

Charakteristika von Unternehmensanwendungen

Wozu werden Unternehmensanwendungen genutzt?

“Enterprise applications are about the **display, manipulation, and storage of large amounts of often complex data** and the **support and automation of business processes** with that data.”

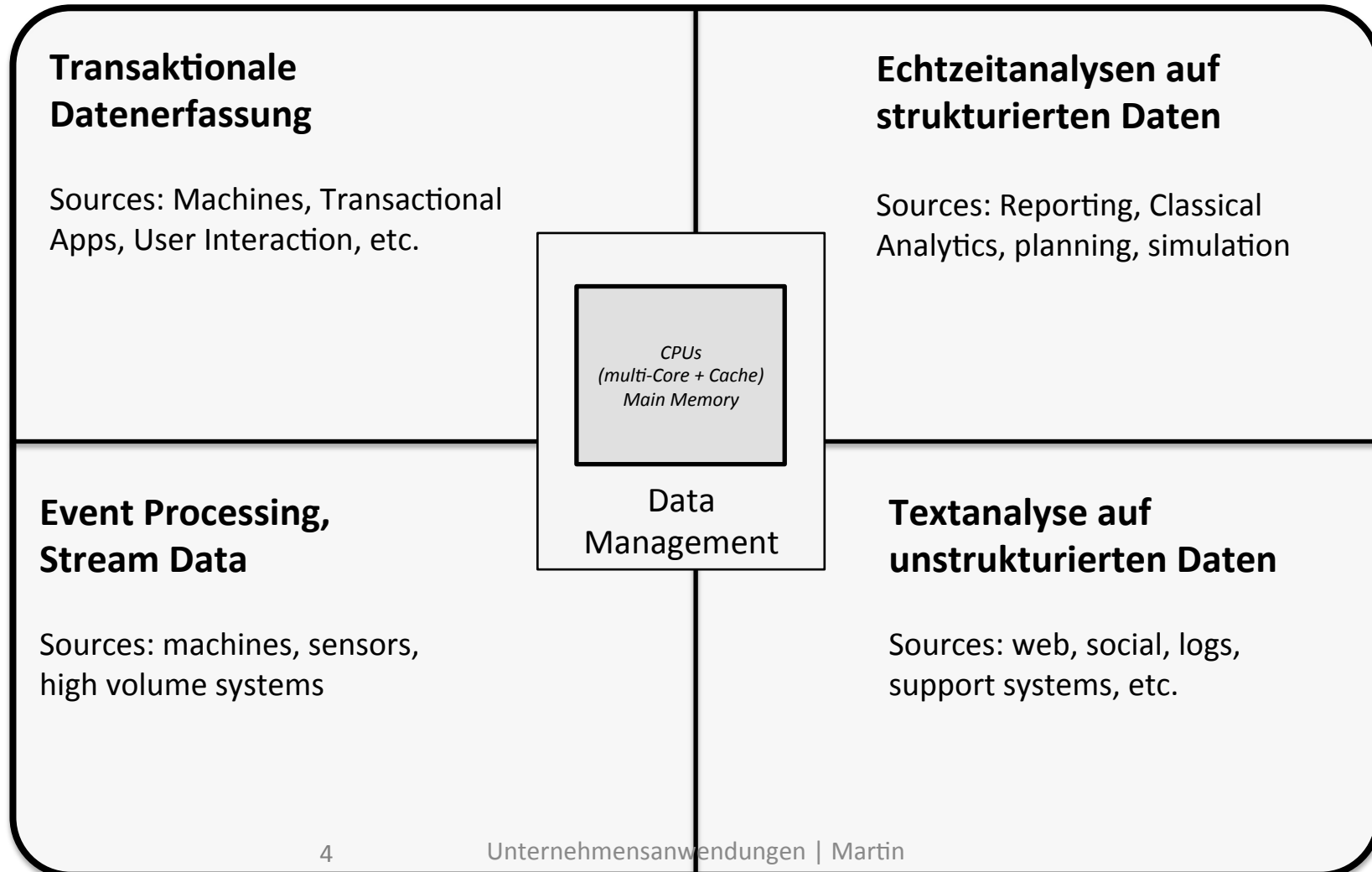
Martin Fowler „Patterns of Enterprise Application Architecture Patterns“ (2002)



