ROMA TRE
UNIVERSITÀ DEGLI STUDI

FAST DATA FOR RAPID DECISION MAKING: THE ROLE OF DATA VIRTUALIZATION

Alberto Pezzi, Emiliano Pierdominici

ISSN 2611-9633

Working Papers (Dipartimento di Economia Aziendale)

[online]

Working Paper Numero 14, 2021

Collana del Dipartimento di Economia Aziendale

I Working Paper del Dipartimento di Economia Aziendale svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali.
La loro pubblicazione è soggetta all'approvazione del Comitato Scientifico.
Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. l del D.L.L. 31 agosto 1945 n. 660 e successive modifiche.
Copie della presente pubblicazione possono essere richieste alla Redazione.
Esemplare fuori commercio ai sensi della Legge 14 aprile 2004 n. 106.

REDAZIONE
Dipartimento di Economia Aziendale
Università degli Studi Roma Tre
Via Silvio D'Amico, 77
00145 Roma – Italia
Email:
ricerca.economiaaziendale@uniroma3.it

COMITATO SCIENTIFICO
Maria Claudia Lucchetti
Carlo Mottura
Mauro Paoloni
Maddalena Rabitti
Carlo Maria Travaglini

ABSTRACT

Fast decision making is becoming a crucial element for companies, as the market is requesting more flexibility and agility from companies in order to be really competitive. The complexity coming from the enormous increase of data volume generated worldwide and the availability of real time data at an unprecedented growing rate, force organizations to adopt new approaches and strategies. The COVID-19 pandemic prompted organizations to re-design their strategy and processes, with particular focus on data analytics and insights, so these capabilities become a key element for companies and organization of any size. The concept of a data driven organization has recently become popular due to the adoption of several new technologies that are supporting new data platforms and architectures worldwide. One of the new enabling technologies is Data Virtualization, that allows organizations to manage and integrate data at a logical level, more efficiently than with the traditional approach, with a 50% to 90% advantage in terms of time and cost reduction. The benefits can be defined as a "Value Chain", providing faster data delivery, time saving, speed of action, and ultimately advantages in terms of cost savings, revenue increase and risk mitigation. Speed is a strategic element to allow faster and cost effective decision making, and speed is what is delivered by innovative software technologies like Data Virtualization. Technology adoption is a key element, but organizations who want to be flexible and data-driven, need to also adopt a clear strategy to manage people and process accordingly. Culture and know-how are required as well as innovative technology to transform companies in fast decision-making organizations. We analyzed 170 projects worldwide in 21 different industries, and presented several case studies where Data Virtualization solution provided real business benefits, for banks, telecom service providers, Government agencies, high-tech and manufacturing organization.

Keywords: Data virtualization, data analytics, decision making

© Alberto Pezzi, Emiliano Pierdominici

Corresponding author: Alberto Pezzi, email address: alberto.pezzi@uniroma3.it

1. INTRODUCTION

The post-COVID-19 era is posing unique challenges to decision makers who are pressed for faster action without breaking the business. The pandemic has shown that having the right data at the right time is crucial to make efficient decisions.

Traditional economic indicators and company data can be inadequate owing to issues of timeliness and granularity. Companies in a wide range of industries and countries have accomplished difficult tasks and achieved faster data delivery, which result in less time to use data and take critical decisions in a very competitive environment.

Re-thinking the entire organization, through the so-called digital transformation, means increasing the speed of decisions and accelerating the scope and scale of innovation. COVID-19 has unveiled the pivotal role of technology and the way data are used and consumed for business purposes. "Technology changes, economic laws do not" (Shapiro and Varian, 1994), this simple idea describes the importance of the real economic value, the effectiveness of a technology to be widely adopted. There are many technologies that have been developed in the last 50 years, in particular in the software and hardware industries, but the ones that are really relevant and have been implemented, are the ones that delivered economic and business benefits, in terms of productivity and cost savings compared to the existing ones.

In the last years, costs of data collection, storage, and analysis have been decreasing, and companies have the potential to change many processes in their organizations (McAfee and Brynjolfsson 2012). At the same time, changing organizational systems to adapt to new technologies is costly, time-consuming, and risky (Bresnahan and Greenstein 1996, Brynjolfsson and Hitt 2003, Tambe and Hitt 2012) and dependent on other systems within firms (Milgrom and Roberts 1990, Rivkin 2000).

Our study emphasizes the managerial practices required to take advantage of new digital resources in the post COVID era. Companies deployed several new architectures and technology (e.g. data lake, AI models, cloud application, analytics tools) that increased complexity and are not designed to follow the current need for speed, flexibility, and innovation.

Therefore, companies need to plan a different architecture for their data platform/stacks to maintain and increase competitiveness, leveraging cloud and new concepts and technologies. Several enabling technologies can be

used to re-architect the existing platform and to reach significant benefits in terms of speed, flexibility, and speed of delivery:

- Cloud service providers (e.g., AWS, Microsoft Azure, Google Cloud);
- SaaS solutions (e.g., SalesForce; Pega, SAP, Oracle);
- Cloud-Based Data Platforms (e.g., Amazon S3, Google BigQuery);
- API management platform (e.g., Google Apigee, Mulesoft);
- Real-Time data processing (e.g., Apache Kafka);
- Data virtualization (e.g., Denodo; Tibco, Dremio);
- Flexible data schemas (e.g., Graph databases, Azure Synapse Analytics).

In particular, we are interested in analyzing managerial practices that are oriented to predictive analytics and can be adopted to integrate data and leverage them in a much more effective way, the so-called data virtualization - with great benefits in terms of operational advantages and strategic opportunities. In fact, almost every company that was interested in predictive analytics accepts there were data related problems that generated a negative effect on business and extra-costs to re-conciliate data. An efficient enabling solution must process the continuous new information needs of a company. During the COVID-19 emergency, data virtualization is very appealing for companies because they can reduce costs and implementation time , while increasing their agility, and reusing the data. Adopting a data integration tools portfolio should respond to various integration needs and should support the data access requirements, thus representing a possible solution to taking real time competitive decisions that can respond to the present market's uncertainty and economic conditions.

We have studied 170 projects in 21 different verticals developed by Denodo, to understand the reasons and the implications of the adoption of data virtualization. The outcome demonstrated the effectiveness of this technology from a business standpoint not only in operations benefits but also at a more strategic level. Data virtualization can bring cost savings in an IT infrastructure, but also allow the creation of new revenue opportunities and re-thinking of the business model.

2. DATA DRIVEN ORGANIZATION BENEFITS AND CHALLENGES

In the recent years, one of the most interesting trends has been the evolution of how data are managed and analyzed in companies. The "Data Driven Company" concept is not new, but highlights the companies' interest to get the highest value from the data and the digital transformation process. Digital transformation of business is highly interconnected with the rise and evolution of several trends of the last 20 years such as the increase of connected devices, higher computing power at lower hardware and storage cost, cloud adoption and the analytics and Big Data concepts.

Data will influence any business decision not only at an operational level, but also at a strategic level. Machine Learning and Artificial Intelligence will play an increasingly crucial role in business operations. Data is at the heart of the digital transformation process, as most companies rely on data and will increase this dependence in the future. Digital transformation is not related to the evolution of devices, but it is about the integration of data into everything we do. Always tracking, always on and always learning (Rydning, 2018).

Companies are collecting data at unprecedented speed and volume, and this increase is not going to slow down. Many organizations are leveraging data and analytics to improve their performance, but there are still some challenges to use data properly, or at their full potential.

