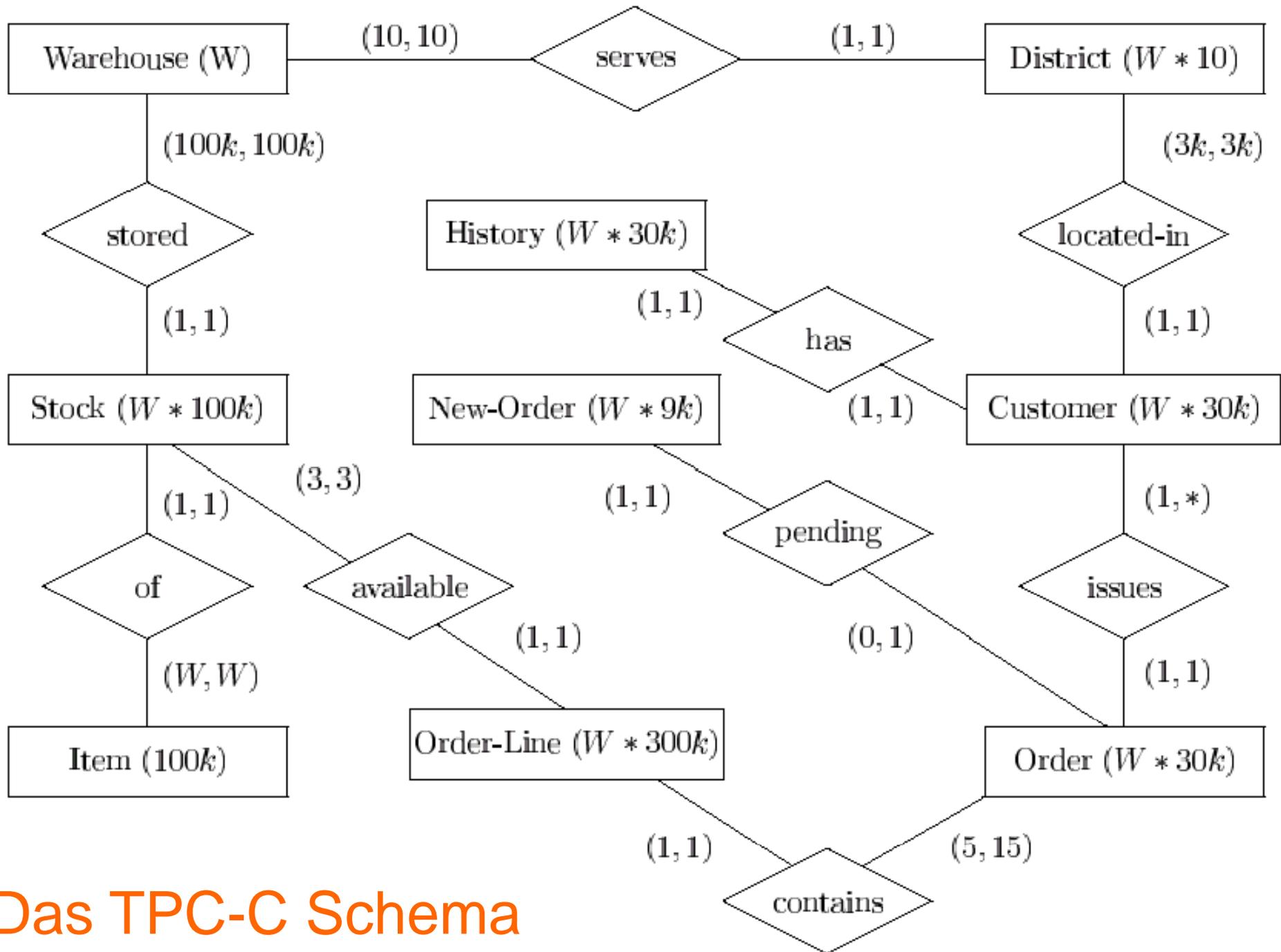


Leistungsbewertung

- TPC-C (online transaction processing OLTP)
- TPC-H/R (online analytical processing OLAP)
- OO7 (objektorientierte Datenbanken)
- TPC-W (Web-Datenbanken)
- Xmark: XML/Xquery-Benchmark
- Neue TPC-Benchmarks
 - TPC-E (ersetzt TPC-C)
 - TPC-App
 - TPC-DS

TPC-C (online transaction processing OLTP Benchmark)

- Kurze Transaktionen
- Wie sie im alltäglichen Betrieb eines Unternehmens anfallen
 - Mission-critical
- Beispielanwendung: Handelsunternehmen
 - Transaktionen sind z.B.
 - Aufgeben einer Bestellung
 - Abarbeiten einer Bestellung
 - ...



Das TPC-C Schema

Erläuterungen zum Schema

- Warehouse: Es werden $W \geq 1$ Warenhäuser durch je ein Tupel modelliert.
- District: Pro Warenhaus gibt es 10 Distrikte, deren Kunden vornehmlich (wenn die bestellten Waren vorhanden sind) von dem zugehörigen Warenhaus beliefert werden.
- Customer: In jedem Distrikt gibt es 3000 (3k) Kunden.
- Order: In der Anfangskonfiguration hat jeder Kunde bereits eine Bestellung aufgegeben. Es kommen dann im Laufe der Benchmark-Durchführung neue Bestellungen hinzu und ausstehende (engl. pending) Bestellungen werden kontinuierlich abgearbeitet.
- New-Order: Eine neu aufgenommene Bestellung wird bis zur Belieferung in dieser Relation eingetragen. Genauer gesagt, die Tupel dieser Relation stellen Verweise auf noch nicht abgearbeitete Einträge in Order dar.

Erläuterungen zum Schema

- Order-Line: Jede Bestellung besteht aus durchschnittlich zehn (variierend zwischen fünf bis fünfzehn) Auftragspositionen.
- Stock: Diese Relation modelliert die Verfügbarkeit von Produkten in den einzelnen Warenhäusern. Stock enthält pro (Warenhaus, Produkt)-Paar einen Eintrag -- also $W \times 100k$ Tupel. Eine Auftragsposition wird aus dem Warenbestand (Stock) eines Warenhauses abgedeckt, was durch die Beziehung 'available' modelliert wird.
- Item: Diese Relation enthält ein Tupel für jedes der 100000 Produkte (Item), die das Handelsunternehmen anbietet. Die Relation
- Item nimmt bei der Skalierung der Datenbasis eine Sonderstellung ein; sie wird in der Größe nicht verändert, auch wenn die Anzahl der Warenhäuser (W) erhöht wird.
- History: Diese Relation enthält Daten zur Bestellhistorie der einzelnen Kunden.

Die fünf Transaktionen

- New-Order: In dieser Transaktion wird eine komplette Neubestellung von fünf bis fünfzehn Auftragspositionen in die Datenbasis eingegeben. Für jede dieser Auftragspositionen wird die Verfügbarkeit des jeweiligen Produkts in der Stock-Relation überprüft.
- Payment: Die Zahlung eines Kunden wird verbucht. Dazu werden zusätzlich Verkaufsstatistiken in den Relationen District und Warehouse fortgeschrieben.

Die fünf Transaktionen

- Order-Status: Dies ist eine reine Lesetransaktion, in der der Status der letzten Bestellung eines bestimmten Kunden überprüft wird.
- Delivery: In dieser Transaktion werden zehn Bestellungen aus der New-Order Relation im Batch-Modus (also ohne Benutzerinteraktion) bearbeitet. Die bearbeiteten Bestellungen werden aus der New-Order Relation entfernt.
- Stock-Level: Dies ist eine Lesetransaktion, die den Warenbestand der in letzter Zeit bestellten Produkte kontrolliert. Der TPC-C-Benchmark erlaubt die Aufspaltung dieser eine große Anzahl von Tupeln lesenden Transaktion in kleinere Datenbank-Transaktionen, um dadurch den Overhead der Mehrbenutzersynchronisation zu reduzieren.

Leistungsgrößen

- Die Transaktion New-Order stellt das "Rückgrat" des TPC-C Benchmarks dar.
- Die Leistungsfähigkeit des Systems wird in der Anzahl der pro Minute abgearbeiteten New-Order-Transaktionen angegeben, wobei natürlich pro New-Order auch eine bestimmte Anzahl der anderen vier Transaktionen gleichzeitig ausgeführt werden muss.
- Weiterhin verlangt der Benchmark, dass 90% der vier erstgenannten Transaktionen eine Antwortzeit von unter fünf Sekunden haben müssen.
- Die Stock-Level-Transaktion muss in 90% der Fälle innerhalb von 20 Sekunden abgearbeitet sein.

Leistungsgrößen

- Der TPC-C-Benchmark hat zwei Leistungskriterien:
 - tpmC: Der Durchsatz von New-Order-Transaktionen pro Minute.
 - Preis/Leistungsverhältnis: Hierzu wird der Gesamtsystempreis, der sich aus Hardware, Software und Softwarewartung für fünf Jahre berechnet, im Verhältnis zum Durchsatz (tpmC) angegeben. Das Leistungsmaß ist dann x Dollar pro Transaktion.
- Bei heutigen Hardware und Softwarekonfigurationen sind folgende Kennzahlen möglich:
 - 2.500 Transaktionen pro Minute bei einem Systempreis von ca. 200.000 US Dollar (also etwa 70 Dollar pro Transaktion im Preis/Leistungsverhältnis)
 - 23.000 Transaktionen pro Minute bei einem Systempreis von ca. 2.750.000 US Dollar (also etwa 120 Dollar pro Transaktion im Preis/Leistungsverhältnis)

Leistungsgrößen

- Man beachte, dass bei beiden Konfigurationen die Hardwarekosten den Systempreis dominieren; die Datenbanksoftware macht i.a. nur einen geringen Prozentsatz des Systempreises aus (meist weniger als 10 %).
 - Diese zwei DBMS-Konfigurationen stellen die beiden Extreme dar: (1) günstiges Preis/Leistungsverhältnis* für eine kleine Konfiguration und (2) hohe Leistungsfähigkeit zu einem entsprechend hohen Preis.
 - Man kann viele weitere Benchmarkergebnisse, in denen auch "Ross und Reiter" genannt sind, über die Webseiten der TPC-Organisation www.tpc.org beziehen -- siehe Literatur.
- * Für eine etwa doppelt so teure Konfiguration (also 400.000 \$) gibt es sogar schon ein noch günstigeres Preis/Leistungsverhältnis von unter 50 Dollar pro Transaktion.

Top Ten TPC-C by Performance

Version 5 Results As of 22-Nov-2009 8:36 AM [GMT]

Note 1: The TPC believes it is not valid to compare prices or price/performance of results in different currencies.

All Results
 Clustered Results
 Non-Clustered Results

Currency

| Rank | Company | System | tpmC | Price/tpmC | System Availability | Data |
|------|---|---|-----------|------------|---------------------|-------------------------------------|
| 1 |  | Sun SPARC Enterprise T5440 Server Cluster | 7,646,486 | 2.36 USD | 12/14/09 | Oracle Database 11g Ent. Ed. w/Re |
| 2 |  | IBM Power 595 Server Model 9119-FHA | 6,085,166 | 2.81 USD | 12/10/08 | IBM DB2 9.5 |
| *** |  | Bull Escala PL6460R | 6,085,166 | 2.81 USD | 12/15/08 | IBM DB2 9.5 |
| 3 |  | HP Integrity Superdome-Itanium2/1.6GHz/24MB iL3 | 4,092,799 | 2.93 USD | 08/06/07 | Oracle Database 10g R2 Enterprise |
| 4 |  | IBM System p5 595 | 4,033,378 | 2.97 USD | 01/22/07 | IBM DB2 9 |
| 5 |  | IBM eServer p5 595 | 3,210,540 | 5.07 USD | 05/14/05 | IBM DB2 UDB 8.2 |
| 6 |  | PRIMEQUEST 580A 32p/64c | 2,382,032 | 3.76 USD | 12/04/08 | Oracle Database 10g R2 Enterprise |
| 7 |  | PRIMEQUEST 580 32p/64c | 2,196,268 | 4.70 USD | 04/30/08 | Oracle 10g Enterprise Ed R2 w/ Part |
| 8 |  | IBM System p 570 | 1,616,162 | 3.54 USD | 11/21/07 | IBM DB2 Enterprise 9 |
| *** |  | Bull Escala PL1660R | 1,616,162 | 3.54 USD | 12/16/07 | IBM DB2 9.1 |
| 9 |  | IBM eServer p5 595 | 1,601,784 | 5.05 USD | 04/20/05 | Oracle Database 10g Enterprise Edi |
| 10 |  | PRIMEQUEST 540A 16p/32c | 1,354,086 | 3.25 USD | 11/22/08 | Oracle Database 10g release2 Ente |

*** - Duplicate results are shown with an asterisk (*) in the Rank column. [Click here](#) for more information about duplicates.

