TRENDS IN BUSINESS INTELLIGENCE & DATA ANALYTICS

ING.
MARIÁN LOJKA

INNOVATION ARCHITECTURE
ACCENTURE RESEARCH

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

DISCLAIMER & TERMS OF USE

This document is proprietary to Accenture. The material, ideas, and concepts contained herein are to be used exclusively for lecture at VŠE (Absolventská středa) on 28.11.2018 and to the audience of this lecture. This slideshow and the information, concepts and ideas herein may not be used for any purpose other than the said presentation of Accenture's services, may not be changed, altered or further distributed to anyone without prior express written permission by Accenture.

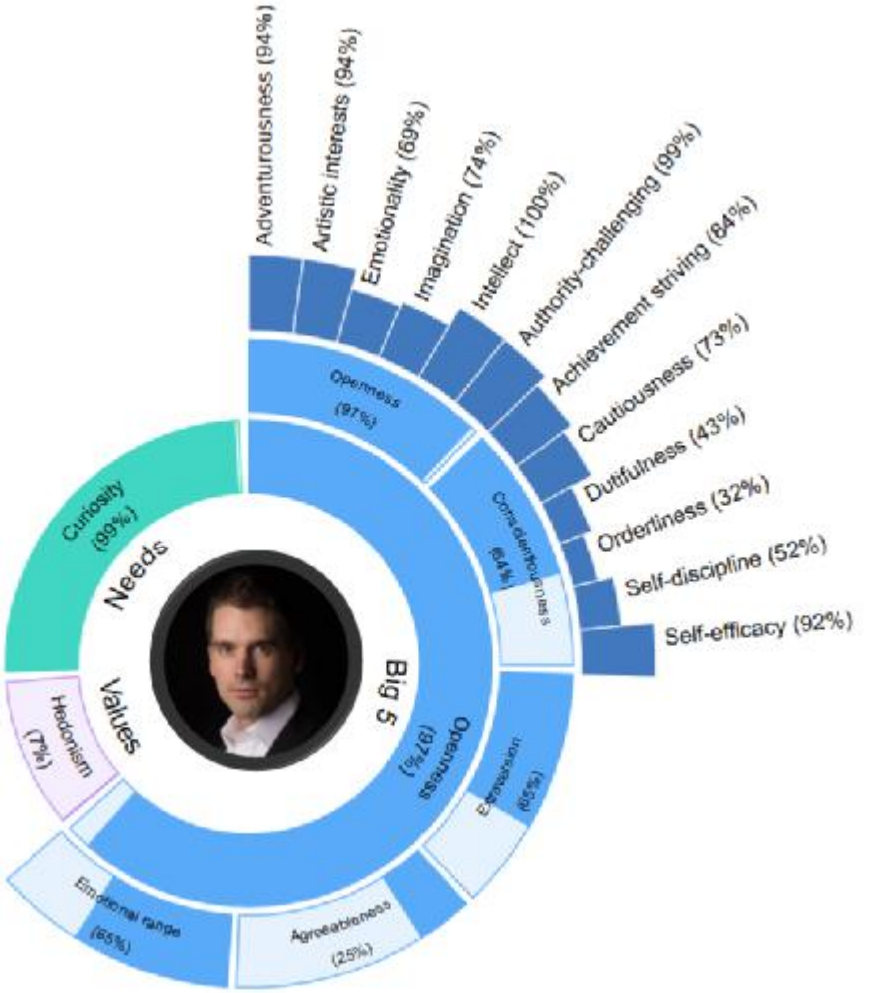
Tato prezentace je vlastnictvím Accenture. Materiály, myšlenky a koncepce v ní obsažené mohou být použity výhradně k prezentaci na přednášce (Absolventská středa) na VŠE dne 28.11.2018 a jeho publikum. Tato prezentace a informace, koncepce a myšlenky v ní obsažené nesmějí být použity k jiným účelům než k prezentaci a uskutečnění kurzu Accenture, nesmějí být bez předchozího výslovného písemného svolení Accenture měněny, pozměňovány nebo nikomu dále distribuovány.

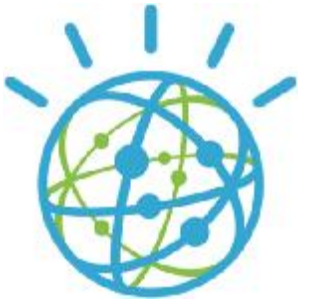
WHAT WE DO

Accenture solves our clients' toughest challenges by providing unmatched services.
We provide end-to-end services for clients across our five businesses.

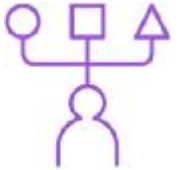
accenturestrategy	accentureconsulting	accenturedigital	accenturetechnology	accentureoperations
SHAPES	TRANSFORMS	DIGITIZES	POWERS	OPERATES
Business Strategy	Management Consulting	Interactive	Application Services	As a Service
Technology Strategy	Technology Consulting	Mobility	Labs	Business Process
		Analytics	Ecosystem Alliances	Cloud
				Security

INTRODUCTION – WATSON WHO AM I ?





IBM WATSON



Personality Insights

This model was generated by IBM Watson Personality Insight try it here > https://personality-insights-demo.ng.bluemix.net/

"IT IS A CAPITAL MISTAKE TO THEORIZE BEFORE ONE HAS DATA."

- Sherlock Holmes

BUT WHAT IF...
DATA IS NOT RELIABLE?

WHAT IS THE DIFFERENCE ?

