LECTURE 1 - INTRODUCTION

<u>CONTEXT + INTRODUCTION COURSE THEME</u>

Digital innovation is a product, process or business model that is perceived as new, requires some significant changes on the part of adopters, and is embodied in or enabled by IT.

Digital technologies have a disruptive, pervasive, and permanent influence on business, work, and society as a whole. The exponential increase of computing power over recent years has given rise to digitization across a broad variety of sectors, ranging from big data analytics in online retailing, to robotization in industry and healthcare, internet of things enabling smart homes, ethical discussions on self-driving cars, digital business models in music and video, blockchains in banking, and smart production technologies and logistics. These and many other forms of digital innovation call for a profound understanding of the opportunities that digital tech offer as well as their potentially adverse effects.
Digital innovation that have huge impact on working and personal life.

Information abundance and the emergence of big data (key external digital trends) are influencing the digital business strategy (speed of decision-making and sources of value creation + capture), and therefore the performance of an organization.

Business Intelligence & Analytics are the techniques, technologies, systems, practices, methodologies, and applications that analyse critical business data to help an enterprise better understand its business and market and make timely decisions.
Business Analytics is concerned with evidence-based problem recognition and solving that happen within the context of business situations. Creating value. Turning data into improves business performance.
Business Intelligence & Analytics and Big Data requires advanced and unique data storage, management, analysis, and visualization technologies because of their large and complex techniques in applications.

BA is a subdomain of data science. It is all about creating value.
Data science is turning data into knowledge/information.

<u>BI&A: Why is it important for organizations?</u>
Big Data is more disruptive than the Internet was, and Big Data is about **four Vs**:

1. Volume: volume of Big Data is huge and enormous
2. Variety: structured: excel sheet how it's stored and unstructured data: FB: we do not know what the meaning is
3. Velocity: speed (=high) in which new data arises
4. Veracity: "can you trust the data that is coming to you as a manager?"

These four interact and give challenges.

What is different? Big Data is *granular* (Data is everywhere), *unintentional* (Data is used with a completely different purpose than it was stored for. FB message), *unmanageable*, and *constant renewal/updating*.

Moore's Law (the increasing power of computer technology) makes that the power of the technology is growing exponentially, and that the data is generated is also growing exponentially. When something enters the second half of the chessboard, this technology gets a big impact and it is difficult to imagine what will happen next.

BI&A is a top priority at "C-Level (Chief Information/Data Officer) in organizations, and innovation and creativity in the use and application of data for decision-making is critical to success as a business executive. Scientific research also proves that data-driven organizations are more successful than organizations that are not data-driven.

<u>BI&A: What does it include?</u>
BI&A is about discovering what you don't know you don't know. At the moment there are algorithms on Facebook that show you things you don't know you don't know (e.g. divorces based on likes from people). BI&A offers you new measurement tools, and these tools bring us new insights.

There are different types of **business analytics capabilities**:
1. Descriptive analytics (access and reporting data → describes what happened within organizations)
2. (Diagnostic: why did it happen?)
3. Predictive analytics (use data to predict businesses): what will happen
4. Prescriptive analytics (use data to prescribe future): how can we make it happen

Davenport (2013) - Analytics 3.0.
Analytics 3.0 is a new resolve to apply powerful data-gathering and analysis methods to company operations and to its offerings.

The common thread in these examples of the paper is the resolution of a management of an organization to compete on analytics not only in the traditional sense (by improving internal business decisions) but also by creating more-valuable products and services. This is the essence of Analytics 3.0!

- Analytics 1.0 = "business intelligence" = analytics is seen as a source of competitive advantage in the form of greater operational efficiency (making better decisions to improve performance/internal business decisions = traditional). Go beyond intuition when making decisions.
- Analytics 2.0 = "era of big data" = companies turned to a new class of databases (NoSQL).
- Analytics 3.0 = "data-enriched offerings" = compete on analytics not only in the traditional sense (decision making) but also by creating more valuable products and services.

(Adapted from Davenport, 2013)

	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 ten requirements for capitalizing on Analytics 3.0:
1. Integrate large and small volumes of data from internal and external sources in structured and unstructured formats to yield new insights in predictive and prescriptive models
2. A new set of data management options: combination of DW, database and big data appliances
3. Faster technologies and methods of analysis
4. Embedded analytics in operational and decision processes
5. Data discovery: Companies need a capable discovery platform for data along with the requisite skills and processes. Make it possible to determine the essential features of a data set without a lot of preparation
6. Cross-disciplinary data teams
7. When analytics are important, companies need senior management oversight
8. Prescriptive analytics require high-quality planning and execution
9. Analytics 3.0 provides opportunity to scale processes to industrial strength
10. You need new ways of decision-making and management

So, the big data model was a huge step forward, but it will not provide advantage for much longer. Analytics 3.0 is the direction of change and the new model for competing on analytics. Companies that want to prosper in the new data economy must again rethink how the analysis of data can create value for themselves and their customers.

BI&A: How to make evidence-based problem recognition and solving happen?
Lee et al. (2014) - A cubic framework for the CDO. A Chief Data Officer can lead the effort to build organizational capability that can energize and sustain the entire organization and extended enterprise. Furthermore, a CDO can be held accountable for a failure of leadership in resolving data problems. There are three dimensions of a CDO role:

1. Collaboration Direction Dimension: inwards (internal business processes with internal stakeholders) versus outwards (collaborating outwards with external stakeholders).
2. Data Space Dimension: traditional data versus big data (offers innovative opportunities to further improve operations/develop new business strategies based on new insights).
3. Value Impact Dimension: service versus strategy.

There are eight CDO roles and a CDO may take multiple roles over time, but a CDO has one primary role.

1. Coordinator: inwards/traditional/service: managing enterprise data resources & set up framework that optimizes collaboration across internal business units
2. Reporter: outwards/traditional/service: deliver high quality enterprise data services for external reporting purposes
3. Architect: inwards/traditional/strategy: using data or internal business processes to develop new opportunities for the organization
4. Ambassador: outwards/traditional/strategy: Promotes the use of inter-enterprise data policy for business strategy and external collaboration
5. Analyst: inwards/big data/service
6. Marketer: outwards/big data/service: Develops relationships with external data partners and stakeholders to improve externally provided data services.
7. Developer: inwards/big data/strategy: Interfaces and negotiates with internal divisions to develop new opportunities to exploit big data
8. Experimenter: outwards/big data/strategy: engages with external collaborators, such as suppliers and industry peers to explore new, unidentified markets and products based on insights from big data

Q: Which dimension do you consider most important for an online retailer?

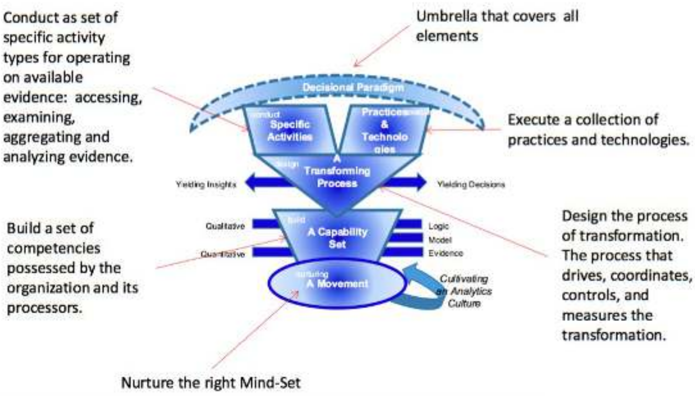
Holsapple et al. (2014) - A unified foundation for BA.
Holsapple et al. wanted to have a definition for business analytics. In the literature were various definitions that combine aspects of the point from the framework.
BA Framework: “As such, the business analytics framework (BAF) serves as an organizing device for practitioners to use *in planning and evaluating their analytics initiatives*, a generative mechanism for researchers to *use in identifying and designing their analytics investigations*, and a guiding template for educators to use *for positioning topics and ensuring full coverage in their analytics curricula*.” (Holsapple et al., 2014)
BA is concerned with evidence-based problem recognition and solving that happens within the context of business situations. There are six classes of business analytics:

1. Movement: people that make things happen related to data-driven decisions (nurture the right mindset)
2. Capabilities: competencies possessed by the organization and its processors to make things happen (it is a capability set that influences what BA transformational process

will be built, and that functions as both means and constraints on what a BA movement can accomplish).
3. Transforming process: design process of the transformation (data into insights + organization to be able to transform data)
4. Specific activities: conducting a set of specific activities for operating on available evidence (accessing, examining, aggregating, and analysing evidence).
5. Practices and techniques: execute a collection of practices and technologies: how things get done when operating in evidence.
6. Decisional paradigm: umbrella that covers all the elements – decision making

There are several reasons why you should pursue BA:

- Achieve a competitive advantage;
- Support of organization's strategic + tactical goals;
- Better organizational performance;
- Better decision outcomes;
- Better/more informed decision processes;
- Knowledge production; and
- Obtaining value from data.



In the end it is: nurturing an analytics movement → adopting a decisional paradigm that stresses analytics → building and growing a managed set of analytics capabilities (movement need the right capabilities) → designing analytics processes for transforming evidence into insights, decisions, and actions → conducting specific evidence-manipulation activities needed within an analytics initiative → and executing practices and technologies.

LaValle et al. (2010) - Analytics: new path to value.

How the smartest organizations are embedding analytics to transform insights into action

- Top-performing organizations use analytics five times more than low performers.
- Improvement of information & analytics is a top priority in many organizations.
- Organizations are pressured to adopt information and analytics approaches.
- Organizations that know where they are in terms of analytics adoption are better prepared to turn challenges into opportunities.

There are three levels of capabilities to achieve long-term advantage:

1. *Aspirational*: focus on efficiency/automation of existing processes + ways to cut costs. Far from achieving desired analytics goals. Assemble the best people and resources to make the case for investments in analytics. To get sponsorship for initial projects, identify the big business challenges that can be addressed by analytics and find the data you have that fits the challenge. Use analytics to justify actions
2. *Experienced*: look beyond cost management. 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. Use analytics to guide actions
3. *Transformed*: use analytics as a competitive differentiator and are already adept at organizing people, processes and tools to optimize and differentiate. Discover and champion improvements in how you are using analytics. You've accomplished a lot already with analytics, but are feeling increased pressure to do more. Focus your analytics and management bandwidth to go deeper rather than broader, but recognize it will be critical to continue to demonstrate new ways of how analytics can move the business toward its goals. Use analytics to prescribe actions.

