



Konzeption und Umsetzung eines Data Warehouse zur Erstellung von Ganglinien und zur Bewertung von Prognosen im Verkehr

Studienarbeit von Ling Kong
Betreuer: Heiko Schepperle

15. Oktober 2004

IPD - Institut für Programmstrukturen und Datenorganisation

 Universität Karlsruhe (TH)

Übersicht

1. Einführung
2. Daten im Verkehr
3. Verkehrsprognose
4. Data-Warehouse-Schema
5. OLAP-Einsatz
6. Zusammenfassung



Einführung: Hintergrund

- **Projekt OVID:** Selbstorganisationsfähigkeit im Verkehr durch I+K-gestützte Dienste
 - **Teilprojekt B1:** Verlässliche Datenbanken für die Informationsbereitstellung im Verkehr
- **Zielsetzung**
 - Ein Datenmodell auf Basis von Data-Warehouse-Technologien ist zu konzipieren, um sich zeitlich verändernde Verkehrsdaten als Ganglinien darzustellen und die Erzeugung weiterer Ganglinien zu unterstützen.
 - Die Bewertung von Prognose soll dadurch auch unterstützt werden.



Einführung: Aufgabe

- Analyse von Daten im Verkehr und Verkehrsprognose
- Untersuchung der Grundlagen für Data Warehouse und OLAP
- Konzeption eines Data-Warehouse-Schemas
- Umsetzung der Konzeption mit Oracle und dem OLAP-Werkzeug der Cognos Serie 7

