
4. WORKSHOP - OSBI

Big Data und Datenvirtualisierung

Dr. Sebastian Streit & Maxim Zehe



F. Hoffmann-La Roche AG



- Gegründet 1896 in Basel
- Über 80.000 Mitarbeitende
- Führende Position in Pharma
- Fokussierung auf den Healthcare-Bereich
- Aktiv in 150 Ländern auf allen Kontinenten
- Weltweit größtes Biotech-Unternehmen mit klinisch differenzierten Medikamenten für die Onkologie, Virologie, Entzündungs- und Stoffwechselkrankheiten und Erkrankungen des Zentralnervensystems

Systemec & Services GmbH

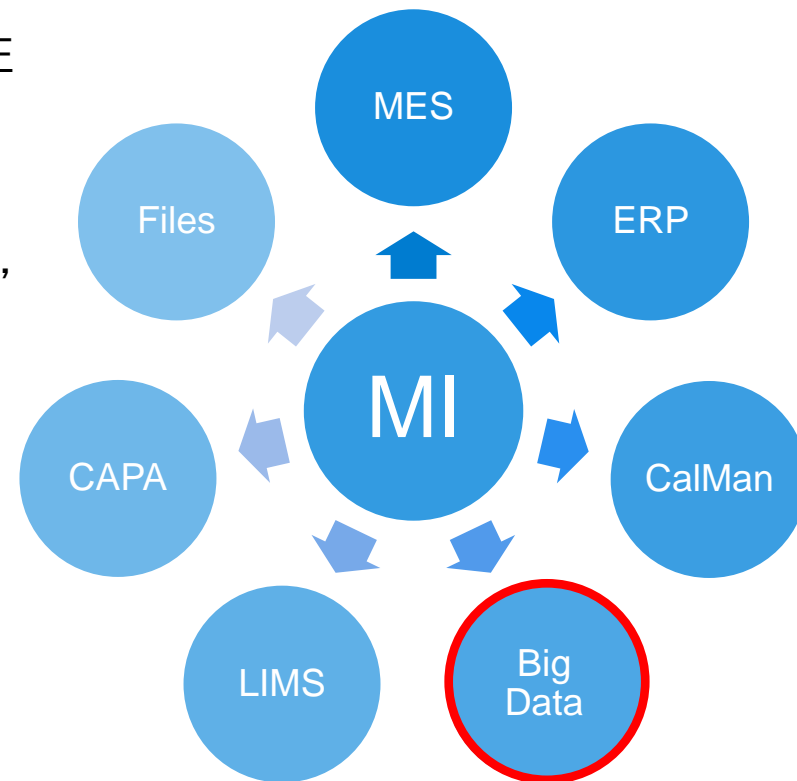


- Hauptsitz
 - Karlsruhe – Deutschland
- Tochtergesellschaft
 - Cary, North Carolina – USA
- Service Center
 - Bad Säckingen – Deutschland
- Mitarbeiter
 - mehr als 55 im In- und Ausland
- Zielmärkte
 - International · Life Sciences · CPG · Kosmetik
- Geschäftsbereiche
 - Services Software · Solutions · Hardware Concepts