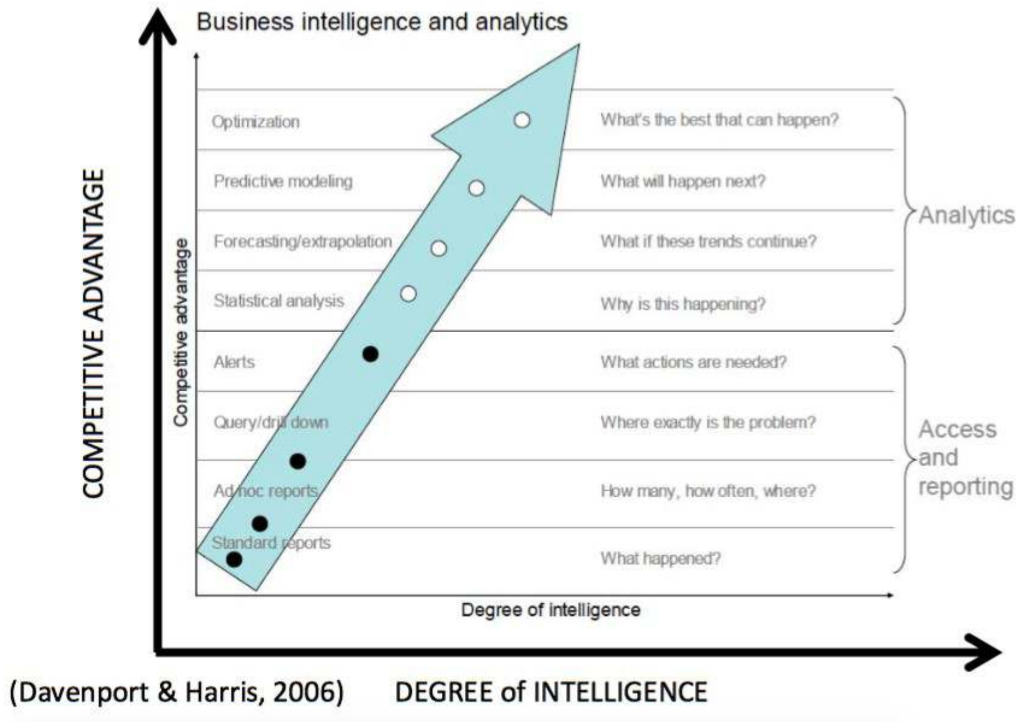
Adoption barriers for analytics are most related to management and culture rather than being related to data and technology. There are five recommendations for successfully implementing analytics-driven management and rapidly creating value:

1. Focus on the biggest + highest value opportunities (PADIE technique).
2. Start with questions, not data.
3. Embed insights to drive actions and deliver value.
4. Keep existing capabilities, while adding new ones.
5. Use an information agenda to plan for the future.

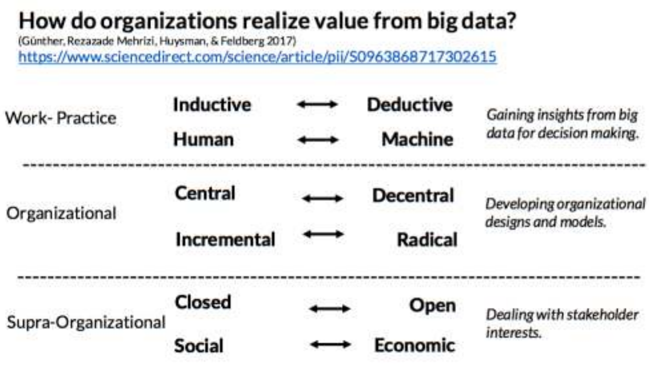
"....*Knowing what happened and why it happened are no longer adequate. Organizations need to know what is happening now, what is likely to happen next and, what actions should be taken to get the optimal results......"* (Lavalle *et al*., 2010: p.1)

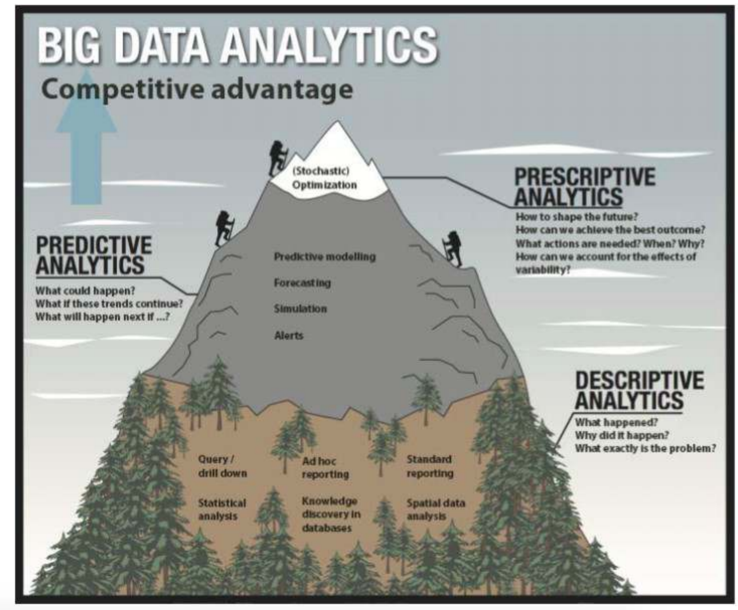
New insights
Microscope: We didn't know that there were bacteria. We knew it after we used the microscope.

Where to invest of what to do?
Types of business analytics capabilities



Based on the article of science direct we found 3 different levels:
Work practice: do we let the data speak for itself, or do we ask questions to the data? (inductive-deductive) Which decisions are made by human and which are made by machines.
Organizational: Do we have to centralize our knowledge? Or do we have to improve our business model on incremental or on radical way.
Supra-organizational: do we share our data or keep it for ourselves.





LECTURE 2 - DATA ARCHITECTURE

Data is about capturing information, and it is raw information. It is pre-factual and pre-analytical, it is different from facts and evidence, and it is different from information and knowledge. **Data** are that which exists prior to argument or interpretation that converts them to facts, evidence and information.

There are five kinds of data:

1. Form: qualitative and quantitative
2. Structure: structured (you know how data points are related to each other: you have prior knowledge and you use this to process the text; you don't need to interpret it before you analyse it → operational reports, OLAP, data-mining and social network analysis), semi-structured and unstructured (text mining and image/optical recognition, everybody can see something else)
   1. You use a hierarchical structure when you describe all attributes
   2. You use a network when you describe relationships between entities
   3. You also have a tabular form where you use rows and columns

b. Source: captured, derived, exhaust and transient (categorizing data)

c. Producer: primary (you created yourself), secondary (someone else produced it) and tertiary

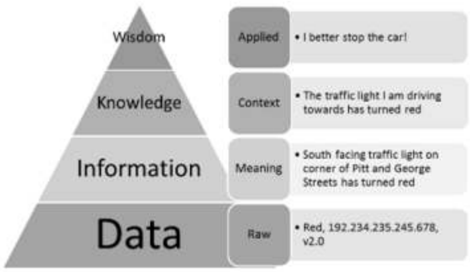
d. Type: indexical (refer to other data), attribute (mostly dealt with in this course, you can count it) and metadata (describes how other data is organized)

External organization: Digital traces, sensor/interaction/self-reported/contributed data and social media
Within organization: Enterprise systems (CRM, ERP = core business processes, the whole organization in one structure, running same kind of system and generate a lot of data), web and e-commerce transaction data.

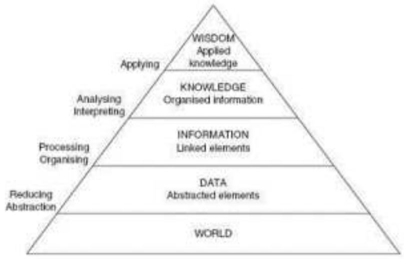
Relational databases are a combination from multiple databases, and is an Online Transactional Process (OLAP): you describe it with the entity-relationship model. You "talk" to a relational database with a **Structured Query Language (SQL)**, which is a high-level, declarative language for data access and manipulation. it is widely used for simple functional reporting, such as: "amount of sales this month".
• Allows asking the database human-like questions (queries)
Widely used for simple functional reporting

• Amount of sales this month
• The number of current female employees
• The age composition of the workforce

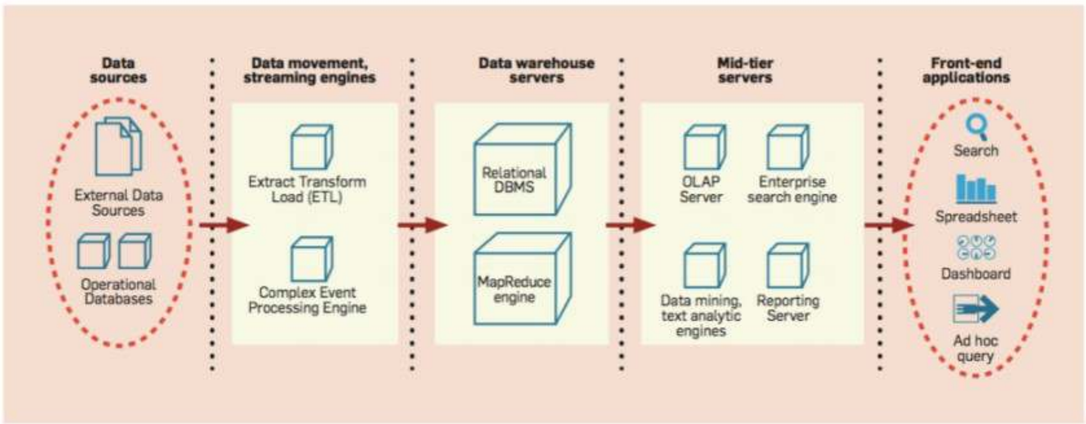
The knowledge pyramid: represents the structural and functional relationship between the layers. We have the world around us and we need to reduce some abstraction to receive data. When we process this data in an organizational way, we will receive information. When we analyse and interpret this information, we will receive knowledge. When we apply this knowledge, we will have wisdom.



Used during the lecture
===================>



Chaudhuri et al. (2011) - An overview of Business Intelligence technology.
Typical business intelligence architecture.



BI Software = a collection of decision support technologies for the enterprise aimed at enabling knowledge workers to make better and faster decisions. Today, it is difficult to find a successful enterprise that has not leveraged BI technology for its business. Data over which BI tasks are performed is typically loaded into a data warehouse that is managed by one or more data warehouse servers.

- Data warehouse servers are complemented by a set of Mid-tier servers: provide specialized functionality for different BI scenarios
  - Multidimensional view: provided by OLAP servers (online analytic processing supports operations such as filtering, aggregation, pivoting, rollup and drill-down)
  - Reporting servers: enable definition, efficient executing and rendering of reports
  - Enterprise search engines: keyword search paradigm
  - Data-mining engines: in-depth analysis of data that provides the ability to build predictive models
  - Text analytic engines: analyse large amounts of text data
  - In-memory BI engines: exploit today's large main memory sizes for ad-hoc queries. They rely on a different set of techniques for achieving good performance than OLAP.