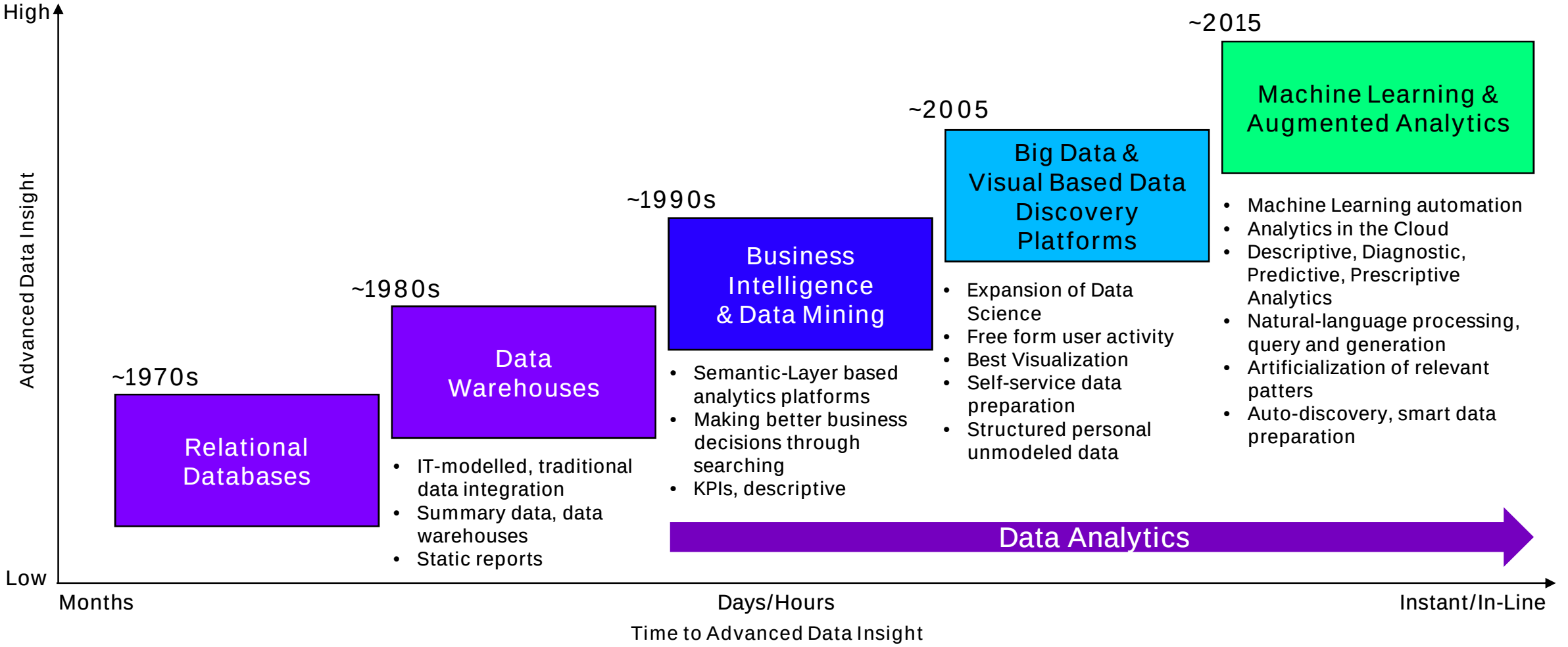
Traditional Business Intelligence

Data Analytics

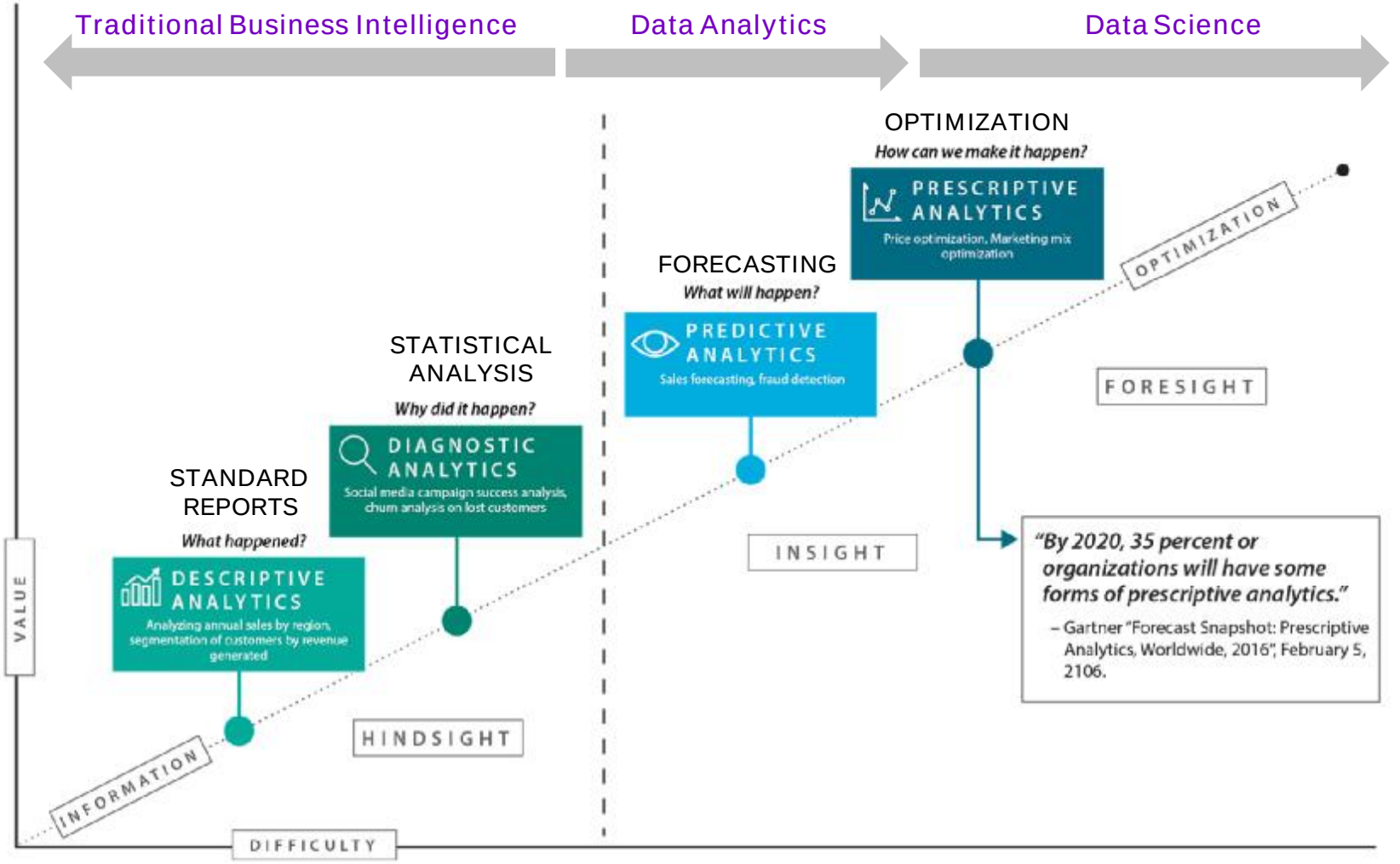
Data Science

Powered by Wordle™

THE EVOLUTION OF ANALYTICAL PLATFORMS



THE DIFFERENCE BETWEEN BI & DATA ANALYTICS



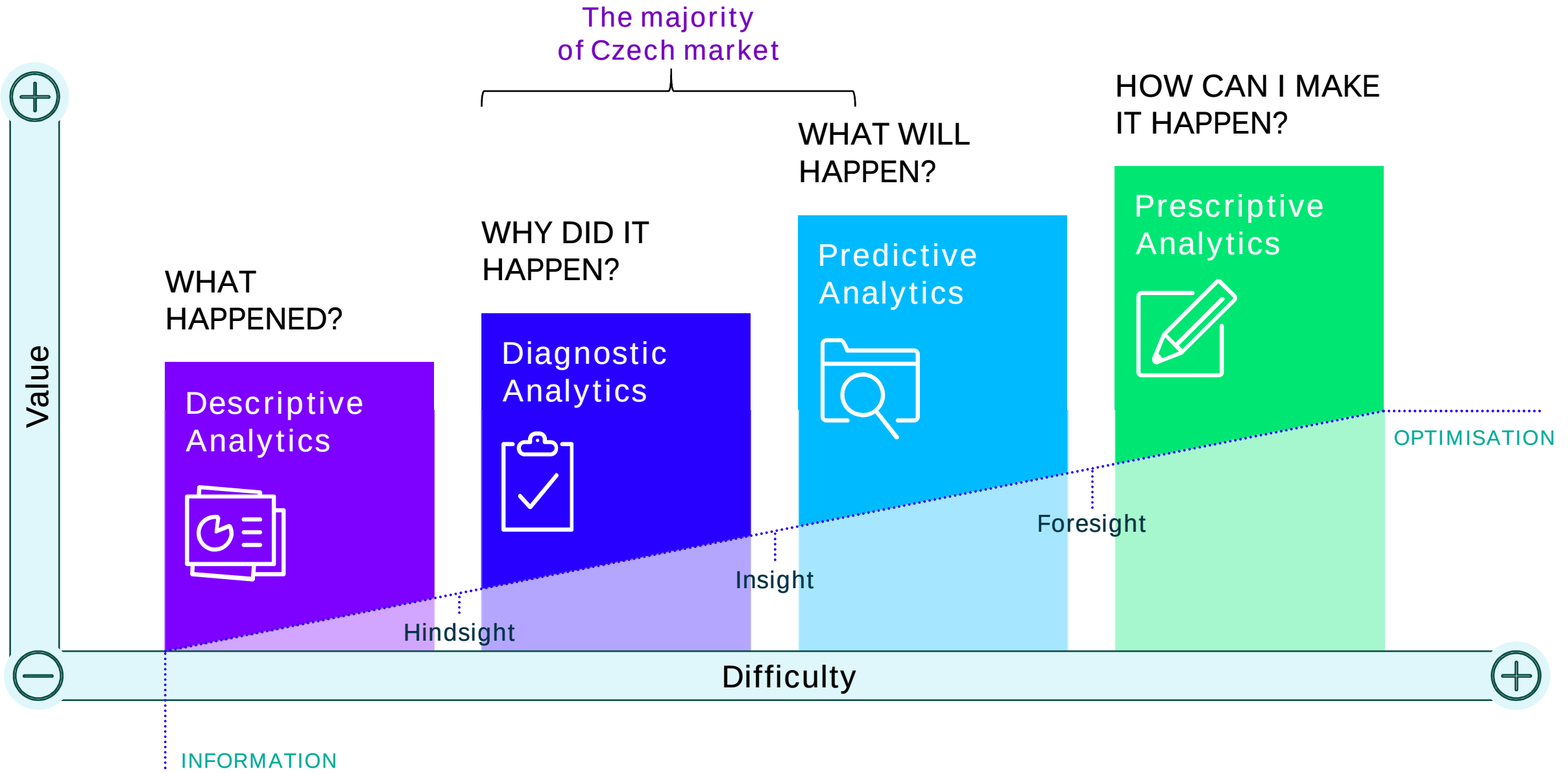
Business Intelligence (Traditional BI)

- Measure past performance and guide business planning
- Includes:
  - Reporting (KPIs, metrics),
  - Dashboards,
  - Scorecards,
  - OLAP,
  - Ad-Hoc query,
  - Operational and Real-Time Query

Data Analytics & Data Science

- Predict future events
- Includes:
  - Data Mining
  - Statistical analysis
  - Predictive modeling
  - Big Data Analytics
  - Simulation
  - Optimization / Data Science
  - Machine Learning

WHERE IS THE MARKET ?



BIG THEMES

50-80% of enterprise data is considered dark, while 90% of unstructured data remains unused”

Companies are failing to maximise the opportunities around internal data for advanced analytics and/or as a new revenue stream through DaaS or Infonomics.

(SOURCE GARTNER, 2018)

BIG THEMES

One of the biggest barriers to adoption of advanced analytics/AI is the availability of data scientists and a new requirement for data engineers

(SOURCE GARTNER, 2018)

BIG THEMES

Data scientists spend 80% of their time on data preparation versus 20% on analytics.

(SOURCE GARTNER, 2018)

BIG THEMES

Data scientists spend 80% of their time on data preparation versus 20% on analytics.

(SOURCE GARTNER, 2018)

BIG THEMES

WHAT SHAPING THE INDASTRY ?

...WHAT BUSINESS NEEDS.

accenture digital

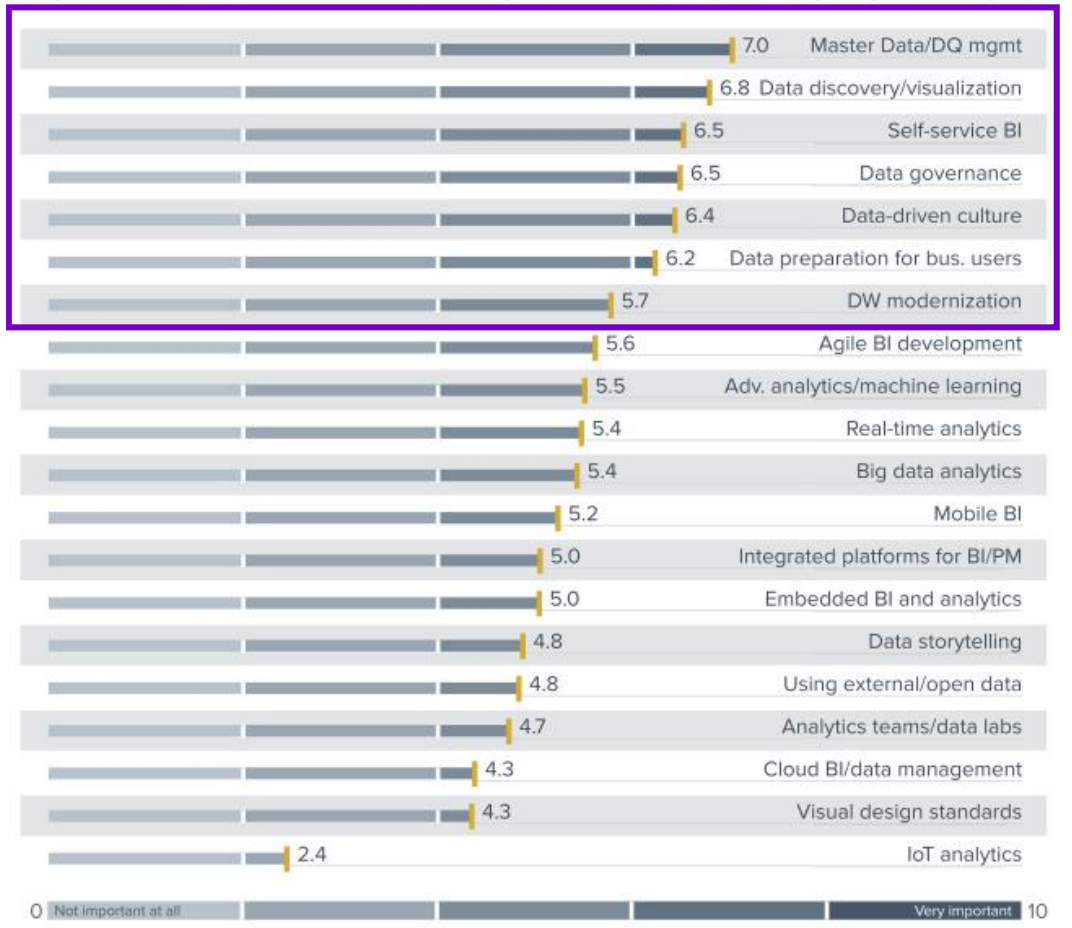
Strategy | Consulting | Digital | Technology | Operations

Accenture Analytics

High performance. Delivered.

WHAT 2,679 BI PROFESSIONALS REALLY THINK IS IMPORTANT, NEEDS IN Y2019

Importance of BI trends from "not important at all" (0) to "very important" (10)



n = 2679

- For the second consecutive year, BI practitioners selected data quality and master data management as the most important trend.
- The significance of relying on high quality data as well as having all important data to hand seems to be consistently high.
- This trend is backed up by the equality of significance of data governance which is ranked on fourth position.
- Data-Discovery and self-service BI, in second and third position respectively, continue to be very important topic.
- Newly introduced topic – "establishing the a data-driven culture" came straight in number five this year, providing itself to be highly significant.
- Other top trend such as data quality, data discovery, self-service BI can be considered to be the foundation of data driven culture

BARC A CXP GROUP COMPANY | Viewpoint

We asked users, consultants and software vendors of BI and data management technology to give their personal rating of the importance of twenty trending topics that we presented to them. The three most important trends remained the same as last year with master data and data quality management in first position, data discovery in second and self-service BI in third. While master data and data quality management builds a strong foundation for handling data, the significance attached to data discovery and self-service BI shows that the empowerment of business users is a consistently strong trend. This impression is also backed up by the newly introduced and fifth ranked topic "establishing a data-driven culture", which depends on greater inclusion of various business departments (aside from IT). Data governance, which remains in fourth position is still seen as an important trend. Here, again, GDPR comes into play. Although data governance covers a much wider spectrum than 'just' providing data protection, the rising significance of data governance can also be traced back to an increase in data security awareness.