However, many executives consider their data's potential hasn't been reached yet, mainly because it is still separated in different siloes and systems. In a research by PWC (2019) it is highlighted that executives believe they can get an enormous value just with the possibility to better access and use data they already have inside the organization, with a potential effect on cost savings by 33% and increased revenue of 31%.

The main benefits expected from companies to be data driven are to get better insights to better address customer needs, to achieve better decision-making and gain process and cost efficiency. The real effectiveness of a data driven organization is multidimensional and depends not only by the adopted technologies, but also by the people's know-how and by the organizational culture.

The biggest barriers companies face in extracting value from data and analytics are organizational, because there are many struggles to incorporate data-driven insights into day-to-day business processes. Leading companies are using their capabilities not only to improve their core operations, but to launch entirely new business models. We believe there is still a great potential in appropriately leveraging data and analytics and the value is still uncaptured in different ways in different verticals. The lack of analytical talents is the typical barrier, alongside technological barriers (i.e. legacy software/hardware systems and infrastructure; lack of centralized platform; data silos; user friendly systems, etc.) which companies must overcome to perform in the data driven economy.

The improvements in data collection, storage, and processing capabilities have created new opportunities for managerial activities - i.e., data-driven decision making (Brynjolfsson and McElheran 2016). Data is considered "the new oil" (Acito and Khatri 2014) and emphasized as a source of competitive advantage (Lambrecht and Tucker 2015).

Brynjolfsson and McElheran (2019) found systematic evidence showing that data-related management practices are causally linked to better performance in a wide range of operational settings. Firstly, data-driven decision making is correlated with improvements in revenue-based productivity of 4-8%. The timing and the IT investments are essential to understand the magnitude of managerial decisions. Data driven decision making has a stronger relationship with performance in earlier years and a stronger correlation with IT infrastructure investments in later years, as diffusion increases (Athey and Stern 1998; Brynjolfsson and Milgrom 2013).

A data driven organization starts from the assumption that imperfect yet "more informative" sets of inputs weakly improve performance. Most companies have reorganized their data architecture, collect data from novel sources, design algorithms to model data and derive insights, and hire or train the talent to do it all. The investments for these efforts have been very high and have led to more-precise information through higher-fidelity measurement. The tools and techniques for data collection, management, and analysis have allowed firms to run online experiments and leverage increasingly powerful empirical techniques to uncover new relationships (McAfee and Brynjolfsson 2012, Chen et al. 2012, Tambe 2014).

Empirical studies found the benefits of using data and automation in decision making by individuals in a wide range of settings (e.g., Pierce et

al. 2015, Hoffman et al. 2017, Kleinberg et al. 2017) and to mitigate the psychological biases in decision making (e.g., Milkman, et al. 2009).

At the organization level, the benefits of using data are associated to IT investments to change organizational design (Brynjolfsson et al. 1994; Brews and Tucci 2004; Ray et al. 2009; Bloom, Garicano et al. 2014; Rawley and Simcoe 2013; Forman and McElheran 2019), business model strategy (Gambardella and McGahan 2010, Casadesus-Masanell and Zhu 2013), and competitive advantage (Bharadwaj 2000, Ray et al. 2009). Technologies that enable greater information collection and facilitate its extraction improve companies' performance and reduce resource misallocations with lower production and coordination costs (David et al. 2016).

A data driven organization helps managers to legitimate their decisions through accurate data and objective quantitative measures that can improve the quality of the decisions. More accurate data causes managers to replace managerial intuition with objective information (Brynjolfsson, and McElheran, 2019). Objective data and information improve the relationship inside the organization and the transformation of tacit knowledge in explicit.

A data driven organization increases employee engagement and results in more effective managerial leadership in companies (Gibbons and Henderson 2012). However, companies have to invest in greater standardization and instrumentation of processes that "involve managerial time and attention, as well as the opportunity cost of worker effort devoted to these discovery, formalization, and socialization processes instead of producing more measurable outputs" (Brynjolfsson, and McElheran, 2019).

Industry and country determinants can affect the benefits and the costs to become data driven. Initial investments are significant and the industries that are already skilled in collecting vast quantities of data (such as financial services, telecommunications, lean manufacturing), can more easily turn to predictive analytics to reduce uncertainty and improve efficiency in the firm, market, and supply chain outcomes (Lambrecht and Tucker 2015, Dilda et al. 2017, Agrawal et al. 2018).

At the same time, the importance of complementary investments and organizational adjustment are strictly dependent on technological progress in terms of computers, software applications, and network infrastructure as well as the government policy and culture (e.g., high speed connection, 5G infrastructures; privacy and competition laws).

Adopting improved technologies and data-centric practices are likely to result in higher relative returns (Karshenas and Stoneman 1993, Stoneman 2002). The sustainability of competitive advantage is longer when the cluster of complementary investments is more complex, embedded, and causally ambiguous. The performance gains should decrease while correlated adoption rises (Brynjolfsson and Milgrom 2013).

3. How Data Virtualization helps companies to become more data driven

Companies and government organizations today intend to use data to improve business and decision-making process, but very frequently data are still difficult to find, extract, integrate and derive insights from.

This is due to the complexity of IT architecture and systems, so often data are hidden inside different data siloes, different applications, databases, on premises or cloud systems, excel files and so on. In the case of multinational companies, data are stored even in different geographies and countries worldwide.

According to IDC research (Rydning, 2018), around 30% of data will be real-time in 2025. This is close to a 600% growth from where we are today and it is based on the same volume of data generated today. This represents a huge change in business scenarios, as it means higher customer expectations and much more complex IT infrastructures.

In addition, today we are facing a critical situation due to the COVID-19 pandemic. The need for speed has never been greater and companies have to change more rapidly than ever before to ensure fast decision making, new production lines, and new customer services; in fewer words: new ways of working.

Traditional data architecture, that was designed 20 years ago, is not able to meet real-time business requirements. The solutions deployed such as data warehouses, self-service business intelligence tools and data lakes are based on a physical approach to manage data, that requires physical movement or data replication. Nowadays, this approach is showing its limits, and a logical approach, that enables faster data delivery, regardless of the physical location of the data sources, is preferred.

Data virtualization is a modern data integration approach, which creates real-time, integrated views of data from disparate, structured, semi-structured, and unstructured data sources, whether on-premises or in the cloud, to support data services for business transactions, analytics, and other workload patterns, without replication. It is an agile data integration approach that presents data as a virtual, abstract layer, independent of underlying database systems, structures, and storage.

A typical application of this new concept is the Logical Data Warehouse (LDW) approach, that is the evolution of the traditional Data Warehouse

architecture developed on a physical data integration methodology: extracting and copying the data in new systems (data warehouse), to make analyses and reporting (Fig. 1).

