

Streaming Systems

Wintersemester 2022/23

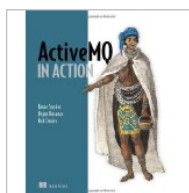
Informatik Master (INM)

Prof. Dr. Bernhard Hollunder

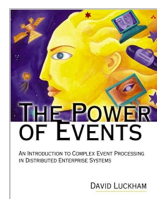
Department of Computer Science
Furtwangen University of Applied Sciences
Robert-Gerwig-Platz 1, D-78120 Furtwangen
<http://www.hs-furtwangen.de/>

Streaming Systems / Event Processing

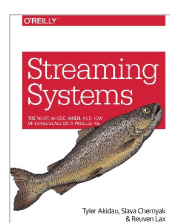
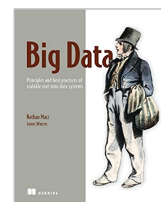
Messaging /
Message Queuing



Complex Event
Processing (CEP)



Classic Big Data



Streaming Systems

A type of data processing engine
that is designed with unbounded
data sets in mind.

Stream Processing

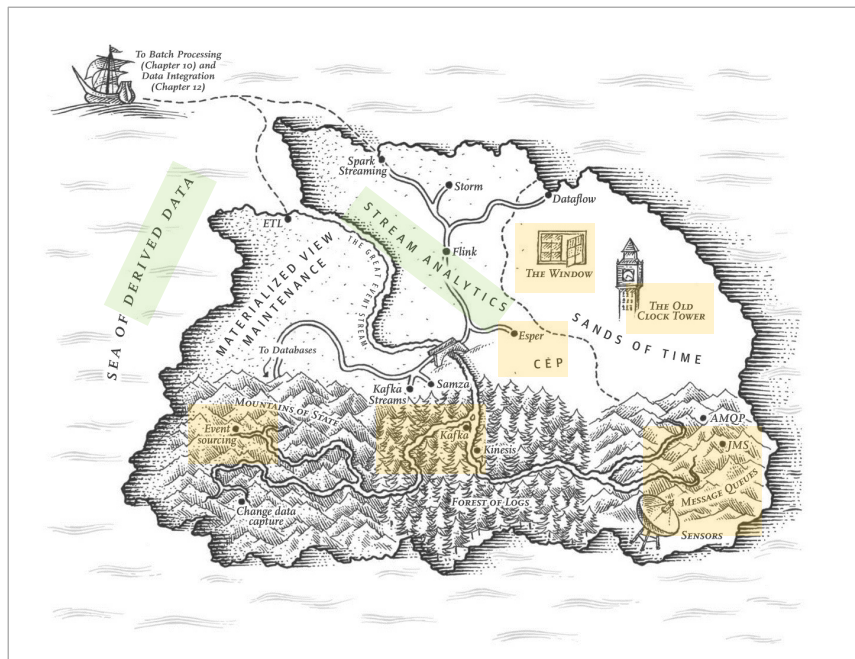


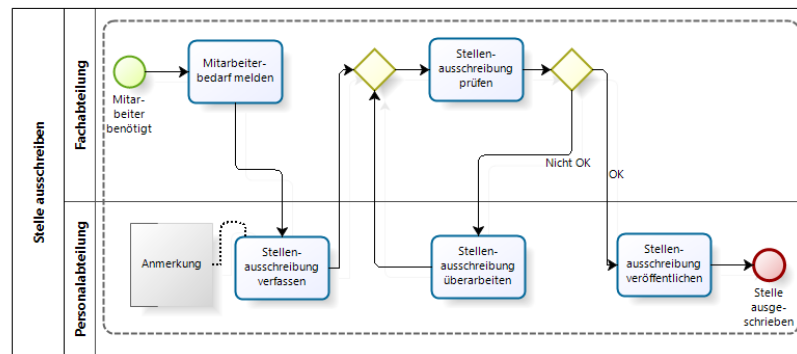
Figure taken from Martin Kleppmann: Designing Data-Intensive Applications. Chapter 11 Stream Processing. O'Reilly. 2017.

Themen

1. Event-driven Computing
2. Event Sourcing and its Application to CQRS
3. Java Messaging Service
4. Messaging and Streaming Platform Apache Kafka
5. Streaming Systems
6. Unified Model for Both Batch and Streaming Processing
7. Complex Event Processing and Event Processing Language
8. Processing Semantics: Read-Process-Write Pattern

