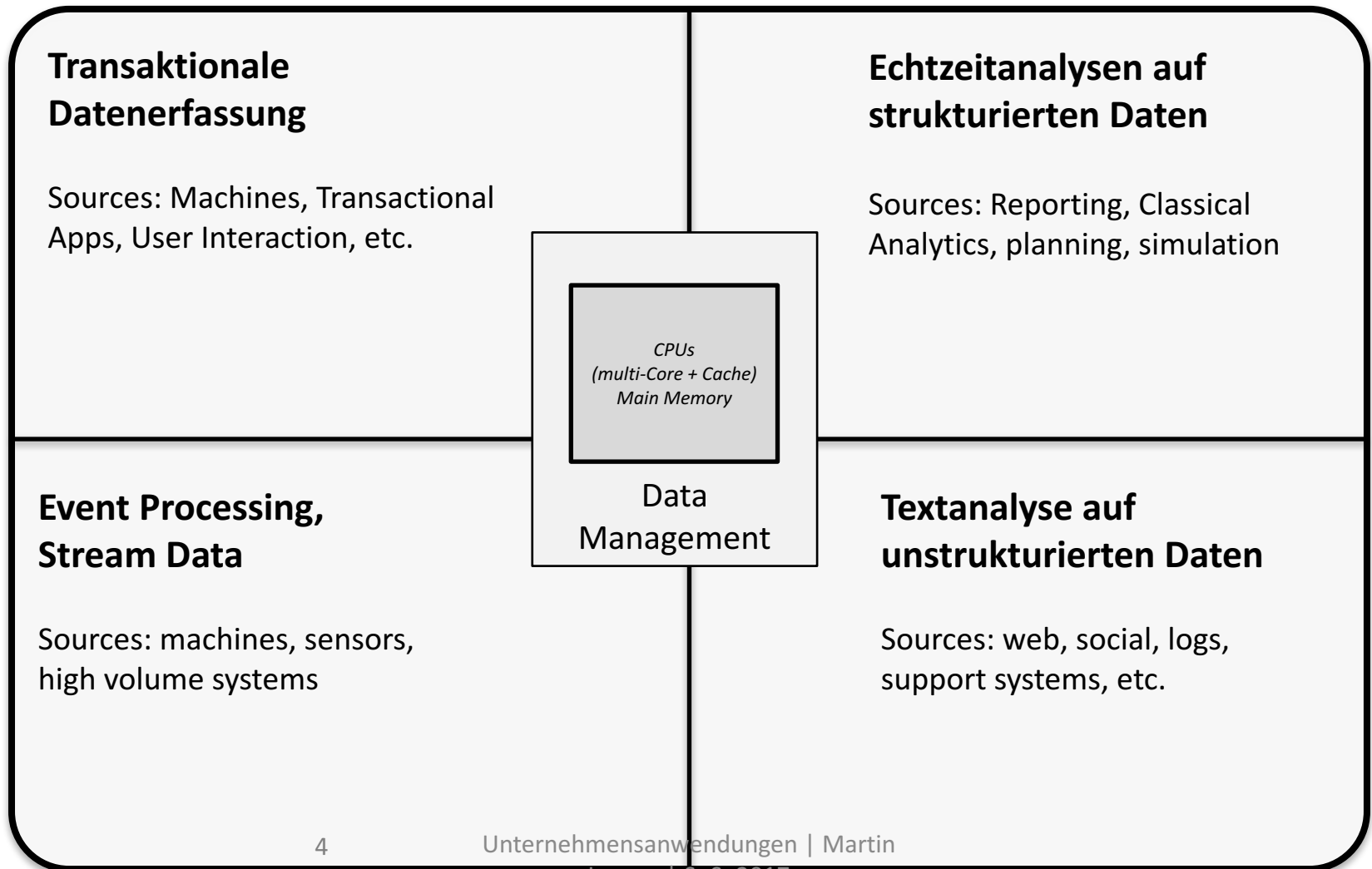


Charakteristika von Unternehmensanwendungen

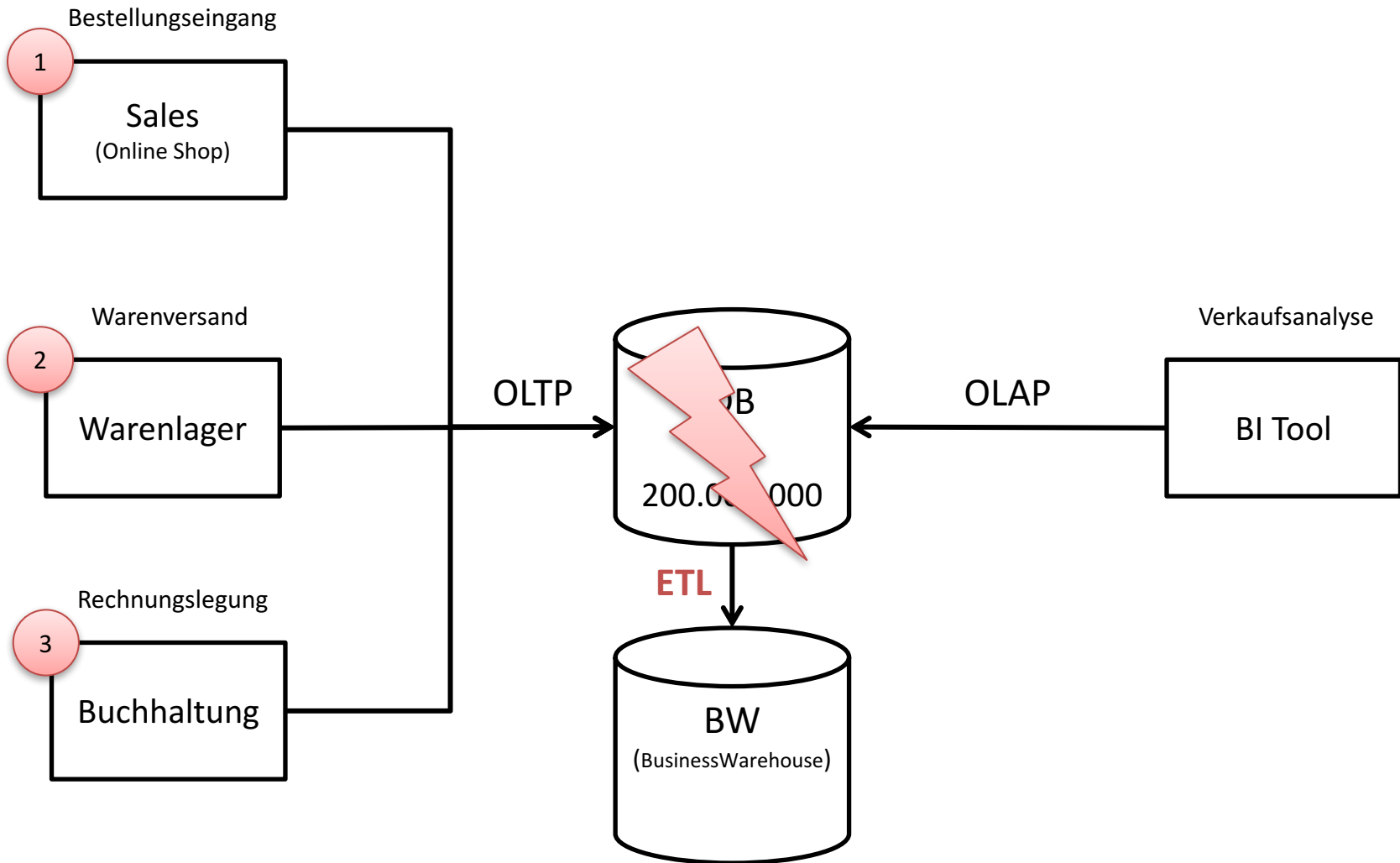
Die Essenz

- Große Mengen komplexer Daten, die in der realen Welt existierende Entitäten abbilden
- Automatisierung von Geschäftsprozessen auf Basis dieser Daten
- Business Intelligence (BI) und Entscheidungsunterstützung, auf Basis dieser Daten

Herausforderung: Fassettenreiche Anwendungen



Rückblick



OLTP vs. OLAP

Online **T**ransaction
Processing

Online **A**nalytical
Processing

- Moderne ERP-Systeme müssen sich aber zunehmend mit **Mixed Workloads** arrangieren.