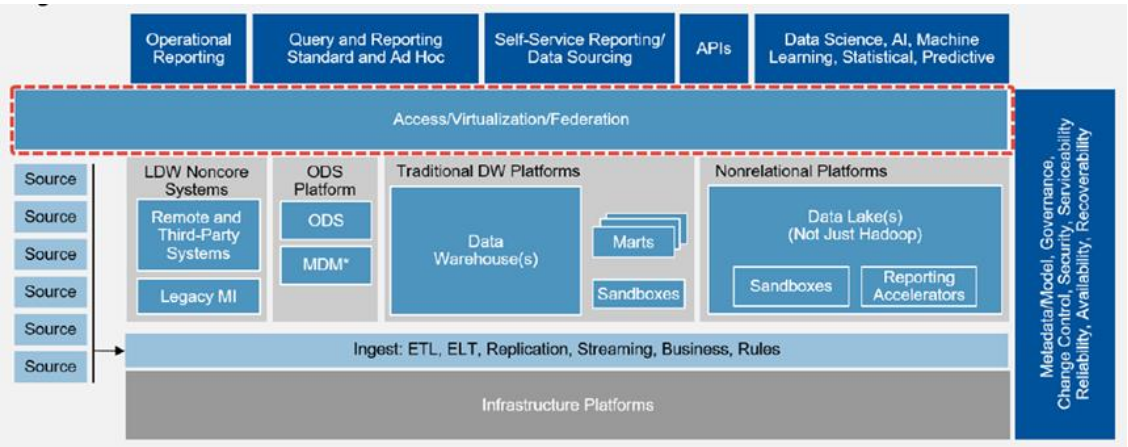

Fig. 1 Logical Data Warehouse Architecture (Gartner, 2018)

Data virtualization complements the capabilities of traditional data integration with the following features:

- Real-time data access without replication and its associated costs.
- Rapid data services delivered using myriad protocols with fine grained security, higher service levels, and detailed monitoring.
- Discovery and governance of data and metadata, data lineage, impact change, and other data management capabilities, enabled by the virtual data layer.

Using data virtualization, it is possible to see data from different systems and sources, as they would be stored in one single, virtual, database.

The typical data integration process based on the physical approach of Extracting, Transforming and Loading (copy) the data somewhere else, can now be modified or replaced by the Logical approach of Connecting, Combining and Consuming the data, without physically moving the data (Fig. 2).

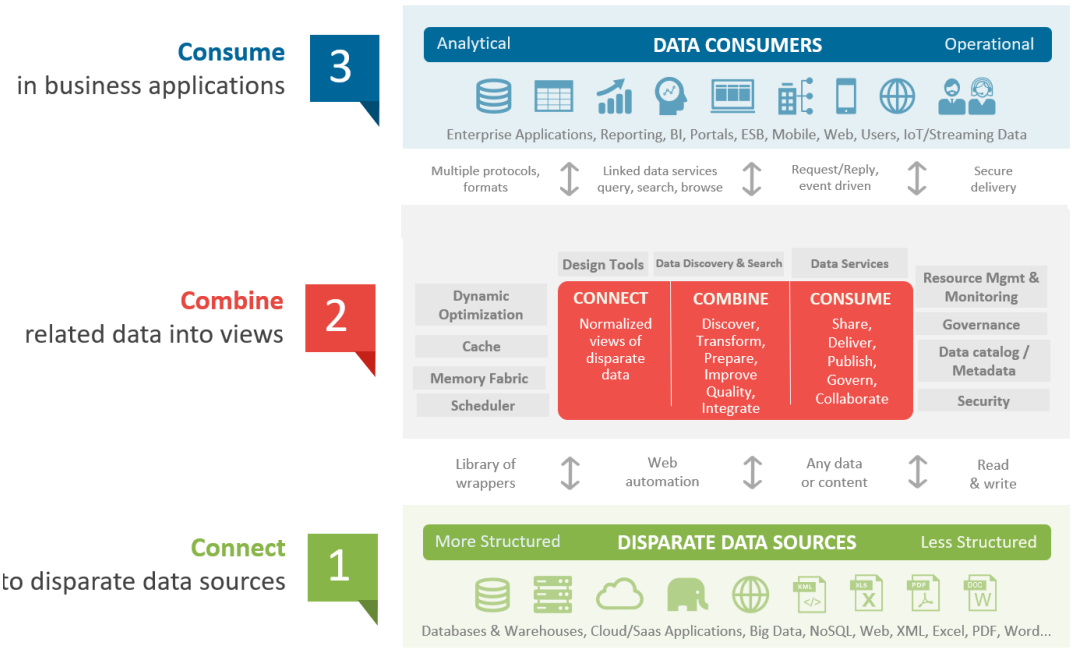

Fig. 2 Data Virtualization solution architecture (source: Denodo Case Study)

In terms of business application, there are many scenarios of possible implementation using data virtualization as one of the technical enablers. The application scenarios can be mainly analytical, for information analysis on historical data, or mainly operational for real time information delivery (Fig. 3).

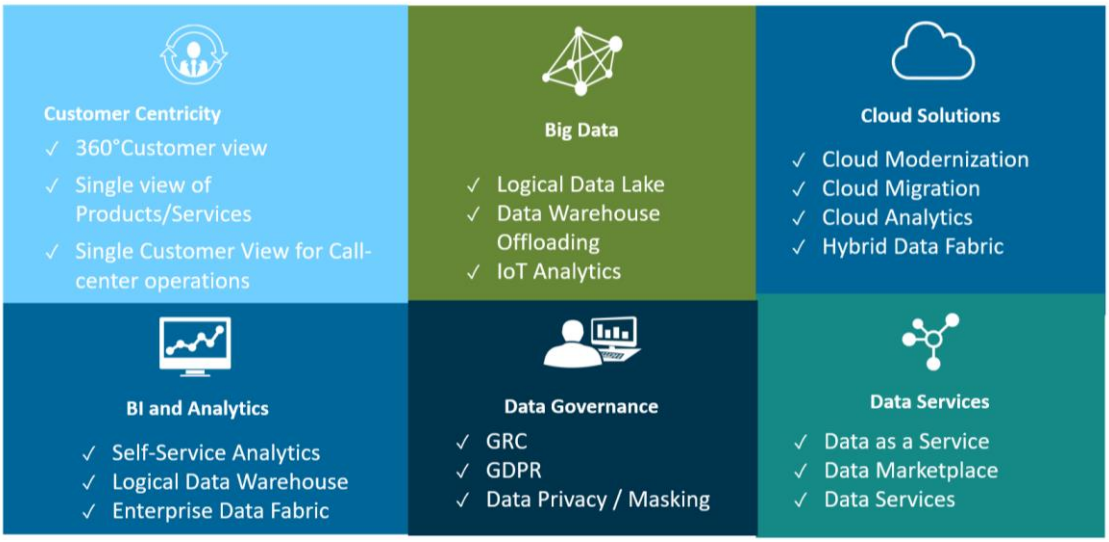

Fig. 3 Data Virtualization main use cases

One of the most typical business use cases is the definition and creation of customer centricity and single customer/product views. Usually, companies store data related to customers in many different systems that are not connected. Differently, this approach focused on an "holistic view" is needed mainly to provide a comprehensive and combined view of customers (or any other entities as products/services), through the aggregation of all of the available data, in order to better analyze and understand their main characteristics, needs and behavioral pattern. Telecom operators are using, for example, this approach to view a unified version of customers' data that are stored in different systems such as, CRM, billing, incident management, mobile apps, social media, network management systems, and so on, and provide a better and more personalized service.

Self-Service Analytics and Agile Business Intelligence are very common requirements for organizations that need agile and flexible data delivery to business users for analytical purpose. As a result, the Logical Data Warehouse and Logical Data Mart" project has been implemented on top of existing traditional Data Warehouse systems, as it is a much more flexible and quicker approach compared to the traditional ones.