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



OLTP vs. OLAP

Online **T**ransaction
Processing

Online **A**nalytical
Processing

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

Beispiele:

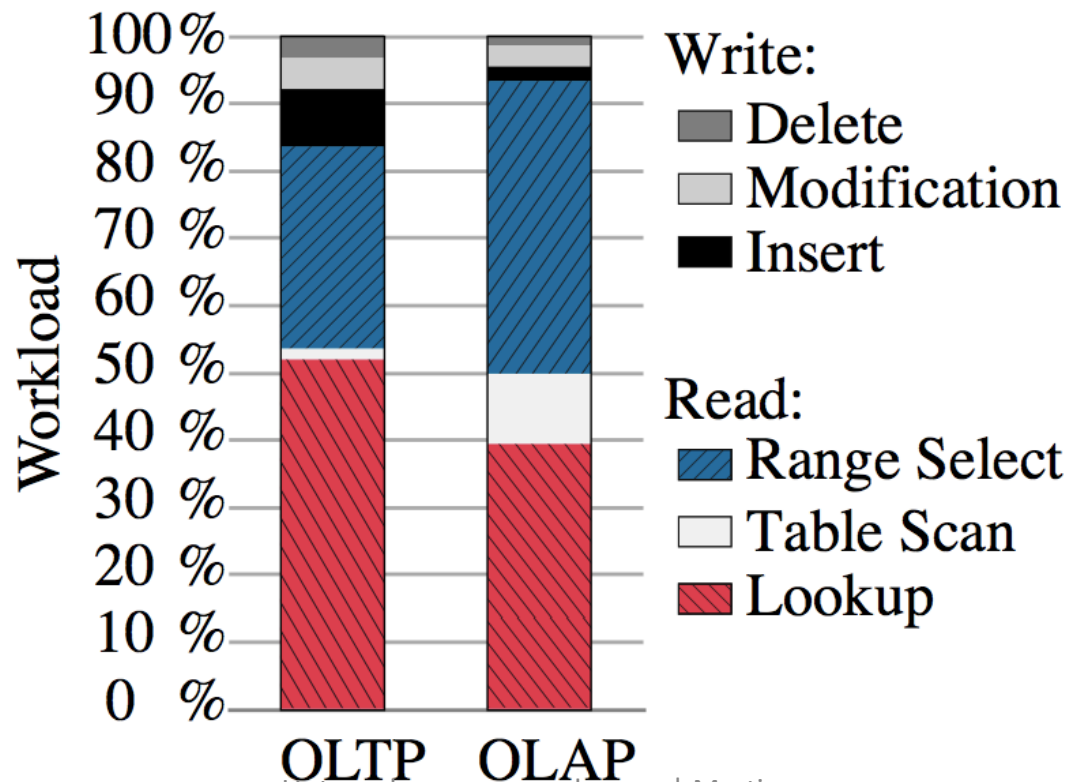
- OLTP: Bestellung anlegen, Rechnung erstellen, Kontenbuchung, Anzeige von Kundendaten oder Bestellungen
 - OLAP: Mahnlauf, Verfügbarkeitsprüfung, Cross-Selling, Operationales Reporting (Auflistung offener Rechnungen)
-
- Aber: Heutige Anwendungen sind entweder für OLTP **ODER** OLAP Workloads optimiert.

