

Informationssysteme für Marketing, Logistik und Produktion

Schule: Hochschule Luzern - Wirtschaft

Studiengang: Wirtschaftsinformatik

Autor: Kevin Stadelmann

Dozent: Klaus-Dieter Gronwald

Datum: Luzern, 8. Juni 2016

Bemerkung: Herr Grondwald hat uns sehr verständlich gemacht, welche Themen der jeweiligen Kapitel prüfungsrelevant sind und welche nicht. So setzt sich der Inhalt dieser Zusammenfassung auch zusammen.

Was ebenfalls nicht in der Zusammenfassung ist, sind die Entscheidungen, welche während dem Beergame getroffen wurden. Herr Gronwald wird spezifische Fragen bezüglich unseren Entscheidungen im Beergame stellen.

Inhaltsverzeichnis

[1 ERP 3](#_Toc453167437)

[1.1 Unternehmerische Ziele 3](#_Toc453167438)

[1.2 ERP-Template 3](#_Toc453167439)

[1.3 Lokalisierung 4](#_Toc453167440)

[1.4 Organisational Readiness 5](#_Toc453167441)

[1.5 IT-Integration Strategien 5](#_Toc453167442)

[2 Supply Chain Management (SCM) 7](#_Toc453167443)

[2.1 Bullwhip-Effekt 7](#_Toc453167444)

[2.2 Demand Forecasting 8](#_Toc453167445)

[2.2.1 Qualitativer Forecast 8](#_Toc453167446)

[2.2.2 Quantitativer Forecast 8](#_Toc453167447)

[2.3 Inventory Management 9](#_Toc453167448)

[2.3.1 Basis EOQ-Modell 10](#_Toc453167449)

[2.3.2 Non-instantaneous Model 11](#_Toc453167450)

[2.3.3 Shortages Model – Fehlmengenmodell 12](#_Toc453167451)

[3 Customer Relationship Management (CRM) 13](#_Toc453167452)

[3.1 Strategisches CRM 13](#_Toc453167453)

[3.2 Analytisches CRM 14](#_Toc453167454)

[4 Big Data Analytics 15](#_Toc453167455)

[4.1 Invertierte Management Pyramide 15](#_Toc453167456)

[4.2 Vier-V 16](#_Toc453167457)

[4.3 Unstrukturierte Daten 16](#_Toc453167458)

[4.4 Text-Analyse 17](#_Toc453167459)

[4.5 Sentiment Analysis – Stimmungsanalyse 17](#_Toc453167460)

# Enterprise Resource Planning (ERP)

Vertriebsstrukturen, Geschäftsprozesse sowie die IT-Infrastruktur sind über ERP miteinander vernetzt und lassen sich somit mit Hilfe einer geeigneten ERP-Strategie standardisieren und mit geeigneten Softwaresystemen realisieren. Die operative Sicht auf ein ERP-System beschreibt es als *System, das sämtliche in einem Unternehmen ablaufenden Geschäftsprozesse unterstützt. Es enthält Module für die Bereiche Beschaffung, Produktion, Vertrieb, Anlagenwirtschaft, Finanz- und Rechnungswesen usw., die über eine gemeinsame Datenbasis miteinander verbunden sind*.

## Unternehmerische Ziele

Ziele wissen

ERP-Systeme werden ausfolgenden drei Grundzielen implementiert:

* Schaffung einer einheitlichen Geschäftsprozessarchitektur
* Standardisierung interner und externer Stammdaten
* Standardisierung der Informationssystemarchitektur

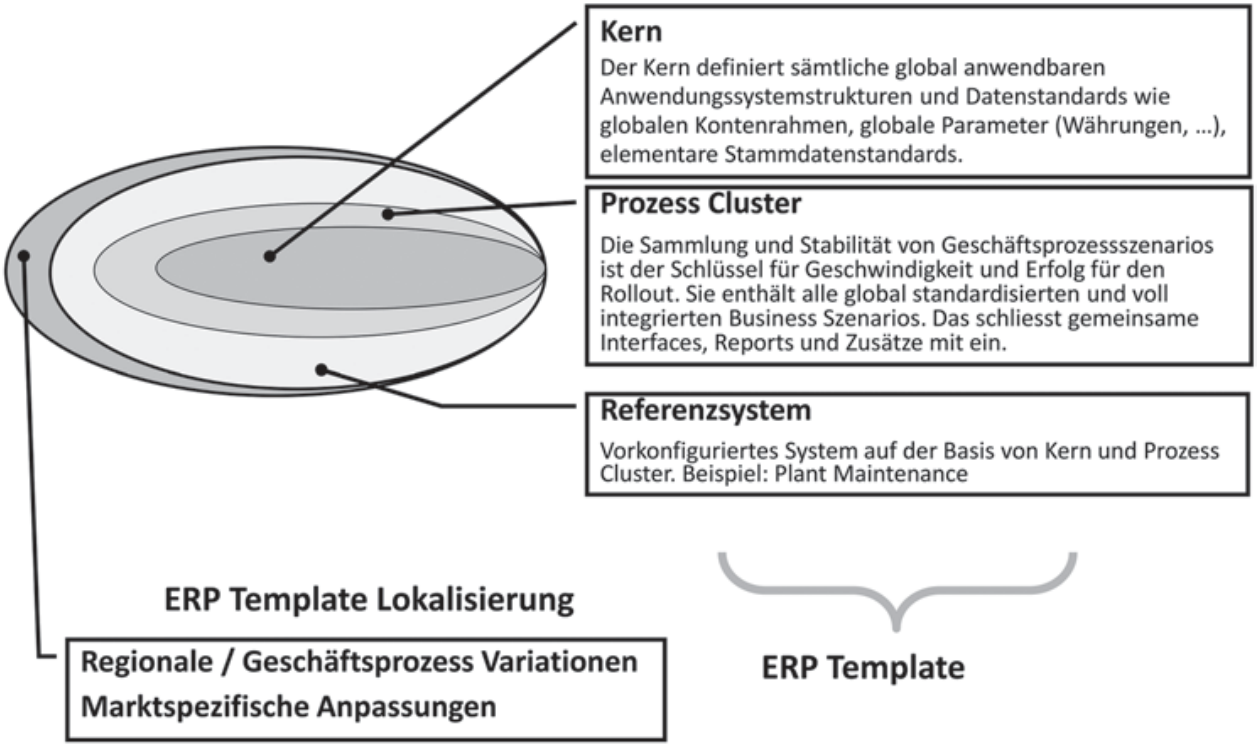
## ERP-Template

Template Grafik kennen

Die Grundidee eines ERP-Systems ist ein Satz standardisierter Geschäftsprozessmodule, die durch Konfiguration von Parametern und Stammdaten an (fast) jede reale Geschäftsprozesssituation ohne Programmierung angepasst werden können (*Configuration*). Diese Module sind um einen zentralen festen Hardware- Softwarekern (Server-Datenbank) angeordnet. In der Praxis lassen sich häufig nicht alle Kundenprozesse mit Standardkonfigurationen abbilden. Es gibt dafür sowohl organisatorische, finanzielle, gesetzliche und firmenpolitische Gründe, die ein aufwendiges Business Process Reengineering verhindern.

