DATA ANALYTICS AND BIG DATA

A DEEP ENOUGH EXPLANATION FOR NON-TECHNICAL ROLES

PART 1

THE BASICS

OR HOW TO BE ABLE TO MAINTAN A CONVERSATION WITH GEEKS WITHOUT FEELING EMBARASSED

By TECHBIZDESIGN.COM

The basics are

About data

The importance of data, types of data, formats, types of Databases…

About analysis

Types of analysis (Diagnostic and descriptive, predictive, prescriptive, experimental…)

Analytical domains

Portals, workbenches, labs…

Analytical capabilities

Reporting, Dashboarding, Data Science…

Roles in Data Analytics

Data produced by IoT devices today

A modern fully instrumented car

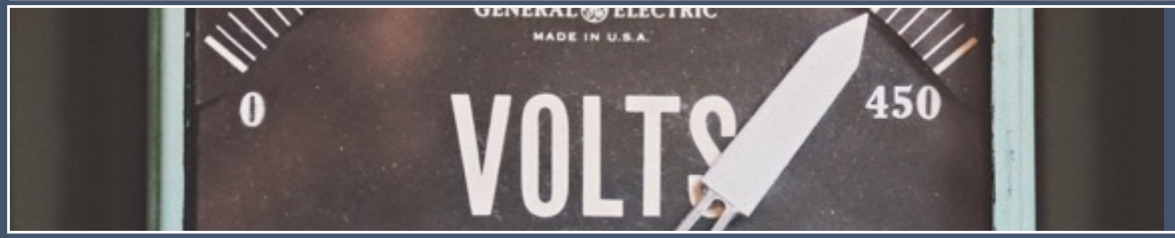
25 Gb per hour

A typical wind farm

150k Data points per second

A fully instrumented jet engine

51,2k Gb per hour



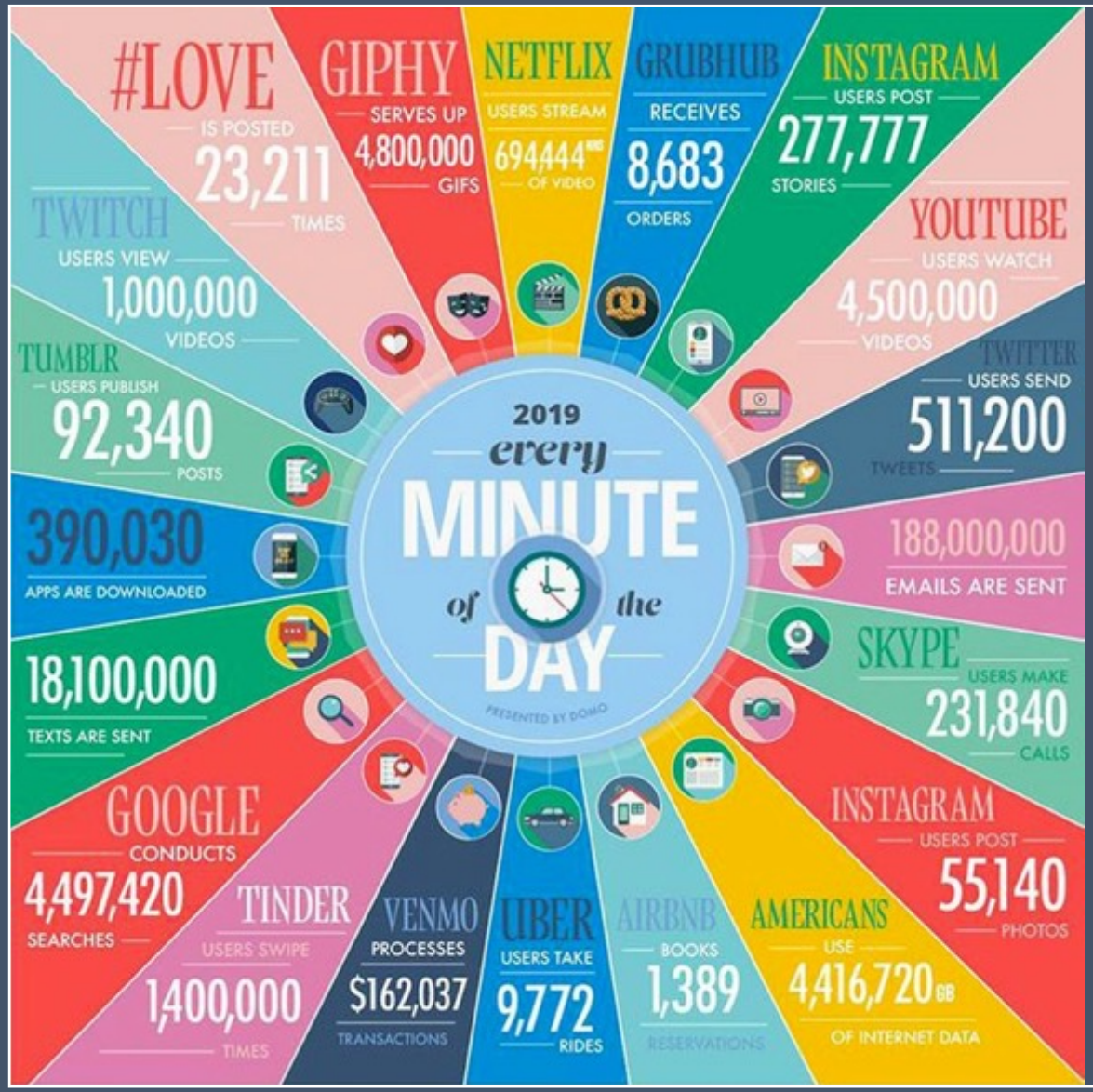
A smart meter project

500M Data readings per day

A single turbine compressor blade

500 Gb per day

What's happening on the net in just one minute today



This activity means:

Internet users generate about **2.5 quintillion bytes of data** each day (as example, 4 new petabytes are generated by Facebook every day)

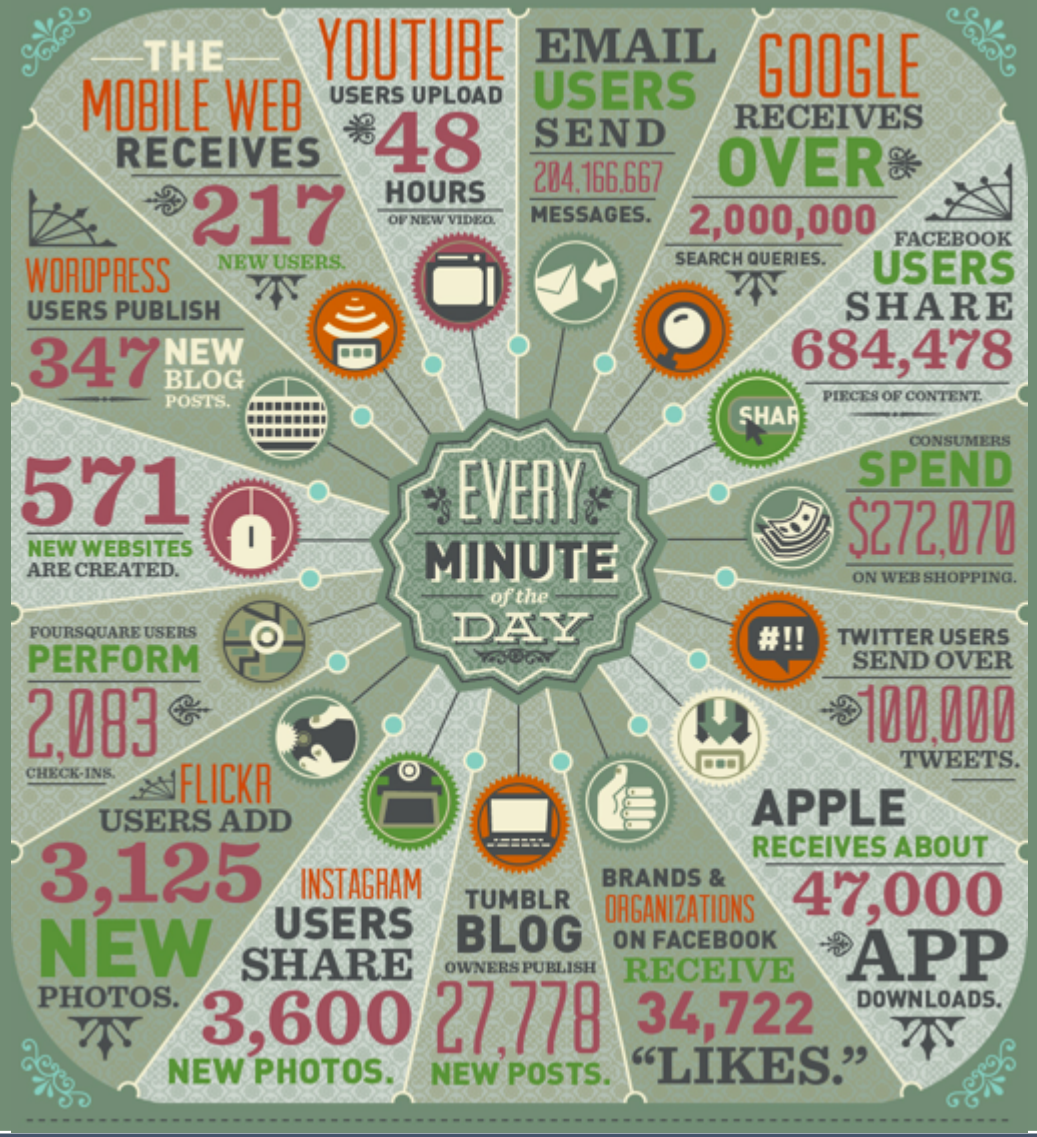
90% of all data has been created **in the last two years**

