Deep Science Publishing
https://doi.org/10.70593/978-93-49910-38-6_8

DeepScience
Open Access Books

Chapter 8: Building scalable and secure data platforms for advisory services

8.1. Introduction

Real-world data is gone unexploited at various companies and organizations, when shared but under controlled access. Data exchange and sharing can result in valuable insights in affected domains, and may boost cooperation and stimulate novel services. Despite many upsides of sharing and exchanging data, significant challenges must be addressed. The importance of data sharing and exchange in DeFi is further fueled by the current architecture of DeFi protocols and related risks of collecting and maintaining data in one location. Full homogenization of data at one location may lead to security issues, as reflected by the various high-profile breaches in crypto. New solutions must safeguard parties involved when sharing and exchanging their data between either analysis or analytics. The lack of both standard representation of knowledge and interconnectedness is a bottleneck for scaling organization-wide and ecosystems-wide applications with the following three interrelated areas as key challenges. A lack of robustness of computational back-end systems, supporting interfaces and analytics of growing complexity and capacity, becomes more challenging with the aforementioned growing complexity of (and demand for) data-processing pipelines. The choice of a primary programming language is pivotal when designing a data platform in terms of performance, the tooling and libraries available, and scalability. Many software engineers have faced the challenge of either designing a new back-end system, or improving and scaling an already existing one. One such area where the design choices are crucial is that of educational and business solutions revolving around the one-off passing, storing and processing of huge amounts of data. In this paper, a common problem in both areas is defined: the lack of heterogeneity of the data sets sent, and the impossibility to practically expand on either system. Many companies and organizations collect, store and analyze on-chain and protocol activity data. Drafts of information representation are standardized and agreed upon for most widely used datasets. However, even while two companies may obtain their data from the same data sources,

they still may not be able to profitably query the same datasets.This session is inspired by and describes a fast, efficient programming model for the future of the finance industry to support scalable computation on the data and input formats of the next decade. This platform will allow clients to capture and store large text/image/audio inputs into a set of document stores and arrays. These document stores will allow clients to perform chunk-level data manipulation taking document-less data into future-compliant formats. Arrays will allow clients to run elastically scalable computation keeping data at rest within their data-compliant storage solutions.

