



Normalization in a Mixed OLTP and OLAP Workload Scenario

Anja Bog¹, Kai Sachs², Alexander Zeier¹, and Hasso Plattner¹

¹Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany

²SAP AG, 69190 Walldorf, Germany

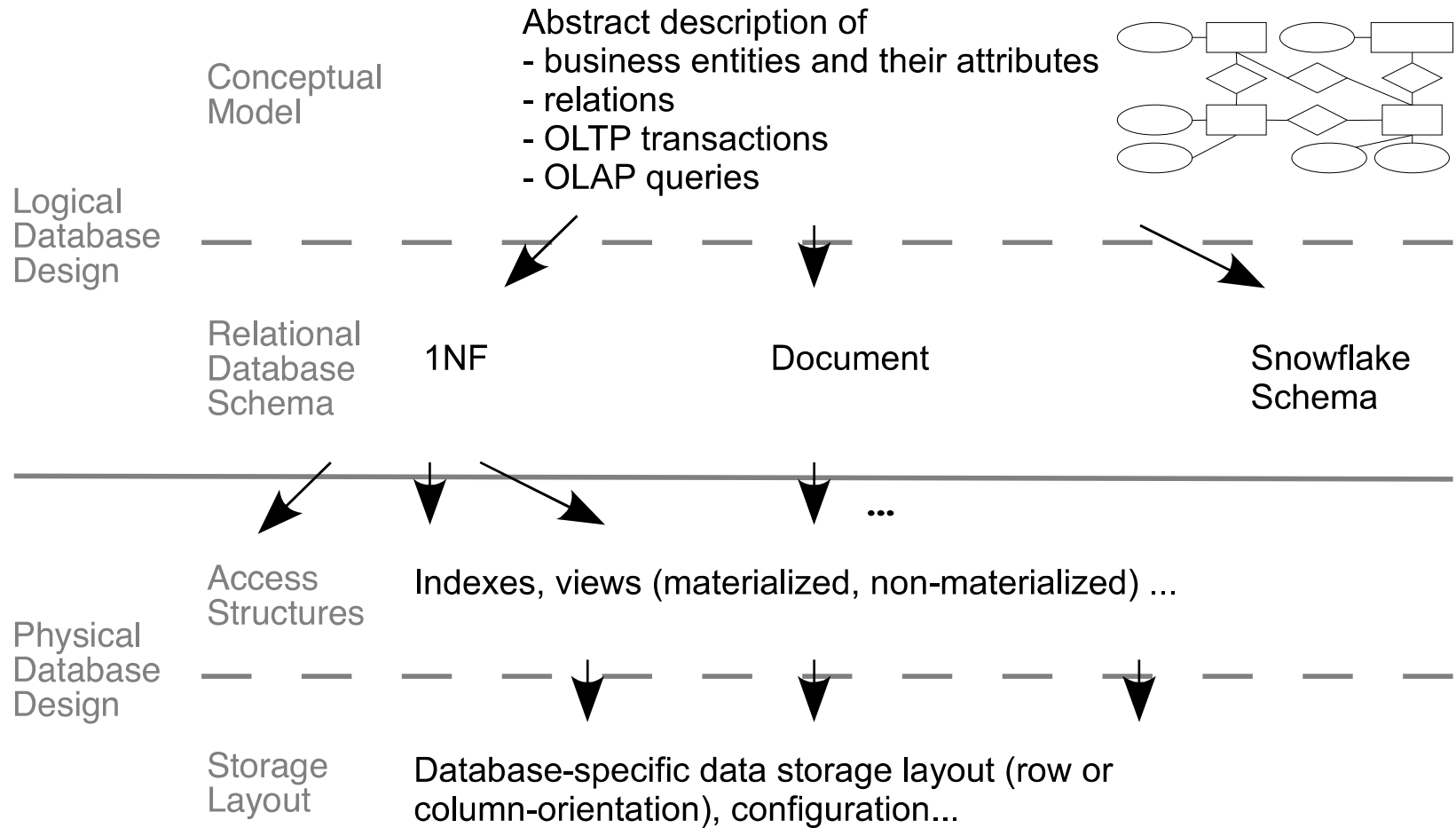
Motivation

2

- Analytics on up-to-date data vs. separated analytical and transactional systems
 - How? Increase update frequency, use synced operational data stores
- Flexible/ad hoc analyses vs. preparation of a data subset for analytics
 - How? Use transactional data directly to answer analytical queries
- But analytics on transactional data immensely intensifies the load on the transactional system?
 - Evaluate impact of adding analytics to the transactional workload
 - Study variants of database design for mixed workloads