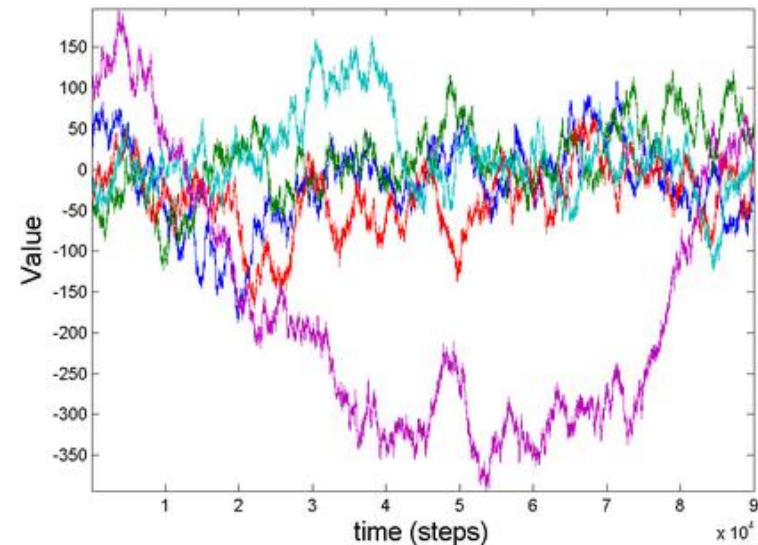
Vision - Manufacturing Intelligence

- Daten-Zusammenstellung aus allen relevanten Systemen
- Datenstrukturierung zur einfachen Interpretierbarkeit
- Data Analyse, z.B. KPI Berechnung wie OEE
- Daten Visualisierung, z.B. in Dashboards
- Automatische vertikale Daten Übertragung, z.B. von Shop Floor Systemen zu SAP



Was sind „Big Data“ im Produktionsumfeld?

- Zeitbasierte Datenreihen, die z.B. Sensordaten oder geloggte Ereignisse abbilden
- Einfache Datenmodelle, meist im Format „Name, Zeitstempel und Wert“ aber nicht relational, relationale Information ist in höhere Schichten verlagert
- Mittlere Datenvolumina im Bereich bis einige Hundert Gigabyte pro Jahr
- Datenaufkommen per Sensor kann im Bereich von mehreren Hundert pro Sekunde liegen
- Starker Echtzeitcharakter in Speicherung und Analyse



Big Data Lösungen und Probleme

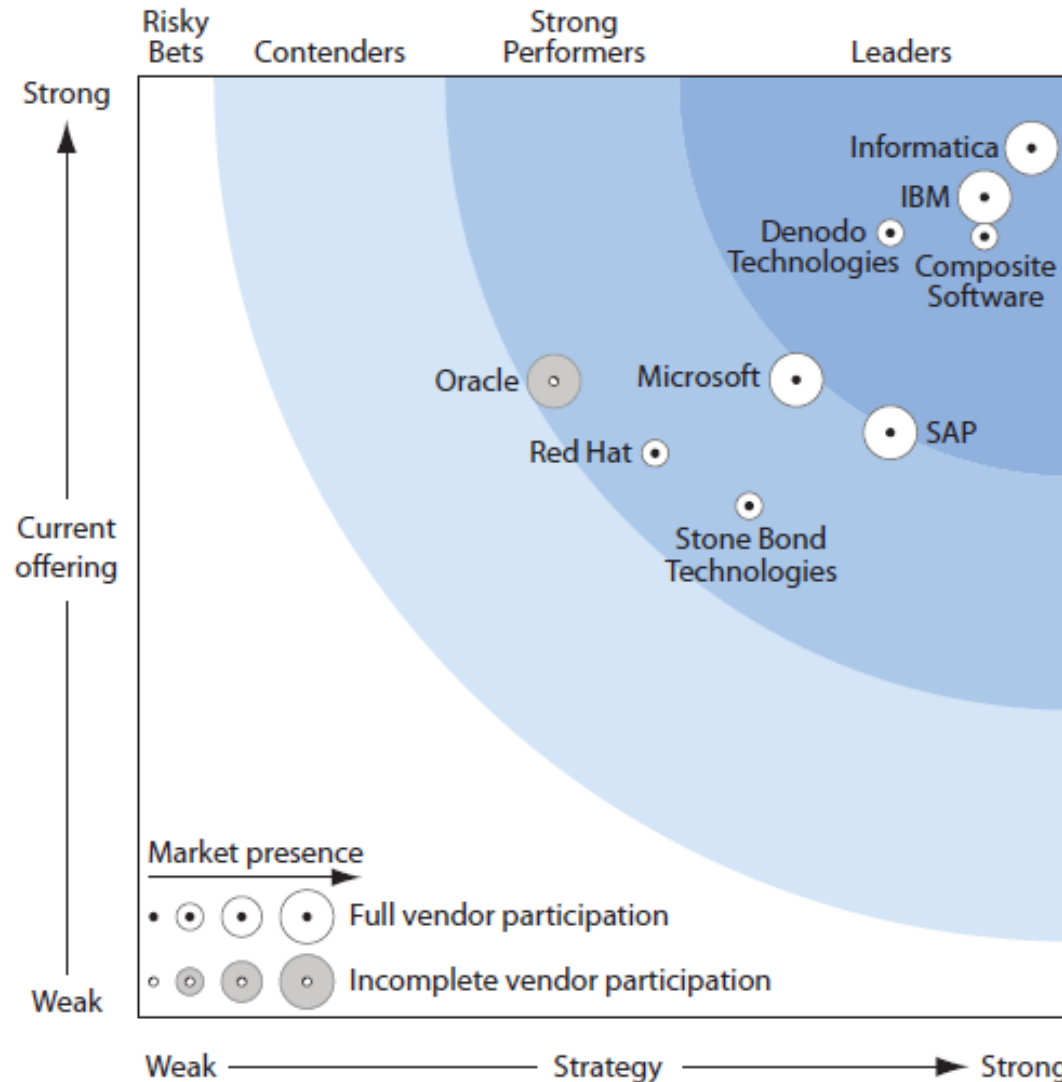
- Kommerzielle Data Historians, z.B. Osisoft PI:
 - Sehr viele Schnittstellen zu Produktionssystemen, z.B. OPC (OLE for Process Control)
 - Sehr aufwendige Kompressionsmöglichkeiten
- Open Source Tools, z.B. OpenTSDB:
 - Wenige Schnittstellen
 - Extrem skalierbar, basiert auf Big Data Lösungen (z.B. HBase)
- Problem der Weiterverarbeitung:
 - Zeitreihen mit vielen Millionen Datenreihen sind nicht gut in relationalen Systemen zu verarbeiten
 - Jedes verarbeitende System muss Echtzeitcharakter der Daten reflektieren