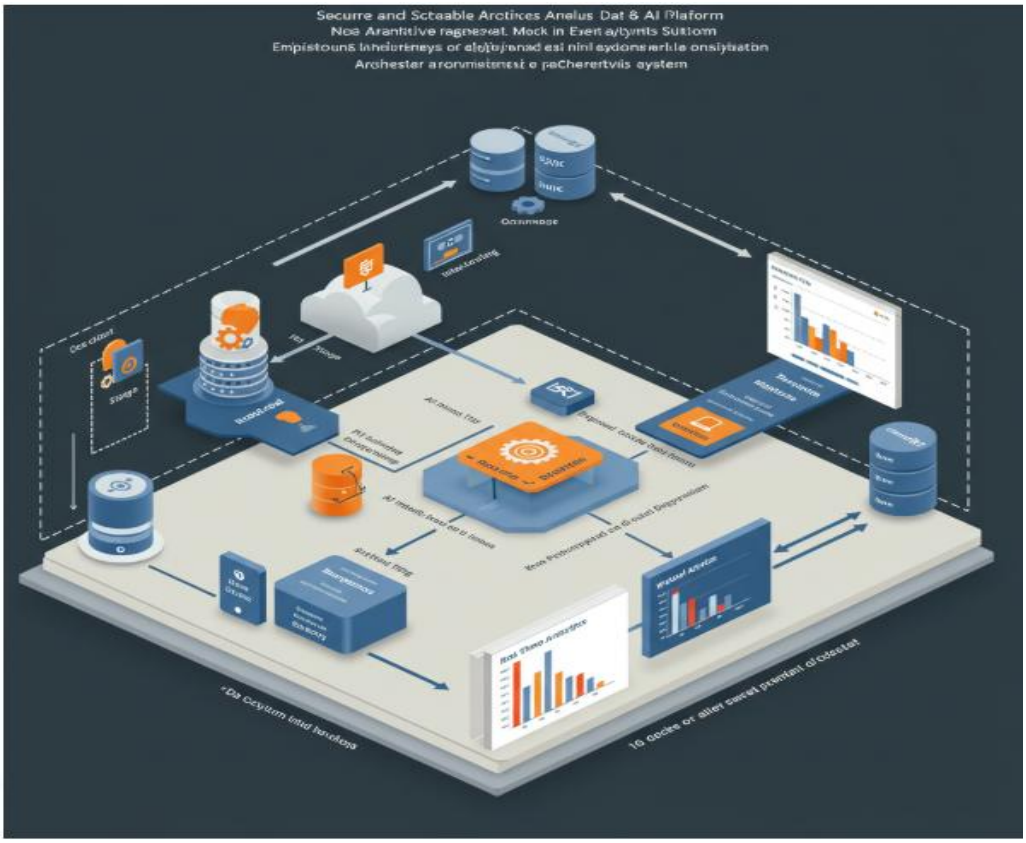

Fig 8.1: Building a Secure and Scalable Data and AI Platform

8.1.1. Background and Significance

Today's organizations are witnessing a renaissance of data platforms to store, process and analyze data at scale. Technology that was previously thought of as too complex or expensive is now becoming commodity. Distributed data processing engines allow organizations to analyze large amounts of data using their existing data warehouses in the cloud. Cleaning, preparing and processing this data using Data Pipelines is easier and

cheaper than ever with various tools. Financial data has become a hot commodity as today's organizations have awoken to the need for data driven decision making going beyond sticking to quarterly balance sheets. This change has further accelerated due to the pandemic where work-from-home has forced traditional organizations to monitor their employees remotely leading to tunneling down on inefficient employee performance tracking software. The Finance industry at large is about to experience a Fintech renaissance where clients will have access to state-of-the-art investments and wealth management. As a by-product of cryptocurrencies and growth in the DeFi space this decade will see bloated, complex but extremely profitable algorithms in exchanges, instability warning models, prediction, with hyperparameters influencing a complex number of Markov Chains evolving fastly toggling on/off.

It will be important for future trading firms to rethink their software stack based on the developments in data integration and speed of analytical capacity. They are stumbling onto a variety of issues while running large-scale data processing workloads such as design partner networks, competing DSPs and backtesting ML. New ideas and concepts such as scalable on-prem data warehouses and future-proof large-scale computation at a fraction of today's costs are sprinkled throughout discussion boards. Most of the proposed solutions will turn into unscalable half-finished projects if not tracked back into principles and verified on benchmark examples.

8.2. Understanding Data Platforms

Continuous access to data is a basic requirement for conducting research-based consulting services. Big data platforms are a new breed of analytical tools that can empower organizations to overcome this challenge. A variety of tools, platforms, and technologies are available that give analysts many options to explore data in new ways and provide novel insights from it. Handling these technologies, however, can be complex and time-consuming. In addition, making the right technological choices is task specific. Therefore, building easy-to-use platforms for data consumption becomes an indispensable task for data-rich organizations. This section describes the data platform technology, its underlying architectural principles, important considerations, and needs for handling big data in consulting services.

The effectiveness of research and the credibility of advisory services depend on the availability of the right type of data at the right time. Organizations with easy access to data can propose innovative and analytical ideas to their customers. Big data platforms are new breeds of analytical tools that can empower organizations to overcome this challenge. They are a combination of databases, tools, and hardware that provide environments for smoothly handling large amounts of data.

8.2.1. Definition and Importance

Data platforms aggregate and analyze information from various sources to create a consolidated repository of knowledge to be publicly reviewed. In the Advisory Services context, information from a significantly higher number of disparate sources can be aggregated into one single place, enabling a more analytical approach to form conclusions on shared problems, as well as verify earlier conclusions. Based on this technical definition, data platforms can both provide significant analytical advantages, while posing a risk to privacy regarding sensitive information. Data platforms are thus trusted repositories, and as such make for a well-defined set of risks for which specific protections need to be provided. This set commonly includes: Availability, Protect against malicious acts (Deniability, and Impairing attacks), and Integrity.

