Business Intelligence & Analytics samenvatting
WEEK 1

BI&A – WHY?
Why important for organizations? What is happening, what is different? Strategy making: new games, new rules! Big data as driver of business model innovation.

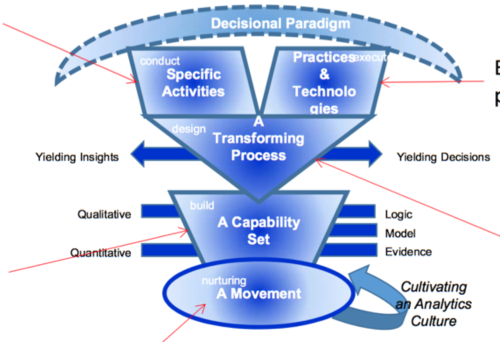
BI&A – WHAT?
What does it include? New insights. Types of business analytics capabilities (descriptive, predictive and prescriptive analytics). The 3 eras of analytics: era of BI until 2000, era of Big Data until 2014 and era of Data-enriched offerings from 2014 onwards.

BI&A – HOW?
The 3 dimensions of a CDO:

1. Collaboration direction dimension: inwards / outwards
2. Data space dimension: traditional data / big data
3. Value impact dimension: service / strategy

> Holsapple et al. (2014): A unified foundation for business analytics, Decision Support Systems

Business perspectives:



Conduct as set of specific activity types for operating on available evidence: accessing, examining, aggregating and analyzing evidence.

Umbrella that covers all elements

Execute a collection of practices and technologies.

Build a set of competencies possessed by the organization and its processors.

Design the process of transformation. The process that drives, coordinates, controls, and measures the transformation.

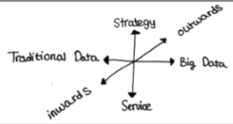
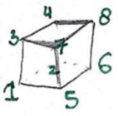
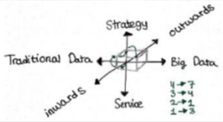
Nurture the right Mind-Set

> Lee et al. (2014): A cubic framework for the CDO: Succeeding in a world of big data.
A Chief Data Officer (CDO) is different from a traditional data manager in two ways:

1. A CDO has the power to build organizational capability that can energize and sustain the entire organization and extended enterprise.
2. A CDO can be held accountable for failure of leadership in solving data problems.

Then there are the following dimensions:

1. Collaboration direction dimension: inwards / outwards
   Inwards = focus on internal business processes associated with internal business stakeholders
   Outwards = focus on stakeholders in external value chain and environment (customer/partner/supplier)
2. Data space dimension: traditional data / big data
   Big data offers innovative opportunities to further improve operations or develop new business strategies based on new insights. Traditional data cannot do this.
3. Value impact dimension: service / strategy

3 CDO dimension	8 CDO roles	CDO role evolution
 Strategy outwards Traditional Data Big Data inwards Service	 Some CDO may take multiple roles, but it has one primary role in most cases.	 Strategy outwards Traditional Data Big Data inwards Service 4→7 3→4 2→1 1→3

1 =	Coordinator	➔ Inward, traditional data, service
2 =	Reporter	➔ Outward, traditional data, service
3 =	Architect	➔ Inward, traditional data, strategy
4 =	Ambassador	➔ Outward, traditional data, strategy
5 =	Analyst	➔ Inward, Big Data, service
6 =	Marketer	➔ Outward, Big Data, service
7 =	Developer	➔ Inward, Big Data, strategy
8 =	Experimenter	➔ Outward, Big Data, strategy

> Lavalle et al. (2010): Analytics: The new path to value – How the smartest … transform insights into action.
Top performing organizations are twice as likely to use analytics in daily operations and future strategies as lower performing organizations. The adoption barriers organizations face most are managerial and cultural instead of related to data or technology. The 3 management needs are:

- Reduced time to value
- Increased likelihood of transformation
- Greater focus on achievable steps

Then there are recommendations to achieve these needs:
1. Focus on biggest and highest value opportunities – attack a big important problem that can demonstrate value and helps the organization toward action.

Using the PADIE technique – Process-Application-Data-Insight-Embedded – helps organizations to operationalize insights drawn from data and helps users across the organization to understand a specific business challenge. Also enables business and analytics teams to work together to create analytic models based on use cases that show analytics in action.

2. Within each opportunity, start with questions, not data – understand the problem and the insights needed to solve it before working on the data that will yield the insights.

Organizations should start by pinpointing to be leveraged, then use readily available data to test the analytic models. Actions based on those insights will help define the next set of insights and data needed. The traditional approach of starting with a comprehensive data program creates too much lag time before insights can be put into action.

3. embedded insights to drive actions and deliver value – ensure the end result impact by making the info come to life. Express use cases and data-driven insights in ways that even nonexperts can understand and act upon.

Organizations expect the ability to visualize data differently will be the most valuable technique in 2 years. There will be several developments, dashboards will show what sales could be next quarter instead of just the actual last quarter sales. Secondly, simulations evaluating alternative scenarios will automatically recommend optimal approaches.

4. Keep existing capabilities while adding new ones - Even as centralized analytics oversight grows, keep distributed, localized capabilities in place.

When executives use analytics more to inform daily decisions and actions, this increasing demand for insights keeps resources at each level involved, expanding analytic capabilities even as activities are shifted for efficiencies.
The frequency with which analytics is used to support decisions increases as organizations transition from one level of analytic capability to the next. At the same time, analytics migrate toward more centralized units, first at the local line of business level and then at the enterprise level, while the portion of analytics performed at points of-need and with IT remain stable.

5. Use an information agenda to plan for the future – Opportunistic application of analytics can create value fast, but it must be part of an enterprise-wide information-and-analytics plan.

All obtained data must be shaped into an information foundation. We can conclude that organizations want data that is integrated, consistent and trustworthy. Information agenda provides a vision and high- level road map for info that aligns business needs to growth in analytics sophistication with the underlying technology and processes spanning: Info governance policies/tool kits/practices, data architecture/currency, data management/integration/middleware, analytical tool kits based upon user needs.

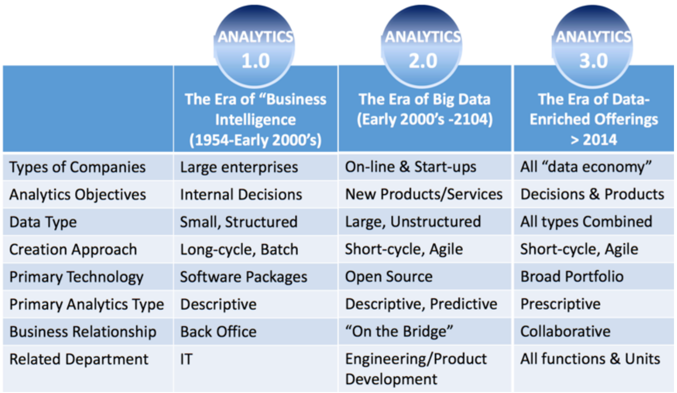
An organization is always in one of the following 3 levels:

1. Aspirational – far from achieving desired analytics goals, organization has few of the necessary building blocks. → To assemble the best people and resources to make the case for investments in analytics. To get sponsorship for initial projects, identify big business challenges that can be addressed by analytics and find data you have that fits the challenge.

2. Experienced – there is some analytic experience, so they can start optimizing the organization.
→ Make the move to enterprise analytics and manage it by keeping focus on the big issues that everyone recognizes. Collaborate to drive enterprise opportunities without compromising departmental needs while preventing governance from becoming an objective unto itself.

3. Transformed – a lot of experience, analytics is used as a competitive differentiator.
→ Discover and champion improvements in how you are using analytics. Accomplished a lot with analytics but want to do more. Focus your analytics and management bandwidth to go deeper not broader, but recognize it will be critical to continue demonstrating new ways of how analytics can move the business toward its goals.

> Davenport, T.H. (2013): Analytics 3.0



	ANALYTICS 1.0 The Era of "Business Intelligence (1954-Early 2000's)	ANALYTICS 2.0 The Era of Big Data (Early 2000's -2104)	ANALYTICS 3.0 The Era of Data-Enriched Offerings > 2014
Types of Companies	Large enterprises	On-line & Start-ups	All "data economy"
Analytics Objectives	Internal Decisions	New Products/Services	Decisions & Products
Data Type	Small, Structured	Large, Unstructured	All types Combined
Creation Approach	Long-cycle, Batch	Short-cycle, Agile	Short-cycle, Agile
Primary Technology	Software Packages	Open Source	Broad Portfolio
Primary Analytics Type	Descriptive	Descriptive, Predictive	Prescriptive
Business Relationship	Back Office	"On the Bridge"	Collaborative
Related Department	IT	Engineering/Product Development	All functions & Units

There are 10 requirements in order to take advantage of analytics 3.0:
1. multiple types of data, often combined
2. a new set of data management options
3. embedded analytics
4. fast technologies and methods of analysis
5. data discovery
6. cross-disciplinary data teams
7. chief analytics officers
8. prescriptive analytics
9. analytics on an industrial scale
10. new ways of deciding and managing

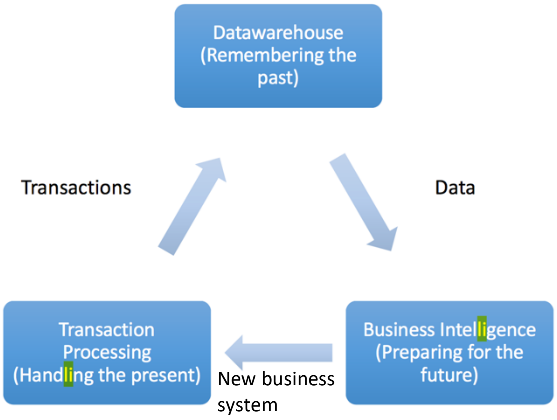
WEEK 2

Def. Data = that what exists before to the argument/interpretation that will translate them into facts, evidence and information

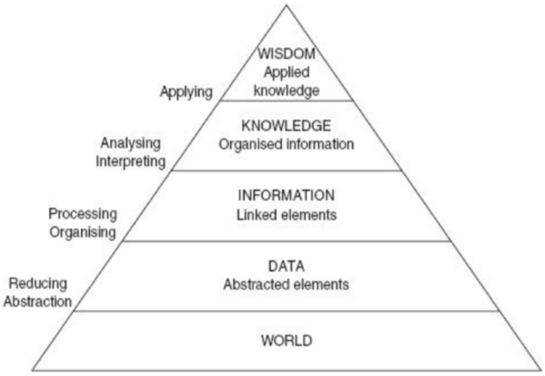
Several types of data are

- Form (qualitative/quantitative)
- Structure (semi-structured/(un)structured)
- Source (captured/derived/exhaust/transient)
- Producer (primary/secondary)
- Type (indexical/attribute/metadata)