Decision Support Systems (DSSs) require operations such *as filtering, join and aggregation*. To support these operations, special structures are developed in relational databases:

Index structures: indexes on columns that help retrieve what you want

- Index intersection
- Materialized views
- Partitioning
- Column-oriented storage
- Workload information

Data compression can have significant benefits for large data warehouses (reduces the amount of data needed to be scanned + can lower storage and backup costs + increases the amount of data that can be cached in memory since):

- Null-compression: data types are treated as variable length for storage purposes
- Dictionary compression: identifies repetitive values in the data and constructs a dictionary

Data platforms based on the map reduce paradigm have attracted strong interest in the context of the “big data challenge”, they provide the ability to support analytics on unstructured data. The competitive pressure of today’s business has led to increased need for near real-time BI, and Complex Event Planning (CEP) engines can enable this.

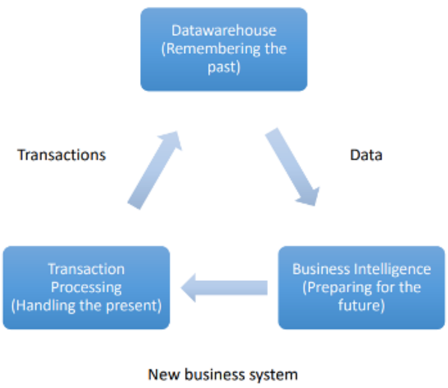
- Data does not first need to be loaded in a warehouse before it can be analysed.

Extract Transform Load (ETL) tools, such as data profiling / extracting structure / deduplication, play a crucial role in helping to discover and correct data quality issues and efficiently load large volumes of data into the warehouse. Data load and refresh are responsible for moving data from operational databases to external sources into the data warehouse quickly and with as little performance impact as possible on both ends.

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

Often, when you are doing an analysis for a company, you don’t want a report that only looks at one function: it is very important to bring different functions (e.g. customers and sales) together to make strategic decisions. To do this, you will use Online Analytical Processing (OLAP), which is a computational approach for answering multi-dimensional analytical queries (MDA), using interactive real-time interfaces.
MDA: drawing on several data domains (several dimensions): “the average amount of sales per branch of carrots on the weekends of the third quarter of 2016 in the Utrecht area”.

Information lifecycle of the firm →



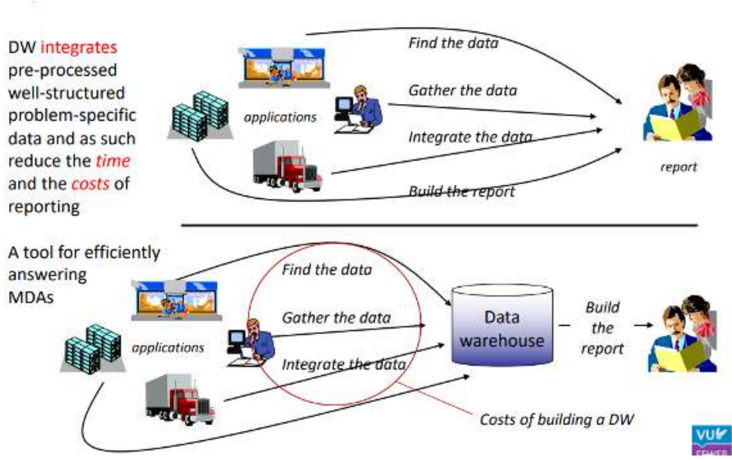
Snowflake schema = A refinement of star schemas where the dimensional hierarchy is explicitly represented by normalizing the dimension tables.
Most ROLAP systems use a star schema to represent the multidimensional data model.
MOLAP combining ROLAP = HOLAP

The competitive pressure today had led to the increased need for near real-time BI. The goal is to reduces the latency between the requirement of data and analysis.

A **data warehouse** is a collection of integrated, subject-oriented databases designed to support DSS function, where each unit of data is relevant to some moment of time. The data warehouse contains atomic and lightly summarised data.

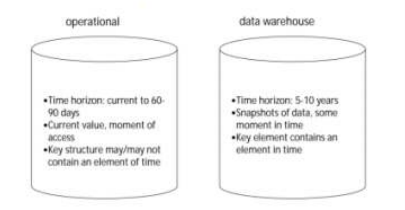
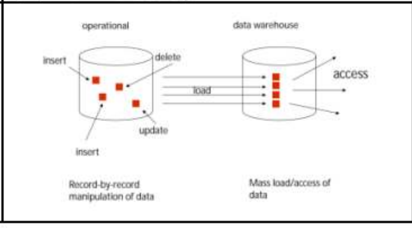
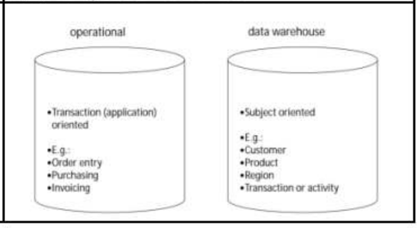
- It integrates **pre-processes well-structured problem-specific data** and as such reduce the time and the costs of reporting (Think of GM case → reduces time)

Data warehousing is in place to improve your business, and to innovate your business. Furthermore, you will save time, you will have more and better information, better decisions, there can be an improvement of your business processes and there is support for the accomplishment of strategic business objectives. When you find/gather/integrate data you can build a report as a result. The greatest benefits of DW apps occur when used *to redesign business processes and support strategic business objectives.*



A data warehouse has five characteristics:

1. Data integration: you do not want to dump the data from multiple applications in a database: you need ETL to transform the data in a purposeful format (= more strategic view)
2. **Subject orientation**: looks at making decisions about customers, product, region, and transaction or activity. It is about putting data together and asking interesting questions where you look at multiple functions, and not only about what the company is currently doing.
3. Historical data: dead archival data is brought back to life for analysis purposes. It is not relevant to know the kind of products you sold 20 years ago, but it is useful for sales over 1 year, which you can collect in the data warehouse.
4. **Non-volatility/non-moving:** mass-loaded and geared towards bulk analytic access. Furthermore, it contains snapshots of values from different moments.
5. **Time-dependency**: time is a key structural element.

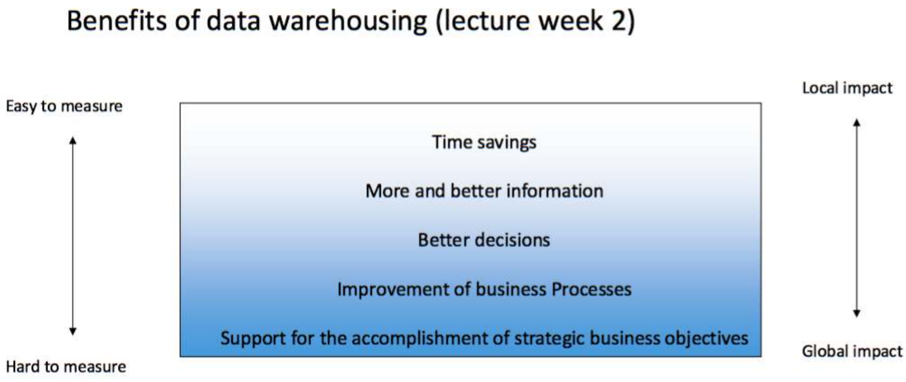
(1) Time-dependency	(2) non-volatility	(3) Subject orientation
		

Watson et al. (2001) - The benefits of data warehousing.

The greatest benefits occur when data warehousing is used to redesign business processes and to support strategic business objectives. For a company to re-design itself, it is necessary to unfreeze the old way of doing business; changing to a new way of doing it; and the freezing it in again. Furthermore, it is best to see data warehousing effort as a means to a strategic end (an appropriate IT infrastructure will follow). *Fundamental business driver behind DW is the desire to improve decision making and organizational performance*. → It involved extracting data from source systems, cleaning, transforming the data and loading it to a data store.

Change and truly momentous benefits are far more likely when the vision comes from the business side of the house rather than the IS side.
Kotter's 8 steps to transforming your organization: how to successfully achieve major change:

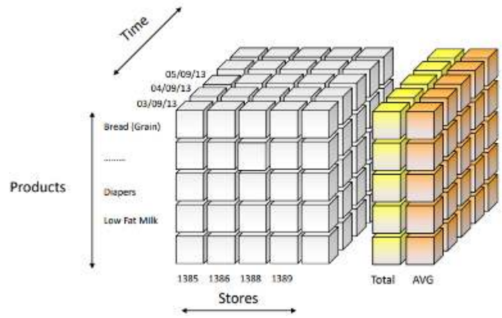
Step 1: Establish a sense of urgency	Step 3: Create a vision	Step 5: Empower others to act on the vision	Step 7: Combine improvements and produce still more change
Step 2: Form powerful guiding coalition	Step 4: Communicate the vision	Step 6: Plan for and create short-term wins	Step 8: Institutionalize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



A case for example might be: do you think it is useful to apply BI&A in an organization? You can answer it with the example of benefits of data warehousing.

A **decision cube** is the equivalent of a pivot table for a data warehouse (multidimensional table = a popular conceptual model used for BI tasks), and allows for summarizing and analysing facts pertaining to a specific subject along different dimensions.

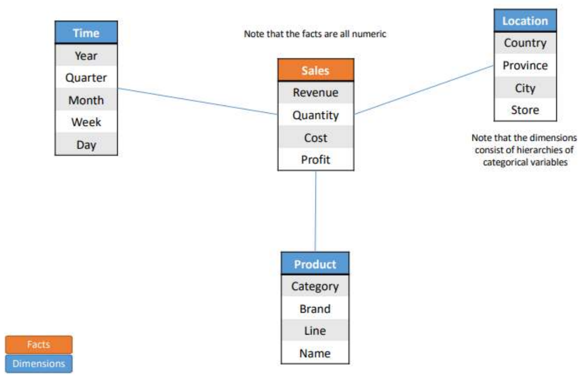
- Fact: what is tracked about the subject “how much” (sales, revenue, units sold)
- Dimension: acts as an index for identifying values within a cube (e.g. time/products/stores) “*when, where, what and how*”. All of which are of a similar type in the perception of the user of the data. Geography/products/time dimensions. Tagging what is tracked (time, product, store of sales).
- Drill down: “from a specific day to a specific time that day”
- Roll up: “from a day to a week”
- Slice and dice: “take specific rows and columns” to answer questions
- Data cube: a group of data cells arranged by the dimensions of the data



The key concepts of a multidimensional data model:
1. Multidimensionality: Organizing/presenting/analyzing data by several dimensions, such as sales by region, by product, by salesperson, and by time (= 4 dimensions).
2. Multidimensional analysis: An analysis of data by ≥ 3 dimensions
3. Multidimensional modelling: A modelling method that involves data analysis in several dimensions.