Availability is concerned with system uptime and resource provisioning. Protecting against malicious acts is an umbrella of protection against acts which would typically aim at undermining the data on the platform, and is thus the most expensive type of protection offered, covering losses in the form of missing or incorrect data. Deniability is an important special case of protecting against malicious acts, of significant relevance in Advisory Services, where some platforms might hold sensitive information on some users against whom both physical act and legal prosecution might be warranted. Deniability implies protecting against lawful acts, and comprises amassing protections against such acts under the umbrella of pleaded plausibility.

A great many protection methods exist for the above protections, ranging from elementary, to very sophisticated methods, and the application of each method depends on the platform process design, the types of sources feeding the data into the platform, and the desired properties and guarantees regarding the platform. Hence, the same set of risks will be encountered by all Advisory Services, but the modulating factors will have significant influence on the design of the requisite protections.

8.2.2. Components of Data Platforms

A data platform is defined as a set of components for collecting, storing, processing, and consuming data. Platforms typically focus on providing facilities for certain aspects of data engineering workflows; however, the platforms are increasingly converging to provide more than one aspect. A reference architecture describes how the components relate to each other to build a data platform. The components within the architecture tend to focus on one aspect of data engineering workflows and can be replaced independently of each other. Some components can be deployed on cloud platforms while others can be deployed on-premises. External connectors are also supported for easy integration with other systems to expand their usage. The architecture is similar to a data platform

reference architecture, where organizations combine components in the cloud or on-premises to establish a homogeneous data platform.

Data platforms have a number of common components. Sources and sinks refer to the systems external to the data platform. They are connectors to commonly used third-party systems. For each source or sink, a variety of connectors are provided. A component refers to a set of tools for specific usage. A component tends to focus on one task; the data pipelines can be easily constructed from the tools provided by the component. Each component can be deployed independently; therefore, scalability can be increased easily. Servers refer to systems to execute the processes produced by the components. Components need to connect to a server to perform data engineering tasks. Platforms can run on public clouds, private clouds, or on-premises. Each has its own advantages and disadvantages. Many platforms provide public cloud offerings. Modifying and adding components can also be done transparently and with minimal disruption to the overall system.

8.3. Scalability in Data Platforms

High transaction workloads in data platforms may lead to unavailability or even data corruption of the entire system when components cannot handle the workloads and/or failures occur. Scalability for Data Platforms identifies single points of failure of any kind and how to eliminate them. It also considers how to minimize (or eliminate) system slowdowns while simultaneously adding computing resources. "It’s better to avoid it than cure it!" is especially true for expensive data platforms. The rise of big data presents significant business opportunities, but demands exponentially more storage and processing capacity. There is no doubt that businesses need help in gathering, analyzing, and utilizing this data at scale. Putting in place the requested solutions (in an efficient way) quickly is a complex challenge for collaborative and/or cloud-based businesses. Current data management solutions with scalable data analytics have a low adoption rate in the big data industry. Non-scalable data platforms having popularity in small to medium data volumes and needed by many small and medium-sized companies have been neglected.

Scalable data platforms are architectural frameworks mainly focused on data management and data analytics, which allow fast prototyping of business solution applications. A new generation of monitoring devices is proposed for the aggregation of huge amounts of numeric data. Customers, SMEs, and very large corporations are provided scalable data management solutions using the currently dominating NoSQL paradigm and deploying fully on the cloud. Collaborative information systems with no single points of failure are presented, which allow gathering big data from different agents and across different devices in a scalable way while keeping it fully secure,

transparent, and privacy-preserving. Proven architectures in detail regarding their implementation details, configuration parameters, and optimization knacks while avoiding needless theory are provided. Both data management and big data analytics are addressed for building end-to-end solutions applicable to thousands of businesses without a team of engineers.