Nachteile dieser Trennung

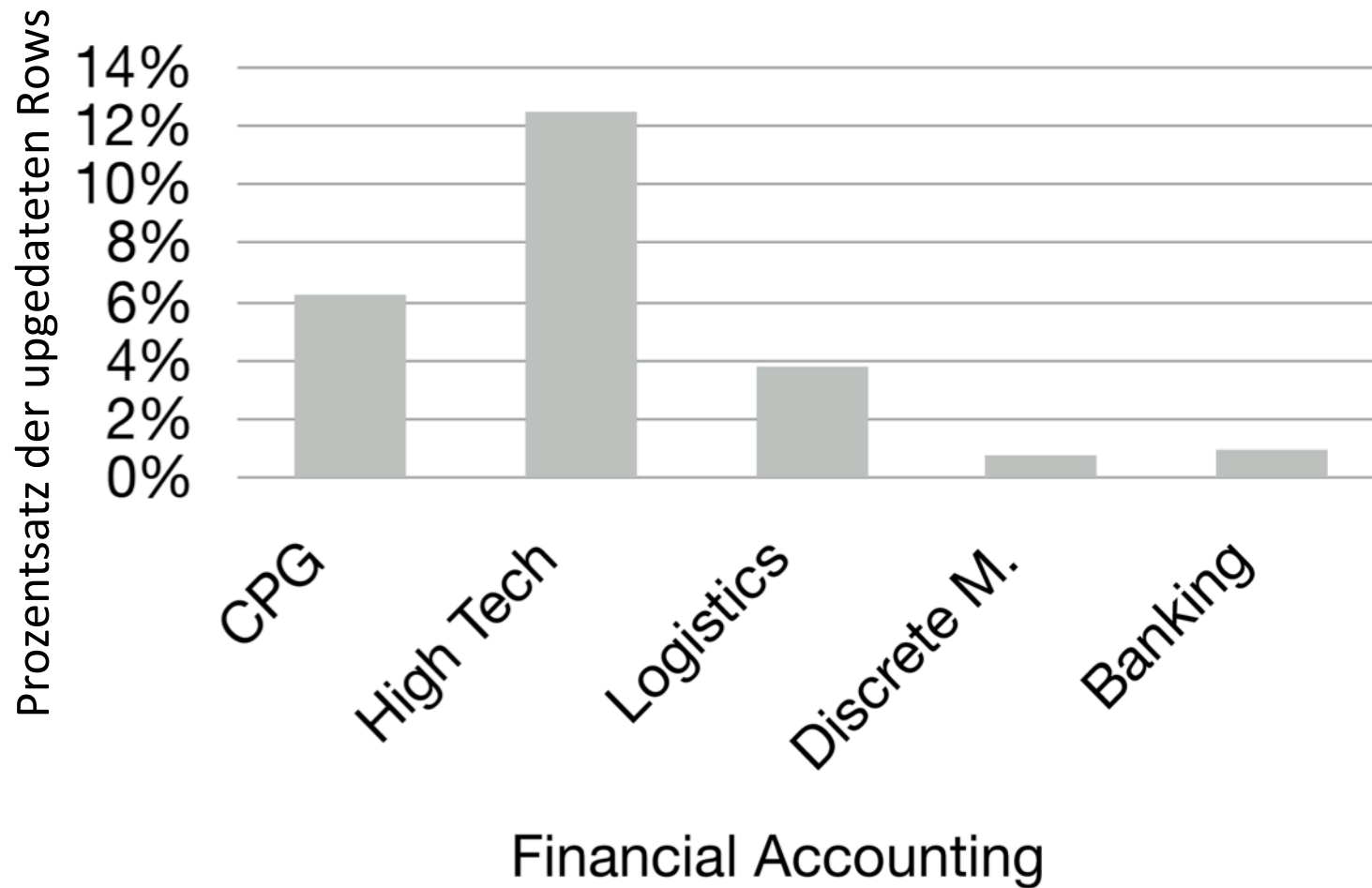
- ❑ OLAP-Systeme haben nicht die **neuesten** Daten
- ❑ OLAP-Systeme halten ein **vordefinierte Untermenge** aller Daten
- ❑ **Kostenintensiver ETL** Prozess muss beide System synchronisieren
- ❑ Es besteht große **Redundanz der Daten**
- ❑ **Unterschiedliche Daten Schemata** erzeugen höhere Komplexität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

OLTP Datenzugriffs Mythos

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



Nur wenig Updates bei OLTP



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.