Top Ten TPC-C by Price/Performance

Version 5 Results As of 22-Nov-2009 8:38 AM [GMT]

Note 1: The TPC believes it is not valid to compare prices or price/performance of results in different currencies.

All Results
 Clustered Results
 Non-Clustered Results
 Currency United States - Dollar (USD) ▾

| Rank | Company | System | tpmC | Price/tpmC | System Availability | Database |
|------|--|-------------------------------|---------|------------|---------------------|---|
| 1 |  DELL | Dell PowerEdge T710 | 239,392 | .50 USD | 11/18/09 | Oracle Database 11g Standard Edition One |
| 2 |  HP invent | HP ProLiant ML350 G6 | 232,002 | .54 USD | 05/21/09 | Oracle Database 11g Standard Edition One |
| 3 |  DELL | Dell PowerEdge 2900 | 104,492 | .60 USD | 02/20/09 | Oracle Database 11g Standard Edition One |
| 4 |  DELL | Dell PowerEdge 2900 | 97,083 | .68 USD | 06/16/08 | Oracle Database 11g Standard Edition One |
| 5 |  HP invent | HP ProLiant ML350G5 | 102,454 | .73 USD | 12/31/07 | Oracle Database 11g Standard Edition One |
| 6 |  HP invent | HP ProLiant ML350G5 | 100,926 | .74 USD | 06/08/07 | Oracle Database 10g Standard Edition One |
| 7 |  HP invent | HP ProLiant ML350G5 | 82,774 | .84 USD | 03/27/07 | Microsoft SQL Server 2005 x64 Enterprise Edt. SP1 |
| 8 |  SYBASE Anywhere | Dell PowerEdge 2950 III | 20,705 | .85 USD | 08/05/08 | Sybase SQL Anywhere 11.0 |
| 9 |  DELL | PowerEdge 2900/1/2.33GHz/2x4M | 69,564 | .91 USD | 03/09/07 | Microsoft SQL Server 2005 Standard Ed. |
| 10 |  HP invent | HP ProLiant DL585G5/2.7GHz | 579,814 | .96 USD | 11/17/08 | Microsoft SQL Server 2005 x64 Enterprise Edt SP2 |

Auswertungsbericht des derzeit (Okt 2009) leistungsfähigsten Systems

- PDF-Datei [Sun_T5440_TPC-C_Cluster_ES_110309.pdf](#)
- http://www.tpc.org/results/individual_results/Sun/Sun_T5440_TPC-C_Cluster_ES_110309.pdf

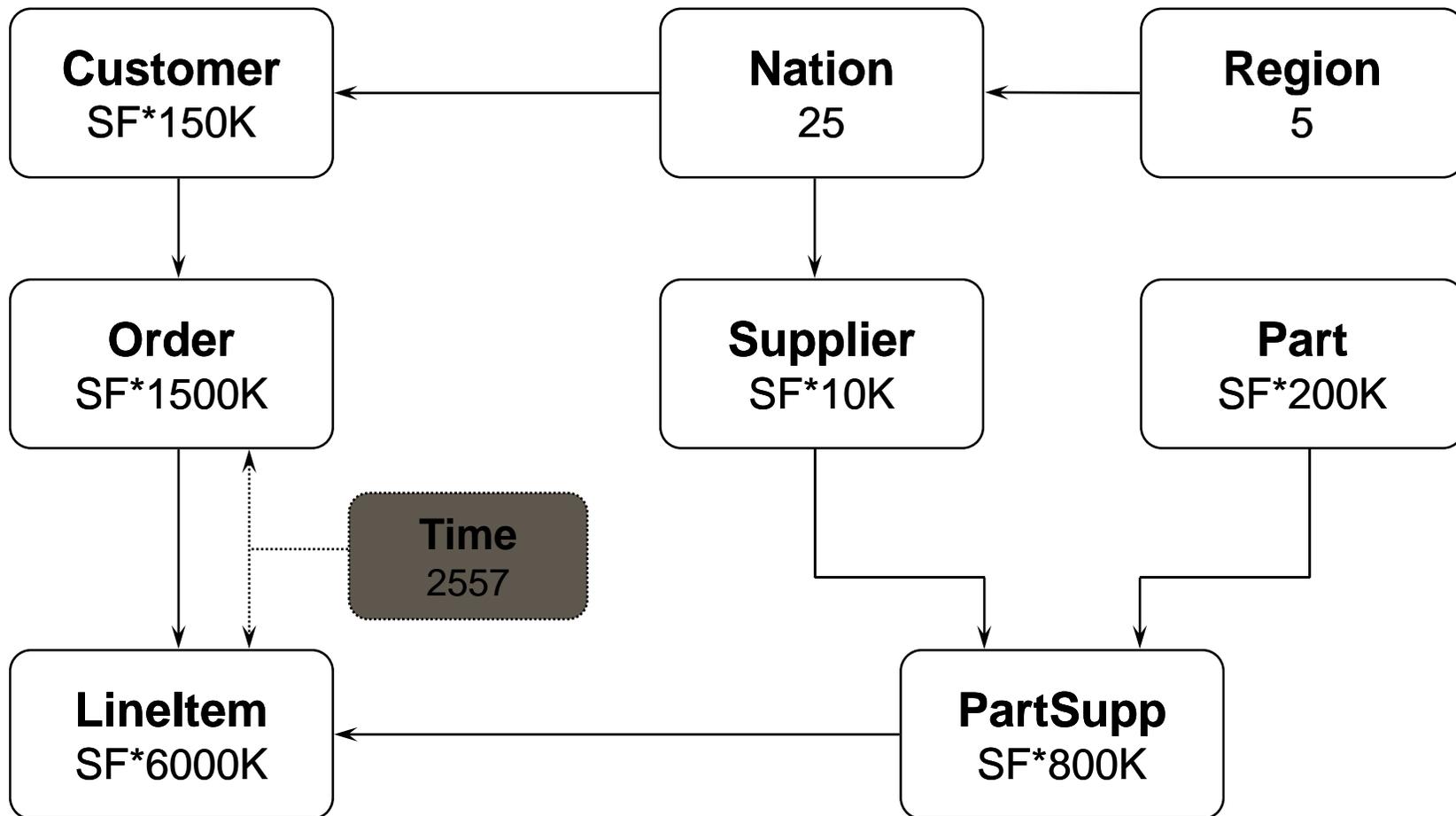
TPC H/R: Online Analytical Processing OLAP-Benchmark

- Decision-Support-Anwendungen
 - Entscheidungsunterstützende Systeme
- Anfrage-lastig
- Wenige Updates

- TPC-H
 - Ad Hoc Queries
 - Keine Anfragespezifischen Optimierungen

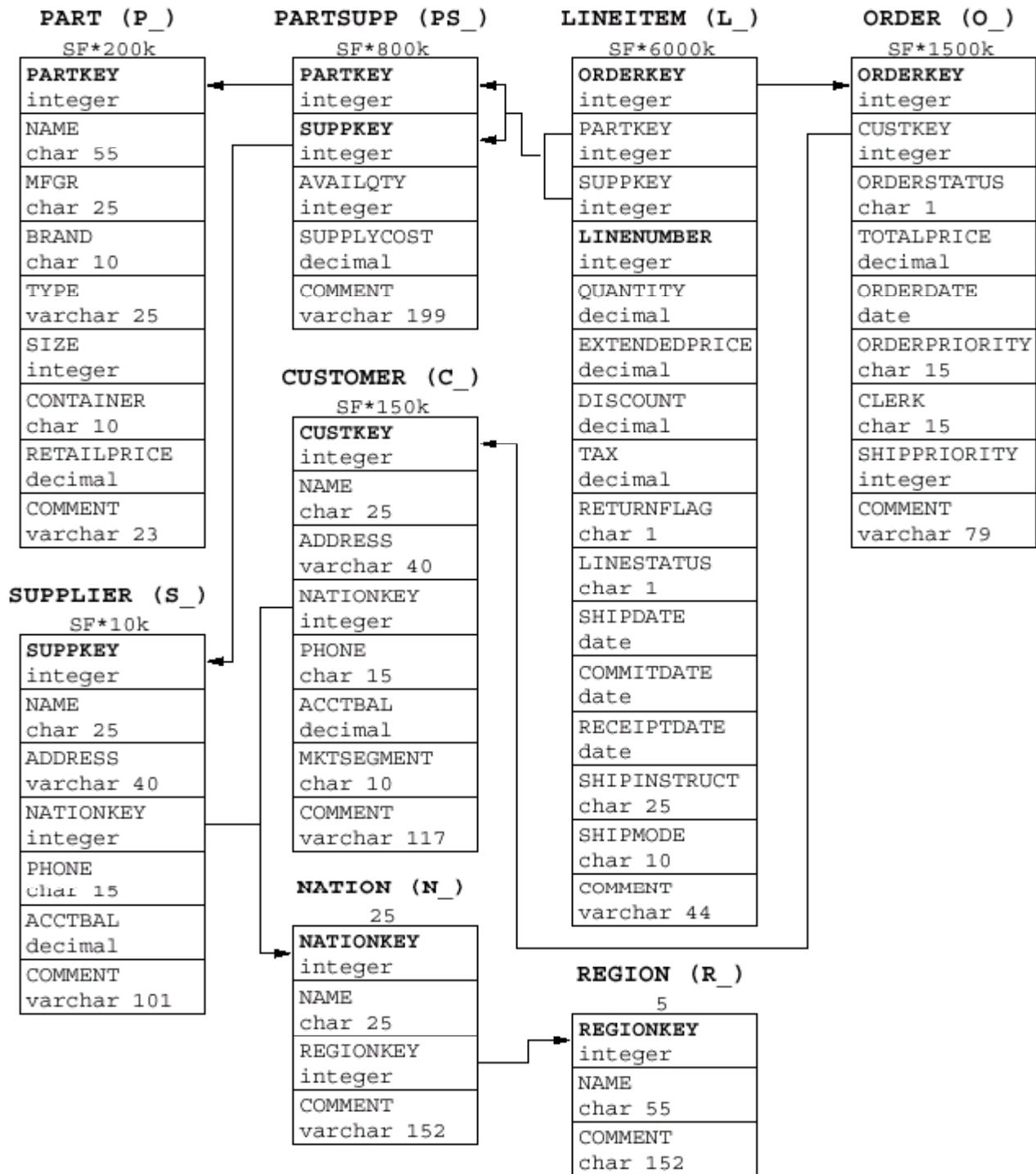
- TPC-R -- OBSOLET
 - Business Reporting
 - Anfragespezifische Optimierungen
 - dennoch: Anfragen sind parametrisiert
 - Sehr viele Parameter-Kombinationen

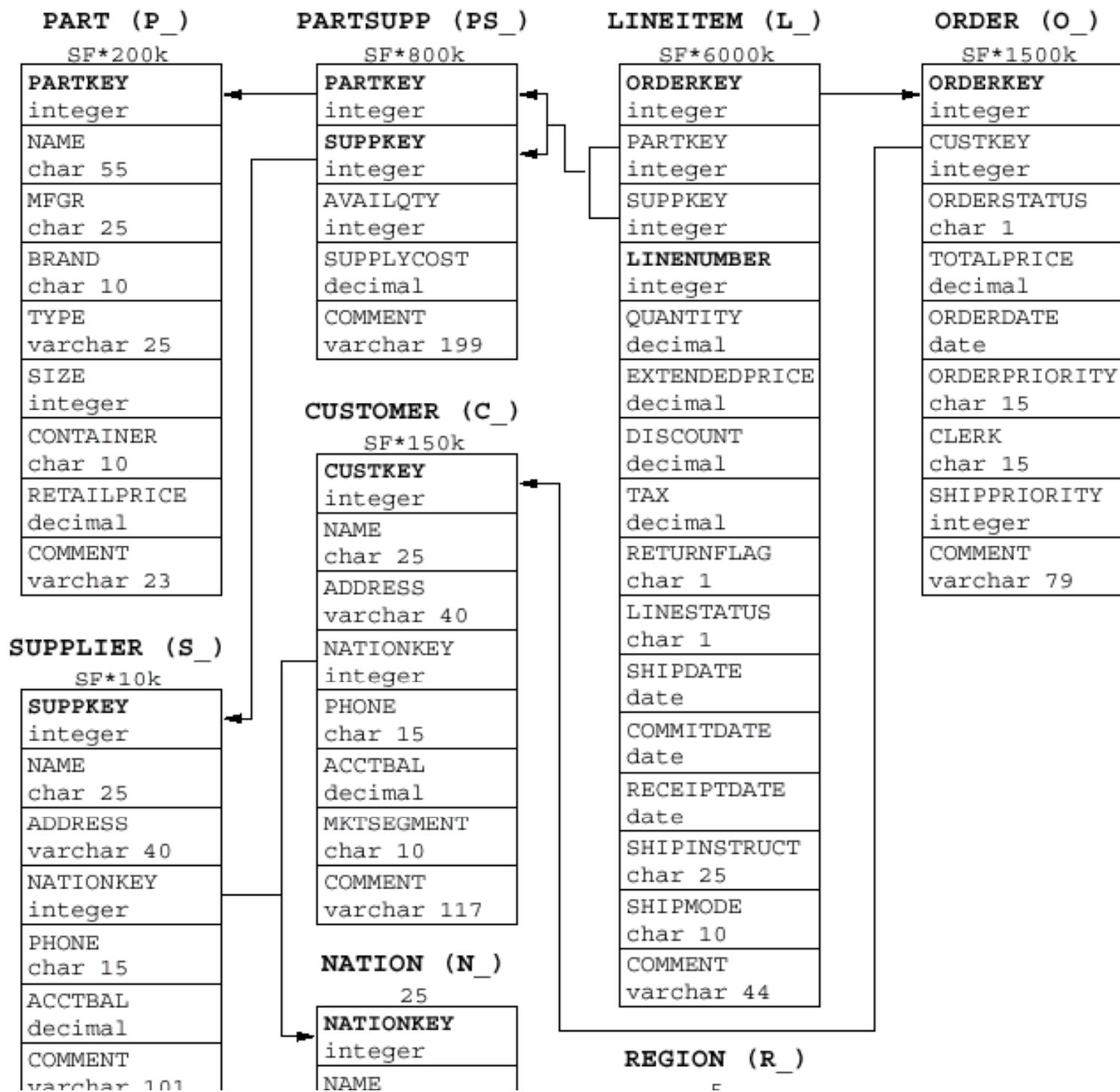
TPC-D Schema



Legend:

- Arrows point in the direction of one-to-many relationships.
- The value below each table name is its cardinality. SF is the Scale Factor.
- The Time table is optional. So far, not used by anyone.