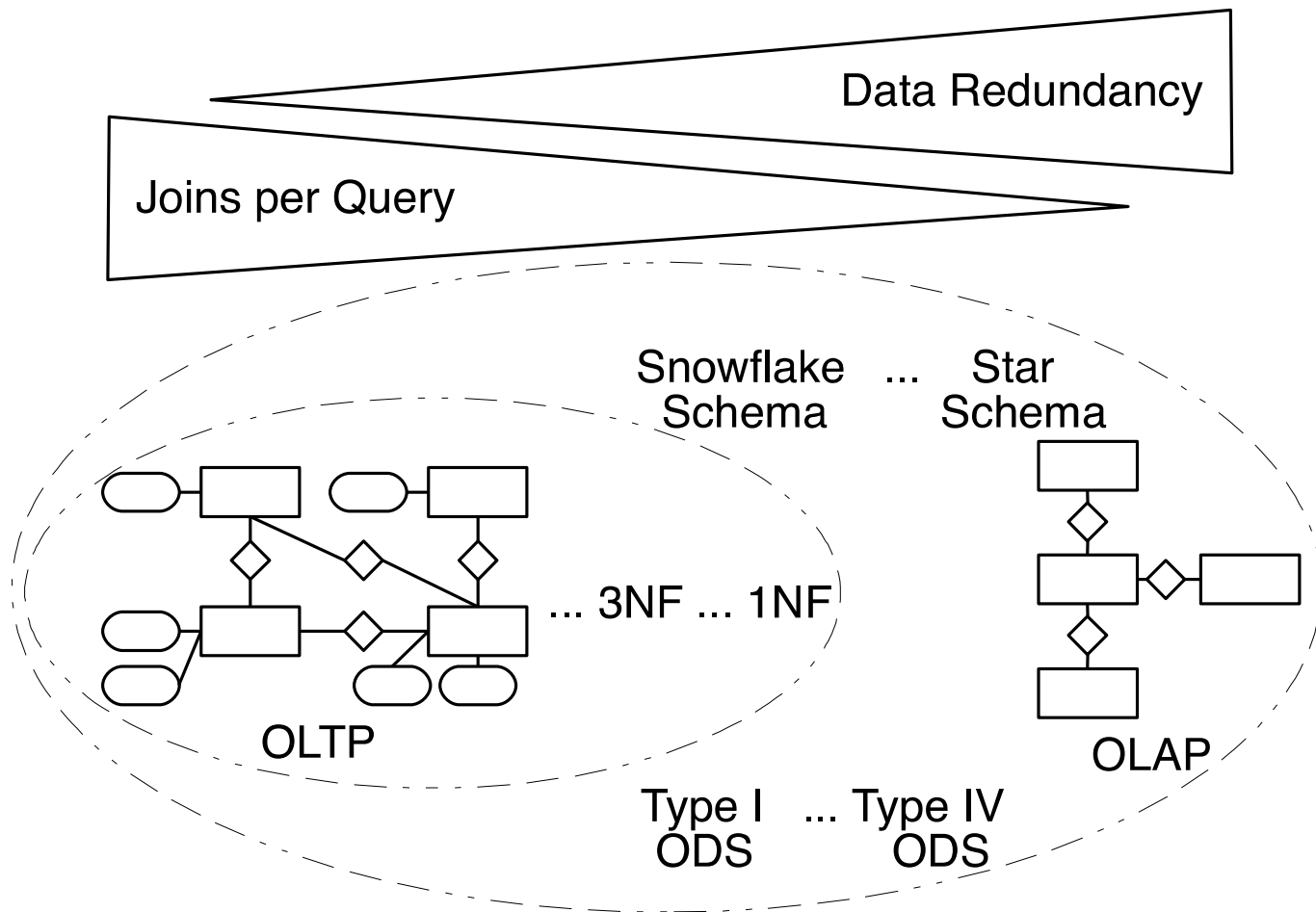
Database Design Steps

3



Database "Schemas" in Analytical and Transaction Processing

4



Case Study

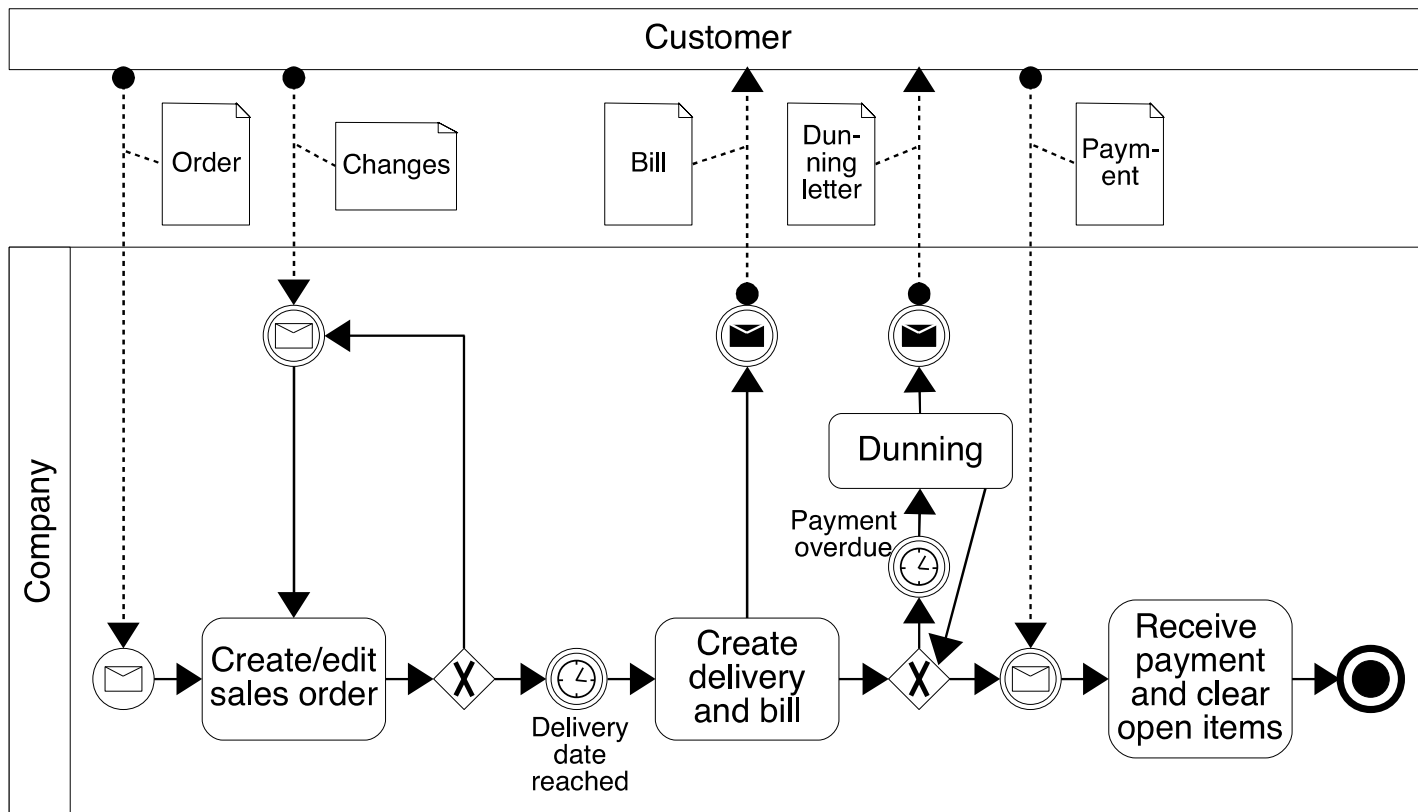
5

- Normalization as a key element that differs in OLTP and OLAP database design
- Analyze the interplay of workload mix and differently normalized database schemas wrt. query performance
- Dynamic database design in varying mixed workload scenarios?

