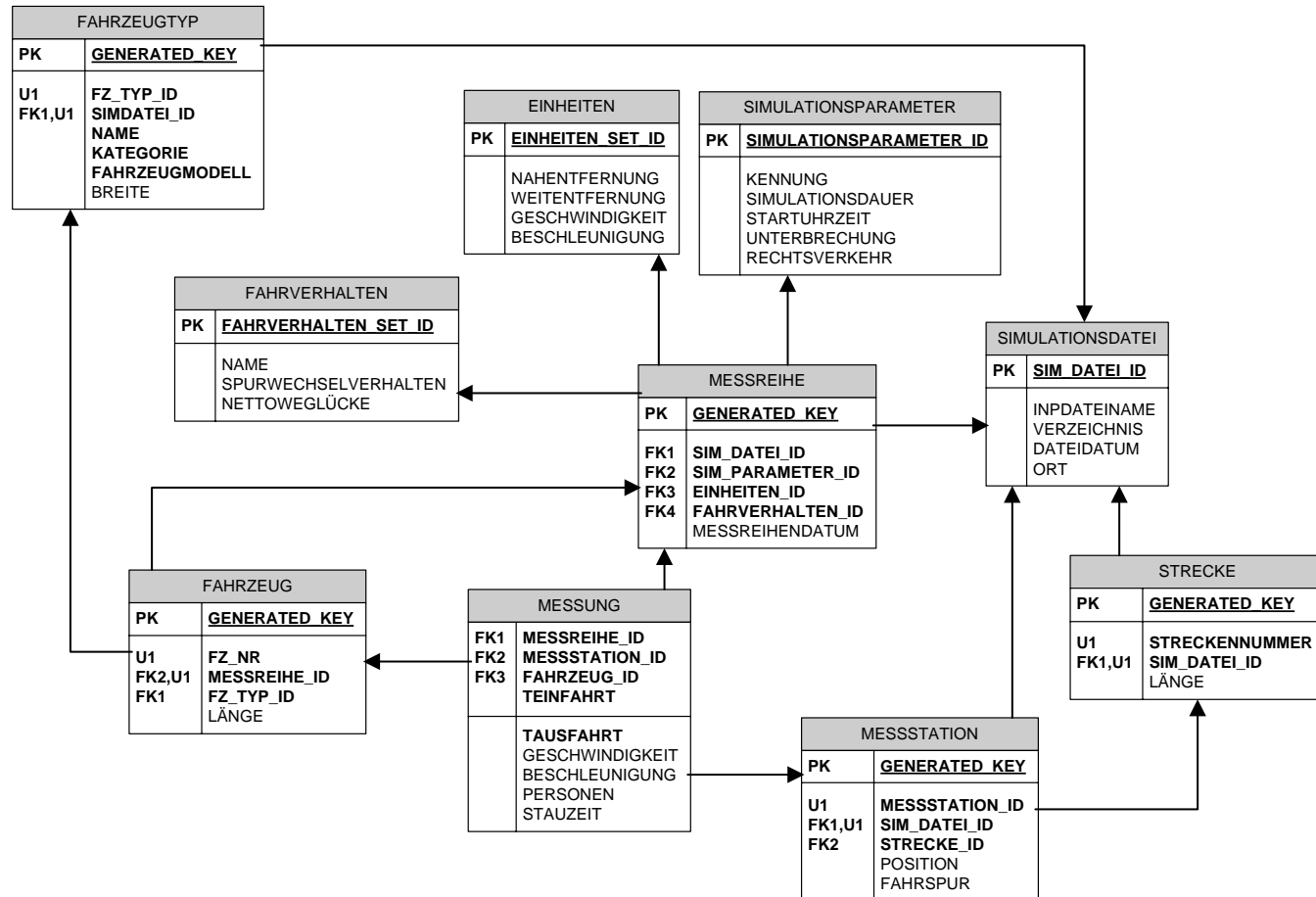
# Besonderheiten einer Simulation

- Problem: Eindeutige Werte in der Datenbank
  - IDs aus der Logdatei sind global nicht eindeutig
  - Eindeutigkeit entweder durch Angabe der Messreihe oder der Simulationsdatei
  - Erzeugen künstlicher Schlüssel
  - Fremdschlüsselverweise auf künstliche Schlüssel





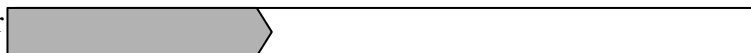
# Das relationale Schema





# Roadmap

- Das Nutzdatenschema
- **Das Füllen der Nutzdatenbank**
- Das Data Warehouse Schema
- Anfragen an das Data Warehouse