By 2020, there will be around **40 trillion gigabytes** of data (40 zettabytes)

By 2020, every person will generate **1.7 megabytes in just a second**

Source: https://www.domo.com/learn/data-never-sleeps-7

What's happening on the net in just one minute in 2012



The thing is in **2012** only 0.5% of all this data was analyzed

But **today**…

97.2% of organizations are investing in big data and AI

Using big data, Netflix saves $1 billion per year on customer retention

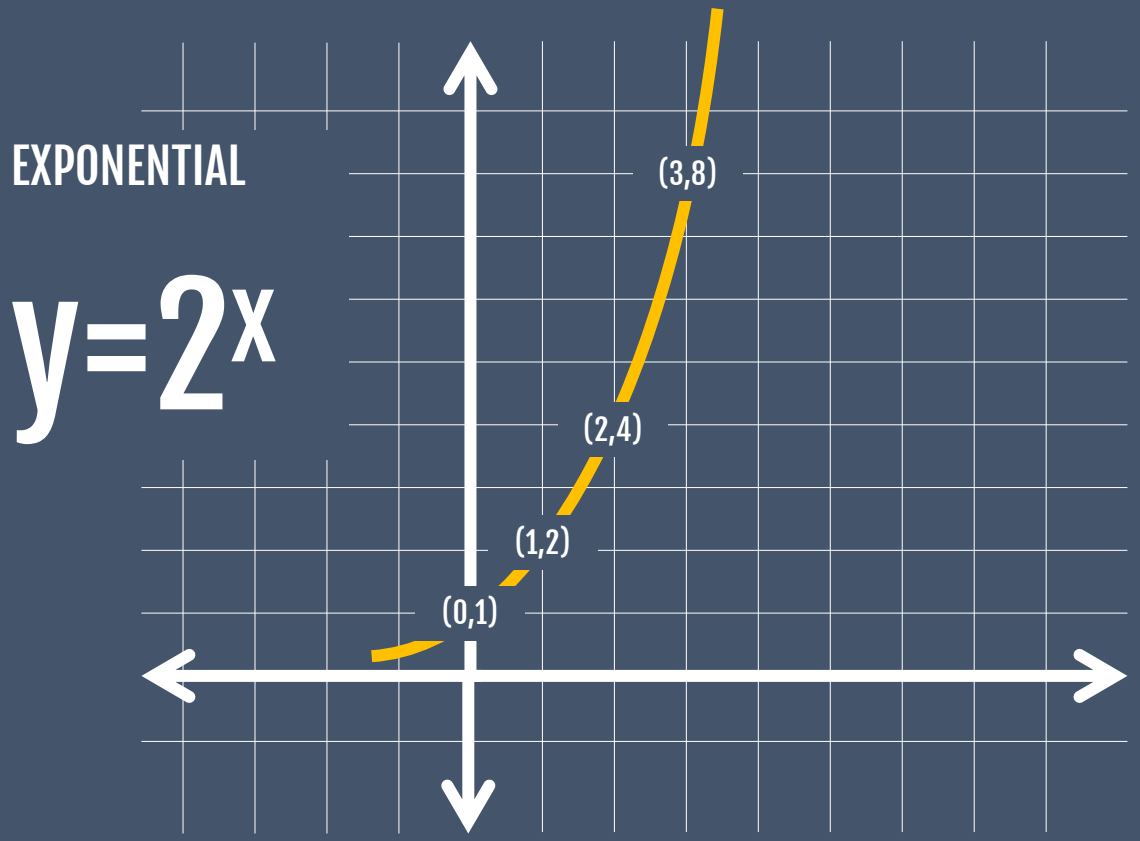
Job listings for data science and analytics will reach around 2.7 million

Businesses will gain $430 billion if they opt for a data-driven approach

Source: https://www.domo.com/learn/infographic-data-never-sleeps#/

But most probably you know that already. Don’t you?

Maybe you don't know it's not a matter of volume. It's just about velocity and business...



Have you ever heard about **Moore's Law**? The number of transistors in a dense integrated circuit doubles about every two years.

In the same time, reduces communications cost half, doubles communications speed or doubles number of servers on Internet.

And every 18 months, doubles our personnel devices' RAM.

Because we live in the "Data age"

Between 1999 and 2009 we treated information as **CONTROL**

Our focus was reporting, predictable things, localized data and finance-driven insights

Between 2010 and 2020 we are treating information as **FUEL**

Our focus is create live intelligence, embed it in our IT systems and enterprise-wide usable

From 2011 we will treat information as **ASSET**

Our focus will be to generate algorithmic business, spread knowledge and make it ubiquitous, build ecosystem and act collaboratively across businesses and governments

Source: Gartner public info and Internet references

Ok then. So, what is DATA?

Structured DATA

Our traditional Excel files and databases

Semi Structured DATA

A paradise for developers

Unstructured DATA

Videos, images, texts, tweets, logs…

Structured DATA

Structured data types are mainly delimited text files like CSV format, tabular data files like Excel and databases

Data respond to an abstract **data model** that organizes principal elements in entities and defines their relations

There's a semantic content assigned to each element according to its relationship with other entities

There has been a previous modeling and formalization process over the data

Semi Structured DATA

Semi structured data types are mainly JSON and XML

Information is structured using **categories or tags** to separate different elements into the doc. Semantic relations are defined on these categories

Data is stored on registers easily identifiable. This characteristic make them flexible and easy to understand for developers .

There's a risk of ambiguity and inconsistency related to this data type

```
{
   "quality":        {
                     "noise_perc": 22
                     "quality_txt": "good"
                     }
   "values":         {
                     "probability": 85
                     "probability_txt": "high"
                     "intensity": 65
                     }
}
```

Unstructured DATA

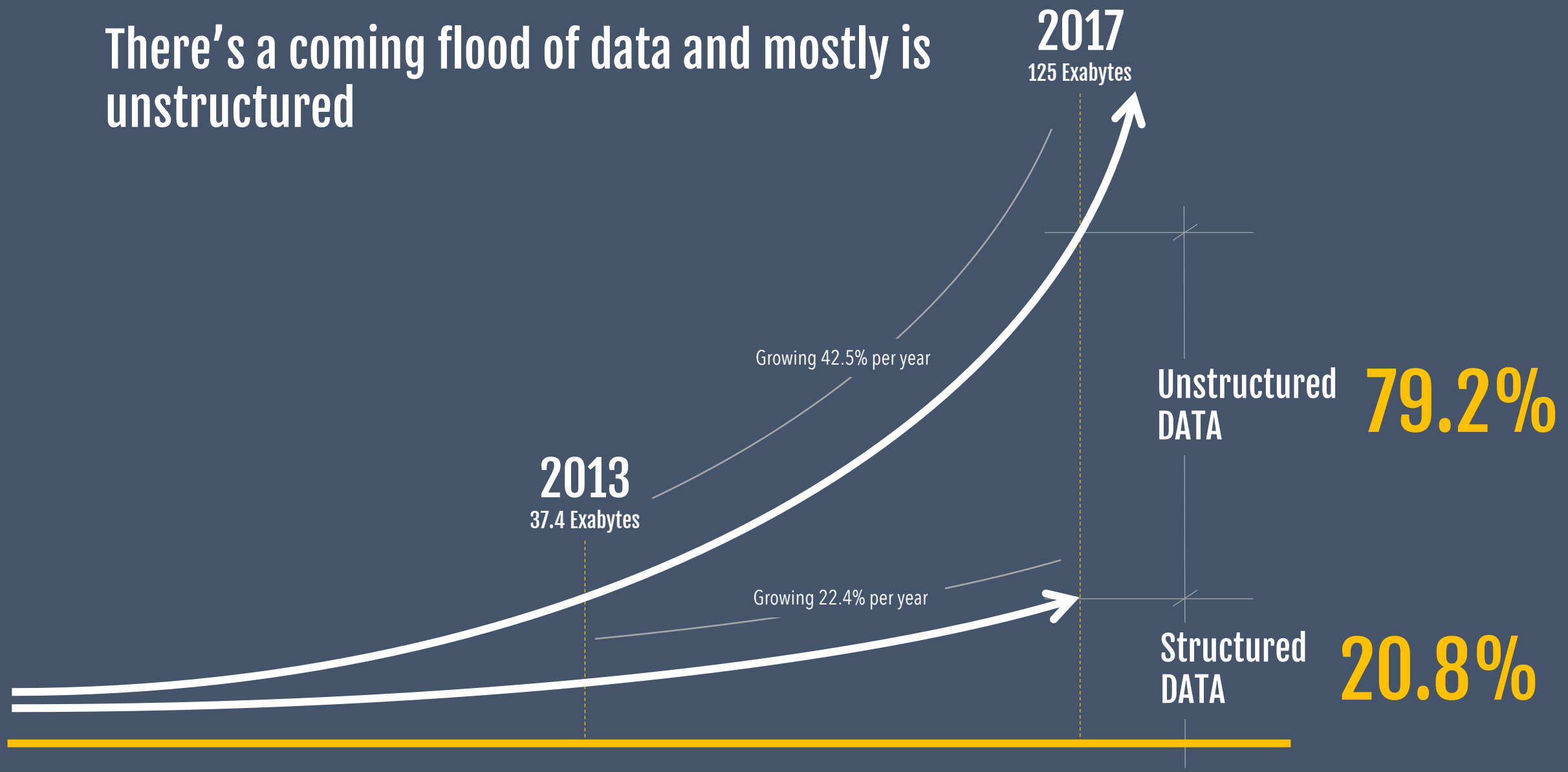
There are many unstructured data types: plain text files and documents (Word, txt…), server, website and application logs, sensor data (Satellite, meteo…), images, video files, audio files, eMails, social media data…