The aim of this section is to clarify the precise meaning of the term scalability, as well as the limitations and difficulties of analyzing the scalability of a given system in detail. In the field of computer science, scalability refers to the capability of a system, network, or process to handle a growing amount of work. Scalability often involves a combination of hardware and software, and it is also referred to as performance scalability. Scalability can be seen as a continuing property (the system is scalable with respect to the considered measure) or as a threshold property (the system supports scalability as long as the quantity or measure does not exceed a given threshold).

Fig 8.2: Modern data platforms and scalability

8.3.1. Concept of Scalability

Scalability is a concept widely used across multiple disciplines, including those based on systems and computer science, which refers to the potential plasticity of a given object. Despite its simplicity, this term has tended to be used in different ways, which is compounded by the fact that other related terms, such as performance or extensibility, have been used as a synonym for scalability. The aim of this section is to clarify the precise meaning of scalability, as well as the limitations and difficulties of analyzing the scalability of a given system in detail. In the field of computer science, scalability refers to the capability of a system, network, or process to handle a growing amount of work. Scalability often involves a combination of hardware and software, and it is also referred to as performance scalability. Scalability can be seen as a continuing property (the system is scalable with respect to the considered measure) or as a threshold property (the system supports scalability as long as the quantity or measure does not exceed a given threshold). A distinction is made between quantitative scalability, where the analysis suggests that performance will change as a predictable function of the scale factor, and qualitative scalability, which indicates that the effects of the scale factor on system performance have a qualitative character. Scalability is a concept widely used across multiple disciplines, including those based on systems and computer science. Scalability refers to the potential plasticity of a given object. Despite its simplicity, this term has tended to be used in different ways, which is compounded by the fact that other related terms, such as performance or extensibility, have been used as a synonym for scalability.

8.3.2. Techniques for Achieving Scalability

Software products are becoming bigger and more complex, mainly due to the increasing number of users, transactions, relational schema size, complexity of transactions, and also the performance of other services involved in it. Distributed databases are a common solution for scaling applications (Leigh, 2023; Khan, 2025). The capabilities of such databases vary, so it is important to correctly analyze system requirements before any hardware or software investments. The analysis allows the distribution of the collected information into several groups, depending on the needs. This is very helpful for getting a good architecture of the system and permits to judge if a distributed database is needed and to point out the appropriate technologies.

In order to keep costs low, many companies choose to store and process their data in-house instead of going for cloud services. In-house data management infrastructures cannot compete with the hardware power of large companies, so this choice usually leads to the need to carefully analyze the system requirements and logic in order to correctly model them and to choose the appropriate software technologies. However, companies that are growing fast or that want to change their internal organization to deal with their

mission more efficiently, often explore the cloud services option. Current cloud data management services offer basic query processors and storage systems with several cloud providers not allowed to have complete control of the systems. Nevertheless, due to the reasonable prices and maintenance costs, this solution is worth considering.

With the increased usage of distributed databases, solutions to guarantee their consistency have emerged. It is important to choose the appropriate consistency level, since it directly affects read/write ratios and therefore the overall performance of the system. Most techniques regard the case of servers located in the same datacenter. With multi-datacenter solutions, the number of clients and caches read replicas need to be increased due to the higher latency of reading with respect to writing. Many of these solutions, though, do not guarantee strong consistency, so it is important to understand the possible trade-offs, such as increased client-side complexity.

8.4. Security Considerations

Data security continues to be a paramount concern as data platforms are used to store, process, and analyze massive amounts of data. Although the main objective of the processing in the context of data platforms is to extract insights that are used to support decision-making in organizations, the platforms often handle sensitive data that must not be exposed to unauthorized access or to eavesdroppers. Although data security is a wide topic, this section will treat data security through privacy preservation, which is a hot topic nowadays as past practical examples have shown big organizations compromised with sensitive data leaks. Furthermore, data privacy preservation also opens up a big opportunity for research on later scheduled research agenda. The description of data security applications is based on the conceptualization of data security and privacy preservation approaches for the consumption of external data in the context of the presented research agenda (Paulson, 2025; Leigh, 2023).

