Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2021, 8(12):239-247**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Enterprise Architecture with Databricks

Ravi Shankar Koppula

Email id: Ravikoppula100@gmail.com

Abstract In order to ensure a unified operational view and robust alignment of IT resources with business objectives, the adaptation of enterprise architecture is essential. The establishment of effective EA presents numerous challenges and holds critical influence on organizational decision-making. This paper explores the integration of Databricks, a unified data analytics platform, as a strategic solution in the EA framework due to its ability to mitigate complexities associated with big data by providing scalability, flexibility, and security for advanced analytics and data management. Within this research scope, an examination is conducted on fundamental components of EA such as business, information, application architectures alongside technology architectures; aiming to understand the multifaceted nature of EA. Specific advantages are explored concerning Databricks' inclusion in EA which encompasses enhanced capabilities for data processing, analytics governance security features coupled with interoperability within existing data lakes and warehouses. To leverage Databricks' strengths effectively requires following structured approach consisting needs assessment tailored architecture design integration efficient pipelining workflows, and continuous monitoring optimization processes sustenance peak system performance.

Keywords Data Lakehouse Enterprise Architecture, Databricks Enterprise Architecture, Data Analytics Architecture Framework

Introduction

Enterprise Architecture (EA) refers to the process of aligning an organization's business strategy and objectives with its IT infrastructure and systems in a formal manner. The aim is to create a comprehensive overview of the organization's operations and identify areas for improvement and optimization. EA serves as a structured framework for understanding the connections between different elements within the enterprise and facilitates informed decision-making regarding technology adoption and implementation. An ongoing challenge faced by organizations is how to effectively develop architecture. It is crucial to recognize that architecture choices have long-term implications for both compute costs and labor costs. Additionally, architecture decisions are often difficult to rectify through subsequent code optimizations and refactoring. As Ralph Kimball once emphasized, the selection of architecture is a foundational and early decision in the system's design.

The landscape of enterprise decision support systems has undergone significant transformations in the past ten years. In this discussion, we will explore these changes and examine how Databricks has emerged as the leading solution to address the challenges accompanying these changes. This guide delves into the role of Databricks in enterprise architecture, highlighting the fundamental concepts, advantages, and recommended approaches for its utilization.

A. Definition of Enterprise Architecture

Enterprise Architecture (EA) is a discipline that focuses on understanding, designing, and implementing the structure, processes, systems, and technologies of an organization. It involves creating and maintaining a comprehensive blueprint that outlines all elements of the enterprise and their interdependencies. EA enables organizations to achieve their strategic goals by aligning business and technology objectives. It provides a structured approach for managing complexity and change, ensuring interoperability, and optimizing resource utilization. Enterprise Architecture refers to the systematic and holistic approach that organizations use to align their business objectives, processes, information systems, and technology infrastructure. Enterprise software, therefore, can be described as software that caters to the unique needs of large-scale enterprise companies.

B. Importance of Enterprise Architecture in Organizations

Describing Enterprise Architecture (EA) and its importance in organizations, it serves as a vital tool in aligning business and IT strategies, improving decision-making processes, and promoting efficiency and agility. EA offers a comprehensive view of the organization, identifying redundancies, gaps, and optimization opportunities. It fosters collaboration between various teams and departments, ensuring that IT investments are aligned with business objectives. Additionally, EA enhances organizational flexibility, mitigates risks associated with technology adoption and implementation, and enables prompt response to market dynamics and customer demands. Enterprise software, which constitutes the components of EA, is a collection of applications and technology stack choices. These software components possess certain characteristics specified in the requirements.

Persistent data is stored in such a way that it can be accessed over a prolonged period of time, primarily due to its expected long-term use.

There is a substantial quantity of data due to the numerous business units and processes that generate it, which increases the probability of having extensive volumes of data.