There's **no predefined data model** in those different data types. So, there's no general and formalized way to extract the data

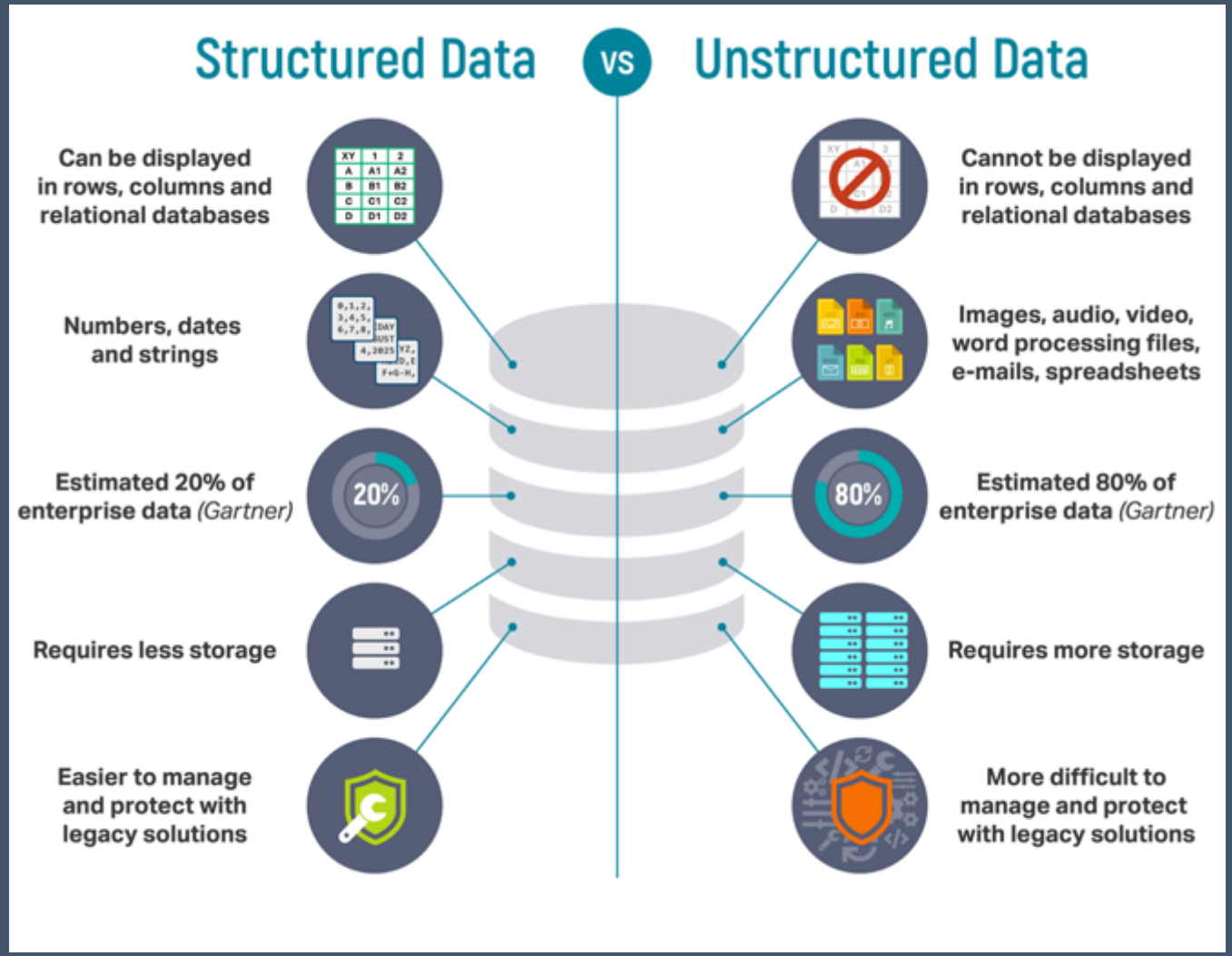
Different data types in the same doc or file, normally with an unknown semantic relation between the different elements

Huge amounts of this type of data are being generated exponentially. They represent the majority of data repositories worldwide

There's a coming flood of data and mostly is unstructured



Remember...



Source: https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/

Database: An organized collection of data, generally stored and accessed electronically from a computer system.

Relational Database – SQL

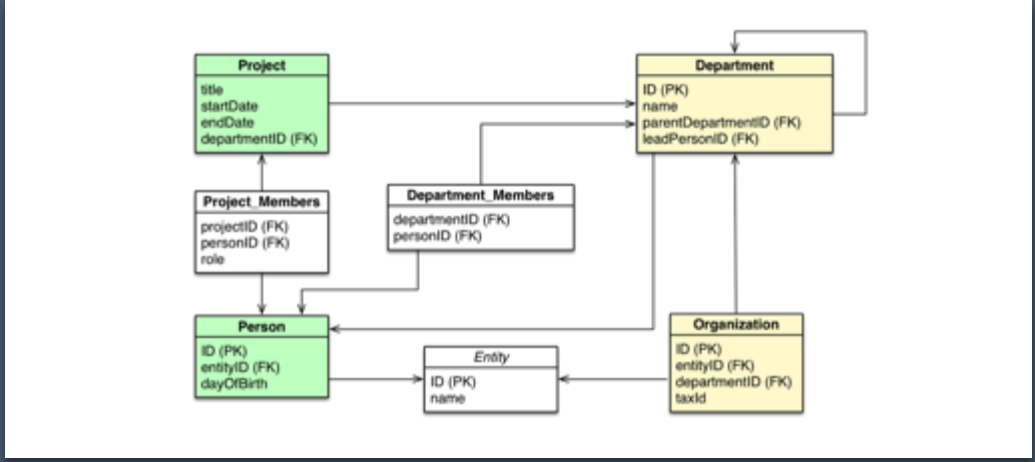
Generally queried via T-SQL for ask questions of the data in a very well-understood way

Indexed which optimizes the query of the data or makes it faster

Transactional oriented which keeps the data in a consistent state

Normally, expensive to scale-up and not designed for real-time

Fundamentally designed to host relational data, not other types of data



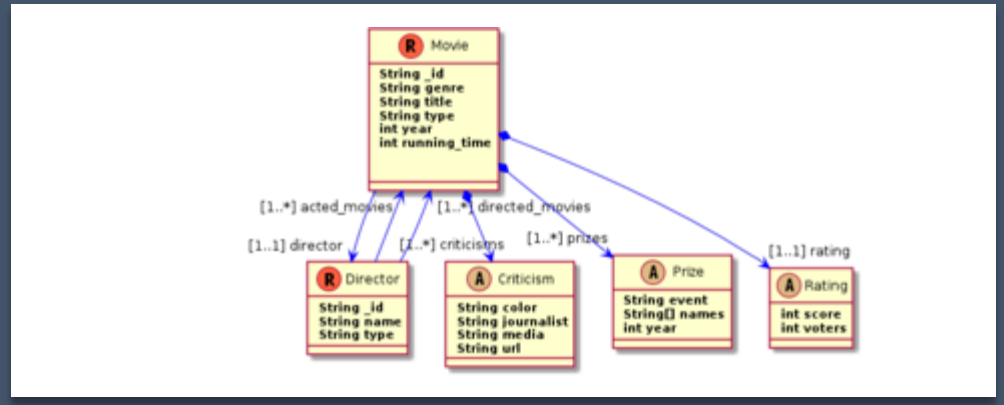
NoSQL

A set of database solutions that are designed to host and work with non-relational data at its core

Good alternative to relational databases or Hadoop (There are around 150 different types of NoSQL databases)

Often focuses on high scalability, high availability and eventual consistency, and designed initially by Google and AWS to store high volumes and large variety

Good for development but uncomfortable for IT once in production



Database: An organized collection of data, generally stored and accessed electronically from a computer system.

Relational Database – SQL

RDBMS two different flavors, OLTP and OLAP

OLTP or **Online Transactional Processing** oriented to operational systems designed to take in new data, inserts, updates, and deletes. Normalized data.

OLAP or **Online Analytical Processing** oriented to reporting systems designed and optimized to read data results out often aggregated or summarized data results. Unnormalized data.

NoSQL

There are different categories for NoSQL databases

We have Key/Value volatile, Key/Value persistent, Wide-column, Document or Graph

PROS

Easy to use and setup | Universal, compatible with many tools | Good at high-performance workloads | Good at structure data

No investment to design model | Rapid development cycles | In general, faster than SQL | Runs well on the cloud

CONS

Time consuming to understand and design the structure of database | Can be difficult to scale

Unsuited for interconnected data | Technically still maturing | Can have slower response time, but improving faster

Source: https://www.quora.com/What-are-the-differences-between-the-SQL-and-NoSQL-databases

I got it. How do we should manage data to get value?

