

Analytische Datenbanken

Test und Evaluierung ausgewählter Produkte

Inhaltsverzeichnis

1. Definition
2. Nutzen
3. Techniken
4. Datenbankauswahl
5. Testszenario
6. Ergebnisse
7. Ausblick

Definition

- Ausrichtung auf analytische Anwendungen
- Abfrageorientierung
- Vernachlässigung der Transaktionsverarbeitung



Nutzen

- Auswertungen größerer Datenmengen
 - Verbesserte Erfolgskontrolle
 - Verbesserte Transparenz
- Hohe Abfrageperformance
 - Echtzeitanalytik
- Geringer Optimierungsaufwand
 - Entlastung der entsprechenden Fachabteilung

Techniken

- Grundlegende Unterscheidung

- Relational vs. NoSQL

- Speichertyp

- Spaltenorientierung vs. Zeilenorientierung

- Speicherort

- Festplatte vs. Arbeitsspeicher(In-Memory)

- Architektur

- MPP vs. Single Server

Spaltenorientierung

- Spaltenweise Datenspeicherung
- Verbesserte Komprimierung
- Keine Verwendung von Indizes
- Datensätze > Attribute

PNR	Name	Alter	Gehalt
1	Josef	56	5000
2	Tom	29	4000



Zeilenorientierung

((1;Josef;56;5000),(2;Tom;29;4000))

Spaltenorientierung

((1;2),(Josef;Tom),(56;29),(5000;4000))

Kompression

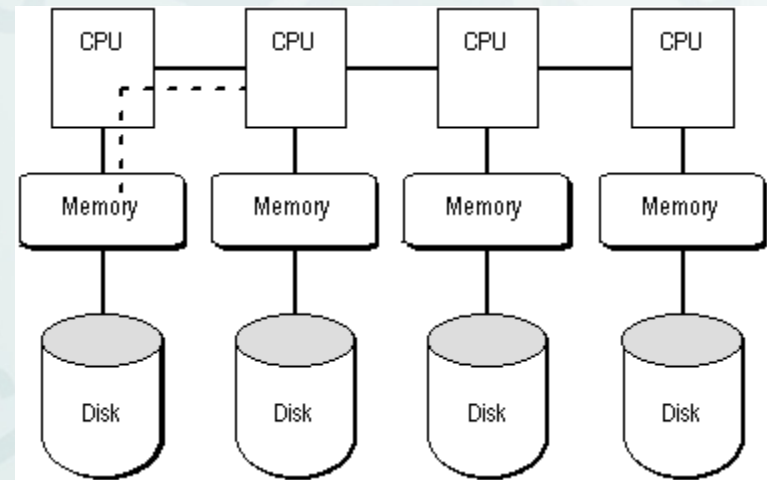
- Einfache, schnelle Techniken
- Komplexe Verfahren unbrauchbar
- Reduzierung des Speicherbedarfs
- Reduzierung der Übertragungszeit

Kompression

	Wörterbuchkompression	Laufängen- kodierung	Deltakodierung
Unkom- primiert	Feuer Wasser Feuer Wasser	F W F F W W	100 101 104 110 112
Kom- primiert	Wörterbuch: Feuer 00 Wasser 01 Daten: 00 01 00 01	1F1W3F2W	1 3 6 2

Massive Parallel Processing (MPP)

- Shared-Nothing-Architektur
- Mehrere baugleiche Serverknoten
- Verteilung des Arbeitsaufwands