Scalability in the data platform is addressed by formulating data processing as a set of maps and reduces with the application of the MapReduce programming model on external data. However, as the data processing is elaborated on potentially sensitive data, it will be dealt with privacy preservation of the data. Most privacy preservation algorithms currently used provide a way of lossy data anonymization prior to the data analysis step, but they cannot assure privacy preservation in the data processing step. As the platforms to analyze the data in peer-to-peer fashion are becoming popular, the question of what to publish as the analysis results arises. As the analysis usually involves discovering some knowledge such as clusters or correlations that hold over the data, another question comprises whether the privacy of the data sources is preserved in the knowledge. Conjointly the data privacy preservation and privacy preserved knowledge discovery on outsourced data streams are discussed. The description of data privacy

preservation is based on the conceptualization of data privacy preservation mechanisms. Data privacy preservation is handled regarding encrypting, anonymizing, and adding noise to the data, which is treated as functions of drawing bounds regarding the audience of the readable data. It is also regarded as a mechanism of formal security for accessing shared data sources with cryptographic protocols on the foundations of the proposed work on accessing relational data sources.

8.4.1. Data Security Fundamentals

Security was given priority in the design of the service and platforms used. This section describes the fundamentals of data security of any setup. Data governance excellence, as a set of data management capabilities, is another important enabler of data platforms for data science. It consists of organizing principles, practices, and standards tailored to the needs of the organization. It ensures that the organization can deliver data to the right people at the right time in the right form and that the proper safeguards are in place to preserve data quality and security as demanded by ethical and other requirements. Data governance excellence was adopted as the basis for maps of good practice in three areas: technical services for data acquisition and pre-processing, setup of data science environments, and processes and standards for data engineering and machine learning. Data Governance Excellence partnered with a platform for the data repository and data science environments. The two partner organizations are part of a service that holds, curates, and enables discovery and access to secondary data at the national and international levels.

To operate the service, there is a need to provision needed resources as a service to service operators and users, encrypted access. Local data repositories are uploaded to map location-based file/data warehousing. On-premises Data Services hosted on the triple-cloud are made available as services. Data cleaning, enrichment, and transformation are performed locally and uploaded as datasets to the Standards are mapped onto risks, best practices and enablers. With access to automatic logging services, data veracity can be ensured. User engagement auditing, problem reporting, and logging of incidents are available as services, and access is logged.

8.4.2. Common Security Threats

As process-oriented, resilient and flexible architectures have become a top priority in managerial, research and societal agendas during the past few decades, large-scale data platforms for data-driven, operations-scenario-based advisory services for manufacturing have been receiving increasing attention from academia and industry. In particular, cloud-based architectures for such platforms have increasingly gained interest, as they provide resilience against cyber-attacks, remain sufficiently adaptive against exogenous shocks and allow applications to scale up effortlessly. However, the demand for secure cloud infrastructures is put forward that strains attention and hampers

operation of data-intensive applications. Consequently, while these applications are technically embraced by the cloud, they risk being under-utilised and, in worst cases, discarded, which motivates the need for tightly integrated and cross-cut, thus scalable, security bids for such platforms. However, a thorough understanding of the security needs of such platforms, as informed by the collaborative use of data between multiple parties, is lacking. Therefore, there is a call for multi-dimensional assessments of these security needs associated with the requirement for accentuated interoperability, while keeping consideration of divergent data governance. These security needs are subsequently addressed through a comprehensive architectural design framework that allows tight integration of security solutions across platform components and scalability in handling multiple events. Its novelty lies in that cross-cut, previously separate elements of security are holistically tackled in a tight, integrated way while differentiating them in handling horizontal, vertical or diagonal aspects of security. Multi-dimensional security threat assessments are critical to the design of feasible and effectual security solutions for data platforms. It identifies common security threats of data platforms that are cross-cut or meta features of other security solutions and initiatives. Subsequently, a thorough theoretical examination of the common security needs of data platforms is conducted.