Information lifecycle of the firm:


Datawarehouse (Remembering the past)
Transactions
Data
Transaction Processing (Handling the present)
New business system
Business Intelligence (Preparing for the future)

Knowledge pyramid:



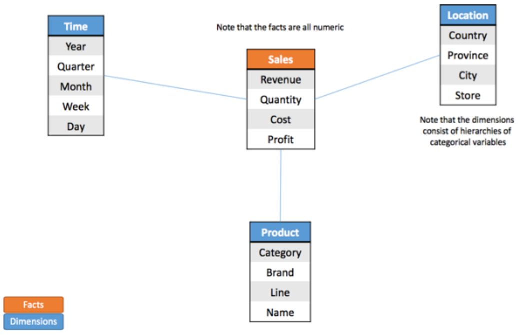
Multidimensional data models are divided into:

- Facts (variable dimensions) → what is tracked sales, revenue, units sold
- Dimensions (identifier dimensions) → tagging what is tracked time, products, store of sale

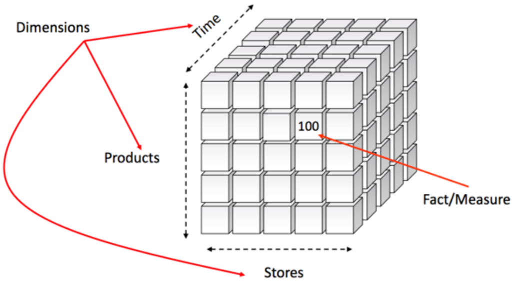
Key concepts of a multidimensional data model:

1. **Multidimensionality**: organizing/presenting/analyzing data by several dimensions (sales by region, time, product, etc.)
2. **Multidimensional analysis**: analysis of data by ≥ 3 dimensions
3. **Multidimensional modelling**: modelling method that involves data analyisis in several dimensions

Star schema:

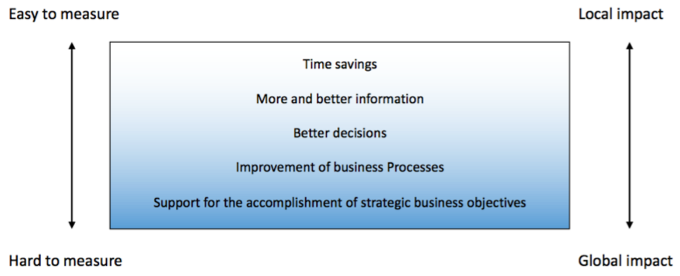


Decision cube:



> Watson (2002): The benefits of data warehousing: Why some organizations realize exceptional payoffs

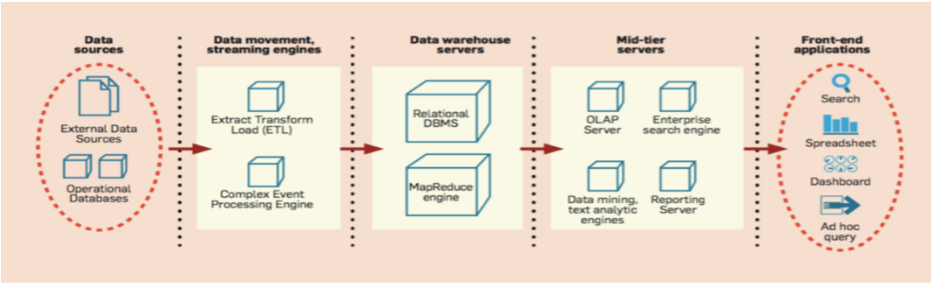
Benefits of data warehousing:


Easy to measure
Local impact
Time savings
More and better information
Better decisions
Improvement of business Processes
Support for the accomplishment of strategic business objectives
Hard to measure
Global impact

8-step model for achieving benefits:
Step 1: Establish a sense of urgency
Step 2: Form a powerful guiding coalition
Step 3: Create a vision
Step 4: Communicate the vision
Step 5: Empower others to act on the vision
Step 6: Plan for and create short-term wins
Step 7: Combine improvements and produce still more change
Step 8: Institutionalize the new approaches

> Chaudhiri et al. (2011)

Business intelligence architecture:



Technologies that are used in BI where research can play/has played an important role:

1. Index structures – Indexes on columns that help you retrieve what you want

- **index intersection**: operations that reduce the need to access base table. Column1=val1 AND column2=val2 is a way of using bitmap indexes
- **materialized views**: greatest strength of this is its ability to specifically target certain queries by effectively hiding their results. But this can also limit its applicability as for a slightly different query it might not be able to use this view. Good physical design is a mix of index and materialized views.
- **Partitioning**: a way to improve the performance and manageability.
- **Column-oriented storage**: ability to have significantly greater data compression than row-oriented storage as column data values are more repetitive than across columns. Also, it is easier to only scan accessed columns than with row-oriented.

2. Data compression – compression can reduce the amount of data that needs to be scanned and reduces the amount of storage required for a database; Increases amount of data that can be cached in memory (pages can be kept in compressed form); Common query operations can be done without decompressing; Compressing data that is transferred over increases network bandwidth.

- **Null suppression**: commonly used data types in database management systems are fixed lengths (integers, money). Only non-null part of the values is stored along with actual length of the value.
- **Dictionary suppression**: identifies repetitive values in data and constructs dictionary that maps such value to more compact. 0,1,2 for example.

3. OLAP Servers

- **MOLAP server**: supports multidimensional view by using multidim. array abstraction. Typically precompute large data cubes to speed up query processing.
- **ROLAP server**: multidim. models have to be mapped into relation/SQL queries. Often use star schema to represent multidim. data model, do not provide support for attribute hierarchies.

- **HOLAP server**: combi of MOLAP and ROLAP, split storage data in MOLAP and a relational store. Perform density analysis to identify sparse/dense sub-regions of multidim. space.

4. Relational servers – DW need to be able to execute complex SQL queries as efficiently as possible against large databases. The first key technology needed to achieve this is query optimization.

- **Query optimization**: a key enabler of BI, converting minutes into days causes a different execution time.
- **Parallel processing and appliances**: play significant role in processing queries over massive databases. Basis paradigm is data parallelism, this has 2 basic architectures: **shared disk**, where each processor has a private memory but a shared disk, and **shared nothing**, where they have private memory/disk.
- **DW appliances**: an integrated set of server and storage hardware, operating system and DBMS software pre-installed and pre-optimized for Data Warehousing.

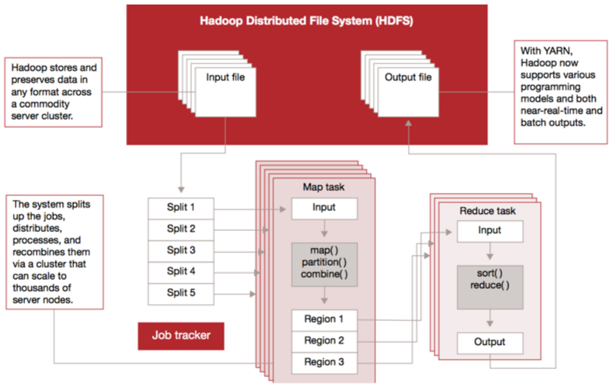
2 key techniques for data distribution: partitioning and cloning.

> Stein et al. (2014): The enterprise data lake: Better integration and deeper analytics

Def. Data lake = a repository for large quantities and varieties of data, both structured and unstructured.

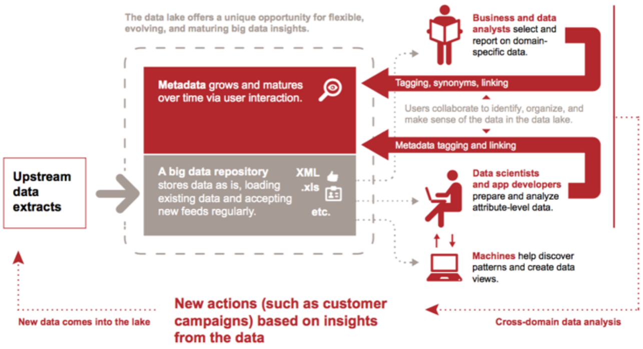
Hadoop allows hospital's disparate records to be stored in native formats for later parsing instead of all-or-nothing integration. Data lake has made possible several data analysis projects including the ability to predict the likelihood of readmissions and take preventive measures to reduce the number of readmissions.

Basic Hadoop architecture for scalable data lake infrastructure:


Hadoop Distributed File System (HDFS)
Hadoop stores and preserves data in any format across a commodity server cluster.
Input file
Output file
With YARN, Hadoop now supports various programming models and both near-real-time and batch outputs.
The system splits up the jobs, distributes, processes, and recombines them via a cluster that can scale to thousands of server nodes.
Split 1
Split 2
Split 3
Split 4
Split 5
Job tracker
Map task
Input
map()
partition()
combine()
Region 1
Region 2
Region 3
Reduce task
Input
sort()
reduce()
Output

Source: Electronic Design, 2012, and Hortonworks, 2014

Data flow in the data lake


The data lake offers a unique opportunity for flexible, evolving, and maturing big data insights.
Metadata grows and matures over time via user interaction.
A big data repository stores data as is, loading existing data and accepting new feeds regularly.
XML
.xls
etc.
Upstream data extracts
Business and data analysts select and report on domain-specific data.
Tagging, synonyms, linking
Users collaborate to identify, organize, and make sense of the data in the data lake.
Metadata tagging and linking
Data scientists and app developers prepare and analyze attribute-level data.
Machines help discover patterns and create data views.
New data comes into the lake
New actions (such as customer campaigns) based on insights from the data
Cross-domain data analysis

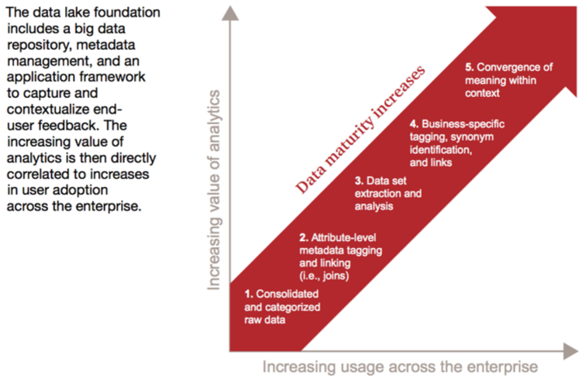
Adoption of Hadoop has several reasons:

- Cost (it costs 10 to 100 times less than conventional data warehousing)
- Opportunity to defer labor-intensive schema development and data cleanup until an organization has identified a clear business need.
- Data lakes are more suitable for the less-structured data companies needed to process

A data lake has to follow four criteria for a good definition:

- **Size and low cost**: big size, low cost.
- **Fidelity**: preserve data in its original form and capture changes to data and contextual semantics throughout the data lifecycle.
- **Ease of accessibility**
- **Late binding**: flexible, task-oriented structuring and does not need up-front data models.