Workload Scenario for the Case Study

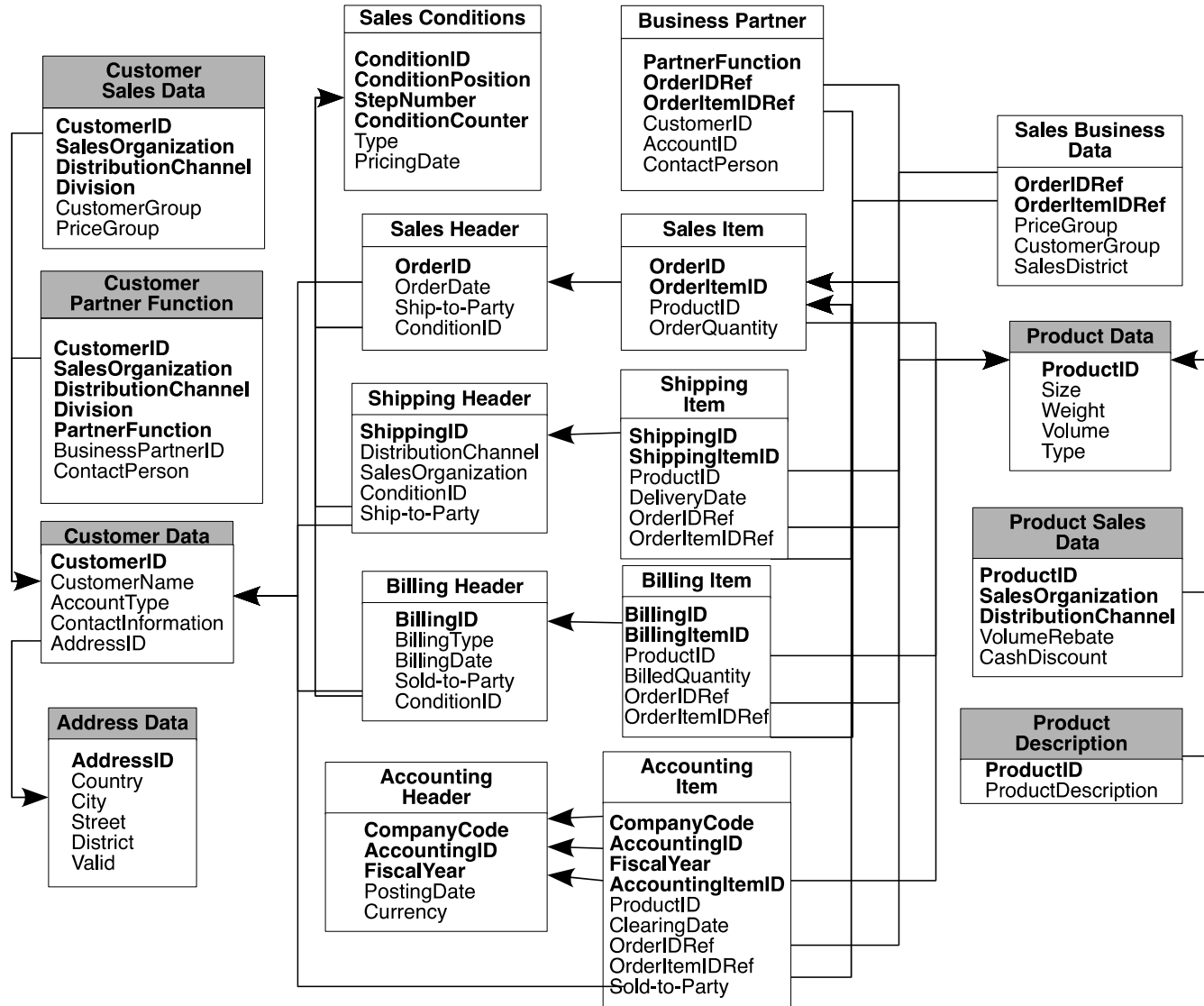
6

- Simulation on the basis of real enterprise data



Sales DB Schema – 1NF

7



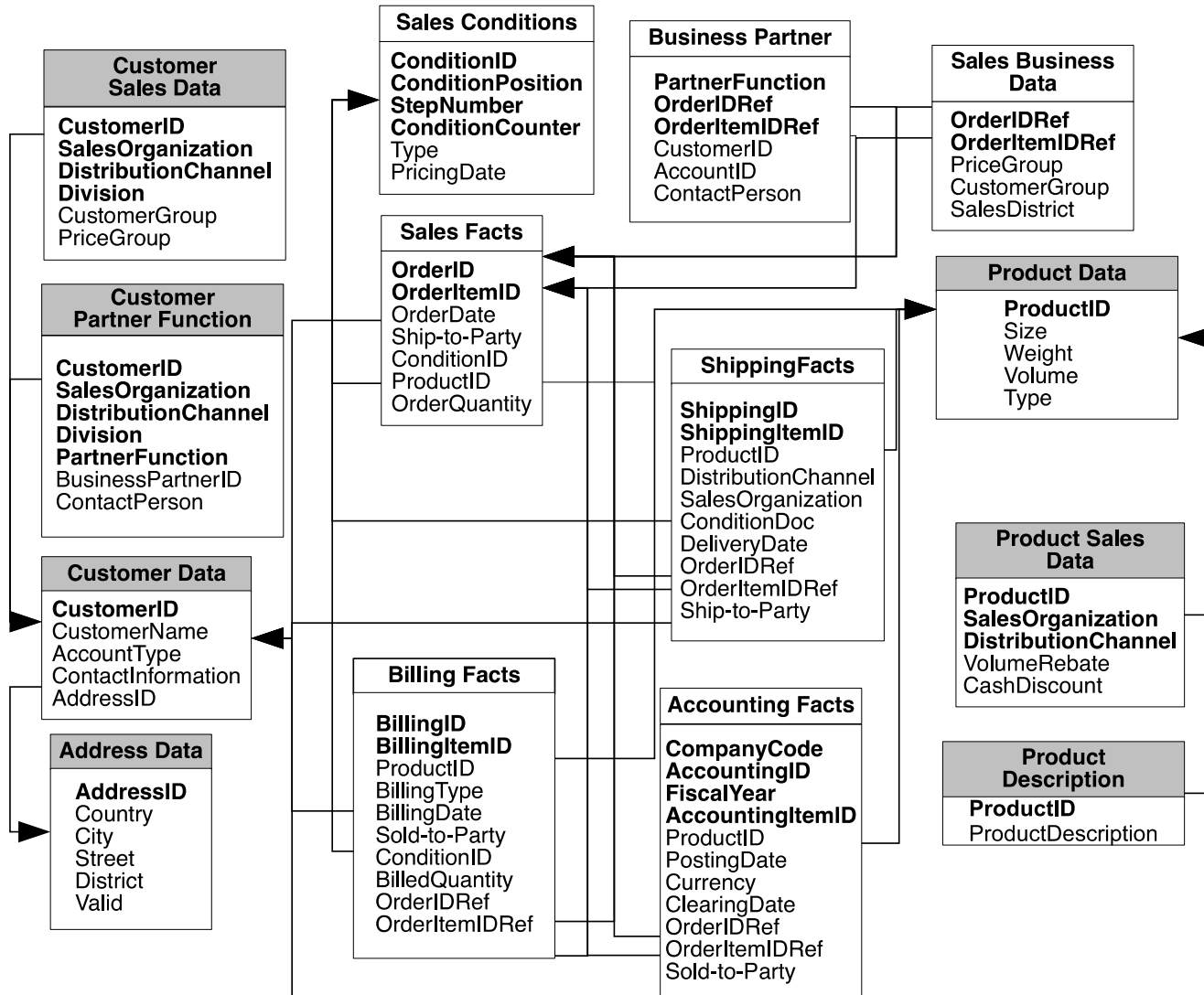
Workload Components

8

Type	Action	Share /Type	Profile
wOLTP	Sales Order	30%	7+ Master Data (MD) selections, 5+ Transaction Data (TD) inserts
	Shipping	27%	2+ TD selections, 2+ TD inserts
	Billing	25%	4+ TD selections, 5+ TD inserts
	Payment	18%	1 TD selection, 1+ TD update
rOTLP	Sales Order by Key	Random/Con-figurable	TD Selection on key, header item join
	Sales Order by Period		Range selection, no join
	Open Bills (Items)		Selection by non-key attributes, header item join, order by FK attribute and PK attribute
	Customer Details		MD Selection on key, 2-table join
	Product Details		MD Selection on key, 2-table join
OLAP	Daily Flash	Random/Con-figurable	TD aggregate range selection, header item join, group by 3 sales hierarchy attributes, order by aggregate
	Avg. Order Processing Time		TD aggregate select on independent range sub-select (4-table header item join), group by sales hierarchy, order by aggreg.
	Order Delivery Fulfillment		TD aggregate range select, 4-table header item join, group by sales hierarchy, order by aggregate, dependent aggregate range sub-select with header item join
	Days Sales Outstanding		TD aggregate range select, header item join, group by currency, order by aggregate, dependent aggregate range sub-select with header item join

