DataBench

Evidence Based Big Data Benchmarking to Improve Business Performance

The market: Assessing Industrial Needs

BDVA Meetup
27 June 2019, Riga

Richard Stevens - IDC

IDC
ANALYZE THE FUTURE

Atos

Institut "Jožef Stefan", Ljubljana, Slovenija

SINTEF

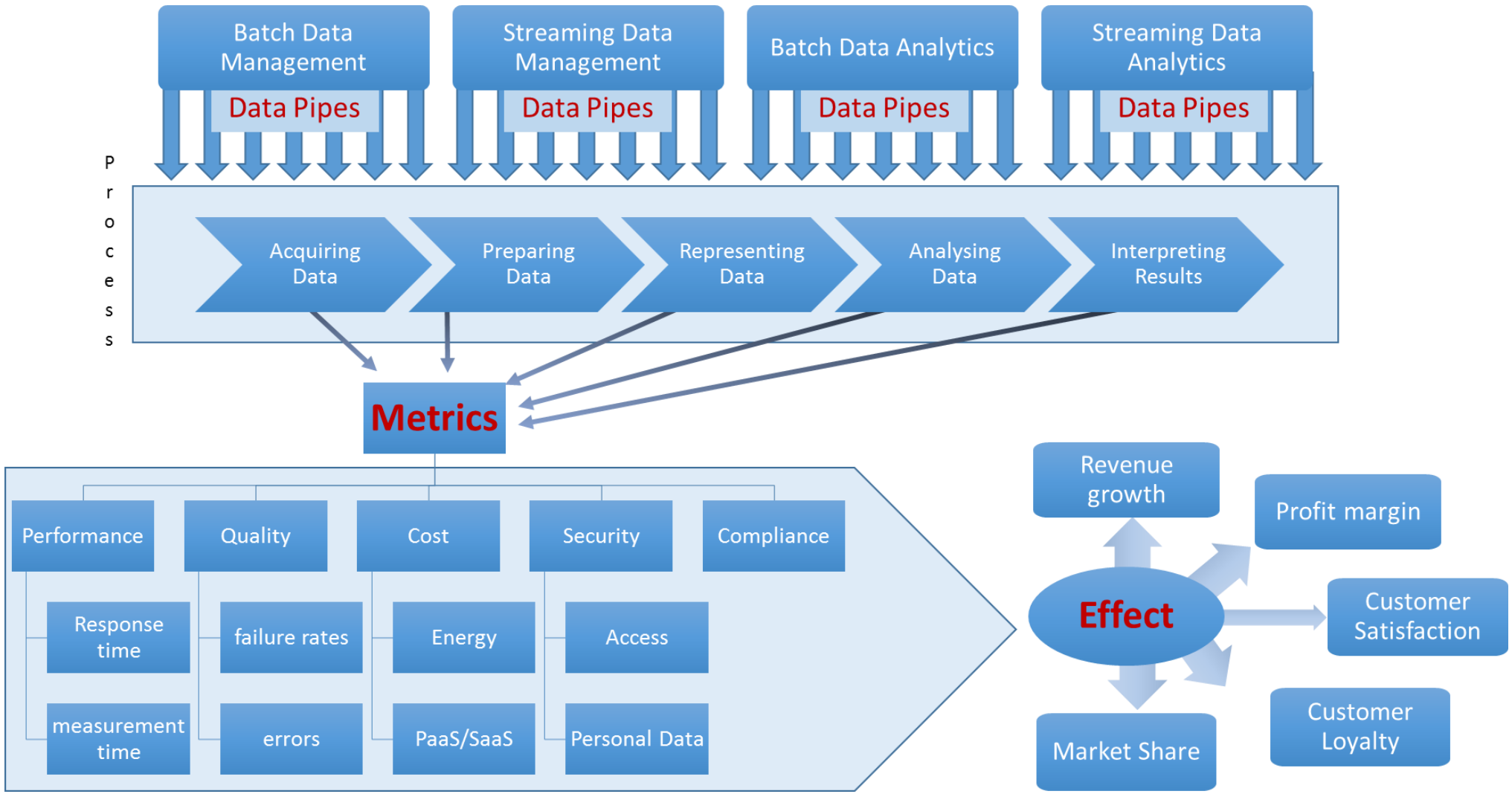
POLITECNICO
MILANO 1863

Frankfurt Big Data Lab
GOETHE UNIVERSITÄT

Outline

- ✓ Objectives
- ✓ Summary of analytics questionnaire
- ✓ Use case selection
- ✓ Case study analysis methodology
- ✓ Preliminary insights from case study analysis

Matching Needs to Measurements



How to Recognize Value of Big Data Technologies

Prepare a preliminary classification of the main business drivers and KPIs for companies using BDT

- Desk research using OECD / Eurostat / IDC data

Survey businesses using BDT and representing to rank importance of these KPIs

- 700 Companies from diverse business sectors
- different types and sizes of companies