Data lake maturity:



The data lake foundation includes a big data repository, metadata management, and an application framework to capture and contextualize end-user feedback. The increasing value of analytics is then directly correlated to increases in user adoption across the enterprise.

Increasing value of analytics
Data maturity increases
5. Convergence of meaning within context
4. Business-specific tagging, synonym identification, and links
3. Data set extraction and analysis
2. Attribute-level metadata tagging and linking (i.e., joins)
1. Consolidated and categorized raw data
Increasing usage across the enterprise

Data Lakes can provide benefits over traditional data with low costs but require many practical considerations and a thoughtful approach to governance, particularly in more heavily regulated industries. Areas to consider include:

- **Complexity of legacy data**
- **Metadata management**
- **Lake maturity**
- **Staging area or buffer zone**

WEEK 3

> Tene et al. (2013): Big Data for All: Privacy and User Control in the Age of Analytics

Big data is a big influence/benefit in the following sectors (data-driven environmental innovation):

- Health Care
- Mobile
- Smart grid
- Traffic management
- Retail
- Payments
- Online

Big data doesn't only come with big opportunities and benefits, but also with big concerns. Some unique privacy risks presented by big data are:

- **Incremental effect**: once any piece of data has been linked to a person's *real* identity, any association between this data and a *virtual* identity breaks anonymity of the latter. Paul Ohm warns that this will lead to a "database of ruin": chewing away on an individual's privacy until his or her profile is completely exposed. For this reason, the European's Commission proposed a "right to be forgotten", which allows users to demand organizations to wipe their data slate clean.
- **Automated decision-making**: Joseph Turow argues that increased personalization based on opaque corporate profiling algorithms poses a risk to open society and democratic speech. By "pigeonholing" individuals into pre-determined categories, automated decision-making compartmentalizes society into pockets (or "echo chambers) of like-minded individuals.
- **Predictive analysis**: can lead to "redlining". Particularly problematic when based on sensitive categories such as health, race, or sexuality.
- **Lack of access and exclusion**: individuals are excluded from the benefits of big data. This manifests in two main ways. First, online interactions are barter-like transactions where individuals exchange personal data for free services. Yet those transactions appear to take place in an inefficient market hampered by steep information asymmetries, which are further aggravated by big data. Second, organizations are seldom prepared to share the wealth created by individuals' personal data with those individuals.
- **The ethics of analytics: Drawing the line**: where is the red line drawn when it comes to ethics of big data?
- **Chilling effect**: discouragement of the real exercise of natural/legal right by the threat of legal sanctions.

A solution for this would be to view the identifiability of data as a continuum as opposed to the current dichotomy. This means adopting a scaled approach, under which data that are only identifiable at great cost would remain within the legal framework, subject to only a subset of fair information principles.
And, according to FTC, as long as (1) a given dataset is not reasonably identifiable, (2) the company publicly commits to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form, the data will fall outside of the scope of the framework.

Other principles of privacy law:

- **Data minimization**: organizations are required to limit the collection of personal data to the minimum extent possible. Also they are required to delete data that isn't used anymore for where they were collected for and implement restrictive policies w.r.t. the retention of personal data in identifiable form.
- **Individual control and context**: consent (individual control) must be specific to the purpose (context). A problem is that users are made to feel that they have control so they will share information, but they don't have control at all. An additional problem is that consent-based processing tends to be regressive because individuals' expectations fall back on existing experiences.

Solutions to the legal framework:

- **Access, portability, and sharing the wealth**: organizations should share the wealth created by individuals' data with individuals in accessible format so the individuals can draw useful conclusions. This will unleash innovation and create a market for personal data apps. The Administration predicted that making user data available to the public would lead entrepreneurs to develop technologies like energy management systems and smartphone apps that can interpret and use such information.
- **Enhanced transparency**
- De-identification as protective measure rather than as solution.

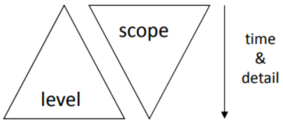
The extent of transparency and access will raise legal and business complexities.

- Organizations (particularly non-consumer facing ones) may argue that in many circumstances providing individual access to massive databases distributed across numerous servers and containing zettabytes of de-identified data is simply not practical.
- To avoid creation of a bigger privacy problem, direct online accessibility to data requires strong authentication as well as secure channeling, which leads to high costs.
- As the ecosystem for personal information expands, building layers of user-side apps over the existing structure will increase security risks of leakage and unauthorized use.
- Access to machine-readable data in usable format promotes data portability.

However, these issues can be diminished:

- If the data identification increases, so should level of access rights.
- Privacy and data require that an individual only gets access to his/her own personal data
- Enhancement of big data with interfaces for user interaction increases number of access points and elevates risk of security breach and data leakage.
- Portability isn't a concept of privacy law but derived from antitrust.
- It is hard for law/policy to keep up with uprising technologies.

Executive Information Systems (EIS) (DSS, BI&A):



Determining information requirements can be done by 4 generic strategies:

- Asking
- Deriving from existing IS
- Synthesizing from characteristics of the utilizing system
- Discovering from experimentation with an evolving IS

The 2 phases in the lifetime of an EIS are the **initial phase** and the **ongoing phase**.

<u>16 methods for determining information requirements:</u>

1. Discussing with executives
2. EIS planning meetings
3. Examinations of computer-generated info
4. Discussions with support personnel
5. Volunteered information
6. Examination of EIS of other organizations
7. Examinations of non-computer-generated info
8. **Critical success factors**:
9. Participation in strategic planning sessions
10. Strategic business objectives methods
11. Attendance at meetings
12. Information systems teams working in isolation
13. Examination of the strategic plan
14. Tracking executive activity
15. Software tracking of EIS usage
16. Formal change requests

CSF = limited number of areas where satisfactory results ensure successful competitive performance for individual/department/organization. They are the few areas where 'things must go right' for the business to show and for the goal of a manager to be achieved.

To measure these CSF you use Key Performance Indicators (KPI):

CSF	**KPI**
Image in financial markets	Price/earnings ratio
Technological reputation with customers	Orders/bid ratio
New market success	Change in market share
Company morale	Employee satisfaction score

Data quality has 6 dimensions:

COMPLETENESS
CONSISTENCY
UNIQUENESS
Data Quality Dimensions
ACCURACY
TIMELINESS
VALIDITY

Completeness: are all data sets/items records?
Consistency: can we match the data set across data stores?
Uniqueness: is there a single view of the data set?
Validity: does the data match the rules?
Accuracy: does the data reflect the data set?
Timeliness: can the data be used anytime?

Issues of low Data Warehousing depends on the level:

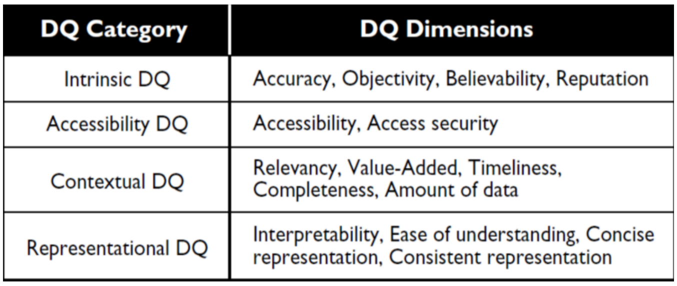
- At operational level: lowered customer/employee satisfaction which caused the increase in cost
- At strategic level: more difficult to set/execute strategy and to contribute to issues of data ownership
- At tactical level: poorer decision making which makes it more difficult to implement DW and to reengineer which causes an increase in the organizational mistrust.

> Strong et al. (1997): Data quality in context
Focus not only on *stored data* (**data custodians**), but also on *production* (**data producers**) and *utilization* (**data consumers**). Consumer in mind, focus on **CONSUMER**!

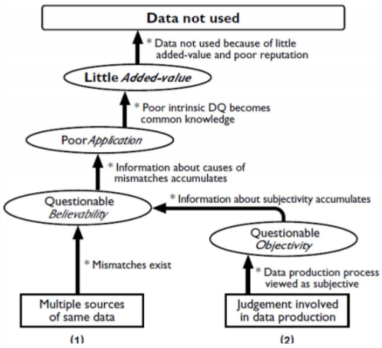
High quality data = data that is fit for use by data consumers

DQ problem = difficulty faced along ≥ 1 quality dimensions that renders completely or largely unfit data for use

DQ project = organizational actions taken to address DQ problems given recognition of poor DQ by the organization



DQ Category	DQ Dimensions
Intrinsic DQ	Accuracy, Objectivity, Believability, Reputation
Accessibility DQ	Accessibility, Access security
Contextual DQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational DQ	Interpretability, Ease of understanding, Concise representation, Consistent representation


Data not used
* Data not used because of little added-value and poor reputation
Little Added-value
* Poor intrinsic DQ becomes common knowledge
Poor Application
* Information about causes of mismatches accumulates
Questionable Believability
* Information about subjectivity accumulates
Questionable Objectivity
* Mismatches exist
* Data production process viewed as subjective
Multiple sources of same data
Judgement involved in data production
(1)
(2)

Intrinsic DQ pattern: mismatches among sources of the same data are a common cause of intrinsic DQ concerns.

Sub-pattern 1: as a reputation for poor-quality data becomes common knowledge, these data sources are viewed as having little added value for the organization, resulting in reduced use.