With the adoption of Data lakes solution, companies started using AI and ML techniques for advanced analytics, predictive and proactive maintenances, risk analysis, production optimization and IoT data integration. In this scenario, Data virtualization delivers faster data collection, preparation and analysis.

Some uses are related to migration projects from old to new systems, for instance to cloud services or applications. In this scenario, it is very frequent to adopt a logical integration approach to save time and ensure business continuity for end users and customers.

Real time data integration through data services is another interesting use as it enables "Real Time" data integration. Several companies are adopting this integration model as it delivers a fast and agile approach.

Finally, Data security is one of the priorities for companies worldwide and the use of an advanced Data integration solution can be really effective to ensure the respect of compliance regulation and safe use of company data. The logical approach dramatically reduces reproduction and export of the data that entails security issues to be managed. This may be extremely useful for GDPR compliance purposes and in M&A operations, where integration of IT systems may be very challenging.

In the following paragraphs the scenarios of possible implementation and the benefits of data virtualization through success business cases from different verticals will be illustrated.

3.1 Data Virtualization Value Chain

Data and analytics technologies are at the heart of many digital transformation projects and are constantly gaining interest and adoption from worldwide market. Machine Learning and Artificial Intelligence (ML/AI) require a high volume of data to process in order to deliver value. More data to process means more complexity to manage from IT and an organizational point of view.

Time is one of the main request of the ML/AI projects. The more the data to be analyzed, the more effective a machine learning algorithm will be. The ability to manage huge volumes of data at a different speed is a source of competitive advantage. Anti-fraud systems, predictive maintenance services based on IoT sensors, and virtual assistant services are based on ML/AI systems that need real time data, and faster data delivery and access.

In the post Covid-19 era, decision making speed and flexibility are becoming a crucial component of business strategy. Redesigning data architecture for more efficient analytics and greater flexibility to react, can be a key component to reach a true differentiation. For example, time may be a source of value creation when retailers need to analyze competitors pricing and integrate online and in-store operations. Time is crucial for pharma and biotech companies to develop new products. Time to market is crucial for financial services to manage financial markets turbulence. Telecom operators need faster data processing to be more effective at call center operations and to provide better customer services. Healthcare services need real time data to manage the ability to treat patients effectively. Every business in the market's present condition would benefit from more time and higher reaction speed.

In the last few years data technology and services have been adopted to achieve more agility, flexibility and speed. Cloud services and new data management solutions allow companies to start and deliver projects in weeks rather than months. Data virtualization is useful to manage the complexity of integration of data and gain efficiency.

Faster and timely data delivery creates several opportunities to generate value within the layers of an organization, and this process can be described as a value chain (Fig. 4). The Data virtualization value chain is based on the ability to provide data faster, in real time or near real time. Speed provides more time and therefore more opportunities and efficiencies

at organizational, business and financial levels. Nowadays, real time data processing is a pre-requisite to compete.

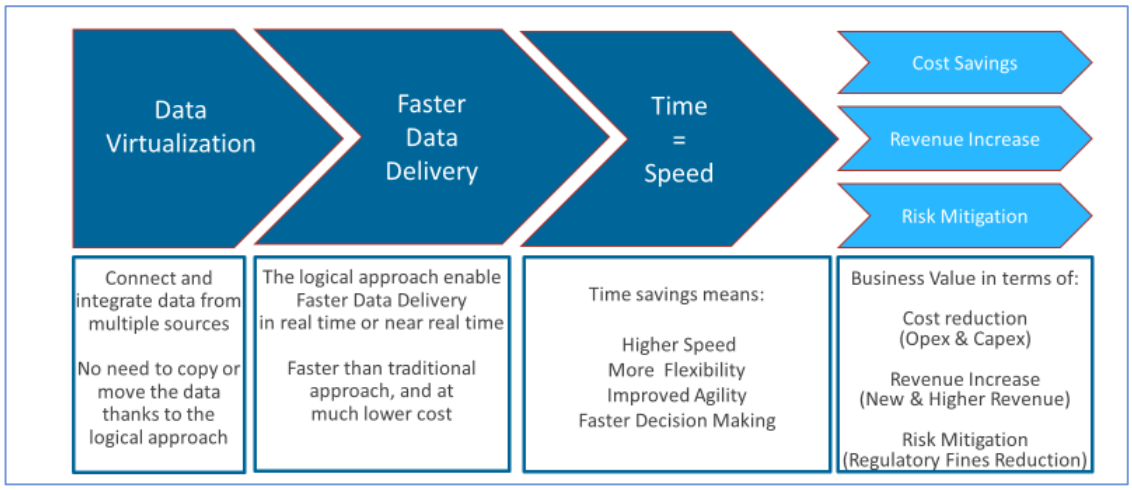

Fig.4 Data Virtualization Value Chain

4. THE BUSINESS BENEFITS OF DATA VIRTUALIZATION

Our analysis focuses on the main benefits of a correct adoption of Data virtualization in terms of efficiencies and tangible economical savings. We have studied 170 projects in 21 different verticals developed by Denodo, and found that Data virtualization creates value by exploiting: cost savings, revenue increase and risk mitigation.

We found that data virtualization has been implemented in many different industries: Agriculture; Automotive; Banking; Business Services; CPG (Consumer Packaged Goods); Defense; Education; Energy; Fashion; Financial Services; Healthcare; High Tech – Software and Hardware; Insurance; Manufacturing; Mining; Oil & Gas; Pharmaceutical & Biotech; Real Estate; Retail; Telecommunication; Transportation.

An additional element considered is the correlation between the size of the organization and the number of data virtualization projects implemented. The segment of large organizations (more than 10,000 employees) is clearly the segment with major numbers of observations (90 companies) and amount of investments depending on the higher needs of data integration and virtualization. It is also important to underline the increasing number of projects implemented by medium-size companies (from 2,000 to 10,000 employees) that is 45% of the samples (77 companies).

The analysis has also considered the geographical location of the projects, to understand if the benefits delivered can be considered a global trend or if it may be just a Silicon Valley (or US) phenomenon. The outcome demonstrates that it is a global trend, with projects delivered in all continents. In terms of location, the projects in the sample have been implemented in: Canada, Colombia, Mexico, US, Austria, Belgium, France, Germany, Island, Italy, Norway, Portugal, Spain, Sweden, Switzerland, The Netherland, United Kingdom, Indonesia, Japan, Singapore, United Arab Emirates, Australia and South Africa.

Data virtualization investments, as generally IT investments, are finalized to improve the company's position within its industry. The investment benefits are created by expanding the gap between buyers' willingness to pay and costs incurred, even reaching, if possible, the dual effect of raising revenues without incurring in extra costs. In our study, we have observed that once the data virtualization platform is implemented in a company, it often creates cost savings and ensures new revenue streams or improvement in data governance and risk mitigation.

4.1 Cost savings benefits

IT departments have used different technologies in the last years to collect, manage and integrate data in order to be more efficient and productive. These activities are still using the "physical" approach to data integration, based on the main concept of data replication or copies. The typical technology used to build the data warehouse is ETL (Extract, Transform and Load), that ensures fast data copy and replication. When the volume and frequency of copies increases, this process becomes more difficult to manage, as it is slower and costly. Adding new sources or new reports typically requires months and this is not acceptable anymore.

Data virtualization helps companies to reduce the time and cost associated to data integration projects, delivering data in real time or near real time, without the need to move or copy them. This logical approach is much more agile, flexible and efficient than a physical one. This approach reduces time, Man/day labor needed, infrastructure (hardware and software) and network costs. Furthermore, the increased speed of action and higher flexibility deliver huge advantages in terms of increased productivity of organization and therefore higher operation margins.