The lack of interest in IoT analytics, which features in last place in its debut year in BI Trend Monitor, leads to the assumption that businesses are neither prepared nor really focused on implementing this special kind of analytics at the moment. With IoT itself just emerging in importance, it will probably take some time before the potential added value of IoT analytics begins to unfold.

THE BIGGEST SURGE IN INTEREST IS SEEN WITH ADVANCED ANALYTICS/MACHINE LEARNING AND ANALYTIC TEAMS/DATA LABS.

n = 2794/2772/2770/2679

- Trends that have clearly increased in importance compared to last year include agile BI development and advanced analytics and analytics teams.
- While agile BI development is connected to a revolutionized cooperative approach between lines of business and IT, advanced analytics expresses the need for businesses to use data in a more beneficial way.
- Also, advanced analytics includes machine learning, which is tightly interconnected to many hyped use cases in the sphere of artificial intelligence.
- Conversely, topics decreasing in importance include real-time analytics and mobile BI. It seems that the perceived practical benefit of these trends has not become as obvious as expected to most BI practitioners yet.
- Low score for trends like Cloud BI can be linked to security concerns / GDPR Europe.

BARC A CXP GROUP COMPANY | Viewpoint

We asked users, consultants and software vendors of BI and data management technology to give their personal rating of the importance of twenty trending topics that we presented to them. The three most important trends remained the same as last year with master data and data quality management in first position, data discovery in second and self-service BI in third. While master data and data quality management builds a strong foundation for handling data, the significance attached to data discovery and self-service BI shows that the empowerment of business users is a consistently strong trend. This impression is also backed up by the newly introduced and fifth ranked topic "establishing a data-driven culture", which depends on greater inclusion of various business departments (aside from IT). Data governance, which remains in fourth position is still seen as an important trend. Here, again, GDPR comes into play. Although data governance covers a much wider spectrum than 'just' providing data protection, the rising significance of data governance can also be traced back to an increase in data security awareness.

The lack of interest in IoT analytics, which features in last place in its debut year in BI Trend Monitor, leads to the assumption that businesses are neither prepared nor really focused on implementing this special kind of analytics at the moment. With IoT itself just emerging in importance, it will probably take some time before the potential added value of IoT analytics begins to unfold.

Top Trends Shaping Industries

By 2022, 80% of AI applications will use cloud-based infrastructures and components, up from 40% today.

1

Agile Adoption

By 2020, 30% of data lakes will be built on standard relational DBMS technology at equal or lower cost than Hadoop.

2

Modernization of DWHs

By 2020, modern BI platforms with augmented data discovery will show 2X adoption over all other modern BI platforms

3

Data Veracity (MDM / DQ)

100% of large organizations will purchase external data by 2019

By 2020, 40% of organizations will be in violation of GDPR; this is expected to be near zero by 2023.

4

CITIZEN AI

By 2019, 50% of analytics queries will be generated using search, NLQ or voice, or will be autogenerated

5

Data Visualization /Discovery & Self-Service BI

Through 2019, 80% of data lakes will not include effective metadata management capabilities, making them inefficient.

Business analytics services spending will increase from $54.5 billion in 2017 to

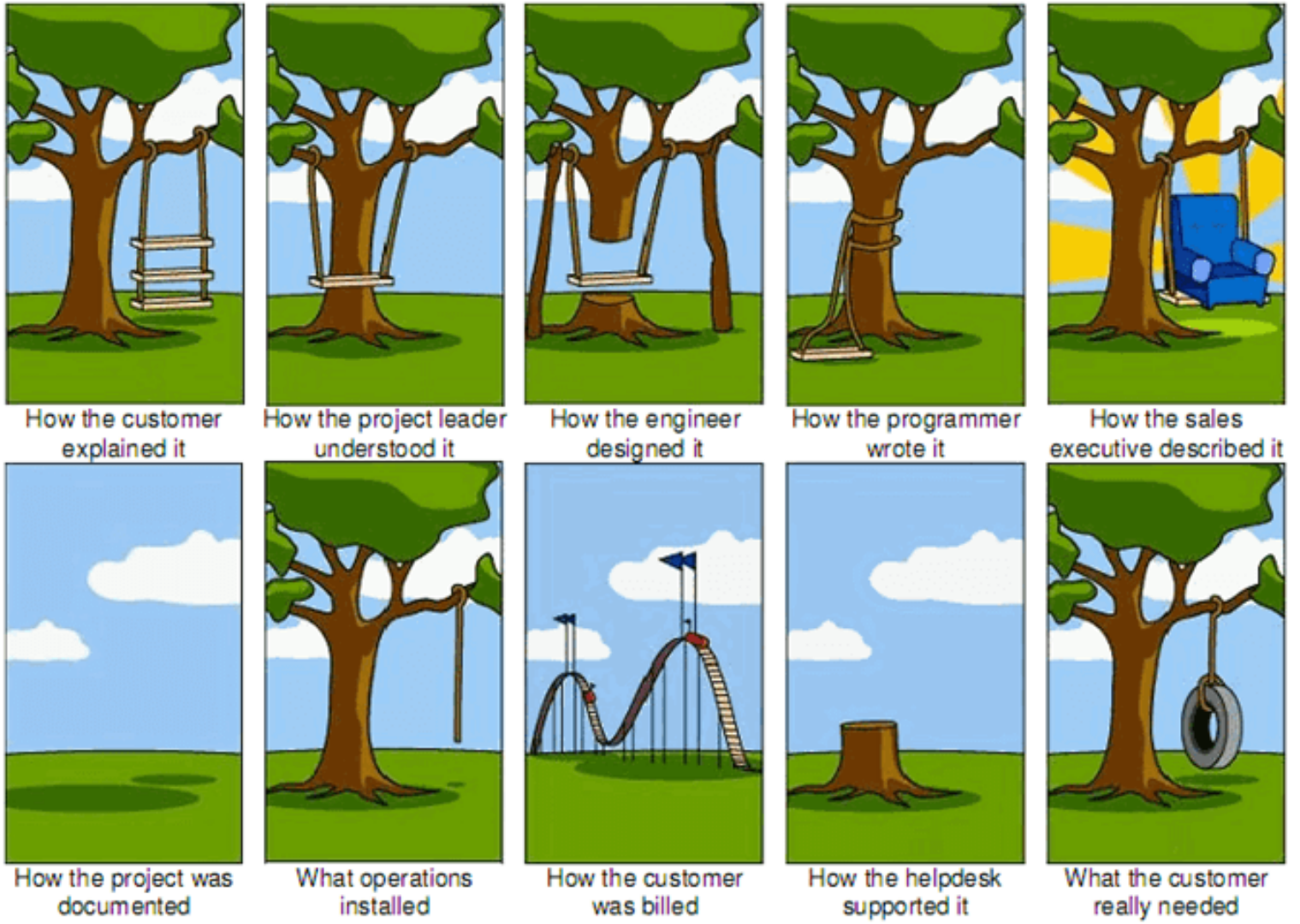
$87.3 billion in 2021

THE AGILE ADOPTION

- Application-focused Development
- Long-term planning and fixed Budgets
- Clear-Cut Business Requirement to IT Handover
- Backwards-looking (follower) view on innovation

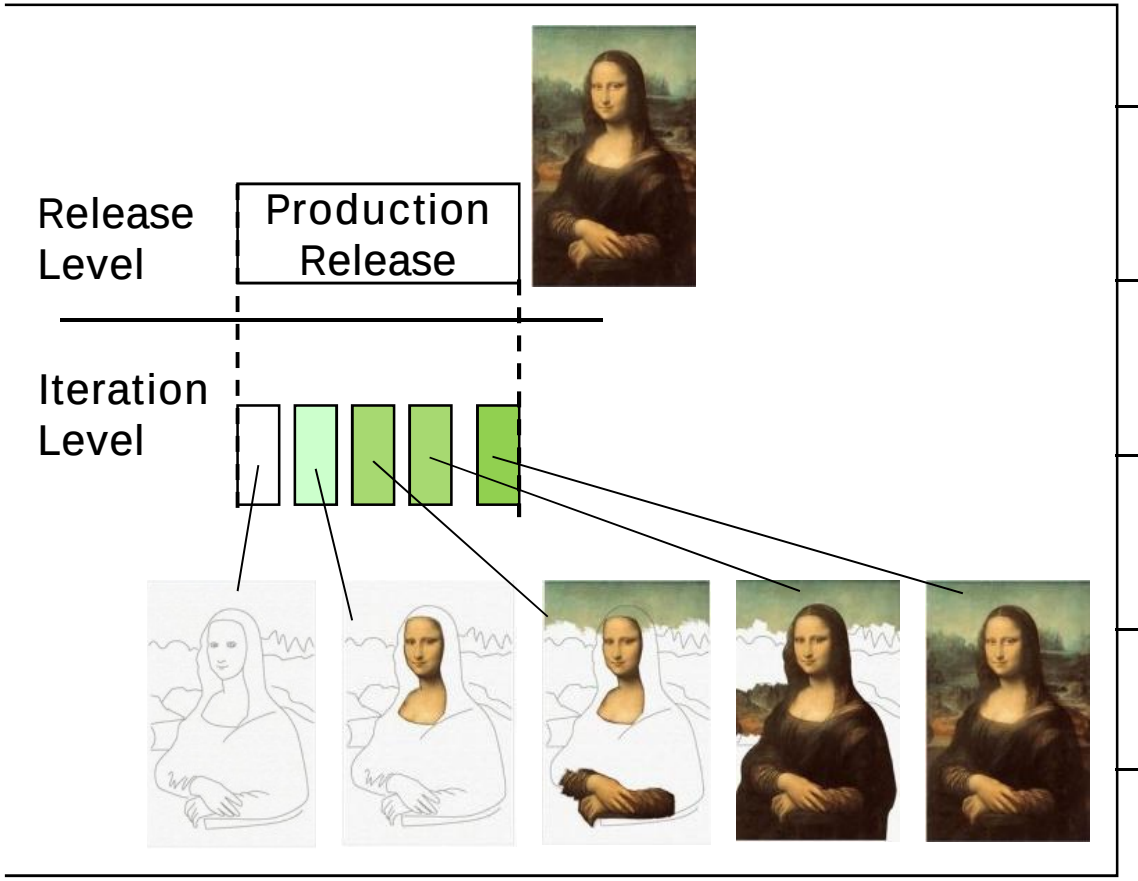
- Product-focused Development
- Rolling Budget based on User Feedback
- Collaborative Ownership of Business and IT
- Exploration of trends, ideas and technologies