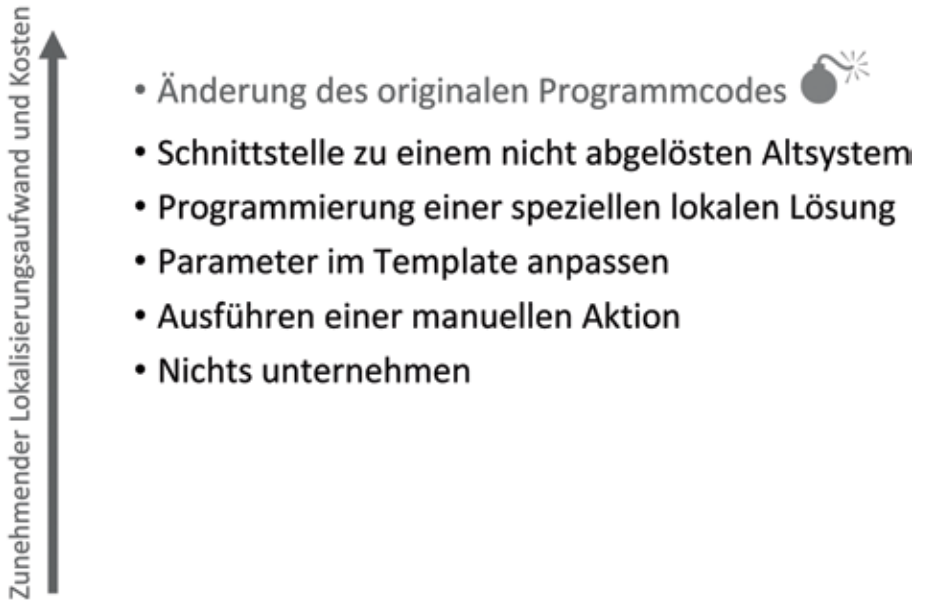
Grundsätzlich gibt es drei Lösungsansätze:

1. Das Unternehmen passt seine Geschäftsprozesse an den ERP-Standard an (Organisational Change Management).
2. Zusätzliche Funktionen müssen programmiert werden ( Customization).
3. Eine Kombination aus a) und b).

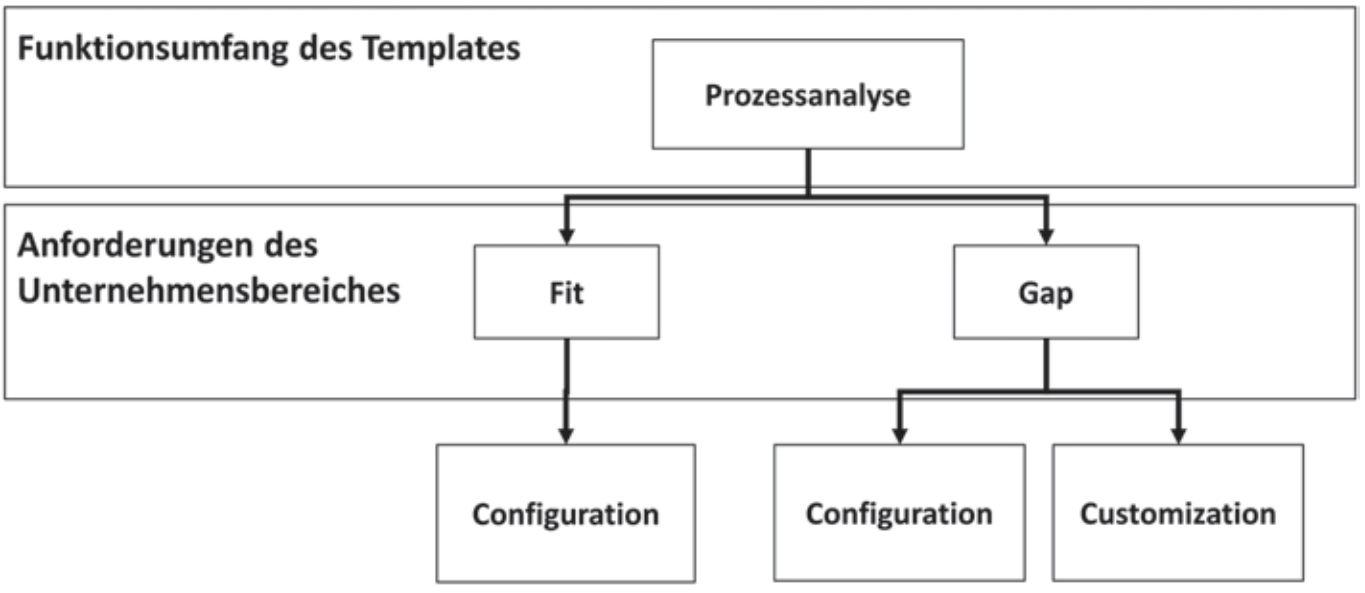


## Lokalisierung

Notwendige Lokalisierungsanforderungen werden bereits während der Template-Entwicklung identifiziert, beurteilt und entweder zurückgewiesen oder implementiert. Die Massnahmen in der folgenden Abbildung sind mit zunehmendem Aufwand und damit verbundenen Kosten aufgeführt. Die obersten Massnahmen, Änderung des originalen Programmcodes in einem Standardsoftwarepaket, ist nur deswegen aufgeführt, weil es Unternehmen gegeben hat, die genau das getan haben. Das ist eine Zeitbombe und natürlich strengstens verboten. Ein gangbarer, wenn auch teurer Weg, ist den Hersteller zu animieren, eine neue Zusatzfunktion in das Programm aufzunehmen. Softwarelieferanten tun das bisweilen für ihre Superkunden oder wenn die neue Lösung eine Marktlücke füllt und sich gut in das bestehende Paket einbauen lässt.



Bevor es aber zu einer Änderung des Programmcodes kommt, wird eine Fit-Gap-Analyse durchgeführt: Entweder kann der Prozess durch Konfiguration «passend gemacht werden» oder die «Funktionslücke» muss durch Konfiguration und Anpassung (Customization) gefüllt werden.