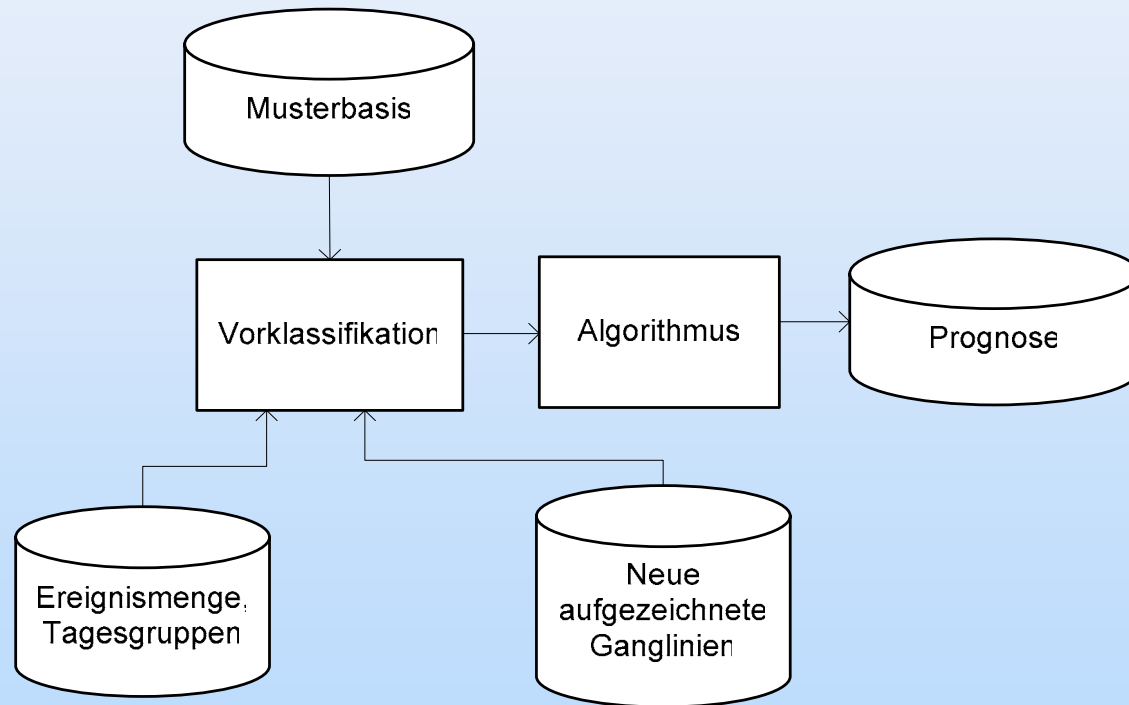


Daten im Verkehr

- **Infrastrukturdaten:** Basisinformationen über das Verkehrsnetz
- **Ereignisdaten und Störung:** Von Menschen beobachtete und aufgenommene Daten
- **Messdaten:** Beschreibung der Verkehrssituation, durch Sensoren ermittelt
 - Querschnittsmessdaten: durch stationäre Messstationen gesammelt