- **Online Analytical Processing (OLAP)**

- A computational approach for answering multi-dimensional analytical queries (MDA), using interactive real-time interfaces.
- Enables users to view the same data in different ways using multiple dimensions.
- Provides online answers to ad hoc questions in rapid time (users need to have a good idea about the information for which they are looking)



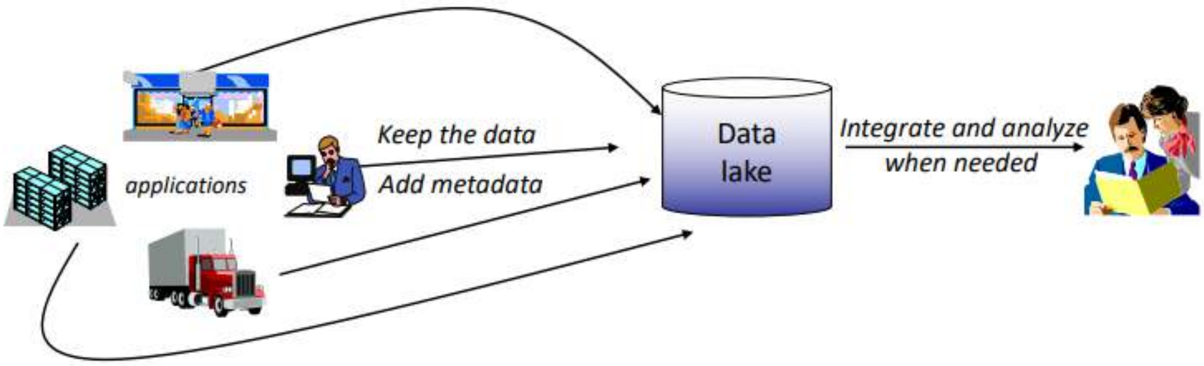
A Star Schema is the simplest style of a data smart schema in computing. It consists of one or more fact tables referencing any number of dimension tables and is called a star schema as it looks like a star, points emitting from a center. A star schema consists of 2 types of tables:
1. Fact tables : The center, consists of foreign keys of dimension tables and measure that contain numeric facts.
2. Dimension tables: Everything around the center with foreign keys.

The main merits/verdiensten of data warehouses become their drawbacks/nadelen in the age of Big Data, and this is where **data lakes** come into play.

- “We keep the original data and analyse it with a data lake; pipeline = new actions based on insights from the data.”

Stein and Morrison (2014) - The enterprise data lake.
Extracting and placing data for analytics in a Hadoop-based repository allows an enterprise’s records to be stored in their narrative formats for later parsing. A **data lake** relaxes standardization and defers modelling (using commodity cluster techniques), which results in a nearly unlimited potential for operational insight and data discovery → as data volumes, data variety, and metadata richness grows, so does the benefit. Data lake support late binding in which users build custom schemas into their queries.
Data lakes are an emerging and powerful approach to the challenges of data integrations as enterprises increase their exposure to mobile and cloud based applications, the sensor-driven internet of things etc..
It is a repository for large quantities and varieties of data, both structured and unstructured. Many companies see it as an opportunity to capture a 360-degree view of their customers of to analyse social media trends.
There are four criteria that underlie a good definition of data lakes: *big size + low cost, fidelity (keeps data in its original form), ease of accessibility, and late binding (*Flexible, task-oriented structuring with no need for up-front data models)
Sourcing new data into the lake can occur gradually and will not impact existing models:
- Stage 1: Consolidate and categorize raw data
- Stage 2: Attribute-level metadata tagging + linking
- Stage 3: Data set extraction and analysis
- Stage 4: Business-specific tagging, synonym identification, and links
- Stage 5: Convergence of meaning within context

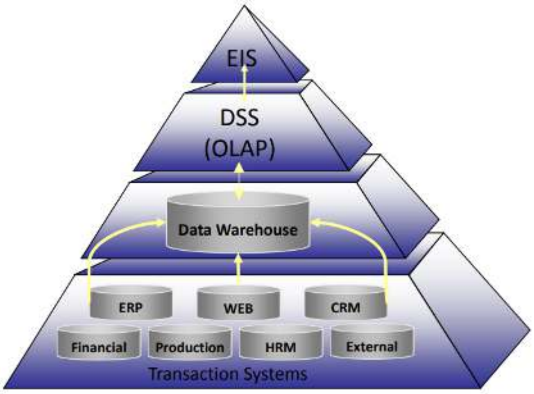
DW have slow-changing data models and do not support well big data volume and variety, they have also big prices tags. But data lakes don’t enforce a rigid metadata schema as do relational DW. Instead, DL support a system in which users build custom schema into their queries. Users can take what is relevant and leave the rest.



LECTURE 3 - DATA MANAGEMENT: INFORMATION REQUIREMENTS, QUALITY, GOVERNANCE, PRIVACY & ETHICS
4 sub topics to understand the basic concepts of data management in the age of BI&A

Information requirements
We are looking into information systems to help them (CDO’s) in their work
Pyramid of information systems: transaction systems are at the bottom of the pyramid (ERP/WEB/CRM/financial/production/HRM/external), and they can all be integrated in a data warehouse. On the data warehouse you can have a DSS (computer based decision support system = OLAP) that supports all the data, and on top of that you have an EIS (executive information system).

- When you look at the domain of IS, you see that EIS are relevant for their critical success factors. They are used to be a DSS and now you see more and more that they are referred to as BI&A.
- Executive work activities are diverse, brief and fragmented. Furthermore, verbal communications are preferred.



Metaphor: Managerial cockpit: there are a lot of levers, and you need a lot of training to be able to handle these leverages. You need to know how to get in the right direction. But what should it look like? When you collect requirements, it is important that you get in touch with an executive, and figure out what information you need for the particular tasks that a person is doing.

There are four **strategies for identifying information requirements**:
1. Asking
2. Deriving from an existing IS
3. Synthesizing from characteristics of the utilizing system: users; what kind of people/is it an organization or do you make it for a group of persons?
4. Discovering from experimentation with an evolving IS: prototyping/ongoing development

Problems that may develop from these strategies: people may not know what information is needed and current systems may not be relevant.

There are two phases in the lifetime of an EIS: the initial phase and the ongoing (modify, add new screens, analysis, reports) phase. An EIS continues to evolve over time in response to: competitor actions, chancing customer preference, government regulations, industry developments, and technological opportunities.

Methods for determining information requirements: there are 16 methods that can be used, and six of them will be discussed:
1. Discussions with executives (most used method and highly useful)
2. Critical success factor sessions (most effective way and is an important part of consultant's toolbox): the limited number of areas in which satisfactory results will ensure successful competitive performance for individual, department or organization. Critical success factors (CSFs) are the few key areas where ‘things must go right’ for the business to flourish and for the manager’s goal to be attained.
    1. You use Key Performance Indicators (KPIs) to measure (quantify) CSFs.
    2. The use of CSFs and KPIs enables measurement, and thus control, of strategic objectives. It is necessary to include KPIs that measure the execution of the strategy and the creation of value.

b. Information systems teams working in isolation (least used method). Not being in alignment with your organization.
c. Tracking executive activity (can lead to someone who is acting differently when they realize that they are being tracked)
d. Software tracking of EIS usage (can only be used when IS is already in place)
e. Formal change requests (can only be used when IS is already in place)

The application of specific method for determining requirements depends on the phase, but each has its pros and cons. It is best to use multiple methods to triangulate (validate) information needs (requirements). Development of BI&A is a journey rather than destination.

The other 10 methods:

- EIS planning meetings
- Examinations of computer-generated information
- Discussions with support personnel
- Volunteered information
- Examination of EIS of other organizations
- Examinations of non-computer-generated info
- Participation in strategic planning sessions
- Strategic business objectives method → related to critical success factors
- Attendance at meetings
- Examination of the strategic plan

<u>Data quality</u>
The quality of data used in BI&A should be high, but what is data of high-quality? There are six dimensions of data quality:

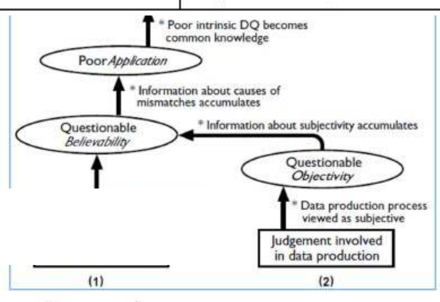
1. Timeliness: is the information there on time? Can the data be used anytime?
2. Completeness: no missing values
3. Consistency: absence of difference: can we match the data set across data stores?
4. Uniqueness: no thing is recorded more than once
5. Validity: defines the way it should be
6. Accuracy: is the measure correct of what is happening?

There are various estimates of what the effects from low data-quality are, but in general people say that it costs you money and that you'll have inaccurate data. At the *operational level* there will be lower customer and employee satisfaction, and increased cost. At the *tactical level* there will be poorer decision making, more difficult to implement data warehouses, more difficult to engineer, and increased organizational mistrust. At the *strategical level* it becomes more difficult to set and execute a strategy, contribute to issues of data ownership, compromise ability to align organizations, and to divert management attention.

Strong et al. (1007) - Data quality in context.
New study reveals businesses are defining data quality with the consumer in mind.
Often, DQ is treated as an intrinsic concept, independent of the context in which data is produced and used. This focus on intrinsic DQ problems in stored data fails to solve complex organizational problems. In contrast to this intrinsic view, it is well accepted in the quality literature that quality cannot be measured independent of consumers who choose and use products. *Data consumers' assessments of DQ are increasingly important because consumers now have more choices and control over their computing environment and the data they use. To solve organizational DQ problems, one must consider DQ beyond the intrinsic view.* Therefore, the article is about other views of DQ like the accessibility, contextual and representational DQ. *This data consumers' perspective is a broader conceptualization of DQ than the conventional intrinsic view.*

High quality data = data that is fit for use by data consumers
DQ problem = difficulty faced along 1 or more than 1 quality dimensions that renders completely of largely unfit data for use
DQ project = organizational actions taken to address DQ problems given recognition of poor DQ by the organization

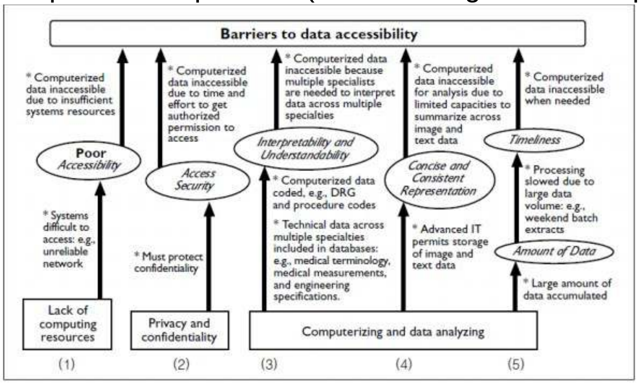
DQ Category	DQ Dimensions
Intrinsic DQ	Accuracy, Objectivity, Believability, Reputation
Accessibility DQ	Accessibility, Access security
Contextual DQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational DQ	Interpretability, Ease of understanding, Concise representation, Consistent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



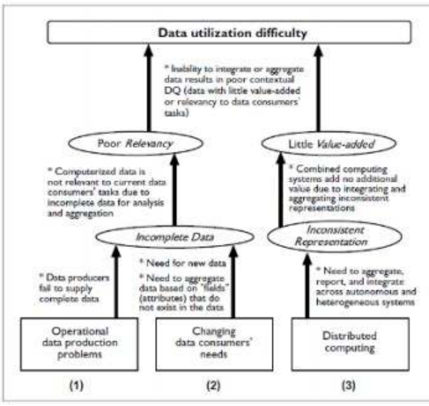
There are four categories of characteristics of problems of high-quality data: intrinsic, accessibility, contextual and representational aspects.
There are two intrinsic data quality patterns:

1. Mismatches among sources of the same data: data consumers do not know the source to which quality problems should be attributed. 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.
2. Judgement/subjectivity in the data production process: initially only those with knowledge of data production processes are aware of the problems. Over time, information about the subjective nature accumulates, which results in data of questionable believability and reputation. It is considered as lower quality than raw uninterpreted data.