Datenbankauswahl

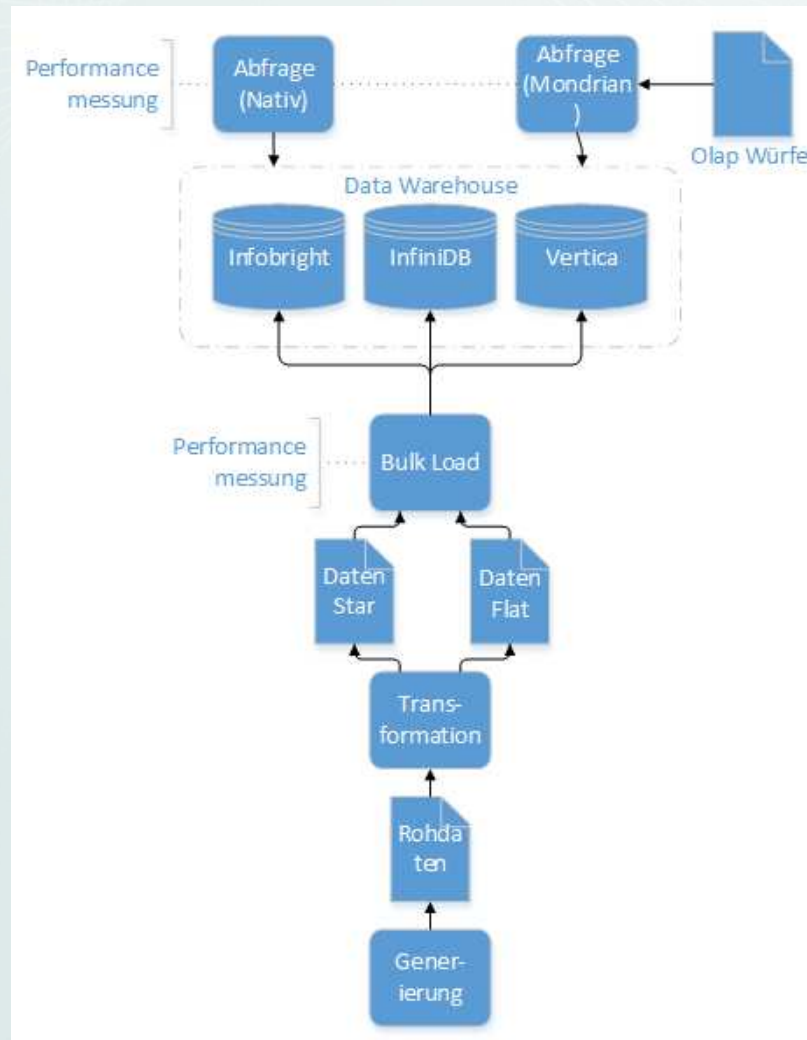
- Infobright
 - Community Edition
- InfiniDB
 - Community Edition
- Vertica
 - Community Edition