The combination of these elements may have a great impact on companies' profitability. In the following paragraphs, we examine several Denodo's data virtualization projects to identify the different advantages that can be reached with different magnitude on cost savings.

Data virtualization is a complex project that is usually finalized to reach several cost advantages. Typically, the project starts with establishing a logical data layer architecture. In one of the largest semiconductor companies in the world, a logical data access layer was created to centralize its data security, data privacy, and data governance rules. The company established also a semantic uniformity with the intention of increasing the efficiency of the cross functional teams. The logical data layer was integrated with HR and supply chain departments' cloud and SaaS applications, filling gaps in its services portfolio. Subsequently, the company rolled out data virtualization as its enterprise wide data access layer, accessed by more than 600 developers and thousands of business users on a regular basis and also standardized API registered services, lightweight data governance, and the enterprise data model (Fig. 5).

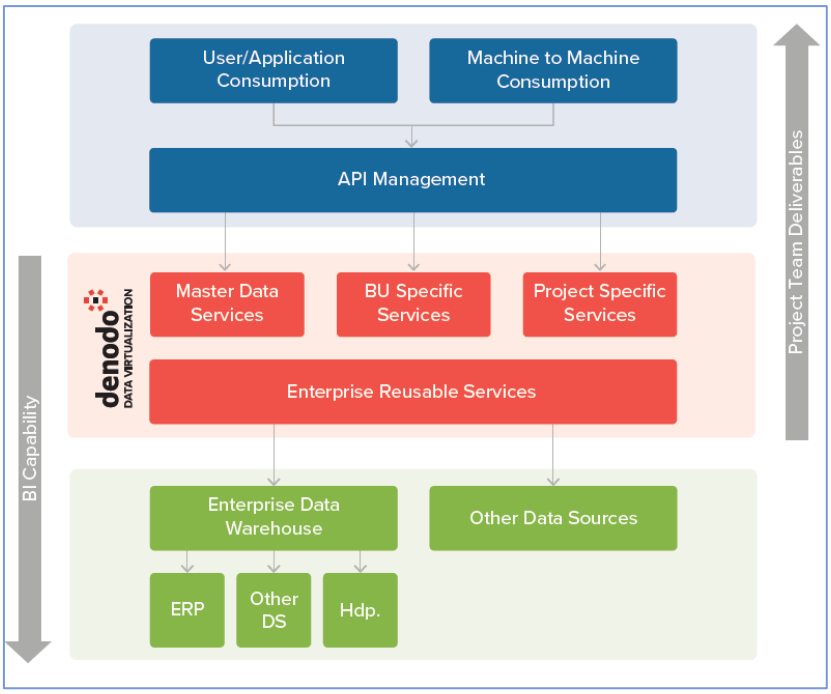

Fig. 5 Leading Semiconductor Project Architecture (Source: Denodo Case Study)

The main results obtained by the data virtualization project can be found in the speed of development and deployment web services. The direct effect of this acceleration is the reduction of the time-to-market for data delivery to business users and the consequent generation of cost savings by the significant cutting of the total cost of ownership (TCO) (Fig.6).

Value Driver	Metric	Goal	Actual
Time to Develop	Time to develop web service in days	50%	90%
Time to Deploy	Time to Deploy web service in days	50%	90%
TTM	Overall time it takes to make web service available for use	60%	90%
Time to Engage	Time it takes for business to engage with IT	75%	75%
Performance	Performance of web services	50%	60%
Impact Analysis	How fast can we perform impact analysis	50%	90%
Enterprise Architectural Alignment	Ease at which data from disparate sources can be integrated	Security, data classification	High

Fig.6 Leading Semiconductor Value Driver Metrics (Source: Denodo Case Study)

The company has also developed a heterogeneous ecosystem of tools and technologies, giving rise to a complex, distributed data ecosystem. The data virtualization project has contributed to deliver business-friendly, consistent information, as well as products and services with enterprise grade security and privacy.

In several other projects we have analyzed, the consolidation of the databases is the main objective to reach for private companies and Government agencies. In an US Government agency, the consolidation has reduced twenty different database sources to five databases (Logical Datawarehouse) in two years' timeframe. Data virtualization has produced cost savings (over $4 million per year) in terms of labor, hardware, software and storage costs and also facilitate the embrace of new cloud platforms (e.g. Amazon) and Data sources (e.g. Amazon Services).

The Italian National Institute of Statistics has set itself the goal of enriching the offer and the quality of the information produced, in order to improve the effectiveness and efficiency of the overall activity. and has combined an increasing awareness of the opportunities generated by the new data sources available, such as web data, data from the world of the Internet of Things or social data. The goal of the organization was to integrate all this new information within its systems and to understand if the latter could, on the one hand improve the quality of existing statistics, and on the other contribute to the development of new ones.

The introduction of a virtualization layer has helped the Institute to overcome the rigidities of dissemination and manipulation of data and, if the need to duplicate or copy the data was not completely eliminated, that approach improved data governance and data quality monitoring actions, concentrating it in only one available access point.

The project carried out by the Institute made it possible to make the investment costs more efficient, to increase the quality of the statistics produced and to reach new opportunities for data enhancement, making the institute ready to seize the information potential of Big Data.

The data virtualization project can be finalized to provide simple, fast and cost-effective access to data, to reduce risk, increase efficiency and increase business scalability through self-service analytics. The reduction of the fragmentation of the IT architecture creates a holistic view for executives and business users and the possibility of re-definition of the business model. For medium-big organization logical data mart creation can be extremely efficient compared to physical data mart, because a logical data mart can cut down the timeline for the development of typical reports, as

achieved by a leading US Investment Bank. Assuming the same organizational structure in terms of employees and the same output required, Table 1 shows the annual labor costs in the case of the adoption of a physical data mart (before) and the adoption of the logical data mart (after).

Physical Data Mart Implementation					
Data Mgmt Role	**Employees**	**Hours/Year**	**Time on Data**	**Rate**	**Total Cost**
Business Intelligence	4	1920	70%	$ 100	$ 537.600
Business Analytics	10	1920	60%	$ 100	$ 1.152.000
Self-Service	20	1920	20%	$ 100	$ 768.000
					$ 2.457.600
Logical Data Mart Implementation using Data Virtualization					
Data Mgmt Role	**Employees**	**Hours/Year**	**Time on Data**	**Rate**	**Total Cost**
Business Intelligence	4	1920	30%	$ 100	$ 230.400
Business Analytics	10	1920	50%	$ 100	$ 960.000
Self-Service	20	1920	10%	$ 100	$ 384.000
					$ 1.574.400
Total Labour Cost Savings					**$ 883.200**

Table 1 Project Annual Labor Savings in US Investment Bank

The physical data mart is a new system that stores some specific data sets (for instance customers' data per region) for analysis purposes, developed using the traditional approach of copying and moving the data physically from source A to destination B. This approach requires time, people, process design and management, hardware and software technologies investment and maintenance costs. The logical data mart developed using data virtualization solutions, can access the same data sets, without the need of copying and moving the data, with great benefits in terms of time saving, human resources needed and overall hardware and software infrastructure costs.

4.2 Revenue Increase

A typical use of data is the analysis of a company's main business entities, such as customer data, sales figures, marketing campaigns results, distribution channel productivity, manufacturing KPI's and many others. The speed of analysis and decision is therefore critical to achieve a business advantage, in many different areas, but revenue growth has always been a company priority.