There are two approaches to problem resolution here: 1) change the system, and 2) change the production process (leads to long-term data quality improvements).



- ← Accessibility data quality problems: these problems are characterized by underlying concerns about technical accessibility; data representation issues interpreted by data



  consumers; and data-volume issues.
- Contextual data quality problems: incomplete data due to operational problems/database design; and problems caused by integrating data across distributed systems. →

To solve organizational DQ problems, IS professionals must attend to the entire range of concerns of data consumers. The 3 patterns for how intrinsic, accessibility, and contextual DQ problems develop in organizations provides an empirical basis for studying organizational choices and actions about DQ improvement.

<u>Data governance</u>
Data governance is the exercise of authority and control over the management of data assets. **Data governance** is the **data management** of all the data which an organization has to ensure that high data quality exists throughout the complete lifecycle of the data.

Weber et al. (2009) - One size does not fit all.
This paper outlines a data governance model that consists of 3 components (DQ roles, Decision areas, responsibilities) which form a responsibility assignment matrix. Data quality is defined as the dependence of perceived quality on the user's need, and the fitness for use. The goal of IT governance is to ensure that IT is a valued and embedded element for business and is enabling a company's overall strategy.

- In centralized models corporate IT performs all IT functions, whereas in decentralized models business units perform these tasks.

The data governance model consists of DQM decision areas and main activities/roles/responsibilities. The 3 components can be arranged in a matrix. The rows of the matrix identify the key decision areas and main activities, while the columns indicate the roles in DQM. The cells of the matrix contain the responsibilities types of interaction, they specify degrees of authority between roles/decision areas. With reference to business engineering, the data governance model addresses DQM on 3 horizontal layers: strategy, organization, and information systems. As opposed, the data governance model described allows for company-specific configuration of data governance. The data governance contingency model is a moderation model with contingencies as covariation effects.

	DQM Role 1	DQM Role 2	DQM Role 3	...	DQM Role n
DQM Task 1	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
DQM Task 2	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
DQM Task 3	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]		[R|A|C|I]
...		DQM Responsibilities (assignment of roles to tasks)			[R|A|C|I]
DQM Task n	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]	[R|A|C|I]

R – Responsible; A – Accountable; C – Consulted; I – Informed

The contingency factors determine the fit between the design of the data governance model and the success of DQM within the organization which means the effectiveness of the organization. The elements of the data governance contingency model are depicted.

The data governance model uses a set of four roles and one committee (data quality board):

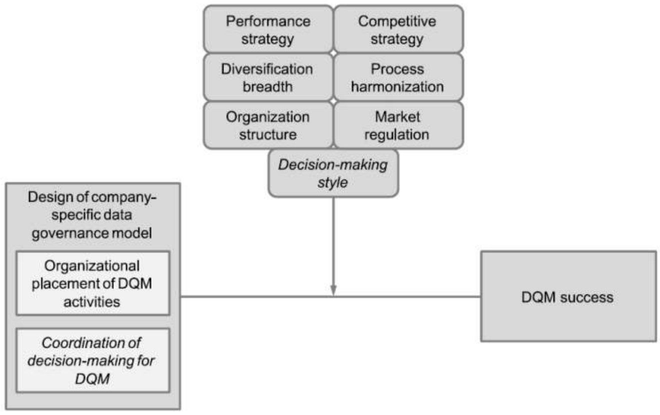
1. Executive sponsor: provides sponsorship, strategic direction, funding, advocacy, and oversight for DQM
2. Chief steward: puts board's decisions into practice, enforces adoption of standards, helps establish DQ metrics and targets
3. Business data steward: details corporate-wide DQ standards and policies for his/her area from a business perspective
4. Technical data steward: provides standardized data element definitions and formats, profiles and explains source system details and data flows between systems
5. Data quality board: defines the data governance framework for the whole enterprise and controls it implementation

The assignment of responsibilities to roles follows the RACI chart: responsible, accountable, consulted, and informed. More than one role may be accountable for authorizing a decision (cooperative culture). However, only one role is ultimately responsible for executing an activity (commitment of a single role for taking care of a certain task).

There are two design parameters that facilitate the comparison of different data governance models:

1. *Organizational structuring* of DQM activities: centralized versus decentralized
2. *Coordination of decision-making* for DQM activities: hierarchical IT governance versus cooperative IT governance

There are seven contingency factors that affect the fit between a company-specific data governance model and successful DQM:



<u>Data privacy & ethics</u>

Tene and Polonetsky (2013) - Big data for all: privacy and user control in the age of analytics.

When considering the risks that big data poses to individual privacy, policymakers should be mindful of its sizable benefits. Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world:

1. Incremental effect: once any piece of data has been linked to a person's real identity, and the association between this data and a virtual identity breaks down the latter

2. Automated decision-making: automated processes based on algorithms and artificial intelligence raises concerns about discrimination, self-determination, and the narrowing of choice
3. Predictive analysis: big data may facilitate predictive analysis with stark implications for individuals susceptible to disease, crime or other socially stigmatizing characteristics or behaviours. → Teen pregnant story
4. Lack of access and exclusion: Individuals exchange personal data for free services and organizations are seldom prepared to share the wealth rated by individuals' personal data with those individuals. Because they don't have access it is impossible.
5. Ethics: where should the red line be drawn?
6. Chilling effect: people will behave different when they realize that they are being tracked

There are also several oplossingen/verbeteringen related to current legal frameworks for privacy:

1. De-identification: Re-identification: over the past few years, computer scientists have shown that even anonymised data can be re-identified and associated with specific individuals. Big Data makes de-identification harder. Maar de-identification is temporary than stable. For example: medical anonymizing research.
2. Data minimizing: in considering the fate of data minimizing, the principles of privacy law must be balanced against societal values and economic efficiency. Organizations are required to limit the collection of personal data to the minimum extent necessary to obtain their legitimate goals. They have to delete data that is no longer used. For big data world, the principle of data minimization should be interpreted differently.
3. Individual control and context: By emphasizing consent existing privacy frameworks impose significant, sometimes unrealistic, obligations on both organizations and individuals → it is also flawed. By just using the words 'privacy policy' customers get a feeling of control. Individuals can be influenced to opt-in or opt-out of an option (organ donation). Consent based processing tends to be regressive because individuals' expectations fall back on existing experiences.

There are some challenges of the legal framework:

1. It is hard for law/policy to keep up with technological developments/possibilities.
2. De-identification as protective measure rather than as a solution.

There are some solutions that can be realized while relaxing data minimization and consent:

1. Organizations should share the wealth created by individuals' data with those individuals in a usable format. This access is most needed where de-identification is weak and the data could therefore provide tangible benefits to individuals. The implementation may require digital signatures and encrypted communication delivery channels.
2. Enhanced transparency: shining the light: Organizations should reveal the criteria used in their decision-making processes. Transparency of data collection/analytics.

Big data boosts the economy, transforming traditional business models and creating new opportunities through the use of business intelligence, sentiment analysis and analytics. This article suggests that to solve the big data privacy quandary, individuals must be offered meaningful rights to access their data in a usable, machine-readable format.

LECTURE 4 - DATA-MINING

Fayyad et al. (1996) - From data mining to Knowledge Discovery in Databases.

KDD (Knowledge Discovery in Databases) is an attempt to address data overload. *It (KDD) refers to the **overall process of discovering useful knowledge from data**, and data mining refers to a particular step in this process.* A driving force behind KDD is the database field, and a relational field is data warehousing (collecting and cleaning transactional data to make them available for online analysis and decision support).

A KDD consists of four main steps: data preparation, search for patterns, knowledge evaluation, and refinement. Furthermore, it consists of nine specific steps:

1. Develop an understanding of the application domain and the relevant prior knowledge + identify the goal of the KDD process from a *customer's viewpoint*
2. Create a target data set on which the discovery is to be performed
3. Data cleaning and pre-processing (warehouse)
4. Data reduction and projection
5. Match the goals of the KDD process to a particular data-mining method
6. Exploratory analysis + hypothesis selection
7. *Data mining*: a data mining algorithm consists of three components = model representation, model evaluation, and search / correlations, patterns and trends
8. Interpreting minded patterns
9. Acting on the discovered data

Data mining is 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 (algorithm) 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.*
It enables in-depth analysis of data including the ability to build predictive models. The set of algorithms offered by data mining go well beyond what is offered as aggregate functions in relational Olap or other hierarchical services.

> Correlation: variation between variables. You have a correlation between two variables. With data mining you'll seek to correlation. → relationship
> Pattern: looking at pattern of two variables.
> Trends: you have two variables → looking at the trend of various attributes.

- "Predict which households will adopt green energy in a neighbourhood"
- "Query via a web search engine for information about Mission Impossible 5"
- "Group together customers with similar purchasing behaviour" = relationships in data. Example: Male consumers with income between 60-70K tent to order new iPhones faster.

You have two types of data (from a miner's view):
- Non-dependency-oriented data → cijfers
    - Numeric: continuous
    - Ordinal: ordered discrete
    - Categorical: unordered discrete
    - Binary: a special case of ordinal or categorical
- Dependency-oriented data: → afhankelijk van iets dus al met elkaar te maken heeft
    - Temporal
    - Spatial
    - Sequences
    - Graphs

Besides these types of data, there is also a distinction that can be made between descriptive and predictive data mining:
- **Descriptive/explorative data mining**: learn about and understand the data / what happened?
- **Predictive data mining**: build models in order to predict unknown values of interest ("a model that predicts how much the customer will spend on the next catalogue order, given a customer's characteristics")

There are two main **applications of data mining**:
1. Customer Relationship Management: CRM

    - *Target marketing* (bol.com/booking.com, problem is use list of prospects for direct email campaign, the solution is use data mining to identify most promising customers, based on past purchase behaviour) → classification
    - *Attrition prediction/churn analysis* (t mobile/tele2/KPN: problem is percent loss of customers and avoid adding churn-prone customers, the solution is use data mining to identify typical patterns of usage of likely-to-defect). → clustering
2. E-Commerce/Mobile Commerce:
    - *Collaborative filtering* (Amazon/Netflix: Problem is how to use information from other users to make inference about a particular user? Solution is use data mining to find users with similar tastes) → groep / clustering
    - *Browse-click-conversion* (de Bijenkorf/Wehkamp: Problem is a large number of people browsing a website, but only a few of them actually make clicks/purchases. Solution: trough clicks, classify customers, adjust website/design features to increase conversion rate) → classification

Mining typologies =
Type of relationship:
- Between attributes (classification)
- Between observations/records (clustering)

Type of algorithm:
- Regression (maps a data item to a prediction variable) (learning a function that maps a data item to a real-valued prediction variable / Takes one variable and maps it into another variable)
- Data clustering (outlier detection) (Common descriptive task where one seeks to identify finite set of categories/clusters to describe data) (unsupervised learning)
- Data classification (decision trees and rules) (Learning a function that maps (classifies) a data item into one of several predefined classes.) (supervised learning)
- Association Rule Discovery (given a set of records, each of which contain some number of items from a given collection: capture the co-occurrence of items) (unsupervised learning) (You type in NY in Google, and because Google knows you are not in NY, you get results like hotels/Airbnb)

Type of learning:
- Supervised (you give the computer some pairs of inputs/outputs, so in the future when new inputs are presented you have an intelligent output - classification) Training/Test set.
- Unsupervised (you let the computer learn from the data itself without showing what the expected output is - clustering)

Examples:
• Customer segmentation (Clustering)
• Intrusion-detection (Classification)
• Fraud (Classification)
• Target marketing (Classification)
• Store product placement (Association rule mining)
• Product recommendation (Collaborative filtering or Clustering + ARM)

You can use data mining in two ways:
1. *As decision support system (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
    - 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
    - Storytelling Manager: Identifying the potentials, evaluating proposals for execution, interfacing with all the stakeholder

2. *As core operation (recommender systems)*

- 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.
  - 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.