Sub-pattern 2: judgment or subjectivity in the data production process is another common cause (coded or interpreted data is considered to be of lower quality than raw uninterpreted data)

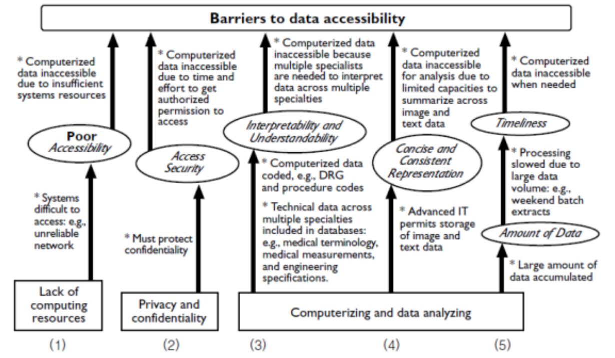
Solutions to sub-pattern 1 - the 2 different approaches: changing systems or changing production processes.

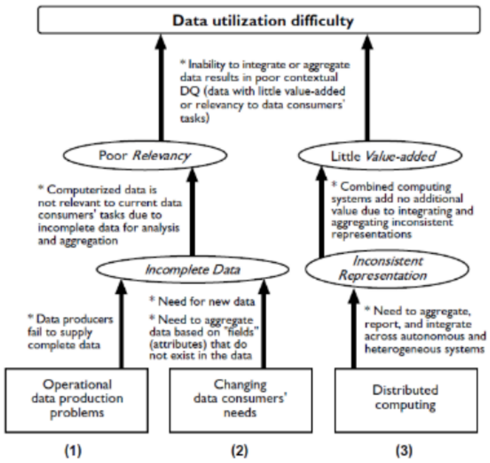
Accessible DQ pattern: data consumers perceive any barriers as accessibility problems. IS professionals must understand the difference between technichal accessibility and broad accessibility concerns. Once clarified, technologies such as DW can provide smaller amount of more relevant data, graphical interfaces can improve access.

Sub-pattern 1&2: concerns about technical accessibility. Simple but costly solution.

Sub-pattern 3&4: data-representation issues interpreted by data consumers as accessibility problems. More difficult to solve than 1,2.

Sub-pattern 5: data-volume issues interpreted as accessibility problems.


Barriers to data accessibility
* Computerized data inaccessible due to insufficient systems resources
Poor Accessibility
* Systems difficult to access: e.g., unreliable network
Lack of computing resources
* Computerized data inaccessible due to time and effort to get authorized permission to access
Access Security
* Must protect confidentiality
Privacy and confidentiality
* Computerized data inaccessible because multiple specialists are needed to interpret data across multiple specialties
Interpretability and Understandability
* Computerized data coded, e.g., DRG and procedure codes
* Technical data across multiple specialties included in databases: e.g., medical terminology, medical measurements, and engineering specifications.
* Computerized data inaccessible for analysis due to limited capacities to summarize across image and text data
Concise and Consistent Representation
* Advanced IT permits storage of image and text data
* Computerized data inaccessible when needed
Timeliness
* Processing slowed due to large data volume: e.g., weekend batch extracts
Amount of Data
* Large amount of data accumulated
Computerizing and data analyzing
(1)
(2)
(3)
(4)
(5)


Data utilization difficulty
* Inability to integrate or aggregate data results in poor contextual DQ (data with little value-added or relevancy to data consumers' tasks)
Poor Relevancy
Little Value-added
* Computerized data is not relevant to current data consumers' tasks due to incomplete data for analysis and aggregation
* Combined computing systems add no additional value due to integrating and aggregating inconsistent representations
Incomplete Data
Inconsistent Representation
* Data producers fail to supply complete data
* Need for new data
* Need to aggregate data based on "fields" (attributes) that do not exist in the data
* Need to aggregate, report, and integrate across autonomous and heterogeneous systems
Operational data production problems
Changing data consumers' needs
Distributed computing
(1)
(2)
(3)

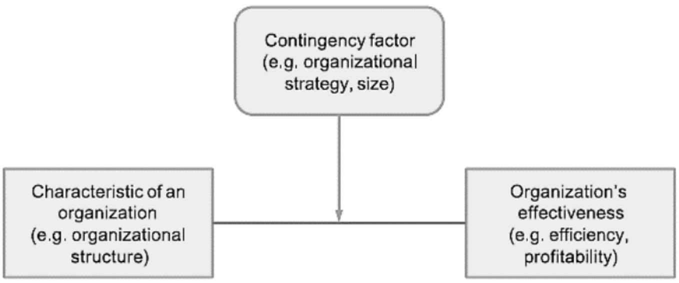
Contextual DQ pattern:
Sub-pattern 1: Addresses incomplete data due to operational problems
Sub-pattern 2: incomplete by design
Sub-pattern 3: problems by integrating data through distributed systems.

Solution: providing HQ data along dimensions of value designing flexible system with easily aggregated/manipulated data.
Alternative: constant maintenance of data/systems to meet changing data requirements.

Result of the research: to confirm the importance of quality categories and dimensions in previous researches. Representational DQ dimensions are underlying causes of accessibility DQ problem patterns.

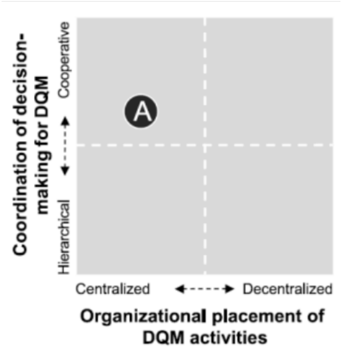
To solve organizational DQ problems, IS professionals must attend to entire range of concerns of data consumers. The 3 patterns for how intrinsic, accessibility and contextual DQ problems develop in organizations provide an empirical basis for studying organizational choices and actions about DQ improvement. Studies of DQ solutions could use the DQ problem patterns identified in this research as solution objectives.

> Weber et al. (2009): One size does not fit all – a contingency approach to data governance
Research aims at starting a scientific discussion on data governance by transferring concepts from IT governance and organizational theory to the previously largely ignored field of data governance. Outlines a data governance model that consists of 3 components: DQ roles, decision areas and responsibilities, which form a responsibility assignment matrix.

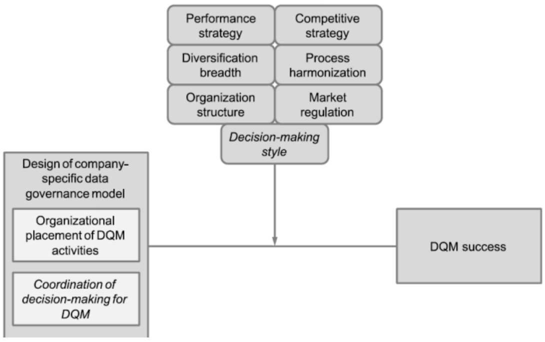

Contingency factor (e.g, organizational strategy, size)
Characteristic of an organization (e.g. organizational structure)
Organization's effectiveness (e.g. efficiency, profitability)

Data Quality Management (DQM) focuses on planning/provisioning/organization/usage/disposal of HQ data. Data governance includes parts of IT governance. Article suggests that **contingencies affect data governance**, and that a data configuration Is specific to a given company. Relationship between characteristic of organization and its effectiveness is determined by contingencies. IT governance research has analyzed 2 domains:
1. organizational structuring of IT activities together with placement of decision making authority
2. effect of multiple contingencies on contribution of IT governance to corporate performance.

Data governance model consists of DQM areas and main activities/roles/responsibilities. 3 components are arranged in a matrix. Rows of matrix are key decision areas and **main activities**, columns indicate **roles**. Cells contain the **responsibilities**, specify degrees of authority between roles/decision areas.

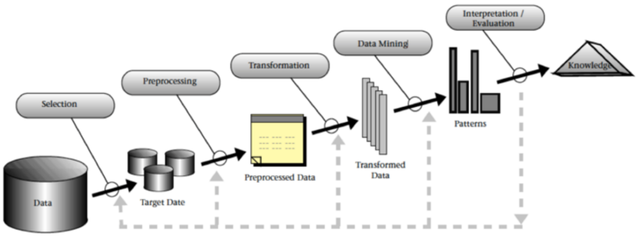

Coordination of decision-making for DQM
Cooperative
Hierarchical
A
Centralized
Decentralized
Organizational placement of DQM activities

Data governance model addresses DQM on 3 horizontal layers: strategy, organization, and information systems.
Data governance contingency model is a **moderation model** with contingencies as co-variation effects. Represented by the design of 2 design parameters **organizational placement of DQM activities** and **coordination of decision-making for DQ** (picture left). This means company A is centralized and cooperative.
Contingency factors determine the fit between design of the data governance model and the success of DQM within organization which means the effectiveness of an organization. The elements of the **data governance contingency model** are shown on the right.



WEEK 4

Data mining = the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques.
OR: the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules.



The view of a minor when considering data types.

- Nondependency-oriented data:	Numeric:	continuous
	Ordinal:	ordered discrete
	Categorical:	unordered discrete
	Binary:	a special case of ordinal or categorical
- Dependency-oriented data ↓:	Temporal (e.g. Time Series)	
Often require different	Spatial (e.g. Maps, Graphics, etc.)	
and often more complex	Sequences (e.g. Genome)	
analytic methods	Graphs (e.g. Networks, Social Networks, etc.)	

Descriptive data mining is learning/understanding the data.
Predictive data mining is modelling building in order to predict unknown values.

Some typical applications:

Customer relationship management:

- Target marketing
  - ♦ Problem: use list of prospects for direct e-mail campaign.
  - ♦ Solution: use data mining to identify most promising customers, based on past purchase behavior.
  - ♦ Example: bol.com, booking.com

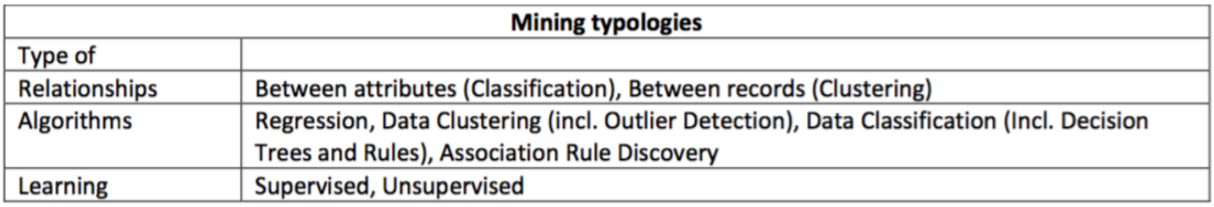
- Attrition prediction / churn analysis
  - ♦ Problem: prevent loss of customers and avoid adding churn-prone customers.
  - ♦ Solution: use data mining to identify typical patterns of usage of likely-to defect / likely-to-churn customers.
  - ♦ Example: T-Mobile, TELE2

E-commerce / mobile commerce:

- Collaborative filtering
  - ♦ Problem: how to use information from other users to make inference about a particular user?
  - ♦ Solution: use data mining to find users with similar tastes.