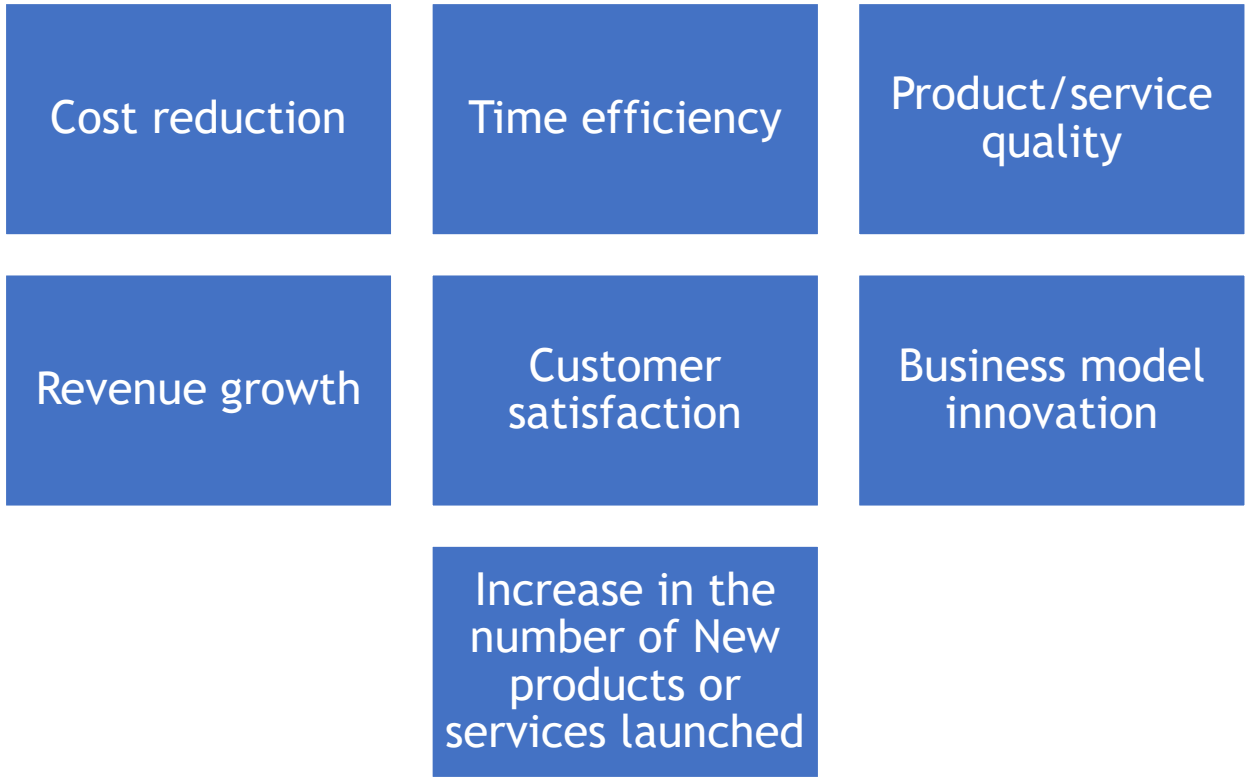
Perform detailed surveys about the technical infrastructure

- Size, type and approach to analytics

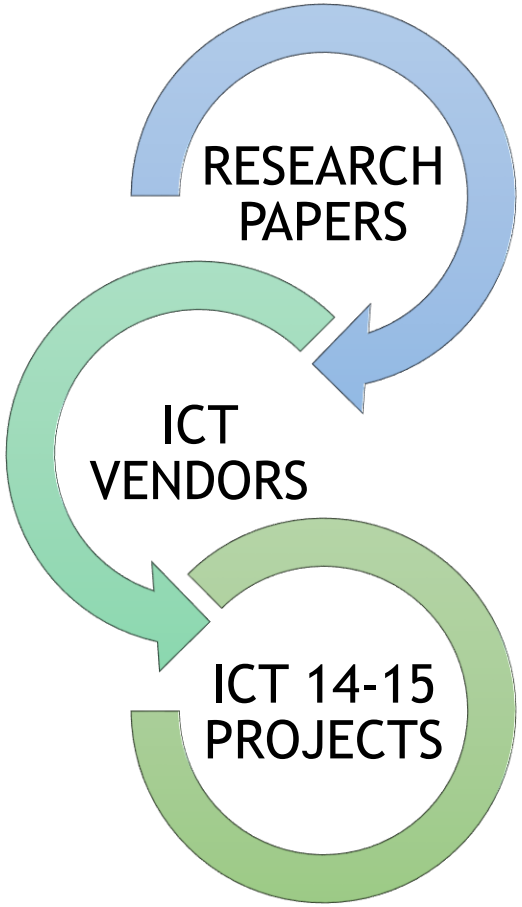
Correlated business data and technical parameters

What's important for companies

Respondents were asked about the 7 main KPI categories selected by the DataBench conceptual framework as measuring the most relevant business impacts:



Desk Research



Thorough examination of

- ✓over 100 research papers centered on measuring business performance
- ✓Characterization of use cases and vertical industry
  - ✓633 use cases in total
  - ✓59 use cases per industry on average

Comparing data from survey with data from the desk analysis provides **mainstream vs. innovation insights**.

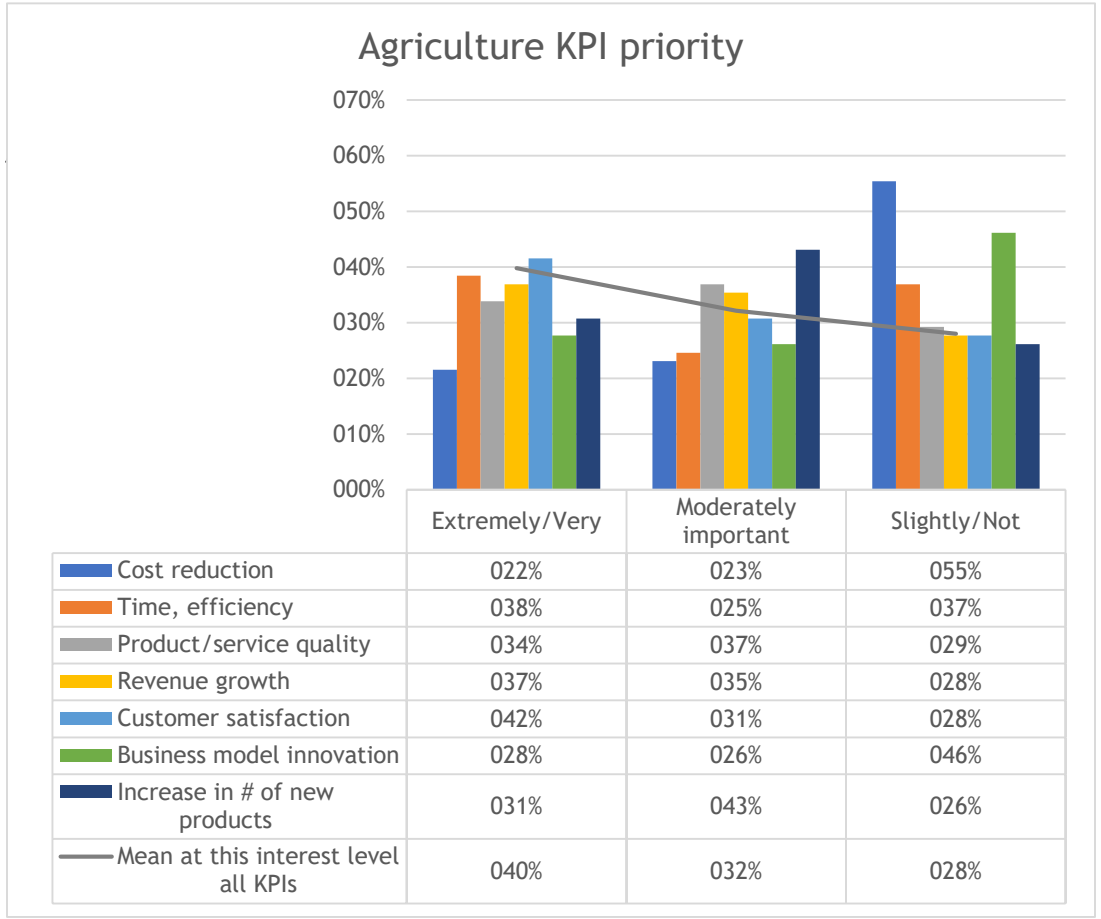
Dimensions and values of business features