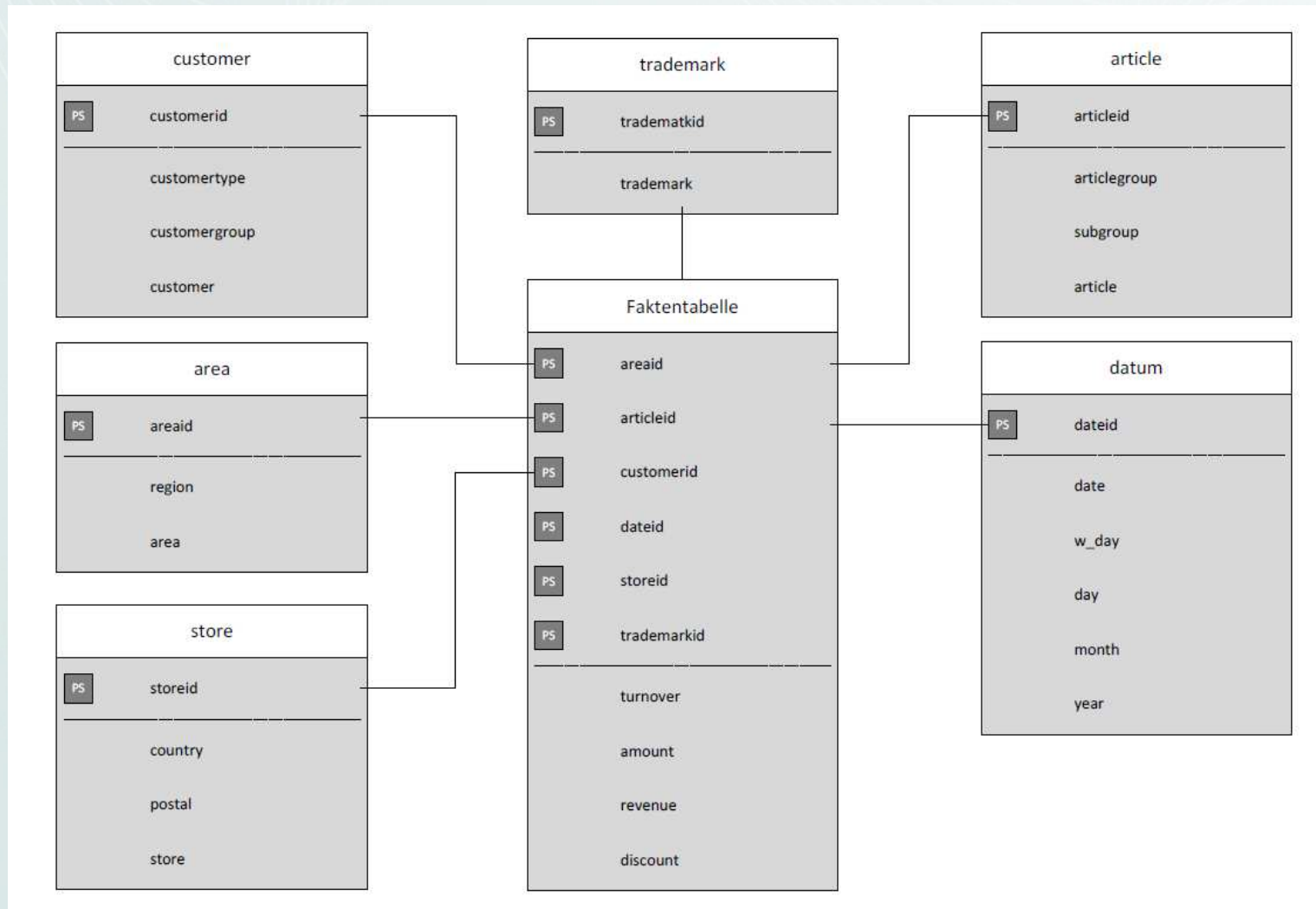
TestszENARIO

- Testsystem(Virtuelle Maschine)
 - Red Hat 5.8
 - 32 GB Arbeitsspeicher
 - 4 Duale Core Prozessoren
- Verschiedene Schemata/ Datengrößen/ Abfragen
- Steigende Abfragekomplexität

Testszenario



Testszenario



Testszenario

● Abfrage 1

```
SELECT t.trademark, sum(f.turnover)
FROM facts_star f
LEFT OUTER JOIN trademark t ON
f.trademarkid = t.trademarkid
WHERE trademark = 'ARBVXI'
Group by t.trademark
;
```

● Abfrage 2

```
SELECT t.trademark,s.postal, s.store, sum(f.turnover),sum(f.amount)
FROM facts_star f
LEFT OUTER JOIN trademark t ON f.trademarkid = t.trademarkid
LEFT OUTER JOIN store s ON f.storeid = s.storeid
WHERE t.trademark = 'ARBVXI'
AND s.postal = 'Austria-1'
Group by t.trademark, s.postal, s.store;
```

Testszenario

● Abfrage 3

```
SELECT t.trademark,s.store,d.year,a.region, sum(f.turnover), sum(amount), sum(revenue)
FROM facts_star f
LEFT OUTER JOIN  trademark t ON f.trademarkid = t.trademarkid
LEFT OUTER JOIN  store s ON f.storeid = s.storeid
LEFT OUTER JOIN  datum d ON f.dateid = d.date_id
LEFT OUTER JOIN  area a ON f.areaaid = a.areaaid
WHERE t.trademark = 'ARBVXI'
AND s.postal = 'Austria-1'
AND (a.region = 'Central' OR a.region = 'North' OR a.region ='West')
Group by t.trademark, s.store, d.year,a.region;
```