Numerous individuals are accessing data, typically in a concurrent manner. Integration with multiple other enterprise applications is an essential aspect of modern data lakes. The transition from traditional data warehousing to data lakes has led to a significant increase in data volumes and sources. As a result, a diverse range of data types is now available, offering valuable insights. It is no longer limited to flat tables. These advancements necessitate updates to existing systems to accommodate the changes. Furthermore, organizations face the challenge of ensuring data quality and reliability, as accurate information is crucial for the single source of truth. Consequently, a well-designed system must address the organization's multifaceted requirements. Moreover, decisions made at the architectural level significantly impact the overall effectiveness. Refactoring an already-architected system poses considerable difficulties. Neglecting intentional architectural planning is not an option, as it will lead to various issues such as duplicated efforts and data, inefficient processes, and complexity.

C. Role of Databricks in Enterprise Architecture

Databricks is a software-as-a-service company that offers a Unified Data Analytics Platform (UDAP) that enables organizations to bring together their data science, data engineering, and business analytics workflows.

Databricks was founded by the same people who created Apache Spark (the de facto standard processing engine in big data processing). In addition to the continued contributions to Apache Spark, Databricks has also created other open-source projects such as MLflow and Delta Lake. These open-source projects are a big reason why the UDAP is effective for data ingestion, munging, and running machine learning experiments.

Working with big data is not easy. Big data is hard to manage. - it is coming in massive volumes, faster than ever before, and in a wide variety of formats.

As data practitioners work to design their organization’s big data infrastructure, they often ask and need to answer questions like:

[1]. Where/how will we store our big data?
[2]. How can we process batch and stream data?
[3]. How can we use different types of data together in our analyses (unstructured vs. structured data)?
[4]. How can we keep track of all of the work we’re doing on our big data?

D. Unified Data Service

The Unified Data Service is the processing engine of Databricks-- it’s what powers all of the work the various data practitioners working in the platform perform. It includes an optimized version of Apache Spark.

E. Enterprise Cloud Service

Databricks is a managed service that can be deployed in your AWS environment, in your Azure environment, or in your Google Cloud environment.

F. Collaborative Data Science Workspace

The Data Science Workspace is the physical online platform that data practitioners work in. Team members log into the UDAP using their own credentials and depending on their roles, access different components within the Unified Data Analytics Platform to perform their daily workflows.

G. Business Analytics

In the case that team members want to use UDAP compute power (the optimized version of Apache Spark) without logging into the Data Science Workspace, Databricks offers BI integrations. These BI integrations allow for this compute power to be accessed through APIs or through BI clients. In other words, if a data analyst wants to use the optimized version of Spark to query data but they want to do their work in Tableau, Looker, Power BI, or others, they can.

H. New Challenges working with big data

Challenge number two - data isn't just tabular anymore. Often, these technology stacks don't work well together due to many different tools introduced that are not compatible with each other, proprietary data formats that are not easy to translate from one tool to another. If you want to enable advanced use-cases, you have to move across these stacks. Think about the work your organization does. You don't typically have a "data warehouse use case" or a "streaming" use case. You most likely have a supply chain logistics use case or a financial risk assessment use case. To address these challenges, you need to move across these technology stacks.

As you can imagine, there are many ways that an organization can set up big data infrastructure -- getting it right is no easy task.

I. Siloed roles lead to organizational inefficiencies

Even once a big data infrastructure is set in place, many organizations suffer from the challenges of having siloed functional roles for individuals on their data science teams. As we mentioned, working with big data is complicated, and without team collaboration and transparency on big data workflows, inefficiencies can ripple through an organization. For example, it is not uncommon for a data scientist to build and train a machine learning model in a vacuum on their own computer, with little to no visibility to related work being done by, for example, the data engineer preparing that data for them, or the data analysts who might be using results from their experiments to produce dashboards. We'll discuss later how there is ongoing discussion in the industry about how to organize data engineering processes and teams, with concepts such as the "data mesh", where separate teams own different data products end-to-end, gaining popularity over the traditional "central data team" approach. Lakehouse designs lend themselves easily to distributed collaboration structures because all datasets are directly accessible from an object store without having to onboard users on the same compute resources, making it straightforward to share data regardless of which teams produce and consume it.