TPC-H/R-Anfragen

Q1

Man erstelle einen aufsummierten Preisbericht über alle Auftragspositionen, die spätestens 90 Tage vor dem 1. Dezember 1998 versandt wurden. Die Ausgabe soll nach RETURNFLAG und LINESTATUS gruppiert und in aufsteigender Reihenfolge nach diesen Attributen sortiert werden. Für jede Gruppe soll die gesamte Menge, der Gesamtpreis, der ermäßigte Gesamtpreis, der ermäßigte Gesamtpreis inklusive Steuern, die durchschnittliche Anzahl, der durchschnittliche Gesamtpreis und der durchschnittliche Nachlass und die Anzahl der Auftragspositionen aufgelistet werden.

Q2

Für jedes Teil aus Messing (engl. brass) mit Größe 15 soll festgestellt werden, welcher Zulieferer in Europa beim nächsten Auftrag ausgewählt werden sollte. Das Kriterium für die Wahl eines Lieferanten sind dabei minimale Lieferkosten. Die Anfrage soll für jeden qualifizierenden Lieferanten den Kontostand, Namen, Land, Teilenummer, Hersteller des Teils, sowie Adresse und Telefonnummer des Lieferanten auflisten.

TPC-H/R-Anfragen

Q3

Man berechne den möglichen Umsatz aus den Aufträgen aus dem Marktsegment "Gebäude" (engl. building), die am 15. März 1995 noch nicht (vollständig) versandt waren. Die 10 Aufträge, die durch Auslieferung der ausstehenden Auftragspositionen den höchsten Umsatz ergeben und deren Lieferpriorität sollen ausgegeben werden.

Q4

Mit Hilfe dieser Anfrage soll überprüft werden, wie gut das Auftragsprioritätensystem funktioniert. Zusätzlich liefert sie eine Einschätzung über die Zufriedenstellung der Kunden. Dazu zählt die Anfrage die Aufträge im dritten Quartal 1993, bei denen wenigstens eine Auftragsposition nach dem zugesagten Liefertermin zugestellt wurde. Die Ausgabeliste soll die Anzahl dieser Aufträge je Priorität sortiert in aufsteigender Reihenfolge enthalten.

TPC-H/R-Anfragen

Q5

Für jedes Land in Asien sollen die Einnahmen aufgelistet werden, die aus Auftragspositionen resultieren, bei denen die Kunden und die dazugehörigen Lieferanten beide aus dem gleichen Land stammen. Anhand dieser Ergebnisse kann festgestellt werden, ob es sich lohnt, in einem bestimmten Gebiet lokale Verteilungszentren einzurichten. Dabei werden nur Aufträge aus dem Jahr 1994 berücksichtigt.

Q6

Es soll berechnet werden, um wie viel sich die Einnahmen erhöht hätten, wenn ein gewährter Nachlass von 5 bis 7 % für Mengen von weniger als 24 Teilen für im Jahr 1994 verschickte Aufträge gestrichen worden wäre.

Sample Query Definition

2.3 Forecasting Revenue Query (Q6)

This query quantifies the amount of revenue increase that would have resulted from eliminating company-wide discounts in a given percentage range in a given year. Asking this type of “what if” query can be used to look for ways to increase revenues.

2.3.1 Business Question

The Forecasting Revenue Change Query considers all the lineitems shipped in a given year with discounts between $\text{DISCOUNT} + 0.01$ and $\text{DISCOUNT} - 0.01$. The query list the amount by which the total revenues would have decreased if these discounts had been eliminated for lineitems with item quantities less than QUANTITY . Note that the potential revenue increase is equal to the sum of $(\text{L_EXTENDEDPRICE} * \text{L_DISCOUNT})$ for all lineitems with quantities and discounts in the qualifying range.

2.3.2 Functional Query Definition

```
SELECT SUM(L_EXTENDEDPRICE*L_DISCOUNT) AS REVENUE FROM LINEITEM  
WHERE L_SHIPDATE >= DATE '[DATE]'  
AND L_SHIPDATE < DATE '[DATE]' + INTERVAL '1' YEAR  
AND L_DISCOUNT BETWEEN [DISCOUNT] - 0.01 AND [DISCOUNT] + 0.01  
AND L_QUANTITY < [QUANTITY]
```

2.8.3 Substitution Parameters

Values for the following substitution parameters must be generated and used to build the executable query text.

1. **DATE** is the first of January of a randomly selected year within [1993-1997]
2. **DISCOUNT** is randomly selected within [0.02 .. 0.09]
3. **QUANTITY** is randomly selected within [24 .. 25]

Sample Query Definition (cont.)

2.8.4 Query Validation

For validation against the qualification database the query must be executed using the following values for the substitution parameters and must produce the following output:

Values for substitution parameters:

- 1. DATE = 1994-01-01**
- 2. DISCOUNT = 0.06**
- 3. QUANTITY = 24**

Query validation output data:

1 row returned

| |
|--------------------|
| REVENUE |
| 11450588.04 |

- Query validation demonstrates the integrity of an implementation
 - Query phrasings are run against 100MB data set
 - Data set must mimic the design of the test data base
 - Answers sets must match those in the specification almost exactly
- If the answer sets don't match, the benchmark is invalid!**

TPC-H/R-Anfragen

Q7

Zur Unterstützung bei der Verhandlung über neue Lieferverträge soll der Wert der zwischen Frankreich und Deutschland transportierten Güter festgestellt werden. Dazu werden jeweils die rabattierten Einnahmen in den Jahren 1995 und 1996 berechnet, die aus Auftragspositionen resultieren, bei denen der Lieferant aus dem einen, und der Kunde aus dem anderen Land stammt (also vier Ergebnistupel).

Q8

Es soll der Marktanteil Brasiliens innerhalb der Region Amerika für den Teiletyp % "economy anodized steel" "STANDARD POLISHED TIN" in den Jahren 1995 und 1996 (ausschlaggebend ist das Bestelldatum) berechnet werden. Der Marktanteil Brasiliens ist definiert als der Anteil am Gesamtumsatz, welcher durch Produkte dieses speziellen Typs, geliefert von einem brasilianischen Lieferanten, in Amerika erzielt wurde.

Q9

Man ermittle den durch eine bestimmte Produktlinie erzielten Gewinn, aufgeschlüsselt nach Zuliefererland und Jahr der Bestellung. Die zu untersuchende Produktlinie besteht aus allen Teilen, die den Teilstring "green" in ihrem Namen enthalten.

TPC-H/R-Anfragen

Q10

Es werden die 20 Kunden gesucht, die durch Rücksendungen (Reklamationen, *RETURNFLAG*= 'R') den größten Umsatzverlust im vierten Quartal 1993 verursacht haben. Es werden dabei nur Produkte berücksichtigt, die auch in diesem Quartal bestellt wurden. Man liste jeweils Nummer und Namen des Kunden, Umsatz durch diesen Kunden, Kontostand, Land, sowie Adresse und Telefonnummer des Kunden auf.

Q11

Man finde durch Überprüfung der Lagerbestände der Lieferanten in Deutschland diejenigen Teile heraus, die einen signifikanten Anteil (mindestens 0,1 %) am Gesamtwert aller verfügbaren Teile in Deutschland darstellen. Man liste Teilenummer und Wert des Lagerbestandes auf, sortiert nach absteigendem Wert.

TPC-H/R-Anfragen

Q12

Diese Anfrage soll feststellen, ob die Verwendung von billigeren Lieferarten kritische Aufträge negativ beeinflusst, und zwar in der Form, dass den Kunden mehrere Produkte erst nach dem zugesagten Datum zugeschickt werden. Zu diesem Zweck zählt die Anfrage für die beiden Lieferarten "MAIL" und "SHIP" und getrennt nach den Prioritätskategorien "hoch" (HIGH, URGENT) und "niedrig" (alle übrigen) all die Auftragspositionen, welche die Kunden im Laufe des Jahres 1994 tatsächlich erhielten, und die zu einem Auftrag gehören, bei dem das RECEIPTDATE das COMMITDATE überschreitet, obwohl die Auftragsposition spätestens einen Tag vor dem angesetzten Liefertermin losgeschickt wurde.

Q13

Man ermittle aufgeschlüsselt nach Jahren die Umsatzverluste, die der Sachbearbeiter (engl. CLERK) Nr. 88 durch zurückgeschickte Aufträge (*RETURNFLAG= 'R'*) verursacht hat.

TPC-H/R-Anfragen

Q14

Die Resonanz des Marktes auf eine Marketingaktion, wie z.B. Fernsehwerbung, soll für den September 1995 bestimmt werden. Dazu muss der Prozentsatz der durch beworbene Produkte (Teilstring "PROMO" im Typ) erzielten Monatseinnahmen am Gesamtumsatz berechnet werden. Es werden nur tatsächlich verschickte Teile betrachtet.

Q15

Der beste Lieferant im ersten Quartal 1996 soll ermittelt werden. Das ist der Lieferant, der in diesem Quartal den größten Anteil zum Gesamtumsatz beigetragen hat. Nummer, Name, Adresse, Telefonnummer des Lieferanten sowie der Umsatz durch diesen Lieferanten sollen ausgegeben werden.

TPC-H/R-Anfragen

Q16

Man finde heraus, wie viele Lieferanten Teile in den Größen 49, 14, 23, 45, 19, 3, 36 oder 9 liefern können, die nicht von der Sorte 45 und nicht vom Typ "MEDIUM POLISHED" sind. Außerdem dürfen für diese Lieferanten keine Beschwerden vermerkt sein, was durch einen Kommentar ausgedrückt wird, der die Teilstrings "Better Business Bureau" und "Complaints" enthält. Man zähle die Lieferanten je Größe, Sorte und Typ und sortiere die Ausgabe nach aufsteigender Sorte und absteigendem Zähler (**count**).

Q17

Man berechne den durchschnittlichen jährlichen Einnahmenverlust, der sich ergeben würde, falls Aufträge mit kleineren Mengen (unter 20% der Durchschnittsmenge für dieses Teil) für die Sorte (brand) 23 im Container "LG BOX" nicht mehr angenommen würden.

Q18

Man ermittle die Auftraggeber (Kunden) der Top 100-Bestellungen (gemäß dem Gesamtpreis (totalprice)) unter denen, die eine Bestellposition im Umfang von mindestens 312 Einheiten enthalten.

TPC-H/R-Anfragen

Q19

Es soll der Gesamtumsatz (unter Berücksichtigung von Rabatten) ermittelt werden, der für drei bestimmte Sorten (brand) von Produkten erzielt wurde. Weiterhin sollen nur die Bestellpositionen berücksichtigt werden, die per Luftfracht (shipmode) und persönlich (shipinstruct) ausgeliefert wurden und eine bestimmte Anzahl (quantity) des Produkts umfasste.

Bei dieser Anfrage handelt es sich um eine typische Anfrage, wie sie von Data Mining-Systemen generiert werden.