Testszenario

● Abfrage 4

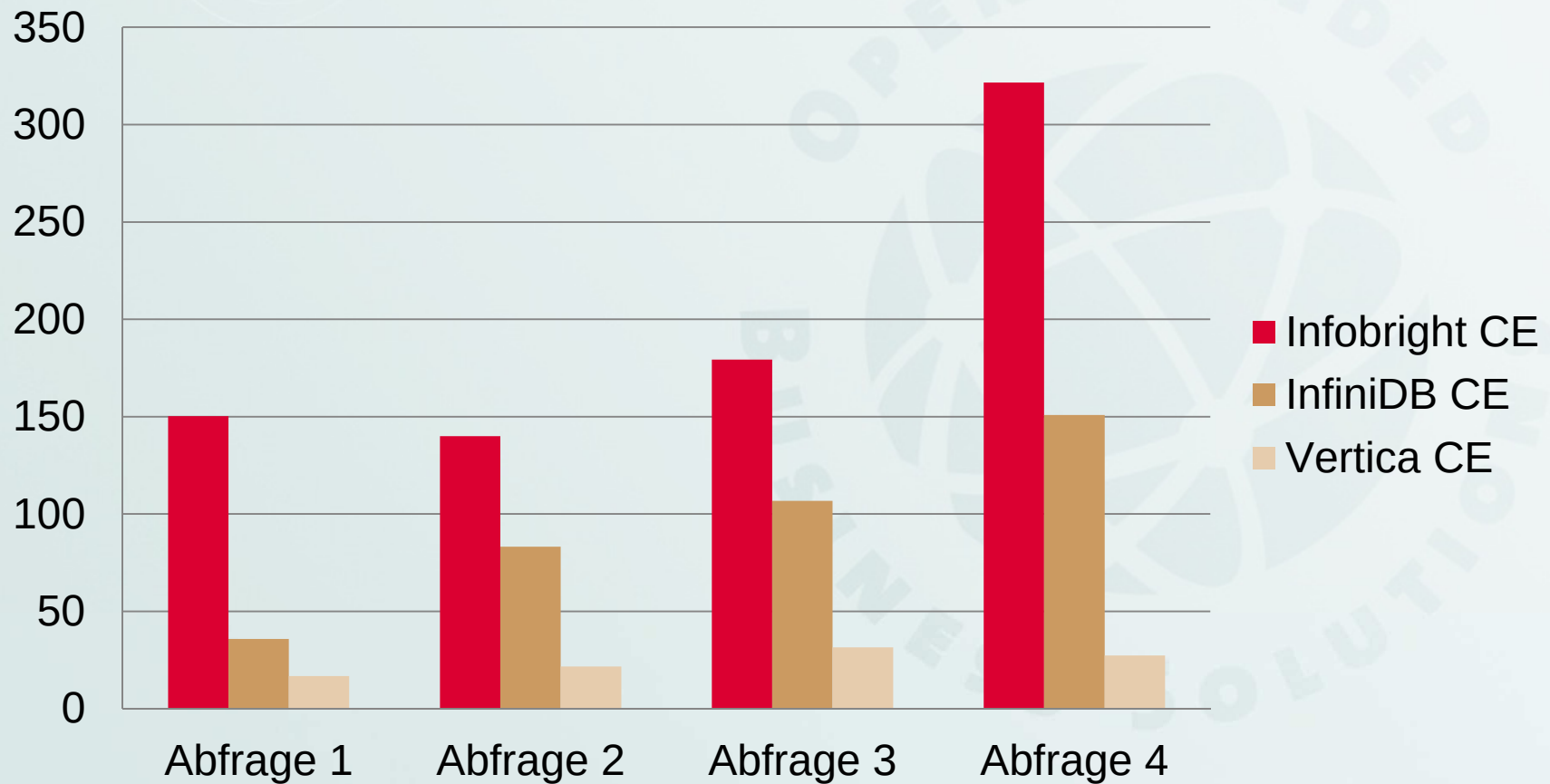
```
SELECT t.trademark,s.store,d.year,a.region,c.customertype,  
f.subgroup, r.articlegroup, sum(f.turnover), sum(amount), sum(revenue), sum(discount)  
FROM facts_star f  
LEFT OUTER JOIN  trademark t ON f.trademarkid = t.trademarkid  
LEFT OUTER JOIN  store s ON f.storeid = s.storeid  
LEFT OUTER JOIN  datum d ON f.dateid = d.dateid  
LEFT OUTER JOIN  area a ON f.areaaid = a.areaaid  
LEFT OUTER JOIN  customer c ON f.customerid = c.customerid  
LEFT OUTER JOIN  article r ON f.articleid= r.articleid  
WHERE t.trademark = 'ARBVXI'  
AND s.postal = 'Austria-1'  
AND (a.region = 'Central' OR a.region = 'North' OR a.region ='West')  
AND d.year = '2004'  
Group by t.trademark, s.store, d.year,a.region,c.customertype, r.subgroup, r.articlegroup;
```

