A 'FUTURISTIC' APPROACH TO DATA MANAGEMENT: GO MODERN OR GO HOME

HOW DO MODERN DATA PLATFORMS UNLOCK THE POWER OF AN ORGANIZATION'S DATA?

This article is part of Capgemini's Applications Unleashed 2022 Report which can be downloaded here.

HIGHLIGHTS

....

- Data grew faster and more complex than ever expected.

....

- Organizations urgently needed a reliable centralized solution to ingest, prepare, store, transform, analyse, and cleanse their big amounts of data.

....

- A modern data platform delivers organizations' overall data needs.

....

- The power of modern data platforms comes from the fact that it is based on cloud computing.

....

Witnessing the powerful effect of data on modern markets, more and more organizations are attempting to adopt a data-driven culture. This attempt raised multiple challenges as data kept growing from both size and complexity perspective, requiring a new way of thinking regarding data platforms.

With both data volume and complexity growing faster than ever expected, organizations were faced with multiple challenges related to data processing, storing, and analytics along with data quality and governance issues. Making data management and analytics harder than ever.

Data provide behavioural insights about consumers which marketers translate into market advantage . It also provides insights used by organizations to enhance their services and products according to their customer's wishes and needs. Making big data and AI a tempting market for organizations. In fact, 97.2% of businesses are investing in data and AI . This high percentage is also justified by the fact that organizations were forced into investing in this area as their data grew faster and more complex than ever expected.

FIRST CHALLENGE: DATA PROTECTION REGULATORS

The challenge started when data protection regulators came into the picture putting pressure on organizations to think twice about their data. For instance, with the EU General Data Protection Regulation (GDPR) taking effect in May 2018, European organizations had to become more aware of the use and storage of their personal data and how it is governed to avoid large financial penalties. GDPR implemented regulations on every aspect of processing, storing, and accessing personal data.

It also dedicates a whole chapter to the security of data processing from encryption to ensure the ongoing confidentiality, integrity, availability, and resilience of processing systems and services.

When not respected, EU's data protection authorities can impose fines of up to €20 million, or 4% of worldwide turnover for the preceding financial year, whichever is higher

when data violations are made. Leading to fines of $150 Billion to Google for example. Google wasn't the only big name associated with GDPR violations. Other big companies such as Amazon, WhatsApp, H&M, and Facebook already paid comparable fines for such violations. Banks also suffered from the same. Danske Bank, the largest bank in Denmark, and a former Fortune Global 500 member faced a fine of 10 million Danish kroner for General Data Protection Regulation (GDPR) violations pertaining to data storage.

SECOND CHALLENGE: GROWTH OF DATA

Those regulations weren't the only challenge organizations were facing. In fact, with the COVID pandemic, more people were forced to work and learn from home. People used more home entertainment options. Causing a higher growth in data than previously expected. According to Statista, the total amount of data created, captured, copied, and consumed globally is predicted to grow to more than 180 Zettabytes over the next five years . While the amount of existing data already increased by 5000% between 2010 and 2020 from 1.2 trillion gigabytes to 59 trillion gigabytes. This data as it is, is meaningless and have no value. It is not understandable by humans. It needs to be processed and cleansed via the correct tools into information. It also needs to be stored in a reliable, easily accessible, and secure way. This information is then used by organizations to create value. The complexity around this data is not only related to its storage, security, and quality but also to the fact that it originates from multiple systems and processes and exists in different formats (structured, semi-structured, and unstructured). 80-90% of it is unstructured, which most organizations cannot manage. In fact, 95% of businesses cite managing unstructured data as one of their business Problems. This came along with the increased threats of data breaches and cyber security attacks.

