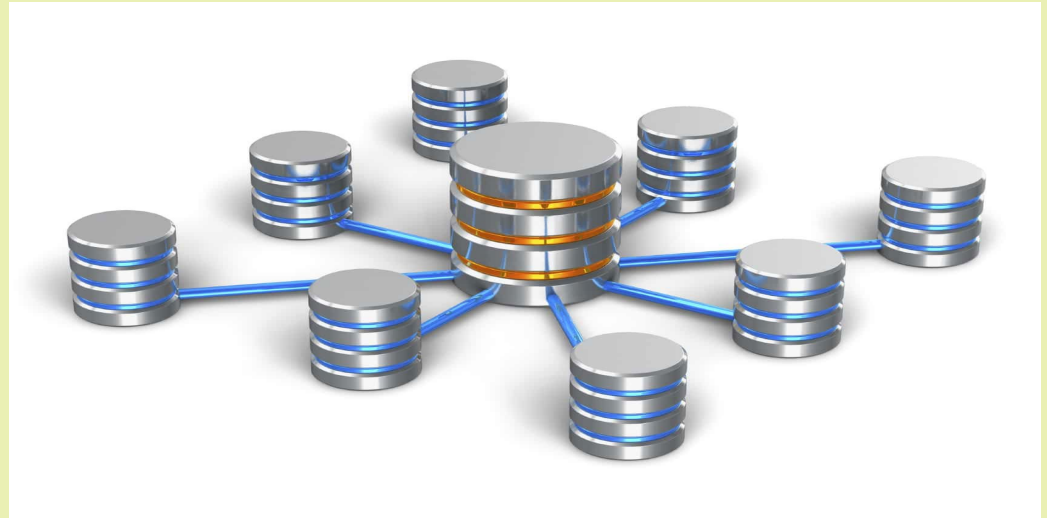


LE1: Einführung Datenintegration

Prof. Dr. Markus Grüne, FB03
Wirtschaftsinformatik



- Den Bedarf für Datenintegration(stechnologien) in Unternehmen verstehen
- Historische Entwicklung der B2B-Integrationstechnologien kennen lernen
 - point-to-point ... cloud-based
- Ziele und Methoden im Kontext der Datenintegration erklären können
- Wesentliche Fachtermini und Definitionen der Kurses kennen lernen

B2B-Integration

- Unternehmen verwenden unterschiedliche Backendsysteme, die untereinander Daten austauschen müssen.
- Die Notwendigkeit für eine B2B-Integrationstechnologie ergibt sich, da die Systeme ursprünglich entworfen wurden, um isoliert zu arbeiten. [Isolation]
- Bei der Integration müssen Datentransformations-Tasks ausgeführt werden, die ein **Mapping** vornehmen, in dem sie
 - Die Darstellung (representation) und
 - Bedeutung (meaning)des einen Systems in die Darstellung und Bedeutung des Zielsystems mappen.
(Bussler 2003)

Hinweis: die Integrationsaufgabe wird schwieriger, wenn auch externe Daten, z.B. Feeds aus der Cloud, angebunden werden sollen.

Wofür brauchen wir Integration?

Der **Technologieeinsatz ändert sich** ständig

- Neue Benutzerfrontends erfordern spezielle Datenrepräsentationen (JSON,...)
- Daten müssen zwischen Systemen transformiert werden (live, batch, realtime, close-to-realtime...)
- Klassische DBS werden durch neue Technologien ersetzt und ergänzt: flat files, XML, JSON, Graphen
- Unterschiedliche Programmiersprachen, Variablentypen und Objekttypen

Moderne Systeme müssen mit Legacy-Systemen (Altsystemen) "kooperieren".

Vielzahl von Architekturen, die ihre Berechtigung haben:

- Relationale Datenbanksysteme
- Unstrukturierte Datenspeicher, z.B. Data Lakes
- Data Streams ...

Business Need for[B2B] Integration (Bussler2003)

"If enterprises have used EDI for over 30 years to connect to their trading partners, why do they still require sophisticated B2B integration technology?

The answer lies in the increasing information technology complexity enterprises face today and will face in the future, where more and more trading partners as well as backend application systems that follow different formats and process standards have to be integrated."