Mondrian Abfragen

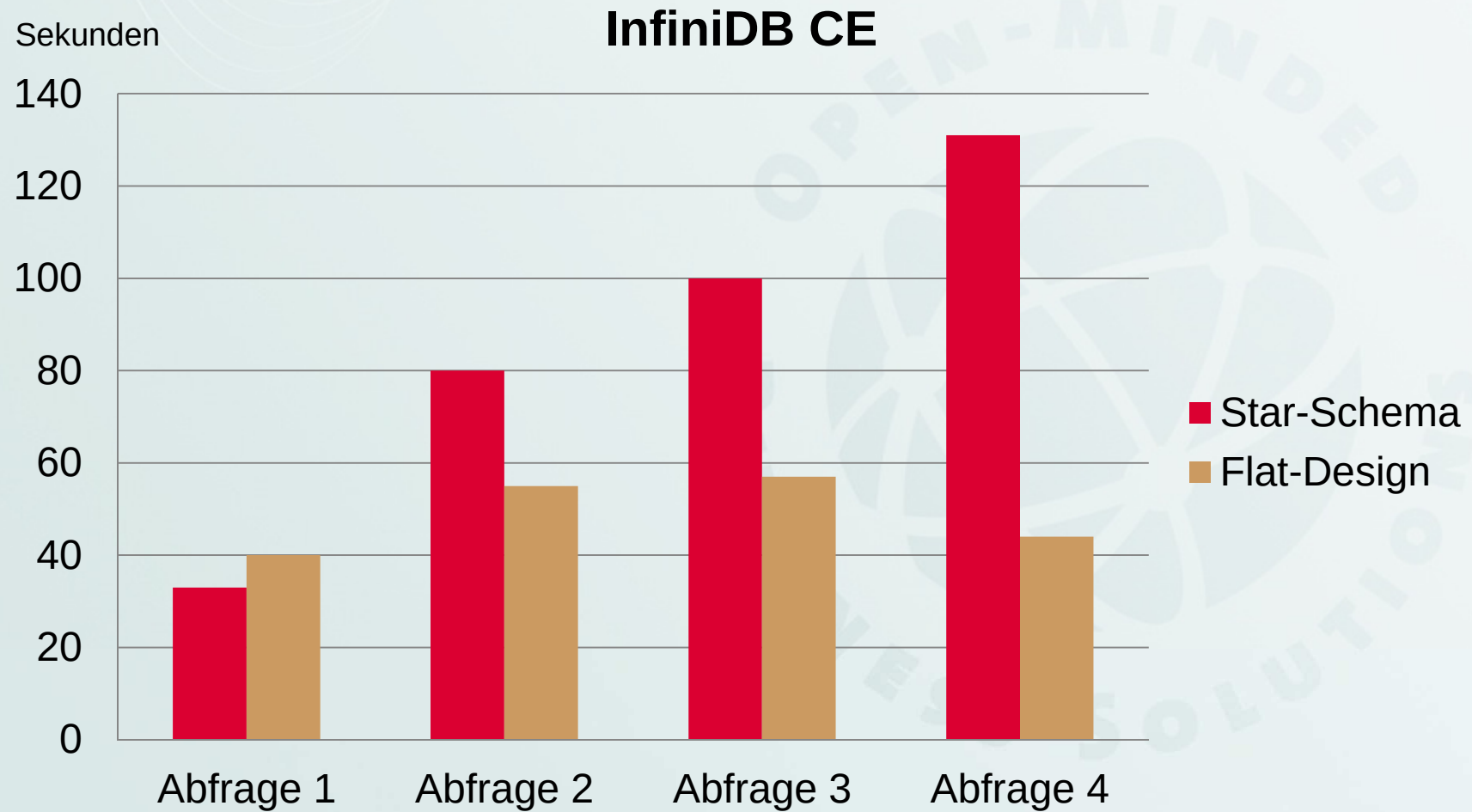
Sekunden

Star Schema

Bei den Infobright Ergebnissen handelt es sich um vorläufige Werte. Die endgültigen Ergebnisse werden auf der Webseite der it-novum veröffentlicht.