Beispiele:

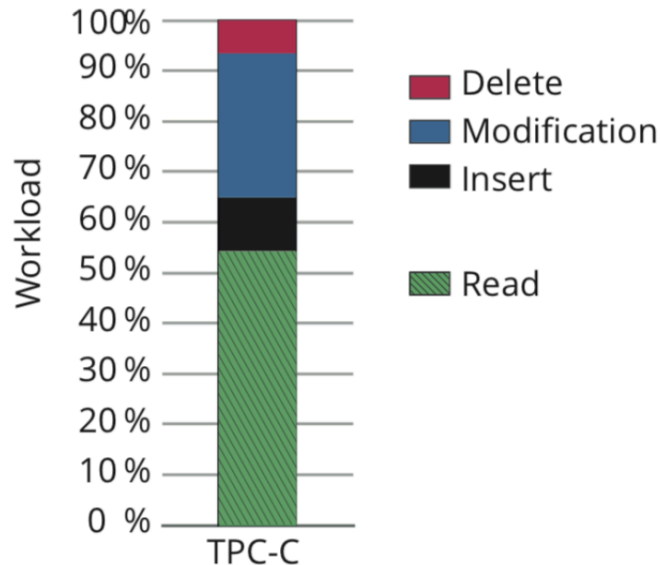
- OLTP: Bestellung anlegen, Rechnung erstellen, Kontenbuchung, Anzeige von Kundendaten oder Bestellungen
 - OLAP: Mahnlauf, Verfügbarkeitsprüfung, Cross-Selling, Operationales Reporting (Auflistung offener Rechnungen)
-
- Aber: Heutige Anwendungen sind entweder für OLTP **ODER** OLAP Workloads optimiert.

Nachteile dieser Trennung

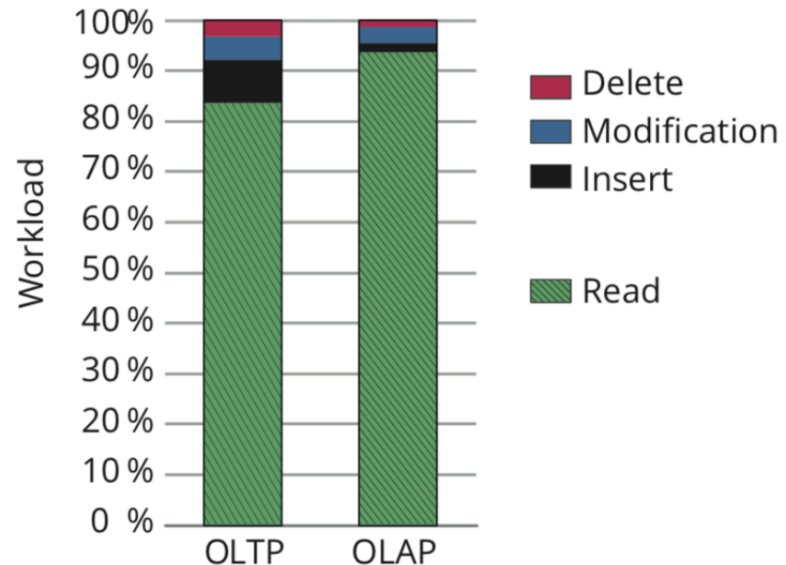
- ❑ OLAP-Systeme haben nicht die **neuesten** Daten
- ❑ OLAP-Systeme halten ein **vordefinierte Untermenge** aller Daten
- ❑ **Kostenintensiver ETL** Prozess muss beide System synchronisieren
- ❑ Es besteht große **Redundanz der Daten**
- ❑ **Unterschiedliche Daten Schemata** erzeugen höhere Komplexität für Anwendungen, die beide Systeme nutzen

OLTP Datenzugriffs Mythos

Workload Analyse von Unternehmensanwendungen zeigt:
OLTP und OLAP Workloads sind **NICHT** unterschiedlich