Q20

In dieser Anfrage sollen Sonderangebots-Kandidaten ermittelt werden: Man finde die Hersteller, die von bestimmten Teilen (gekennzeichnet durch die Farbe "forest") mehr als die Hälfte des kanadischen Jahresabsatzes (von 1994) auf Lager haben.

TPC-H/R-Anfragen

Q21

In dieser Anfrage sollen säumige Lieferanten aus Saudi Arabien ermittelt werden. Es sollen die Lieferanten dieses Landes ermittelt werden, die in einer mehrere Lieferanten betreffenden, verspätet ausgelieferten Bestellung als einzige den Status (linestatus) 'f' haben.

Q22

Finde potentiell reaktivierbare Kunden. Für bestimmte Länderregionen (identifiziert durch die ersten beiden Ziffern der Kunden-Telefonnummer) finde die Anzahl der Kunden, die in den letzten 7 Jahren keine Bestellung mehr aufgegeben haben, aber dennoch ein überdurchschnittliches Bestellskonto (acctbal) aufweisen.

Der Umsatz durch einen *LINEITEM* berechnet sich aus $L_EXTENDEDPRICE * (1-L_DISCOUNT)$.

Der Gewinn durch einen *LINEITEM* berechnet sich aus Umsatz minus dem Einkaufspreis:

$L_EXTENDEDPRICE * (1-L_DISCOUNT) - L_QUANTITYPS_SUPPLYCOST$.

Update-Operationen

- UF1
- Mit Hilfe dieser Updatefunktion werden neue Verkaufsinformationen in die Datenbank eingefügt. Dazu lädt sie zusätzliche Datensätze in die Tabellen *ORDER* und *LINEITEM*, welche zuvor mit dem Programm DBGEN erzeugt wurden. Insgesamt müssen $SF * 1500$ neue Tupel in die Relation *ORDER* und pro neuer Bestellung eine zufällig im Bereich 1 bis 7 gewählte Anzahl von zugeordneten *LINEITEM*-Tupeln eingefügt werden.
- UF2
- Diese Funktion entfernt überholte bzw. überflüssige Informationen aus der Datenbank, indem sie die entsprechenden Datensätze in den Tabellen *ORDER* und *LINEITEM* löscht. Insgesamt werden $SF * 1500$ Tupel aus *ORDER* gelöscht und alle zu diesen gelöschten Bestellungen gehörenden Einträge aus *LINEITEM*.

Leistungsgrößen

- Der Systempreis
 - wiederum bestehend aus Hardware, Software und Softwarewartung für fünf Jahre.
- Die TPC-H/R Powermetrik $Q_{ppH/R@Size}$, die in Anzahl von sequentiell ausgeführten Anfragen und Änderungen pro Stunde angegeben wird.
 - Der Parameter `Size` gibt die Datenbankgröße an.
- Der Powerwert wird auf der Basis des inversen geometrischen Mittels der Anfrage- und Update-Laufzeiten ermittelt, damit die weniger aufwendigen Anfragen nicht von den sehr komplexen Anfragen des Benchmarks dominiert werden. Dieser Durchschnittswert wird mit dem Skalierungsfaktor der Datenbasis gewichtet (d.h. multipliziert), um den unterschiedlichen Datenbankgrößen Rechnung zu tragen.

Leistungsgrößen

- Der Durchsatz $Q_{thH/R@Size}$, der sich aus der Anzahl bearbeiteter Anfragen pro Stunde ergibt, wenn die Anfragen in parallelen Strömen bearbeitet werden. Wiederum ist dieser Wert gewichtet mit dem Skalierungsfaktor.
- Die kombinierte Leistungsmetrik $Q_{phH/R@Size}$ wird berechnet aus dem geometrischen Mittel aus $Q_{ppH/R@Size}$ und $Q_{thD@Size}$.
- Das Preis/Leistungsverhältnis, das in Dollar pro Anfrage pro Stunde angegeben wird.

10,000 GB Results

| Rank | Company | System | QphH | Price/QphH | System Availability | Database |
|------|---|--|---------|------------|---------------------|--|
| 1 |  | IBM System p 570 | 343,551 | 32.89 USD | 04/15/08 | IBM DB2 Warehouse 9.5 |
| 2 |  | HP Integrity Superdome/Dual-Core Itanium/1.6 GHz | 208,457 | 27.97 USD | 09/10/08 | Oracle Database 11g Enterprise Edition |
| 3 |  | IBM System p5 575 with DB2 UDB 8.2 | 180,108 | 47.00 USD | 08/30/06 | IBM DB2 UDB 8.2 |
| 4 |  | HP Integrity Superdome-DC Itanium2/1.6GHz/64p/128c | 171,380 | 32.91 USD | 04/01/07 | Oracle Database 10g R2 Enterprise Edt w/Partitioning |
| 5 |  | Sun Fire[™] E25K server | 108,099 | 53.80 USD | 01/23/06 | Oracle 10g Enterprise Ed R2 w/ Partitioni |
| 6 |  | HP Integrity Superdome □ Itanium2/1.5 GHz-128p/128 | 86,282 | 161.24 USD | 04/06/05 | Oracle Database 10g Enterprise Edition |
| 7 |  | Unisys ES7000 Model 7600R Enterprise Server(16s) | 80,172 | 18.95 USD | 02/17/09 | Microsoft SQL Server 2008 Enterprise x6- |
| 8 |  | HP Integrity Superdome | 63,650 | 38.54 USD | 08/30/08 | Microsoft SQL Server 2008 Enterprise Edi |
| 9 |  | HP Integrity Superdome □ Itanium2/1.5 GHz-64p/64c | 49,104 | 118.13 USD | 03/25/04 | Oracle Database 10g Enterprise Edition |

30,000 GB Results

| Rank | Company | System | QphH | Price/QphH | System Availability | Database |
|------|---|--|---------|------------|---------------------|---------------------|
| 1 |  | HP Integrity Superdome - Itanium2/1.6 GHz/18MB iL3 | 150,960 | 46.69 USD | 06/18/07 | Oracle Database 10g |

Beispiel-Ergebnis

- HP / Oracle
 - 30 TB Datenbankgröße
 - 30 * 14 TB Speichervolumen (mit Indexes, materialisierten Sichten, etc)
 - Kosten: ca. 10 Mio. \$
 - [hp_tpch_sd_30TB_es.pdf](#)
 - http://www.tpc.org/results/individual_results/HP/hp_tpch_sd_30TB_es.pdf

Query Variations

Formal Query Definitions are ISO-92 SQL

- EQT must match *except* for Minor Query Modification

Date/Time Syntax

AS clauses

Table Naming Conventions
By/Order By

Ordinal Group

Statement Terminators
(space)

Coding Style (I.e., white

Any other phrasing must be a Pre-Approved Query Variant

- Variants must be justifiable base on a criteria similar to 0.2
- Approved variants are include in the specification

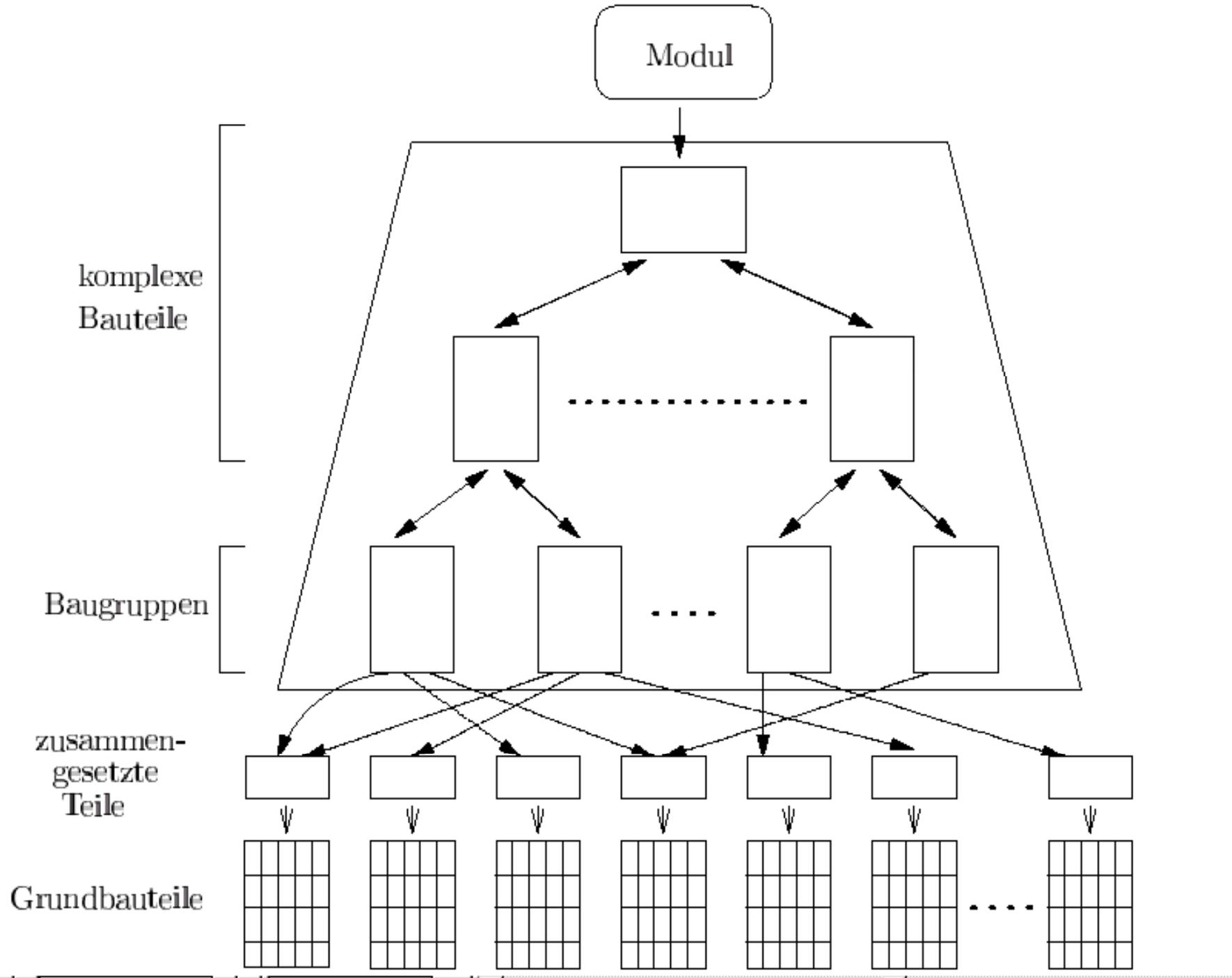
An implementation may use any combinations of Pre-Approved Variants, Formal Query Definitions and Minor Query Modifications.

TPC-D Update Functions

- Update 0.1% of data per query stream
 - About as long as a medium sized TPC-D query
- Implementation of updates is left to sponsor, *except*:
 - ACID properties must be maintained
 - The update functions must be a set of logically consistent transactions
- New Sales Update Function (UF1)
 - Insert new rows into ORDER and LINEITEM tables equal to 0.1% of table size
- Old Sales Update Function (UF2)
 - Delete rows from ORDER and LINEITEM tables equal to 0.1% of table size

Der OO7-Benchmark

- Zur Bewertung objektorientierter Datenbanken



Erläuterungen zum OO7-Schema

Der OO7-Benchmark modelliert Objekthierarchien, wie sie in ingenieurwissenschaftlichen Anwendungen (z.B. CAD, CAM) vorkommen.