Revenue growth may generate up to 70% of total shareholder returns in five year (Candelon et al., 2020) and, during the COVID-19 crisis, artificial intelligence projects are mainly focused on generating revenue growth. The possibility of having more data available in real time or near real time, is enabling new possibilities powered by ML/AI for sales and marketing actions. These are the main areas where data virtualization can deliver significant value: upselling/cross selling opportunities on existing customer base; personalized/customized offers and solutions; new product/services design and development; real time competitive pricing adjustment; improvement of customer retention (churn rate reduction); customer satisfaction.

In our analysis, we found that banks and insurances have defined specific upselling or cross selling opportunities of new services within their customer base, as for example home insurance packages for car insurance customers. Telecommunication operators have implemented more efficient call center operations to develop specific offers to propose during customer interactions.

Real-time access improves customer services, generates higher customer satisfaction, retention and churn reduction. In some projects, telecom operators defined specific customized products for their customer to offer value added services such as digital content. Pharma & biotech companies speeded up the design and development time to get to new products faster to the market. Manufacturers developed new predictive maintenance services to increase customer satisfaction and create new revenue flow. Retailers collected competitive pricing info and then modified the price list of their available special offers in specific stores or restaurants menus in case of restaurant chains. Online and offline pricing adjustment may be critical to manage demand peak and fall.

To better understand the benefits in terms of revenue enhancement, we analyze the case of a leading telecommunications and data transmission carrier with its own infrastructure in Spain. One of the main goals of the

operator was reaching the highest level possible of customer service and therefore a clear differentiator against competitors.

The operator decided to use an internal call center team and implemented a software solution that required the agents to use many different systems (CRM, Billing or Monitoring and Diagnostic applications) to get useful information about the customer who was calling and provide the required information. Due to the complexity of this process the Operator decided to implement a new software solution to integrate data across the company applications and provided the agents with a "Single Customer View", with the right data, in the right format at the right time. This new approach achieved several relevant benefits, such as a clear improvement in the quality of customer service and an increased call center productivity, even with the possibility to integrate external data sources from the partner ecosystems (Fig. 7).

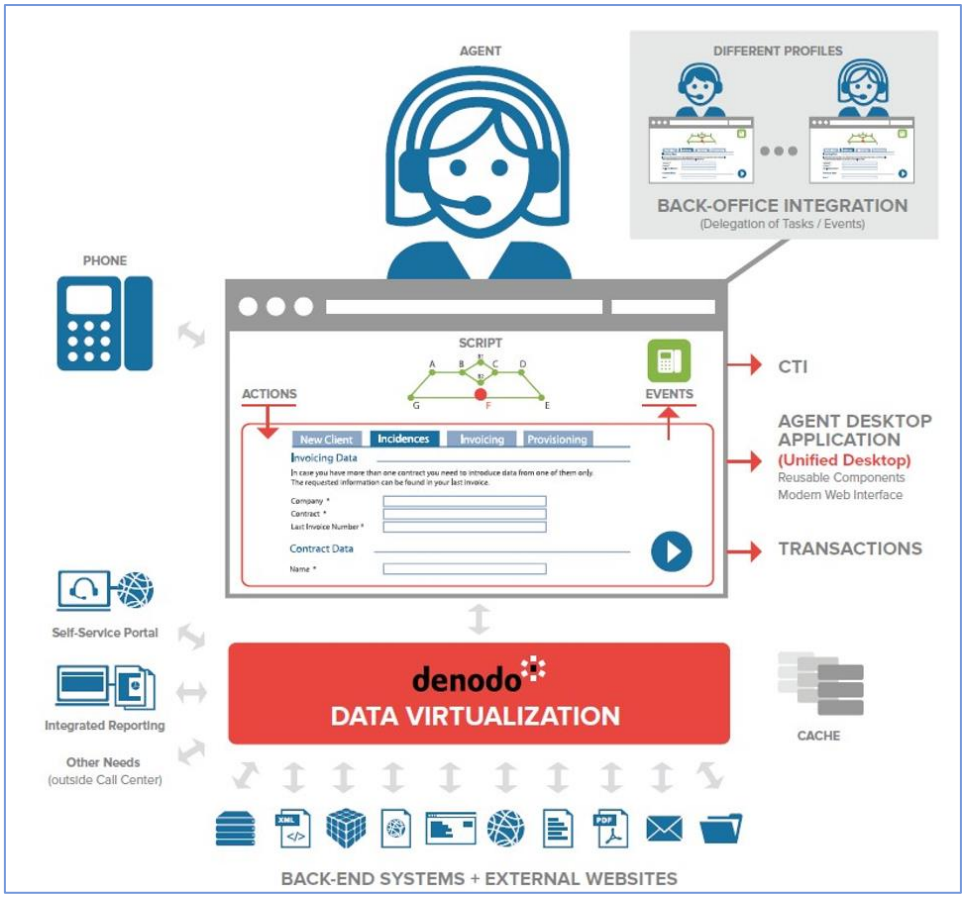

Fig. 7: Telecom Operator Solution Architecture (Source: Denodo Case Study)

The main results obtained by the projects based on Data Virtualization can be summarized in the increased quality of customer service and the new opportunities for revenue growth. In terms of revenue improvement, the most important result achieved is the higher customer satisfaction (satisfaction index of over 94%) and the improved customer retention level (doubled), that prevented customer churn (increased first call resolution rate to over 90%). An additional benefit of the project can be measured by the improved productivity of the agents, that saved time in getting useful customer data from multiple applications (average handle time reduced by 10%- 1,2 Mil € per year). Furthermore, an over 50% improvement was achieved in the back-office workload.

To understand how data virtualization can help to redefine the business model and to create new opportunities for growth, we have studied the project implemented by the world's largest construction equipment manufacturer. It provides machines and technology for construction, transportation, mining, energy, logistics, and electric power generation. The company distributes its products and services through a dealer network in 190 countries, with a total revenue of approximately 39 B $ per year.

The company decided to design a new proactive maintenance service to reduce or eliminate machines breakdown, that would better differentiate its "high quality" offering against an increasing number of low-cost competitors. In order to achieve this goal, the company invested in technologies and tools for telematics, IoT sensors sending data to a data lake and advanced analytics (Fig. 8).

The complexity of this scenario required a new data integration architecture to combine operational and analytics data, real-time sensors data and historical data, managing all the needed collections and combinations of different data sets, formats and sources. In other words, a new Data Service Layer to provide the needed information and reporting at the right time.

In this way, the company created a proactive maintenance service for its customer base, achieving higher customer satisfaction, a new "highly differentiated service" against low cost competitors, and an increased revenue flow. The revenue growth was due to the increased sales of parts and services, the new telematics services offered in DaaS (Data as a service) model, and the increased long-term revenue flow.