## Organisational Readiness

«Resistance against change» verstehen

ERP-Template Rollouts erfordern ein hohes Mass an Veränderungsbereitschaft von den Mitarbeitenden und Unternehmensorganisationen (weil sie das tägliche Geschäft und die Art und Weise wie die Mitarbeite arbeiten müssen teilweise extrem verändern -> Mensch ist ein «Gewohnheitstier»). Die Veränderungsbereitschaft von Menschen in einer Organisation herzustellen, Restistance against Change abzubauen, ist Aufgabe und Ziel von Organisational Change Management.

**Organisational Change Management**

OCM findet in drei Schritten statt:

* Vorbereitung
* Implementation
* Verankerung in der Organisation

Die Ziele sind:

* Erhöhung der Akzeptanz (Readiness) bei Endbenutzern und Organisation während der Vorbereitung
* Erhöhung der Geschwindigkeit mit der sich die betroffenen Bereiche an die neue Situation anpassen bei gleichzeitiger Minimierung der Unterbrechungen und Einbussen bei der Produktivität während der Implementation
* Verankerung in der Organisation bei gleichzeitiger Erreichung der Businessziele

## IT-Integration Strategien

Die Integration der IT-Umgebungen von verschiedenen Unternehmenseinheiten ist die grösste Herausforderung an Volumen und Zeit, da Informationstechnologien die grösste Zahl von Abhängigkeiten und Funktionen beinhalten. Die IT-Integration Strategie soll eng mit den strategischen und operationellen Zielen der neuen Unternehmensgruppe abgestimmt werden. So wird eine möglichst nahtlose Implementierung erreicht, welche die Stabilität und Dynamik der Geschäftsbereiche während der Implementation nicht gefährdet.

Ziele von IT-Integrationen sind unter vielen anderen:

* Standardisierung: Prozesse, Geschäftsbereiche und Funktionen homogenisieren
* Kostensenkung durch Synergieeffekte
* Elimination von Redundanzen bei Prozessen, Organisation und Infrastruktur

Herausforderungen an die IT-Integration:

* Identifikation und Lösung von IT-Konflikten zwischen den Organisationen
* Analyse und Planung der Integration aller IT-Infrastrukturen ohne Verlust an Effizienz und Best Practises: Total Cost of Ownership (TCO) soll reduziert werden
* Identifikation aller Datenquellen
* Sicherheitsstandard und gesetzliche Vorschriften weiter erfüllen

**Total Cost of Ownership**: Gesamtkosten von Investitionsgütern (Computersysteme in unserem Fall)

Bestehen aus

* direkten Kosten (ca. 60%): Administration, Fixkosten, Kapitalkosten, …
* indirekte Kosten (ca. 40%):Endanwendersupport, Systembetrieb, Verfügbarkeit garantieren,..

Es werden folgende vier IT-Integrationsarten unterschieden:

|  |  |
| --- | --- |
| **Koexistenz / Symbiose**  Erhaltung existierender IT-Systeme. Portale oberhalb der aktuellen Systeme, um Informationen zu aggregieren     * Integrationsgeschwindigkeit: hoch * Kosteneinsparung: gering * Synergien Potential: gering | **Absorption / Übernahme**  Die dominante IT-Organisationsform wird erhalten durch eine vollständigen Integrations- und Konsolidierungsprozess     * Integrationsgeschwindigkeit: niedrig * Kosteneinsparung: hoch * Synergien Potential: mittel |
| **«Best of Breed» / Standardisierung**  Erzielung von Mehrwert durch Kapitalisierung von Synergien aller Organisationseinheiten     * Integrationsgeschwindigkeit: niedrig * Kosteneinsparung: niedrig * Synergien Potential: mittel | **Transformation**  Installation einer vollständig neuen IT-Plattform und Infrastruktur     * Integrationsgeschwindigkeit: niedrig * Kosteneinsparung: niedrig * Synergien Potential: mittel |

Vor allem Absorption und Transformation verstehen, IT-Integration im Beergame hat ihm nicht gepasst, darum wird er nachfragen.

# Supply Chain Management (SCM)

Supply Chain Management (SCM) ist das Steuern von Material- Informations- und Finanzflüssen innerhalb einer Versorgungskette (Lieferkette) vom Rohstofflieferanten über den Hersteller, den Zwischenhandel bis zum Endkunden.

Die ultimativen Ziele von SCM sind:

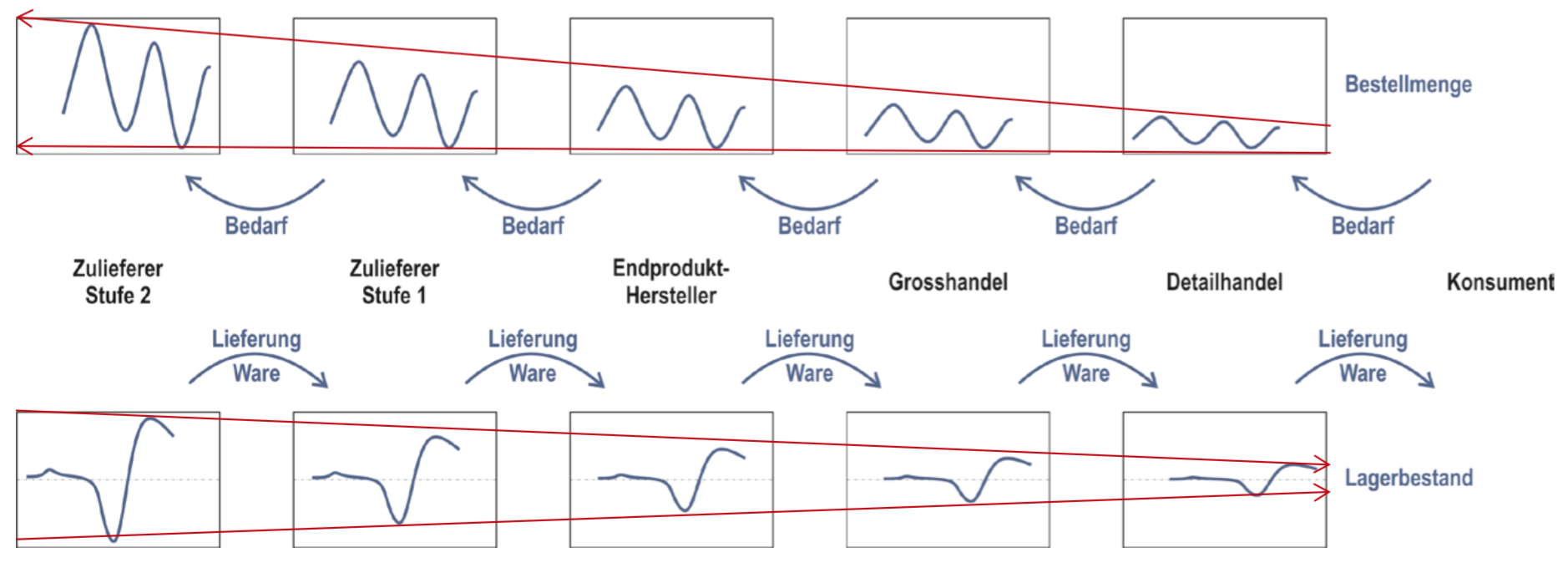
* Lagerbestände minimieren
* Sicherstellung der Lieferbereitschaft

Die zwei wesentlichen Komponenten für die Optimierung und Automatisierung von SCM sind Demand Forecasting und Inventory Management.