- Example: amazon, Netflix

- Browse-click-conversation
  - Problem: a large number of people browsing a website, but only a few of them actually make clicks/purchases.
  - Solution: through clicks, classify customers, adjust website/design features to increase conversation rate.
  - Example: Bijenkorf, Wehkamp



Mining typologies	
Type of	
Relationships	Between attributes (Classification), Between records (Clustering)
Algorithms	Regression, Data Clustering (incl. Outlier Detection), Data Classification (Incl. Decision Trees and Rules), Association Rule Discovery
Learning	Supervised, Unsupervised

Def. **Supervised learning** = ML task of inferring a function from labeled training data, e.g. Classification

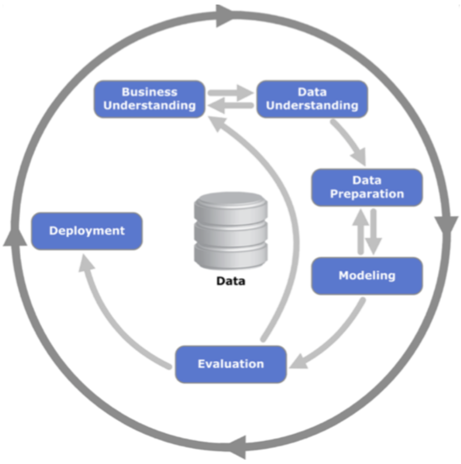
Def. **Unsupervised learning** = ML algorithm used to draw inferences from datasets consisting of input data without labeled responses, e.g. Clustering.

Def. **Regression** = A function that maps a data item to a prediction variable

Def. **Clustering** = Given a set of data points, each having a set of attributes, and a similarity measure among them, finds clusters such that data points in one cluster are more similar to each other and data points in different clusters are less similar to each other.

Def. **Classification** = Given a collection of records (training set) each record contains a set of attributes, one of the attributes denotes the class. Find a model for class attribute as a function of the values of other attributes. The goal is to assign previously unseen records to a class as accurately as possible.

Data mining as DSS:
Business Understanding: Understand project objectives and data mining problem identification
Data Understanding: Capture, understand, explore your data for quality issues
Data Preparation: Clean data, merge data, derive attributes etc.
Modeling: Select the data mining techniques, build the model
Evaluation: Evaluate the results and approved models Deployment: Put models into practice, monitoring and maintenance



Roles in data mining projects are:
Data Scientist: Doing the actual modelling.
Visual Artist: Info graphs, visualizations, visual representation.
Business Collaborator: Translating from business to the execution and interpretation and storytelling
Manager: Identifying the potentials, evaluating proposals for execution, interfacing with all the stakeholders

Data mining as core:
1. Recommender systems are systems for recommending items (e.g., books, movies, CDs, web pages, newsgroup messages) to users based on examples of their preferences → Content-based, Collaborative Filtering, Hybrid.

2. Collaborative filtering: Maintain a database of many rating of the users of a variety of items. For a given user, find other similar users whose ratings strongly correlate with the current user. Recommend items rated highly by these similar users, but not rated by the current user. Almost all commercial recommenders use this approach.

What can go wrong?
- Problem formulation
  - Need to understand the business well
- Inappropriate use of methods
  - Lack of sufficient and high-quality data
  - Computational issues
- Evaluation

  - Need domain experts throughout the process to provide indispensable input and validate results
- Inability to act upon pattern because of political or ethical reasons
  - Securities Trading models
  - Data mining in clinical evaluation
  - Privacy (Insurance & credit)
  - Admission interviews

> Fayyad (1996): From data mining to knowledge discovery in databases

Data mining and **knowledge discovery in databases** (KDD) are related both to each other and to ML, statistics and DB. For these apps, this form of manual probing of a dataset is slow, expensive and subjective. KDD is an attempt to address data overload. KDD is overall process of discovering useful knowledge from data. Data mining = particular step in this process. Main KDD application areas:

- **Marketing**: primary app is DB marketing systems, analyzes customer DB to identify different groups and forecast their behavior.
- **Investment**: when systems uses expert systems, neural nets, and genetic algorithms to manage portfolios.
- **Fraud detection**: watching over millions accounts to identify suspicious financial transactions
- **Manufacturing**: applied by 3 European airlines to diagnose and predict problems with Boeing 737.
- **Telecommunication**: using novel framework for locating frequently occurring alarm episodes from the alarm stream and predicting them as rules.
- **Data cleaning**: identification of duplicate welfare claims.

Def. **Model representation** = Language used to describe discoverable patterns. Increased representational power for models increases the danger of overfitting the training data, resulting reduced prediction accuracy on unseen data.

Def. **Model-evaluation criteria** = Quantitative statements (or fit functions) of how well a particular pattern (a model and its parameters) meets the goals of the KDD process.

How is KDD different from pattern recognition or ML? Answer: These fields provide some of the data-mining methods that are used in the data-mining step of the KDD process. KDD focuses on the overall process of knowledge discovery from data, including how the data are stored/accessed, how algorithms can be scaled to massive data sets ultimate and still run efficiently and how results can be interpreted and visualized. A related field evolving from databases is DW, which refers to the popular business trend of collecting and cleaning transactional data to make them available for online analysis and decision support. Data warehousing helps set the stage for KDD by data cleaning and data access. A pattern is knowledge if it exceeds some interestingness threshold. The KDD process can involve significant iteration and can contain loops between any 2 steps.

> Jones, M. T. (2013): Recommender systems, Part 1: Introduction to approaches and algorithms
Goal of the article is to explains the ideas behind recommender systems and introduces the algorithms that empower them.

Some well-known examples:
LinkedIn uses specialized collaborative-filtering and Apache Hadoop
Amazon uses item-to-item collaborative filtering
Hulu uses item-based collaborative filtering and Hadoop
Netflix uses Cinemach. Built an ensemble of 107 recommendation algorithms that resulted in a single prediction.

Most recommender systems take either of two approaches: **collaborative filtering** or **content-based filtering**. Other approaches (such as **hybrid approaches**) also exist.

Collaborative filtering is based on a model of prior user behavior. Can be constructed from a single user's behavior or from the behavior of other users who have similar traits.

Content-based filtering constructs a recommendation on the basis of a user's behavior. E.g. historical browsing information.

Hybrid approaches combine collaborative and content-based filtering. Hybrid approach could also be used to address collaborative filtering that starts with sparse data – known as *cold start* – by enabling the results to be weighted initially toward content-based filtering, then shifting to collaborative filtering as the available user data set matures.

<u>Algorithms that recommender systems use</u>:

- **Pearson correlation**: calculating similarity between users and their attributes (e.g. articles read). Measures the linear dependence between two variables (users) as function of their attributes. Doesn't calculate this over entire population of users. Population must be filtered down to *neighborhoods* based on a higher-level similarity metric. Algorithm for <u>collaborative filtering</u>.
- **Clustering algorithms**: unsupervised learning that can find structure in a set of seemingly random (unlabeled) data. Identify similarities among items (e.g. blog readers), by calculating their distance from other items in a *feature space*.
  - ♦ *K-means algorithm*: partitions items into *k* clusters. Initially, items are randomly placed into clusters. Then a *centroid* or *center* is calculated for each cluster as function of its members. Then each item's distance from the centroids is checked. If an item is closer to another cluster, it's moved to that cluster. This keeps repeating until no movements can be done, and the algorithm ends.
  - ♦ *Euclidean algorithm*: to treat each item as a multidimensional vector and calculate the distance
  - ♦ *Adaptive Resonance Theory* (ART) *family*
  - ♦ *Fuzzy C-Means*
  - ♦ *Expectation-Maximization*
- **Bayesian Belief Nets**: directed acyclic graph, with arcs representing associated probabilities among variables.
- **Markov chains**: similar to BBN but treat recommendation problem as sequential optimization instead of prediction.
- **Rocchio classification**: exploits feedback of item relevance to improve recommendation accuracy.

Some challenges:

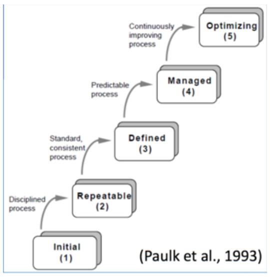
- Some users do not exhibit the predicted behavior
- Users can exploit a recommender system to favor one product over another
- Scalability: once the data grows (particularly on-line), it gets harder to analyze.
- Privacy-protection considerations

WEEK 5

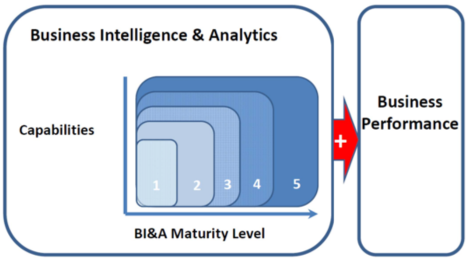
The **maturity** of a BI&A object is, is how developed that object (person, thing, IS) is and the differences between objects. Can be modeled. More mature is not always better!

Models:

- In IT: Capability Maturity Model (CMM) for software
- Different levels of maturity
- Multidimensional maturity
- Alignment

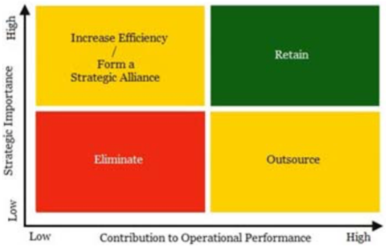


(Paulk et al., 1993)



In-house is when something is made by the organization itself. Out-source is when the organization spent money on some aspects.

Outsourcing matrix: ---------------------------------->

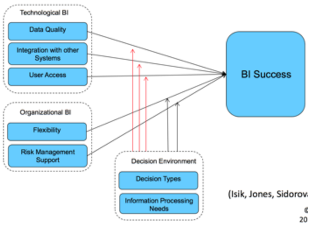

High
Strategic Importance
Low
Increase Efficiency / Form a Strategic Alliance
Retain
Eliminate
Outsource
Low
Contribution to Operational Performance
High

Two types of organizations:

- Analytically challenged: quick and easy way to access analytic capability/skills. Do not worry about IP, like to collaborate in this area.
- Analytically superior: see analytics as important 'core competence' leading to competitive advantage. Will be more hesitant to outsource analytics.

(not) outsourcing will impact organization:

- New organizational units and restructuring business processes occur
- Job satisfaction changes
- Job can become stressful/anxious
- Activities and performance of a manager changes.