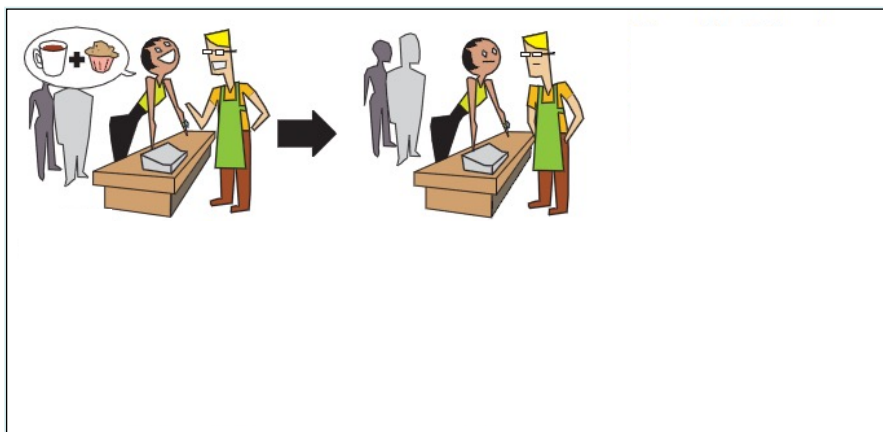
Geschäftsprozess

- Geschäftsprozess (Business Process)
 - Folge von Schritten zur Erreichung eines angestrebten Arbeitsergebnisses.
 - Dient der Leistungserbringung in einem Unternehmen.
- Beispiel: Stellenausschreibung eines Unternehmens:



Ereignisgesteuerte Geschäftsprozesse

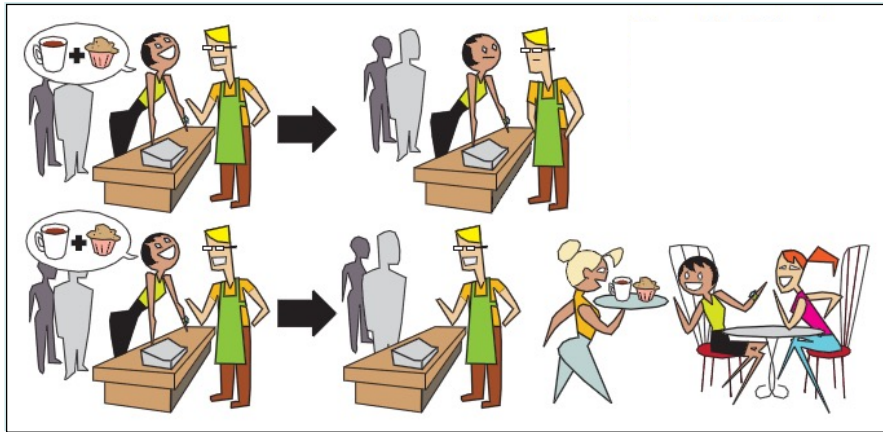
- Ereignisgesteuerter Geschäftsprozess
 - Die spezifische Verarbeitungsfolge wird durch das Auftreten von Ereignissen ausgelöst oder substantiell beeinflusst.
- Synchroner vs. ereignisgesteuerter Ablauf:



Aus: Etzion, Niblett: Event Processing in Action, Manning, 2011.

Ereignisgesteuerte Geschäftsprozesse

- Ereignisgesteuerter Geschäftsprozess
 - Die spezifische Verarbeitungsfolge wird durch das Auftreten von Ereignissen ausgelöst oder substantiell beeinflusst.
- Synchroner vs. ereignisgesteuerter Ablauf:



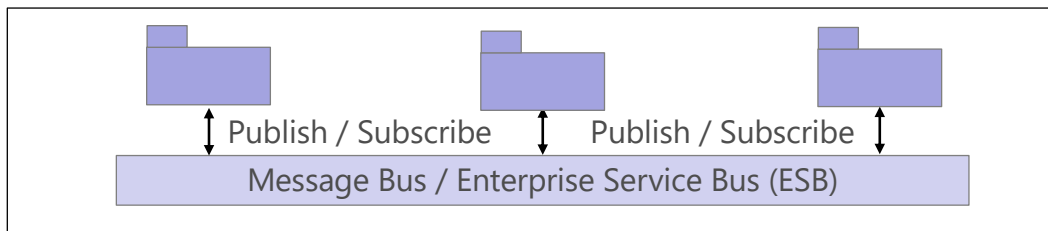
Aus: Etzion, Niblett: Event Processing in Action, Manning, 2011.

Terminologie

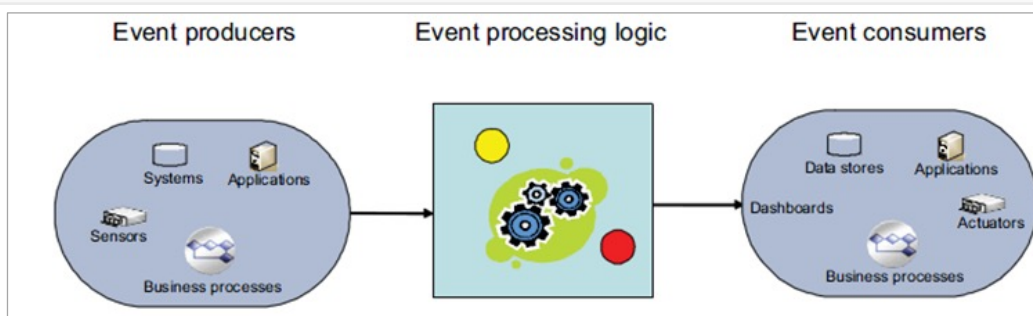
- Ereignissteuerung
 - Als ereignisgesteuert wird das Verhalten eines Systems oder Softwaremoduls bezeichnet, dessen Verarbeitungsprozess durch das Eintreffen von Ereignissen von internen oder externen Quellen ausgelöst wird.
[Bruns und Dunkel, Event-Driven Architecture, Springer, 2010]
- Event processing
 - Event Processing is computing that performs operations on events. Common event processing operations include reading, creating, transforming, and deleting events.
[Etzion, Niblett: Event Processing in Action, Manning, 2011]
- Streaming system
 - A type of data processing engine that is designed with infinite datasets in mind.
[Akidau, Chernyak, Lax: Streaming Systems, O'Reilly, 2018]