J. Protecting customers and their data is difficult

According to Gartner, 80% of organizations will fail to develop a consolidated data security policy. This leaves them and their data vulnerable to security breaches. Think about the ramifications of a security breach. Beyond just the immediate monetary cost, there is a long-lasting loss in customer trust and company reputation. If you've ever been a customer of a company that has suffered a security breach, you know first-hand how long it can take to rebuild trust. In addition to protecting data from leaking out, organizations must also make sure they're compliant with data protection regulations like GDPR (European Union's General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act), or that they have required certifications to run their businesses. And, there can be hefty penalties involved if they are not compliant.

K. Traditional architectures for working with big data need improvement

Not all architectural patterns work well for big data management and analytics. For example, older architectural patterns might struggle to simultaneously process batch and streaming data. This means that anytime a data engineer needs to validate, reprocess, or update batch and streaming data, they might deal with:

Complexities from having to manage separate code bases and workflows. Difficulties merging/reconciling data for one single source of truth. Aside from this, using older architectural patterns can make it difficult to guarantee data availability for everyone (who can access it and when), implement security controls or know which data can be trusted. In summary, it means that data teams end up spending more time processing and managing data than actually working with it to derive insights.The emergence of unified data analytics stemmed from helping organizations overcome these challenges. We'll go in great depth on a new paradigm in architecture - the Lakehouse - but at a high level:

A data lakehouse is a new, open data management paradigm that combines the most popular capabilities of data lakes and data warehouses. Per their design, they implement similar data structures and data management features to those in a data warehouse, directly on the kind of low-cost storage used for data lakes. Merging these ideas into a single system means that data teams can move faster as they can use data without accessing multiple systems. Data lakehouses also ensure that teams have the most complete and up-to-date data available for data science, machine learning, and business analytics projects.

Key Concepts of Enterprise Architecture

Enterprise Architecture (EA) is a strategic framework utilized by organizations to align their business and technology strategies and accomplish their objectives. It encompasses numerous significant concepts that are vital for a successful implementation. These concepts include business architecture, information architecture, application architecture, and technology architecture. Each of these concepts is instrumental in developing a comprehensive understanding of the organization and its systems. By comprehending these key concepts, organizations can proficiently strategize, design, and oversee their enterprise architecture to foster innovation, enhance efficiency, and facilitate digital transformation.

A. Business Architecture

Enterprise Architecture does not exist as a thought experiment. It is the result of careful planning on behalf of the data team to provide the organization with relevant information. While requirements gathering, we should not only be mindful of the needs the organization currently has, but also mindful of the needs the organization will have in the near to long term future. There are many ways to approach this. We could:

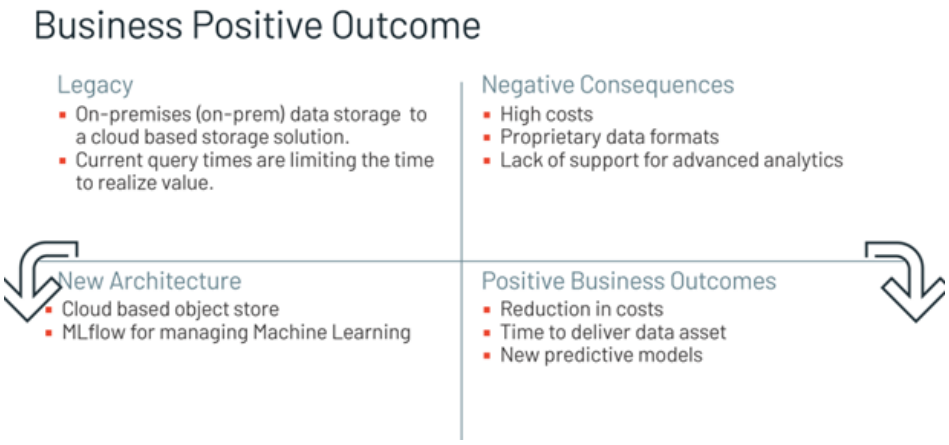
[1]. Make a list of the questions which members of the organization ask.
[2]. Actively poll potential stakeholders and understand their pain points.
[3]. Examine various personas throughout the organization and their data needs.
[4]. Write down use cases from other organizations in similar industries.