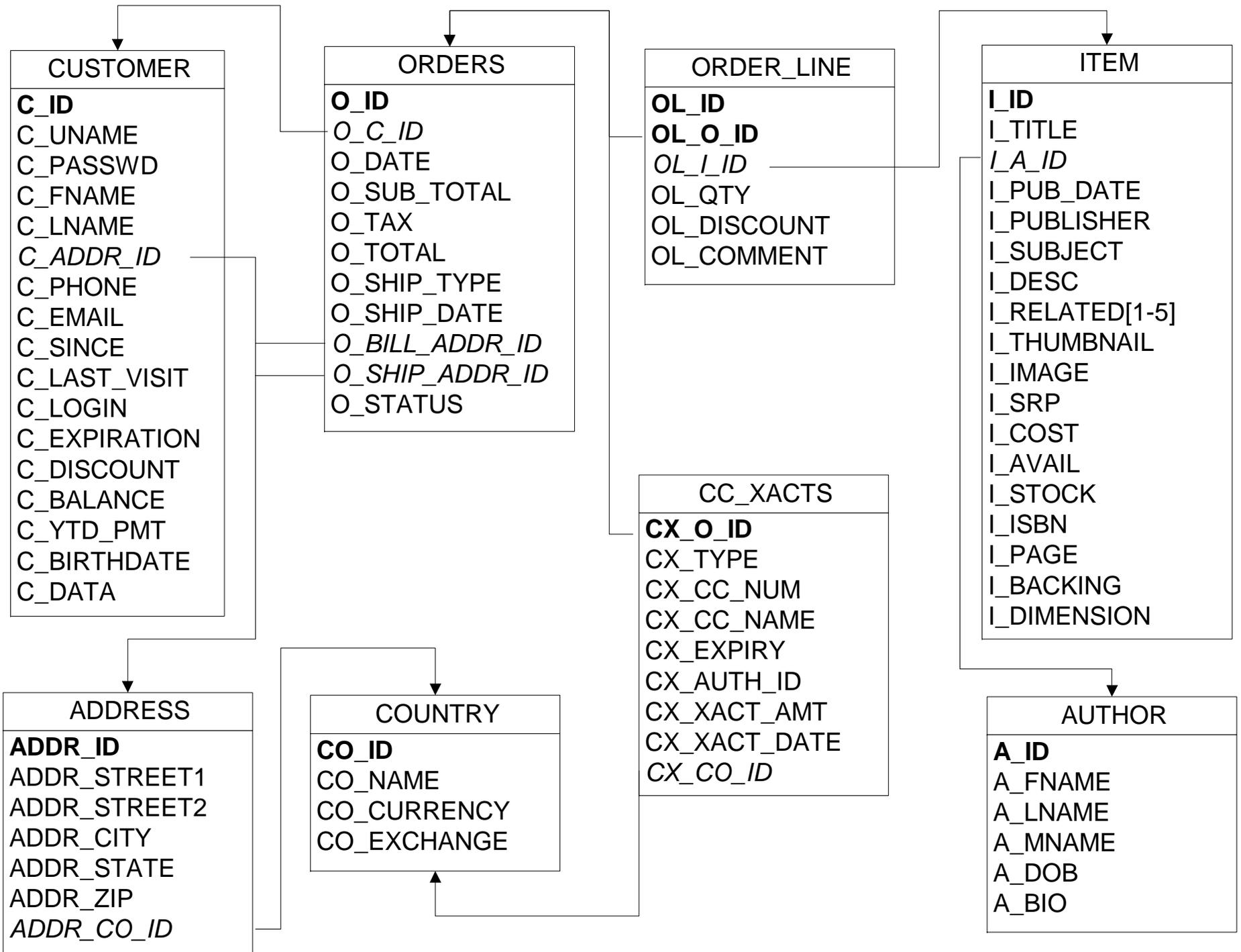
Ein zusammengesetztes Objekt (composite part) besteht aus einem Textdokument und einem Objektnetz von 20 (in der größeren Datenbankkonfiguration 200) Grundbauteilen (atomic parts).

Jede Baugruppe hat eine bidirektionale Beziehung zu drei zusammengesetzten Teilen, die ihrerseits aber Bestandteil mehrerer Baugruppen sein können (shared subobjects).

Baugruppen sind nochmals in einer Hierarchie zu komplexen Bauteilen verknüpft. Den Einstiegspunkt zu dieser Hierarchie bildet ein Modul.

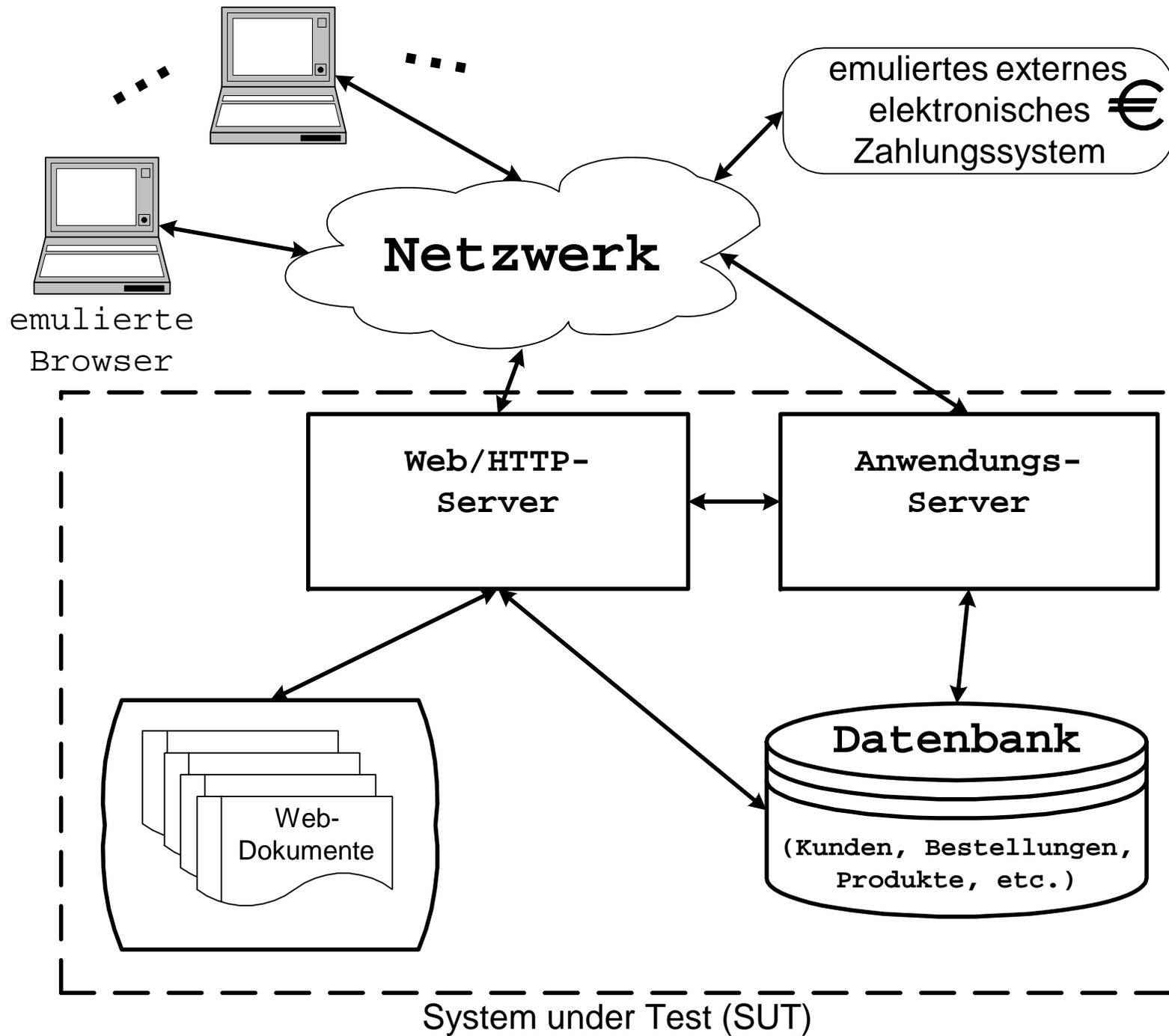
Der TPC-W-Benchmark

- Modelliert ein E-Commerce-Unternehmen
- Konkret wird ein Buchhändler (Amazon?) simuliert
- Die Datenbank wird über Web-Schnittstellen angesprochen
- Es wird nicht nur die Datenbank analysiert, sondern das gesamte System
 - DBMS
 - Applicationserver
 - Webserver



Datenbank des TPC-W Benchmarks

- Die Datenbank kann auf unterschiedliche Größen skaliert werden, wobei statisch die Anzahl der Produkte variiert wird.
- Gültige Größen sind Datenbanken mit 1000, 10000, 100000, und 1000000 Produkten.
- Die Anzahl der Kunden, Bestellungen und Transaktionen ergibt sich dynamisch, je nach der Anzahl der abgearbeiteten Aufträge, die von simulierten Browsern initiiert werden.



Anwendungsbeschreibung des TPC-W

Die Last wird dabei durch sogenannte emulierte Web-Browser generiert.

Dabei wird das menschliche Verhalten beim virtuellen Besuch eines online-Buchladens simuliert.

Interaktionen werden zufällig aus einer Menge von insgesamt 14 sogenannte Web-Interaktionstypen generiert.

Jedem Benutzer wird zunächst dieselbe statische Einstiegsseite mit HyperLinks auf Seiten mit Neuerscheinungen und Bestseller für die unterschiedlichen Produktkategorien (Attribut *I_SUBJECT*) präsentiert.

Die Anforderung dieser Einstiegsseite wird als Home-Web-Interaktion bezeichnet.

Beim Zugriff auf die Bestseller- bzw. die Neuerscheinungen-Seiten muss die Information natürlich dynamisch aus der Datenbank generiert werden. Dabei könnte beispielsweise eine Datenbankanbindung via Servlets oder Java Server Pages, wie wir sie in Kapitel 16 vorgestellt haben, zum Einsatz kommen.

Anwendungsspezifikation -- Fortsetzung

Benutzer können gezielt nach Büchern suchen, wobei zwischen der Spezifikation (Search Request Web Interaktion) und dem Präsentieren der Ergebnisse (Search Result Web Interaktion) unterschieden wird.

Bei der ersten Produktauswahl wird den Benutzern ein virtueller Einkaufswagen (shopping cart) zugeordnet, in dem die ausgewählten Artikel gespeichert werden.

Der eigentliche Einkauf besteht auch aus zwei Web-Interaktionen: Einmal der Anforderung (Request) und dann der Zustimmung (Confirm).

Die Abwicklung einer Bestellung verlangt zusätzlich die Kommunikation des E-Commerce-Systems mit einem externen, simulierten elektronischen Zahlungssystem (Kreditkarten-Firma).

Bei diesem externen Zahlungssystem muss die Autorisierung für die betreffende Transaktion eingeholt werden.

Die übermittelte Autorisierungsinformation wird in der Transaktionen-Relation (CC_XACTS) gespeichert.

Zusätzlich zu den simulierten Benutzern gibt es parallel auszuführende Administrationsaufgaben, die darin bestehen, Produktinformationen zu ändern.

Leistungsgrößen

- Aus Benutzersicht ist sicherlich die Antwortzeit eine entscheidende Größe -- wie viele Leser aus eigener leidvoller Erfahrung im "World Wide Wait" einschätzen können.
- Die Antwortzeit für Benutzeraufträge errechnet sich als die Zeitspanne zwischen der Übertragung der Anforderung und dem Erhalt der Bestätigung.
- Aus System-Administratorsicht ist der Durchsatz, also die Anzahl der Web-Interaktionen pro Zeiteinheit ein wichtiger Leistungsparameter.
- Wie bei den anderen TPC-Benchmarks wird auch der Systempreisermittelt.
- Daraus abgeleitet ergibt sich das Kosten/Leistungs-Verhältnis.

Erzielbare Werte

- TPC-W gilt als obsolet
 - Deshalb sind die letzte Benchmarkergebnisse aus dem Jahre 2002
 - Nicht mehr aktuell für derzeitige Hardware/Software-Infrastrukturen
- Bei einer Datenbankskalierung auf 10.000 Produkte sind heute Leistungszahlen von ca. 21000 Web-Interaktionen pro Sekunde bei einem Systempreis von ca. 700.000 US-\$ möglich.
- Bei größer skalierten Datenbanken auf 100.000 Produkte benötigt man schon ein System im Wert von ca. 1,2 Millionen US-\$, um die gleiche Anzahl von 10.000 Web-Interaktionen pro Sekunde zu erzielen.