Enterprise Information Integration (EII)



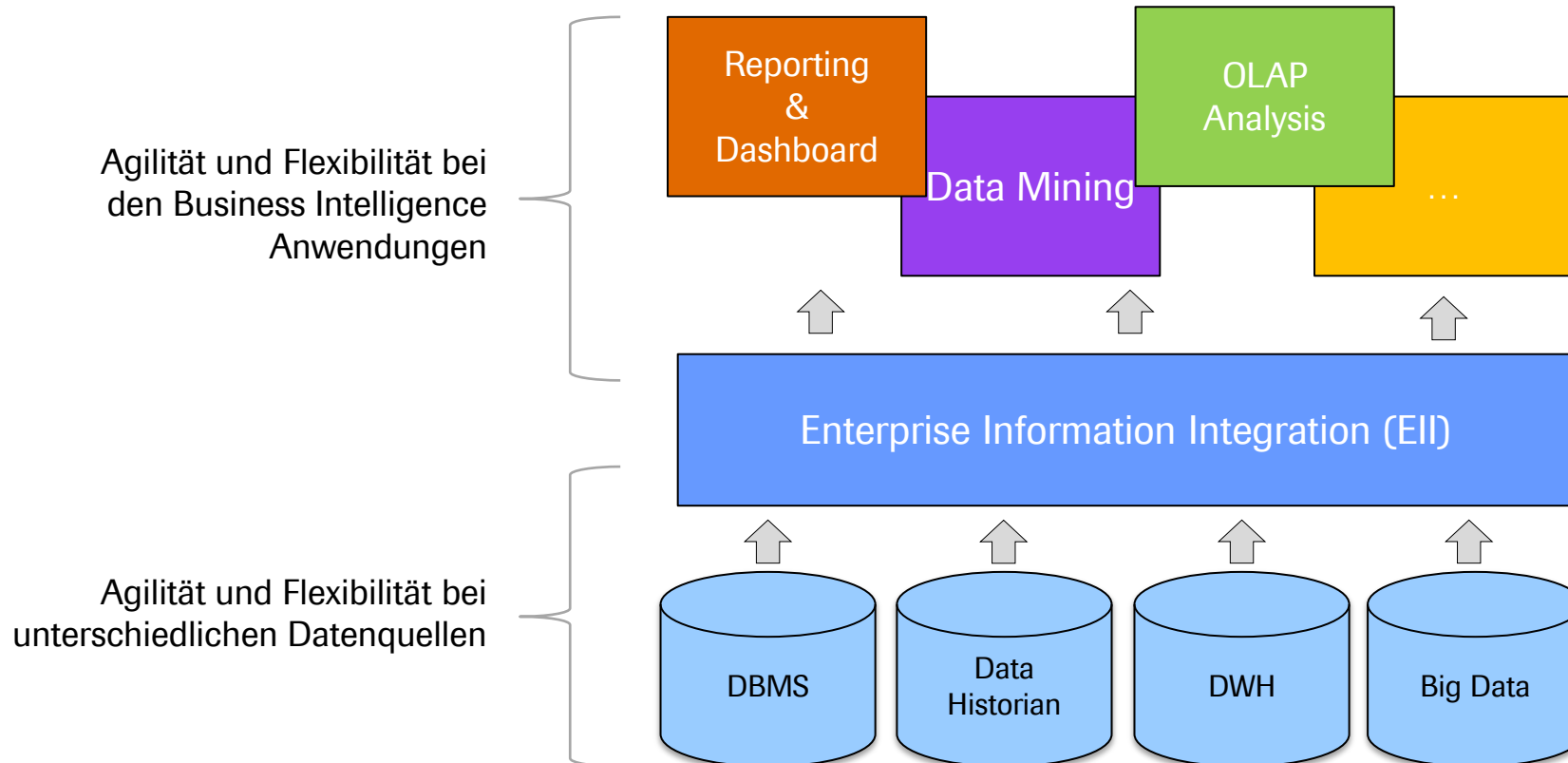
- EII kann den etablierten ETL-Prozess durch *Data Virtualization*, also Integration der Informationen mittels Datenabstraktion, ersetzen
 - Alternative zum traditionellen Data Warehouse Ansatz
- *Data Virtualization* ist unter dem Begriff *Information as a Service* (IaaS) bekannt
 - Ziel: „A single Version of Truth“
 - Annäherung an Echtzeitanalyse
- *Abstraktionsschichten* ermöglichen eine einfache und flexible Datenstruktur für die Integration von Daten aus heterogenen Datenquellen
- Forrester prognostiziert einen Anstieg der Umsätze im EII Bereich auf *8 Milliarden* US-Dollar im Jahr 2014