Der Austausch von Daten ist bspw. essenziell, um Logistikprozesse durchzuführen.

Homegrown Integration

- Backendsysteme rufen sich synchron gegenseitig auf, um Daten auszutauschen.
- Zwischenspeicher (Dateien oder Datenbanken) werden verwendet, um Daten asynchron auszutauschen. Der Speicherort wird vorab vereinbart.
- Wenn drei und mehr Systeme involviert sind, muss jedes Paar einen eigenen Speicherort für den Austausch von Daten definieren.
- Letztendlich müssen die Backendsysteme sich selbst um den Datenaustausch kümmern.

Die Probleme aus der "Homegrown Integration" führten zur Entwicklung der ersten Integrationssoftware-Systeme.

(Bussler 2003)

Point-to-Point Integration

- Paarweise Integration der Backendsysteme
- Direkter Datentransfer wird durch Integrationssoftware realisiert
- Verschiedene Arten: synchron / asynchron
- ETL-Software extrahiert und lädt Daten. Im asynchronen Fall werden die Daten nach der Extraktion gespeichert. Die Transformation wird durch die Integrationssoftware vorgenommen.

Einschränkungen

- Neue Anwendungen müssen 2 Datenübertragungsverbindungen aufsetzen (in/out)
- Die Transformationsschritte müssen definiert werden.
- Komplexe Prozesse mit mehr als 2 abhängigen Systemen sind schwer umzusetzen.

(Bussler 2003)

Hubs-and-Spoke Integration

- Zentraler, gemeinsamer Speicher für alle Backendsysteme (hub).
- Die Speichen (spoke) sendet Daten und der Hub verteilt diese an die empfangenden Spokes.
- Spokes können durch "subscribe" Datenlieferungen buchen.
- Der Hub sendet Nachrichten entsprechend der hinterlegten Regeln und analysiert die Header der ankommenden Nachrichten.
- Backendsystem können vergleichsweise einfach hinzugefügt werden.

Einschränkungen

- Keine Geschäftslogik
- Keine Integration über mehrere Stufen
- One-way-Integration: nur Senden, aber kein Warten auf Antwort

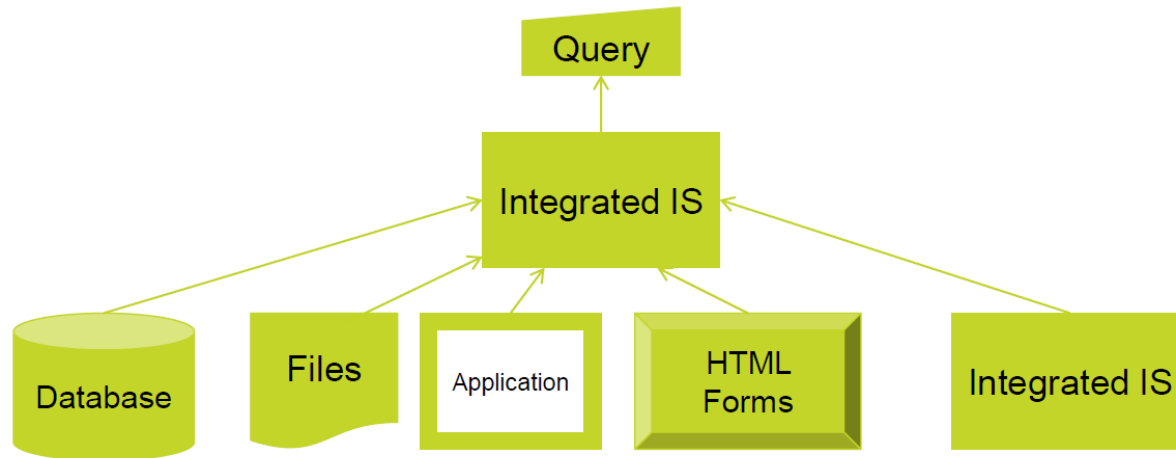
(Bussler 2003)

Integrationstechnologien -Trends

- Integration basierend auf Workflows / Geschäftsprozessen
- Integration zwecks Reporting aus unterschiedlichen Backendsystemen, z.B. DWH
- Integration in Supply-Chain-Netzwerken
- Cloud-basierte Integrationssoftware, die bestehende on-premise Integrationssoftware in die Cloud migriert
- Technologien zur Auszeichnung von Daten und zum Management der Datenherkunft (Data Provenance)
- Einbeziehung von Daten aus dem Web (Twitter, ...)
- Integration unterschiedliche Datenformate (Datenströme, Complex Events, Data at Rest)
- Fehlerbereinigung beim Datenaustausch
- ...