Weitere/neuere Benchmarks

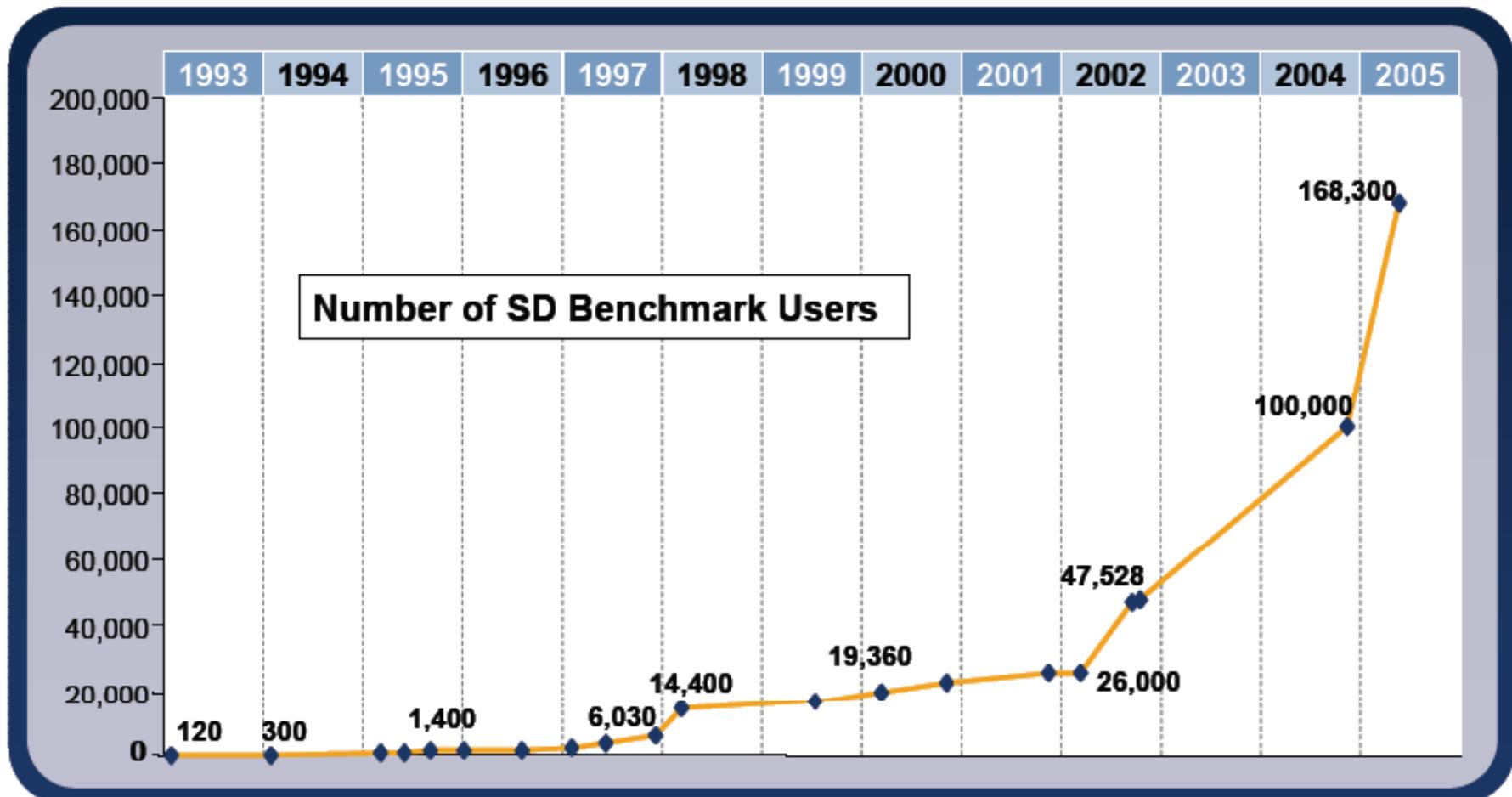
- SD Benchmark für SAP R/3
 - Sales and distribution
- XMark-Benchmark für XML-Datenbanken
 - Orientiert sich an den Daten eines Auktionshauses (ebay)
- TPC E : OLTP
 - Löst TPC C bald ab
- TPC Application
 - Löst TPC W ab
 - B2B mit Web Services
- TPC DS
 - Decision support
 - Mehrere Faktentabellen (Verkäufe über verschiedene Kanäle: Geschäft, Web, Telefon, etc)
 - 99 Queries

SAP-Charakteristika: OLTP

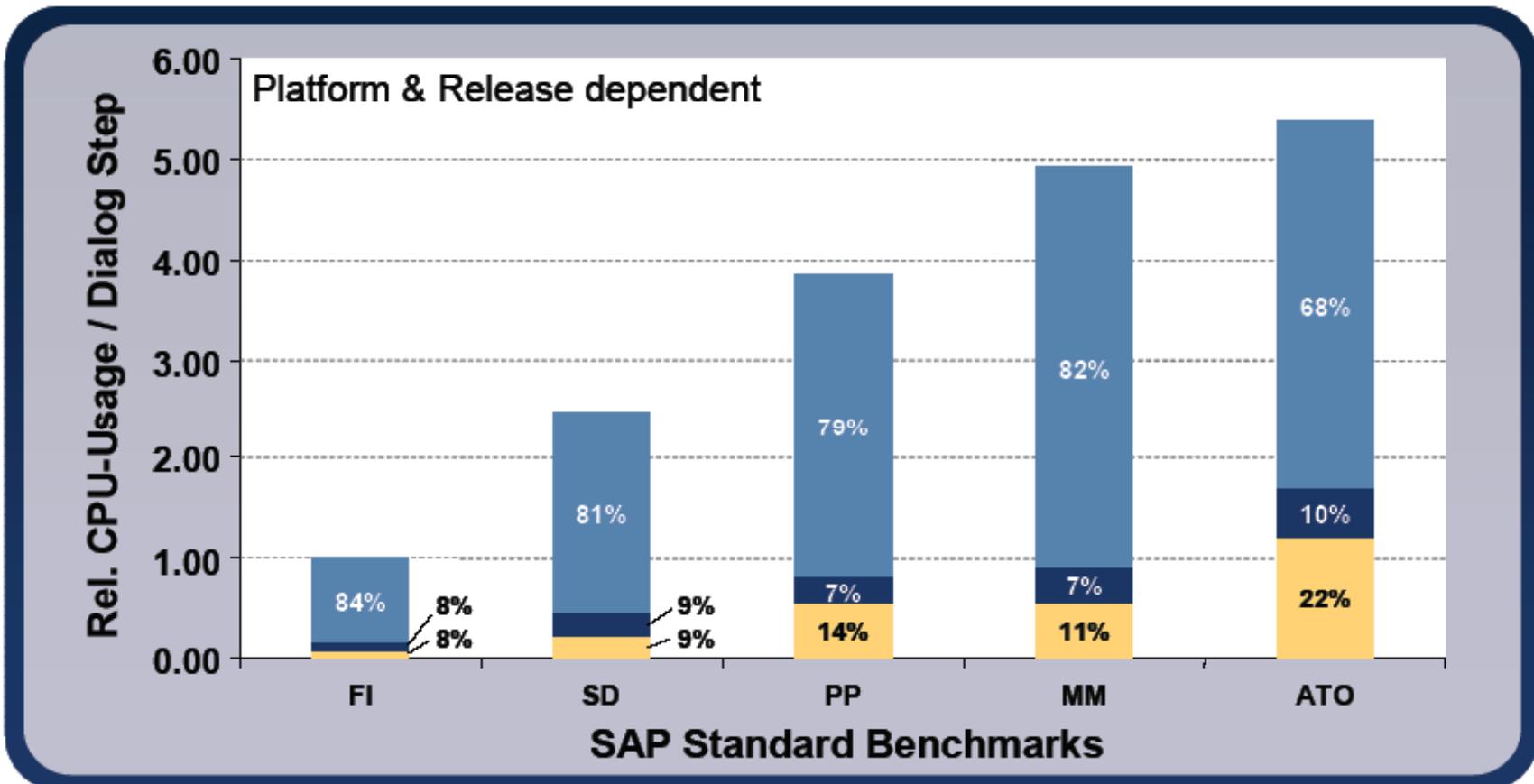
- 67.000 tables
- 700.000 columns
- 10.000 views
- 13.000 indexes
- 100.000.000 rows (initial size)
- 57 GB disk footprint (initial size)
- 270 millions lines of code

SD-Leistungskennzahlen ... über die Jahre (number of users)