The Forrester Wave 2012



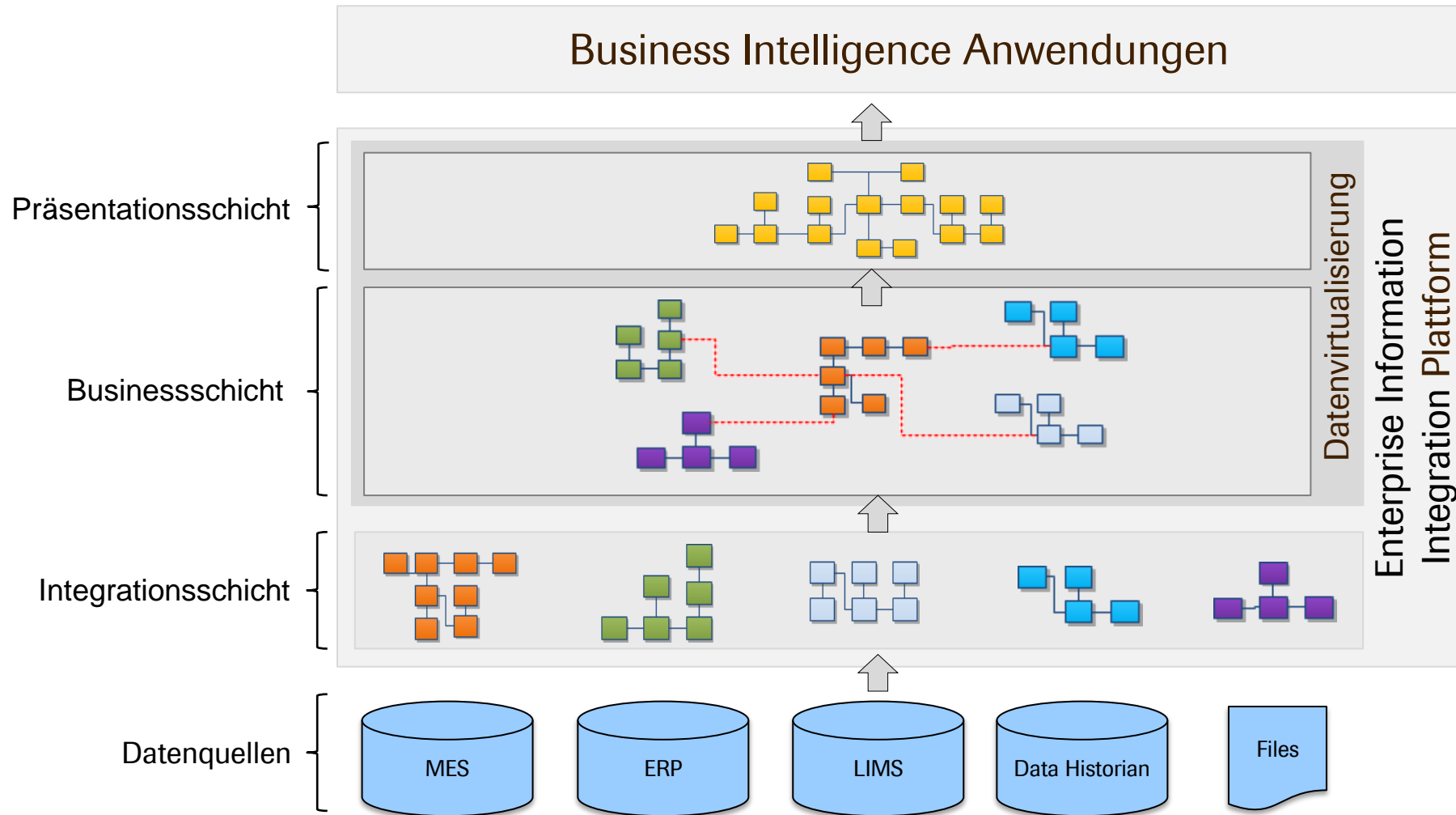
Quelle: Forrester Wave 2012a Yuhanna, Noel; Gilpin, Mike: The Forrester Wave: Data Virtualization, Q1 2012.