Ereignisgesteuerte Kommunikation

- Besteht aus unabhängigen Komponenten, die über den asynchronen Austausch von Nachrichten miteinander interagieren.
- Komponenten
 - produzieren Nachrichten und senden diese an alle oder an ausgewählte Komponenten
 - können sich für bestimmte Nachrichtentypen registrieren und werden automatisch informiert, wenn entsprechende Nachrichten versendet wurden.
- „Message Bus“ als technische Infrastruktur, um Ereignisse (Events) zwischen den Komponenten auszutauschen.



Ereignisgesteuerte Systeme – Einfache Sicht



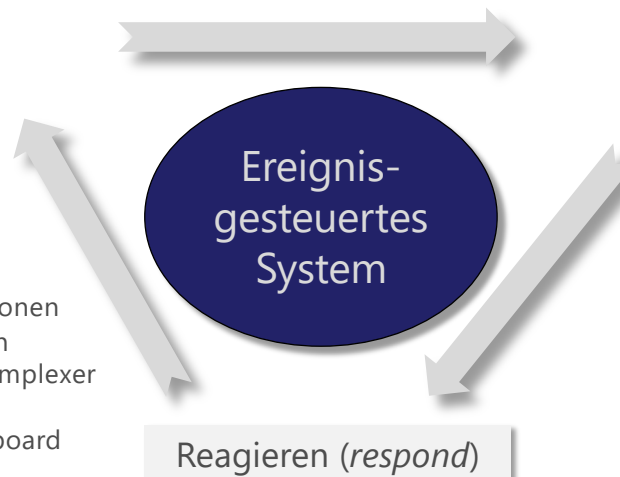
- Ereignisquellen:
 - Sensoren, z.B. Umweltdaten
 - RFID-Lesegeräte
 - Datenticker
 - Positionsdaten
 - Nutzerinteraktionen, z.B. Klicks
 - Kontrollmeldungen von Anwendung
 - Log-Dateien
 - ...
- Ereignissenken und typische Reaktionen zur Ereignisbehandlung:
 - Aufruf von Diensten und Auslösen von Prozessen
 - Aktualisierung einer Datenbank
 - Publizieren einer Nachricht und Weiterverarbeitung von anderen Komponenten
 - Anzeige auf einem Dashboard
 - ...

Charakteristik ereignisgesteuerter Systeme

Erfassen von technischen / fachlichen Ereignissen aus Ereignisquellen (*sense*)

Analyse von Ereignissen, Zusammenhänge explizit machen (*analyse*)

- Initiierung von Aktionen
- Aufruf von Diensten
- Erzeugen neuer, komplexer Ereignisse
- Monitoring / Dashboard