Business KPI	Industry	Application Area	Level of business process integration
• Cost reduction • Time efficiency • Product/service quality • Revenue and profit growth • Customer satisfaction • Innovation	• Agriculture • Financial services • Business/IT services • Healthcare • Manufacturing • Retail/wholesale trade • Telecom/media • Transport/logistics • Utilities/oil & gas	• Customer service and support • R&D • Product innovation (new business initiatives) • Maintenance and logistics • Marketing • Finance • HR & legal • Sales • Product management • Governance, risk, and compliance • IT and data operations	• Low • Medium • High

Dimensions and values of the technical features

Data size	Data type	Platform type	Analytics type	Processing type	*Application level performance*
• Gigabytes • Terabytes • Petabytes • Exabytes	• Tables and structured data • Graph and linked data • Geospatial and temporal data (including time series and IoT data) • Media (image, audio or video) • Text and semi-structured data (XML, genomic data, etc.)	• Distributed • Centralized	• Descriptive • Diagnostic • Predictive • Prescriptive	• Batch (not in real-time) • Streaming (real-time) • Interactive/(near) real time • Iterative/in-memory	• Cost • Throughput • End-to-end execution time • Accuracy/quality/data quality/veracity • Availability

Example: Agriculture



	Extremely/Very	Moderately important	Slightly/Not
Cost reduction	022%	023%	055%
Time, efficiency	038%	025%	037%
Product/service quality	034%	037%	029%
Revenue growth	037%	035%	028%
Customer satisfaction	042%	031%	028%
Business model innovation	028%	026%	046%
Increase in # of new products	031%	043%	026%
Mean at this interest level all KPIs	040%	032%	028%

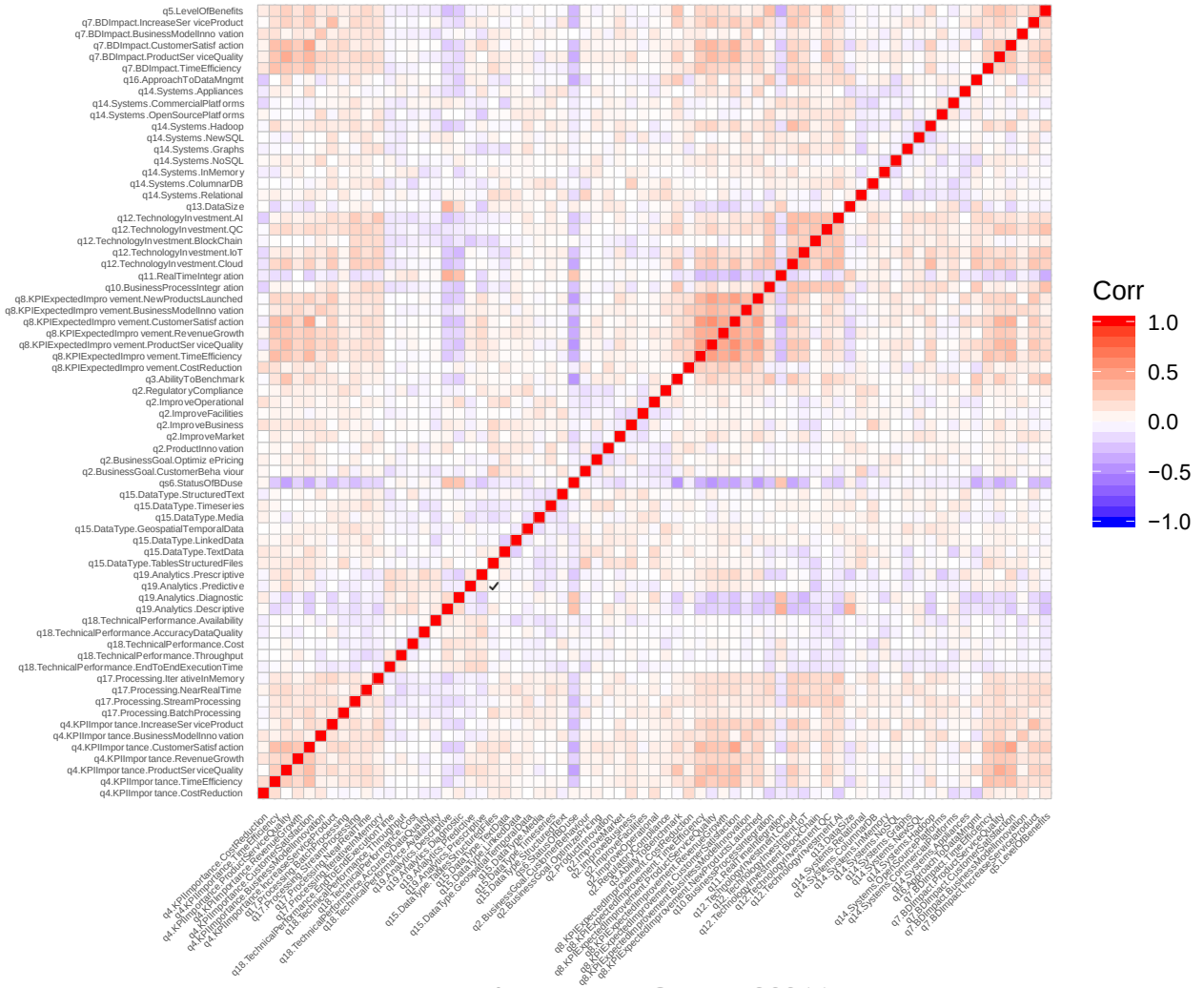
- KPIs

Importance	KPI
-	Cost Reduction
-	Time Efficiency
-	Product/service quality
-	Customer satisfaction
-	Business Model innovation