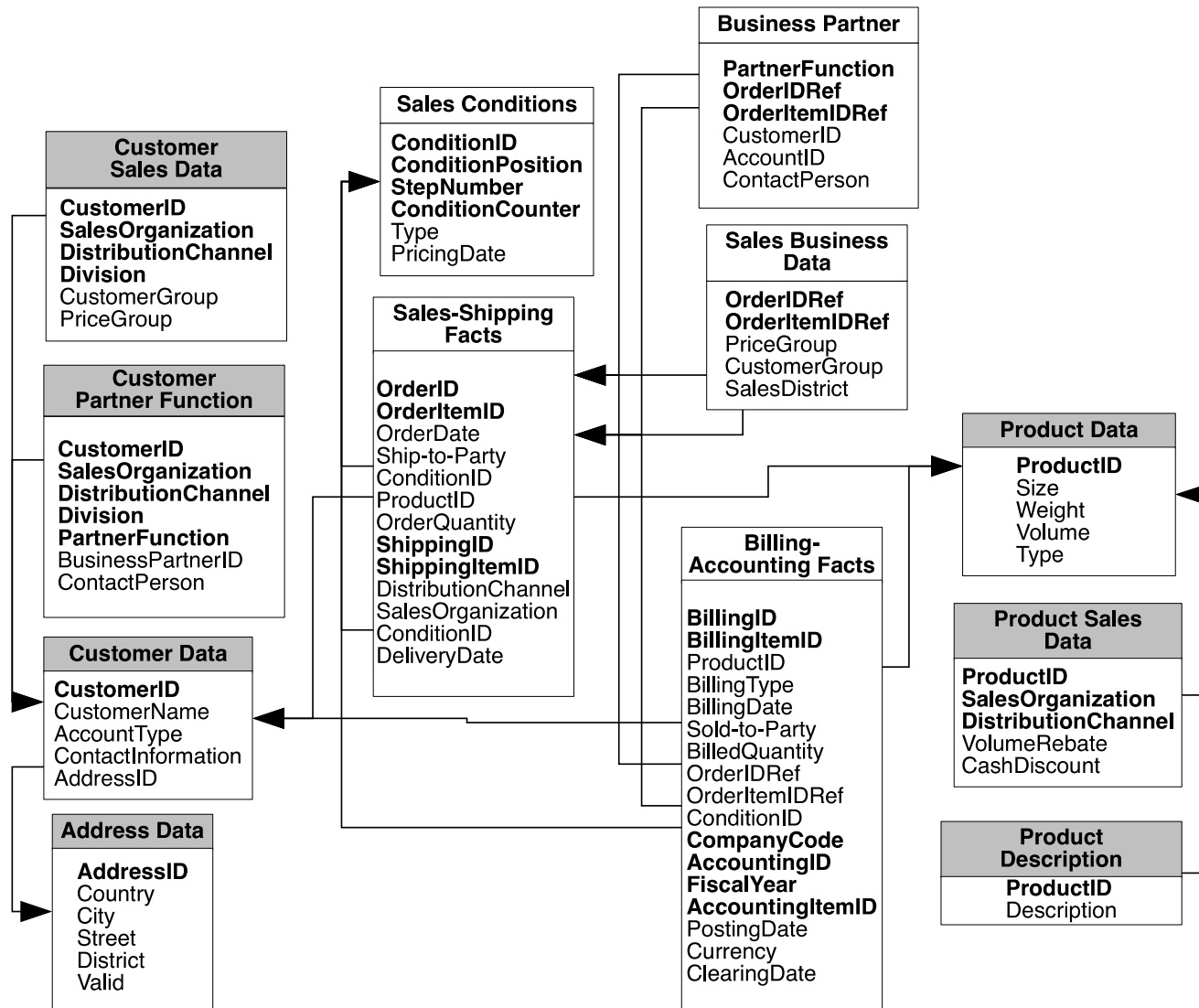
Schema Variant I – Document Style

9



Schema Variant II – Star Schema Style

10



Experimental Setup

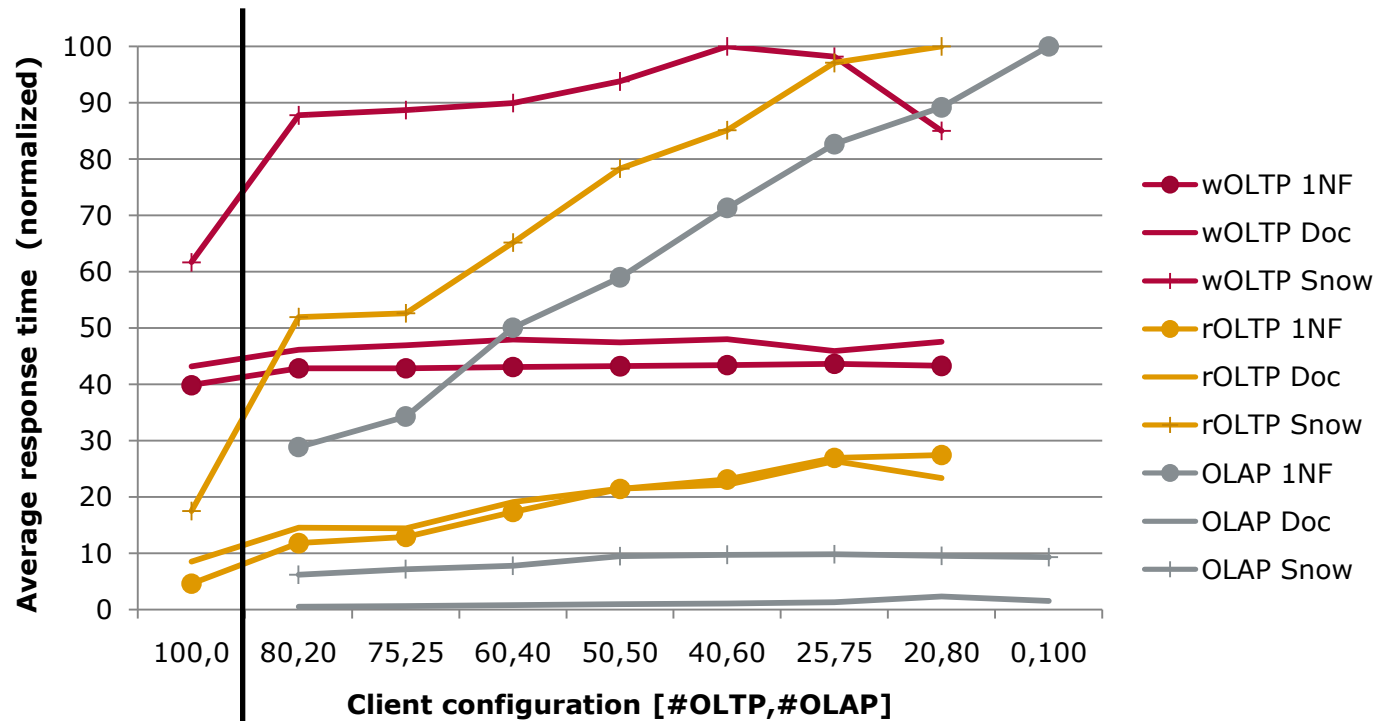
11

- Database types in test (chosen according to their usage in productive enterprise systems)
 - In-memory column-oriented
 - Disk-based column-oriented
 - Disk based row-oriented

- Workload simulation
 - 100 concurrent clients sending requests to the database
 - Varying share of OLTP and OLAP clients