Zentral: Ereignisse / Events

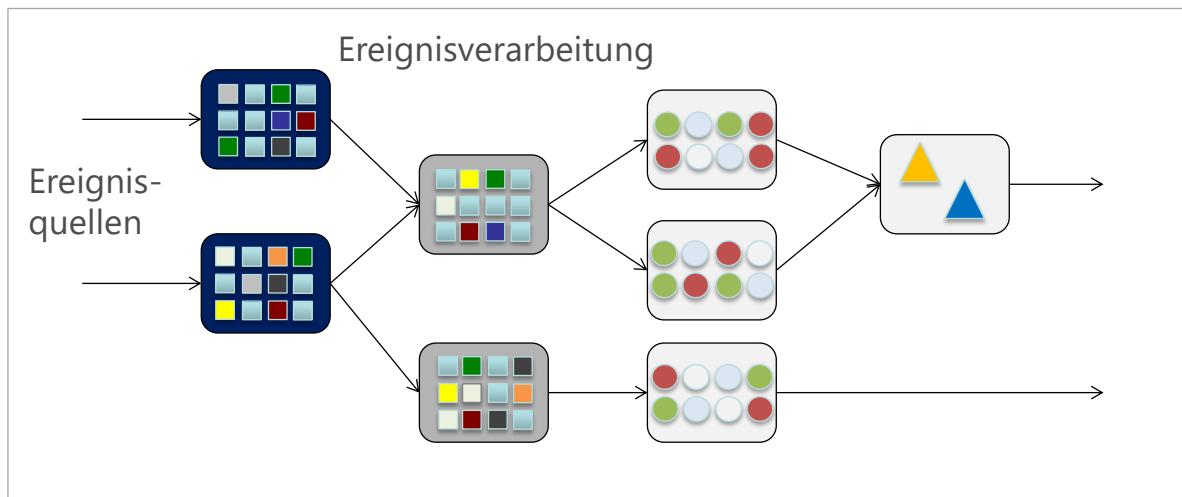
- Ereignis (*Event*, je nach Kontext / Framework auch *Message*, *Record*, *Tuple*).
 - Ein Ereignis kann alles sein, was aufgetreten ist. Z.B.: Start oder Beendigung einer Aktion, Positionsänderung eines Objektes, Mausklick.
 - Im Allgemeinen bezieht sich ein Ereignis auf die Veränderung eines Zustands, typischerweise auf die Änderung des Wertes einer Eigenschaft eines realen oder virtuellen Objektes.
 - Das Ausbleiben einer Änderung kann auch ein relevantes Ereignis sein!
- Ereignisse können Vorkommnisse auf unterschiedlichen Abstraktions-ebenen repräsentieren wie etwa
 - Technische Ereignisse: gemessene Temperatur, CO₂-Gehalt der Luft, Scannen eines Bar/QR-Codes, Geodaten eines Objektes, ...
 - Systemereignisse: Eingang eines HTTP-Requests, Eintrag in einer Log-Datei, fehlgeschlagene Authentifizierung, Beendigung einer Berechnung, ...
 - Geschäftsereignisse: Storno einer Bestellung, Unterschreiten des Lagerbestands einer bestimmten Ware, Übergabe einer Warensendung, ...

Ereignisverarbeitung

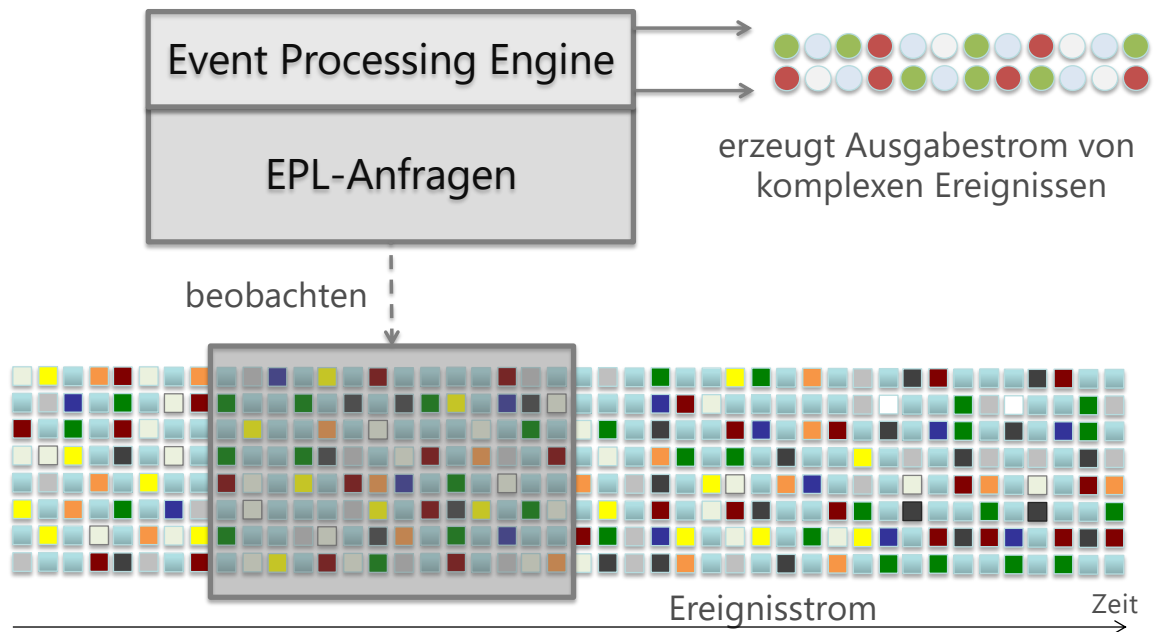
- Eingabe
 - Kontinuierliche Ströme von Ereignissen, oft aus mehreren Quellen.
- Ausgabe
 - Daten mit fachlichem Mehrwert, aufbereitete Daten, aggregierte Daten, „komplexe“ Ereignisse.
- Ansätze und jeweilige Schwerpunkte
 - Streaming Systems, Event Stream Processing (ESP), Event Processing Networks (EPNs), ...
 - Mechanismen zur verteilten, parallelen Verarbeitung.
 - Skalierbarkeit und Echtzeit-Verarbeitung sehr großer Datenmengen.
 - Imperative Umsetzung der Verarbeitungslogik.
 - Complex Event Processing (CEP)
 - Regeln, die kontinuierlich von einem Interpreter auf die eingehenden Ereignisströme angewendet werden.
 - Deklarative Beschreibung der Verarbeitungslogik, z.B. mit CQL und EPL.