Inspired by a widely embraced architecture framework that allows meticulous classification of threat solutions, existing security solutions and initiatives for data platforms are systematically reviewed and analysed with regards to their languages, edges and properties. A comprehensive security architecture framework that allows tight integration of security solutions across data platform components and scalability for handling multiple events is derived. Subsequently, a qualitative inquiry of powerless data platforms in kicking off the mind-set change needed to embrace cross-cut security solutions is reported. In shaping an imagined future for the evolving discourse, four modes of hope—narrative, visual, and sonic; interpersonal, participatory, and transformative; prescriptive; and provocative, enactment, and playful—are identified and elaborated.

8.5. Technological Frameworks

The Data Health Monitoring Framework (DHMF) assessment and monitoring framework consists of four main sections: compliance rules, monitoring indicators, compliance indicators, and quality-of-life measures. It can encompass any number of specific rules in its general specification. Each compliance rule has been formulated using simple propositions constructed from the indicators available to the monitoring agent. A monitoring agent can combine many rules to create a powerful assessment tool adaptable to different application contexts. Environmental change, systems evolution, or

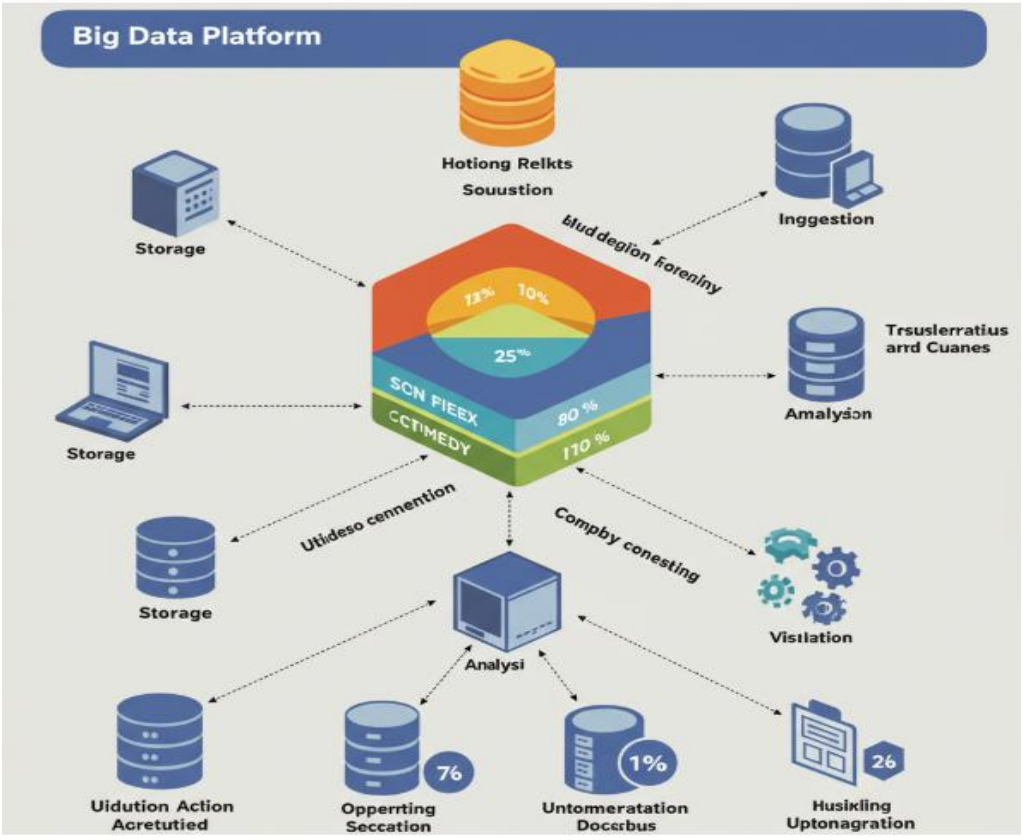

Fig : Ideal Suite of Solutions to Set up a Big Data Platform

modifications in the goals and policies of the governing agent can result in the introduction, removal, or change of compliance rules. The tools for compliance assessment outlined above impact the system by modifying which rules are monitored or how the monitoring agents interpret or take action on various indicators. This has the potential to allow for a more application-focused customization process for tuning the monitoring processes within the limits set by the governing agent. All systems are capable of general modification within their structure, implied or concrete. Actions taken to modify or define structures (change definitions, rules, policies, reallocate resources) can either be static or dynamic (i.e., they can modify systems peripherally or simultaneously) . In a concrete modification process, elements accessible to the governing agent within a candidate structure can be changed.