WATERFALL APPROACH



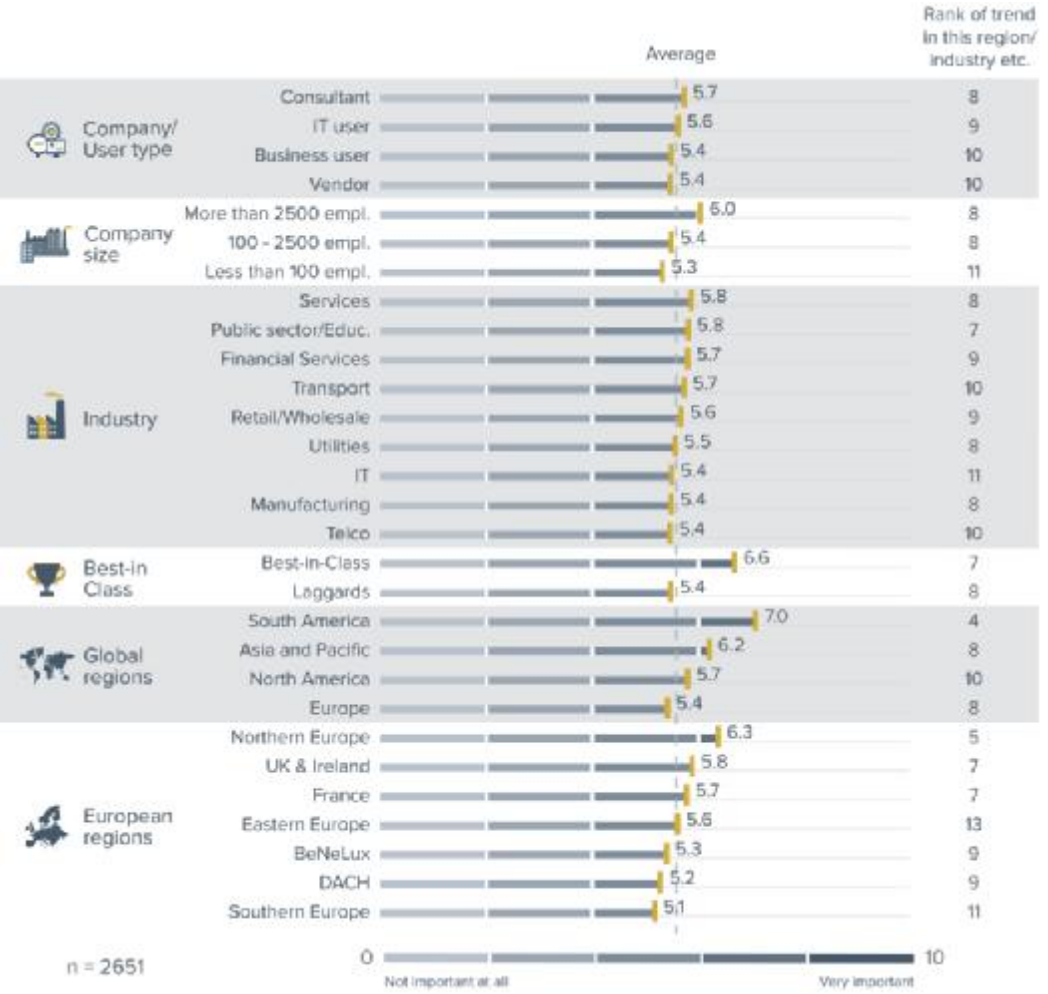
JOURNEY FROM WATERFALL TO AGILE ADOPTION

The journey begins with release cycles and continous improvement



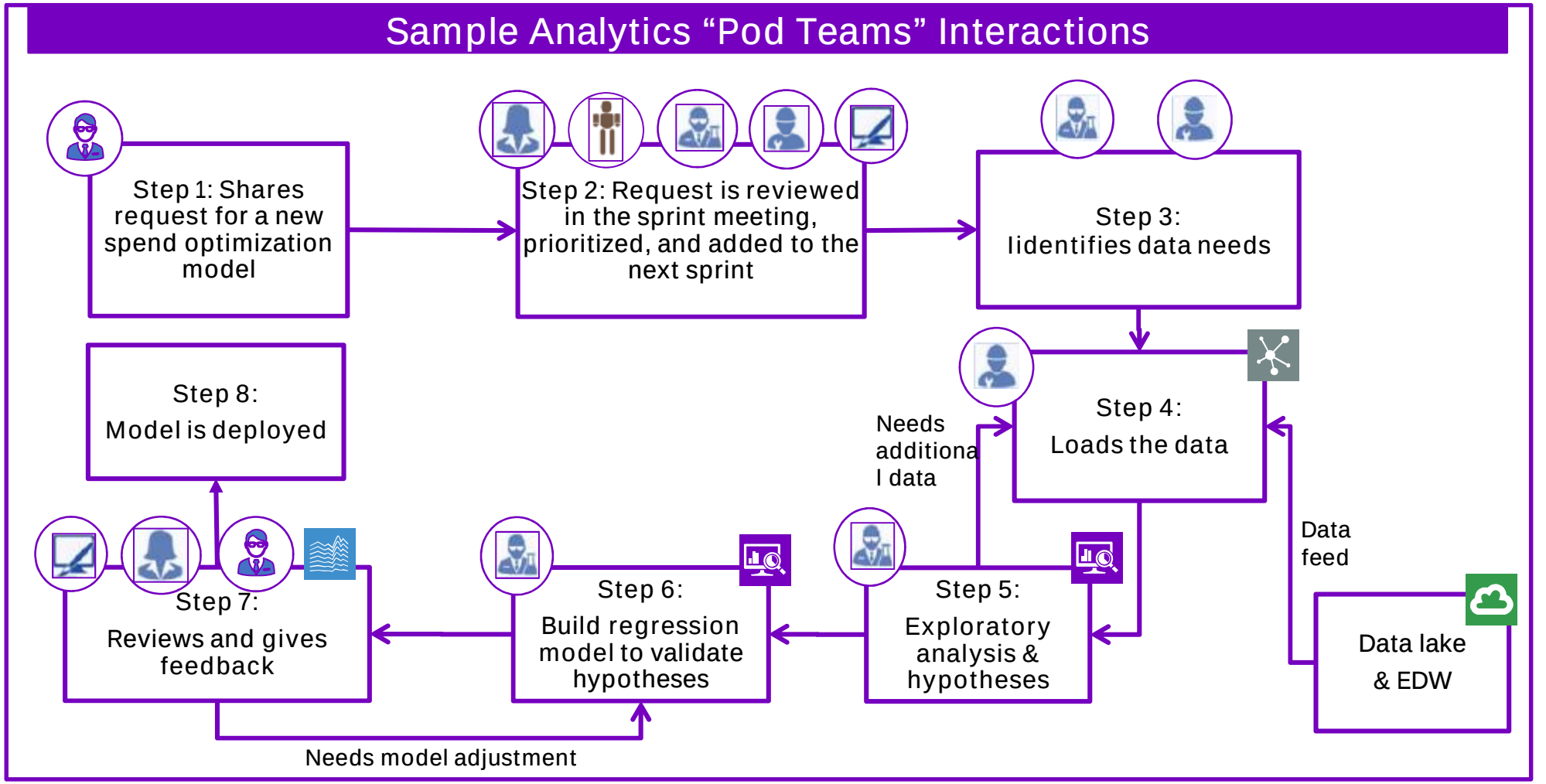
- Waterfall requires single big bang delivery
- Fixed budget for fixed Scope
- Quality can only be ensured when big bang is delivered
- Agile handles changes iteratively
- Quality in Agile continuously improved

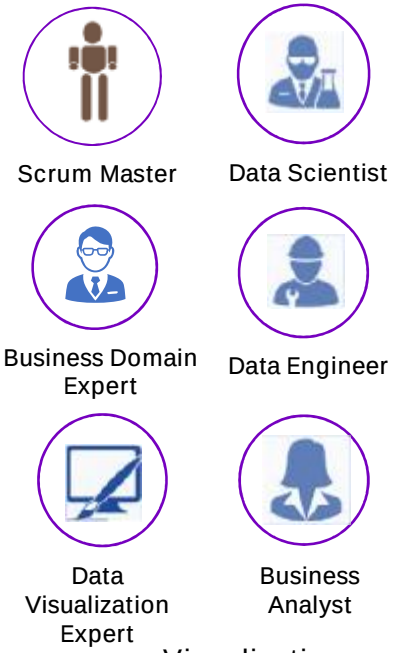
JOURNEY FROM WATERFALL TO AGILE ADOPTION



- The term "agile" has increasingly been adopted in the context of business intelligence in recent years. The "agile" moniker is now often used as a requirement for the development of new data models, reports, dashboards or visualizations.
- Arguably, most users requesting "agile BI" have very little understanding of the agile development methodology and use the term as a synonym for "flexible", indicating a pressing need for faster development cycles.
- Agile BI requires organizations to adopt an iterative development approach Instead of the traditional waterfall method, by which requirements are gathered before the development process starts, close collaboration between business and IT, using rapid prototyping, enables organizations to increase development speed while better responding to business needs.
- Many companies are not set up organizationally for this approach, however, and some changes in organizational structures may be required. Ideally, the agile BI development approach is also supported by agile project management, by which planning, requirements collection, development, but also functional, regression and usability testing are managed in an iterative manner.
- An important aspect, and one that is often considered a bottleneck, is the availability of business users to collaborate in the development process.