- USE CASES

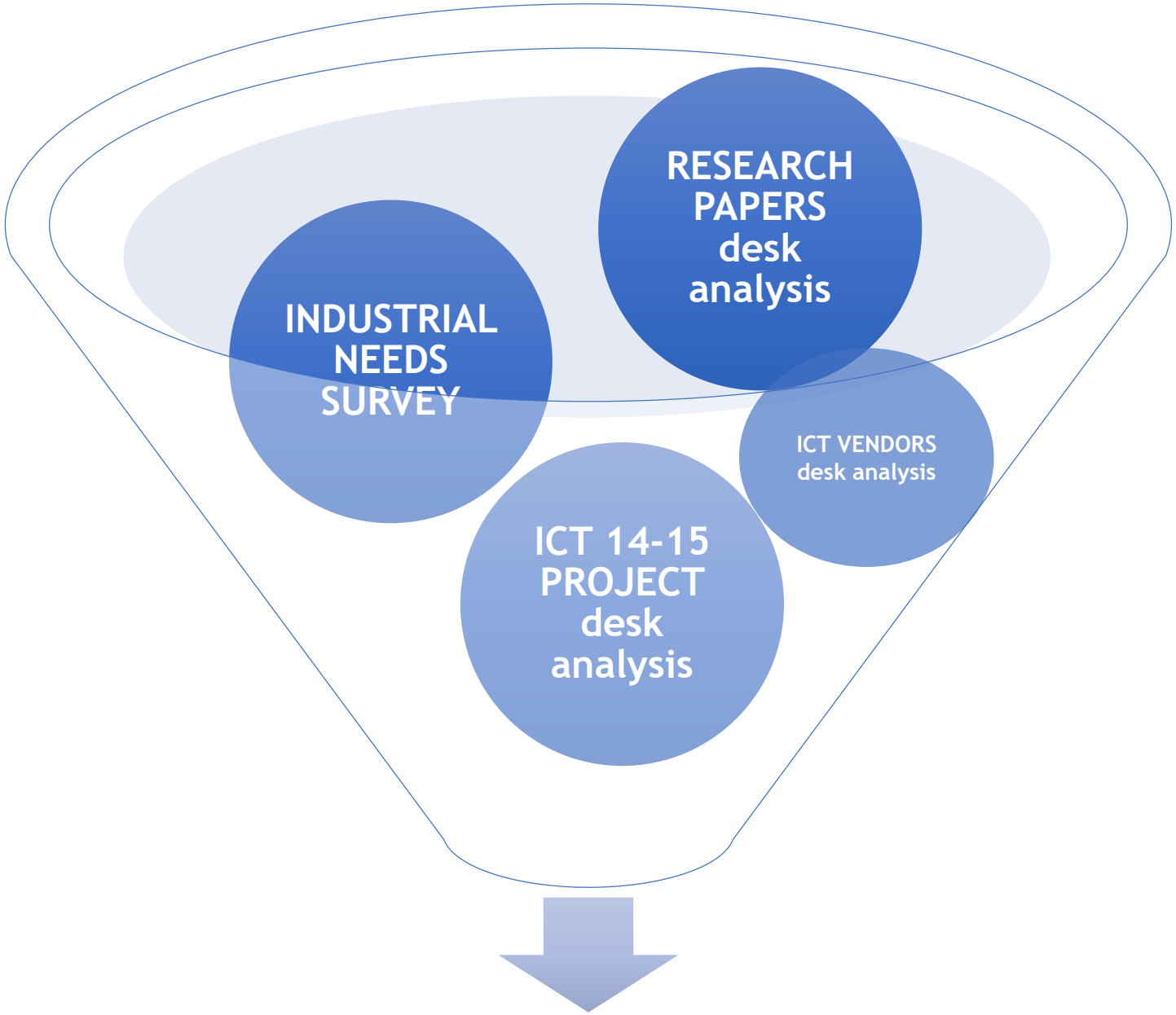
Use Case	# Responses	% Responses
Field mapping & crop scouting	44	68%
Price Optimization	42	65%
Inventory and service part optimization	42	65%

Statistical analysis: correlation matrix



Statistical analysis: factor analysis

- The factor analysis stresses that companies that have already obtained and measured business benefits from BDT projects are focused on traditional batch processing.
- In contrast, companies that experiment with more advanced real time applications of BDTs have not yet measured business benefits.
- Moreover, companies that have not yet exploited BDTs or have a traditional exploitation of BDTs (batch) are technology enthusiast and/or plan to explore more innovative applications of BDTs, but do not view future business benefits as measurable with economic KPIs at this stage of development of BDTs.