Agilität und Flexibilität von EII



Quelle: Lans 2012 Van der Lans, Rick F.: Data Virtualization for Business Intelligence Agility 2012

Abstraktionsschichten



■ Benefits

- Schnelle und einfache Datenintegration neuer oder temporärer Datenquellen
- Ergebnisse mit Annäherung an Real-Time-Abfragen durch gleichzeitigen Zugriff auf operative Systeme
- Einheitliche Ergebnisse bei Reporting- und Analysetools
- Flexibilität gegenüber Änderungen des Datenmodells
- Kostenersparnis bei Hardware, Lizenzen und Wartung

■ Caveat

- Performance kann nur begrenzt durch Caching-Mechanismen erreicht werden
- Tatsächliche Performance bestimmt die Quellsysteme (Flaschenhalseffekt)
- Komplexe Abfragen für Analysen können die Performance verringern
- Daten werden zur Ausführungszeit eines Reports konsolidiert oder verdichtet

- Aktuelle Situation: Einsatz in validierten und nicht validierten Umgebungen
- EII-Lösungen erweitern existierende Funktionalitäten (ETL)
- Open Source und proprietäre Lösungen sind auf dem Markt verfügbar
- Governance und Support Modelle erschaffen neue Herausforderungen, da die Flexibilität auch schnell zu Fehlkonfigurationen führen kann
- Einsatz von Spiegelservern muss genau evaluiert werden, um Produktionssysteme zu entlasten (PI2PI, Oracle GoldenGate)
- Der Einsatz von EII ermöglicht eine flexible Integration von verschiedenen Technologien und Ansätzen

Kontakt



- Dr. Sebastian Streit
Expert Solution Architect

F. Hoffmann-La Roche AG
Dept. FPPE - Bldg. 239 / 1.13
CH-4070 Basel

Phone: +41 61 68 79850
Fax: +41 61 68 87635
sebastian.streit@roche.com

- Maxim Zehe
Technical Consultant

Systemec & Services GmbH
Emmy-Noether-Str. 17
76131 Karlsruhe

Phone: +49 721 6634 310
Fax: +49 721 6634 333
maz@systemec-services.com