DELIVERY IN ACTION, ADVANCE ANALYTICS WITH "POD" TEAMS





Visualization
Best of breed self-service BI and Geo-spatial visualization

Analytics
Building blocks for modelling & processing analytics

Storage
Cloud architecture and other validated storage

Data Ingestion
Integrates data preparation and extraction tools

MODERNIZATION OF DATA WAREHOUSES

MODERNIZATIO OF DATA WAREHOUSES - NEW DATA DNA

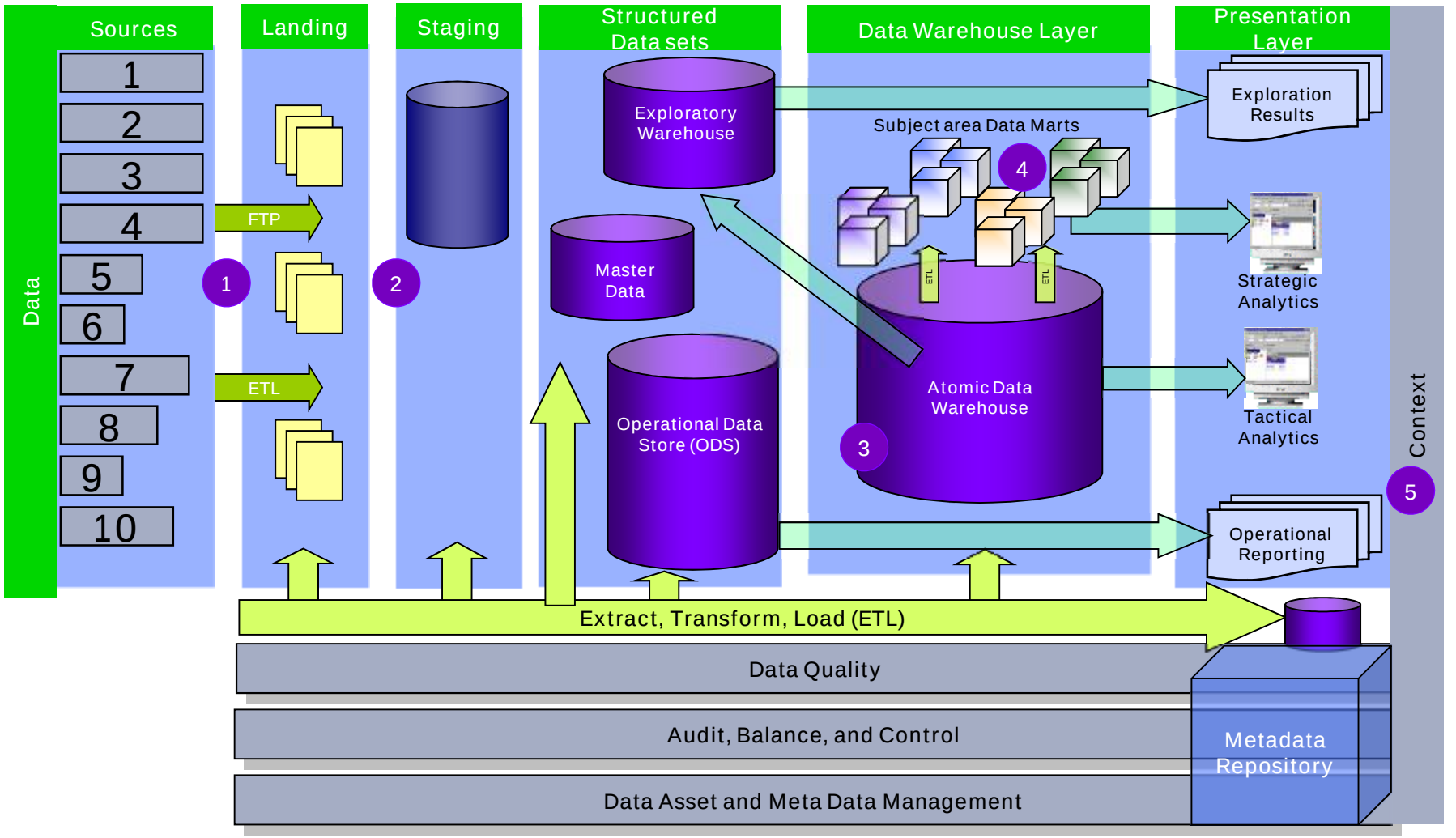
Dimensione	Range	
VELOCITY	**Batch** (eg. Subscrition of new agreement)	*Streaming* (eg. Markets data)
VOLUME	**Gigabyte** (es. client record)	*Petabyte* (eg. multichannel navigation data)
VARIETY	**Structured** (eg. Field, table, column)	*Non strutturato* (eg. text, email)
SOURCE	*Internal* (es. quotations)	*External* (eg. Social Media)

NEW DATA

MODERNIZATIO OF DATA WAREHOUSES - NEW DATA DNA

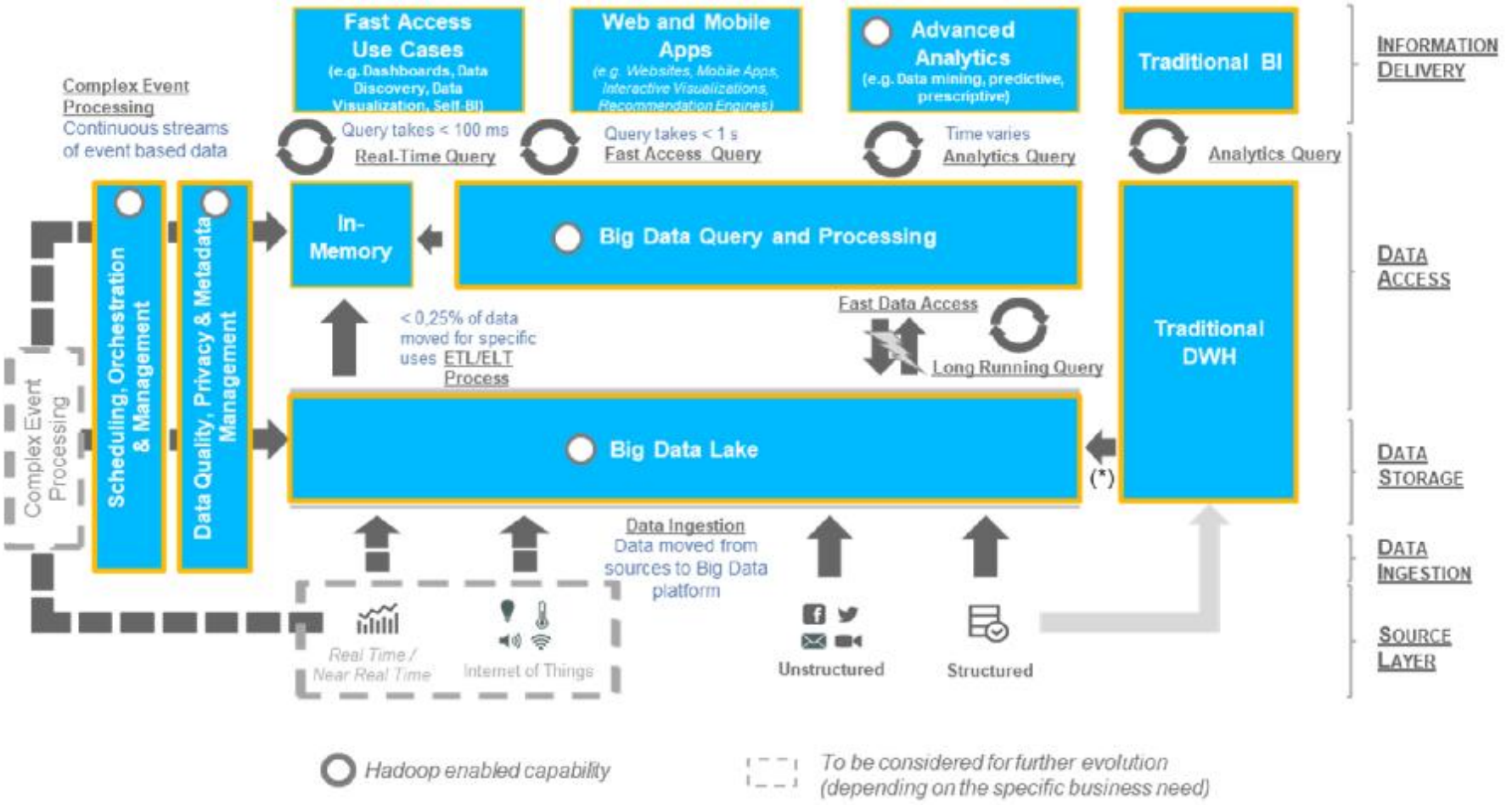
- New analytical challenges, increasing data variety, rising data volumes, faster decision processes, process automation and decreasing hardware costs are all having major effects on how companies store their data.
- Firstly, older data warehouse landscapes have become too complex to support agile development, or too expensive to have their functionality extended to accommodate modern analytics requirements.
- Furthermore, the type of the implementation of many data warehouse landscapes were originally designed and optimized does not cover the way analytics is currently moving forward in the direction of exploration and operational processing alongside classical BI requirements.
- Now, organizations are beginning to understand the new challenges and the potential of alternative methodologies, architectural approaches and utilizing other technical options like in-memory, cloud storage or data warehouse automation tools. IT must prepare for faster, changing analytical requirements, and they must also compete against new and cheaper implementation options from external service providers.
- Collaborative approaches are needed to cover the increasing expectations of the business to pull maximum business value from data. It is now time to assess historically grown data warehouses against present requirements and evaluate how updated hardware and technology could make life easier.

THE TRADITIONAL DATA WAREHOUSE ARCHITECTURE



1. Data Load of selective data elements based on precise requirements
2. Data loaded to Stage prepared for transformation in target integration model (3rd NF)
3. Mapping data into the target model requires exact knowledge of data
4. After data transformation to Atomic Data Warehouse or DataMart Business can do data analysis the first time
5. This typically takes too long, despite the fact that hopefully all data is correctly selected and fulfills content expectations

THE HYBRID DATA WAREHOUSE ARCHITECTURE



- A hybrid data warehouse architecture includes traditional enterprise data warehouse (EDW) and BigData Data Warehouse (BDW) to support new needs of unstructured data and Advanced Data Analytics
- BDW and EDW can exist separately. But they work best as complements.
- Multiple traditional EDWs, BDWs, and data lakes have become the new norm to support the variety of analytical workloads.

BY 2020, 40% OF ORGANIZATIONS WILL BE IN VIOLATION OF GDPR; THIS IS EXPECTED TO BE NEAR ZERO BY 2023

RIGHT TO ACCESS

Individuals can get confirmation of what personal information is being processed, where it is being stored, and why their information is being held. If EU citizens wish to know, a Controller must provide electronic copies of this data to the individual, free of charge.

RIGHT TO BE FORGOTTEN

Individuals are entitled to have their data erased, ceased from further dissemination, and potentially have third parties halt processing of data. In the case that their data is no longer relevant to why they originally gave their information, they may also have their data erased.

RIGHT TO DATA PORTABILITY

The right to data portability allows individuals to obtain and reuse their data for their own purposes across different services. It allows them to move, copy, or transfer personal data easily from one IT environment to another in a safe and secure way.

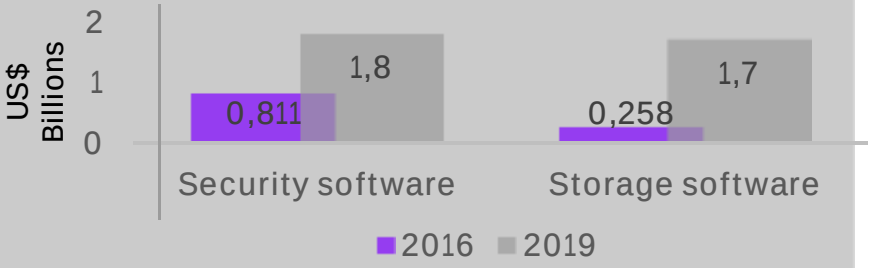
RIGHT TO NOTIFICATION

In the event of a data breach, businesses are required to notify their Data Protection Authority (DPA) within 72 hours of the breach. Individuals are also entitled to be notified in the event of a breach of their personal data.