Use Case Analysis

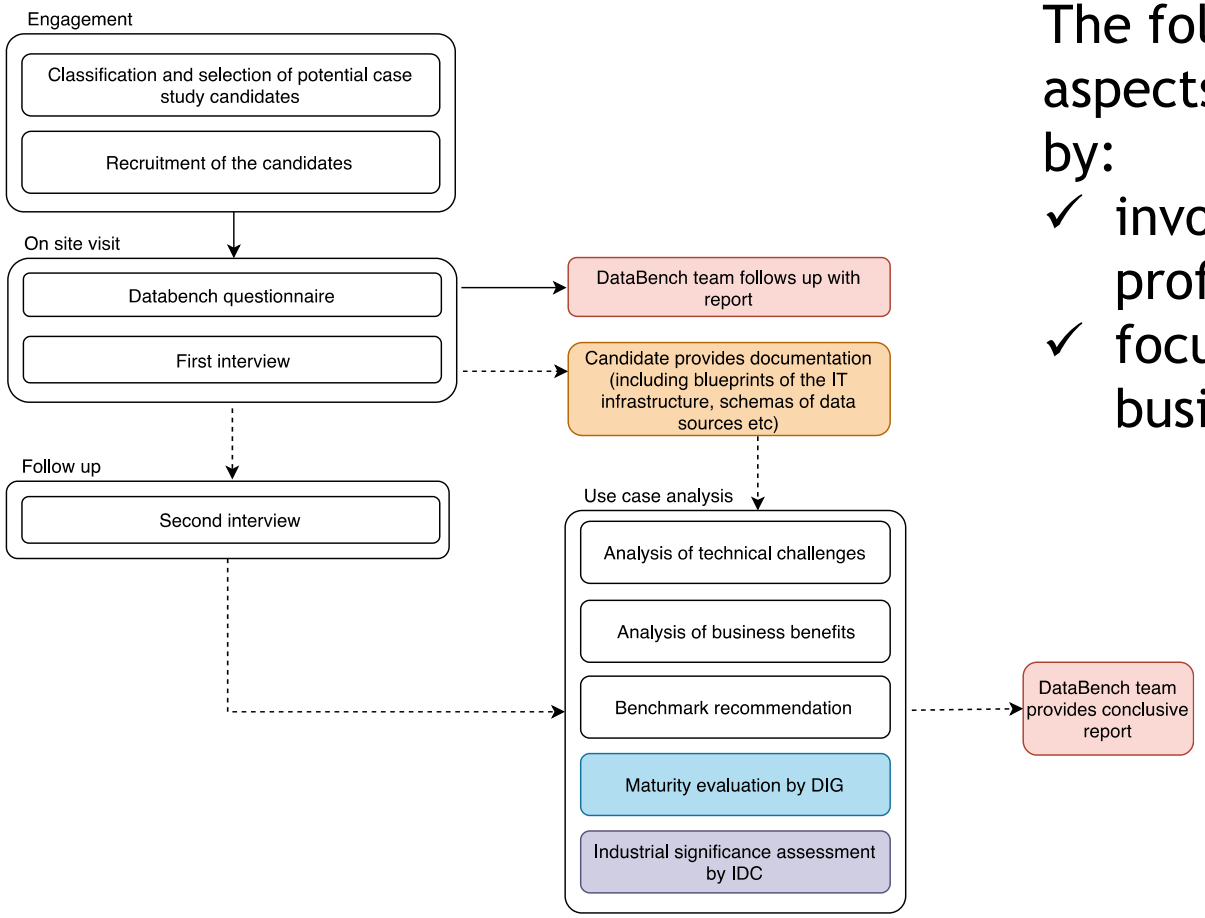
Use case selection criteria

The list of *use cases* is based on the IDC industrial needs survey.

The list of *DataBench use cases* was defined by:

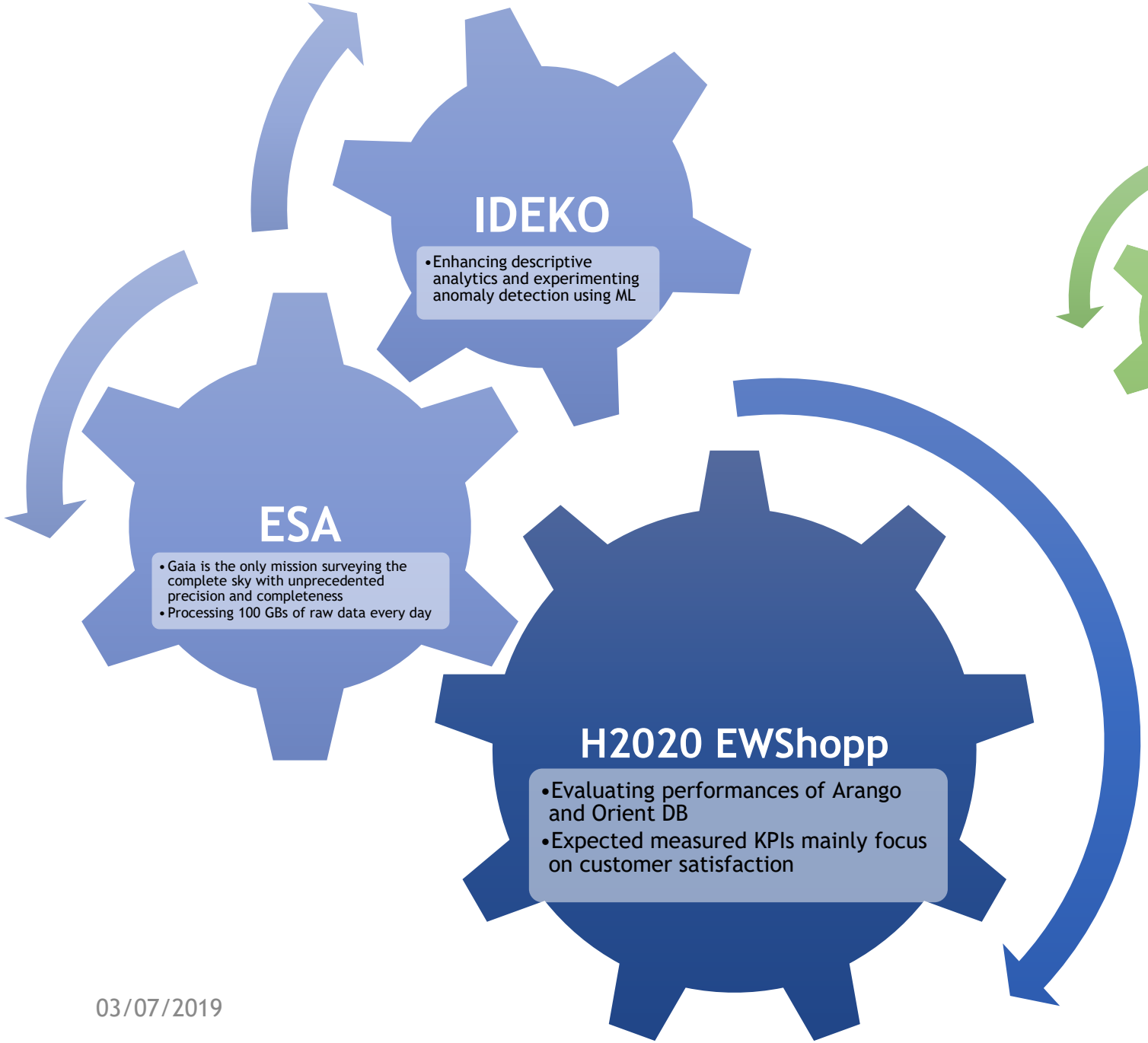
- ✓ using one use case from *state of the art use cases list* -> to be able to assess the business KPIs,
- ✓ using one use case from the desk analysis -> to account for research and emerging use cases,
- ✓ preferring use cases specific to the industry -> to make it easier to identify pilots with quantitative business KPIs,
- ✓ keeping some cross industry use cases, e.g., supply chain optimization is a topic in manufacturing as well as in retail.