TPC-C Benchmark



Workload Analysis
Krueger et al. VLDB'11

Vision

Kombination von OLTP und OLAP Daten

durch Nutzung moderner Hardware und
Datenbanksysteme

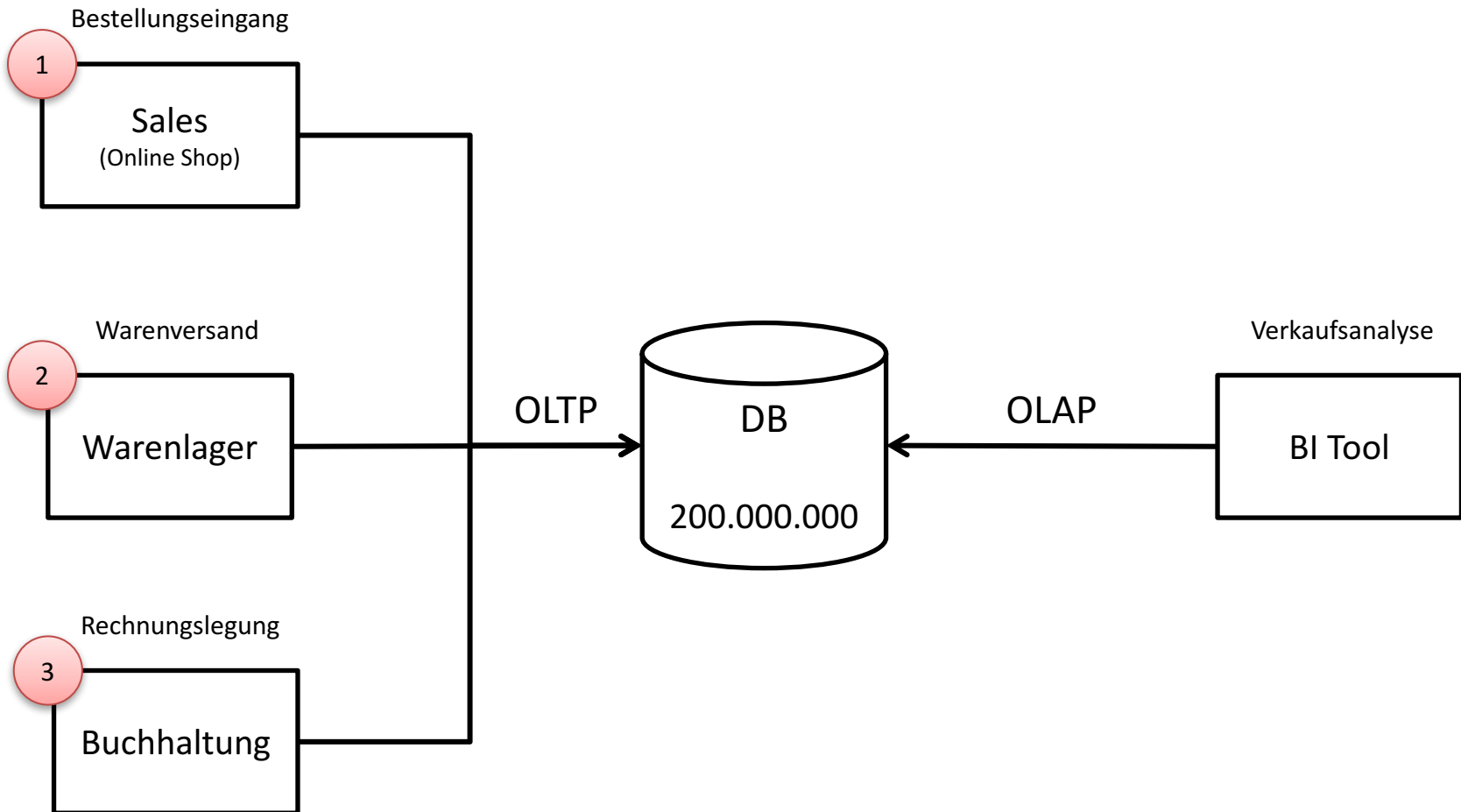
um eine **Single Source of Truth,**

für **Echtzeitanalysen, vereinfachte** Anwendungen und
Datenbankstrukturen zu nutzen.