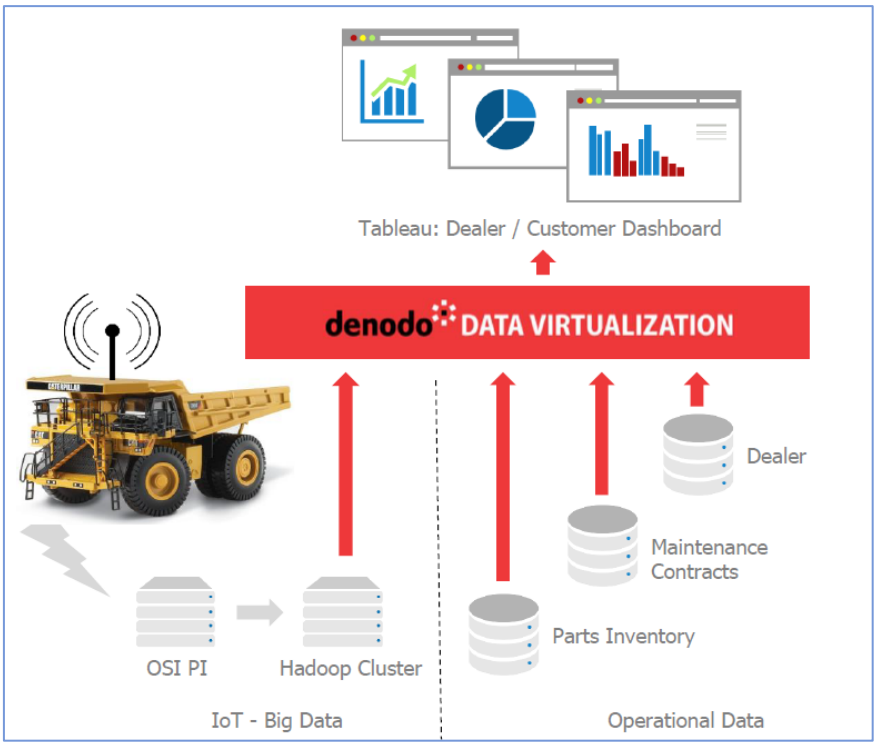

Fig. 8 IoT and Historical Data Integration Solution Architecture (Source: Denodo Case Study)

Flexibility and speed are key elements also when particular market conditions require new solutions to be implemented in a very limited time as it happened during COVID-19 emergency. A leading US bank, part of a global financial services group, decided to implement a new data platform to modernize its data infrastructure to sustain growing business.

During the pandemic emergency, the Small Business Administration (SBA) and US Treasury Department created a Paycheck Protection Loan designed to provide a direct incentive for small businesses to keep their workers on the payroll. The program was delivered via US banks, and the borrower had to send them the applications. The main problem was the enormous amount of applications received from the bank in a very short timeframe. The bank was expecting around 2,000 applications in a few weeks, but it received around 9,000 applications in 2 days, and around 21,000 in 3 days. The bank loan application system was overwhelmed and not responding anymore due to volume of transactions requested. IT department of the bank, used the Data virtualization platform already in use for a Logical Data Warehouse project, to rapidly deliver a solution in less than two days. Data Virtualization platform was used for its flexibility to incrementally add sources to meet the changing needs of the business as they adjusted the processing of loans. Speed and flexibility were key components of the

solution as the business requirements changed in real time as they dealt with the exploding volume of applications. And IT was able to respond in real time as the business requested to add new data sources to help categorize the applications. The project would have never been implemented on time using a traditional physical data integration approach. The main benefit delivered was the creation of a new revenue channel thanks to speed and flexibility of IT infrastructure. The bank implemented a new service (new application process) in just two days. This created a new revenue flow and a customer acquisition opportunity for the company in an extremely reduced timeframe,

4.3 Risk Mitigation

Reducing risk has always been one of the strategic priorities for executive management in any verticals. In the last years, companies are facing additional types of risk associated for instance to brand identity, social responsibility and regulatory compliance.

Data virtualization is one of the technologies that can help companies to reduce risk in some areas of IT organization, but also from a business point of view to mitigate compliance and data governance risks (e.g. GDPR, IFRS 17). Banks needs to adhere to several regulatory requirements, and sometimes time is a critical factor to ensure compliance and reporting. A luxury and fashion company needed to protect its brand reputation and therefore being compliant is a business imperative. The priority of GDPR requirements for customer data must be managed very carefully, therefore implemented a data virtualization solution to mitigate the risk of GDPR fines.

Another interesting area of application is in the security and corporate social responsibility. An Asian Telecommunication operator implemented data virtualization as part of its emergency plan. It needed to produce reports in case of emergency and the communication network are the most critical ones during emergency and natural disasters. Being able to analyze fast data may help to save lives and enable better rescue operations.

In our study, we found important benefits in terms of risk mitigation when we analyzed the data management architecture of insurance companies, which is extremely complex by nature, with multiple operational source systems and multiple stakeholders from different business groups that use their tools to access data. In one leading USA Insurance, there was no clear

policy or control on data access, and no possibility to track real access responsibility. The entire IT architecture was therefore vulnerable to security breaches. Compliance to insurance market specific regulations and GDPR was not satisfied and so the company was exposed to potential fines of millions of dollars.

The data virtualization platform has improved the IT architecture, acting as the single point of entry to all the different systems, and so reducing the associated risk of multiple access to multiple systems. The new data architecture solution improved consistently security, governance and assured regulatory and GDPR compliance. Another advantage was related to business flexibility as it improved reporting time for business users.

Another interesting area of application is during the mergers and acquisitions' integration. Mergers & acquisitions' value creation is strictly dependent on the integration between merging entities and organization. Data virtualization helps the acquirer to speed the integration process and reduce the complexity of IT infrastructure integration minimizing this risk. For example, the data virtualization project in a semiconductor company was finalized to facilitate the post-M&A data integration. The company was involved in several acquisitions every year. It was crucial to quickly integrate the acquired companies in the IT systems. In particular, the company' needs were to distribute new self-service entitlements to downstream acquisition applications, and to distribute directory identities to the acquisition directory. The company also required to architect its enterprise data access layer for a single point of entry for HR and supplier data consumption, seamless support for data source migration, and scalable interaction among on-premises and cloud data sources. The major issue is the salary payment process that in the US is a 2 week cycle. Once the company is officially acquired, IT has only 2 weeks to integrate the newly acquired company data in the HR systems to allow payments. Data virtualization solution has helped the company to find an agile data access solution when challenges became overwhelming.

5. CONCLUSION

The data driven decision making process is now extremely important for any organization to compete and be successful in the market. In particular, nowadays fast decision making is even more critical due to atypical market conditions.

The technology landscape changed dramatically in last five years and will evolve even more in the next 10 years, due to the exponential growth of data volume, formats and sources. Real time data will be six times faster than today and will be 30% of overall data volume generated.

Companies and organization worldwide are adopting new technologies and solutions to become more data driven, but it is also clear that the transformation is not only technology related. Organizations (people) and processes have to be adapted and managed in new ways to ensure that data will become a real company asset to be analyzed, used and leveraged to compete more efficiently. Organization should be empowered by specific know-how and culture, that is why a real data driven leadership is also a key element in the transformation process. This is becoming even more important now that the COVID-19 emergency is requesting companies to adopt new organization, new process and new business model. The need of agility, flexibility and speed has never been greater than now. And company data is a new competitive asset to be managed faster than ever before.

Big Data with ML/AI solutions and analytics technologies are the enabling technologies to innovate the data platform architecture. Data virtualization is one of the enabling technologies of the data architecture evolution and innovation as it is focused on providing a logical approach to data integration, that is extremely efficient and effective.

Data virtualization value chain is the value creation process enabled by implementation of a Data virtualization platform in an organization. It creates value because it delivers time savings and speed in data use and analysis. Time savings bring more agility to organization, enable faster decision-making and greater flexibility to adapt strategies and business models. Speed means faster actions to compete more effectively, while reducing costs to become more efficient. From a business point of view, this is translated in cost savings, revenue improvement and risk mitigation.