## Bullwhip-Effekt

Erklären können und Massnahmen zum Verhindern aufzählen können

* auch als ‚Peitscheneffekt‘ oder ‚Demand Applification' bekannt
* leichte Schwankungen von 3-5% in der Endkonsumentennachfrage schaukeln sich über die Wertschöpfungskette bis zu den Rohstoffherstellern bis zu 70% auf!
* Die Bestellungen seitens des Handels zeigen keinen Zusammenhang mehr mit dem effektiven Kundenbedarf aufweisen.



* leichte Schwankungen in der Konsumentennachfrage schaukeln sich über die Wertschöpfungskette zu den Rohstoffherstellern auf!
* entsprechend nehmen die Lagerbestände zu

**Ursachen**

Mehrfach aktualisierte Absatzprognosen:

Prognosen werden in regelmässigen Intervallen aktualisiert. Dabei werden kurzfristige Schwankungen geglättet. Dies führt zu einer Multiplikation der Überschätzung durch die ganze Supply Chain.

Bündelung von Aufträgen:

Zeitbündelung (konstante Zeitintervalle), Mengenbündelungen (konstante Mengen). Neben den Bestellfixkosten sind auch die Transportkosten zu beachten.

Preisschwankungen:

Kunden horten Produkte aufgrund von Preisschwankungen (Lager werden bei tiefen Preisen gefüllt). Dies verstärkt den Peitscheneffekt.

Mengenkontingentierung und Engpasspoker:

Kunden realisieren, dass in der Supply Chain Engpässe bestehen und bestellen deshalb übermässig viel. Dies führt zu überhöhten Lagerbeständen

Zeitverzögerte Bestellvorgänge:

Die zeitlichen Verzögerungen durch Bedarfsmeldefristen und Lieferung (inkl. Transport) forcieren den Bullwhip-Effekt.

Grosse Distanzen zwischen den einzelnen Stufen der Supply Chain

Führen einerseits zu langen Lieferzeiten (erhöht den Bullwhip-Effekt). Andererseits werden höhere Sicherheitsbestände einkalkuliert, was wiederum den Bullwhip-Effekt verstärkt.

## Demand Forecasting

Forecasting ist ein bedeutendes Element des Demand Management. Es ist die Basis für die Planung und Grundlage für Managemententscheide. Das Ziel ist also Techniken zu entwickeln, welche die Lücken zwischen dem aktuellen Bedarf und dem Forecast minimieren. Forecasting ist somit auch ein iterativer, repetitiver Prozess, welcher die Vorhersage laufend mit der Realität vergleicht und entsprechend anpasst.

In diesem Zusammenhang dient Demand Forecasting vor allem der Unterdrückung des Bullwhips-Effekts.

### Qualitativer Forecast

basieren auf Intuition und Erfahrung der Forecaster. Sie werden vornehmlich für Langzeit Projektionen wie z. B. für die Einführung neuer Produkte verwendet, wenn aktuelle Daten  
nicht relevant sind oder nicht existieren.

* Qualitative Methoden sind u. a. Meinung eines Executive Commitees
* Sales Force Erfahrung
* Kundenbefragungen

### Quantitativer Forecast

lassen sich in zwei dominierende Klassen unterteilen:

**Ursache und Wirkung Methoden** (auch kausale Methoden genannt)

werden verwendet, wenn ein oder mehrere äussere Faktoren (unabhängige Variablen) in Beziehung zum Demand stehen und somit Einfluss auf den Forecast haben, wie Einkommensentwicklung, Wirtschaftsprognosen, etc.(ökonometrische Faktoren). Sie benötigen zum Teil aufwendige statistische Methoden (Regression, Ökonometrische Modelle, Neuronale Netze). Diese Verfahren werden im Abschnitt Big Data Demand Management behandelt.

**Zeitserienanalyseverfahren**

* basieren auf der Annahme, dass die Zukunft eine Fortschreibung der Vergangenheit ist.
* Es werden historische Daten für die Vorhersage der Zukunft verwendet ohne Berücksichtigung äusserer Faktoren

Es gibt folgende Zeitserienanalyse-Verfahren:

* Naiver Forecast
* Einfacher gleitender Mittelwert
* Gewichteter gleitender Mittelwert
* Exponentiell geglätteter Forecast
* Linearer Trend Forecast

Dabei haben alle Verfahren etwas gemeinsam: Sie glätten die Bedarfsschwankungen und mindern so den Bullwhip-Effekt.

## Inventory Management

Beide Modelle verstehen und auf Beergame anwenden können

Durchschnittlich 30–35 % des Materialwertes, 6–15% vom Umsatz (branchenabhängig)  
sind Lagerkosten. Es wird zwischen folgenden zwei Lagerhaltungsmodellen unterschieden:

* **Bestellpunkt-Modell (fixed-order quantity model)**

Fixe (optimale) Bestellmenge abhängig vom Sollbestand. Der Bestellzeitpunkt ist variabel beim Erreichen des Meldebestandes.

* **Bestellrythmus-Modell (fixed-time period model)**

Variable Bestellmenge, abhängig vom aktuellen Verbrauch. Der Bestellzeitpunkt ist regelmässig zu einem fest definierten Zeitpunkt

### Basis EOQ-Modell

EOQ steht für Economic Order Quantity. Das EOQ ist eine einzelne Formel für die Bestimmung der optimalen Bestellgrösse, welche die Summe der Lager- und Wiederbeschaffungskosten minimiert.

|  |  |
| --- | --- |
| Folgende Annahmen werden für dieses Modell getroffen:   * Demand ist bekannt und relativ konstant über die Zeit * Keine Fehlmengen * Konstante Lieferzeit (Lead Time) * Bestellmengen werden vollständig geliefert |  |

Bestimmung der optimalen Bestellmenge unter Berücksichtigung aller relevanten Kostenkomponenten: Bestellkosten und Lagerkosten

Optimale Bestellmenge xOpt: Minimum der Gesamtkosten -> B = L und nach x auflösen!