Once we have potential scenarios and use cases which the organization has identified as an area which requires information, we will need to frame a potential data asset in terms which would require initiative and buy in. For example:

The organization would like to move from legacy on-premises (on-prem) data storage to a cloud-based storage solution.

Current query times are limiting the time to realize value. Data assets like a dashboard need a faster runtime to provide actionable insight

Once we have potential scenarios and use cases which the organization has identified as an area which requires information, we will need to frame a potential data asset in terms which would require initiative and buy in. For example:



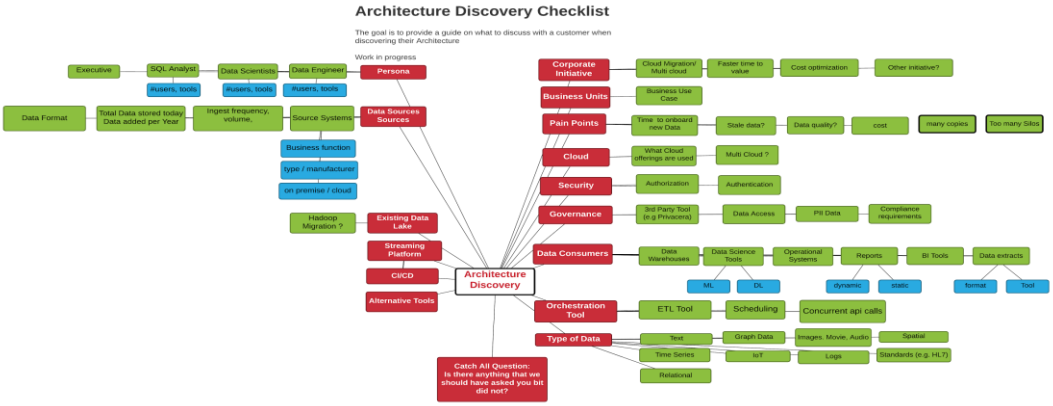
In order to assess whether a data asset is required by the organization, you can ask some high-level questions to begin with:

[1]. What is the top line revenue impact?
[2]. What are the operational cost savings?
[3]. Is there business risk mitigation associated with this new architecture?

From there, we can suggest metrics. In respective order to the points just listed:

[1]. Is there a percentage that we can attribute to a faster delivery of a product or service?
[2]. What does that percentage of total impact in terms of dollars?
[3]. What would that reduce underlying costs by?

Currently, we want to ask questions that pertain to the top right of this diagram:



In identifying positive outcomes, we might instead begin our journey with some challenges we see in the organization. Note, in the next section, we'll discuss challenges on the road to implementation. Those implementation challenges are different from what we're talking about here. Here, we're talking about the challenges that the current architecture and tech stack that the organization has in place is impacting the organization.

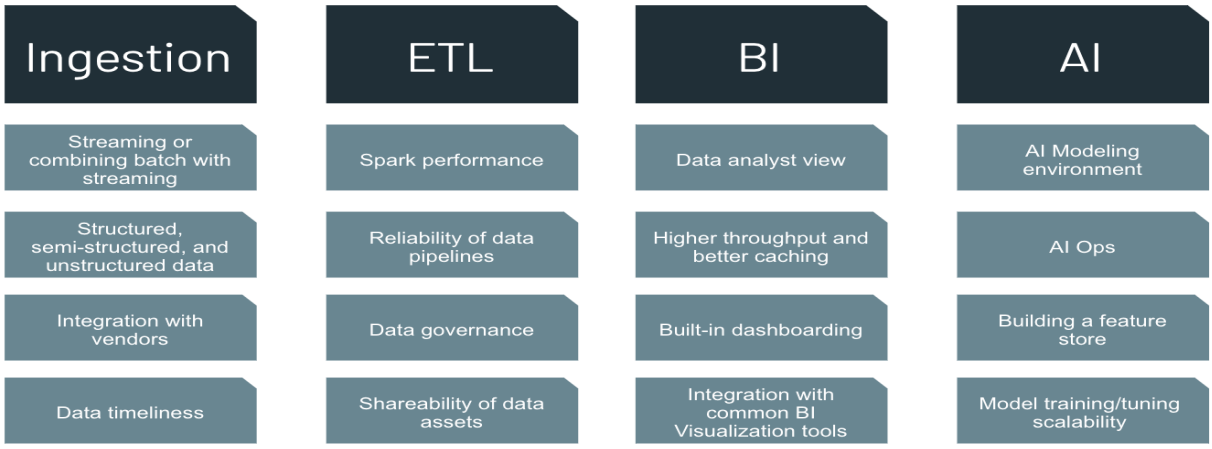
[1]. How is the organization's current challenge measured? If it's not, we are architecting from scratch.
[2]. Who owns the outcome of that challenge?
[3]. What are the numbers (AKA metrics or KPIs) that are the evidence of the challenge?
[4]. What would good numbers look like?
[5]. Who else is affected by the challenges?
[6]. How high in the organization does this pain reach?
[7]. What effect does this have on the business? How often? What does that keep the organization from doing? If you could solve, what would that allow? What good things would result?