We have analyzed 170 projects implemented by Denodo in many verticals and geographies worldwide. The most interesting results of this analysis are:

Data virtualization creates business value in three main areas: cost savings, revenue increase and risk mitigation. In terms of cost savings, the benefits come mainly from faster data delivery and enhanced productivity, reduced project development cycle and lower labor cost associated. Other cost savings elements are the IT infrastructure overall cost in terms of both investments and total cost of ownership for software, hardware and network traffic.

Revenue increase is related to more effective sales operations, the creation cross-selling and up-selling opportunities, new product and services development, personalized and profiled offers and promotion, new business model enablement, better customer services and improved customer satisfaction, and churn reduction.

Risk Mitigation is achieved with higher data governance and security that reduces risk for compliance and regulatory fines, as well as brand protection against regulatory actions. An additional advantage may come for M&A process improvement due to faster and easier IT systems integration.

Data virtualization enables up to 10 times faster data delivery. Data is delivered where it is needed, when it is needed, and in the right format for the right users. Data virtualization enables a much more agile and flexible organization because it can be implemented for different purposes or cases, and this maximizes ROI.

One of the key points emerged from the analysis is the fact that this technology is horizontal, and it has been implemented in many different industries such as automotive, banking, defense, education, energy, Government, insurance, manufacturing, pharma, retail and telecommunication.

Becoming a real Data-Driven organization is not an easy task for any organization. The implementation of the data virtualization software solution, is one of the possible approaches to achieve the goal of being a fast data decision maker. Technology adoption cannot be considered the unique option, as it is clear that other components must be taken into consideration.

Organizations should be managed by leaders that truly believe in this digital process innovation built around company data. This is a

fundamental element in building the right data-oriented culture, that should lead all company activities and projects.

The other critical element is the know-how and talent that must be created, hired and managed in order to leverage and exploit the technology investments. The human factor is still critical and is at the heart of any change and innovation in an organization.

REFERENCES

Acito F, Khatri V (2014) Business analytics: Why now and what next? Business Horizons. 57(5):565–570.

Agrawal A, Gans J, Goldfarb A (2018) Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press, Brighton, MA.

Athey S, Stern S (1998) An empirical framework for testing theories about complementarity in organizational design. Working paper, National Bureau of Economic Research, Cambridge, MA. No. 6600.

Bharadwaj AS (2000) A resource-based perspective on information technology capability and firm performance: An empirical investigation. MIS Quarterly. 24(1):169–196.

Bloom N, Garicano L, Sadun R, Van Reenen J (2014) The distinct effects of information technologies and communication technologies on firm organization. Management Science 60(12):2859–2885.

Brews PJ, Tucci CL (2004) Exploring the structural effects of internetworking. Strategic Management Journal 25(5):429–451.

Brynjolfsson E, Malone TW, Gurbaxani V, Kambil A (1994) Does information technology lead to smaller firms? Management Science 40(12):1628-1644

Brynjolfsson E, McElheran K (2016) The rapid adoption of data-driven decision making. American Economic Review. 106(5):133–139.

Brynjolfsson E, Milgrom P (2013) Complementarity in organizations. Gibbons R, Roberts J, eds. The Handbook of Organizational Economics (Princeton University Press, Princeton), 11-55.

Brynjolfsson, E, McElheran, K (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in US Manufacturing. Available at SSRN 3422397.

Candelon F, Reichert T, Duranton S, Charme di Carlo R, Stokol G (2020) Deploying AI to maximize revenue - 2020) Boston Consulting Group -BCG Henderson Institute, https://www.bcg.com/it-it/publications/2020/deploying-ai-artificial-intelligence-to-maximize-revenue

Casadesus-Masanell R, Zhu F (2013) Business model innovation and competitive imitation: The case of sponsor-based business models. Strategic management journal 34(4):464-482.

Chen H, Chiang RH, Storey VC (2012) Business intelligence and analytics: From big data to big impact. MIS Quarterly. 36(4): 1165-1188.

David JM, Hopenhayn HA, Venkateswaran V (2016) Information, misallocation, and aggregate productivity. Quarterly Journal of Economics 131(2):943-1005.

Dilda V, Mori L, Noterdaeme O, Schmitz C (2017) Manufacturing: Analytics unleashes productivity and profitability. Report, McKinsey & Company.

Forman C, McElheran K (2019) Firm organization in the digital age: IT use and vertical transactions in US Manufacturing. Working paper, SSRN Working Paper Series. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3396116.

Gambardella A, McGahan AM (2010) Business-model innovation: General purpose technologies and their implications for industry structure. Long range planning. 43(2-3): 262-271.

Gartner, (2018) Adopt the Logical Data Warehouse Architecture to Meet Your Modern Analytical Needs. www.gartner.com/en/documents/3871182/adopt-the-logical-data-warehouse-architecture-to-meet-yo0

Gibbons R, Henderson R (2012) Relational contracts and organizational capabilities. Organization Science 23(5):1350-1364.

Harvard Business Review Analytic Services. (2019). An inflection point for the data-driven enterprise. https://hbr.org/resources/pdfs/comm/snowflake/AnInflectionPoint.pdf

Hoffman M, Kahn LB, Li D (2017) Discretion in hiring. Quarterly Journal of Economics. 133(2): 765-800.

Karshenas M, Stoneman PL (1993) Rank, stock, order, and epidemic effects in the diffusion of new process technologies: An empirical model. RAND Journal of Economics 24(4): 503-528.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2017) Human decisions and machine predictions. Quarterly Journal of Economics. 133(1): 237-293.

Lambrecht A, Tucker CE (2015) Can big data protect a firm from competition? Working paper, SSRN. https://ssrn.com/abstract=2705530.

McAfee A, Brynjolfsson E (2012) Big data: the management revolution. Harvard Business Review. 90(10):60-68.

Milkman KL, Chugh D, Bazerman MH (2009) How can decision making be improved? Perspectives on Psychological Science 4(4): 379-383.

Pierce L, Snow DC, McAfee A (2015) Cleaning house: The impact of information technology monitoring on employee theft and productivity. Management Science. 61(10): 2299-2319.

PWC (2019) 22nd Annual Global CEO Survey: CEOs' curbed confidence spells caution, https://www.pwc.com/gx/en/ceo-survey/2019/report/pwc-22nd-annual-global-ceo-survey.pdf

PWC (2019) Data trust pulse survey. Monetizing and trusting data: 2019 is the year to seize the prize, https://www.pwc.com/us/en/services/consulting/cybersecurity-privacy-forensics/library/pulse-survey.html

Rawley E, Simcoe TS (2013) Information technology, productivity, and asset ownership: Evidence from taxicab fleets. Organization Science 24(3): 831-845.

Ray G, Wu D, Konana P (2009) Competitive Environment and the Relationship Between IT and Vertical Integration. Information Systems Research. 20(4): 585-603.

Rydning DRJGJ (2018) The digitization of the world from edge to core. Framingham: International Data Corporation.

Shapiro, C, Varian, HR (1998). Information rules: A strategic guide to the network economy. Harvard Business Press.

Stoneman P (2002) The Economics of Technological Diffusion (Blackwell Publishers, Oxford, UK).

Tambe P (2014) Big data investment, skills, and firm value. Management Science. 60(6):1452-1469.