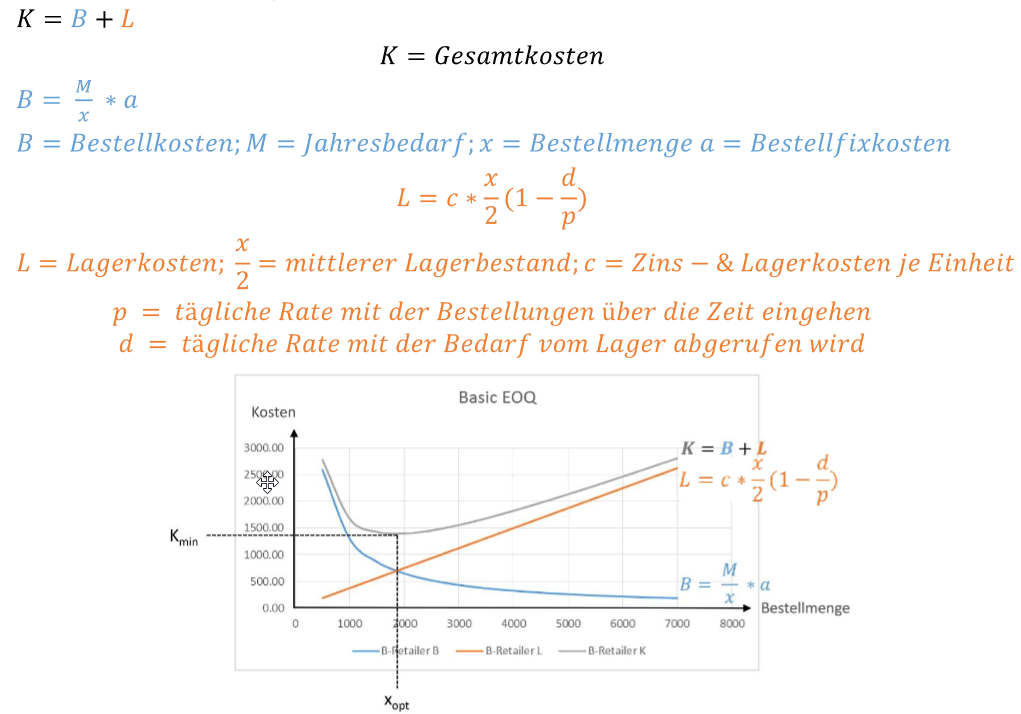
|  |
| --- |
|  |
|  |
|  |
|  |

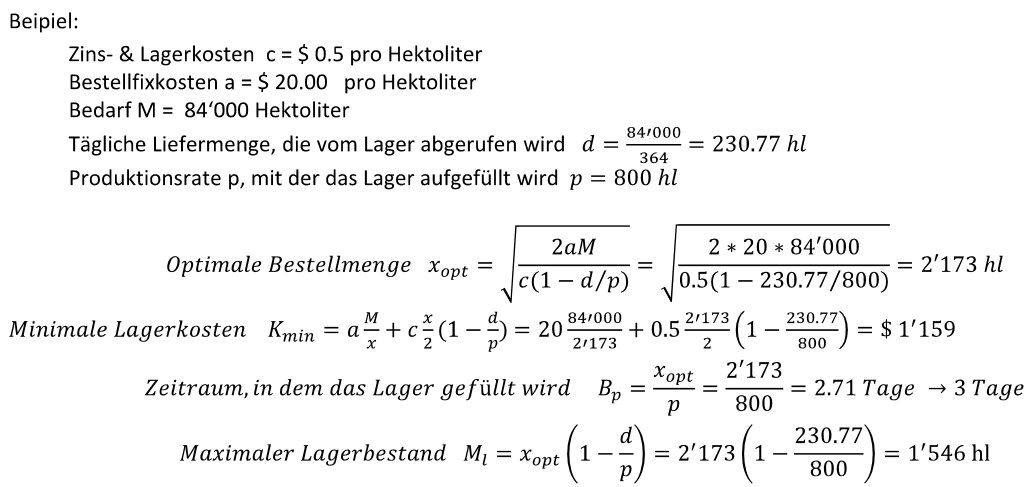
### Non-instantaneous Model

Kommt 100% an der Prüfung

Ist eine Variation des Basis EOQ Modells. Lager Ein- und Ausgänge werden kontinuierlich geführt mit der Rate p, mit der Lieferungen über die Zeit eingehen und dem Bedarf d, der kontinuierlich vom Lager abgerufen wird.

|  |  |
| --- | --- |
| Folgende Annahmen werden für dieses Modell getroffen:   * Demand ist bekannt * Keine Fehlmengen erlaubt * Konstante Lieferzeit (Lead Time) * Bestellmengen werden in Tranchen verteilt über den Zeitraum zwischen zwei Bestellzyklen geliefert |  |