 - Floating Car Daten: durch Sensoren in FCD-Fahrzeugen gesammelt
 - Güterverkehrsdaten: Güterverkehrstransporte in Deutschland (für die Berichtsjahre 1999, 2001 und 2002)

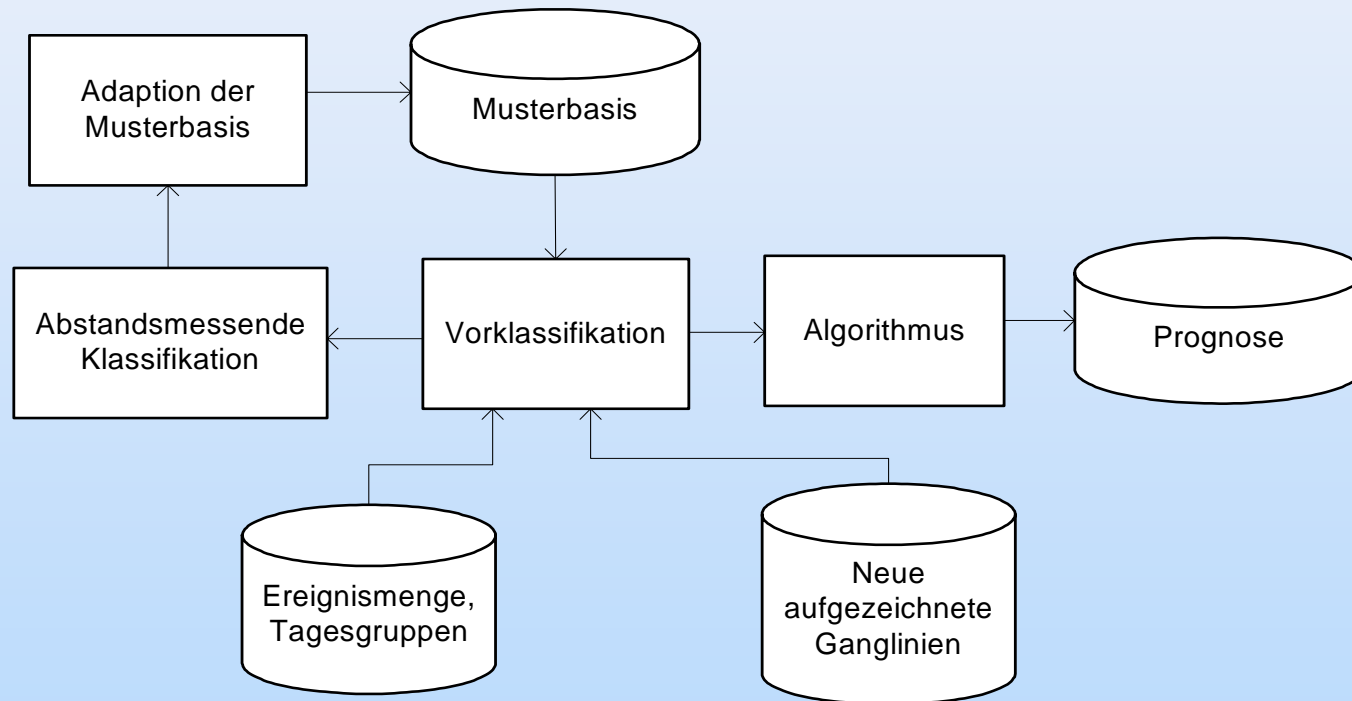
Verkehrsprognose

- Ein generelles Modell für Verkehrsprognose (adaptiert von [Wild96])