Challenges

Often, there are challenges that come with implementing an architecture, especially if that architecture contains components which don't integrate well with newer and more modern technologies.

Common Challenges	Impacts of Challenges on Organization
Silo reporting systems so it is hard to have unified view across different business processes	Inconsistent Reporting
Complexity in providing real-time analytics	Stale Data
Complexity in joining streaming and batch data	Siloed data
DevOps utilization is really high because of job management: team needs to move data across different systems	Resource allocation
Data science teams are working on silo and/or stale data	Inaccurate forecasting
It is hard to discover and find the related data.	Resource allocation

We should begin with a survey of the technology components that exist and will not go away. Many times, this will determine a large number of considerations such as: "which third party tools and services can we reliably connect into our existing tech?" What in the following image currently has blockers?



B. Information Architecture

Information architecture is an essential aspect of enterprise architecture that focuses on how information is structured, organized, and managed within an organization. It encompasses defining strategies for data assets, data flows, data governance, and data integration to ensure that data is consistent, accurate, available, and secure. Effective information architecture allows organizations to efficiently store, manage, and retrieve data, facilitating better decision-making, gaining data-driven insights, and enhancing operational efficiency. By implementing a strong information architecture, organizations can fully utilize the potential of their data and utilize it as a valuable strategic resource.

C. Application Architecture

Application architecture is an essential aspect of enterprise architecture that concentrates on the design and administration of applications and software systems in an organization. It involves comprehending the business requirements, determining suitable application components, establishing integration patterns, and guaranteeing the compatibility and scalability of the applications. Application architecture permits organizations to align their technological solutions with their business demands, enhance application development and deployment procedures, and enhance overall system performance. By implementing a strong application architecture, organizations can enhance flexibility, decrease intricacy, and promote digital advancement.

D. Technology Architecture

The technology architecture of an organization is a critical element of its overall enterprise architecture. It is primarily concerned with the careful selection, integration, and control of all technology components within the organization. This includes determining the necessary hardware, software, networking, and infrastructure that will best support the organization's business objectives. An effective technology architecture allows organizations to have a clear understanding of their technology environment, choose the most suitable technologies, mitigate technology-related risks, and ensure that their technology infrastructure remains stable and reliable. By establishing a strong and comprehensive technology architecture, organizations can maximize the return on their technology investments, enhance their operational efficiency, and promote innovation and growth.

Benefits of Using Databricks in Enterprise Architecture

Incorporating Databricks into your enterprise architecture brings a multitude of advantages. Not only does it enhance your data processing and analytics capabilities, but it also improves data governance and security, streamlines data integration and management, and offers scalability and flexibility for growing organizations. By harnessing the power of Databricks, your enterprise can effectively utilize its data, make more informed decisions, and gain a significant competitive edge in your industry.

A. Enhanced Data Processing and Analytics Capabilities

Databricks offers an enterprise architecture that enhances data processing and analytics capabilities. Its robust distributed computing framework enables organizations to process large amounts of data efficiently and quickly. The advanced analytics functionalities of Databricks facilitate the extraction of valuable insights from data, performing intricate data transformations, and conducting complex data modeling and machine learning algorithms. Consequently, businesses are empowered to make informed decisions based on data and unveil concealed patterns and trends.