The DATA LIFE CYCLE or how to get knowledge and value from data

DATA SOURCES

Integrate data sources

- . Sensors
- . Mobile
- . Logs
- . Tabular data
- . Social
- . Telemetry
- …

INGEST

Build Data Pipelines

- . REAL TIME
- . NEAR REAL-TIME
- . BATCH

PREPARE

Process and transform data

- . Data CLEANSING
- . Data WRANGLING
- . Data QUALITY
- . Data ENRICHMENT

STORE

Store data

- . Data LANDING AREAS
- . Data LAKES
- . EDW (Enterprise Data Warehouse)
- . Data VIRTUALIZATION (Logical Datamarts)

DISCOVER

Exploratory analytics

Dashboard and reports

Statistical modelling

Machine learning

- . Information PORTAL
- . Analytics WORKBENCH
- . Data Science LAB
- . Operational Intelligence

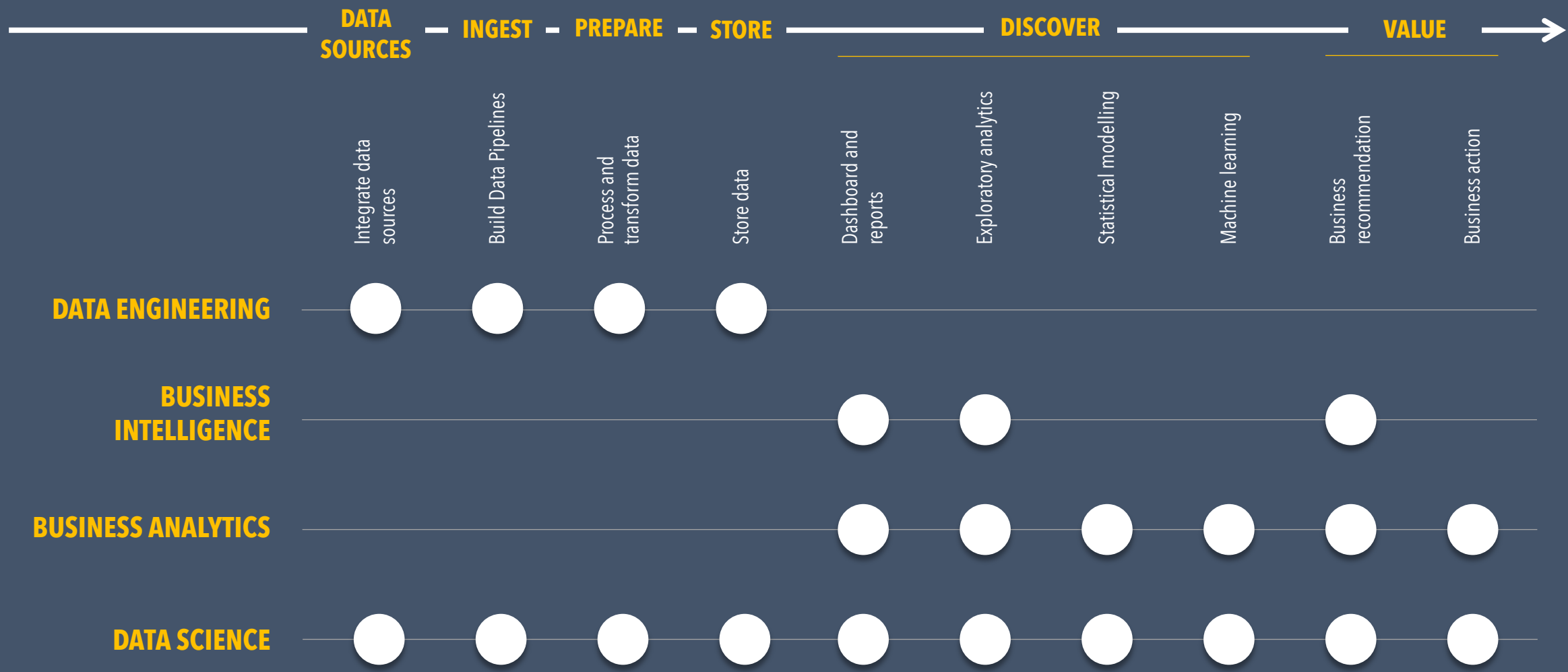
VALUE

Business recommendation

Business action

GOVERNANCE | Security, maintenance, quality, operations, lineage…

The DATA LIFE CYCLE and Data Analytics disciplines



Interesting. How can I get from Data? What kind of analysis I can do?

Types of Data Analysis

1. Descriptive

The goal of this analytics is to **present numerical and summarized facts** about the performance of the business in the past . It helps analysts understand the events that happened during the past period.

This is the earliest form of analytics also called "reporting" that summarizes data to understand how did in a given time period

His goal is to compare different segments and time periods.

Presented normally predefined and pre-canned

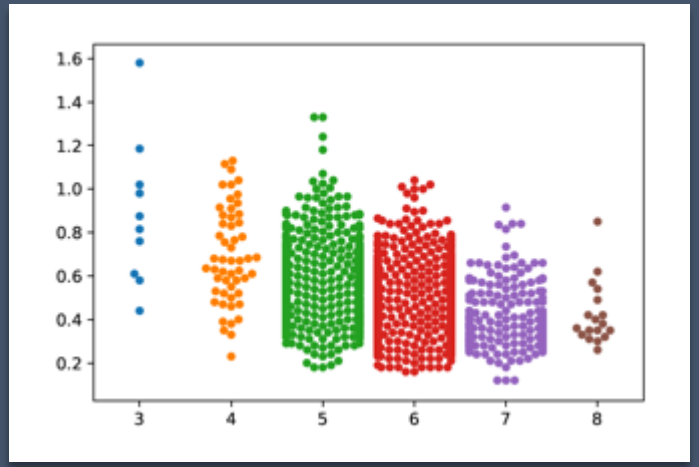
2. Diagnostic (Exploratory)

This is essentially a **deep dive into the data** in an ad-hoc, yet structured manner to understand patterns and confirm hypothesis. The best analogy for exploratory analytics is a hound that picks up a scent and chases it.

It's the way to get familiar with the data and make a deep dive.

It's normally done in an ad-hoc manner by **analysts using SQL** and specific analytical tools (segmentation and profiling, graphical tools…) instead standard reports or dashboards.

It's not regularly done. It's needs base.



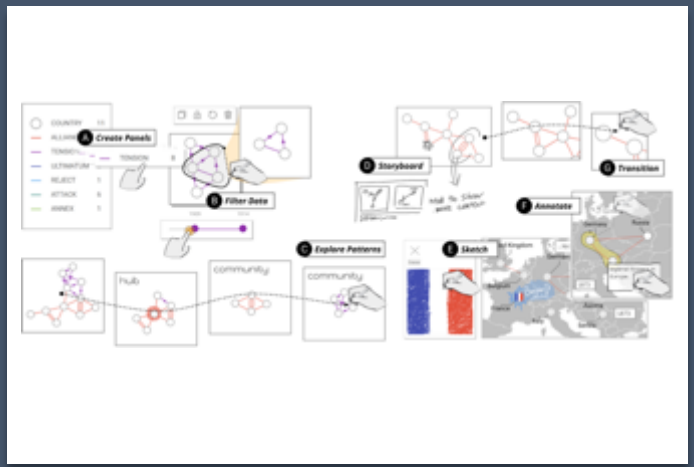
3. Diagnostic (Explanatory)

Explanatory analytics seeks **to tell stories** with data.

The differences with Exploratory are very few, except that here we focus on trying to find the root cause, not just patterns

It's just about answer questions and present to an audience, something normally done by **analysts and managers**

This stage is usually a prelude to next actions taken in the business based on business results and answers produced by analytics



4. Predictive

The goal of predictive analytics is to identify the **likelihood of future outcomes** based on historical data, statistics, and machine learning

It's fully data –driven prediction (not logic or intuition) that uses historical data to understand past performance.

Dealing with large quantities of data, required for accurate predictions

It's a kind of Analytics that uses **automation and machine learning** intensively

Capable to draw real-time predictions of trends

Source: https://www.sapanalytics.cloud/resources-predictive-forecasting/

5. Prescriptive

The goal of prescriptive analytics is to identify ways and means to **take advantage of the findings and predictions** provided by earlier stages (exploratory, explanatory, and predictive)

Intensively usage of patterns and predictions produced earlier requires additional analytics for better business outcomes

To start with this kind of analytics, budget, time and human resources need consideration. In addition, cost and benefits need to be evaluated

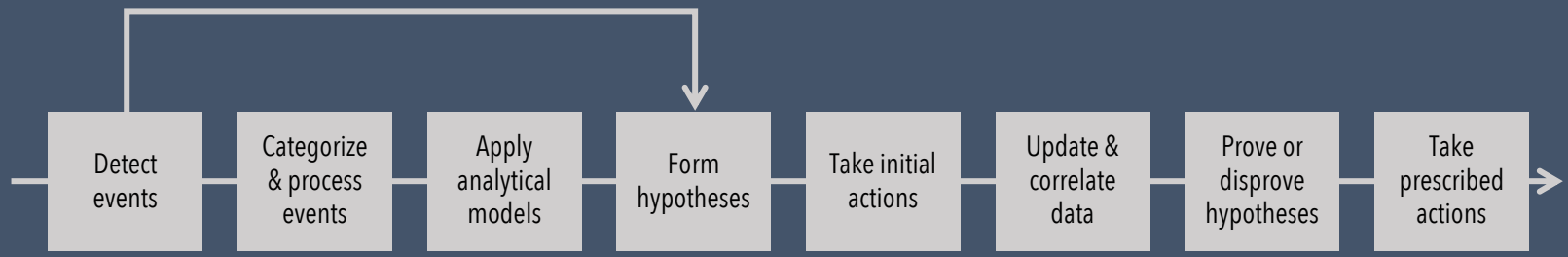
What kind of insights do we seek?