Mixed Workload Charakteristika

Full row operations **and** retrieve small number of columns

Simple queries **and** complex queries

Tuple reconstruction **and** aggregation and group by

Inserts, updates **and** mainly selects

Short transactions **and** long transactions

Small result sets **and** large result sets

Pre-determined queries **and** adhoc queries

Real time changes

“Single Source of Truth”

Enterprise Data Characteristics

- ❑ Many columns are not used even once
- ❑ Many columns have a low cardinality of values
- ❑ NULL values/default values are dominant
- ❑ Sparse distribution facilitates high compression

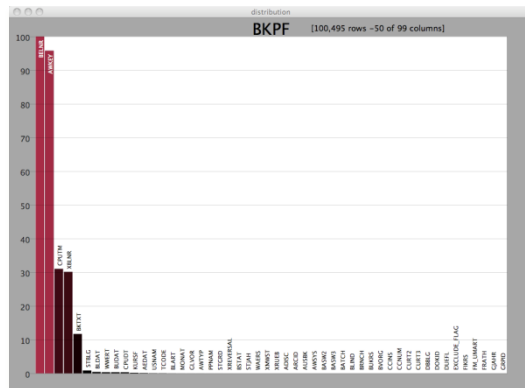
Standard enterprise software data is **sparse and wide**

Low Cardinality of Values Within Many Columns

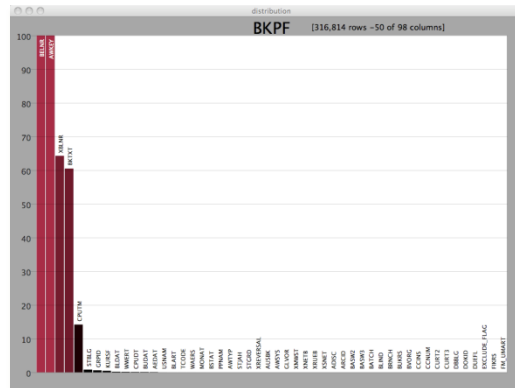
Results from analyzing financials

Distinct values in accounting document headers (99 attributes)