Streaming Systems / Event Stream Processing

- Topologie von Komponenten, die Ereignisse entgegennehmen, verarbeiten und neue Ereignisse erzeugen.



Regelbasierte Ereignisverarbeitung



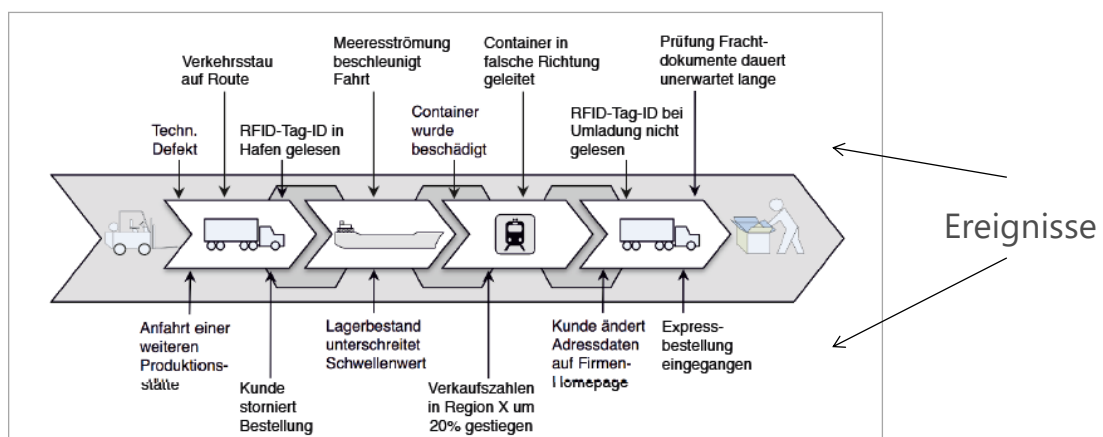
Anwendungsszenarien

Ereignisverarbeitung: Typische Einsatzgebiete

- Logistik
- Maut-Systeme
- Sport und Freizeit, Heat Maps
- Finanzmärkte
- Trends und Trendanalyse in Soziale Netze
- Empfehlungssysteme
- Gesundheitsbereich, Überwachung von Vitalparametern
- Business Activity Monitoring (BAM)
- Verkehrskontrolle und -steuerung
- IoT-Infrastrukturen wie etwa Smart Home
- ...

Ereignisgesteuerte Anwendungen (1)

- Logistik
 - Versand von Waren und Auslieferung beim Empfänger.
 - Nutzung unterschiedlicher Verkehrsträger.
 - Netzwerk verschiedener Dienstleister, Institutionen und Organisationen. (Spediteure, Fertigungsunternehmen, Zollbehörde, Wetterdaten, ...).



[Bruns und Dunkel, Event-Driven Architecture, Springer, 2010]

Ereignisgesteuerte Anwendungen (2)

■ Maut-Systeme

- Nutzung von gebührenpflichtigen Strecken.
- Auswertung und Kontrolle der erfassten Daten (z.B. Abrechnungs- und Mehrwertdienste).
- Zusammenspiel unterschiedlicher Teilnehmer (Toll-Collect Zentrale, Verkehrsministerium, Spediteure, ...).
- Mautkontrollbrücken und mobile Kontrolleure.

■ Ereignisse

- Erfassen von Nummernschildern von Fahrzeugen, die bestimmte Stellen passieren.
- Kommunikation der Daten.
- Anmeldung von Fahrten.
- ...



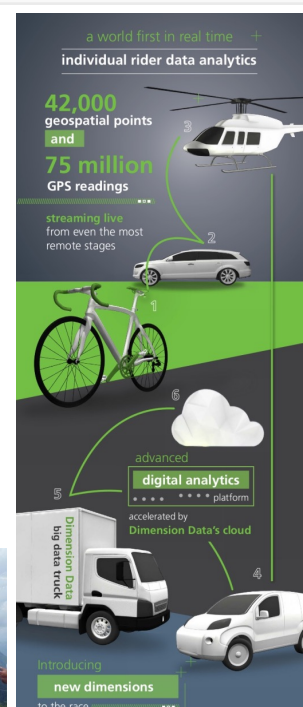
Ereignisgesteuerte Anwendungen (3)

■ Sport und Freizeit

- Erfassung von Lauf-, Wander- und Radstrecken.
- Visualisierung auf Landkarten.
- Kombination mit Zeitstempeln.
- Zeitnahe Auswertung und Erstellung von „spannenden“ Statistiken.
- Darstellung mit „Heat Maps“.