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. This overall KDD process includes the evaluation and possible interpretation of the mined patterns to determine which patterns can be considered new knowledge.

What can go wrong:

- Problem formulation
    - Need to understand the business well → uniquely human feature. You are the one that asked the question, as a human. Algorithm can't ask the question. You have to formulate the question properly
- 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

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.
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.

Understanding data mining and model induction at this component level clarifies the behavior of any data-mining algorithm and makes it easier for the user to understand its overall contribution and applicability to the KDD process. *An important point is that each technique typically suits some problems better than others*. Thus, there is no universal data-mining method, and choosing a particular algorithm for a particular application is something of an art.

Jones (2013) - Introduction to approaches and algorithms. Learn about the concepts that underlie web recommendation engines.
Recommender systems identify recommendations for individual users based on past purchases and searches, and other users' behaviour:

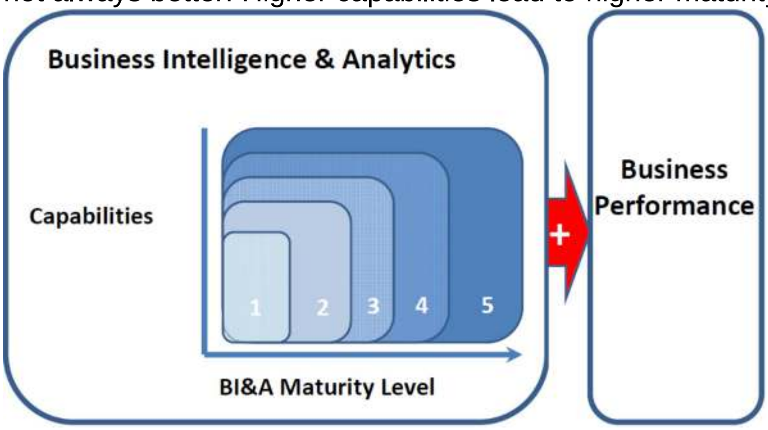
- Collaborative filtering: use group knowledge to form a recommendation based on like users (clustering) / Collaborative filtering arrives at a recommendation that's based on a model of prior user behaviour. Your behaviour or other users behaviour who have similar traits. For example: group users who have similar interests (blogs).
- Content-based filtering: use historical browsing information to identify and recommend similar content
- Hybrid approaches: starts (cold start) with content-based filtering, and ends with collaborative filtering. Increasing the efficiency (and complexity) of recommender systems. → more accurate recommendation.

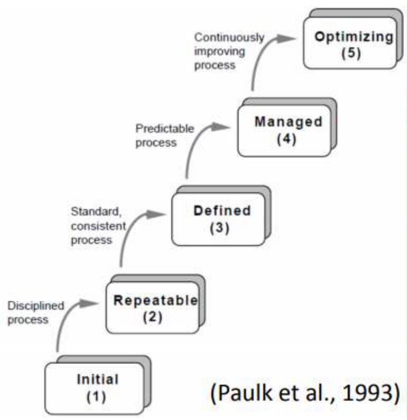
Algorithms that recommender systems use

- Pearson correlation: measures the linear dependence between 2 variables/users as a function of their attributes. Popular for collaborative filtering.
- Clustering: unsupervised learning that can find structure in a set of seemingly random data. (k-means)

LECTURE 5 - ORGANIZING FOR BI&A: MATURITY, SUCCESS, AND CULTURE
It is about determining if BI&A function is mature and if it is convenient to outsource it. You should know if/when the BI&A is a success and how to advice on the BI&A outsourcing decision. The way of improving/measuring/understanding BI&A maturity of an organization will be addressed as well.

<u>BI&A Maturity (models)</u>
The maturity of a BI&A model is the way an object is developed and the differences between objects. The manner this evolves over time can be modeled by using maturity models.
You can use maturity for benchmarking and for roadmapping: when you are in level 1 (initial) you want to go to level 5 (optimizing), in this way you make a plan for the future. BI&A is like a mountain and the higher the better, but also the more complex it becomes, more mature is not always better! Higher capabilities lead to higher maturity and business performance.





(Paulk et al., 1993)

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

This paper develops a holistic, theoretically-grounded and practically relevant **business analytics capability framework (BACF)** that specifies, defines and ranks the capabilities that constitute an organisational BA initiative. That framework *help explain how organisation achieve benefits with BA systems*. Capability frameworks are important instruments for information management and BA is an important application field for these frameworks. Findings shows that IT capabilities are more likely to yield competitive advantage than IT resources.

BACF was developed in 2 phases

1. Conceptual framework was developed based on resource-based view theory.
2. Conceptual framework was further developed and refined using a 3 round Delphi study involving 16 BA experts

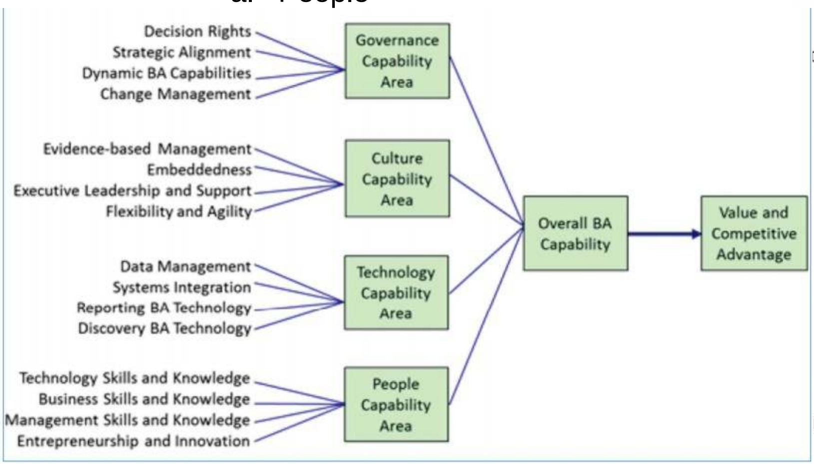
So, the Delphi study saw significant changes made to capability descriptions in the framework, minor but important changes to the overall structure, and a high level of consensus achieved after 3 rounds.
VRIN IS capabilities are developed from interactions that occur between VRIN IT resources and other VRIN organizational resources. The publications were analysed thoroughly in order to identify any keywords, themes of phrases that corresponded to the previously mentioned definition of a BA capability. There are four groups of analytic capabilities based on similarities:

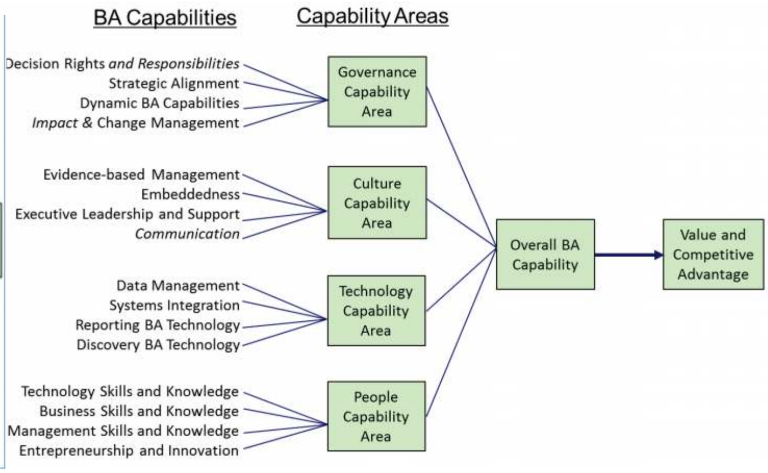
1. Governance: managing the use of BA resources and the assignment of decision-rights and accountabilities to align BA initiatives with organizational objectives

a. Culture: tacit and explicit organizational norms, values and behavioural patterns that lead to systematic ways of gathering, analysing and disseminating data

a. Technology: development and use of hardware, software and data within BA activities

a. People





<u>Outsourcing</u>

Fogarty (2014); Should you outsource analytics? .

Outsourcing relates to the question "make or buy", where make means that you perform something in-house (when something is made by the organizations itself), and buy (organization spent money on some aspects) means that you outsource it. It is important to consider whether an activity is high or low in its contribution to an organization, and to look at how strategically important the activity is for an organization.

Whether you outsource the BI&A function of an organization depends on the *type of organization.*

1. **Analytically challenged** companies outsource analytics because it is cheaper to outsource these actions than learning employees how to

do it. They see outsourcing as a quick and easy way to access capability/skills, and generally do not worry about intellectual property, like to collaborate in this area.
2. **Analytically superior:** companies (analytic as a core operation) only outsource their "basic" analytics to make more time for internal analysts. It is hard to outsource their core analytic operations because this has a high strategic value for these companies (it is the whole part of their operating). When they do outsource their strategic analytics, they need to share information and create a partnership to protect intellectual property. They see analytics as important core competence leading to competitive advantage.

Outsourcing analytics requires a carefully constructed relationship. Managers should know certain things before outsourcing their analytics. You should design the outsourcing relationship while considering the different motivations of them.

Outsourcing is an easier way to get capabilities quickly and at a low cost. But, the cost of setting up own analytics teams and the risk of failure if the organization invested too little made setting up offshore analytics teams appear may be too risky. Remember that outsourcing analytics was seen as a way to achieve advantage over competitors.

Companies that perceive the need to grow their analytic capabilities can gain advantages by using an offshore analytics BPO. By choosing the right analytics BPO to match their culture and business requirements they can develop competitive/unique analytic capabilities that can lead to competitive advantage. But, it is important to maintain a degree of control over the analytics BPO his marketing efforts and avoid giving the BPO too much credit for their joint work.