SD Benchmark (three-tier): Highest number of users

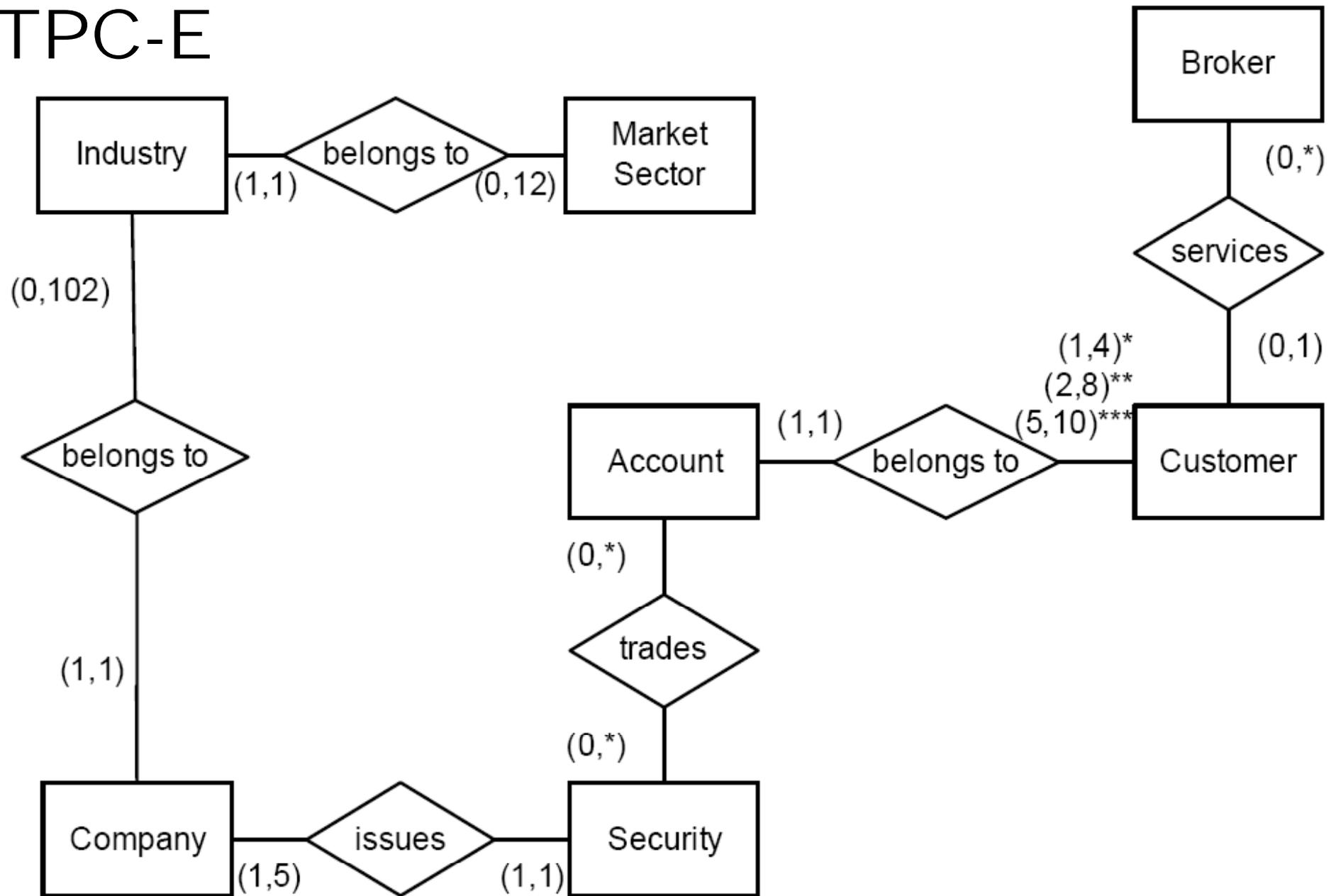


Typical OLTP CPU Load Distribution



■ + ■ = Application Server ■ = Database Server

TPC-E



20.6: Das Schema des TPC-E-Benchmarks

TPC-E: Testumgebung

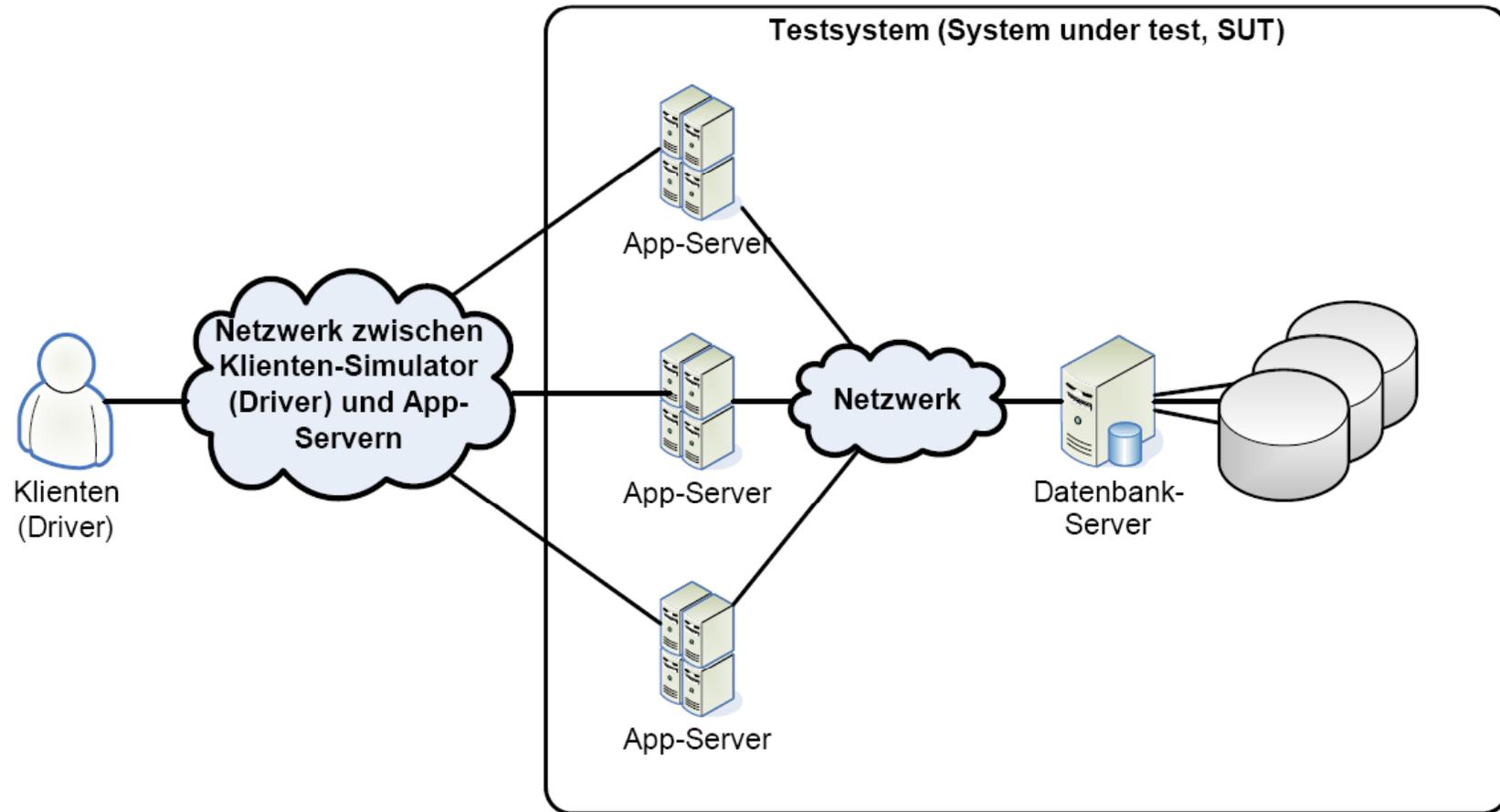
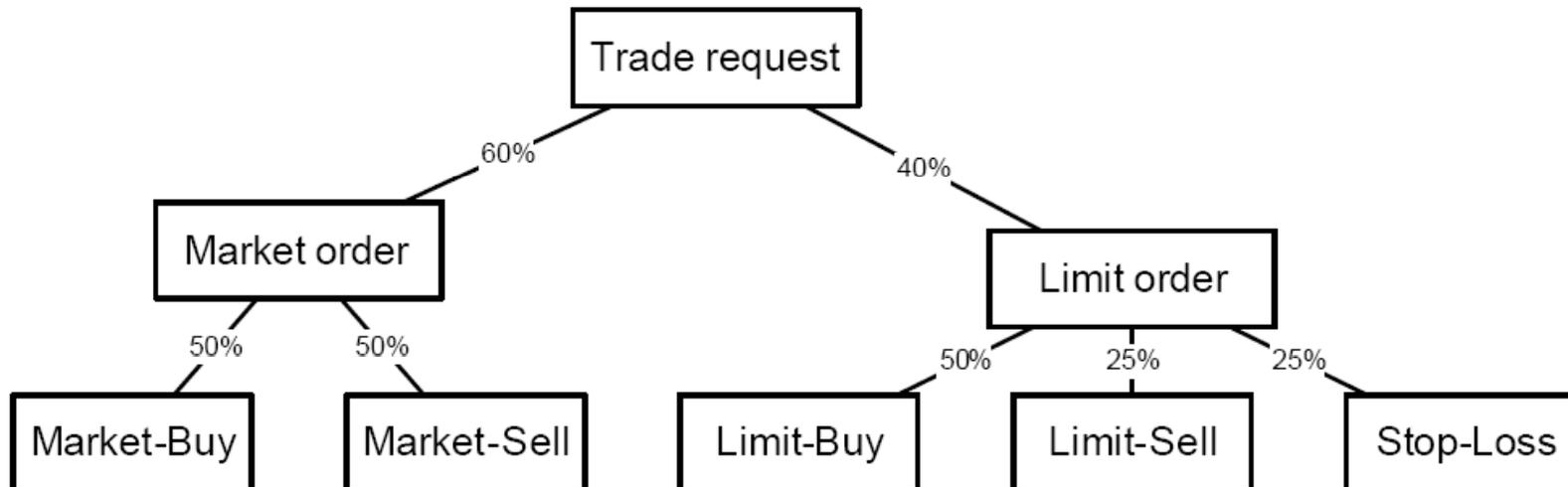


Abb. 20.7: Die Testumgebung des TPC-E-Benchmarks

Auszuführender Transaktionsmix

Der vom System auszuführende „Mix“ von Handels-Transaktionen (trade requests) ist nachfolgend illustriert:

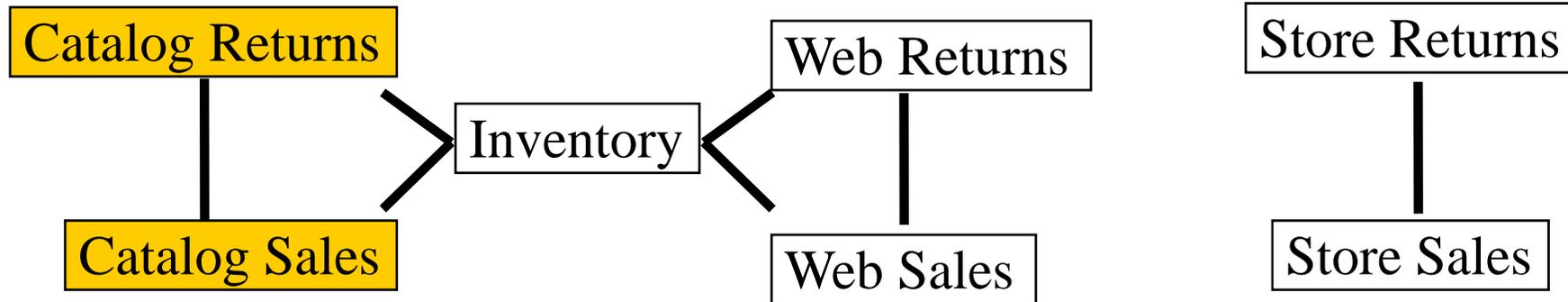


TPC-H/DS



TPC-DS Subcommittee Reports
Meikel Poess

Schema: TPC-DS - Fact Tables



- 3 sales channels: Catalog - Web - Store
- 7 fact tables
- 2 fact tables for each sales channel
- 24 tables total
- Basic (TPC-H like) auxiliary data structure are allowed on all tables
- Aggregation and “fancy” auxiliary data structures are only allowed on **Catalog Sales and Catalog Returns**

Schema: Store Channel w/ Dimensions

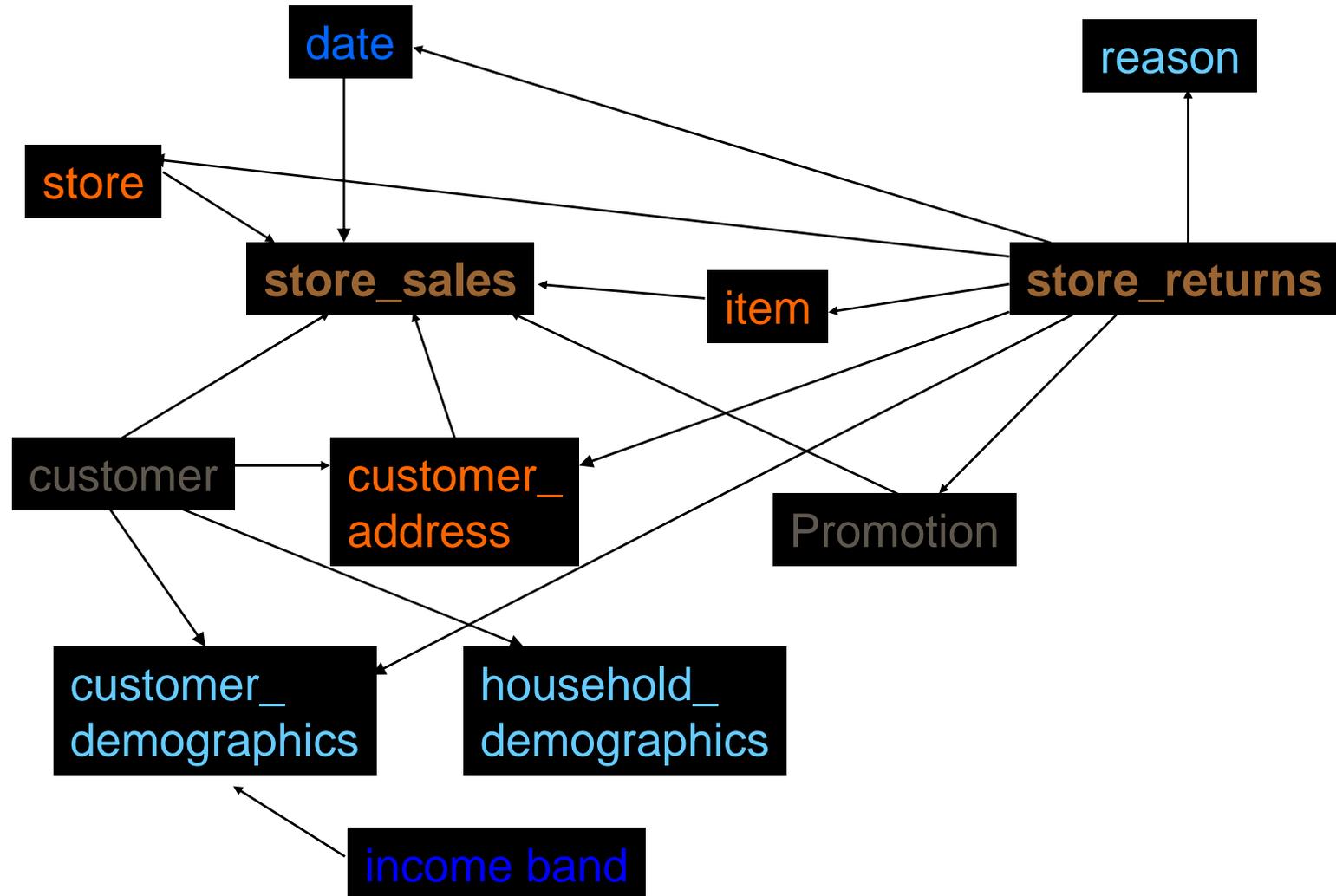
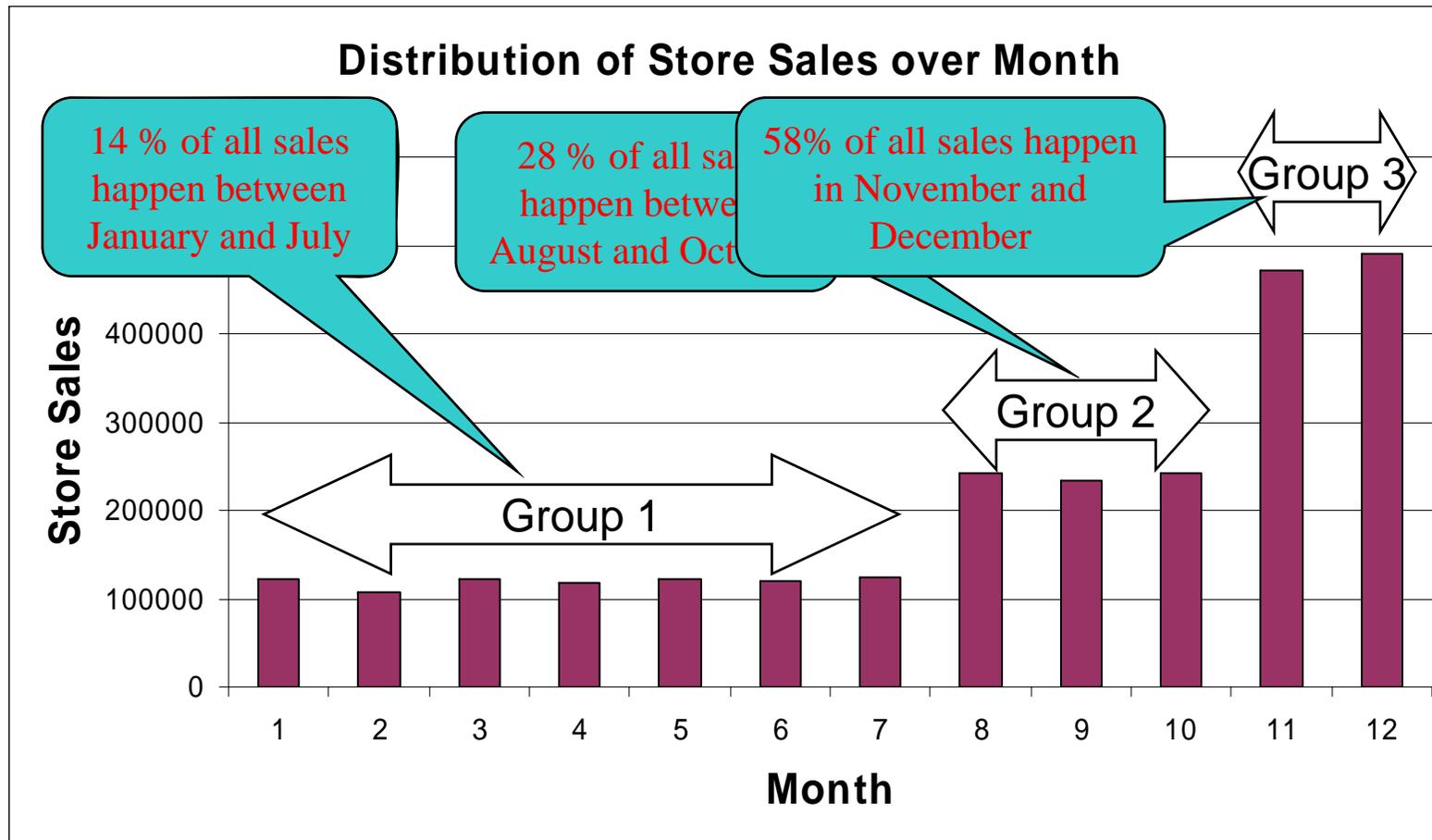