■ Ereignisse

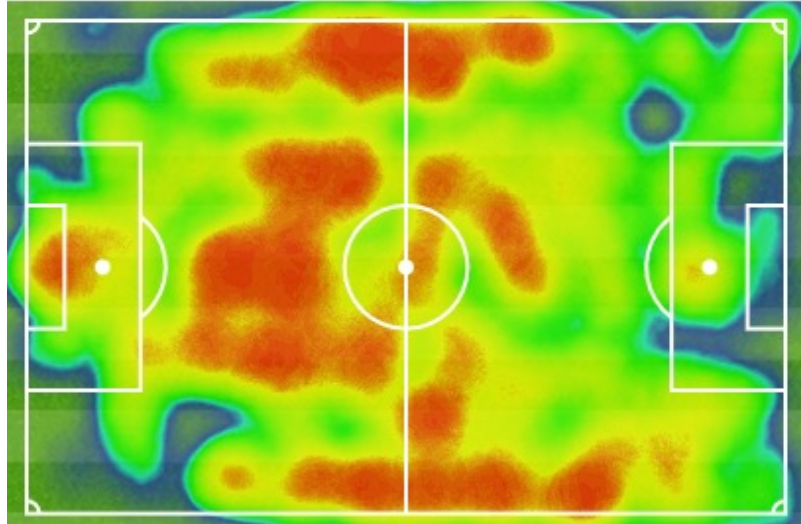
- Erfassung von GPS-Daten.
- Erreichen bestimmter Geopunkte.
- Signifikante Änderungen von Geschwindigkeiten (z.B. auf Grund eines Sturzes).
- ...



Ereignisgesteuerte Anwendungen (4)

■ Heat Maps

- Visualisierung von dynamischen Strukturen und Abläufen.
- Zeitnahe Aktualisierung.



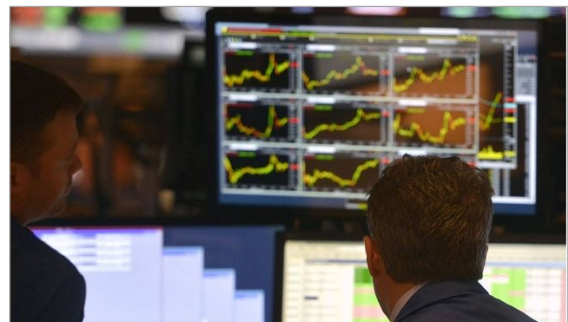
Ereignisgesteuerte Anwendungen (5)

■ Finanzmärkte / Wertpapierhandel

- Kauf- und Verkaufsentscheidungen orientieren sich an der aktuellen Entwicklung unterschiedlicher Kennzahlen wie Devisenkurse, Rohstoffpreise, Börsenkurse und Aktienindizes.
- Beobachtung der Kursentwicklung, um Kauf- und Verkaufsentscheidungen in Bruchteilen von Sekunden treffen zu können.
- Analyse des Zahlungsverkehrs, um automatisch Betrugsversuche (*fraud detection*) zu entdecken.
- Geldwäsche und auffällige Finanztransaktionen.

■ Ereignisse

- Kursänderungen.
- Buchungen.
- Aufträge.
- Kauf- und Verkauf.
- ...



Ereignisgesteuerte Anwendungen (6)

■ Soziale Netze

- Virtuelle Nähe zu interessierten Personen und Institutionen.
- Darstellung von Text- und Multimediainhalten.
- Kommunikation.
- Analyse von Inhalten und Klickverhalten.
- Profilbezogene Werbung.



■ Ereignisse

- Online/offline-Meldungen von Teilnehmern.
- Publikation neuer Inhalte.
- Schalten von Werbung.
- Anzeige bzw. Lesen von Informationen.
- Kommentare abgeben.
- ...



Ereignisgesteuerte Anwendungen (7)

■ Monitoring von Vitalparametern

- Diagnosesystemen im Medizinsektor.
- Überwachung des Gesundheitszustands, insbesondere bei Operationen und auf der Intensivstation.
- Sensoren als technisches Bindeglied zwischen der physikalischen Welt und der Informationsverarbeitung.
- Abgleich mit kritischen Werten (Blutdruck, Puls, Atmung, ...).
- Veränderung von Werten.

■ Ereignisse

- Pulsschlag, Herzfrequenz und Atemfrequenz.
- Blutdruckmessung.
- Körpertemperatur.
- Messung des Körpergewichts.
- Medikamenteneinnahme.
- ...



Ereignisgesteuerte Anwendungen (8)

- Business Activity Monitoring (BAM)
 - Geschäftsdaten werden in Echtzeit verdichtet und zu Unternehmensindikatoren aufbereitet.
 - Warenlieferungen und -ausgänge werden ohne Unterbrechung überwacht, um den Lagerbestand zu optimieren
 - Nutzung von ortsbezogenen Smartphone-Daten, um Kundenwünsche frühzeitig zu verstehen oder aufkommende Trends im Vorfeld zu erkennen.
 - Steuerungsinformationen im Handels- und Dienstleistungsbereich.
- Ereignisse
 - Auftragseingang.
 - Zahlungseingang.
 - Fehlender Zahlungseingang.
 - Bestell- und Lieferabwicklung.
 - Änderung des Lagerbestands.
 - ...