# Füllen der Datenbank

Implementierung in Java mittels JDBC

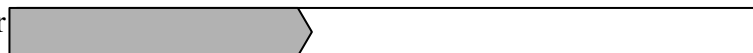
1. Einfügen der Simulationsumgebung

2. Einfügen der Messstationen

3. Einfügen der Messungen

Jeweils zwei Zeilen werden zusammengefasst

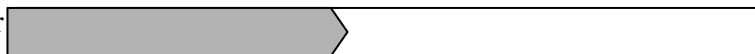
Messung	t(Einf)	t(Ausf)	Fz-Nr	Fztyp	Linie	v[m/s]	b[m/s <sup>2</sup> ]	Bel	Pers	tStau	FzLaenge
2811	600.62	-1.00	960	100	0	10.8	-1.69	0.08	2	0.0	4.40
2811	-1.00	601.04	960	100	0	10.0	-1.68	0.04	2	0.0	4.40





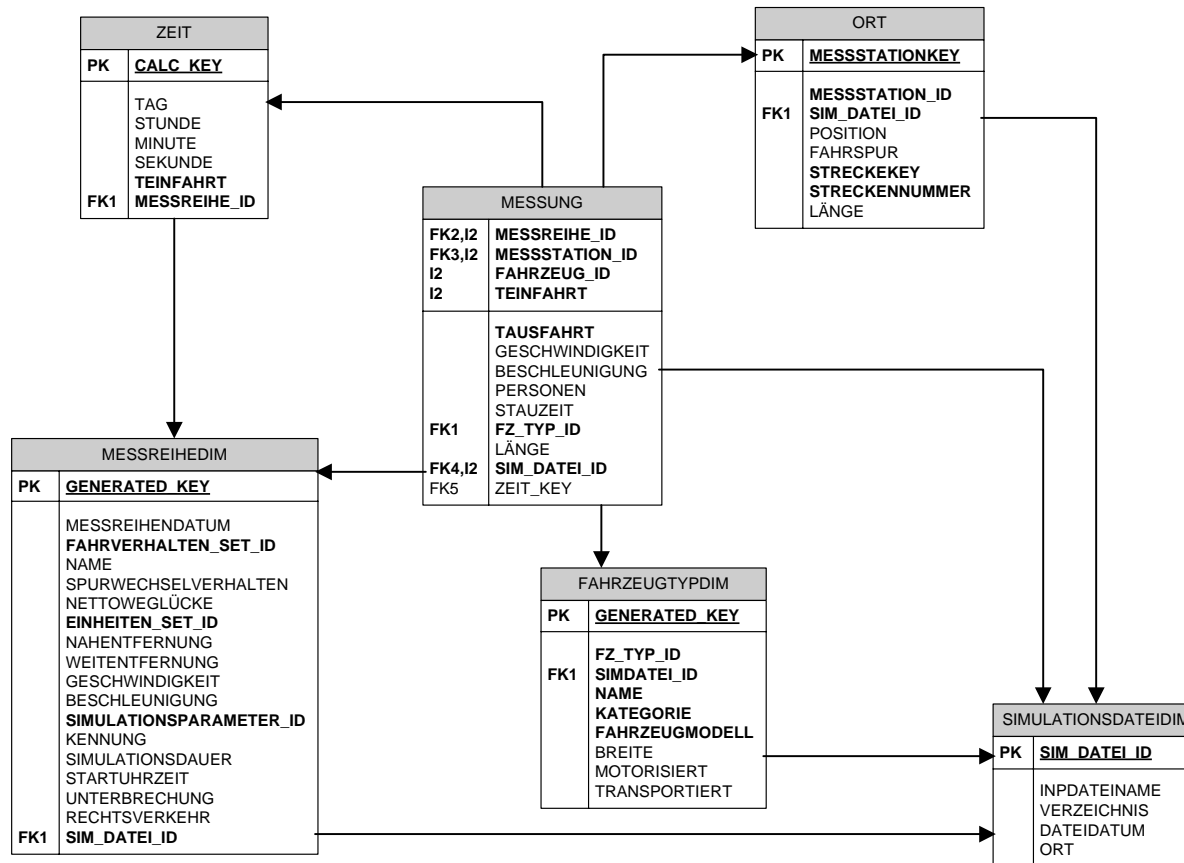
# Roadmap

- Das Nutzdatenschema
- Das Füllen der Datenbank
- **Das Data Warehouse Schema**  
Materialisierte Sichten
- Anfragen an das Data Warehouse





# Warehouse Schema





# Warehouse Schema

- Faktentabelle Messung
- Restliche Relationen: Dimensionstabellen
- Besonderheiten:
  - Fahrzeugtyp enthält zwei neue Felder zur Aggregierung
  - Zeit enthält nur redundante Information, Felder zur Aggregierung