GARTNER BELIEVES

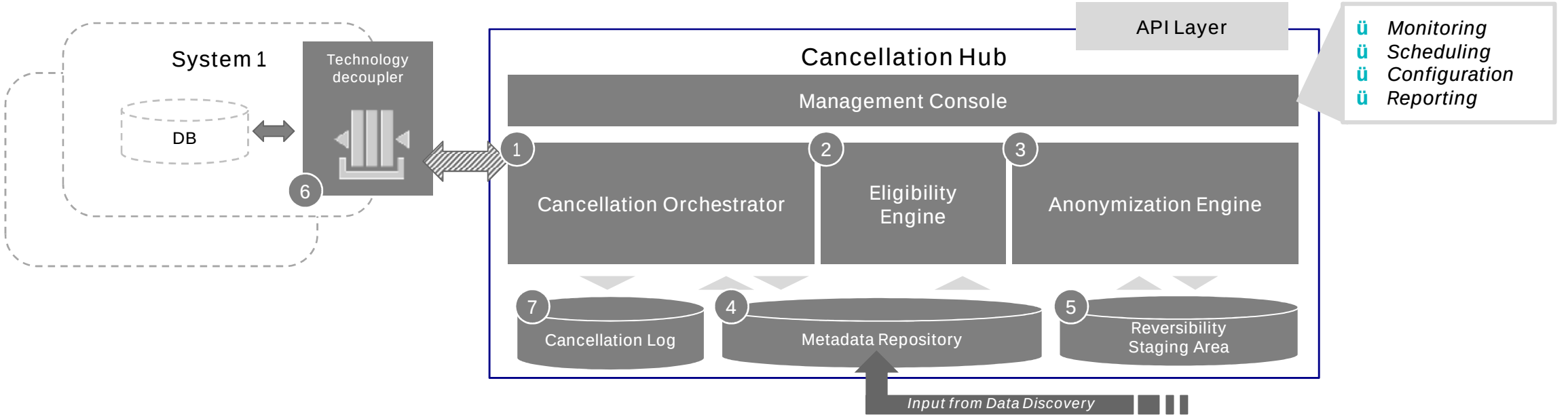
GARTNER - 40% OF ORGANIZATIONS WILL BE IN VIOLATION OF GDPR; THIS IS EXPECTED TO BE NEAR ZERO BY 2023

One in three firms believes they are GDPR-compliant today — but they may not be. Forrester believes that just a portion of these firms have actually engaged in data discovery and classification exercises as well as built data flow maps and run gap analysis.

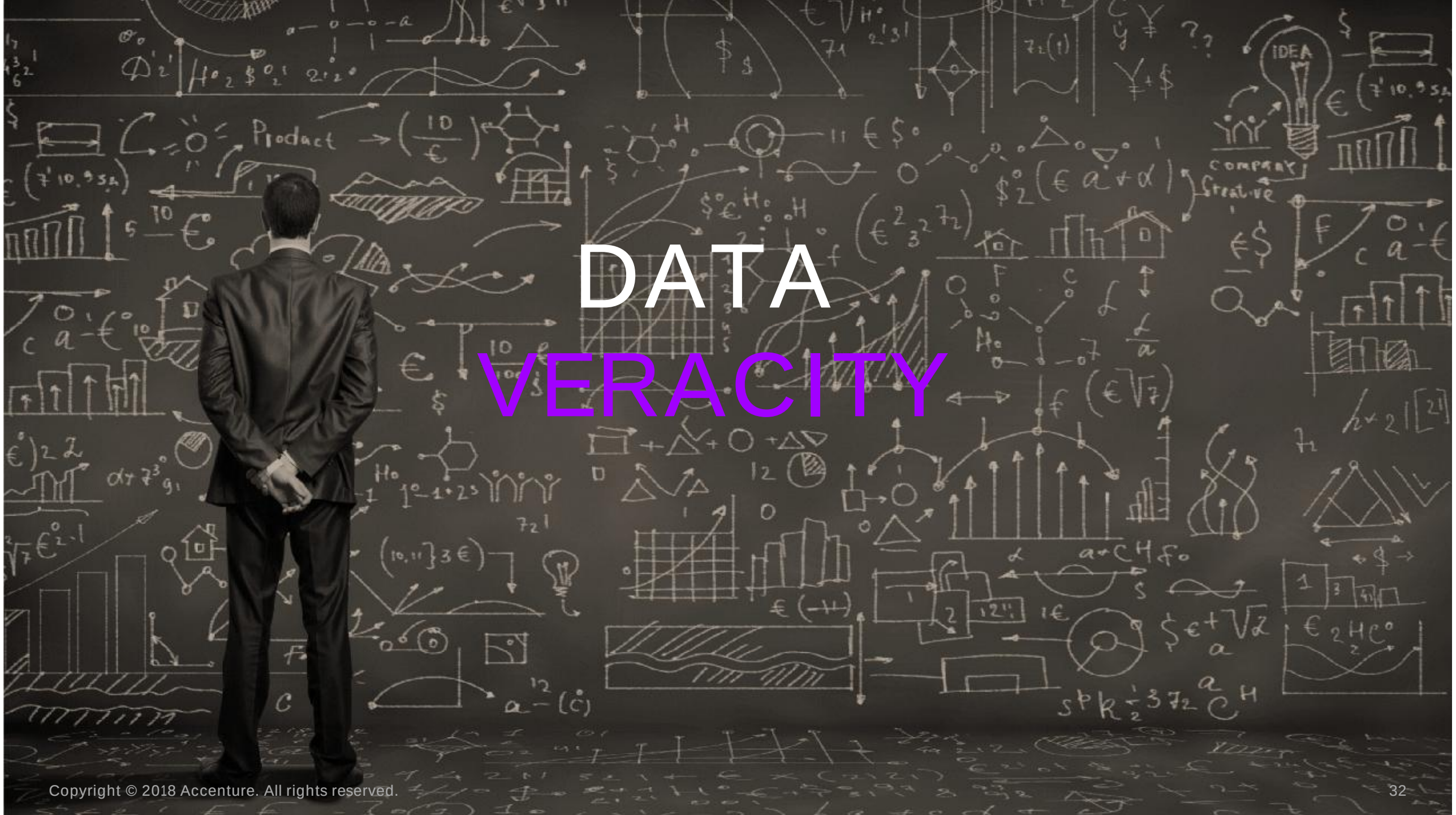


DATA DELETION ENGINEES

Logical Architecture

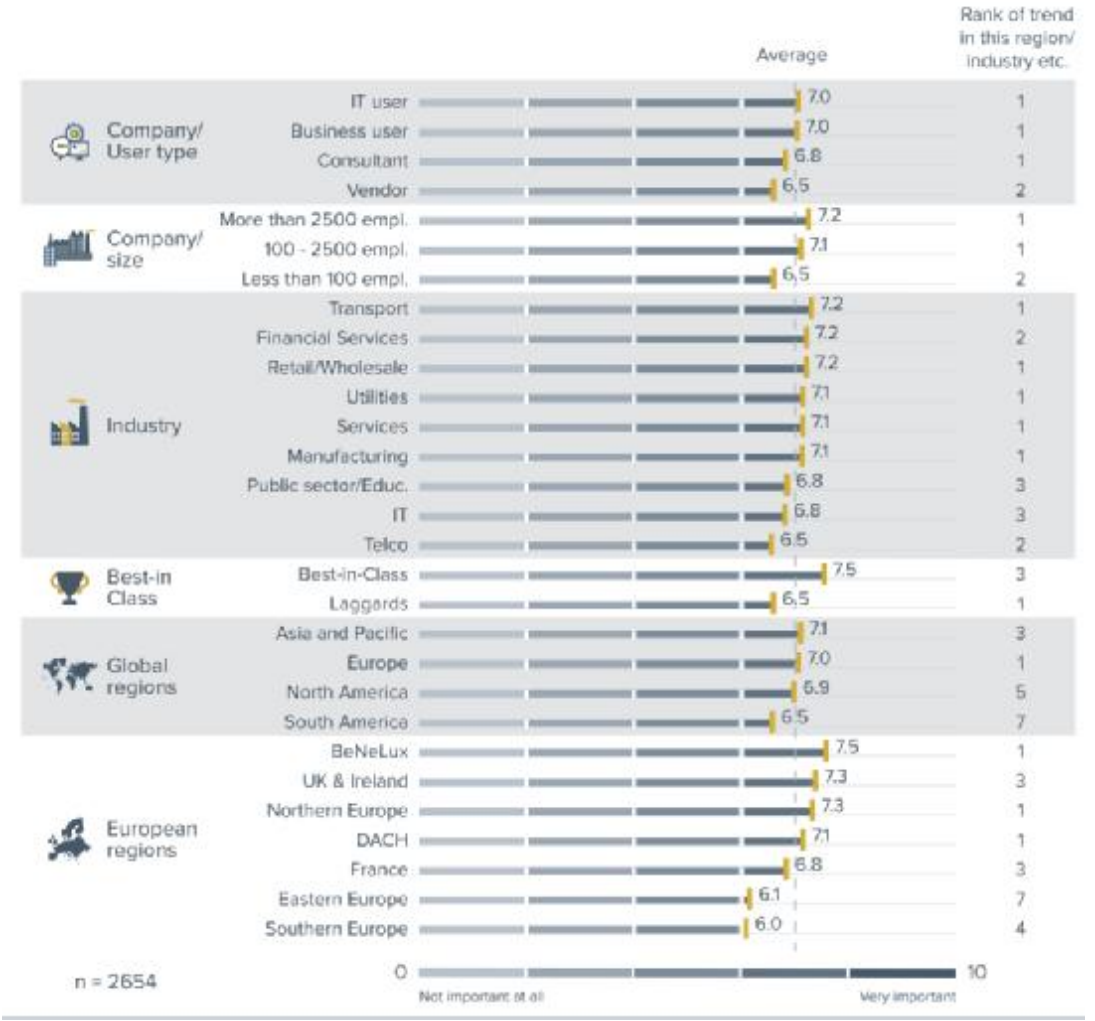


1. Orchestrates the cancellation across applications taking into account customer's constraints, eligibility, dependencies and topologies
2. Selects eligible personal data to cancel based on retention rules, subject type, purposes of the treatment and data types
3. Applies the right cancellation techniques to personal data based on field types, retention period and other customer's preferences
4. Stores metadata about tables, fields, databases and applications dependencies needed to calculate retention and to be delete
5. Temporary stores an encrypted version of the data cancelled to enable recovery in a defined period of time
6. Stores cancellation messages to be consumed by technology-specific cancellation procedure
7. Stores all the cancellation actions in order to report on past actions



DATA VERACITY

INCREASE DATA VERACITY WITH MASTER DATA & DATA QUALITY MANAGEMENT



- The importance of data quality and master data management can be explained very simply: people can only make the right decisions based on correct data. Decision-making processes and operational actions depend on reliable data. Through their aggregation mechanisms, BI reports and analyses can help to reveal data quality issues.
- The goal of master data management is to bring together and exchange master data across multiple systems. Aside from a "master" ERP system, many companies also work with other CRM or SCM systems, use web services, or need to merge systems following corporate mergers, or to co-operate as partners effectively.
- There are proven concepts for increasing data quality and implementing master data management. One example is the Data Quality Cycle, which many software vendors have implemented in their tools.
- In today's digital age, in which data is increasingly emerging as a factor of production, there is a growing need to use and produce high quality data to make new services and products possible. The critical success factors for sustainable high data quality are defined roles and responsibilities, quality assurance processes and continuous monitoring of the quality of a company's data.