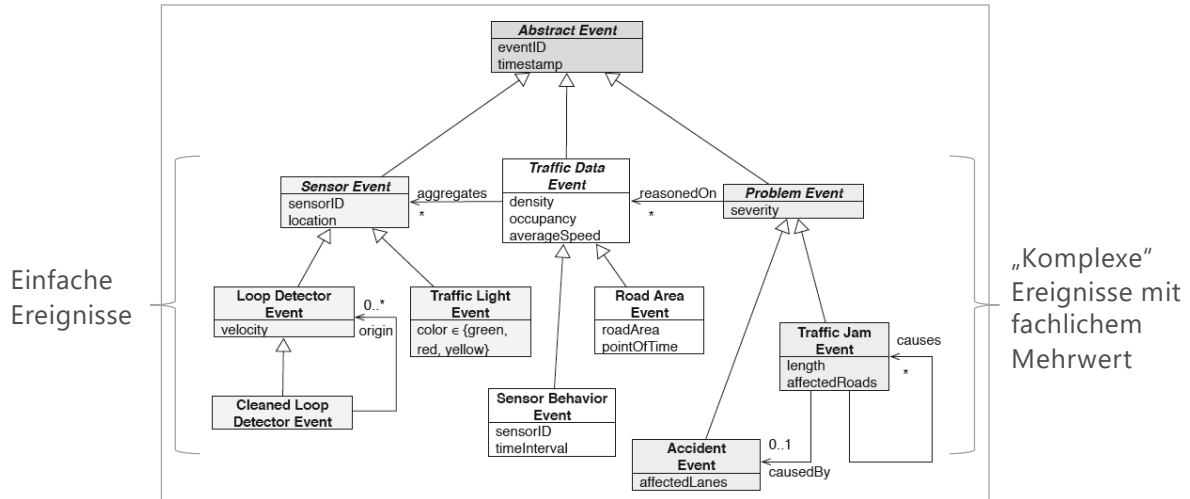


Ereignisgesteuerte Anwendungen (9)

- Verkehrskontrolle und -steuerung:
- Ziele
 - Intelligente Steuerung der Verkehrsflüsse.
 - Erhöhung der Effizienz der bestehenden Verkehrsinfrastruktur.
 - Vermeidung von Verkehrsbehinderungen, insbesondere Staus.
 - Erhöhung der Verkehrssicherheit.
 - Senkung von verkehrsbedingten Umweltbelastungen.
- Grundlage
 - Zeitaktuelle Informationen über die Verkehrswege (Baustellen, Straßensperrungen, Behinderungen, Gefährdungen, ...).
 - Verkehrsdaten (Verkehrsdichte, Durchschnittsgeschwindigkeit, ...).
 - Umweltdaten (Temperatur, Niederschlag, Nebel, ...).

Ereignisgesteuerte Anwendungen (10)

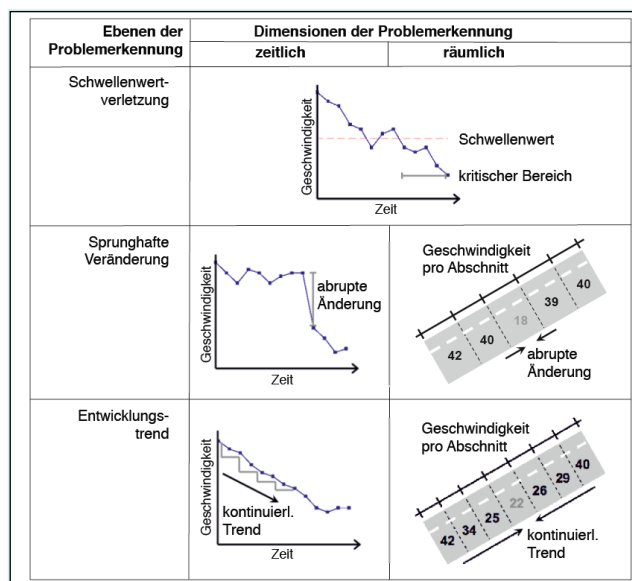
- Verkehrskontrolle und -steuerung: mögliches Ereignismodell, d.h. Spezifikation der Typen der zu verarbeitenden Ereignisse:



[Bruns und Dunkel, Event-Driven Architecture, Springer, 2010]

Ereignisgesteuerte Anwendungen (11)

- Analyse und Visualisierung einer Verkehrssituation:



[Bruns und Dunkel, Event-Driven Architecture, Springer, 2010]

Ereignisverarbeitung: Gründe und Herausforderungen

Gründe für den Einsatz der Ereignisverarbeitung

- Anwendungen, die über Sensoren und andere Quellen kontinuierlich Daten erhalten, diese analysieren, weiterverarbeiten und darauf reagieren.
- Anwendungen müssen zeitnah bestimmte Situationen erkennen können.
- Existierende Anwendungen sollen auf minimal invasive Weise (z.B. Auswerten von Log-Dateien) erweitert werden, um weitere fachliche Anforderungen umzusetzen.
- Logik zur Verarbeitung der Ereignisse kann von anderen Systemkomponenten weitestgehend isoliert werden.
- Anwendungen müssen in großen, sich verändernden Datenmengen nach bestimmten Mustern suchen.
- Große Technologie- und Werkzeugunterstützung, insbesondere im Open Source Bereich.