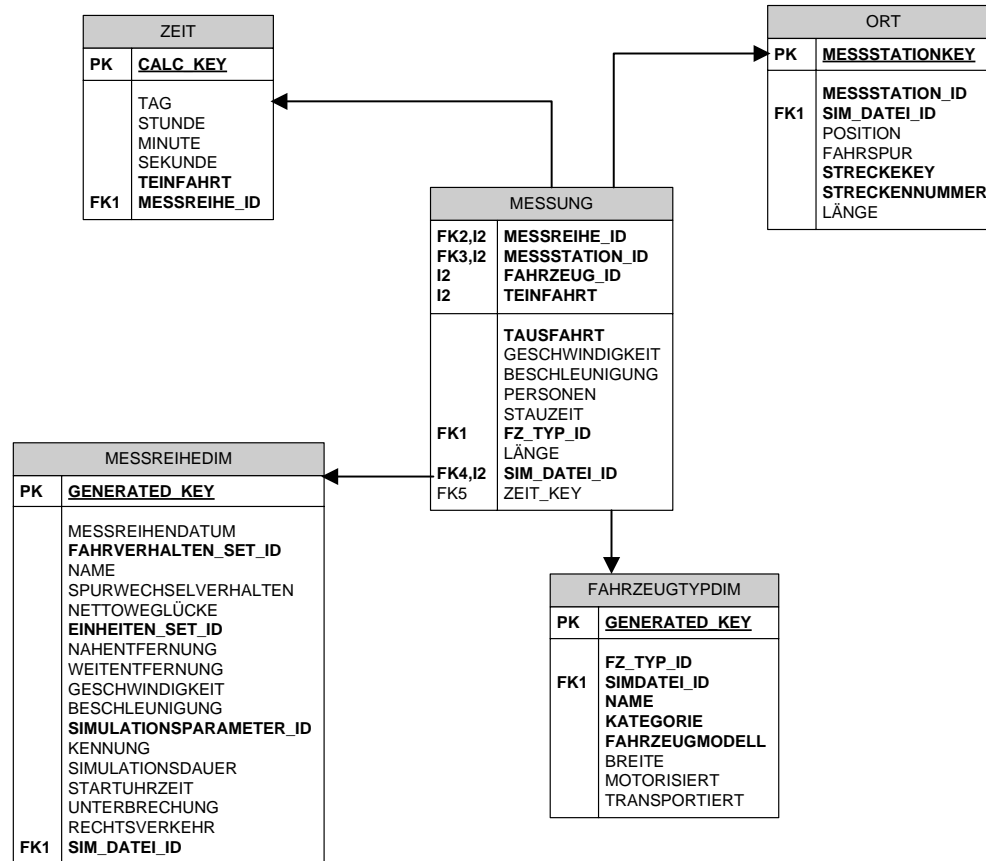
# Star Schema

- Ohne Berücksichtigung der Besonderheiten einer Simulation: Star Schema
- Mit Berücksichtigung dieser Besonderheiten:
  - „Simulationsdatei“ ist Top-Element jeder Dimensionstabelle
  - Daher ist die Orthogonalität der Dimensionstabellen unmöglich umzusetzen.





# Star Schema





# Roadmap

- Grundlagen
- Das Nutzdatenchema
- Das Füllen der Nutzdatenbank
- Das Data Warehouse Schema
- Anfragen an das Data Warehouse
  - Einfache SQL Anfragen
  - Anfragen mit OLAP Werkzeugen





# SQL Anfragen

Zur Verdeutlichung der Funktion der Dimensionstabelle Zeit:

Wie lange steht ein Fahrzeug im Feierabendverkehr zwischen 17:00 und 18:00 Uhr durchschnittlich im Stau?

Anfrage möglich über

- SQL Datentypen (Timestamp, Intervall)
- Eigene Repräsentation von Zeit





# OLAP

Anfragen mit dem OLAP Werkzeug von Oracle wurden bisher noch nicht durchgeführt.

Vorteile von OLAP Werkzeugen:

- Aggregierungsoperationen wie Roll Up oder Drill Down
- Auswerten von Dimensionen und Hierarchien





# Abschliessende Arbeiten

- Im Rahmen dieser Studienarbeit muss die Evaluation des Data Warehouse mittels OLAP Werkzeugen noch durchgeführt werden.
- Außerhalb dieser Studienarbeit könnten für die Implementierung des Data Warehouse Schemas Alternativen wie „Oracle Warehouse Builder“ oder Cognos Produkte herangezogen werden.





# Fragen – Anregungen?

Vielen Dank für Ihr Interesse!