(Isik, Jones, Sidorov
20

Relativity of success depends on who/when you ask, the success dimensions and ES implementation.

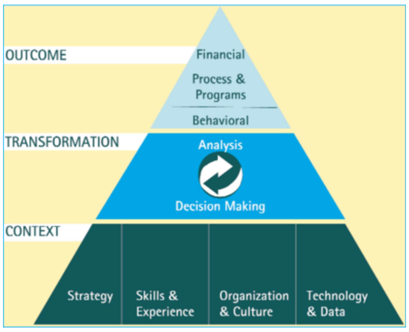
→ Success dimensions: Management/Project/User/Correspondence/System success
→ Phases of ES implementations: Chartering/Project/Shakedown/Onward/Upward phase

A **data-driven culture** is the capability to aggregate, analyze, and use data to make informed decisions that lead to action and generate real business value. You'll need technical/human capabilities. This will lead to data-driven decision making.

Context: fundamentals of success in this process. Strategic/skill/organizational & cultural/technological/data related factors must be present for an analytical effort to succeed. Constantly refined & affected by other elements

Transformation: where data is actually analyzed and used to support a business decision

Outcomes: the result of transformation part. Example outcomes: behaviors, processes&programs, financial conditions.

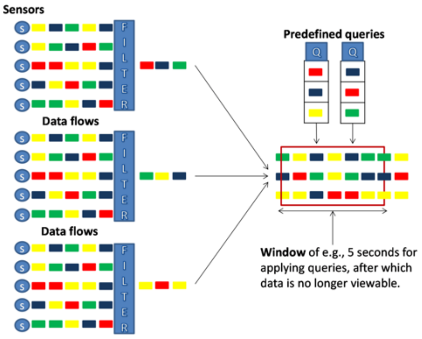


Fast data (FD) = the ability to gain insights from (near) real-time data streams and derive value from these insights. It is 'Big Data in Extrema'. Caused by the increasing number of sensors & data sources. Big Data V's are: Volume, Velocity, Variety, Veracity, Variability, Value. FD is important because it is a source of value for organization and plays a role in rapidly changing environment and increasing customers expectations.

To achieve successful FD, some changed are needed built on 4 pillars: Technology, Strategy, Culture, Skills & Experience.

Technology:

- Analyze and process data as events
- Process incoming fast data directly by using 2 techniques:
  - Splitting
  - Filtering
- Maintain 2 separate DB:
  - Historical data
  - Fast data
- Combine these events for recognition


Sensors
S
FILTER
Data flows
Data flows
Predefined queries
Q
Q
Window of e.g., 5 seconds for applying queries, after which data is no longer viewable.

Strategy:

- Ensure a clear strategy for required data
- Define what ‘real-time’ means
- Determine the required length for ‘windows’ of incoming stream data
- Translate org. strategy into business rules
- Be able to adapt those every moment

Culture:
Data-driven *and* agile
So:

- Ensure trust in available data
- Let employees practice and learn with FD decisions
- Give employees autonomy to respond based on FD
- Be prepared for rapid changes based on FD

Skills and Experience:
Knowledge of & experience with:

- Technology: systems and software
- Algorithms and pattern meaning
- The data
- The organization and -strategy
- Communication: be able to convey the data and patterns found

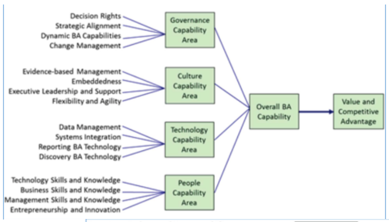
> Fogarty (2014): Should you outsource analytics?
Companies with superior analytic capabilities will approach outsourcing differently than companies that are analytically challenged. Success of business process organizations (BPO) partnerships depends on factors like geographic distance between on/offshore hubs, existence of suitable infrastructure and connectivity, acceptable language/technical skills, and proper contingency planning.

Companies like **analytically challenged** are usually happy to outsource their analytics requirements, and companies like **analytically superior** want to expand their internal analytic capabilities.

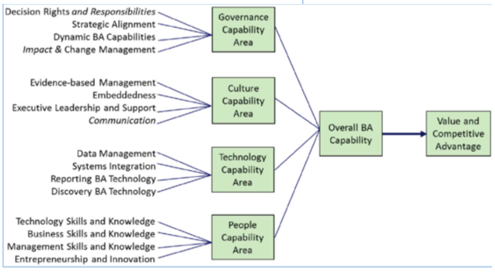
> Cosic et al. (2015): A business analytics capability framework

Paper develops a holistic, theoretically-grounded and practically relevant business analytics capability framework (BACF) that specifies/defines/ranks the capabilities that constitute an organizational BA initiative. The BACF was developed in 2 phases.

Phase 1: conceptual framework was developed based on the Resource-Based View theory of the firm and thematic content analysis of the BA literature. Result: -->


Decision Rights
Strategic Alignment
Dynamic BA Capabilities
Change Management
Governance Capability Area
Evidence-based Management
Embeddedness
Executive Leadership and Support
Flexibility and Agility
Culture Capability Area
Data Management
Systems Integration
Reporting BA Technology
Discovery BA Technology
Technology Capability Area
Technology Skills and Knowledge
Business Skills and Knowledge
Management Skills and Knowledge
Entrepreneurship and Innovation
People Capability Area
Overall BA Capability
Value and Competitive Advantage

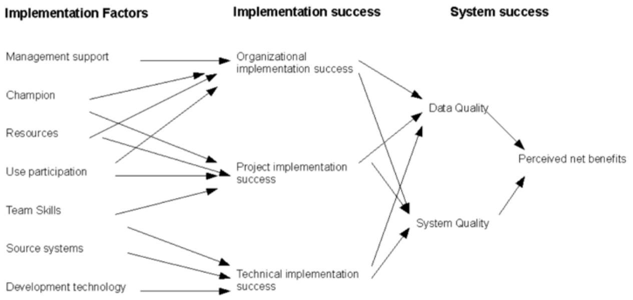
Phase 2: the CF was further developed and refined using 3 round Delphi study involving 16 BA experts. By ranking of capabilities based on importance. Result: ------>


Decision Rights *and Responsibilities*
Strategic Alignment
Dynamic BA Capabilities
Impact & Change Management
Governance Capability Area
Evidence-based Management
Embeddedness
Executive Leadership and Support
Communication
Culture Capability Area
Data Management
Systems Integration
Reporting BA Technology
Discovery BA Technology
Technology Capability Area
Technology Skills and Knowledge
Business Skills and Knowledge
Management Skills and Knowledge
Entrepreneurship and Innovation
People Capability Area
Overall BA Capability
Value and Competitive Advantage

Direct support for **technology** and **culture** capability areas. Together with **governance** and **people** they're common themes in IS literatures.

> Wixom et al. (2001): An empirical investigation of the factors affecting DW success

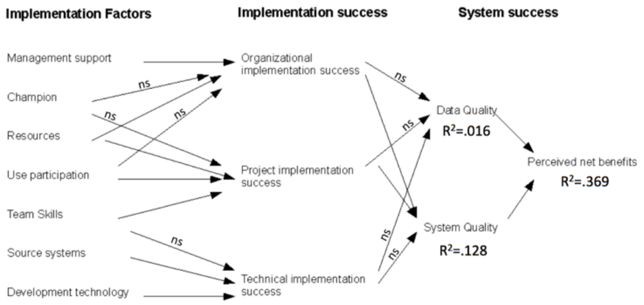
Research model according to critical success factors:



Implementation success = when a project team has the organization convinced to accept DW and completed it according to the plan and overcome technical obstacles that arose. Affects overall success of system.

Using DW literature, initial survey and 3 interviews, 3 facets of DW implementation success were identified:

1. Success with organizational issues
2. Success with project issues
3. Success with technical issues



↑ Research model was tested using a structured modeling technique, PLS, that can be used for highly complex predictive models. Strengths: its ability to handle formative constructs and its small sample size requirements.

A DW with good data/system quality improves the way data is provided to decision-support apps and decision makers.

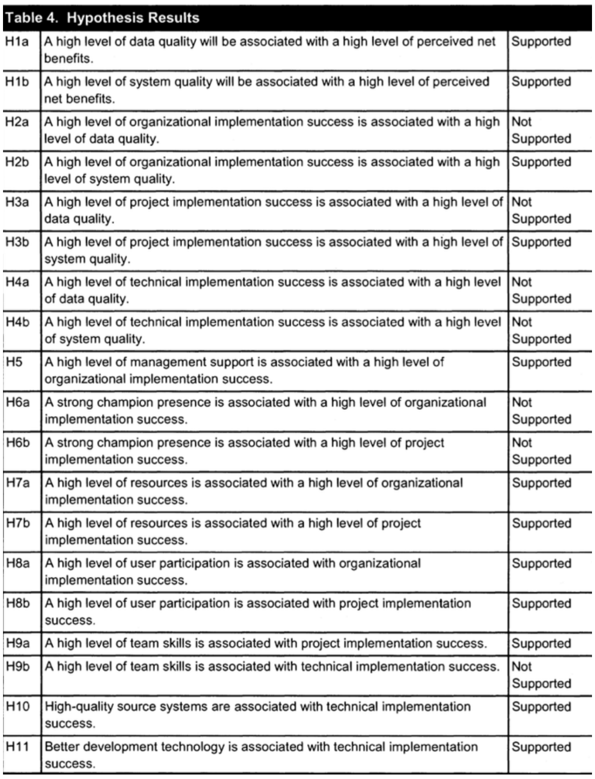
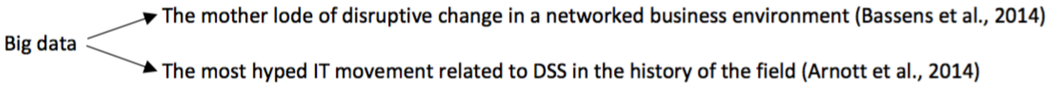