Therefore, organizations urgently needed a reliable centralized solution to ingest, prepare, store, transform, analyze, and cleanse their big amounts of data. They needed this solution to be secure and to allow them to govern their data in a compliant way. This is when modern data platforms
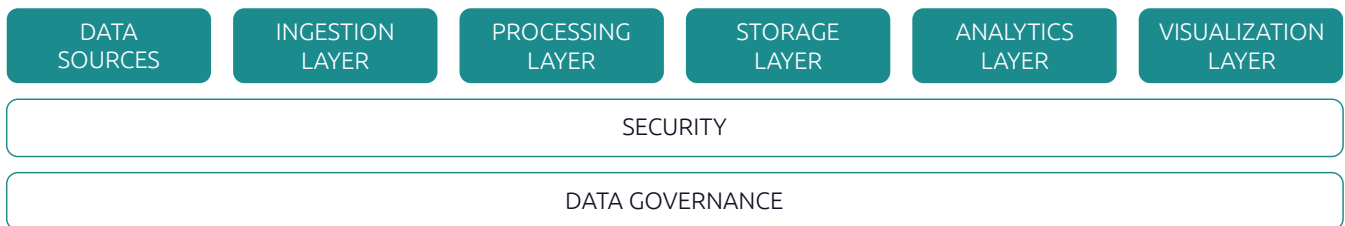
came into the picture and to the rescue. A modern data platform is a combination of interoperable, scalable, and replaceable technologies working together to deliver an enterprise's overall data needs . It enables the acquisition, storage, preparation, delivery, and governance of data and is a security layer for users and applications. Modern data platforms are usually designed in layers as described in figure 1: Modern data platform layout.

The power of modern data platforms comes from the fact that it is based on cloud computing

Figure 1: Modern data platform layers

DATA SOURCES	INGESTION LAYER	PROCESSING LAYER	STORAGE LAYER	ANALYTICS LAYER	VISUALIZATION LAYER

SECURITY

DATA GOVERNANCE

The power of modern data platforms comes from the fact that it is based on cloud computing and so through that, it provides redundant, inexpensive, and scalable storage and computing in a pay-as-you-go way to meet an organization's data requirements. It also allows the creation of scalable data pipelines that can handle large volumes of data. More connectors, tools, and features became available giving more control to organizations over their own data to access, manage, and easily add new data sources when required. ETL processes became simpler and data complexity was reduced, which accelerated and matured data analytics resulting in reliable insights that can be used for reliable decision-making and business capabilities such as compliance and annual reporting. Making these modern platforms popular and more in demand. Adopting them became a matter of business agility to keep up with customers' needs and competitors.

Multiple Cloud providers offer all the components needed to build such a platform. Azure for instance introduced Azure data lake storage gen 2(ADLS Gen 2) as the foundation for building data platforms on Azure. It is a fully managed cloud-based set of capabilities dedicated to big data analytics that enables data ingestion, preparation, storage, and monitoring. It integrates more than 90 built-in connectors with up to 5 GB/s throughput giving the capability to collect data in a scalable secure way from multiple sources to central storage eliminating data siloes. Data encryption is supported both for data at rest and at transit. Cost efficiency in this case is explained by the cloud pay-as-you-go model and by the possibility to scale storage and compute separately, which is not possible with on-premises data lakes.

DESIGN MODERN DATA PLATFORMS

To fully benefit from those capabilities, organizations should invest in designing modern data platforms considering the organization's whole strategy before building it. Organizations should look at their data as a whole rather than splitting it down into parts: how different systems are producing data and how that data will eventually be used, stored, and processed together. They should consider different scenarios and think ahead of relationships rather than creating data silos and data unnecessary dependency chains, causing new governance issues for themselves. Transformations along the way should be clear. Who did what to which dataset should be logged and traced. That is the only way to comply with regulations and to be able to trace data back to the source leading to the required level of data lineage. This 'as whole' perspective should also apply when choosing the modern data platform components. With the large pool of options for data tools evolving and getting wider, technical expertise became a challenge. In fact, 50% of US executives and 39% of European executives reported a lack of technical expertise in the 5 top challenges when it comes to data. This is when software vendors and consultancy partners came to the rescue and joined efforts together with their clients to define data strategies, and best practices, enhance the tools in a joint effort, and spread the knowledge among employees.

Adopting modern architectures and technologies along with the recent pandemic caused multiple data challenges. Those challenges can be solved with a proper modern data platform. While the combination of data platforms and cloud computing gave several possibilities and solutions, most of these data challenges persist. Organizations can only take full advantage of modern data platforms if they invest the correct time and effort in their design. It takes real data thinking and architecture to overcome modern data challenges.

ABOUT THE AUTHOR

Asma Cherni

Data community lead

Asma is a Design and implementation lead of modern data platforms using Azure Cloud. She Support clients modernizing data solutions and strategies.

https://www.linkedin.com/in/asma-cherni/

asma.cherni@capgemini.com

About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of 325,000 team members in nearly 50 countries. With its strong 55 year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2021 global revenues of €18 billion.

Get the Future You Want | www.capgemini.nl

For more information:

Capgemini Nederland B.V. P.O. Box 2575, 3500 GN Utrecht Tel. + 31 30 203 05 00 www.capgemini.nl