Verkehrsprognose

- Ein generelles Modell für Verkehrsprognose (adaptiert von [Wild96])



Übersicht

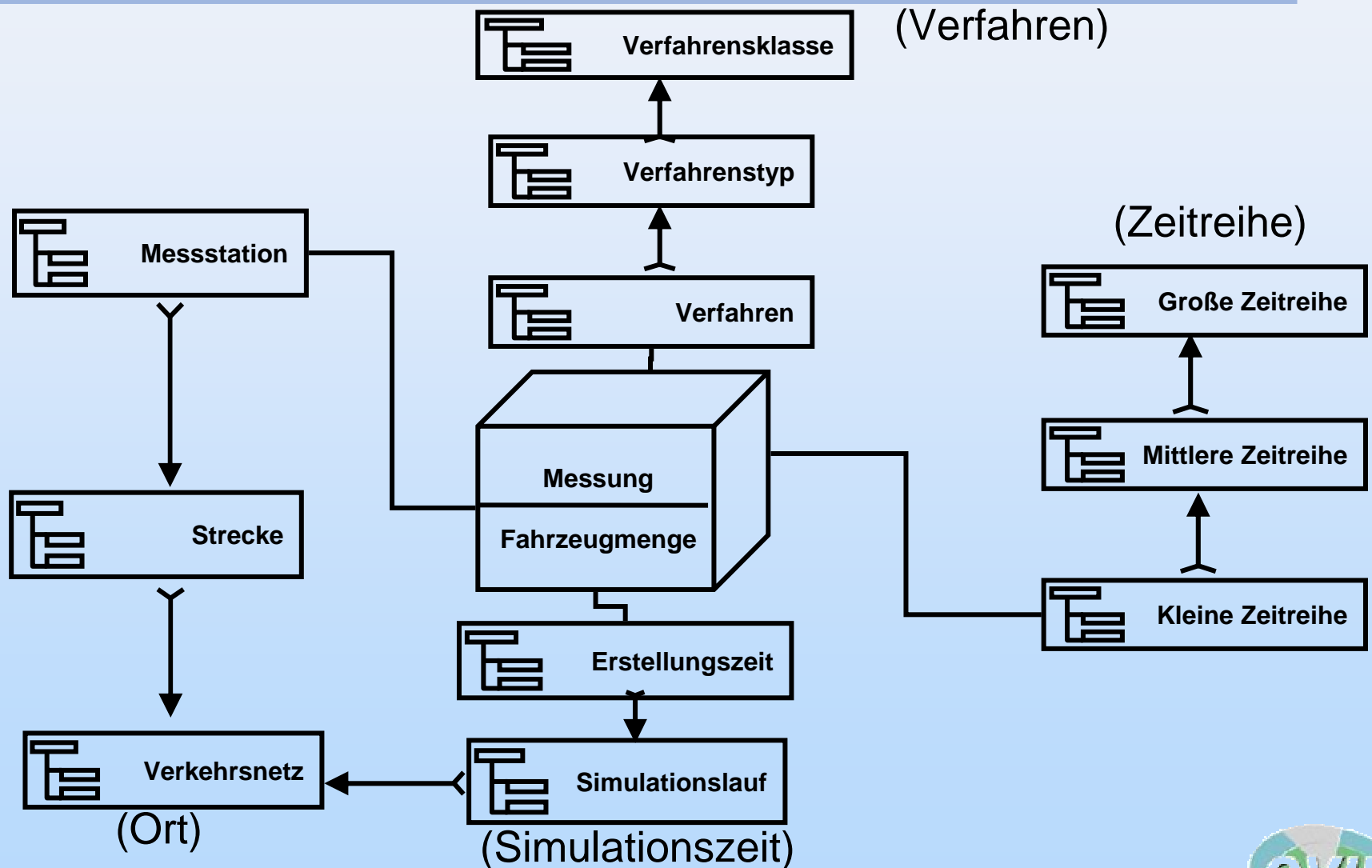
1. Einführung
2. Daten im Verkehr
3. Verkehrsprognose
4. Data-Warehouse-Schema
5. OLAP-Einsatz
6. Zusammenfassung