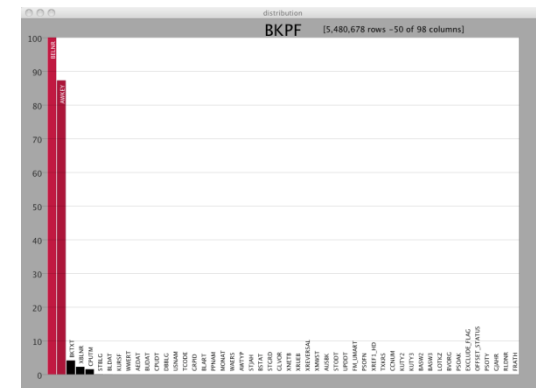
CPG



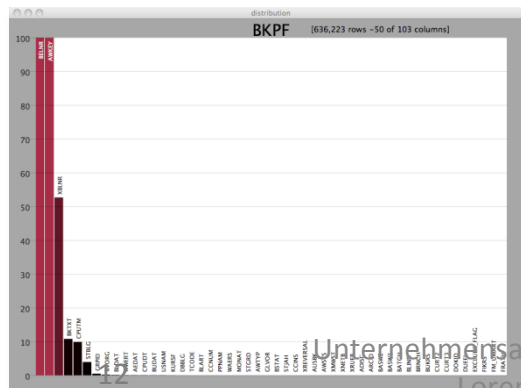
Logistics



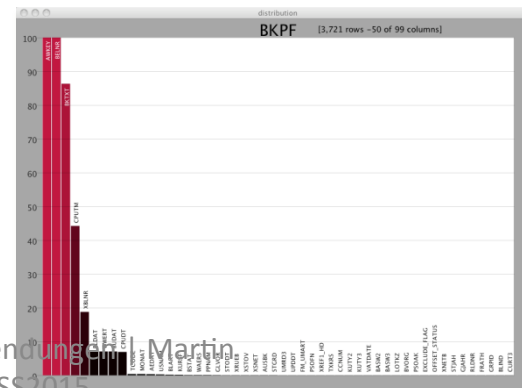
Banking



High tech



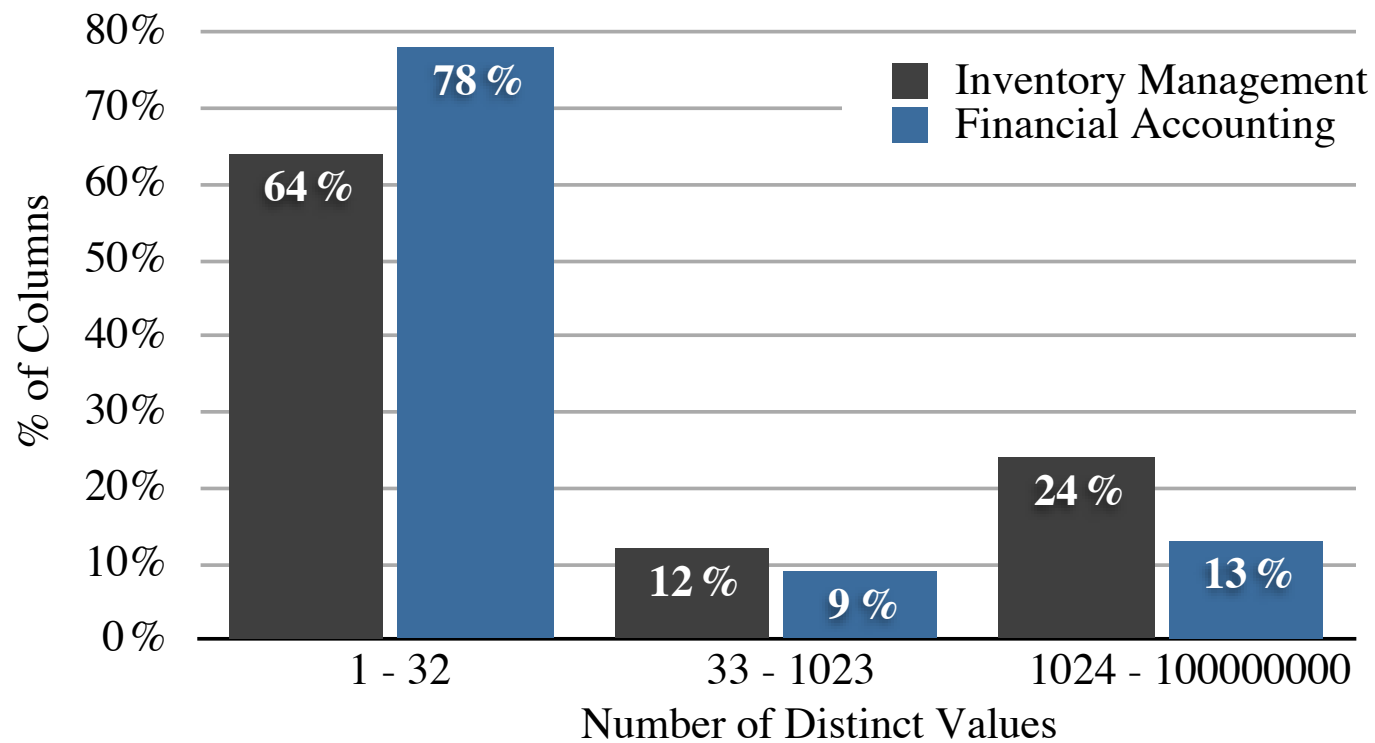
Discrete manufacturing



Many Columns are not Used Even Once

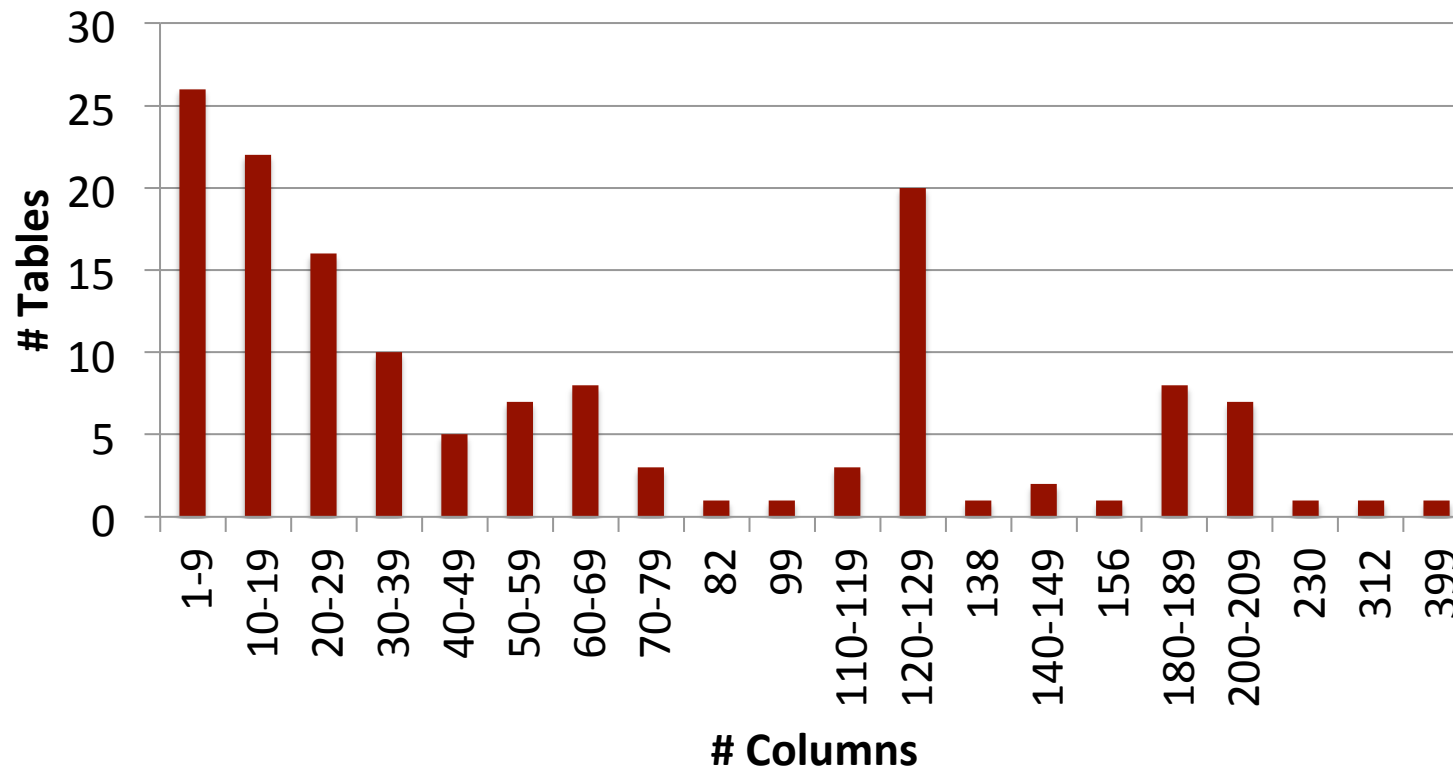
55% unused columns per company in average

40% unused columns across all companies



Wide Tables

Analysis of width of 144 most used* tables



Warum ist das wichtig? I/III

Die Machbarkeit und die Performanz von Geschäftsprozessen werden durch die Performanz und die Machbarkeit ihrer Aktivitäten bestimmt.