DATA QUALITY PROCESSES

No matter how powerful are data quality tools and sophisticated the methodology, without effective process data quality cannot be really achieved

Business Functions & IT

Data Quality Process

1. **Controls execution**
   First level controls are automatically executed by IT applications where data are first recorded
   Monitoring controls are defined, created and executed with a given frequency to guarantee a certain level of quality

2. **Anomalies identification**
   Through control dashboards checks results can be reviewed.
   In case of anomalies it is possible to trigger the analysis process to identify the root cause and define the necessary actions

3. **Remediation processes**
   Remediation processes are classified based on anomaly characteristics:
   - Issue in data acquisition (e.g. wrong data entry), handled by related business organizational structure
   - IT issue (e.g. elaboration problem), handle by related IT application owner
   - Issue that highlights the need to modify the control itself (e.g. false positive)

4. **Remediation monitoring**
   Following the remediation actions it is required to monitor the effectiveness of the interventions, always leveraging on Data Quality Dashboards

Chief Data Officer

Data Quality Monitoring

5. **Data Quality monitoring and adequacy evaluation**

CDO informs constantly all stakeholders on the overall behavior of data quality with the aim to

- Monitoring data quality trends
- Support remediation processes and actions
- Support a continuous improvement of the Data Quality Framework, with a constant enhancement of controls

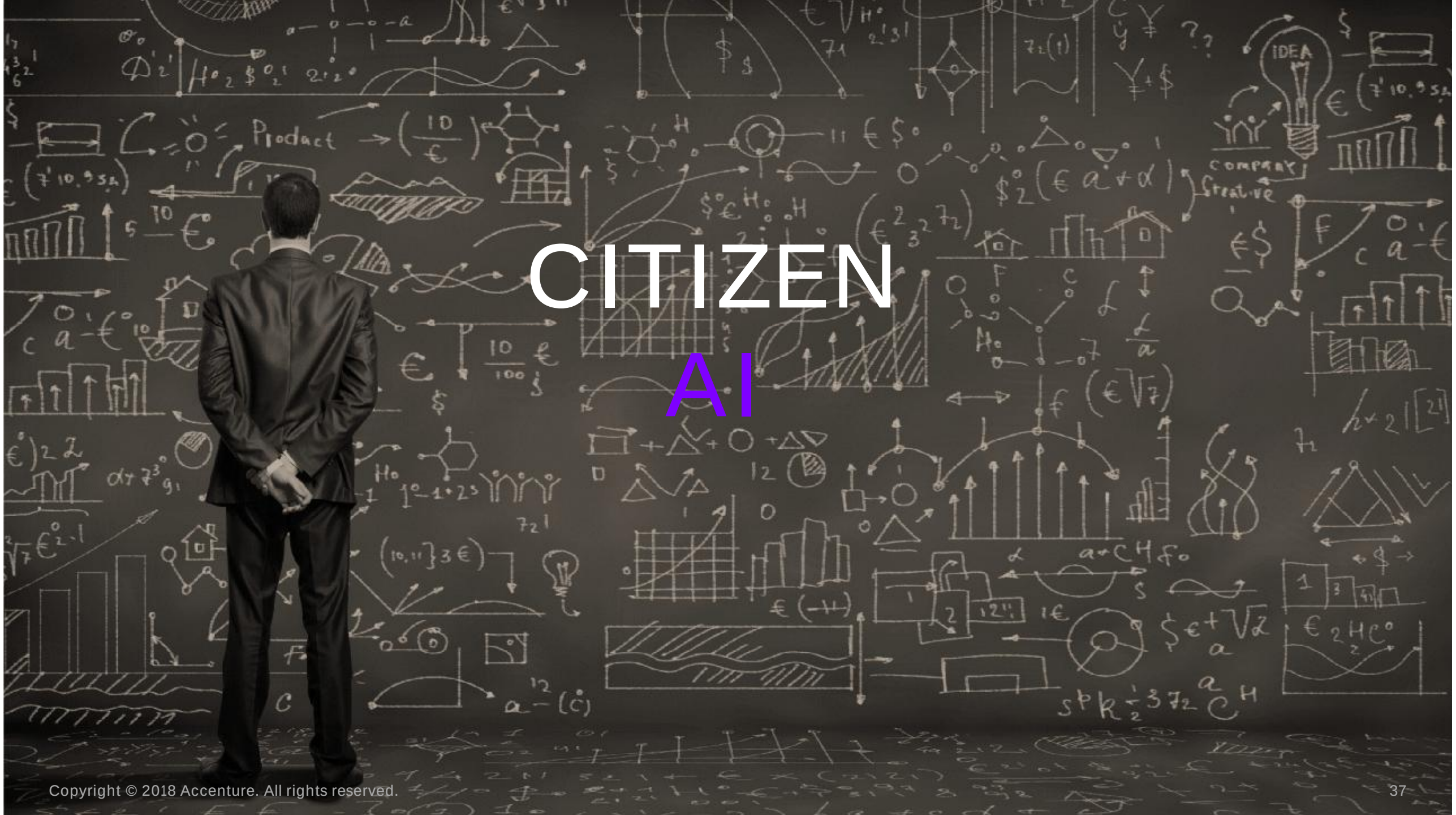
Periodically CDO reports to Top Management with synthetic metrics to provide a clear and exhaustive view about the status of Data Quality and the future evolution process

DATA GOVERNANCE IS CLOSELY RELATED OBJECTIVES

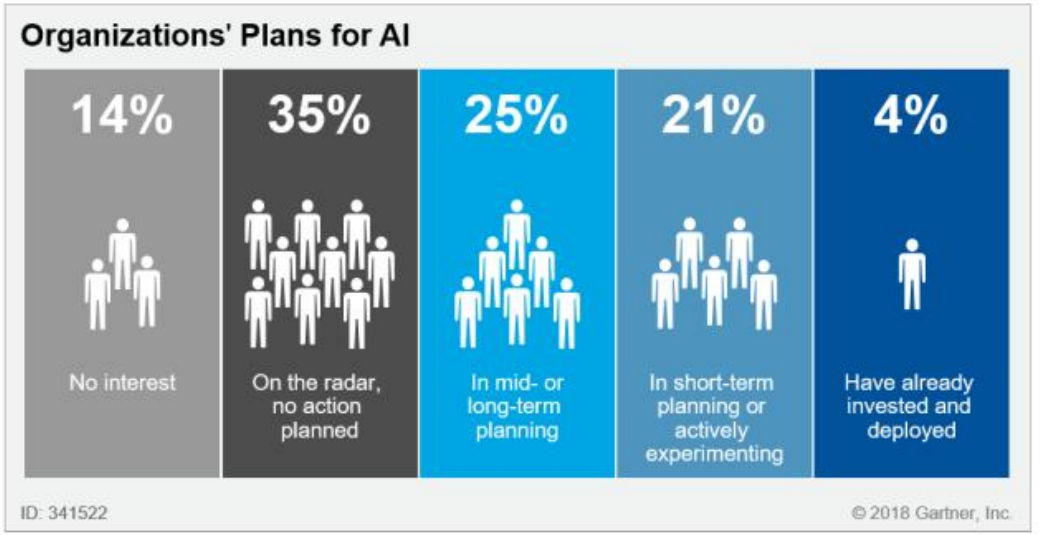
Data Governance leads to the definition of a Data Architecture able to satisfy the following objectives

Data Governance Objectives

Drivers		Description
A Data Management	A.1 Intelligibility	§ Increase the degree of comprehensibility and user awareness
	A.2 Documentability and Traceability	§ Document the entire chain of data lifecycle within the architectures in order to enhance the traceability requirements § Draw transformations, aggregations and changes occurred along the production chain of the data
	A.3 Data «Protection»	§ Document and report the results of data management activities in data migration, data back-up and remediation in order to ensure the protection and integrity of the data handled
B Data Quality and Data Checks	B.1 Governance Indicators	§ Understand and evaluate the level of effectiveness of the Data Governance, measure the coverage of the functional areas, as well as the timeliness of resolution of requests for intervention
	B.2 Data Quality Indicators	§ Continuously monitor the quality of the data, the performance of the control processes and the status and results of the action plans
	B.3 Formal technical checks	§ Report and measure the first level controls results



AS ORGANIZATIONS RELY MORE ON ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING MODELS, HOW CAN THEY ENSURE THEY'RE TRUSTWORTHY?



Question: "What are your organization's plans in terms of artificial intelligence?"
n = 3,138 respondents to the 2018 Gartner CIO Survey (excluding "Don't knows")

Source: Gartner (March 2018)

As Artificial Intelligence expands further into society, the business accountability around raising a responsible and explainable AI will rapidly grow.

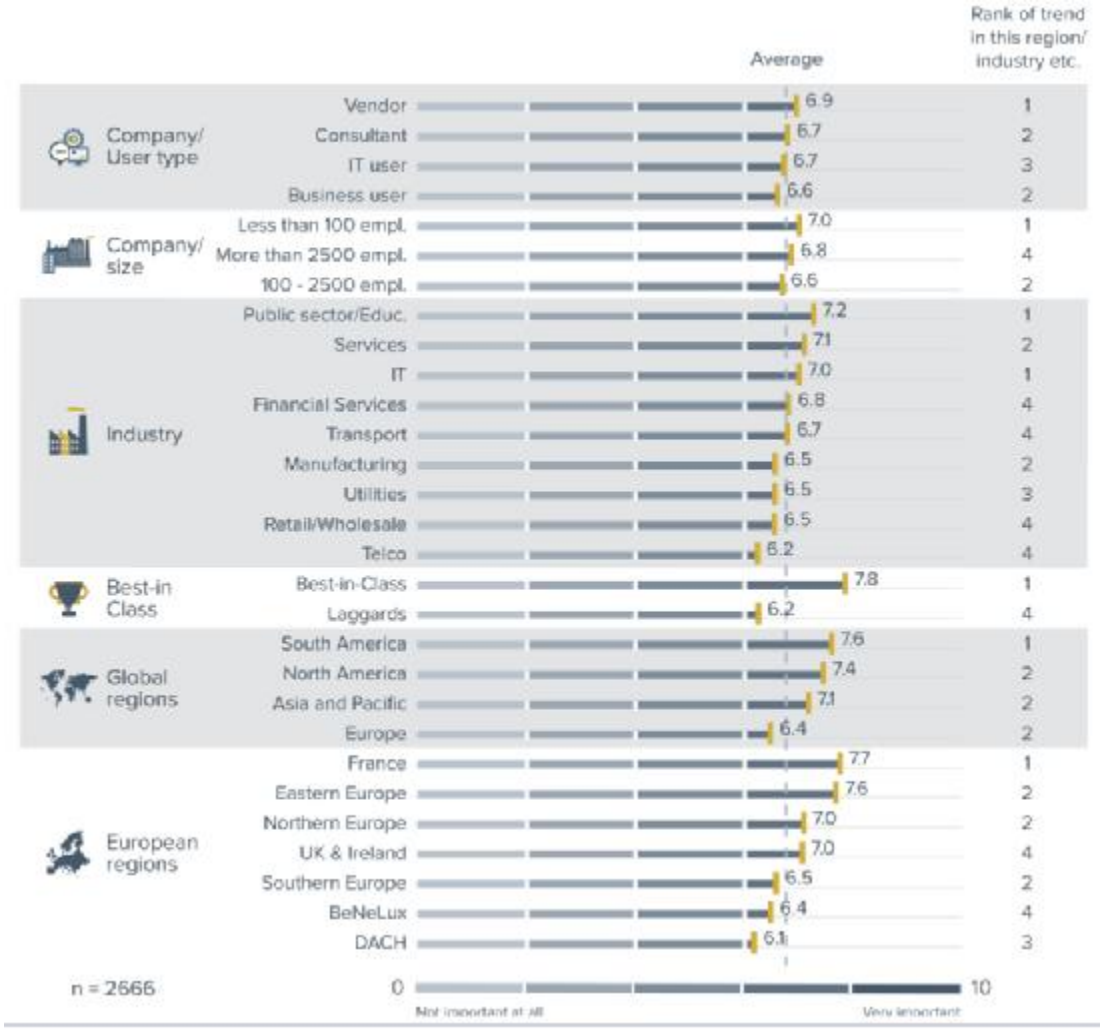
- Raising responsible AI means **addressing many of the same challenges faced in human education and growth**. Companies can look to milestones of human development for guidance. People learn how to learn, then they rationalize or explain their thoughts and actions, and eventually they accept responsibility for their decisions.
- Many machine learning applications don't currently have a way to "look under the hood" to understand the algorithms or logic behind decisions and recommendations, so organizations piloting AI programs are rightfully concerned about widespread adoption.
- Line of business leaders in organizations—particularly organizations concerned with risk like financial services and pharmaceutical companies—are demanding data science teams to use models that are more explainable and offer documentation or an audit trail around how models are constructed.