Data Warehouse System

- Zur Analyse und Auswertung
 - operative Daten
 - große Datenmengen
- Multidimensionales Datenmodell
 - Konzeptuelles Modell: z.B. mE/R-Modell
 - Logisches Modell: Star-Schema oder Snowflake-Schema
- Eine Fakten und mehrere Dimensionen
 - Fakt (Kennzahl, Kenngröße): quantifizierende Daten
 - Dimension: qualifizierende Daten
- Redundante Datenhaltung zur Beschleunigung von Datenabfrage

Konzeptuelles Modell

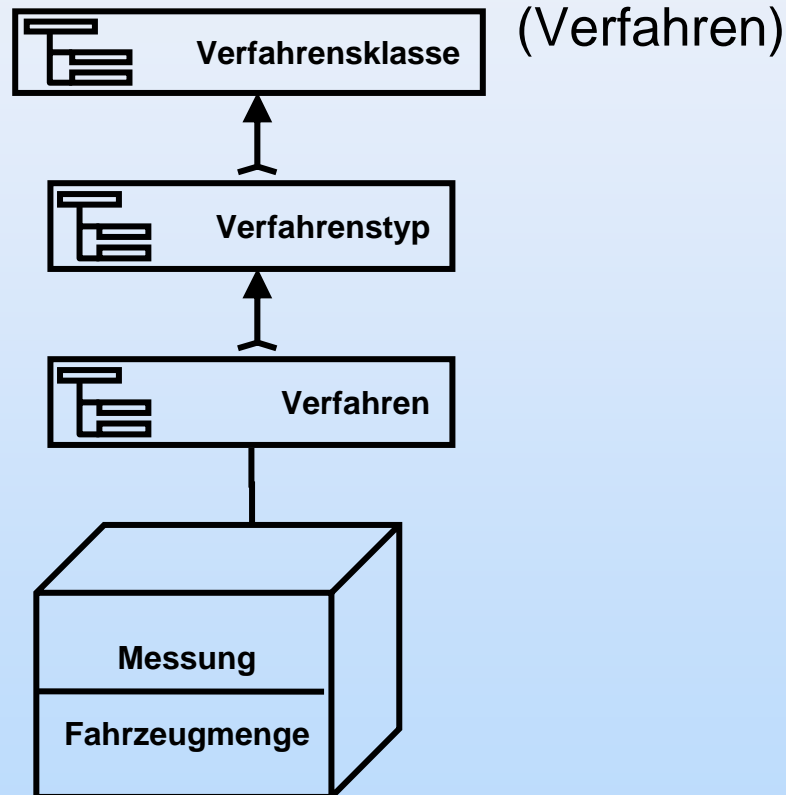


Konzeptuelles Modell - Verfahren

Erwartete Daten

Tatsächliche Daten

Prognosedaten



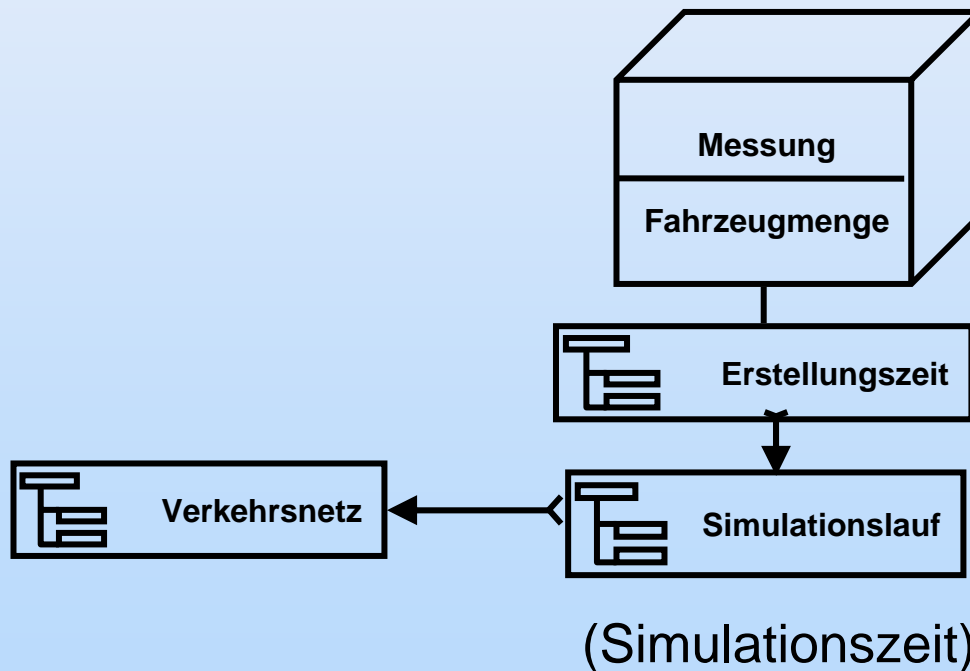
Konzeptuelles Modell - Simulationszeit

Erstellungszeit **01.11.2005 10:23**

ist nicht eindeutig!

Erstellungszeit **01.11.2005 10:23 Lauf 17**

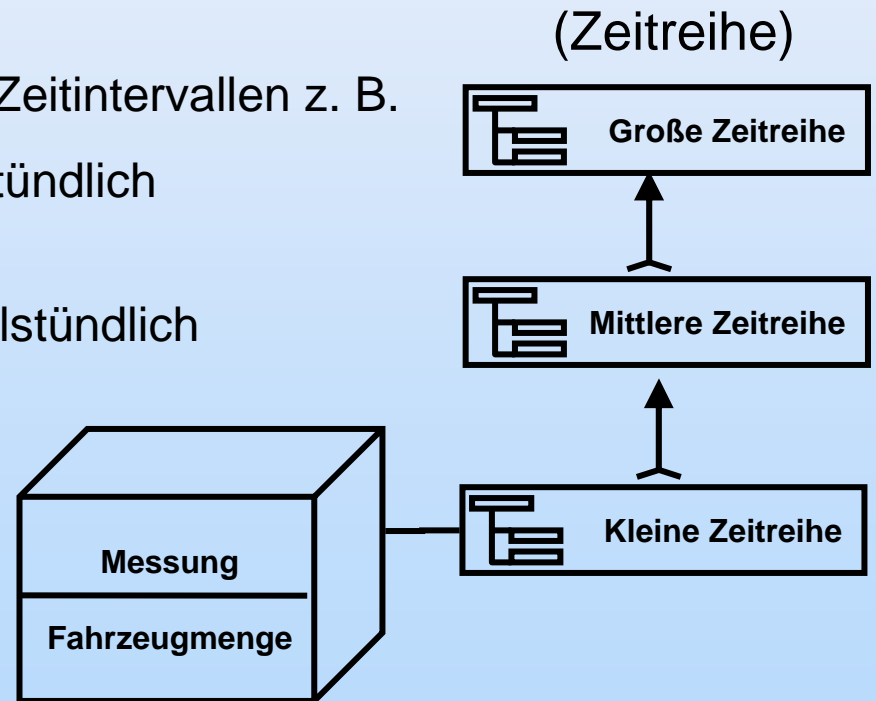
ist eindeutig!