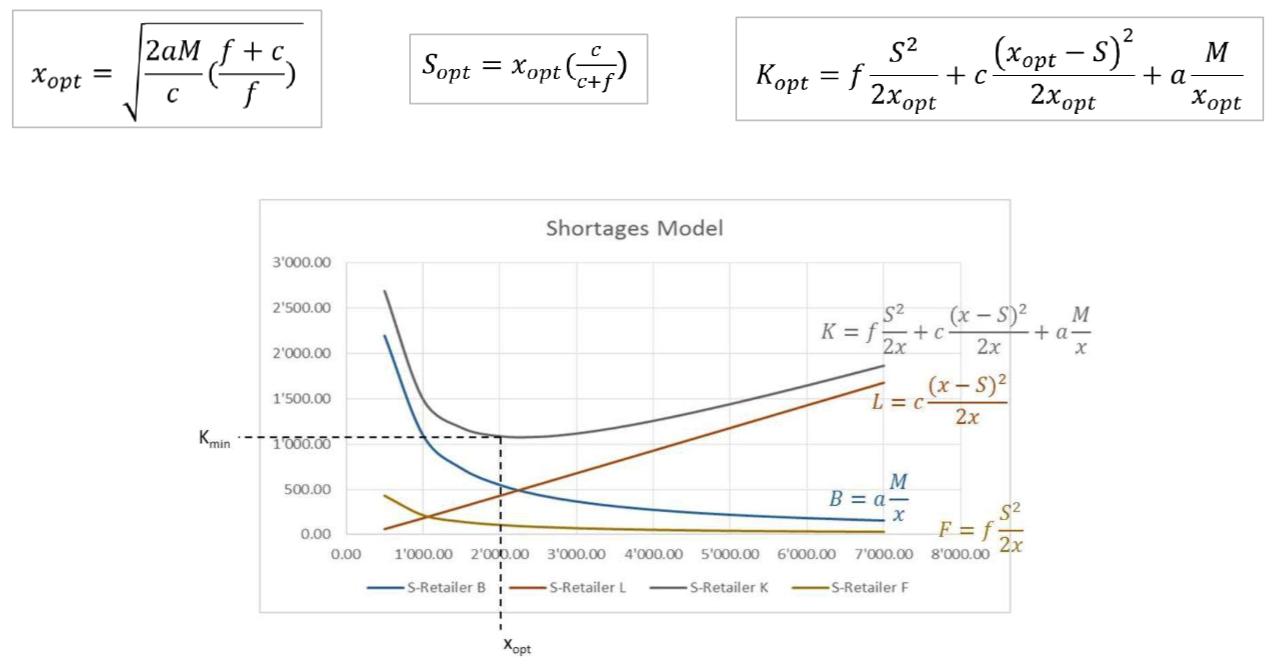


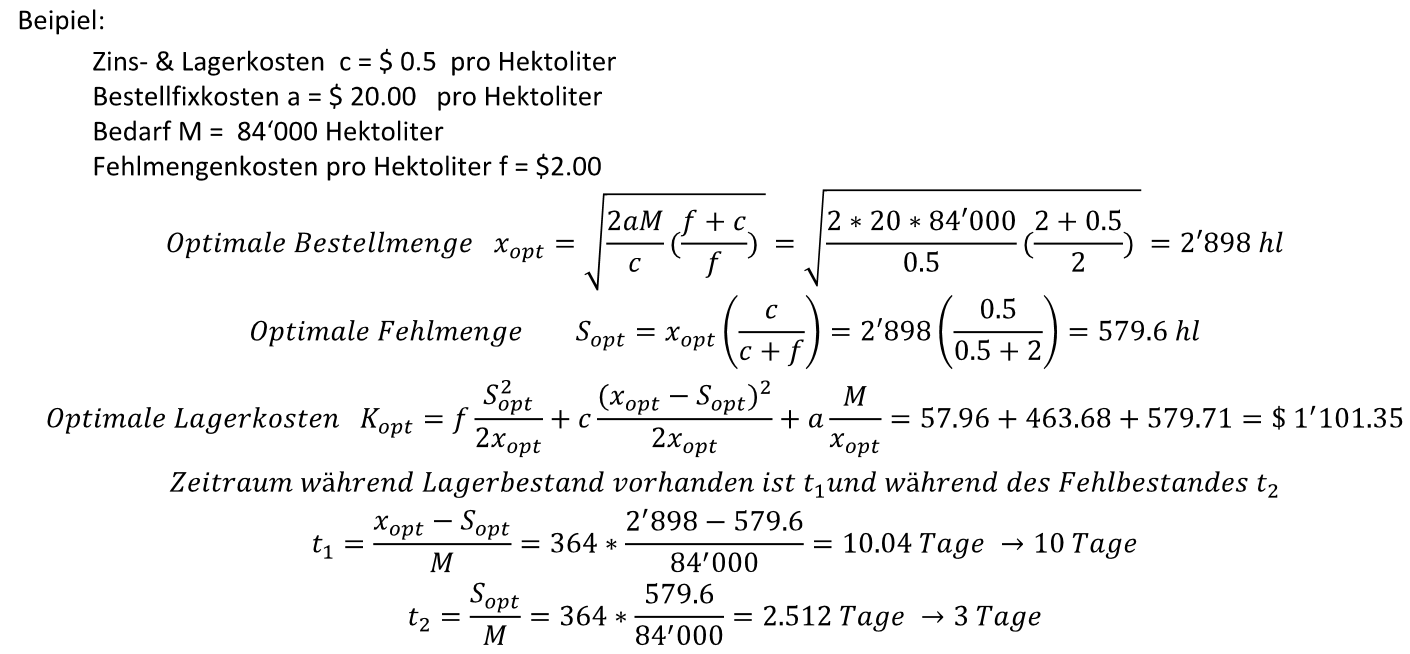
### Shortages Model – Fehlmengenmodell

Entspricht dem Basis EOQ-Modell, dieses Mal sind jedoch Fehlmengen ausdrücklich vorgesehen. Es wird jedoch angenommen, dass der Gesamtbedarf gedeckt wird, indem Fehlmengen nachgeliefert werden.

|  |  |
| --- | --- |
|  |  |

Die drei Kostenkomponenten schneiden sich nicht in einem gemeinsamen Punkt. Für die Bestimmung der optimalen Bestellmenge und des optimalen Fehlbestandes S muss durch Differentiation der Totalkostenkruve K nach x und S herausgefunden werden.





# Customer Relationship Management (CRM)

Customer Relationship Management ist die Pflege von Kundenbeziehungen. Hauptziel von CRM ist die Gewinnmaximierung in einem gesättigten Markt. Dies wird durch folgende Punkte erreicht:

* Kundenbindung: Bestehende Kunden halten
* Kundenselektion: Profitable Kunden identifizieren
* Kundengewinnung: Kundenstamm vergrössern

CRM ist als Frontend System zum Kunden integrierter Bestandteil von Supply Chain Management Systemen und beeinflusst mit Marketingaktionen direkt den Demand (Bedarf). Forecasting ist das Bindeglied zwischen CRM als Demand Generator und SCM als Methode Demand Fulfillment.

Gründe, welche für CRM sprechen:

* Wert eines Unternehmens wird von seinen Kunden bestimmt
* Produkte werden zunehmend austauschbarer
* Wettbewerbsdruck nimmt bei sinkenden Margen zu
* Es ist 6x teurer an einen neuen Kunden zu verkaufen als an einen bestehenden
* Ein unzufriedener Kunde wird 8-10 andere Menschen (negativ) beeinflussen
* 70% der sich beschwerenden Kunden bleiben, wenn man ihr Problem löst

Es werden vier Vorgehensmodelle unterschieden:

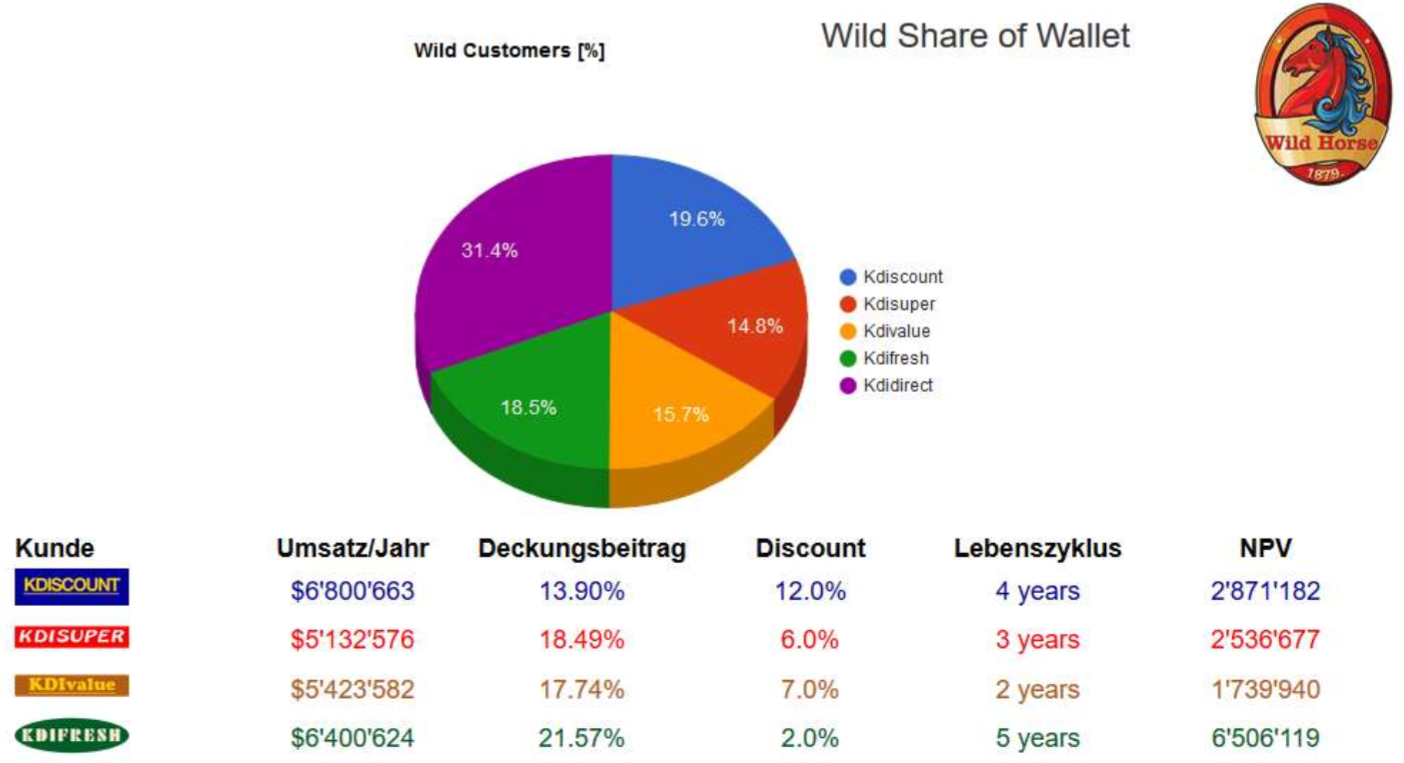
* Strategisches CRM
* Analytisches CRM
* Operatives CRM (hat Gronwald gestrichen)
* Kommunikatives CRM (hat Gronwald gestrichen)

## Strategisches CRM

Das Ziel von strategischem CRM ist, möglichst umfassendes Wissen über Kunden aufzubauen. Dieses Wissen soll verwendet werden, um die Interaktion zwischen Unternehmen und Kunden zu optimieren. Ziel ist es den Customer Lifetime Value (CLV) für das Unternehmen zu maximieren.

**Customer Lifetime Value (CLTV)** = Kundenwert über den Gesamtzeitraum einer Geschäftsbeziehung

**Share of Wallet** = Anteil der Kaufkraft eines Kunden, der beim Unternehmen bleibt.



Langjährige Kundenbindungen -> ist das heute überhaupt noch aktuell? -> 100% Prüfungsfrage

Antwort: Ja! Heute umso mehr, siehe Gründe für CRM -> Globalisierung (Wettbewerbsdruck steigt) und Produkte sind immer mehr austauschbar. Deshalb ist eine starke Kundenbindung sehr wichtig.

## Analytisches CRM

Beim analytischen CRM nutzt man Kundendaten, um profitable Beziehungen zwischen Kunden und Unternehmen zu bilden. Es nutzt traditionelle Business Intelligence (BI) Methoden wie Data Warehouse, Data Mining und Online Analytical Processing Systeme (OLAP), um Kundenzufriedenheit, Kundentreue (Share of Wallet) zu bestimmen und daraus aktive Massnahmen zur Optimierung der entsprechenden Parameter abzuleiten.

# Big Data Analytics

Themenbereich BI wurde gestrichen (OLAP, OLTP, ETL Prozess, BI-Data Mining). Der BI-Themenbereich wird in diesem Kapitel jedoch trotzdem angeschnitten und erwähnt.

Big Data ist keine Technologie. Big Data verwendet Technologien, um auf die richtigen Fragen Antworten in Echtzeit zu finden und diese im Geschäftsalltag gewinnbringend umzusetzen. Die richtige Kombination von Big-Data-Methoden, Tools und Technologien wie Hadoop, In-Memory-Computing, NoSQL-Datenbanken, Social Media und traditionellem Data Mining ermöglichen Unternehmen jeder Grösse, Antworten auf ihre Fragen zu finden. Die Schwierigkeit besteht darin, die richtigen Fragen zu stellen. Big Data ist primär kein Datenmengenproblem, sondern ein Komplexitätsproblem. Dies betrifft sowohl die Fragestellungen als auch den Zugang zu den richtigen Daten. Big Data setzt das kundenorientierte Geschäftsmodell der *invertierten Pyramide* voraus. Damit verbunden ist ein Umdenken von einem allein auf Intuition und Erfahrung basierenden Entscheidungsprozess zu einem datendominierenden Prozess als Teil der Unternehmenskultur. Big Data erfordert Flexibilität und ein für Innovation offenes Unternehmen sowohl von den Produkten als auch von dem Geschäftsmodell her – und vor allem ein *Customer Relationship Management*, das kein Tool, sondern die Basis für ein aktives persönliches Kundenverhältnis ist. Die Rolle der IT wechselt von der Bereitstellung und Aufarbeitung grosser Datenmengen hin zur Sicherstellung des Zugangs zu Daten, Netzen und Dienstleistungen, der notwendigen Infrastruktur und Bereitstellung von *Big Data Self Services*

Big Data Analytics ist das Resultat von vier globalen Trends:

* Moore’s Law,
* Mobile Computing
* Social Networking
* Cloud Computing

Moore’s Law postuliert (seit 1965), dass die Anzahl Transistoren auf einem Microchip sich alle 24 Monate verdoppelt. INTEL schätzt, dass sich dieser Trend bis 2029 fortsetzen wird. Ergebnis: Computerleistung wird ständig billiger.