B. Improved Data Governance and Security

Databricks enhances data governance and security within enterprise infrastructure. It ensures the safeguarding of sensitive data through robust security measures and access controls, both during transit and at rest. Moreover, Databricks presents all-encompassing auditing and monitoring capabilities, enabling organizations to monitor data usage, detect possible security breaches, and adhere to regulatory obligations. By integrating Databricks, organizations can establish a more secure and compliant data ecosystem.

C. Streamlined Data Integration and Management

Databricks plays a crucial role in streamlining data integration and management within enterprise architecture. It offers seamless integration with diverse data sources and systems, enabling organizations to effectively consolidate and manage their data assets. Not only that, Databricks provides robust data transformation capabilities, empowering organizations to cleanse, modify, and prepare their data for analysis and usage. By leveraging Databricks, enterprises can establish a centralized data platform and optimize their data integration and management workflows.

D. Scalability and Flexibility for Growing Organizations

Databricks is known for its ability to help growing organizations within enterprise architecture. With its cloud-based infrastructure, organizations can effortlessly adjust the scale of their data processing and analytics capabilities to meet their specific needs without having to commit to expensive hardware or infrastructure investments upfront. Additionally, Databricks allows seamless integration of different data tools and technologies, giving organizations the freedom to choose the solutions that best suit their evolving requirements. This scalable and flexible approach empowers organizations to effectively handle their expanding data volumes and easily adapt to changing business demands.

Implementing Databricks in Enterprise Architecture

Incorporating Databricks into the enterprise architecture necessitates the seamless integration of the platform within the existing organizational structure. This task entails a meticulous evaluation of the organization's needs and requirements, followed by the creation of an efficient architecture framework. The integration of both data

lakes and data warehouses is vital for ensuring the smooth integration and management of data. Additionally, establishing data pipelines and workflows facilitates effective data processing and analytics. Lastly, it is crucial to continuously monitor and optimize the Databricks environment in order to maintain optimal performance and fully leverage the advantages of using Databricks within the enterprise architecture.

A. Assessing Organizational Needs and Requirements

Before integrating Databricks into the enterprise architecture, it is crucial to thoroughly evaluate the organizational needs and requirements. This entails gaining an understanding of the organization's unique challenges and goals, as well as assessing the current infrastructure and technology landscape. Through this comprehensive assessment, the organization can identify the specific areas where Databricks can make the biggest difference. This evaluation process also aids in determining the necessary resources, skills, and technology integrations needed to successfully implement Databricks within the enterprise architecture.

B. Designing an Effective Architecture Framework

Creating a successful architecture framework for Databricks requires establishing a structure that is in line with the objectives and needs of the organization. This framework encompasses defining the overall data architecture, patterns of data flow, and connections with other systems. It also involves deciding on the necessary data governance and security measures to safeguard and comply with data regulations. Moreover, the architecture framework takes into account scalability, flexibility, and performance considerations to meet the changing requirements of the organization. By designing a comprehensive architecture framework, the groundwork is laid for the effective implementation and utilization of Databricks in the enterprise.

C. Data Lake and Data Warehouse Integration

Integrating data lakes and data warehouses with Databricks holds significant importance in the overall implementation of the platform in enterprise architecture. Data lakes act as a centralized storage hub for both structured and unstructured data, while data warehouses offer a well-structured and optimized storage system. The integration of these two crucial components with Databricks enables organizations to effortlessly and effectively integrate and manage data. This integration process expedites data exploration, processing, and analytics, allowing data scientists and analysts to extract valuable insights from the integrated data ecosystem.

D. Building Data Pipelines and Workflows

Developing data pipelines and workflows plays a crucial role in effectively implementing Databricks in enterprise architecture. These pipelines serve as a mechanism for effortlessly transferring and converting data across different sources and destinations. By crafting and constructing data pipelines, businesses can optimize their data ingestion, transformation, and integration procedures, ultimately enhancing overall efficiency. Additionally, workflows provide a structured framework for orchestrating and automating data processing and analytics tasks, ensuring a streamlined approach. By leveraging the power of Databricks, organizations can construct robust data pipelines and workflows, enabling them to effortlessly manage vast amounts of data and extract valuable insights promptly.