Konzeptuelles Modell - Zeitreihe

Zwei Möglichkeiten zur Modellierung:

- über OLAP-Werkzeug (→ Werkzeug-Funktionalität)
- über Datenbankmodell
 - Zeitreihen mit unterschiedlichen Zeitintervallen z. B.
 - von 22:00 - 04:00 Uhr zweistündlich
 - von 06:00 – 09:00 Uhr und von 16:00 – 18:00 Uhr viertelstündlich
 - sonst stündlich

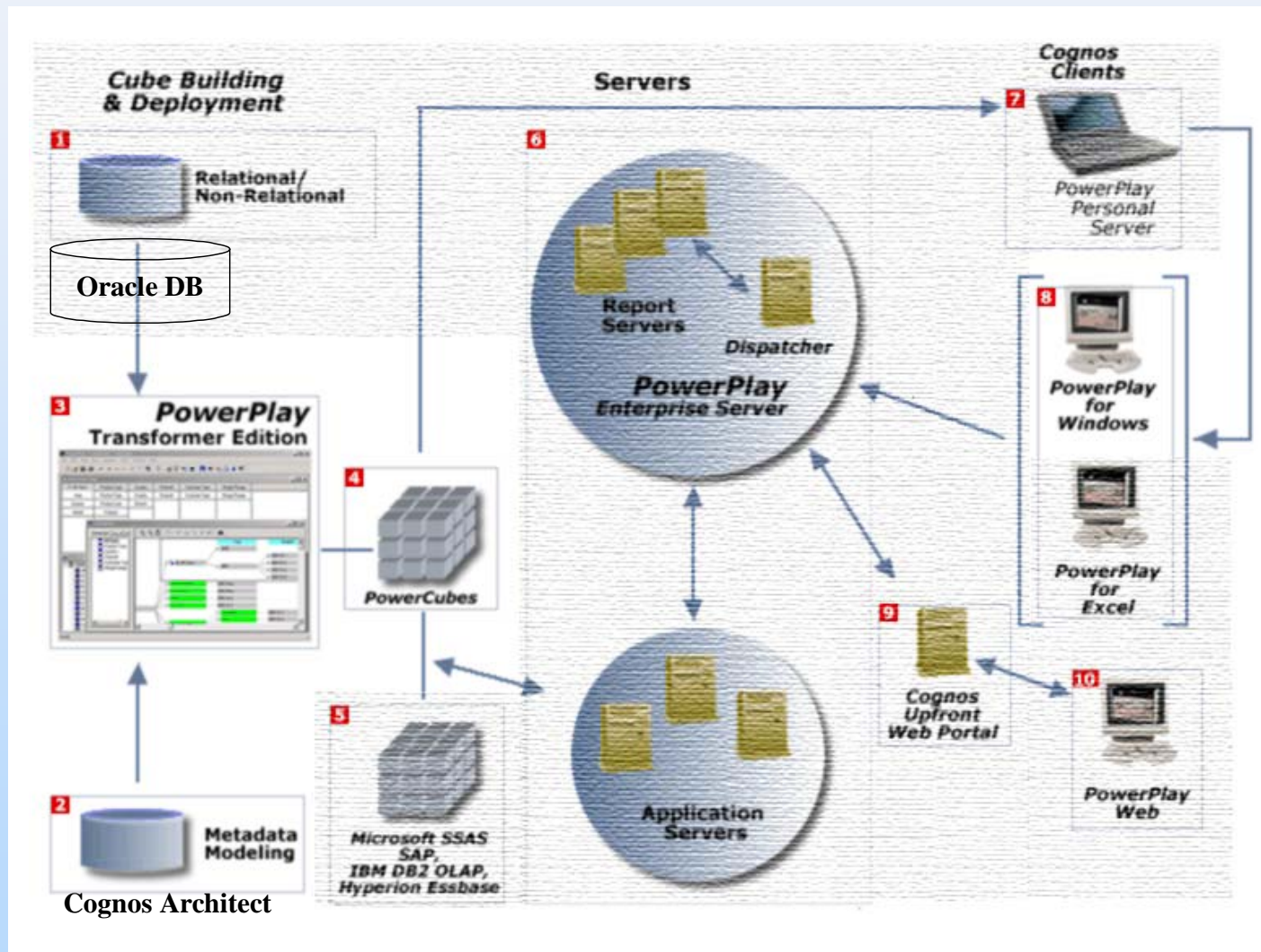


Übersicht

1. Einführung
2. Daten im Verkehr
3. Verkehrsprognose
4. Data-Warehouse-Schema
- 5. OLAP-Einsatz**
6. Zusammenfassung



Überblick: Cognos Series 7



Ganglinien: Kreuztabelle

PowerPlay - [PPlay1 von kong_modell_20040811 (Explorer)]

Datei Bearbeiten Ansicht Einfügen Analysieren Berechnen Format Extras Fenster Hilfe

3 MESSREIHE MESSSTATION 20040503 VON-BIS MENGE

d:\cognos_series7\cer2\bin\ko

- FLAG_ID
- MESSREIHE
- MESSSTATION
- SIMULATIONSdatum
- VON-BIS
- KENNZAHLEN

		13:00-14:00	14:00-15:00
Messdaten nach Gewichtungsfaktoren	21	72	14
n 4(0.1), 21(0.2), 41(0.3), 66(0.4)	22	42	12
	23	78	9
	24	50	0
	25	111	
	26		
	27		
	28		
	29		

Ganglinien: Kreuztabelle

PowerPlay - [PPlay1 von kong_modell_20040811 (Explorer)]