In terms of enforcing behavioral rules, three main choices can be adopted. Actions can be taken to force compliance amongst the agents, making it impossible for them to disobey or deviate from the provided rules. Alternatively, agents can actively ensure compliance by checking newly proposed actions. This approach is similar to regulatory approaches in the real world. A softer approach that allows agents some leeway is to

create incentives for compliance and/or temporary disincentives for non-compliance. Classic examples of this are monetary bonuses or tax reductions for those adhering to environmental regulations.

8.5.1. Cloud-Based Solutions

Cloud computing has become a modern-day necessity for many organizations providing research, analytics, data, and software services. Cloud-based solutions provide a space to store large amounts of data, as well as processing power to compute and analyze the data. Systems have been built on cloud-based solutions to recruit participants, screen them for eligibility, and manage the preparation, collection, and storage of patient-reported outcome measures from subjects prior to clinical trials. Cloud-based solutions are the only plausible services for consumer-driven data-sourcing that are able to meet the needs of technology forward researchers and industries, especially when prioritizing speed, ease-of-use, and huge processing capabilities.

Medicare has phased out paper-based claims data including the enrollment, eligibility, and payment history of the kit recipients. A data platform has been created on AWS by researchers from the Emory University School of Nursing in Georgia that will help identify success factors in broader outcomes across multiple insurers. Interdisciplinary teams from the Centers for Medicare and Medicaid Services (CMS), MBL Technologies, Emory University, and Emory's IT department collaborated to develop the Proteus platform at Emory. Proteus is a secure, reliable, and scalable platform that stores and analyzes identifiable research data stored in Amazon Web Services (AWS) Simple Storage Service (S3) using the AWS Elastic MapReduce (EMR) and Redshift services, and RStudio Server. Statistical computing language R is used for data management and analysis, using the RStudio IDE. Collaboratively involved in the transition to the cloud-based platform were programmers who successfully implemented the technical infrastructure, as well as policy advisors who ensured compliance with the complex and challenging requirements of the CMS policy. Important lessons learned throughout the AWS transition process, and the decisions made, challenges faced, and solutions discovered to meet CMS policy requirements are documented in the hope that they will provide guidance for other organizations facing similar circumstances.

8.5.2. On-Premises Solutions

The evolution of scalable data platforms solves the challenge of data growth through massively scalable and distributed databases, data storage platforms, modern data processing engines, and cloud-based data analysis. These are closed systems designed, implemented, or deployed on a single site where all resources, memory, and disk space

are managed by a single machine or in a local area network. In contrast, the scale-out strategy relies on flexible data storage, access, and analysis distributed in a geographically redundant manner over different machines to achieve availability and fault-tolerance, with respect to the fact that, even if the distributed system design is more complex, it presents considerably better cost scaling compared to a single site. Protocols are applied to various levels of distributed data access, enforcing consistency and transactionality guarantees. Here, consistency and transactionality refer to the properties of correctly designed transactional data storage, including first normal form, the ACFD properties, and strict 2PL concurrency control protocols. Given that many levels are required in distributed systems to translate independently designed competitors, it follows that distributed systems can rarely provide full-fledged industrial-grade ACID guarantees. These guarantees are nevertheless required for critical advisory based services that demand high-level integrity and confidence. Global data privacy and data protection laws and regulation restrictions stipulate the locality of the data stored or processed on a given within a jurisdiction to assess the risk of data offers to data losses or breaches. Missing or wrongly configured legal and contractual precautions and track governance of suspicious data handling often lead to proof-of-compliance errors. Storing and transferring sensitive organizational data assets to critics' trust third-party systems raise these concerns. In the early phases of the cloud paradigm, some organizations did not transfer their data and service outsourcing to cloud solutions based on a general distrust of complex third-party-offered systems. Recent efforts have led to the offering of cloud solutions supported by several compliance certifications and global insurance creations to guarantee minimum service level expectations. However, this has not covered the pitfalls due to heightened security concerns over secret data handling. Data audit trails highlight critical security concerns, including potentially destructive behavior by junior admins, critical data deletion or corruption errors while being transferred or processed, loss or damage due to power outages or other failures, or secondary or insider use overrunning contractual precautions.