Abschaffung von,

- Extraction, Transformation, und Loading (ETL) Prozesse
- Vorberechnete Aggregate und ausmaterialisierte Views.

Vision



Mixed Workload Charakteristika

Komplette Zeilenoperationen **und** Materialisierung weniger Spalten

Einfache **und** komplexe Abfragen

Tuple Wiederherstellung **und** Aggregationen und Group Bys

INSERTS, UPDATES **und** viele SELECTS

Kurze **und** lange Transaktionen

Kleine **und** große Ergebnisse (Anzahl Zeilen)

Vordefinierte **und** adhoc Abfragen

Echtzeitveränderungen

“Single Source of Truth”

Enterprise Data Characteristics

- Viele Spalten werden NICHT benutzt (DEFAULT Werte)
- Viele Spalten haben eine geringe Kardinalität
- NULL values/default values are dominant
- Zerstreute Verteilung erlaubt hohe Kompression

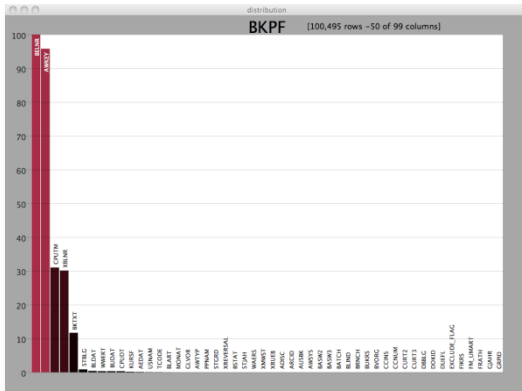
Standard enterprise software data is **sparse and wide**

Viele Spalten mit geringer Anzahl von Werten

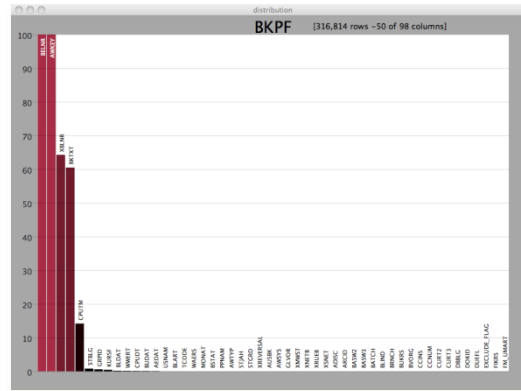
Analyse eines Buchhaltungssystems

Distinct values in accounting document headers (99 attributes)