Table 4. Hypothesis Results

H1a	A high level of data quality will be associated with a high level of perceived net benefits.	Supported
H1b	A high level of system quality will be associated with a high level of perceived net benefits.	Supported
H2a	A high level of organizational implementation success is associated with a high level of data quality.	Not Supported
H2b	A high level of organizational implementation success is associated with a high level of system quality.	Supported
H3a	A high level of project implementation success is associated with a high level of data quality.	Not Supported
H3b	A high level of project implementation success is associated with a high level of system quality.	Supported
H4a	A high level of technical implementation success is associated with a high level of data quality.	Not Supported
H4b	A high level of technical implementation success is associated with a high level of system quality.	Not Supported
H5	A high level of management support is associated with a high level of organizational implementation success.	Supported
H6a	A strong champion presence is associated with a high level of organizational implementation success.	Not Supported
H6b	A strong champion presence is associated with a high level of project implementation success.	Not Supported
H7a	A high level of resources is associated with a high level of organizational implementation success.	Supported
H7b	A high level of resources is associated with a high level of project implementation success.	Supported
H8a	A high level of user participation is associated with organizational implementation success.	Supported
H8b	A high level of user participation is associated with project implementation success.	Supported
H9a	A high level of team skills is associated with project implementation success.	Supported
H9b	A high level of team skills is associated with technical implementation success.	Not Supported
H10	High-quality source systems are associated with technical implementation success.	Supported
H11	Better development technology is associated with technical implementation success.	Supported

WEEK 6

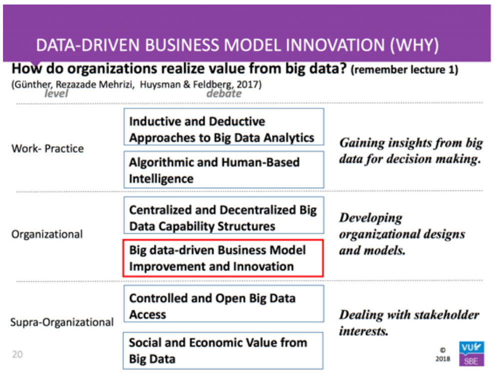
Discussion question: How do organization realize value from Big Data? 2 possible views:


Big data
The mother lode of disruptive change in a networked business environment (Bassens et al., 2014)
The most hyped IT movement related to DSS in the history of the field (Arnott et al., 2014)

Big Data has 2 specific features:
1. **Portability**: possibility of transferring/remotely accessing digitized data from one context of application to be used in other contexts.
2. **Interconnectivity**: the possibility to synthesize various data sources.

Our future 6 D's:
- Digitized
- Demonetized
- Democratized
- Delocalized
- Deceptive
- Disruptive
- **Datafied** → added.

But also:
Internet as a Service
Cloud computing
High performance computing
Crowd sourcing
Powerful software (open source!)
High-performance analytics



DATA-DRIVEN BUSINESS MODEL INNOVATION (WHY)

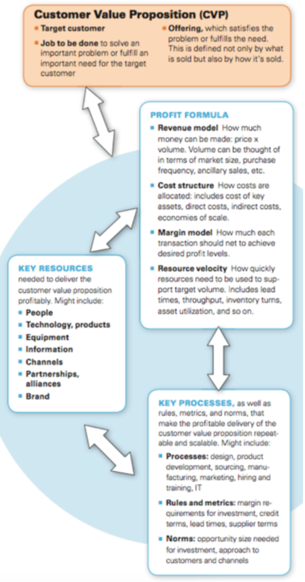
How do organizations realize value from big data? **(remember lecture 1)**
(Günther, Rezazade Mehrizi, Huysman & Feldberg, 2017)

level	*debate*	
Work- Practice	**Inductive and Deductive Approaches to Big Data Analytics** **Algorithmic and Human-Based Intelligence**	***Gaining insights from big data for decision making.***
Organizational	**Centralized and Decentralized Big Data Capability Structures** **Big data-driven Business Model Improvement and Innovation**	***Developing organizational designs and models.***
Supra-Organizational	**Controlled and Open Big Data Access** **Social and Economic Value from Big Data**	***Dealing with stakeholder interests.***

20 © 2018

Def. Business model: A business model articulates how an organization ***creates*** value for its customers and ***appropriates*** value from its markets.

Business model framework: →
1. **Customer value proposition** includes important problems/needs satisfied by offering for the target customers
2. The **profit formula** defines how the company creates value for itself and consists of the revenue model, cost structure, margin model and resource velocity
3. **Key resources** are those needed to deliver value proposition
4. **Key processes**



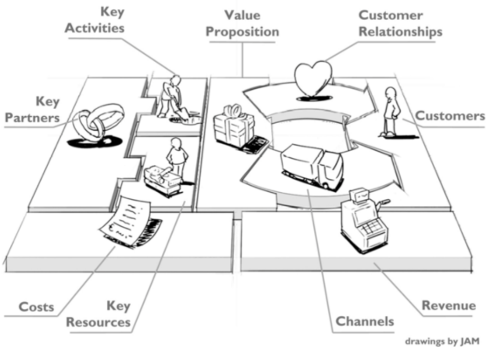
Customer Value Proposition (CVP)
- **Target customer**
- **Job to be done** to solve an important problem or fulfill an important need for the target customer
- **Offering,** which satisfies the problem or fulfills the need. This is defined not only by what is sold but also by how it's sold.

PROFIT FORMULA
- **Revenue model** How much money can be made: price x volume. Volume can be thought of in terms of market size, purchase frequency, ancillary sales, etc.
- **Cost structure** How costs are allocated: includes cost of key assets, direct costs, indirect costs, economies of scale.
- **Margin model** How much each transaction should net to achieve desired profit levels.
- **Resource velocity** How quickly resources need to be used to support target volume. Includes lead times, throughput, inventory turns, asset utilization, and so on.

KEY RESOURCES needed to deliver the customer value proposition profitably. Might include:
- **People**
- **Technology, products**
- **Equipment**
- **Information**
- **Channels**
- **Partnerships, alliances**
- **Brand**

KEY PROCESSES, as well as rules, metrics, and norms, that make the profitable delivery of the customer value proposition repeatable and scalable. Might include:
- **Processes:** design, product development, sourcing, manufacturing, marketing, hiring and training, IT
- **Rules and metrics:** margin requirements for investment, credit terms, lead times, supplier terms
- **Norms:** opportunity size needed for investment, approach to customers and channels

Business model canvas: →



Building blocks:
- **Customer segment:** different groups of people or organizations an organization aims to reach and serve.
- **Value proposition**: value created for customers through offering. Describes products/services that create value for specific customer segment.
- **Channels**: the ways organizations communicate/reaches its customer segments to deliver a value proposition.
- **Customer relationships**: relationship types a company establishes with specific customer segments.
- **Revenue streams**: the cash a company generates from each customer segment (earnings=revenue-/-costs)
- **Key resources**: most important assets required to make a business model work.
- **Key activities**: most important things an organization should do to make the business model work.
- **Key partnerships**: network of suppliers/partners/stakeholders that make the business model work.
- **Cost structure**: all costs incurred to operate a business model.

Search for ideas to innovate business models:
- **Competency based**: how can we build on the capabilities and assets that already make us distinctive to enter new business and markets?
- **Customer focused**: what does a close study of customers behavior tell us about their tacit, unmet needs?
- **Changes in business environment**: if we follow "megatrends" or other shifts in their logical conclusion, what future business opportunities will become clear?
- **A 4th approach complements the existing, focuses on opportunities generated by the explosion in digital info and tools.**

How can we create value for customers using data and analytic tools we own or could have access to?

5 patterns (Parmar et al.):

Pattern 1: Augmented products to generate data → Sensors, digital twins

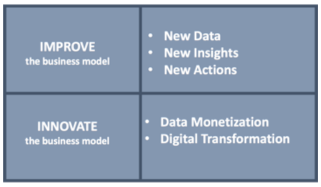
Pattern 2: Digitizing assets → Digital models (processes/products) Management of Digitization (scoping/switching up)

Pattern 3: Combining data within/across industries → Enhanced data integration, coordinate data across/within industries in new ways. Search for ideas to innovate business models

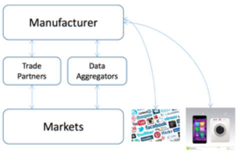
Pattern 4: Trading data → Combine disparate datasets (TomTom sources)

Pattern 5: Codifying a distinctive service capability → Automate business processes (Air crew scheduling KLM)

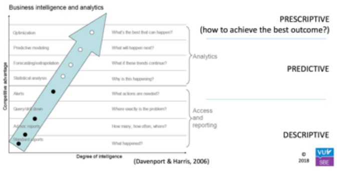
The use of big data to craft strategy & business models



New data



New insights



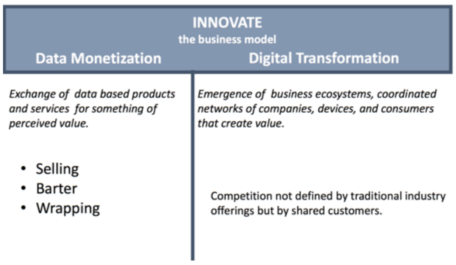
(Davenport & Harris, 2006)

New actions:

New data + new insights
=
New actions

Prescriptive

The use of big data to craft strategy and business models:



Exchange of data based products and services for something of perceived value.

- Selling
- Barter
- Wrapping

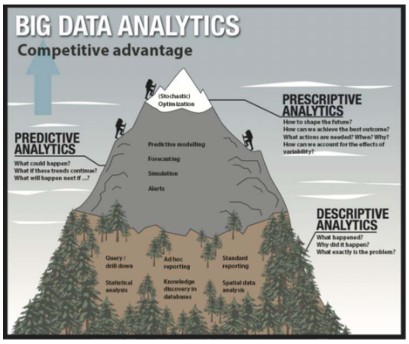
Emergence of business ecosystems, coordinated networks of companies, devices, and consumers that create value.

Competition not defined by traditional industry offerings but by shared customers.

Examples: 10 top apps for eating healthy, health apps, myfitnesspal, runkeeper
Selling the withdrawal behavior, giving discounts when customers share their data (**barter**).

Challenges/Risks of big data:

- Data privacy & ethics
- Data obsession (dictatorship of data)
- Data quality
- Skills: data scientists
- Energy
- Security



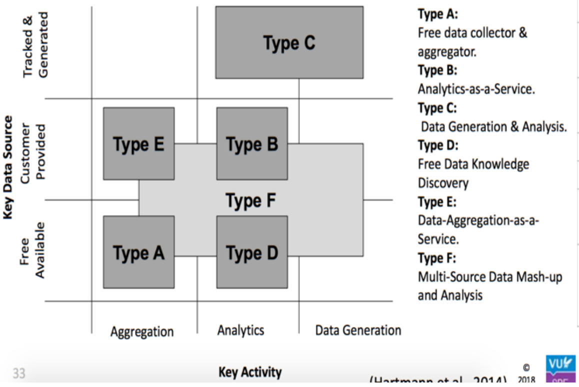
Cognitive computing (CC) describes technology platforms that, broadly speaking, are based on the scientific disciplines of Artificial Intelligence and Signal Processing. These platforms encompass machine learning, reasoning, natural language processing, speech and vision, human- computer interaction, dialog and narrative generation and more.