. Different **timeframes** | Past, present and future

. Different **precision** | Cold hard facts, strong hypotheses or fuzzy hypotheses

. Different **paradigms** | Specific questions, digging through data or alternate history

. Different **scope** | Organization-specific, process-specific or entreprise-wide



The prescriptive analytics recommended workflow

Its primary risks are:

. Available data isn't fully processed and analyzed

. Data processed but hypotheses aren't formulated

. Hypotheses not definitively proved or disproved

. Critical insights are formulated but not action taken

Other risks involved could be:

. Insufficient data management tech

. Immature software tools

. Unbalanced emphasis in projects on insights in the past

. Poorly architected systems

Source: Gartner public info and Internet references

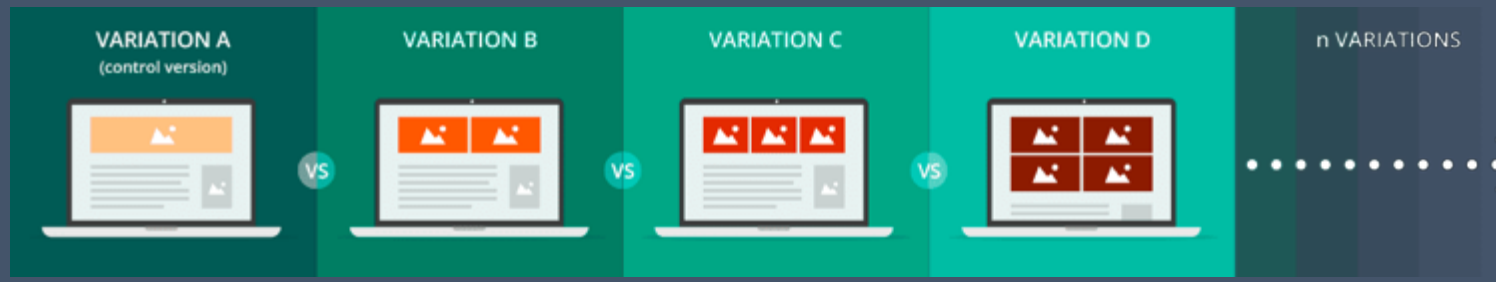
6. Experimental

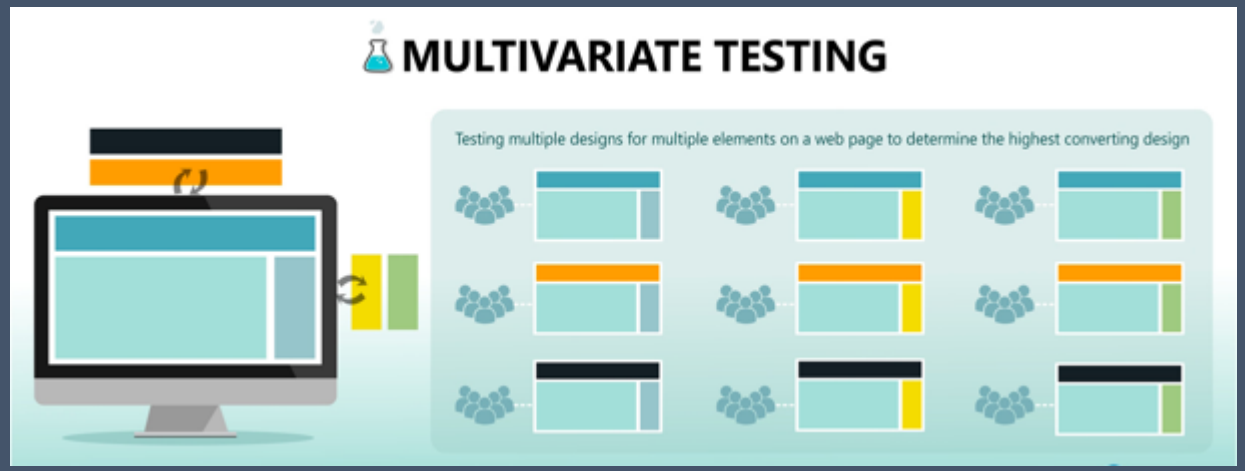
They key difference between prescriptive analytics and experimental analytics is, prescriptive analytics are used with simulation, within the enterprise. Experimental analytics deals with actual **experiments conducted on the field** with actual customers, or subjects

This kind of analytics pursues to implement plan on a subset and test multiple alternatives to find best results.

It helps a lot to identify environmental variables hitherto unknow

Normally we use tools and techniques like sample testing, A/B testing or multivariate testing, among others





Source: Gartner public info and Internet references

Which are the ways to interact with Data? How can I be able to apply those types of Analytics?

Navigate through the Analytics Domains

The information portal

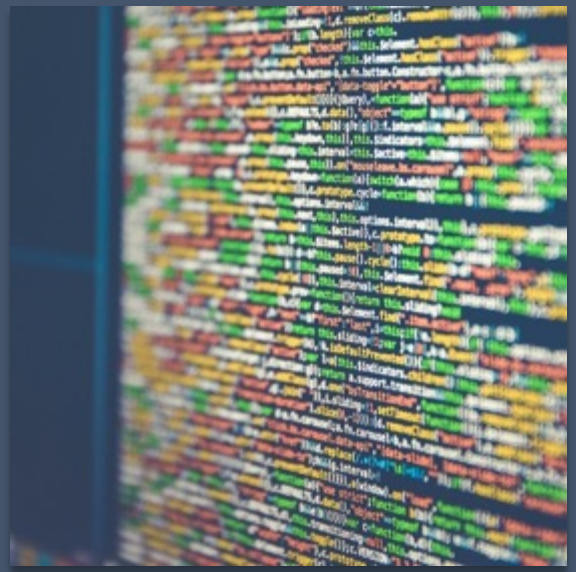
Provides trusted information to business users in the form of the **reports or dashboards** offered by traditional business intelligence (BI) tools

The analytics workbench

Empowers business users with the ability to autonomously produce and publish insights, mainly through self-service data preparation and **visual data discovery** tools.

The Data Science laboratory

Supports the business in the delivery of advanced analytics outputs, using predictive modeling, prescriptive analytics, machine learning and other **sophisticated analytics** capabilities.



The Artificial Intelligence HUB

Offers technology that appears to emulate human performance

Source: Gartner public info and Internet references

The information portal

It's usually driven by technology specialists like IT crew, who access data sources and produce and publish Business Intelligence content.

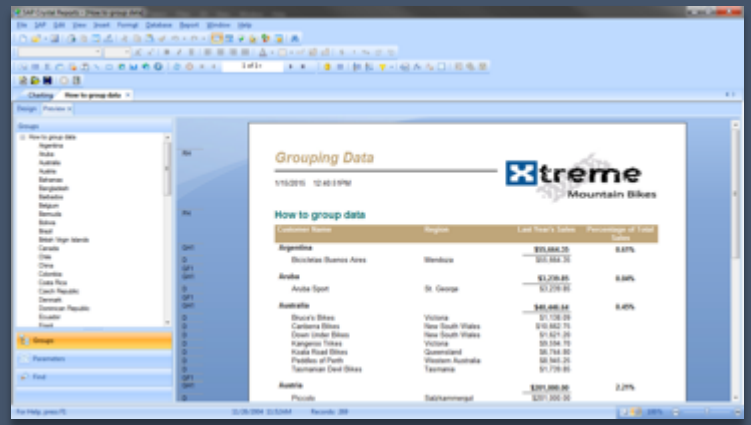
The analytics capabilities from this domain should be selected when reliable metrics are required and not others as agility or flexibility.

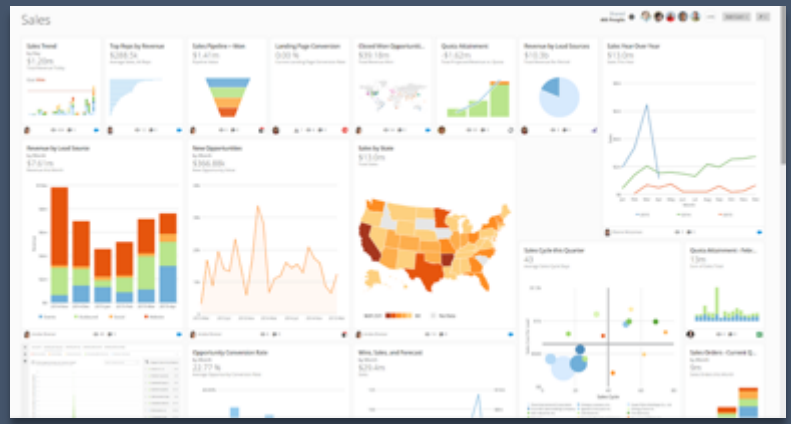
It's a domain heavy driven where autonomy to get the answers is not a decisive factor.

The **ANALYTICAL CAPABILITIES** here are Enterprise reporting, Dashboarding, OLAP – Online Analytical Processing, Ad Hoc Query, Mobile BI and Real-Time reporting – Continuous intelligence.

The business oriented roles involved in this domain are information consumer and Data Steward

The IT or tech specialists oriented roles involved in this domain are Data modeler, BI developer and Data Quality Manager





ENTERPRISE REPORTING

The Operational reporting are trusted, sanctioned and highly controlled production reports and dashboards. These are automatically distributed to large numbers of business users and external customers, or embedded in applications.

These kind of Analytic tool provides a multipage reporting view of data and embedded visualizations in an application

DASHBOARDING

Dashboards represent a style of reporting that aggregates and displays performance metrics (KPIs), enabling them to be examined at a glance by all manner of users.

They also display metrics using infographics such as dials, gauges and "traffic lights," which indicate progress toward defined targets. In some cases, dashboards are evolving to become "applications" on their own

Dashboarding is a powerful visualization tool for business users because provide single version of truth and communicate information in the context of a business process or user's role.