Results – Disk-based row-oriented

12



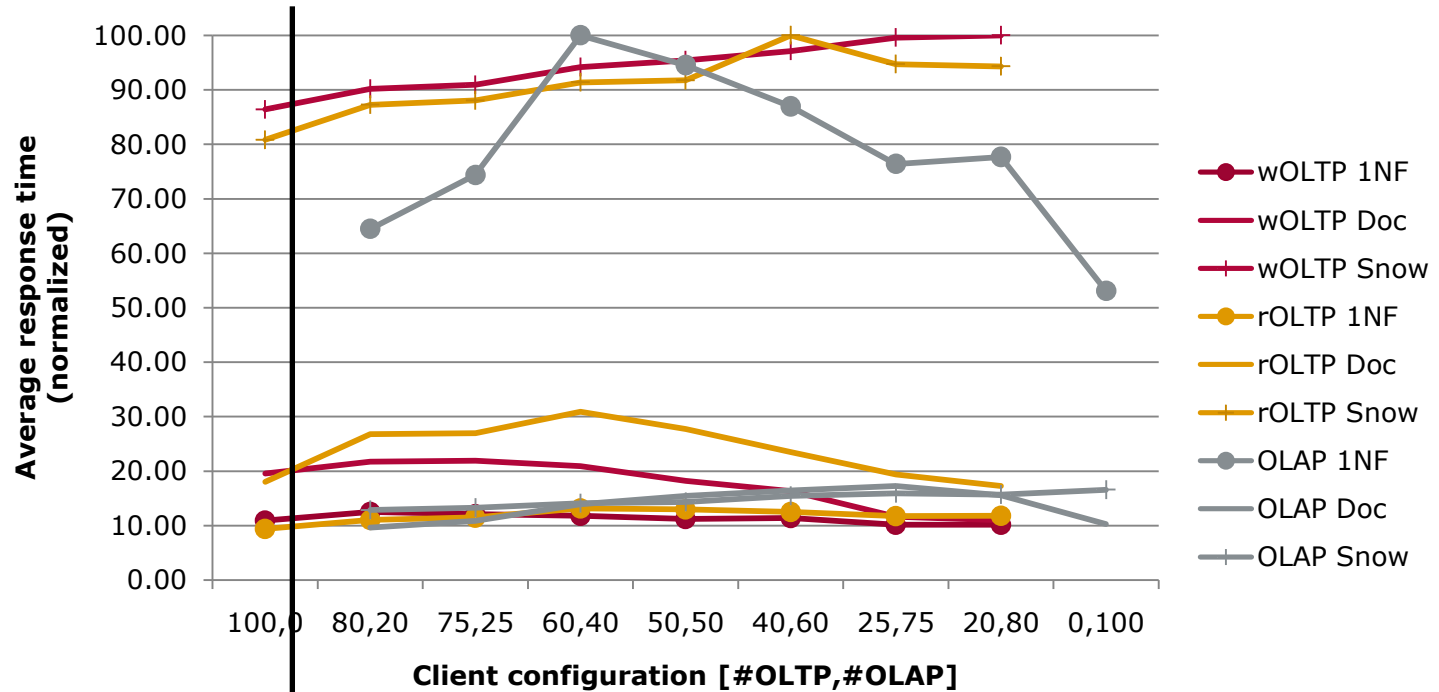
Favored Design: 1NF

Document

Type	Doc/1NF	Snow/1NF	1NF/Doc	Snow/Doc
wOLTP	+8%	+54%	-10% - -5%	+78% - +114%
rOLTP	+85%	+279%	-19% - +17%	+240% - +327%
OLAP	N/A	N/A	+3681% - +6220%	+305% - +1041%

Results – Disk-based column-oriented (I)

13



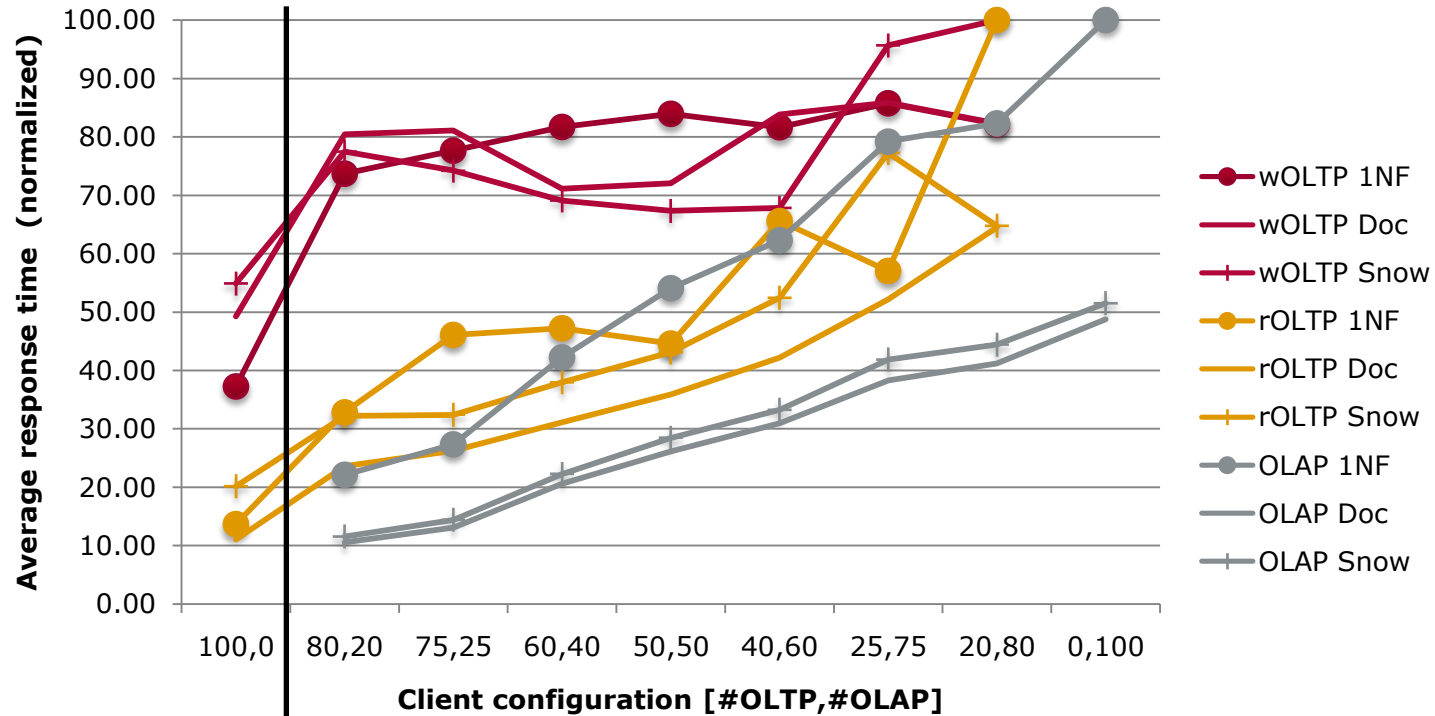
Favored Design: 1NF

Document

Type	Doc/1NF	Snow/1NF	1NF/Doc	Snow/Doc
wOLTP	+79%	+692%	-44% - -7%	+314% - +812%
rOLTP	+91%	+760%	-58% - -31%	+196% - +760%
OLAP	N/A	N/A	+342% - +619%	-7% - +61%