E. Monitoring and Optimization of Databricks Environment

Regularly analyzing the Databricks environment is imperative to achieve prime performance and fully leverage the advantages of Databricks within the enterprise architecture. Establishing streamlined monitoring mechanisms is pivotal for organizations to track system performance, resource utilization, and data integrity. Through proactive monitoring of the Databricks environment, organizations can promptly detect and address any issues or hindrances that may emerge. Furthermore, constantly refining the Databricks environment entails fine-tuning system configurations, optimizing data processing workflows, and implementing industry-leading methodologies to enhance overall efficiency and productivity.

Best Practices for Enterprise Architecture with Databricks

In order to optimize enterprise architecture with Databricks, it is imperative to adhere to recommended guidelines. These include establishing a well-defined governance framework and setting clear standards to ensure consistent and standardized usage of Databricks. In addition, collaboration between business and IT teams is vital for harnessing the maximum potential of Databricks, enabling a thorough comprehension of business requirements and technical capabilities. Data quality and consistency can be guaranteed through the implementation of robust processes and tools that validate and cleanse data, thereby ensuring accuracy and reliability. Finally, constant learning and improvement are essential for staying updated on evolving technologies and best practices in enterprise architecture, enabling organizations to continually enhance their utilization of Databricks.

A. Establishing Clear Governance and Standards

Creating clear governance and standards is crucial when implementing enterprise architecture with Databricks. It is essential to establish and record policies, procedures, and guidelines for using Databricks within the organization. By doing so, organizations can guarantee consistency, security, and compliance with regulations.

Standards may involve directives for handling data, storing and processing it, and controlling access. They can also encompass best practices for coding, testing, and deploying. The establishment of governance and standards serves as a framework that promotes efficient collaboration, decision-making, and risk management among teams engaged in the enterprise architecture process.

B. Collaborating Across Business and IT Teams

Facilitating cooperation between business and IT teams is an essential practice in enterprise architecture using Databricks. This cooperative method helps bridge the divide between business objectives and technical implementation. Business teams provide valuable perspectives on the organization's specific requirements and goals, while IT teams contribute their technical expertise and understanding of Databricks. The collaborative efforts result in a deeper comprehension of the business context, guaranteeing that technical solutions align with business requirements and that Databricks is effectively implemented to maximize its potential for the organization. Ongoing communication, collaborative planning, and joint decision-making enable business and IT teams to optimize the utilization of Databricks and achieve successful outcomes.

C. Ensuring Data Quality and Consistency

One critical aspect of enterprise architecture with Databricks involves the meticulous maximization of data quality and consistency. To obtain accurate insights and make well-informed decisions, organizations must possess unwavering confidence in the integrity and caliber of their data. This necessitates the implementation of stringent data quality checks and validation processes within the Databricks environment. By firmly establishing and rigorously enforcing meticulous data quality standards, organizations can effectively identify and rectify issues such as missing or incongruous data, outliers, and errors. Additionally, it is imperative to establish robust data governance practices to ensure proper management, meticulous documentation, and authorized access to data. By steadfastly upholding data quality and consistency, organizations can unlock the boundless potential of Databricks and drive remarkable business outcomes.

D. Continuous Learning and Improvement

A vital aspect of effectively implementing Databricks in an enterprise architecture setting is the persistent commitment to continuous learning and improvement. Given the ever-evolving nature of technology and business requirements, it becomes crucial for organizations to remain well-versed in the latest advancements and best practices. This necessitates a dedicated investment in ongoing training and development for teams that operate within the Databricks framework, a vigilant monitoring of industry trends, and an active pursuit of opportunities to bolster skills and expand knowledge. The pursuit of continuous improvement also entails the regular evaluation and optimization of the Databricks environment, with a keen focus on identifying areas that can be enhanced and subsequently implementing changes that drive efficiency, performance, and security. By embracing a culture that prioritizes continuous learning and improvement, organizations can effectively leverage the capabilities of Databricks and maintain a competitive edge in the ever-changing realm of enterprise architecture.