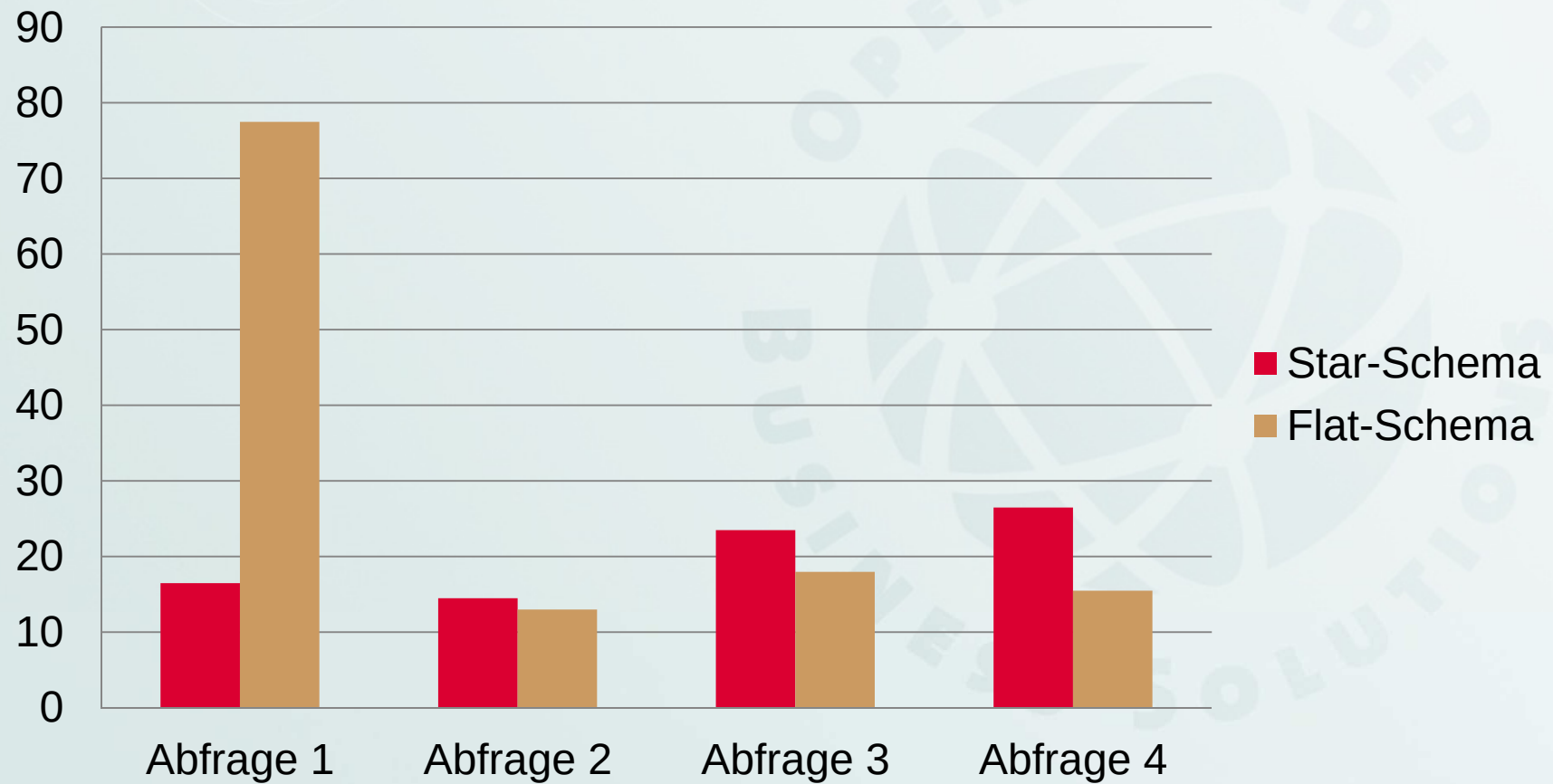
Native Abfragen



Native Abfragen

Sekunden

Vertica CE

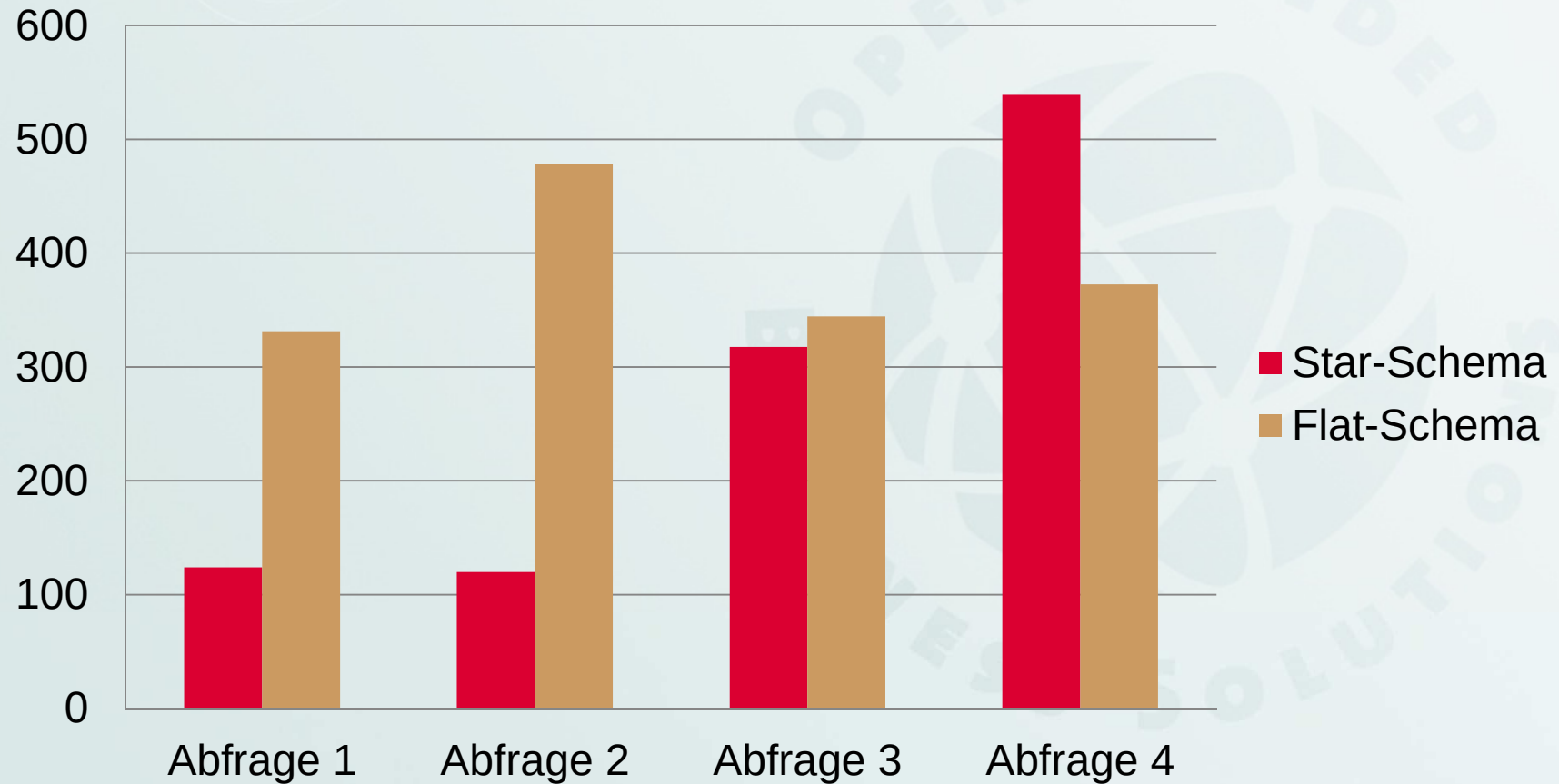


Native Abfragen

Sekunden

Infobright CE

Bei den Infobright Ergebnissen handelt es sich um vorläufige Werte. Die endgültigen Ergebnisse werden auf der Webseite der it-novum veröffentlicht.



Ausblick