Results – Disk-based column-oriented (II)

14



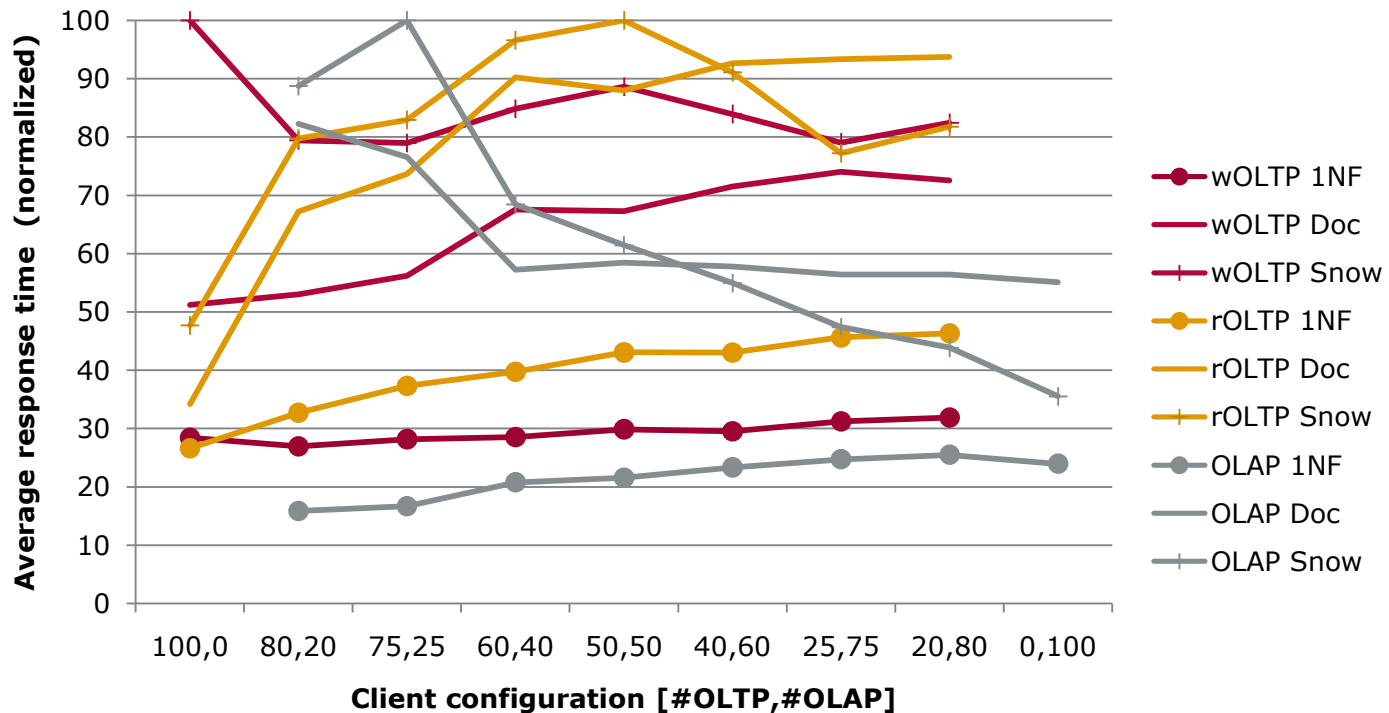
Favored Design: 1NF

Document

Type	Doc/1NF	Snow/1NF	1NF/Doc	Snow/Doc
wOLTP	+32%	+47%	-0.2 - +16%	-19% - +47%
rOLTP	-18%	+47%	+9% - +75%	+0.4 - +48%
OLAP	N/A	N/A	+99% - +109%	+5% - +9%

Results – In-memory column-oriented

15



Favored Design: 1NF

Type	Doc/1NF	Snow/1NF
wOLTP	+80% - +143%	+153% - +251%
rOLTP	+28% - +127%	+69% - +143%
OLAP	+121% - +418%	+72% - +498%

Conclusion

16

- Manually created schema variants, tested via benchmark framework that simulates real company order-to-cash composite analytical and transactional workloads
- Not necessarily beneficial to have a static database design in a mixed workload scenario when the workload varies
- In Future: Utilize strategies from database design advisors for autonomous schema variation to adapt normalization in schemas to changing workloads?

Thanks!

Please visit

<http://epic.hpi.uni-potsdam.de>

to find out more about our research.