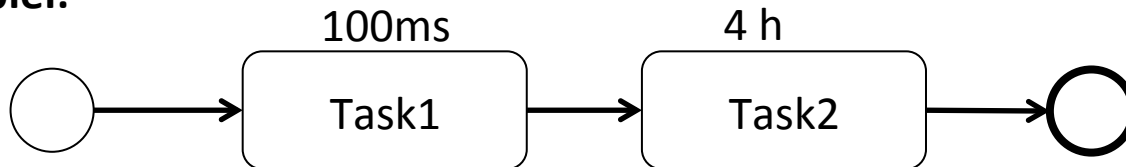
Arten von Aktivitäten:

- **Manuell** – Ausgeführt durch Mitarbeiter oder Vergabe nach Extern
 - „Kunden anrufen“
 - „Kundendaten in System einpflegen“
 - „Durchführung einer Testfahrt vor Auslieferung“
- **Automatisiert** – System führt die Schritte aus
 - „Benachrichtigung aller Kunden per Email“
 - „Verfügbarkeitsprüfung“
 - „Mahnlauf“
 - „Jahresabschluss“

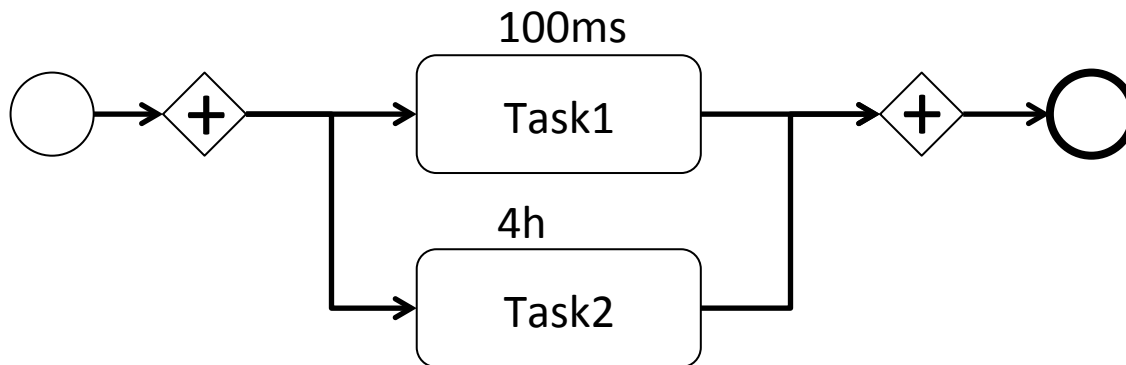
Wichtig: Unter der Annahme, dass aus dem Prozess alle unnötigen Schritte entfernt wurden, kann die Performanz des Prozesses nur durch erhöhte Parallelisierung (Optimierung des Workflows) oder Verbesserung der Performanz der einzelnen Aktivitäten gesteigert werden.

Warum ist das wichtig? II/III

Beispiel:



----- Optimierung



Die Performanz von manuellen Aktivitäten ist meist schwer vorherzusagen. Sie hängt von der Art der Tätigkeit, der benötigten Ressourcen und der Mitarbeiter ab.

Die Performanz automatisierter Aktivitäten hängt von der Menge der zu verarbeitenden Daten und der Art des Algorithmus der verwendet wird ab.

Warum ist das wichtig? III/III

Um die Machbarkeit des einzelnen Prozesses abzuschätzen muss man sich über die Machbarkeit der Aktivitäten und ihrer Zusammenhänge im Klaren sein.

Bei rechenintensiven automatisierten Aktivitäten brauche ich ein **Mengengerüst** um die Laufzeit der Aktivität abschätzen zu können.

Problem:

Entwickler wissen meist nicht genau wie das System beim Kunden aussehen wird und entwickeln mit völlig falschen Annahmen.

Besonderheiten

- Business Logic
- Data Characteristics
- Workload
- Misc, Input/Requirements PIL von Thomas

Challenges for Enterprise Software

- Massive code base
- Handle a lot of data
- Access data concurrently
- Conceptual dissonances within data
 - Customer structure for dept. 1 <> customer structure for dept. 2
- Complex business “illogic”
- Data schema evolution triggered from
 - inside (e.g., change DB)
 - outside (entity evolution, new business functions)