3 | MESSREIHE | MESSSTATION | 20040503 | VON-BIS | MENGE

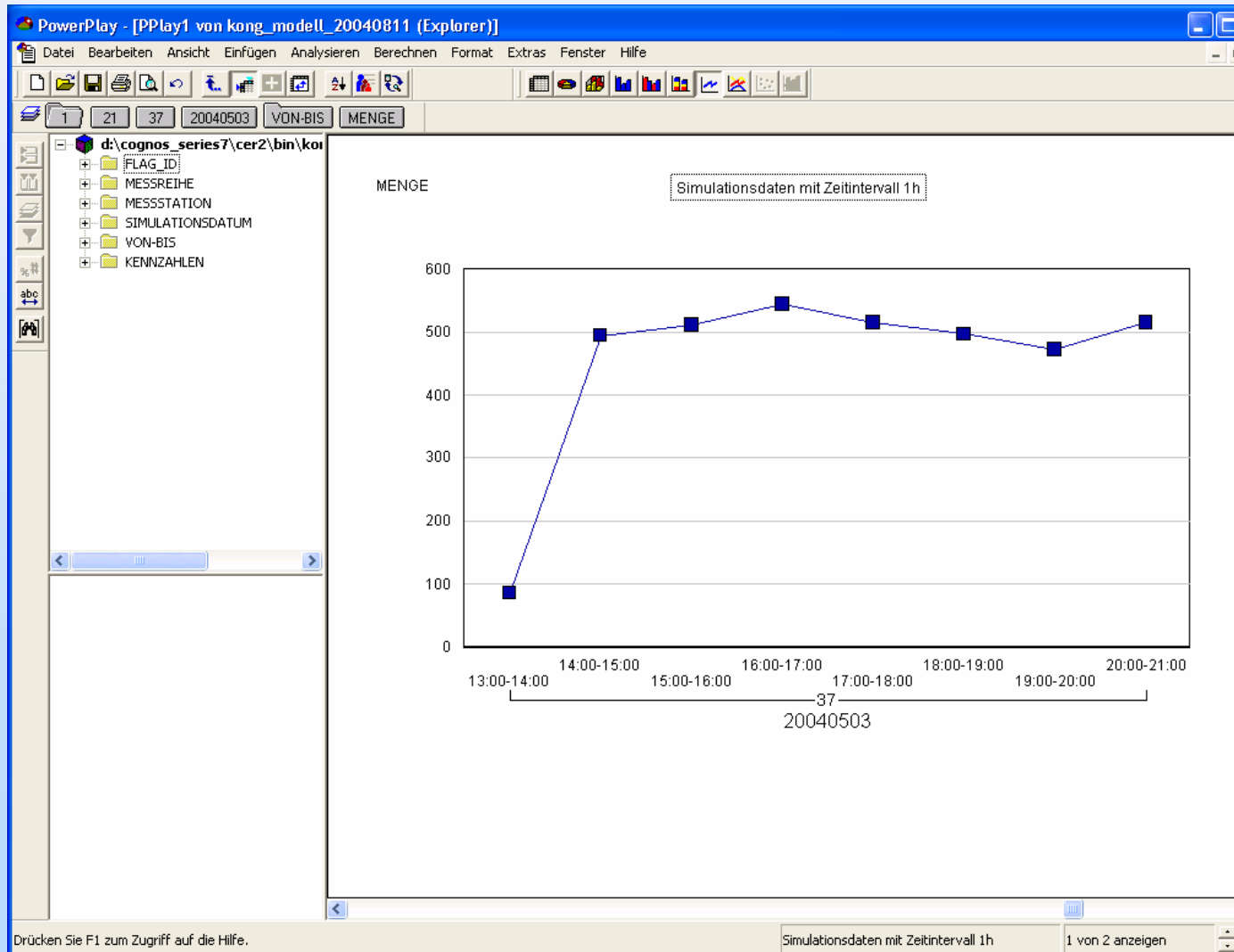
d:\cognos_series7\cer2\bin\koi

- FLAG_ID
- MESSREIHE
- MESSSTATION
- SIMULATIONSdatum
- VON-BIS
- KENNZAHLEN

		20040503					
		13:00-14:00	14:00-15:00	15:00-16:00	16:00-17:00	17:00-18:00	18:00-19:00
Messdaten	21	72	488	515	475	468	464
nach	22	14	99	115	108	99	104
Gewichtsfaktoren	23	42	203	178	202	212	221
n 4(0.1),	24	12	36	36	40	39	45
21(0.2), 41(0.3),	25	78	518	536	506	484	488
66(0.4)	26	9	68	90	81	78	79
	27	50	239	212	239	249	263
	28	0	4	2	0	3	3
	29	111	732	729	758	734	705
	30	13	117	86	102	122	103
	31	41	236	281	238	261	245
	32	90	522	527	486	528	500
	33	70	426	423	446	429	423
	34	21	135	135	159	153	117
	35	25	135	167	132	156	140
	36	84	493	517	473	501	486
	37	77	445	459	490	464	447
	38	24	140	126	141	142	122
	39	16	67	87	74	66	63
	40	115	695	742	686	717	711
	41	66	396	392	410	371	358
	42	99	522	481	504	497	508

Drücken Sie F1 zum Zugriff auf die Hilfe.

Ganglinien: Liniendiagramm



Zusammenfassung

- Ein Data-Warehouse-Schema wurde konzipiert, damit Messdaten im Verkehr als Ganglinien dargestellt werden können.
- Integrierte Behandlung verschiedener Datenarten: Erwartete Daten, tatsächliche Daten und Prognosedaten können mit dem Schema integriert dargestellt werden. Dadurch wird die Bewertung von Prognosen möglich.
- Das Data-Warehouse-Schema wurde mit Oracle 9i umgesetzt.
- Ganglinien-Abfragen wurde mit OLAP-Werkzeug der Cognos Serie 7 erstellt.



Vielen Dank für Ihre Aufmerksamkeit!