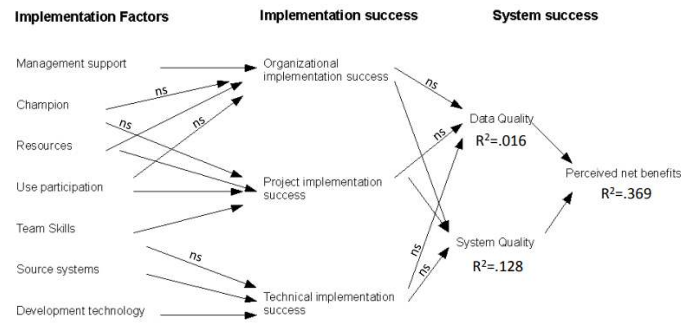
Service-oriented DSS: there are three types of service-oriented thinking:
1. Data-as-a-service (Information sources, data management (ETL + warehouses + marts))
2. Information-as-a-service (routine business reporting, OLAP, dashboards, intranet search for content)
3. Analytics-as-a-service (hardest to get and to outsource) (optimization, data mining, text mining, simulation, automated decision systems)

A service oriented focus entails reusability, substitutability, extensibility, scalability, customizability, composability, and reliability. Organizations need to respond quickly to their (changing) environment, and a service-oriented DSS makes this possible. There is also a difference between tight and loose coupling:
- due to a more service-oriented way of thinking you get more loose coupling between systems, and between systems and business processes.

When you talk about BI&A and the decision to outsource, it is the "as-a-service" concept that leads to:
1. New organizational units
2. Restructuring of business processes
3. A possible change in job satisfaction
4. A possible increase in job stress and anxiety
5. A possible change in manager's activities and their performance: it may require more knowledge, and middle management can become automated

<u>BI&A Success</u>
Wixom and Watson (2001) - An empirical investigation of the factors affecting data warehousing success.
A data warehouse is a repository of data created to support decision-making; *data are extracted from source systems, cleaned/scrubbed, transformed, and placed in data stores.*
There are three facets that influence the success of a data warehouse implementation:

- Organizational issues (people may tend to resist the implementation due to considerable organizational change that the data warehouse requires)
- Project issues (how well meets the team it's critical time, budgetary and functionary goals)
- Technical issues (precluding warehouse from high-quality data + no flexible system/not integrated as required).



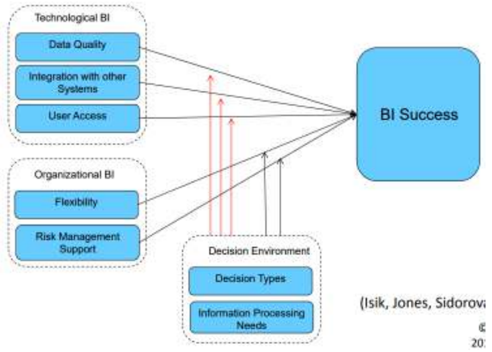
The success of the implementation in turn affects the system success (quality of data: accuracy, completeness and consistence & system quality: flexibility, integration, response time and reliability).

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. It affects the overall success of the system.
Implementation factors
Management support – High organizational implementation success
Champion – Actively supports and promotes the project and provides information, material resources, and political support. (high organizational & project success)
Resources – The money, people, and time that are required to successfully complete the project (high organizational & project)
User participation – When users are assigned to project roles and tasks, which leads to better communication of their needs and helps ensure that a system is implemented successfully. (high organizational & high project)
Team skills – People are important when implementing a system and can directly affect its success or failure.
Source systems – A good system makes it possible to load data into the warehouse properly. (technical)
Development technology – The hardware, software, methods and programs used in completing a project. (technical)

This impacts in turn the perceived net benefits from the use of the data warehouse.
Information systems success: a high level of data quality/system quality is associated with a high level of perceived net benefits
According to the findings, having resources, appropriate people on the team, and user participation have positive effects (significant effects) on the project's outcome.

The relativity of success **also** depends on "who" you ask within an organization, "when" you ask the question, the success dimensions and the ES implementation.
There are five different success dimensions:
1. Management success: related to performance
2. Project success: extend to which it is delivered in correct time and budget
3. User success: extend to which users have positive attitude towards the system
4. Correspondence success: extend to which there is a fit between system and business processes
5. System success: data accuracy, maintenance and reliability

There are four phases of ES implementation:
1. Chartering phase (defining what the system should look like)
2. Project phase (building the system)
3. Shakedown phase (system is implemented)
4. Onward and upward phase (receiving benefits)



To achieve BI success some aspects should be right (DQ, User access, Flexibility etc). In order to get those aspects 'right' it is necessary to make some decisions.
A DW with good data/system quality improves the way data is provided to decision-support applications and decision makers.

<u>(Data-driven) Culture</u>
To be able to transform data into knowledge and into results → build an analytic capability. You need a capability to aggregate, analyse, and use data to make informed decisions → action and generate real business value. In order to achieve this you need technical/human capabilities. Eventually this → Data-driven decision making.

Data to knowledge to results - MODEL
Davenport: most companies are not succeeding in turning data into knowledge and then results. Why? Have neglected the most important step in the data transformation process – the human realm of analysing, interpreting data and then acting on the insights. Ignoring the organizational, cultural and strategic changes necessary to leverage their investments.

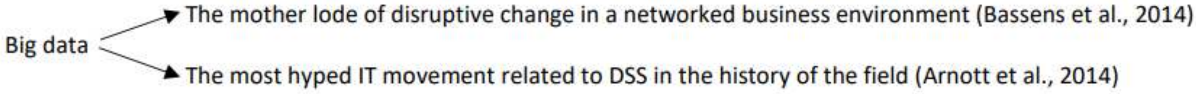
You need to have a solid foundation (*context* = bottom of the pyramid): prerequisites of success in this process: the strategic, skill, organizational & cultural, technological and data related factors that must be present for an analytical effort to succeed (continually refined and affected by other elements). *Transformation*: where the data is actually analysed and used to support a business decision. *Outcomes*: changes that result from the analysis and decision-making: behaviours, process and programs, and financial conditions.

Fast Data is the ability to gain insights from (near) real-time data streams and derive value from these insights.
You can also say that it is '*Big Data in extrema*'. *FD is caused by the increasing number of sensors & data sources.* As known, the Big Data 'Vs' are: Volume, Velocity, Variety, Veracity, (Variability), (Value). It is important because: source of value for organizations, rapidly changing organization environment, inversing customer expectations, competitive advantage: Data-driven business innovation.
An organization that responds successful to fast data, needs to be built on four pillars:
1. Technology: you need to analyse and process the data as events, need to process incoming fast data by using splitting and filtering, maintain 2 separate DB (historical and fast) and combine these for event recognition.
2. Strategy: you need to define a way to respond to the fast data → You need to have a clear strategy for the required data; define what real-time data means; determine the length for 'windows' for incoming data streams; and you need to be able to adapt business rules at every moment.
3. Culture: in order to be able to adapt at every moment; you need a *data-driven culture* and an agile culture. Let employees practice/learn with FD decision, give them autonomy to respond and be prepared for rapid changes based on FD.
4. Skills and experience / knowledge: people need the knowledge and experience with the technology, algorithms and pattern meaning, the data, the organization, and the strategy. You have to be able to transfer the data and the patterns that are found.

LECTURE 6 - DATA DRIVEN BUSINESS INNOVATION
"How can we use data to improve our business and innovate our business?"

<u>Data-driven business model innovation: WHY</u>

Why does it make sense for organizations to think about data-driven business innovation? / How do organizations realize value from Big Data? To answer → discuss role big data.

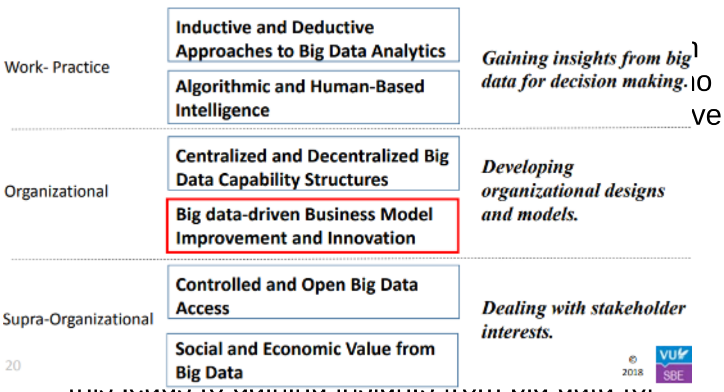


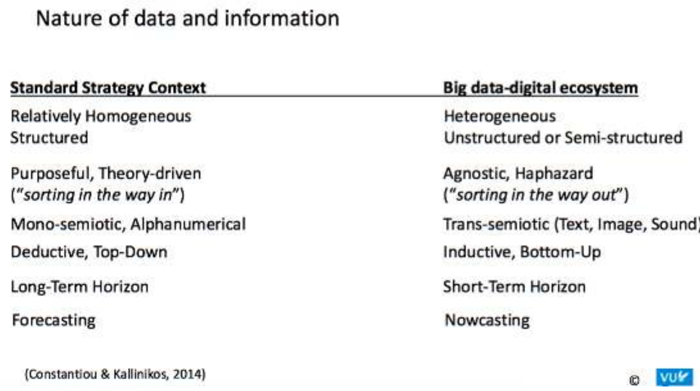
The new industrial revolution is driven by data analytics (cyber-physical streams). It is related to capturing value from the Internet of Things and Big Data.
Big Data is not a source of value in itself, and it has two specific features:

1. **Portability**: Possibility of transferring/remotely accessing digitized data from one context of application to be used in other contexts. Can be used in other country/context, use information
2. **Interconnectivity**: Possibility to synthesize various data sources. Data can be put in a new context and becomes new information; data can be connected to another derive from data sources. Can become other data when used in different context/when you ask a different question.

Our future will consists of six Ds:

1. Digitized: Technology can be represented as ones and zeroes
2. Demonetized: exchanging products and services for money
3. Democratized: involvement in decision-making
4. Delocalized: processes are time and place dependent
5. Deceptive: Exponential growth is deceptive. – there is a period during which exponential growth goes mostly unnoticed, until suddenly it becomes visible.
6. Disruptive: it might disrupt industries (Uber might become an advisory company because they know exactly where traffic is, and this might be interesting for policy makers/other parties)
7. Datafied (additionally added).