The goal of cognitive computing is to simulate human thought processes in a computerized model. Using self-learning algorithms that use data mining, pattern recognition and natural language processing, the computer can mimic the way the human brain works.

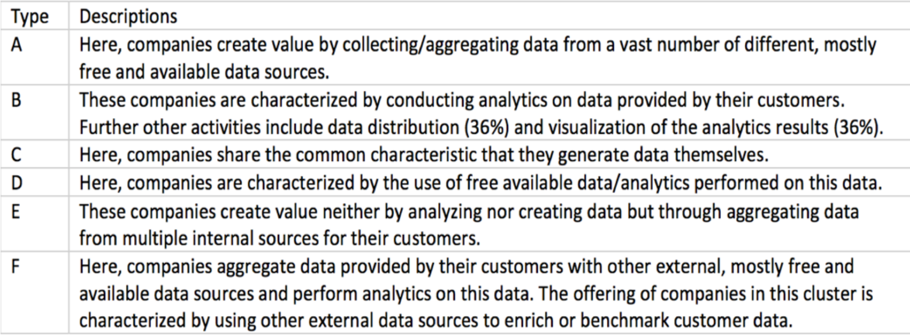
In general, the term cognitive computing has been used to refer to new hardware and/or software that mimics the functioning of the human brain and helps to improve human decision-making.
In this sense, CC is a new type of computing with the goal of more accurate models of how the human brain/mind senses, reasons, and responds to stimulus. CC applications link data analysis and adaptive page displays (AUI) to adjust content for a particular type of audience. As such, CC hardware and applications strive to be more affective and more influential by design.

IBM describes the components used to develop, and behaviors resulting from “systems that learn at scale, reason with purpose and interact with humans naturally.” According to them, while sharing many attributes with the field of artificial intelligence, it differentiates itself via the complex interplay of disparate components, each of which comprise their own individual mature disciplines.

Business model types:



Type A:
Free data collector & aggregator.
Type B:
Analytics-as-a-Service.
Type C:
Data Generation & Analysis.
Type D:
Free Data Knowledge Discovery
Type E:
Data-Aggregation-as-a-Service.
Type F:
Multi-Source Data Mash-up and Analysis



Type	Descriptions
A	Here, companies create value by collecting/aggregating data from a vast number of different, mostly free and available data sources.
B	These companies are characterized by conducting analytics on data provided by their customers. Further other activities include data distribution (36%) and visualization of the analytics results (36%).
C	Here, companies share the common characteristic that they generate data themselves.
D	Here, companies are characterized by the use of free available data/analytics performed on this data.
E	These companies create value neither by analyzing nor creating data but through aggregating data from multiple internal sources for their customers.
F	Here, companies aggregate data provided by their customers with other external, mostly free and available data sources and perform analytics on this data. The offering of companies in this cluster is characterized by using other external data sources to enrich or benchmark customer data.

> Woerner et al. (2015): Big data: extending the business strategy toolbox
This article addresses how Big Data challenges properties and the time horizons of strategy making. Also, it suggests that studying how companies use data to improve company choices/operations help creating actionable practices that will help companies overcome the limitations of big data.

BD helps improvements to business models between industries. The following 3 examples suggest this too:
1. **New data**: used to be difficult for manufactures to feed data into internal product development. When they started using sensors in the product this is no longer the problem.
2. **New insights**: big data analytics can be used to identify outliers based on data inconsistencies
3. **New action**: as companies become well-armed with big data and proficient at making insights based on that data, they will act differently (faster/more wisely).

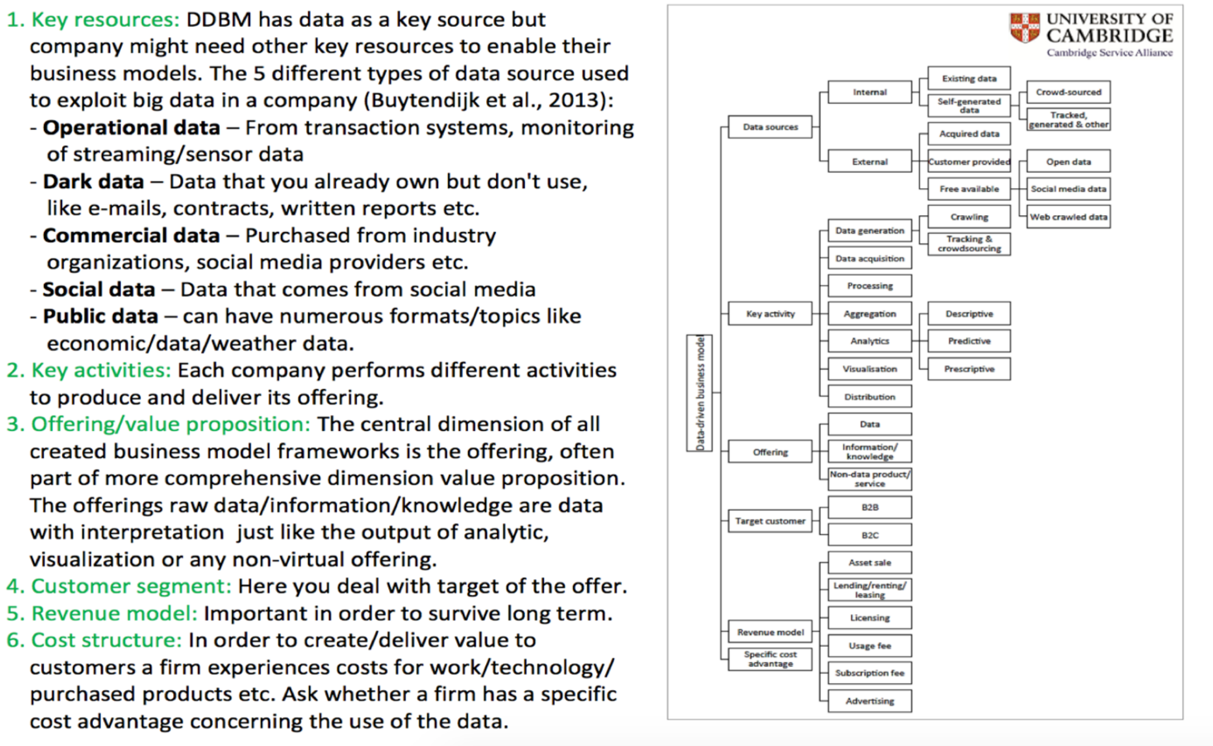
BD may be used by 2 different approaches:
1. **Data monetization** = the act of exchanging information-based products/services for legal tender or something of apparent equivalent value. Done by selling, bartering, wrapping.

Selling = when companies receive money in exchange for information offerings.
Bartering = when companies choose to trade information in return for new tools, services or special deals.
Wrapping = wrapping information around other core products and services.

2. **Digital transformation**: doesn't just increase new/different types of data, may also shape company boundaries so it becomes more difficult to tell where a partner organization commitment begins/ends. BD itself doesn't create strategic issues, only when generating business value from it.

> Hartmann et al. (2014): Big Data for Big Business? Catalog of data-driven Business Models, used by start-up firms.
The term **data-driven business model** (DDBM) is commonly used by practitioners and has 3 implications:
1. It is not limited to companies directing analytics, also includes companies that are aggregating/collecting data.
2. A company may sell its data/info and also other product/service that relies on data as a key resource.
3. It is obvious that company uses data in some way to conduct business, the size doesn't matter at all.

DDBM framework consists of 6 dimensions common to most of the business model framework.



1. Key resources: DDBM has data as a key source but company might need other key resources to enable their business models. The 5 different types of data source used to exploit big data in a company (Buytendijk et al., 2013):
   - **Operational data** – From transaction systems, monitoring of streaming/sensor data
   - **Dark data** – Data that you already own but don't use, like e-mails, contracts, written reports etc.
   - **Commercial data** – Purchased from industry organizations, social media providers etc.
   - **Social data** – Data that comes from social media
   - **Public data** – can have numerous formats/topics like economic/data/weather data.
2. Key activities: Each company performs different activities to produce and deliver its offering.
3. Offering/value proposition: The central dimension of all created business model frameworks is the offering, often part of more comprehensive dimension value proposition. The offerings raw data/information/knowledge are data with interpretation just like the output of analytic, visualization or any non-virtual offering.
4. Customer segment: Here you deal with target of the offer.
5. Revenue model: Important in order to survive long term.
6. Cost structure: In order to create/deliver value to customers a firm experiences costs for work/technology/ purchased products etc. Ask whether a firm has a specific cost advantage concerning the use of the data.

> Parmar et al. (2014): The new patterns of innovation
The search for new business ideas and models may cause failures in most corporation. Management scholars have considered reasons for this failure. Important question to consider: "How can we create value for customers using data and analytic tools we own or could have access to?" Out of the advances in the IT came the hunt for new business values in 5 distinct patterns. They form the basis of the framework and by examining them methodically, managers can conceive ideas for new businesses.

Pattern 1: Augmenting products to generate data – Using data that physical objects now generate to improve a product or service or create new business value. Example: Rolls Royce's engine health management capability. Allowed them to identify airplane engine problem at early stage.

Pattern 2: Digitizing assets – adapting to the more and more digitizing world. Slashes distributing costs and makes it able to move physical inventory efficiently or secure favorable store locations less critical. Offering customers more choices/tailored service will become more and more important.

Pattern 3: Combining data within and across industries – Example: IBM developed a network of sensors in home that monitor not only conditions like temperature/water level etc. but also what constitutes "normal" behavior patterns – e.g. regular cooking times. When abnormal, it will call a relative or friend, so they can check whether it's all okay.

Pattern 4: Trading/Exchanging data – Example: the partnership between Vodafone and TomTom. Vodafone knows when and where their users are driving, and TomTom uses this to pinpoint traffic jams.

Pattern 5: Codifying a distinctive service capability – Example: a major UK retailer has developed an efficient system for designing an online catalog. This lets it offer a much bigger range of products while maintaining less than half the stock of competitors.

Actual business initiatives include 2/3 of the patterns, some examples involve more than 1 pattern. To uncover new business opportunities, you should know what data you have, what data can be accessed, what data could be created from the products/operations, what helpful data could be used and what others have for data that you may want to use. Based on this, modification/combination of patterns could be applicable in business context of the company.