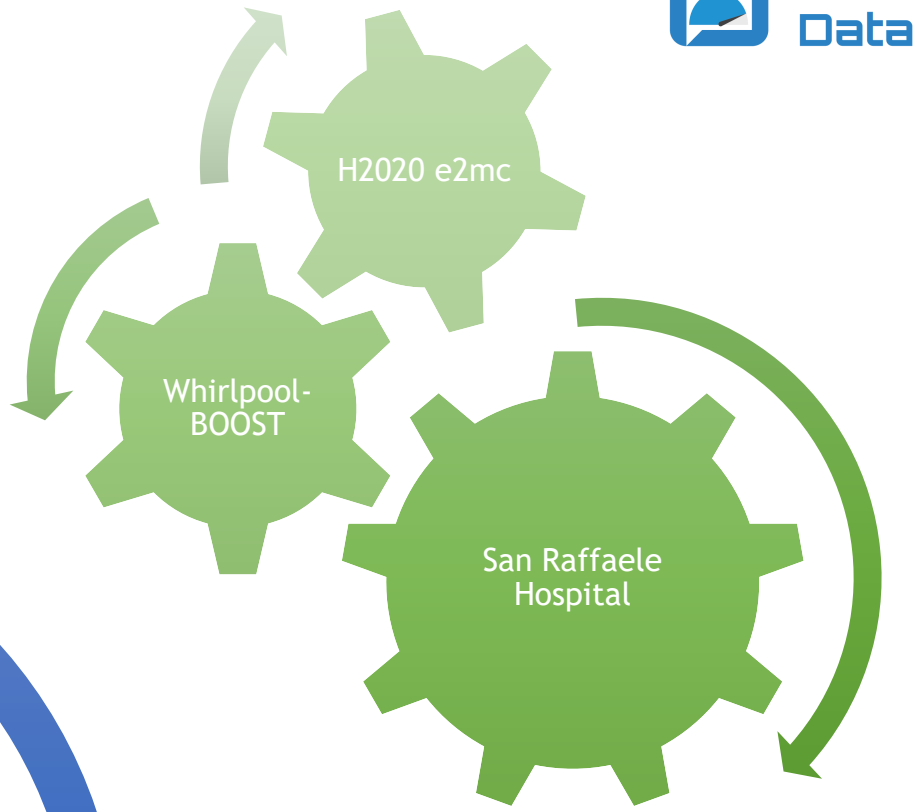
Case study analysis methodology



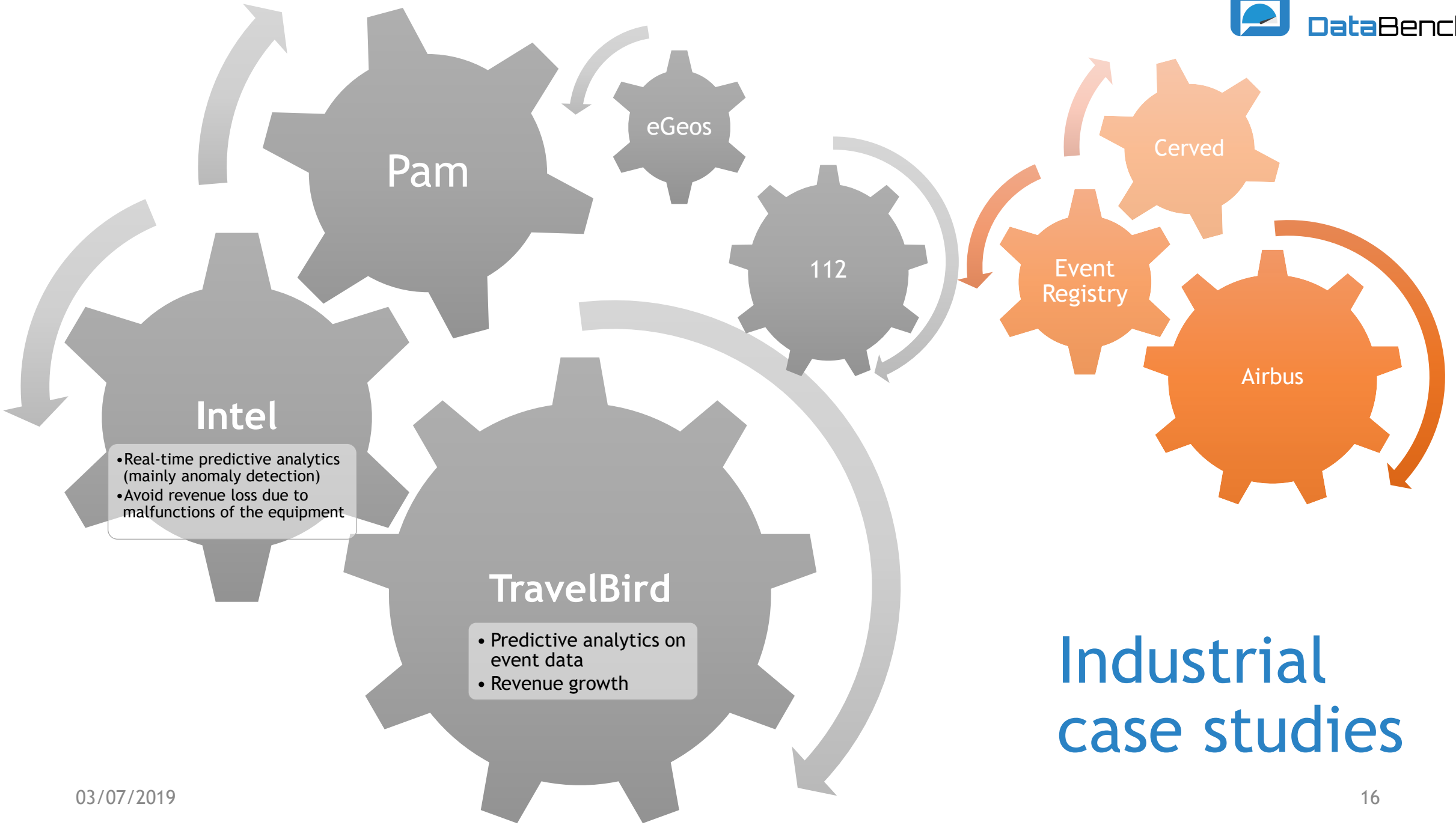
The follow-up interview should cover the aspects/perspectives missed by the first interview by:

- ✓ involving respondents with a more specific profile,
- ✓ focusing on the collection of quantitative business KPIs.





Research case studies



Industrial case studies

Case study analysis: cases by industry

Retail & Wholesale	•Pittarosso •Pam •H2020 EW-Shopp
Agriculture/EO	•eGeos •ESA
Manufacturing	•INTEL •Fater •Whirlpool
Healthcare	•San Raffaele Hospital
Business / IT Services / AI	•TravelBird •Event Registry •Ideko
Transport and Logistics	•Siemens •112
Financial Services	
Telecom/Media	
Utilities / Oil & Gas	

Early Insights

- From the evidence that has been collected so far, an important lesson learnt is that **most companies believe that technical benchmarking requires highly specialized skills** and a considerable investment. We have found that very few companies have performed an accurate and extensive benchmarking initiative. In this respect, using **DataBench like Solutions grants** them with an easier access to a broader set of technologies that they can experiment with.
- On the other hand, they acknowledge the variety and complexity of technical solutions for big data and envision the following **technical risks**:
  - The risk of realizing that they have **chosen a technology** that proves non scalable over time, either technically or economically.
  - The risk of relying on cloud technologies that might create a lock in and require a considerable redesign of software to be migrated to other cloud technologies.
  - The risk of discovering that cloud services are expensive, especially as a consequence of scalability, and that technology costs are higher than business benefits.

Evidence on business KPIs from case study analysis

- We have evidence of business KPIs for case studies where we have reached the pilot stage according to our case study methodology.
- Evidence is aligned with results from survey (business benefits are in the 5-8% range).
- We have already performed 4-5 case studies at pilot stage confirming results
- Business KPIs are seldom quantified in cases studies from the literature (most projects are at POC level).
- From the desk analysis, multiple business KPIs are affected simultaneously and the benefits from a single project are often difficult to isolate from other factors affecting the same business KPI.

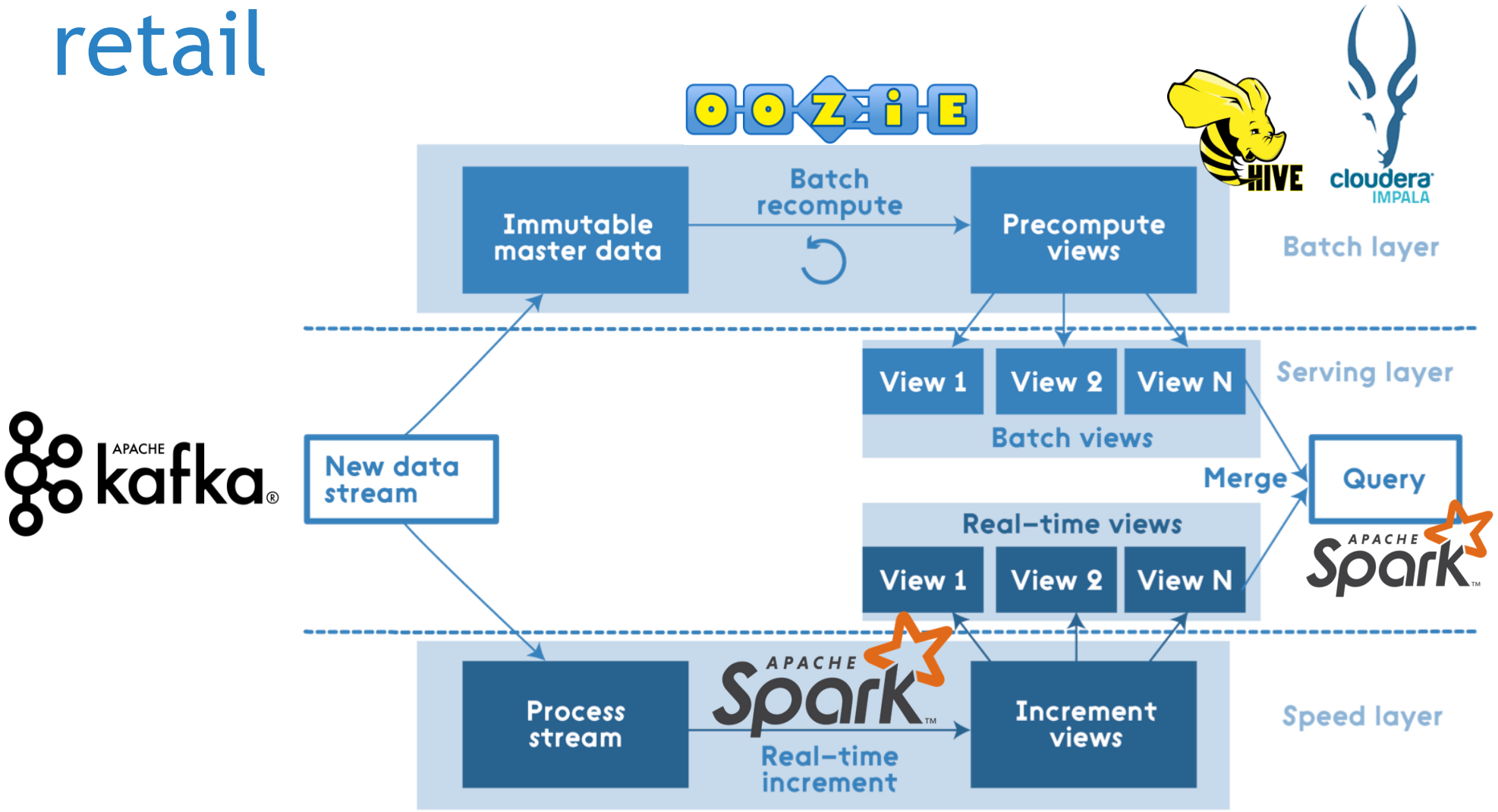
Sample case study: intelligent fulfilment in retail