Datenströme in vernetzten Systemen

- „Klassische“ Big Data Anwendung werten vor allem *statische* Datenbestände aus.
- Im Gegensatz dazu werden beim Event Processing Ströme (*Streams*) kontinuierlich eintreffender Daten berücksichtigt.
- Datenströme weisen spezifische Eigenschaften auf, die bei ihrer Verarbeitung zu berücksichtigen sind:
 - Die Ströme enthalten aktuelle Live-Daten, d.h. jeder Datensatz bezieht sich auf ein gerade stattgefundenes Vorkommnis.
 - Die Auftrittszeitpunkte sowie Reihenfolgen der Ereignisse spielen eine wichtige Rolle.
 - In der Regel sind die Daten feingranular (d. h. von geringer Komplexität) und hochfrequent.
 - Datenströme sind unbegrenzt, da fortlaufend neue Vorkommnisse geschehen.
 - Datensätze stehen zueinander in impliziten Beziehungen, die nicht offensichtlich sein müssen, sondern in den Datendetails versteckt sein können.

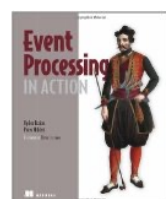
Ziele der Ereignisverarbeitung

- Verarbeitung
 - massiver Datenströme und
 - großer Daten-Volumina mit über hunderttausend Daten pro Sekunde.
- Schritt halten der Verarbeitungsgeschwindigkeit der Datenströme mit den eintreffenden Datenmengen.
- Reaktion in (nahezu) Echtzeit auf eingetretene Ereignisse.
- Verarbeitungsgarantien, z.B. *At-least-Once Delivery*.
- Unterstützung komplexer Datenanalyse, wie z.B.
 - umfangreiche Aggregationen über homogene / heterogene Daten
 - Korrelationen von aktuellen und historischen Daten
 - Erkennen von Ereignismustern
 - auch: Aktionen aufgrund der *Abwesenheit* von Ereignissen.

Literatur

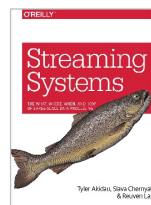
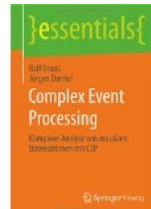
Literatur (1)

- Effektive Software-Architekturen, ein praktischer Leitfaden.
 - Gernot Starke.
 - Hanser Verlag, 2015.
- Software Architecture in Practice.
 - Len Bass, Paul Clements, Rick Kazman.
 - Addison-Wesley, 2009.
- Event Processing in Action.
 - Opher Etzion, Peter Niblett.
 - Manning, 2011.



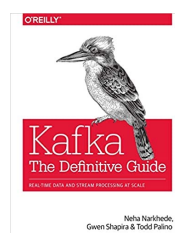
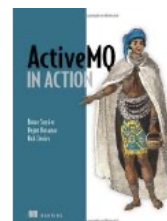
Literatur (2)

- Event-Driven Architecture.
 - Ralf Bruns und Jürgen Dunkel.
 - Springer, 2010.
- Complex Event Processing: Komplexe Analyse von massiven Datenströmen mit CEP.
 - Ralf Bruns und Jürgen Dunkel.
 - Springer, 2015.
- Streaming Systems
 - Tyler Akidau, Slava Chernyak, and Reuven Lax.
 - O'Reilly, 2018.



Literatur (3)

- Complex Event Processing
 - Ulrich Hedtstück.
 - Springer, 2015.
- ActiveMQ in Action.
 - Bruce Snyder, Dejan Bosanac, Rob Davies.
 - Manning, 2011.
- Kafka: The Definitive Guide: Real-time data and stream processing at scale
 - Neha Narkhede, Gwen Shapira, Todd Palino
 - O'Reilly, 2017.



Kontrollfragen und Aufgaben

- Erläutern Sie den Unterschied zwischen prozess- und ereignisgesteuerter Abarbeitung anhand eines ausgewählten Geschäftsprozesses.
- Beschreiben Sie den grundlegenden Aufbau und das Verarbeitungsmodell ereignisgesteuerter Systeme.
- In welchen Situationen kann der Einsatz der Ereignissteuerung besonders hilfreich sein?
- Erklären Sie anhand von gängigen Anwendungsszenarien den Mehrwert, der sich durch den Einsatz von Techniken der Ereignisverarbeitung ergibt.
- Welche Herausforderungen ergeben sich beim Einsatz in einem produktiven Anwendungsumfeld?
- Geben Sie jeweils eine grobe Charakterisierung von *Streaming Systems* und *Complex Event Processing*.
- Erklären Sie den Unterschied zwischen einem „einfachen“ und einem „komplexen“ Ereignis.
- In einem Bibliothekssystem müssen verschiedene Ereignisse wie etwa Buchausleihe und -rückgabe abgebildet werden. Erstellen Sie ein Ereignismodell.