8.6. Conclusion

All these considerations show that designing, implementing and maintaining a data platform that provides wide scalability while respecting ACID and robust availability guarantees is a very complex task. Solutions built around eventual consistency, when used wisely, can provide a good tradeoff between availability, scalability and implementation simplicity, but it is expected that they will not reach the QPS levels of social networks. Intermediate solutions, like sharding with strong consistency and limited scale violations, can easily work within tight bounds on the sizes of the entities, but a price must be paid in terms of complexity and operational burden. Finally, it is possible to provide guaranteed low latencies and very high QPS levels over workloads

that tightly control the access patterns, but designing satisfactory systems for more complex sets of queries is really a tough problem that has not been properly solved yet. An alternative and increasingly popular approach to deal with high volume and high velocity data is to build systems whose purpose is to collect a non-trivial amount of data and periodically run processing together with their application-dependent state information. The goal of data platforms is to provide a versatile set of off-the-shelf technology modules whose behaviour may be tuned to a great extent within the boundaries of a robust performance and availability behaviour. They should keep a generic data selection service providing a limited but robust forward-looking selection. Support and documentation to ease such tuning, technology selection and proper architecture design, focusing on the performance and continual availability properties of the systems, should be included. A group of technology modules to be included in external packages will be acknowledged in advance for proper splitting of work and joint design and testing.

8.6.1. Future Trends

Data platforms have been consistently evolving and improving in terms of scalability and security. It is expected that these platforms still have a lot of room for improvement. In this section, the related ongoing trends are introduced. It is highly expected that these trends will significantly improve the current data platforms and the proposal suggested in Chapters 5–7.

Future scalability challenges are mainly categorized into three aspects: (1) speed of data growth, (2) ask for more sophisticated systems, (3) systems are composed of hybrid technologies. As for the speed of data growth, it has been imagined that storage systems will evolve in accordance with Moore's Law. However, this trend is not observed in network systems. The expectations are that network traffic and access frequency will rise 100-fold in the coming 10 years. With the rapid growth of web technologies, data can be processed in a high frequency manner. New data technologies will be required to process transactions in a time window of millisecond accuracies.

This raises some barriers to the existing distributed data platforms. Only one master server can respond to requests and append the logs for the single transaction group. For data horizontally, the possibility of replica locations is limited by the summation of the replicas to estimate the global load balance. There is also a problem with multi-region replication in synchronism with the Cloud technology. The cause is that the cloud platform becomes hybrid and data fragmentations can be tiered on cloud systems. Some recent developments tend to increase efficiency to process massive requests in terms of global load balancing and scalability.

As for future security challenges, the explosive adoption of web, internet of things, and web services has arisen the alarmingly increasing frequency of data breaches and privacy violations. The existing decrypted solutions are still insecure because of the central key management mechanism. The entire system becomes insecure as soon as the master key is exposed due to some reasons. New schemes are expected to be secure against leakage of the entire key and latency of system transactions in the cloud environment or outsourcing environment. It is also challenged to keep the data highly efficient and to maintain the back-end compatibility with minimal modification of applications and systems.

References

Khan, M. (2025). *The Role of AI and ML in Tax Analytics for Strategic Decisions*. Apex Accountants & Tax Services.

Leigh. (2023). *Embracing the Power of AI: Transforming Corporate Taxation*. Concise.

Paulson, D. (2025). *AI in Tax Advisory: A Revolution in the Making*. Paulson and Partners.