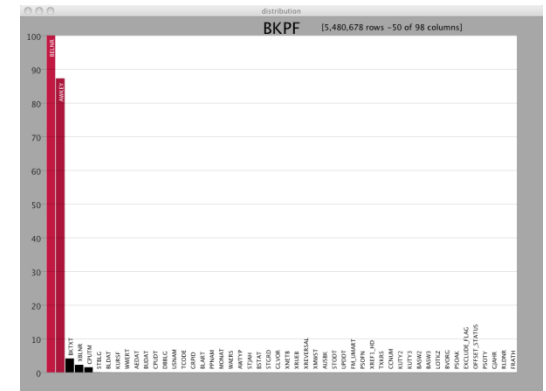
CPG



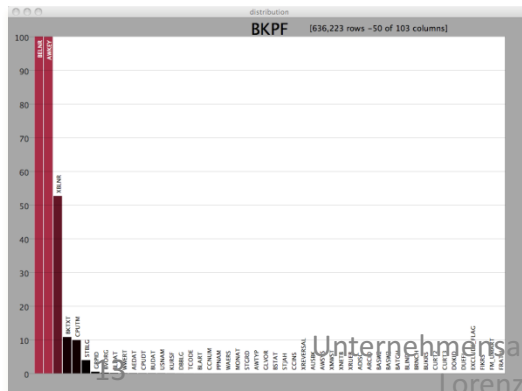
Logistics



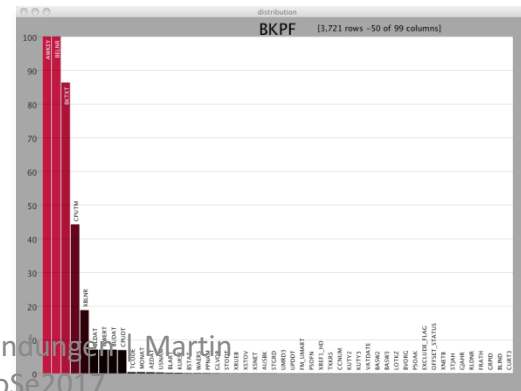
Banking



High tech



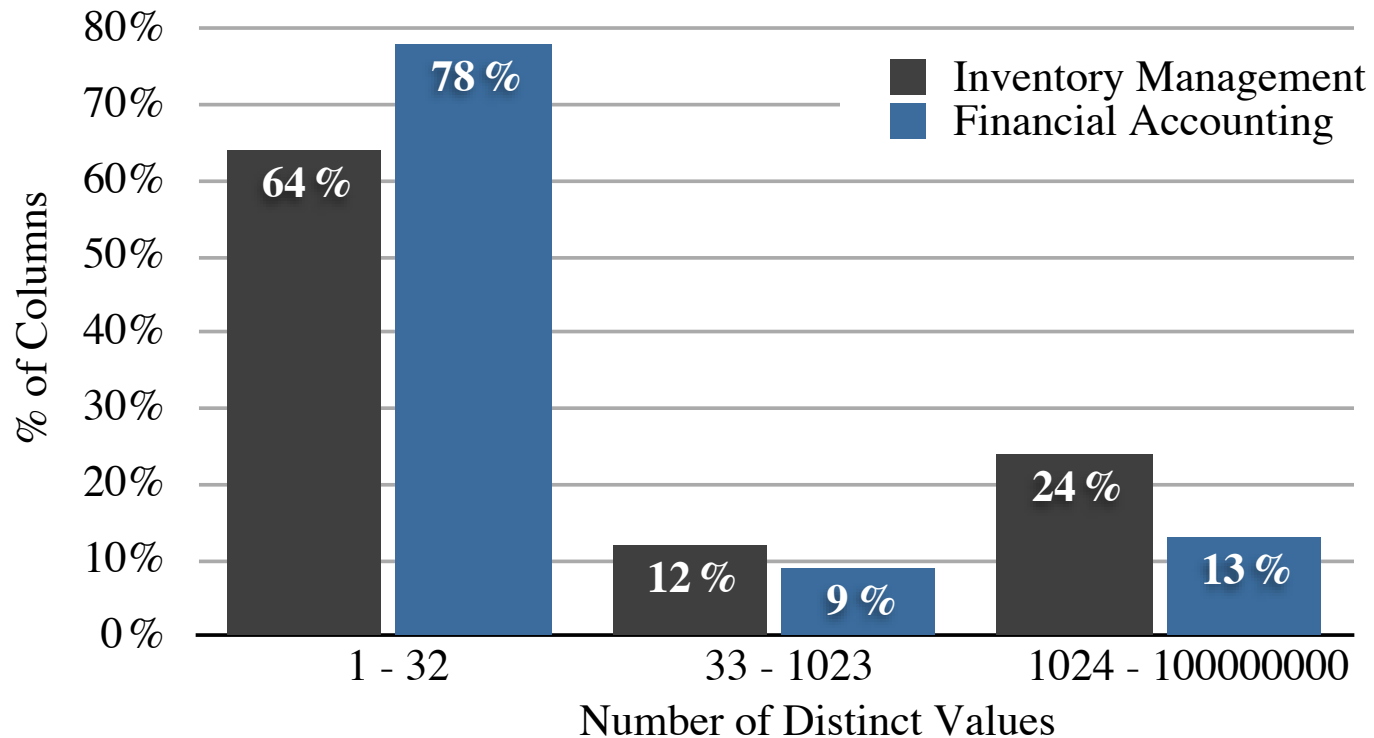
Discrete manufacturing



Many Columns are not Used Even Once

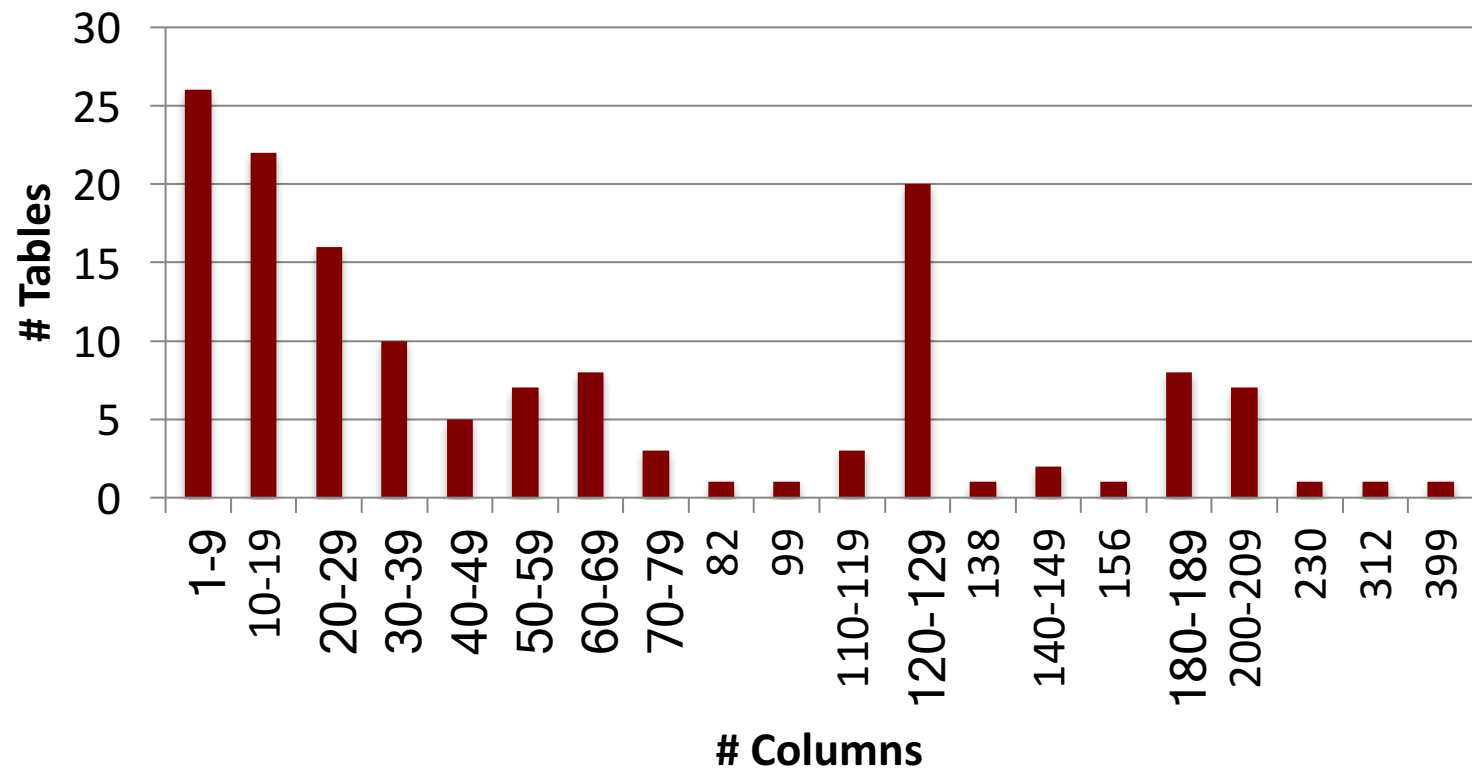
55% aller Spalten werden (im Durchschnitt) nicht benutzt

40% aller Spalten werden durch gar kein Unternehmen benutzt



Wide Tables

Analysis of width of 144 most used* tables



Herausforderungen für Unternehmensanwendungen

- Riesige Codebasis
- Verarbeitung riesiger Datenmengen
- Verteilter/paralleler Zugriff
- Konzeptionelle Dissonanzen
 - Kundenstruktur für Abt. 1 <> Kundenstruktur für Abt. 2
- Komplexe “Geschäfts(un)logik”
- Veränderungen des Datenschema durch
 - von Innen (e.g., Optimierung)
 - von aussen (Evolution von Entitäten, neue Funktionalität)