## Invertierte Management Pyramide

|  |  |
| --- | --- |
| Business Intelligence hat sich aus Management Support Systemen (MSS) entwickelt. Data Mining wurde ursprünglich verwendet, um CEOs und Controllern detaillierte Business Information zu liefern, die typische Darstellung als Pyramide mit den Kunden an der Basis, der *Kundenbasis,* und den CEOs oben als Entscheider. Mit der Einführung von CRM als strategisches Marketinginstrument wird Data Mining verwendet, um das unternehmerische Denken und Handeln kunden- und datenbasierend zu gestalten und somit die Entscheidungskompetenz der Kunden*basis* zu geben. |  |

## Vier-V

Um Big Data gegenüber BI abzugrenzen wird im Allgemeinen von vier Dimensionen (Vier-V) gesprochen.

**Volume**

Hierbei ist nicht das Datenvolumen, welches Big Data charakterisiert, sondern dass die riesigen Mengen an Daten in Echtzeit anfallen und interpretiert werden müssen. Die Verarbeitung grosser Datenmengen ist nicht neu. Es gab sie schon, bevor Big Data als Begriff und Disziplin auftauchte

**Velocity**

Die Geschwindigkeit, mit der Daten bei einer Organisation ankommen und wie schnell sie verarbeitet werden.

**Variety (Vielfalt und Komplexität der Datenquellen)**

Das Ziel ist der Versuch, sämtliche Datenquellen, strukturiert, unstrukturiert, semi-strukturiert realtime in Entscheidungsprozesse einzubeziehen.

**Veracity (Datenunsicherheit)**

Unsicherheit über die Datenverfügbarkeit, Schwankungen von Streaming Data, die richtigen Daten in der richtigen Menge zur richtigen Zeit. Die Glaubwürdigkeit von Daten. Zum Beispiel das Erkennen von Sarkasmus in Text Mining und Sentiment Analysis.

Als fünfte Dimension wird manchmal auch von Value (ökonomischer Nutzen) gesprochen. Dort stellt sich die Frage, wer den direkten Nutzen aus gewonnen Informationen zieht oder welche Business Entscheidungen getroffen werden müssen.

## Unstrukturierte Daten

Big Data lässt sich ebenfalls über das V *Variety* definieren. Big Data ist die Summe strukturierter und unstrukturierter Daten aus dynamischen unterschiedlichen Quellen und Medien, die realtime verarbeitet werden. An dieser Stelle ist das Datenvolumen von Bedeutung. 90 % aller Daten seit Anbeginn der Zeit wurden zwischen 2012 und 2014 generiert. Die meisten dieser Daten stammen aus unstrukturierten Quellen. Diese konnten bereits seit ca. 2004 nicht mehr konventionell verarbeitet, da sie nicht mehr zwischengespeichert werden konnten.

Nach dem Grad der Strukturiertheit, können Daten in vier Klassen unterteilt werden:

1. **Strukturiert (Structured)**  
   Daten enthalten einen definierten Datentyp, ein wohldefiniertes Format, und eine  
   Datenstruktur. Beispiel: Transaktionsdaten und OLAP.
2. **Semi-Strukturiert (Semi-Structured)**  
   Textdateien mit einem identifizierbaren Muster, einer Struktur, die sich analysieren  
   lässt. Beispiel XML Dateien, die selbsterklärend sind und durch ein XML Schema  
   beschrieben werden.
3. **Quasi-Strukturiert (Quasi-Structured)**  
   Textdateien mit inkonsistenten, variierenden Datenformaten, die mit Aufwand und den  
   entsprechenden Tools formatiert werden können. Beispiel: Web Clickstream Daten, die  
   Inkonsistenzen bei Werten und Formaten haben können.
4. **Unstrukturiert (Unstructured)**  
   Daten haben keine sichtbare oder leicht identifizierbare Struktur und sind normalerweise in unterschiedlichen Datenformaten gespeichert. Beispiel: Textdokumente,  
   PDFs, Bilder und Videos.

## Text-Analyse

Textanalyse (Text Mining) ist primär der Prozess unstrukturierten Text so aufzubereiten, dass er mit anderen analytischen Methoden weiterbehandelt werden kann, um daraus Information zu gewinnen. Text-Analyse umfasst die folgenden Prozesse:

* Content Categorization: Einteilung von Textdokument in Kategorien
* Text Mining: Erkennen von Mustern und Strukturen und daraus Vorhersagen zu machen oder das Verhalten zu verstehen
* Sentiment Analysis: Bewertung von Textinhalten als positiv oder negativ (Polarisierung)

Text Mining lässt sich in sieben Kategorien unterteilen, die zwar separat als Prozesse identifizierbar aber miteinander vernetzt:

1. Search and information retrieval: Speicherung und Retrieval von Textdokumente (z. B. Suchmaschinen, Keywordsuche).
2. Document Clustering: Gruppierung und Kategorisierung von Begriffen, Abschnitten, Paragraphen und Dokumenten mit Hilfe von Data Mining Clustering Methoden.
3. Document Classification: Gruppierung und Klassifizierung von Abschnitten, Textpassagen, Paragraphen und Dokumenten mit Hilfe von Data Mining Klassifizierungsmethoden.
4. Web Mining: Daten und Text Mining im Web.
5. Information Extraction: Identifikation und Extraktion relevanter Fakten und Beziehungen aus unstrukturiertem Text, was die Transformation unstrukturierten und semistrukturierten Texten in strukturierte Daten beinhaltet.
6. Natural Language Processing: Mensch-Maschine-Interaktion, welche es Computern erlaubt, Bedeutung aus natürlicher Sprache zu ermitteln und daraus Handlungen abzuleiten (computational linguistics).
7. Concept Extraction: Gruppierung von Worten und Sätzen in semantisch ähnliche Gruppen.

## Sentiment Analysis – Stimmungsanalyse

Sentiment ist ein Prozess der Bewertung von polarisierten Meinungen in Texten. Das können Textdokumente, Websites, Social Media, etc. sein. Die Bewertungskriterien sind *positiv, negativ* oder *neutral.* Sentiment Analysis kann sich auf einzelne Begriffe, oder Phrasen beziehen. Beispiel *Alkohol.* Eine Twitter Textanalyse zum Thema wird eine Reihe von Aussagen dazu liefern. Diese werden nach Begriffen analysiert, die positiven oder negativen Charakter haben. Dazu gibt es mittlerweile Bibliotheken in vielen Sprachen. Diese sind geteilt in separate Wortsammlungen, die als positiv bzw. negativ bewertet werden.

Sentiment Analysis lässt sich ebenso auf Texte einzelner Personen anwenden und kann so ein Persönlichkeitsprofil zu bestimmten Ansichten, Werten und Themen liefern.