Conclusion

To summarize, the inclusion of Databricks as part of Enterprise Architecture is a significant stride meaningful business value. Additionally, following best practices like implementing clear governance, managing data quality, and continuous learning and improvement is indispensable for maintaining security, compliance, and operational excellence. Databricks' flexible and cost-effective solutions enable enterprises to handle diverse data formats and scale alongside their growth. As organizations continue to develop, the adoption of Databricks towards aligning an organization's business strategy with its IT infrastructure to fully capitalize on data analytics. The incorporation of Databricks necessitates a strategic approach that takes into account the needs of the organization, available resources, and scalability. By making use of the platform's robust features, such as enhanced data processing, streamlined data management, improved governance, and scalability, organizations can effectively address the challenges posed by big data and extract valuable insights. The achievement of success with Databricks relies heavily on creating a cohesive environment through collaboration between IT, business, and data teams. This collaboration guarantees that the implementation aligns with the organization's goals and generates within the framework of Enterprise Architecture provides a dynamic and powerful tool to remain competitive in the data-driven economy. By effectively utilizing Databricks' capabilities and adhering to best practices, enterprises can ensure that their architecture not only supports current data analytics needs but also adapts to future requirements.

References

[1]. E. Niemi and S. Pekkola, "The benefits of enterprise architecture in organizational transformation," Business & information systems engineering, 2020.springer.com

[2]. https://www.databricks.com/blog/2020/04/28/new-study-databricks-delivers-nearly-29-million-in-economic-benefits-and-pays-for-itself-in-less-than-six-months.html
[3]. F. Saleem and B. Fakieh, "Enterprise architecture and organizational benefits: a case study," Sustainability, 2020.mdpi.com
[4]. D. Dumitriu and M. A. M. Popescu, "Enterprise architecture framework design in IT management," Procedia Manufacturing, 2020.sciencedirect.com
[5]. Y. Gong and M. Janssen, "Roles and capabilities of enterprise architecture in big data analytics technology adoption and implementation," Journal of Theoretical and Applied Electronic, 2021.mdpi.com
[6]. M. P. Uysal and A. E. Mergen, "Smart manufacturing in intelligent digital mesh: Integration of enterprise architecture and software product line engineering," Journal of Industrial Information Integration, 2021.HTML
[7]. M. M. Alvord, F. Lu, B. Du, and C. A. Chen, "Big data fabric architecture: How big data and data management frameworks converge to bring a new generation of competitive advantage for enterprises," 2022.eapj.org
[8]. P. Ghavami, "Big data management: Data governance principles for big data analytics," 2020.HTML
[9]. M. Armbrust, T. Das, L. Sun, B. Yavuz, S. Zhu, et al., "Delta lake: high-performance ACID table storage over cloud object stores," in Proceedings of the, 2020.stanford.edu
[10]. R. K. Batwada, N. Mittal, and E. S. Pilli, "Uncovering Data Warehouse Issues and Challenges in Big Data Management," in Big Data, Machine Learning, and …, Springer, 2020.HTML
[11]. D. Oreščanin and T. Hlupić, "Data lakehouse-a novel step in analytics architecture," in 2021 44th International Convention on…, 2021.HTML
[12]. M. Armbrust, A. Ghodsi, R. Xin et al., "Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics," Proceedings of …, 2021.cmu.edu
[13]. P. Ghavami, "Big data management: Data governance principles for big data analytics," 2020.HTML
[14]. S. Cisneros-Cabrera, A. V. Michailidou, S. Sampaio, et al., "Experimenting with big data computing for scaling data quality-aware query processing," Expert Systems with Applications, Elsevier, 2021.auth.gr
[15]. K. H. Hu, M. F. Hsu, F. H. Chen et al., "Identifying the key factors of subsidiary supervision and management using an innovative hybrid architecture in a big data environment," Financial Innovation, 2021.springeropen.com