- Automatic replenishment optimization is a fundamental process in the retail industry, it involves multiple departments, e.g., logistics, order management, etc., and affects sales through stockouts and, thus, revenues and customer satisfaction.
- Currently, in the analyzed case study the replenishment process is carried out manually by the point of sale (POS) employees that periodically check for the presence of a sufficient number of items/products to fulfill the expected demand for the following days.
- The manual process has evident drawbacks, as it is error prone and suffers of the bias introduced by the judgment of the employee.

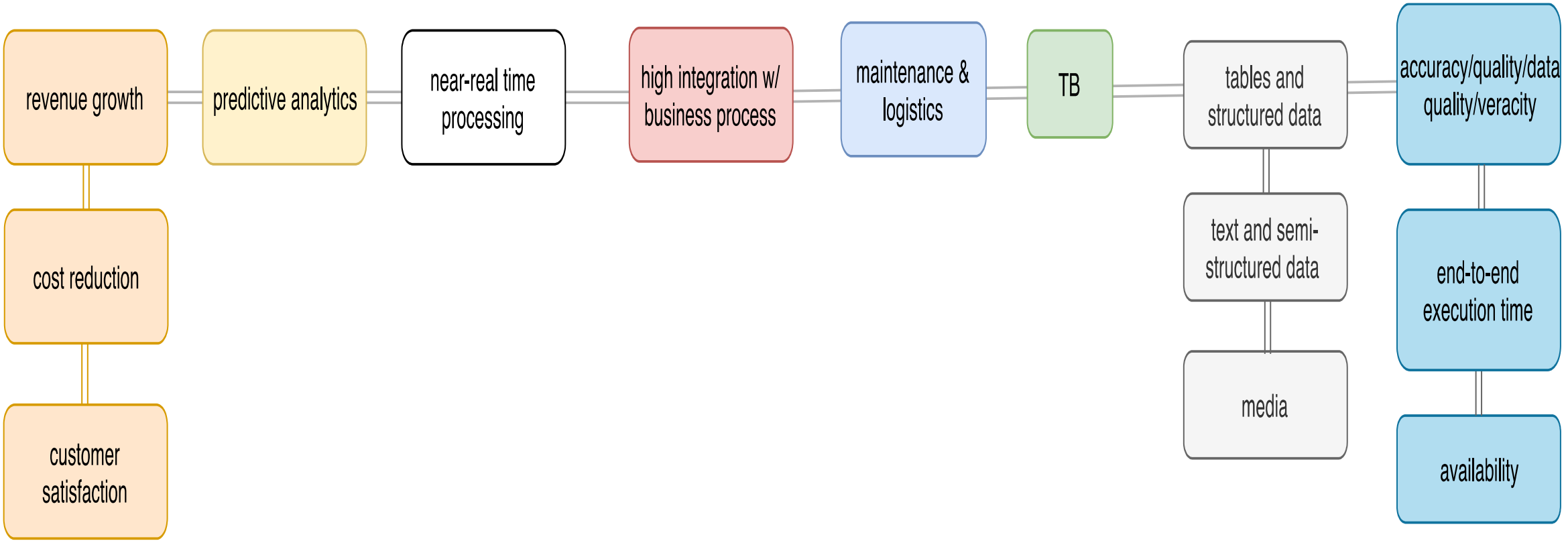
Sample case study: intelligent fulfilment in retail

- The company is piloting an automatic replenishment procedure that will optimize the order scheduling process by using a set of sensors to detect the number of items on the shelves and by adopting a machine learning algorithm to forecast products demand.
- Overall, the main goals of the case study are to improve the quality of the service provided to customers and to improve the efficiency of the replenishment process.
- From a technical perspective, innovative aspects of the new automatic replenishment procedure include:
  - ✓ the adoption of an ad-hoc prediction algorithm for product demand forecasting,
  - ✓ the positioning of a set of image sensors able to monitor in real-time the number of items on the shelves,
  - ✓ the adoption of an image recognition algorithm able to identify the number of products on the shelves, and consequently to identify mis-placed items, items positioned inaccurately, etc.

IT architecture: intelligent fulfilment in retail



Blueprint


revenue growth
predictive analytics
near-real time processing
high integration w/ business process
maintenance & logistics
TB
tables and structured data
accuracy/quality/data quality/veracity
cost reduction
text and semi-structured data
end-to-end execution time
customer satisfaction
media
availability

Business KPIs: intelligent fulfilment in retail

- Business benefits carried by the intelligent fulfillment are manifold.
- General indicators, useful to assess the effectiveness of the process, include:
  - ✓ revenue growth due to avoided lost sales,
  - ✓ customer satisfaction,
  - ✓ improvement in the efficiency of the fulfillment process, that results to be more structured and organized,
  - ✓ more specific indicators useful to assess the efficiency of the intelligent fulfillment process include number of stockouts, inventory turnover and, with a focus on logistic efficiency, mean time between orders. In this context, the inventory turnover measures the time spent by an item in the warehouse, high levels of storage represent an undesirable condition because they increase storage management costs.
- The case study is still in its piloting stage and it has not yet delivered quantitative business benefits. Nonetheless, it is providing deep insights to the whole POS management and efficiency.
- Economic scalability issue has been recognized: cameras acquiring images have to be placed on shelves in 250 POS (over 40 Km of cameras).
- Technical scalability issue: is the lambda architecture scalable with image processing? At what costs?

Get in touch with us!


info@databench.eu


@DataBench_eu


DataBench


DataBench Project


DataBench


DataBench Project


DataBench

www.databench.eu