Source: Gartner public info and Internet references

The information portal

It's usually driven by technology specialists like IT crew, who access data sources and produce and publish Business Intelligence content.

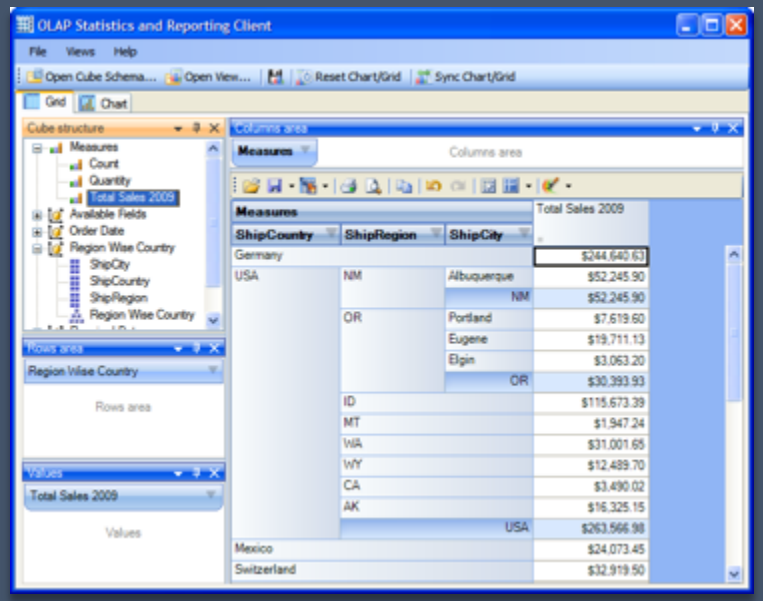
The analytics capabilities from this domain should be selected when reliable metrics are required and not others as agility or flexibility.

It's a domain heavy driven where autonomy to get the answers is not a decisive factor.

The **ANALYTICAL CAPABILITIES** here are Enterprise reporting, Dashboarding, OLAP – Online Analytical Processing, Ad Hoc Query, Mobile BI and Real-Time reporting – Continuous intelligence.

The business oriented roles involved in this domain are information consumer and Data Steward

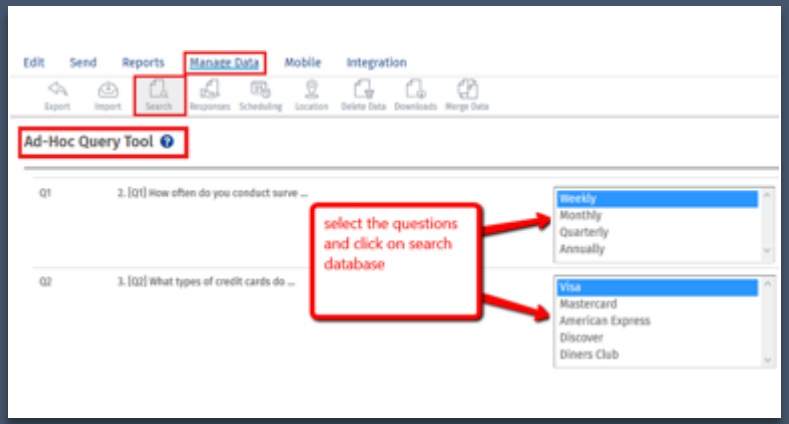
The IT or tech specialists oriented roles involved in this domain are Data modeler, BI developer and Data Quality Manager



OLAP – AnalytiCaL Processing

OLAP analysis enables users to analyze data that has been pre-aggregated to answer specific business questions in an OLAP cube with multiple dimensions to enable faster query and calculation performance.

It supports a style of analysis known as "slicing and dicing." Users are able to navigate multidimensional drill paths going deep down or up according to their needs.



AD-HOC QUERY

An ad hoc query solution provides a highly governed, self-service Business Intelligence module that allows developers or business power users to create their own queries, reports and charts from a browser.

They are able to create queries, reports and chart against a modeled data source. It may be aggregated or granular, with some latency and multiple subject areas. Output requirements may change

Source: Gartner public info and Internet references

The information portal

It's usually driven by technology specialists like IT crew, who access data sources and produce and publish Business Intelligence content.

The analytics capabilities from this domain should be selected when reliable metrics are required and not others as agility or flexibility.

It's a domain heavy driven where autonomy to get the answers is not a decisive factor.

The **ANALYTICAL CAPABILITIES** here are Enterprise reporting, Dashboarding, OLAP – Online Analytical Processing, Ad Hoc Query, Mobile BI and Real-Time reporting – Continuous intelligence.

The business oriented roles involved in this domain are information consumer and Data Steward

The IT or tech specialists oriented roles involved in this domain are Data modeler, BI developer and Data Quality Manager

MOBILE BUSINESS INTELIGENCE

Mobile business intelligence is the delivery of capabilities such as reports, dashboards, visual data discovery and basic analytics capabilities through mobile devices (tablets and smartphones).

This is an increasingly option in modern BI and analytics tools, providing opportunities to connect and explore data to create new insights everywhere with any device.

Navigation on these tools is achieved through touch- based interfaces that query remote or local data.

REAL-TIME REPORTING

This is a broader set of building blocks that include event stream processing tools, messaging middleware/event brokers and more diagnostic analytics tools that allow users to dive deeply into an event once it is detected.

These tools are 100% focused to provide operational intelligence for situations in which real-time data from the last few seconds or minutes significantly improves business decisions.

It's an obvious trend that Real-Time reporting is trying to offer prescriptive information about the best available action to be taken in response to the situation, closing the prescriptive cycle.

Source: Gartner public info and Internet references

The information portal

It's usually driven by technology specialists like IT crew, who access data sources and produce and publish Business Intelligence content.

The analytics capabilities from this domain should be selected when reliable metrics are required and not others as agility or flexibility.

It's a domain heavy driven where autonomy to get the answers is not a decisive factor.

The analytical capabilities here are Enterprise reporting, Dashboarding, OLAP – Online Analytical Processing, Ad Hoc Query, Mobile BI and Real-Time reporting – Continuous intelligence.

The **BUSINESS ORIENTED ROLES** involved in this domain are information consumer and Data Steward

The **IT OR TECH SPECIALISTS ORIENTED ROLES** involved in this domain are Data modeler, BI developer and Data Quality Manager

INFORMATION CONSUMER

This a common role through entire domains.

He represents the end user who gets value from analytic tools and capabilities and uses them to take data-driven decisions

DATA STEWARD

A data steward is a role within an organization responsible for utilizing an organization's data governance processes to ensure fitness of data elements - both the content and metadata

DATA MODELER

This role understand and translate business needs into data models supporting long-term solutions.

He works to implement data strategies, build data flows and develop conceptual data models

BI DEVELOPER

He's responsible for designing enterprise-level solutions for very large multidimensional databases and plan BI solutions.

He also creates and deploys reports and writes relational and multidimensional database queries

DATA QUALITY MANAGER

He's is responsible for coordinating activities to complete standards of quality and advise on how these quality systems are managed.

He reports performance and measure against set standards, and often sets policy standards

Source: Gartner public info and Internet references

The analytical workbench

The business plays the main role with IT enabling and supporting the infrastructure.

These analytics capabilities should be selected when the objective is to support user autonomy, agility and flexibility, without requiring advanced analytics skills.

The analytics workbench domain should be avoided when the use cases don't tolerate doubts around the trust level of insights created by users

The **ANALYTICAL CAPABILITIES** here are Visual Data Discovery, Location Analytics, Data preparation, Citizen Data Science, NoSQL based discovery, Stream Analytics, Graph Analytics and Embedded Analytics.

The business oriented roles involved in this domain are Analyst and Citizen Data Science

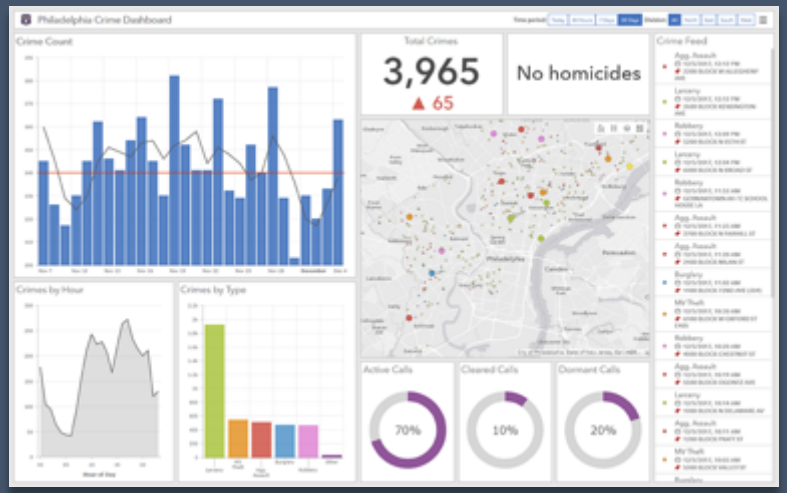
The IT or tech specialists oriented roles involved in this domain are analytics support expert and Data Engineer.

VISUAL DATA DISCOVERY

These powerful tools blend data from multiple sources into a proprietary in-memory store that is tightly coupled with an interactive visualization layer. They enable agile, rapid prototyping of interactive data visualizations.

It contrasts with the traditional BI platform, which relies on a more modular architecture and is dependent on three distinct technologies to integrate, store and present data.