- 4/5 **executives (81%) agree** within the next two years, AI will work next to humans in their organizations, as a co-worker, collaborator and trusted advisor.
- **72% of executives report** that their organizations seek to gain customer trust and confidence by being transparent in their AI-based decisions and actions.

Source: © Gartner: Lessons Learned From Early AI Projects, Dec 2017

DATA DISCOVERY / & SELF-SERVICE BI

DATA DISCOVERY / SELF-SERVICE BI & VISUALIZATION

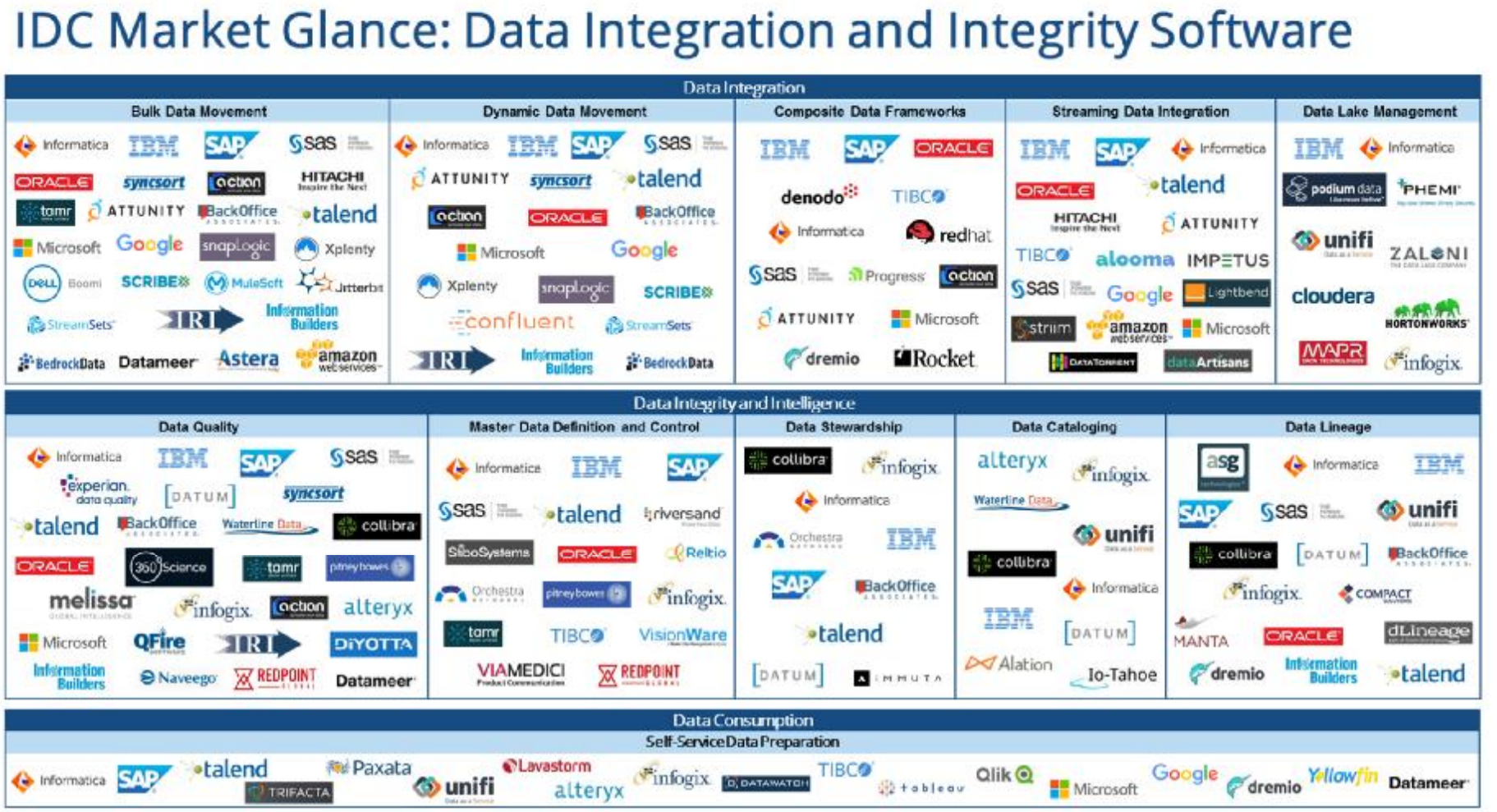


- Data discovery is the business user driven process of discovering patterns and outliers in data. It must cover and integrate at least three functional areas to efficiently and effectively identify patterns and outliers in an iterative approach. Business users must be well equipped with features for connecting diverse sources, cleaning, enriching and shaping data to create data sets for analytics (data preparation).
- Machine learning is increasingly being added to data discovery tools to guide business analysts through all steps from preparation over analysis to presentation.
- Self-service BI promises quicker and more efficiently prepared analyses and reports by empowering the business users involved to gain insight from data and make better informed decisions. The number of implementations that allow business users to build their own reports and dashboards or even explore data with guided advanced analytics and build data assets (data preparation) is increasing. Not all business users take part in actively creating BI and analytics content. It is important to understand self-service BI as a complement, not a supplement, to serviced or 'silver service' BI.
- Self-service BI elevates agility and speeds up the time to insight, but the quality of results or efficiency must not be sacrificed for agility.
- These data sets can be explored by using visual analysis or sifted by guided advanced analytics to find patterns not visible to the human eye in large data sets with a high number of variables.

SELF-SERVICE DATA PREPARATION UNDER THREAT AS INTEREST IN DATA CATALOGS GROW

CONTEXT

- Self-service data preparation software vendors are adding data visualization capabilities, while data visualization software vendors are adding deeper data preparation capabilities
- AI/ML is finding its way into automation of mundane data integration and integrity activities



Source: IDC, 1Q18 *For areas that IDC covers, the top 3-5 market share leaders are represented. For areas that IDC does not cover, vendor selection is up to analyst discretion.*

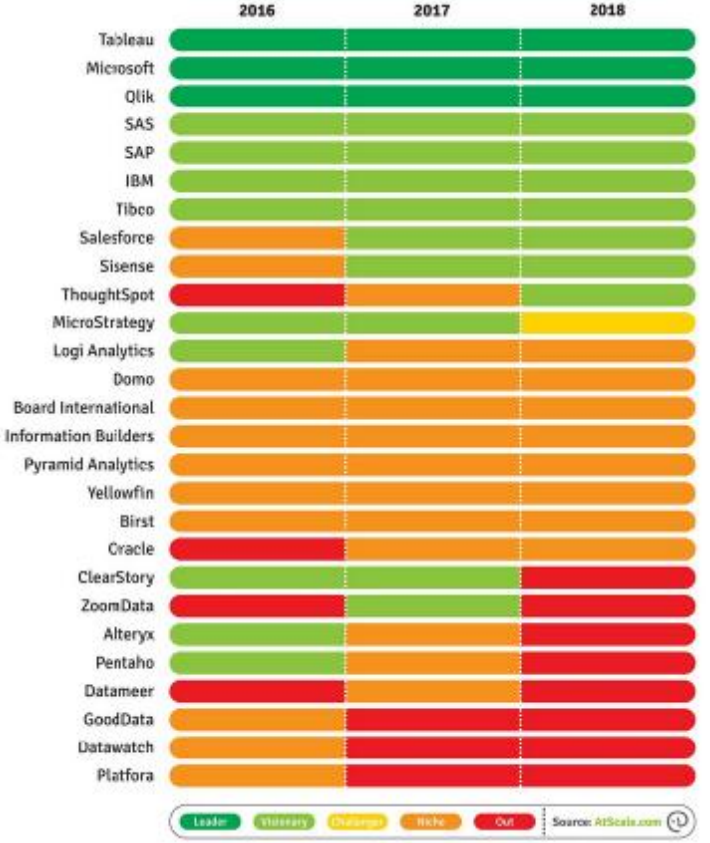
Source: © IDC: Market Glance: Data Integration and Integrity Software, 1Q18:

GARTNER MAGIC QUADRANT FOR BUSINESS INTELLIGENCE AND ANALYTICS PLATFORMS

It will not be possible to offer the full range of benefits with a single modern BI and analytics platform deployment because no tool or vendor offers the full spectrum of required capabilities

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2018)



- Key Considerations -

- Self-service BI has shaken the industry and has resulted in several unintended consequences. Business units have bought all the tools they could buy, and now I.T departments have to bring order to their BI chaos
- Data Lakes have become present everywhere. The average Chief Information Officer spends close to $20M a year to orchestrate Big Data analytics workloads across traditional data warehouses like Teradata, modern data platforms like Hadoop and next-generation serverless technologies like Google BigQuery.
- The Cloud, highlighted by Big Data, will drive 72% of enterprise CIOs to modernize their enterprise data architecture and economically scale their deployments.

Source: © Gartner: Business Intelligence & Analytics Platform, 1Q18:

DATA IS EVERYTHING.

How well you use your data can determine the degree of your success.

QUESTIONS ?

THE SOURCE

European research and advisory firm.

BARC is an European research and advisory firm. For over twenty years, BARC analysts have combined market, product and implementation expertise to advise companies and evaluate BI, Data Management, ECM, CRM and ERP products.

American research and advisory firm.

Gartner, Inc. is an American research and advisory firm providing information technology related insight for IT and other business leaders located across the world. Its headquarters are in Stamford, Connecticut, United States.

THE COMMUNITY

THANK YOU FOR ATTENTION!

Please, fill in the Feedback.

bit.ly/lojka_vse