Nature of data and information

Standard Strategy Context	Big data-digital ecosystem
Relatively Homogeneous Structured	Heterogeneous Unstructured or Semi-structured
Purposeful, Theory-driven (*"sorting in the way in"*)	Agnostic, Haphazard (*"sorting in the way out"*)
Mono-semiotic, Alphanumerical	Trans-semiotic (Text, Image, Sound)
Deductive, Top-Down	Inductive, Bottom-Up
Long-Term Horizon	Short-Term Horizon
Forecasting	Nowcasting

(Constantiou & Kallinikos, 2014)

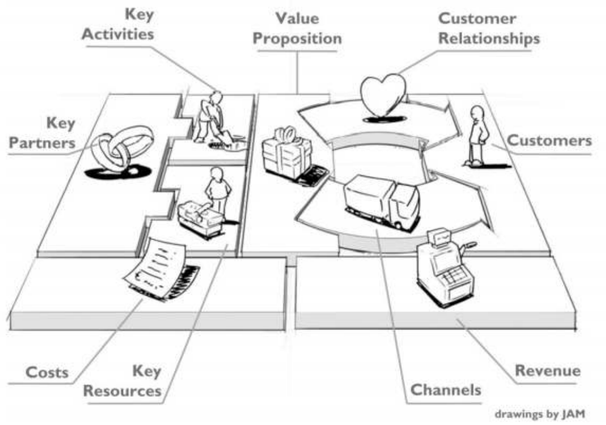
this leads to gaining insights from big data for decision making.
On an organizational level you talk about big data-driven business model improvement and innovation, and this leads to the development of organizational designs and models.

<u>Data-driven business model innovation: WHAT</u>
A **business model** articulates how an organization creates value for its customers and appropriates value from its markets.

Business model framework by Johnson, Christensen et al. (2008):
1. *Customer value proposition* includes important problem/need satisfied by offering for the targeted 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*

Building blocks business models:



- ➢ Customer Segment: Different groups of people or organizations aims to reach and serve.
- ➢ Value Proposition: Value created for customers through the offering. Describes the products and services that create value for a specific customer segment
- ➢ Channels: The ways organizations communicate/reaches its customer segments to deliver a value proposition.
- ➢ Customer Relationships: Describes relationship types a company establishes with specific customer segments.
- ➢ Revenue Streams: The cash a company generates from each customer segment (earnings= revenues -/- costs).
- ➢ Key Resources: Most important assets required to make a business model work.
- ➢ Key Activities: Most important things an organization must do to make its business model work.
- ➢ Key Partnerships: The network of suppliers/partners/stakeholders that make a business model work.
- ➢ Cost Structure: All costs incurred to operate a business model.

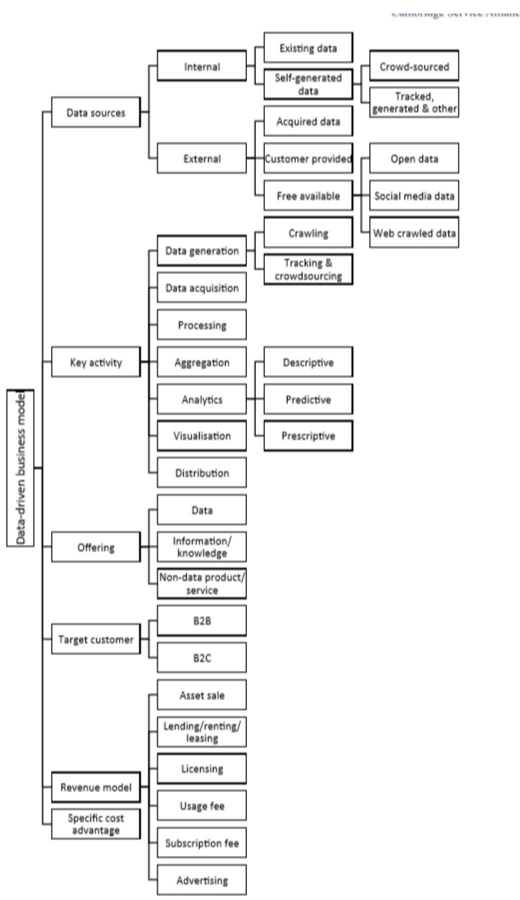
Hartmann et al (2014) - Taxonomy of data-driven business models used by start-up firms.
The paper suggests a catalog of business models used by start-up firms that rely on data as a major resource of their business. It proposes a DDBM framework to systematically analyze/compare DDBMS in the start-up world..

The term data-driven business model is commonly used by practitioners. This definition has 3 implications.

1. It is not limited to companies directing analytics, it 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. Kinsa, for instance, sells thermometers for the iPhone and provides a service to constantly monitor the body temperature.
3. It is obvious that company uses data in some way to conduct business, the size does not matter at all.

A data-driven business model is a business model that relies on data as a key resource. This business model consists of six dimensions that are common to most business model frameworks:

1. Key resources (data sources) - internal and external 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):

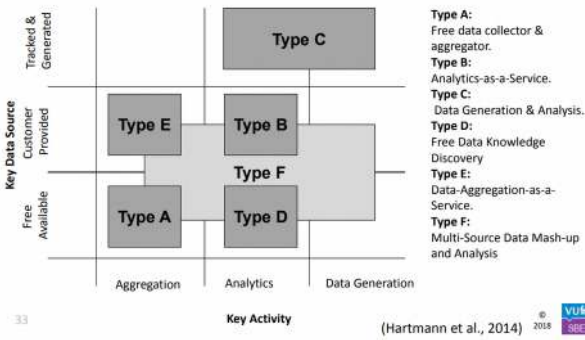


igure 1 The data-driven business-model framework (DDBM)

    - 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 (analytic activities): each company performs different activities to produce and deliver its offering: data generation, data acquisition, processing, aggregation, analytics, visualization and distribution
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 - B2B or B2C
5. Revenue model: important in order to survive long term - asset sale, lending/renting/leasing, licensing, usage free, subscription free and advertising
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.

These dimensions (key activities and key resources) can be presented in a 3x3 matrix and lead to 6 types of companies:

- Type A: *Free data collector and aggregator*: Create value by collecting and aggregating data from a vast number of different, mostly free available data sources
    - Key activity = aggregation
    - Key data source = free available
- Type B: *Analytics-as-a-service*: characterized by conducting analytics on data provided by their customers. Further other activities include data distribution and visualization of the analytics results.
    - Key activity = analytics
    - Key data source = customer provided
- Type C: *Data generation and analysis*: companies share the common characteristic that they generate data themselves rather than relying on existing data
    - Key activity = analytics and data generation
    - Key data source = tracked & generated
- Type D: *Free data knowledge discovery*: companies are characterized by the use of free available data and analytics performed on this data.
    - Key activity = analytics
    - Key data source = free available



- Type E: *Data-aggregation-as-a-service*: companies create value neither by analysing nor creating data but through aggregating data from multiple internal sources for customers
    - Key activity = aggregation
    - Key data source = customer provided
- Type F: *Multi-source data mashup and analysis*: 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 (free) external data sources to enrich or benchmark customer data.

- o Key activity = aggregation, analytics and data generation
- o Key data source = free available and customer provided

The framework allows identification and assessment of available potential data sources that can be used in a new DDBM and it provides comprehensive sets of potential key activities as well as revenue models.

<u>Data-driven business model innovation: HOW</u>

Parmer et al. (2014) - New patterns of innovation.

The search for new business ideas and new business models may cause failures in most corporations, despite the amazingly pressure on executives to grow their businesses. Management scholars have considered reasons for this failure. It is important to consider the next questions: *How can we create value for customers using data and analytic tools we own or could have access to?* The researchers noticed the advances in IT facilitate in the hunt for new business value in 5 distinct (maybe overlapping) patterns. Those patterns form the basis of the framework and by examining the patterns methodically, managers in most industries can conceive solid ideas for new businesses.

There are five patterns that can be used to find *new ways of innovation* based on (big) data:

1. *Augmenting products to generate data*: Use data that physical objects generate to improve a product or service to create new business value (sensors/wireless communities/big data). Example: Rolls-Royce's engine health management capability. Their new sensor technology and data management allowed them to identify airplane engine problems at an early stage so maintenance and repair schedules could be optimized and the engine design could improve.
2. *Digitizing (physical) assets*: Digital models (processes/products) and management of digitization (scoping/switching up). Example: Spotify/Uber. It slashes distribution costs and makes it able to move physical inventory efficiently of secure favourable store locations less critical. Offering customers more choices / tailored service will become increasingly important.
3. *Combining data within and across industries*: Due to *portability* and *interconnectivity*, data can be transferred to "our" business and combined with other data (smart cities). Example: IBM developed a network of sensors in the home that monitor not only conditions like temperature etc but also abnormalities (if someone falls etc).
4. *Trading data*: Selling data (telephone provider and TOMTOM). TOMTOM buys traffic from Vodafone, so that they can pinpoint where traffic jams are.
5. *Codifying a distinctive service capability:* Selling of perfected and standardized processes to other companies (processes that are best in class but not central to a company's competitive advantage are sold). Example: A major UK catalog retailer has developed an efficient/agile system for designing/producing an online catalogs. This lets it offer a much bigger range of products while maintaining less than half the stock of competitors. Standard process which they can sell to others.

Companies that had successful initiatives have four things (success factors) in common: strong technology presence, inputs from external parties, motivated leadership and emotional commitment. These patterns are a helpful way to structure a conversation about new business ideas. Actual initiatives often include 2/3 of the patterns, some examples involve more than one pattern.

Woerner and Wixom (2015) - Big data: extending the business strategy toolbox.

Digitization creates challenges, for most companies it is unevenly distributed through the organization, this makes it difficult to combine/simplify the increasing amount of data caused by digitization. Deriving insight and proceed on it based on that becomes difficult too. The 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.

When you improve the business model, you use big data to refine and optimize business processes and decision-making. You can do this in three ways:

1. New data: Used to be difficult for manufactures to feed data into internal product development. When they started using sensors, mining social media, and gathering customer-produced contend this is no longer the problem.
2. New insights: Big Data analytics can be used to identify outliers based on data inconsistencies.
3. New actions: As companies become well-armed with big data and proficient at making insights based on that data, they will act differently: faster and more wisely
   New data + new insights = new actions

When you innovate the business model, *you look for new ways to generate revenues*.
You can do this in two ways:

1. *Data monetization:* exchanging information-based products and services for legal tender or something of perceived equivalent value
    1. Selling: health apps that sell data to retail stores; the apps know the customers better
   → app owners can advertise the retail store which products to sell.
    2. Bartering: discount on car insurance when you share data from your driving
   Behaviour. Trade information in return for new tools, services or special deals.
    3. Wrapping: wrapping data/information around existing products (download an app
   when you are skiing and share your skiing behaviour; this can give you information about which slopes you need to use)

b. *Digital transformation:* When companies leverage digitization to move into completely new industries or create new ones. Emergence of business ecosystems, coordinated networks of companies, devices, and customers that create value. The competition is not defined by traditional industry offerings but by shared customers.
   This does not just increase the new/different types of data, it may also shape company boundaries so it becomes more difficult to tell where a partner organization commitment begins/ends. Big Data is not necessarily an obstacle to a strategy as it offers rich/exciting opportunities to leverage/extend the toolbox of a business of a company. Big data itself does not create strategic issues but when generating business value from it.

- Digital twins: the engine from Verstappen has a digital twin in the pit lane, and due to portability and interconnectivity this data can be shared all over the world → tesla heft dat ook.

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. → Watson, asking question shows a model.

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.