Because is a combination of the in-memory analytics and interactive visualization technologies, Visual Data Discovery provide self- service analytics with deeper diagnostic analytics capability.



LOCATION ANALYTICS

These analytical techniques uncover previously unseen spatial/ location relationships, resulting in improved operational efficiencies and decision making.

Geospatial and location intelligence includes applications, infrastructure, tools and best practices that enable access to and utilization of geospatial and location data of people, things and information for location-referenced analysis.

Source: Gartner public info and Internet references

The analytical workbench

The business plays the main role with IT enabling and supporting the infrastructure.

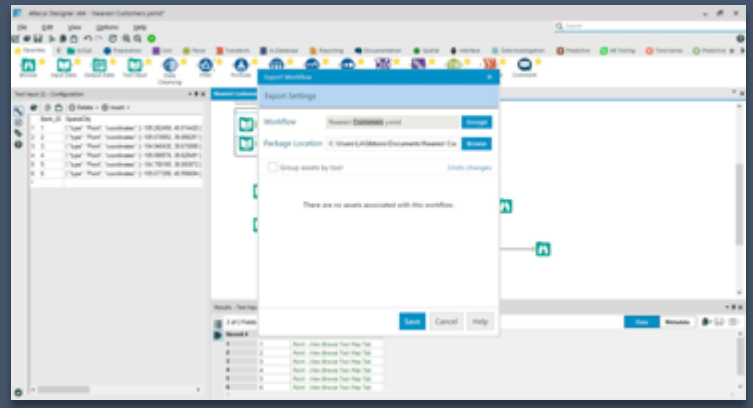
These analytics capabilities should be selected when the objective is to support user autonomy, agility and flexibility, without requiring advanced analytics skills.

The analytics workbench domain should be avoided when the use cases don't tolerate doubts around the trust level of insights created by users

The **ANALYTICAL CAPABILITIES** here are Visual Data Discovery, Location Analytics, Data preparation, Citizen Data Science, NoSQL based discovery, Stream Analytics, Graph Analytics and Embedded Analytics.

The business oriented roles involved in this domain are Analyst and Citizen Data Science

The IT or tech specialists oriented roles involved in this domain are analytics support expert and Data Engineer.

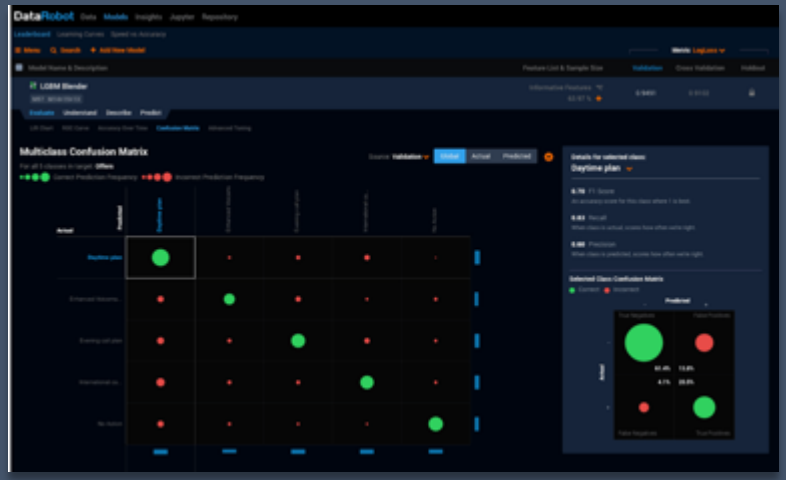


DATA PREPARATION

This is an iterative and agile process for exploring, combining, cleaning and transforming raw data into curated datasets for self-service data integration and data science.

Data preparation tools speed time to insight by allowing users to reduce the complexity of data preparation, find patterns in their integrated datasets and share their findings for further analysis, all without extensive IT support or coding knowledge.

It's a powerful because simplify access and collaboration on reusable components.



CITIZEN DATA SCIENCE

These kind of analytical capabilities enable business users to extract advanced analytics insights from data without the need for extensive data science expertise

Central to enabling citizen data science are rapidly progressing augmented analytics capabilities that streamline data preparation, provide user guidance for data science operations (correlations, clustering, predictions), augment user insights through automated modeling and pattern detection, and enable collaboration and sharing

Source: Gartner public info and Internet references

The analytical workbench

The business plays the main role with IT enabling and supporting the infrastructure.

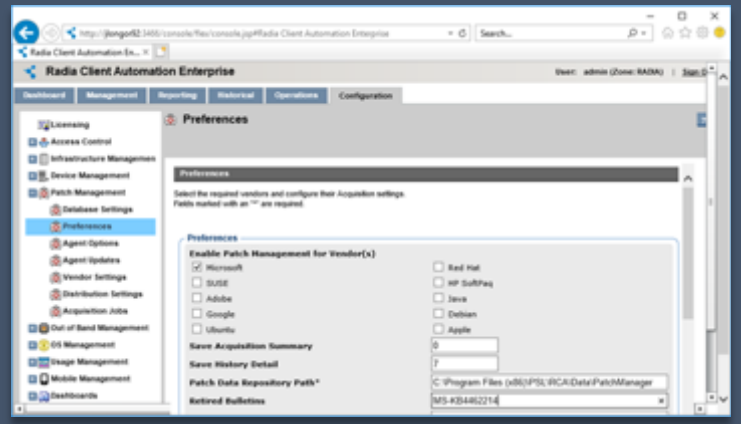
These analytics capabilities should be selected when the objective is to support user autonomy, agility and flexibility, without requiring advanced analytics skills.

The analytics workbench domain should be avoided when the use cases don't tolerate doubts around the trust level of insights created by users

The **ANALYTICAL CAPABILITIES** here are Visual Data Discovery, Location Analytics, Data preparation, Citizen Data Science, NoSQL based discovery, Stream Analytics, Graph Analytics and Embedded Analytics.

The business oriented roles involved in this domain are Analyst and Citizen Data Science

The IT or tech specialists oriented roles involved in this domain are analytics support expert and Data Engineer.



NoSQL BASED DISCOVERY

It helps to query and analyze data with Apache Hadoop and NoSQL data sources. Non-relational-based discovery products enable users to query and analyze data across data lakes built using object stores, distributed file systems and NoSQL data stores.

Products here typically have begun in the big data discovery sector and have since expanded to include support for relational data sources.

Key functionalities included are semantic modeling, performance enhancement and distributed processing

STREAM ANALYTICS

These tools are designed to detect threat and opportunity situations for decision support and decision automation systems

They're deployed for time series data visualization and enable users to easily create and customize dashboards that connect to streaming data sources.

Source: Gartner public info and Internet references

The analytical workbench

The business plays the main role with IT enabling and supporting the infrastructure.

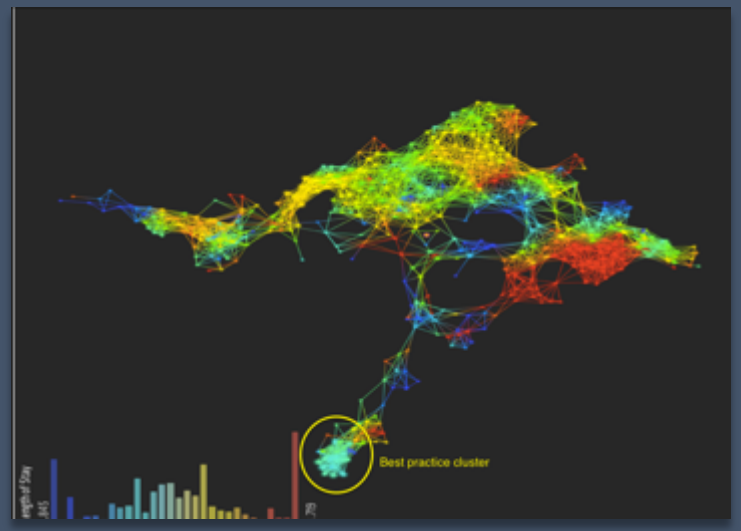
These analytics capabilities should be selected when the objective is to support user autonomy, agility and flexibility, without requiring advanced analytics skills.

The analytics workbench domain should be avoided when the use cases don't tolerate doubts around the trust level of insights created by users

The **ANALYTICAL CAPABILITIES** here are Visual Data Discovery, Location Analytics, Data preparation, Citizen Data Science, NoSQL based discovery, Stream Analytics, Graph Analytics and Embedded Analytics.

The business oriented roles involved in this domain are Analyst and Citizen Data Science

The IT or tech specialists oriented roles involved in this domain are analytics support expert and Data Engineer.



GRAPH ANALYTICS

It's a powerful analytical tool to explore indirect relationships between entities across multi-structured data

Perfecto to enhance pattern analysis through complex networks.

EMBEDDED ANALYTICS

This analytical capability delivers real-time reporting, interactive data visualization and/or advanced analytics, including machine learning, directly into an enterprise business application.

The data is managed by the analytics platform, and reports are placed directly within the application UI to improve the context and usability of the data for business users

Source: Gartner public info and Internet references

The analytical workbench

The business plays the main role with IT enabling and supporting the infrastructure.

These analytics capabilities should be selected when the objective is to support user autonomy, agility and flexibility, without requiring advanced analytics skills.

The analytics workbench domain should be avoided when the use cases don't tolerate doubts around the trust level of insights created by users

The analytical capabilities here are Visual Data Discovery, Location Analytics, Data preparation, Citizen Data Science, NoSQL based discovery, Stream Analytics, Graph Analytics and Embedded Analytics.

The **BUSINESS ORIENTED ROLES** involved in this domain are Analyst and Citizen Data Science

The **IT OR TECH SPECIALISTS ORIENTED ROLES** involved in this domain are analytics support expert and Data Engineer.