Table Sizes at 100GB

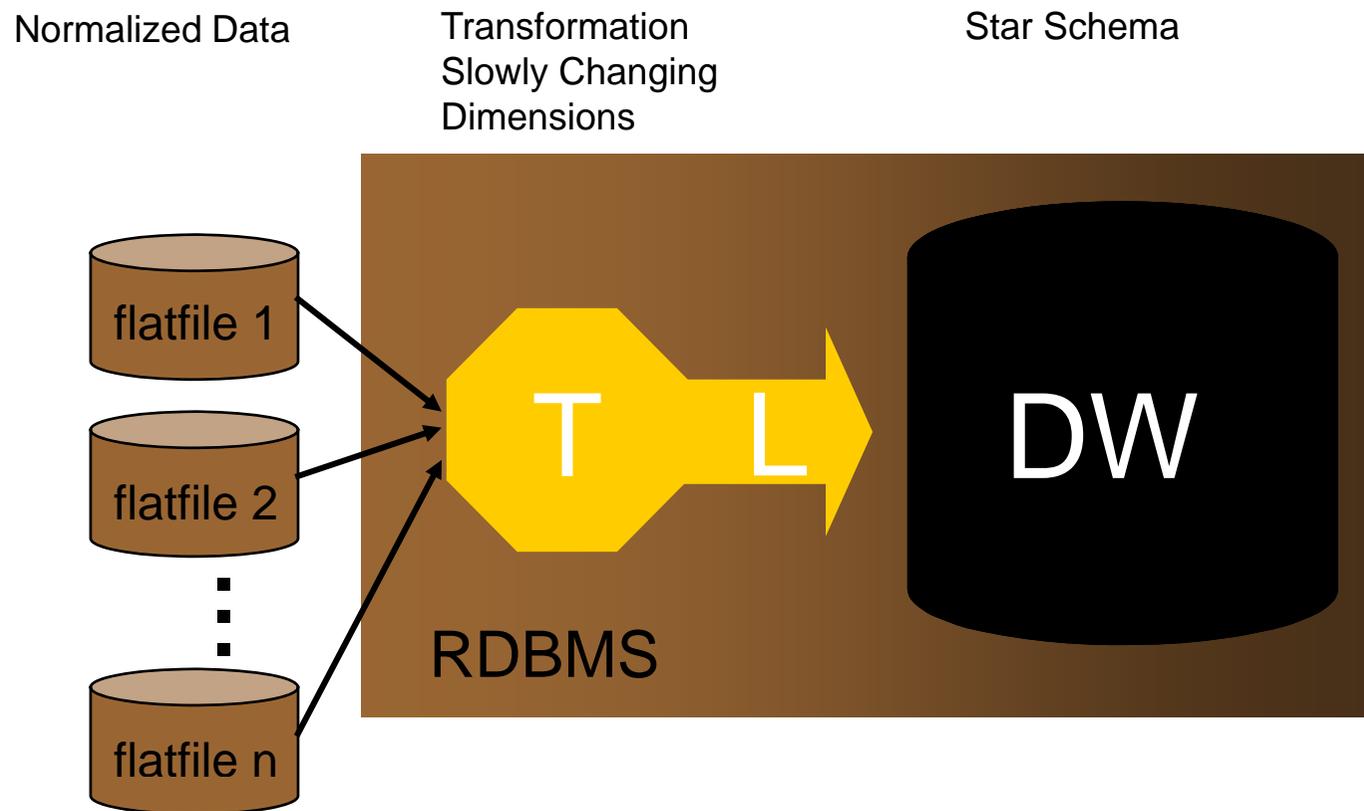
| Table | #Rows | Raw Size (GB) |
|-----------------|--------------|----------------------|
| Store Sales | 288 Million | 47.8 |
| Store Returns | 14.4 Million | 2 |
| Catalog Sales | 144 Million | 33.3 |
| Catalog Returns | 14.4 Million | 2.6 |
| Web Sales | 72 Million | 16.1 |
| Web Returns | 7.2 Million | 2.6 |
| Inventory | 390 Million | 6 |
| Customer | 2 Million | 0.438 |
| Item | 100,000 | 0.047 |
| Catalog Page | 24,000 | 0.006 |

SS_SOLD_DATE Distribution



Data Maintenance Model

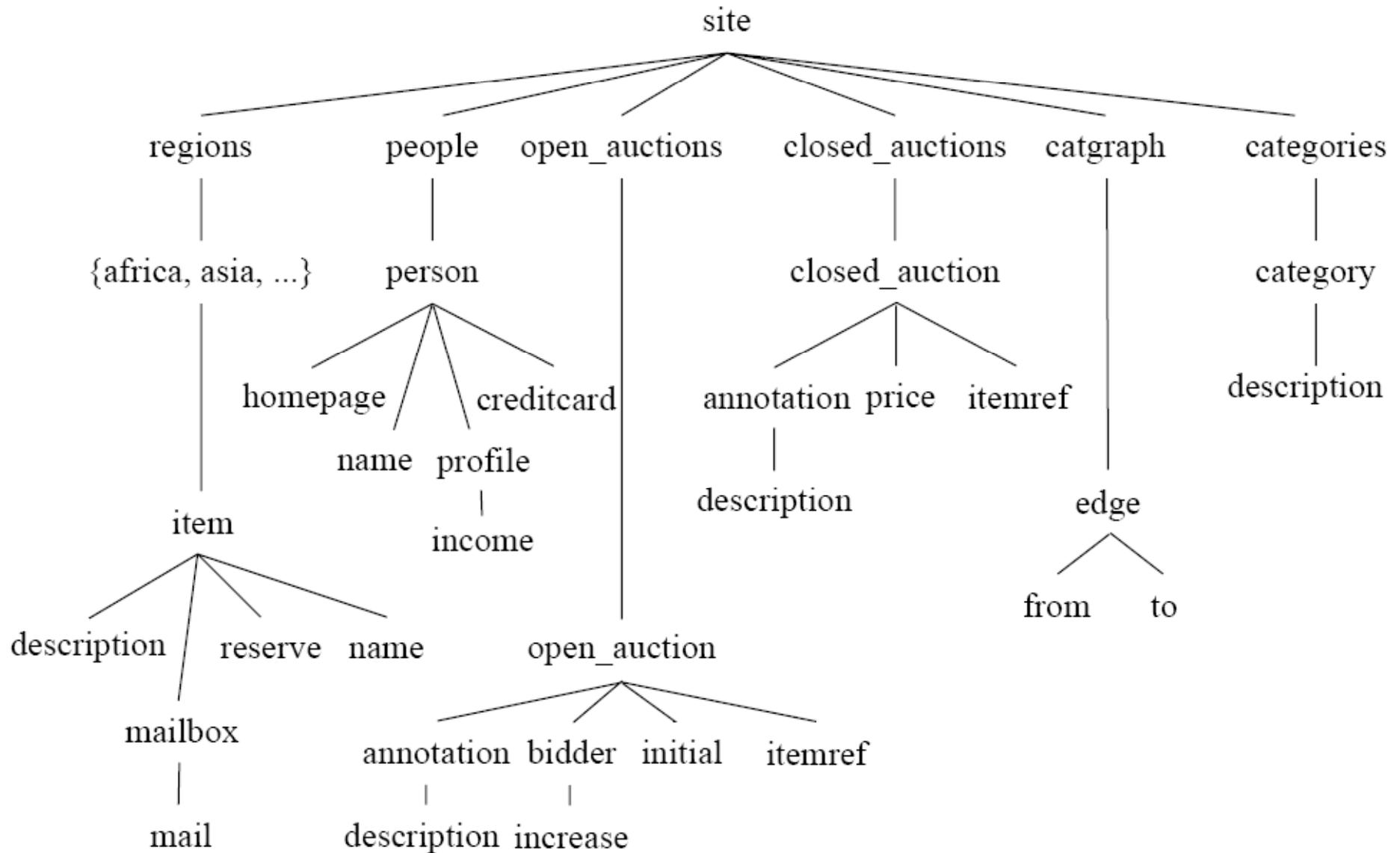
(Extraction) - Transformation - Load



TPC-App

- Mehrere parallele Anwendungs-Sessions, die jeweils mehrere transaktionale Services für einen Klienten bündeln
- Kommerziell verfügbare Anwendungsplattform (also „off the shelf“ und keine Eigenentwicklung)
- Nutzung von XML-Dokumenten und SOAP für den Nachrichtenaustausch
- Nutzung eines relationalen Datenbanksystems als Backend (also trotz SOAP-Kommunikation kein XML-Datenbanksystem)
- Business-to-Business Anwendungslogik
- Transaktionale Integrität gemäß ACID-Paradigma
- Zuverlässige und dauerhafte Nachrichtenverwaltung (messaging)
- Dynamische Webservice-Interaktionen, bei denen die Antworten (responses) dynamisch generiert werden
- Gleichzeitige Ausführung vieler Transaktionstypen, die die gesamte Bandbreite eines B2B-ecommerce abdecken (Browsing im Katalog, Bestellungen aufgeben, Statusabfragen, etc.)
- Datenbankschema mit mehreren Relationen unterschiedlicher Größe und Anzahl von Attributen

XMARK



Q6. How many items are listed on all continents?

```
FOR $b IN document("auction.xml")/site/regions
RETURN COUNT ($b//item)
```

A good evaluation engine should realize that there is no need to traverse the complete document tree to evaluate such expressions.

Q7. How many pieces of prose are in our database?

```
FOR $p IN document("auction.xml")/site
LET $c1 := count($p//description),
    $c2 := count($p//mail),
    $c3 := count($p//email),
    $sum := $c1 + $c2 + $c3
RETURN $sum;
```

Also note that the COUNT aggregation does not require a complete traversal of the tree. Just the cardinality of the respective relation is queried. Note that the tag <email> does not exist in the database document.

5.5 Chasing References

References are an integral part of XML as they allow richer relationships than just hierarchical element structures. These queries define horizontal traversals with increasing complexity. A good query optimizer should take advantage of the cardinalities of the sets to be joined.

Q8. List the names of persons and the number of items they bought. (joins person, closed_auction)

```
FOR $p IN document("auction.xml")/site/people/person
LET $a := FOR $t IN document("auction.xml")/site/closed_auctions/closed_auction
        WHERE $t/buyer/@person = $p/@id
        RETURN $t
RETURN <item person=$p/name/text()> COUNT ($a) </item>
```

| Query | RP | Monet XML |
|----------|------|-----------|
| Query 1 | 1000 | 1217 ms |
| Query 2 | 1 | 2945 ms |
| Query 3 | 1 | 3891 ms |
| Query 4 | 1 | 153 ms |
| Query 5 | 1 | 161 ms |
| Query 6 | 1000 | 762 ms |
| Query 7 | 1000 | 2167 ms |
| Query 8 | 1 | 469 ms |
| Query 9 | 1 | 977 ms |
| Query 10 | 1 | 22292 ms |
| Query 11 | 1 | 8672 ms |
| Query 12 | 1 | 7464 ms |
| Query 13 | 1 | 25 min |
| Query 14 | 1 | 9223 ms |
| Query 15 | 100 | 762 ms |
| Query 16 | 100 | 2018 ms |
| Query 17 | 1000 | 250 ms |
| Query 18 | 100 | 7799 ms |
| Query 19 | 1 | 493 ms |
| Query 20 | 10 | 346 ms |

Figure 5: Performance Impression: Monet XML at scaling factor 1.0