Ziele der Integration

Einheitlicher Zugriff auf Geschäftsdaten, unabhängig von unterschiedlichen Datenformaten in einer Vielzahl von Backendsystemen



(Leser und Naumann 2007)

Daten-Integration: Anwendungen teilen Informationen auf Basis von Daten

Methoden-Integration: Anwendungen teilen Information, indem sie Methoden anderer Anwendungen aufrufen

Prozess-Integration: Prozesse werden lose miteinander verbunden, um Dienste zu konstruieren, die bei Aufruf Daten zurückliefern

Integration per **Frontend**: Einheitlicher Zugriff auf alle Applikationen (z.B. gemeinsames Dashboard)

Herausforderungen der Datenintegration

- Verteilung der Daten
- Heterogenität hinsichtlich
 - Data model (textuell, relational, XML,...)
 - Zugriff (SQL, text retrieval, proprietäre Zugriffsmethoden)
 - Schemas (Attribute, Attributbezeichner, Datentypen)
 - Semantische Unterschiede
- Blurred / fuzzy / unsaubere Daten: unvollständig, widersprüchlich, Duplikate
- Integrierte Informationssysteme (IIS) müssen ein einheitliches Datenmodell definieren und die Übersetzung der Quelldaten in dieses Modell übernehmen.

Datenquelle

- Daten, die integriert werden sollen, stammen aus einem Quellsystem.

Integriertes Informationssystem (IIS)

- Informationssystem, das den Zugriff auf die integrierten Quellen ermöglicht
- Implementiert Regeln für den Zugriff und die Konvertierung der Daten

Transparenter Zugriff

- Users benötigen kein Wissen über interne Datenstrukturen
- Sie müssen nicht wissen, welche Datenquellen integriert wurden und aus welcher Quelle die Daten ursprünglich stammen.

Metadaten

Daten über Daten / Daten, die Daten beschreiben

Ein IIS speichert Metadaten über

- sich selbst
- die Systeme, die es integriert
- Transformationsregeln, um syntaktische und semantische Transformationen in ein Ziel-Datenmodell vornehmen zu können

Szenarien – Wann benötige ich Datenintegration?

Unternehmensfusionen (Mergers and Acquisitions)

- Unternehmen übernimmt ein anderes
- Ähnliche Geschäftsbereiche
- Z.B. Integration von anderen Banken in die Deutsche Bank

Kundenbeziehungsmanagement (CRM)

- Kundinnen und Kunden kontaktieren ein Unternehmen über unterschiedliche Kanäle:
 - Zweigstelle, Telefon, Web, ...
 - Oft haben Einheiten des Unternehmens eigene Anwendungen zur Verwaltung von Kundendaten

Verteilte Datenbanken und föderierte Datenschemata

Data Warehouses

Unternehmensdatenmodelle

Anwendungsintegration (Application Integration)

- Enterprise Application Integration (EAI)
- Service-oriented Architectures (SOA)

Key Takeaways

- Es gibt viele Arten von Systemen, die untereinander Daten austauschen.
- Die Datenmodelle unterscheiden sich zwischen den Systemen teils erheblich.
- Es gibt eine Vielzahl von Methoden zur Integration von Daten.
- IIS

Think! (Diskussion)

- Kennen Sie ein Beispiel eines Integrierten Informationssystems?
- Welche Daten werden dort integriert und woher stammen diese? Intern, extern, andere Applikationen ...
- Wie werden die Daten integriert (falls Sie das wissen)?

Bussler C (2003) B2B integration; Concepts and architecture ; with 4 tables. Springer, Berlin

Leser U, Naumann F (2007) Informationsintegration; Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen. dpunkt-Verl., Heidelberg