ANALYST

He delivers value to their companies by taking information about specific topics and then interprets, analyzes, and presents findings in comprehensive reports.

He can be more Data or Business oriented

CITIZEN DATA SCIENCE

They're power users who can perform both simple and moderately sophisticated analytical tasks that would previously have required more expertise.

Typically, a citizen data scientist is not a member of an analytics team

DATA ENGINEER

The data engineer is someone who develops, constructs, tests and maintains architectures, such as databases and large-scale processing systems. The data scientist, on the other hand, is someone who cleans, massages, and organizes (big) data

ANALYTICS SUPPORT EXPERT

This is an IT role fully oriented to provide a consistent tech support to business analytics advanced roles as Analysts and Citizen Data Scientist

Source: Gartner public info and Internet references

The Data Science Lab

This domain requires very specialized users as data scientists with a strong intuition for data, analytics and the business domain, with IT acting as a support role.

The analytics capabilities in this domain should be selected when there are clear requirements that justify their potential complexity, and when the resources with the right advanced skills are available.

By reaching this domain, organizations demonstrate a high level of mature on Data Analytics and Big Data.

The **ANALYTICAL CAPABILITIES** here are Machine learning, Deep learning, Predictive analysis, Prescriptive analysis and Simulation and optimization

The business oriented roles involved in this domain are Statistician, Data Science and Analytics project manager.

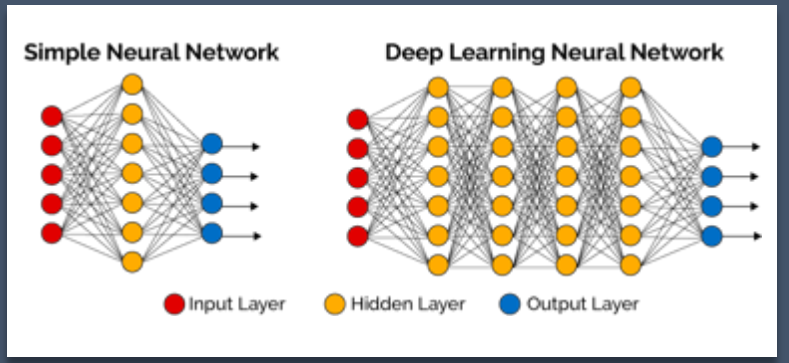
The IT or tech specialists oriented roles involved in this domain is Analytics system integrator.

MACHINE LEARNING

Machine learning is the art of study of algorithms that learn from examples and experiences.

Machine learning is based on the idea that there exist some patterns in the data that were identified and used for future predictions.

The difference from hardcoding rules is that the machine learns on its own to find such rules.



DEEP LEARNING

Deep learning is a sub-field of machine learning. Deep learning does not mean the machine learns more in-depth knowledge; it means the machine uses different layers to learn from the data. The depth of the model is represented by the number of layers in the model. For instance, Google LeNet model for image recognition counts 22 layers.

In deep learning, the learning phase is done through a neural network. A neural network is an architecture where the layers are stacked on top of each other.

Source: Gartner public info and Internet references

The Data Science Lab

This domain requires very specialized users as data scientists with a strong intuition for data, analytics and the business domain, with IT acting as a support role.

The analytics capabilities in this domain should be selected when there are clear requirements that justify their potential complexity, and when the resources with the right advanced skills are available.

By reaching this domain, organizations demonstrate a high level of mature on Data Analytics and Big Data.

The analytical capabilities here are Machine learning, Deep learning, Predictive analysis, Prescriptive analysis and Simulation and optimization

The **BUSINESS ORIENTED ROLES** involved in this domain are Statistician, Data Science and Analytics project manager.

The **IT OR TECH SPECIALISTS** oriented roles involved in this domain is Analytics system integrator.

DATA SCIENTIST

DT helps companies interpret and manage data and solve complex problems using expertise in a variety of data niches. They generally have a foundation in computer science, modeling, statistics, analytics, and math - coupled with a strong business sense.

STATISTICIAN

This role develops and initiates innovative statistical techniques and protocols. At the same time, he develops easy-to-analyze sampling techniques and processes.

Under the coordination of Data Scientist, he executes statistical operations in total fairness to derive zero-error results.

ANALYTICS PM

Standard project management techniques do not work well for analytics projects, and therefore, nor do standard project managers. This is why Project manager role in analytics is critical for Analytics adoption success in organizations.

This is a role quite difficult to find because IT managers who make the transition into analytics often struggle with this aspect.

ANALYTICS SYSTEM INTEGRATOR

For traditional systems integrators, continuous improvement is second nature. Service delivery is their core business and continuous improvement is how they generate profits.

Work with Big Data and Analytics, close to IoT, needs an oriented and specialized role in that critical area.

Source: Gartner public info and Internet references

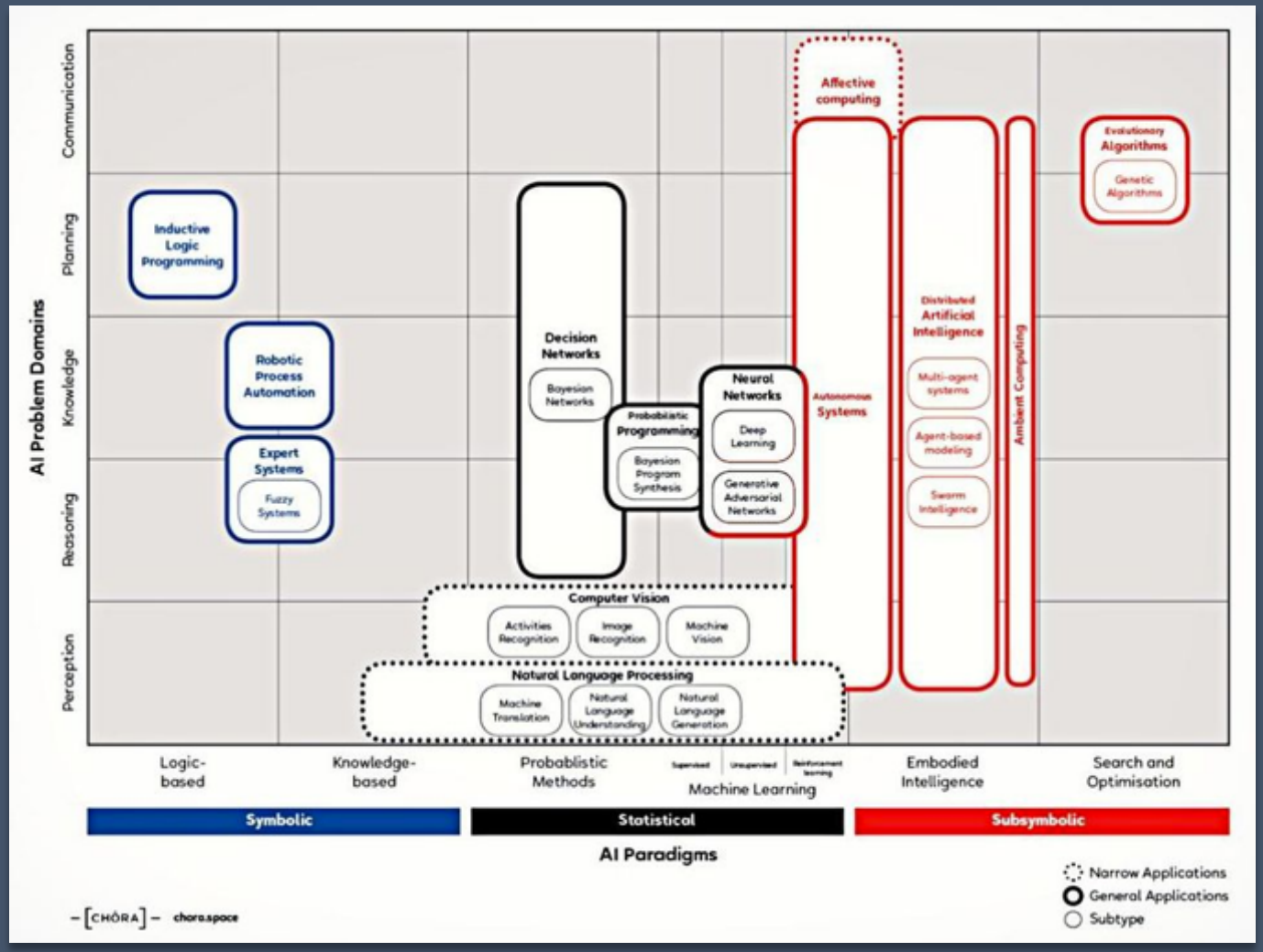
The Artificial Intelligence HUB

The Hub offers technology that appears to emulate human performance, typically by learning, coming to its own conclusions, appearing to understand complex content, engaging in natural dialogues with people, enhancing human cognitive performance or replacing people for nonroutine tasks.

Here IT will play a key role in the integration of analytics with business applications and the automation of processes

There are a lot of emerging analytical capabilities at this domain. We can see great advances on Personal Digital Assistants, new digital services around Cognitive computing or business applications using Knowledge graphs.

The roles involved here are being created constantly to cover all the new features discovered. A new one is the Analyst Enterprise Architect, absolutely key to build a seamless end-to-end Analytical process.



Source: https://www.forbes.com/sites/cognitiveworld/2018/08/22/ai-knowledge-map-how-to-classify-ai-technologies/#6fbbffa17773

Coming soon

DATA ANALYTICS AND BIG DATA – INSIDE THE ORGANIZATION