A Sneak Peek on the Data Architecture

Arun Uniyal

arunkuniyal@yahoo.com

Accenture Solutions Pvt. Ltd.

Gurugram, Haryana, India

Dr. Indrapal Singh Oberoi

Himalayan Garhwal University

Dept. of Computer Science & Applications

Abstract—

Data is of profound use in any organization. Management of data requires a strong data strategy and foundation which should drive value to the enterprise. Data Management cuts across many dimensions of data which includes architecture, acquisition, ingestion, storage, transformation, consumption, governance etc. While every dimension has its own significance in the field of data, data architecture sets a strong foundation for the enterprise to drive business decisions which can impact the bottom or top line in delivering value.

The first and foundational step to implement the data strategy is to define the right data architecture. The target state data architecture will confirm the efficiency, efficacy and worth of data for meaningful actions or outcomes for the business. According to open source architectural principle TOGAF®, a data architecture augments the architecture vision of an organization by enabling business architecture in alignment to stakeholders expectations. The data architecture hence allows the business to define their logical and/or physical data assets with associated management processes structurally.

It is one of the toughest skills to acquire. In today's emerging technology world where data is fueling the growth of businesses, stakeholders can be benefitted from basic understanding of data movement across various functions in the enterprise which can derive insights for meaningful business decisions.

This paper is an attempt to explore the data architecture principles, available options and considerations while evaluating the limitations.

Keywords — Data Architecture, Data Management, Enterprise Data Warehouse, Data Ingestion, Pipelines, Data Security, Data Consumption. Dark Data

---

I. INTRODUCTION

Data has always been at the midst of technical architecture for businesses. From the older days, where the objective was solely focused on capturing transactional data from consumers to modern age where data is driving transformation to enterprises, data has played a pivotal role in business growth. With every changing nature of data around volume, shape, size and nature, businesses have faced challenges to manage it; may it be small or large dataset. Business, Digital or Data Transformations are dependent on how the data is architected in the existing ecosystem which can drive the path to success when migrated. Data Architecture, hence, plays a very important role in the entire data supply chain, to align data strategic needs to platform realities. A well architected data platform will not only guarantee adoption by end consumers, but it also lays a foundation for successful business outcomes or decisions. Use cases deployed on mature data architecture reap value by uplifting profits, increasing sales, better consumer experience, reduced customer churn, successful product roadmap, product price sensitivity, and many other benefits depending on the implementation.

Since ages, Data Architecture, defines the way data is consumed in any enterprise. From traditional applications where the focus was centered around capturing all possible datasets; which can help in day-to-day transactions to modern applications where data is driving innovation, businesses, and transformations; data has always helped businesses to understand how they are placed against the business objectives from time to time. On the similar notes, the data architecture has matured significantly over the past which has helped to relook at how to position data in an enterprise. The progression in data architecture has repositioned data from mere bare metal to cloud computing infrastructure along with other applications. The usage is not limited to these use cases but also looking at strategic compliance and regulatory need of the organizational or non-organizational data to secure it. A right data architecture helps an organization to augment their architecture vision and enhance the business, technical and application architecture.

The core components of any data architecture are based on following key principles:

1) Data Accquisition

2) Data Storage (Commonly known as storage)

3) Data Computations (Commonly known as compute)

4) Data Consumption (Internal, external or through any other form)

There are additional considerations while designing any data architecture which complements the organizational architecture vision by ensuring continuous availability of data in any adverse situation to ensure business continuity. The principles which define the abilities of proposed data architecture will

guarantee that the business, technology or data strategy is not overseen while observing usage of data in the enterprise.

II. DATA ARCHITECTURE: IMPORTANCE AND ALIGNMENT TO OTHER ARCHITECTURAL PRINCIPLES

Data Architecture is a high-level view of an enterprise to handle data. The handling of data includes its categorization, integration, storage, etc. It is one of the initial steps in process of defining the enterprise architecture which help business to achieve its strategic goals. The right data architecture ensures timely, efficient, reliable, relevant, and healthy data to be made available to businesses people. A data architecture is important for a business in many reasons:

1) To gain better understanding, insights of data

2) Better management of data across the data supply chain

3) Provide a structure to data which is strategic and useful to drive business insights

4) Ensure the data is compliant to internal or extrnal policies and procedures

5) Gurantee data security and privacy as per local or global regulations

6) Augment business intelligence for proactive or reactive decision making

7) Setting up a foundational data platform in alignment with data or business strategy

8) Derive predictions or analytical use cases for exporatory data anlaysis otherwise insights

9) And many more strategic or business needs.

TOGAF® from open group defines four architectural domains to be encompassed in any enterprise architecture. These four architectural areas are:

1) **Business Architecture**: defines the business strategy, governance, organization, and key business processes of the organization. Business Architecture is often necessary mean to demonstrating the value of Technical Architecture work to be carried out for key stakeholders. It also helps to evaluate the ROI (return on investment) for these stakeholders to support and participate in the subsequent architectural work for the enterprise.

2) **Data Architecture**: Accroding to TOGAF® a data architecture describes the structure of an organization's logical and physical data assets and its association with data management resources. It is further elaborated as process that enables business architecture and architecture vision in a way that it addresses the key organizational or stakeholder concerns.

3) **Technical Architecture**: Technology Architecture defines the necessary components to suppost mission crtical applications. The key components of any technical architecture are underlying hardware,

software, network infrastructure components which ensures the deployment of the core for relevant applications for the strategic needs.

4) **Application Architecture**: The application architecture provides a blueprint for individual appliacations or systems which will be deployed in an organization. It further illustrates the communication or interactions between these application systems, their relationships and flow of infromation in alignment to core business processes of the organization. The application architecture clearly articulates the way applications should be deployed, secured, protocols for communication etc.

For any enterprise, it is prudent to define a data architecture which should be strategically positioned to augment its architecture vision and supports the enterprise architecture. Whereby business architecture is the first step in defining an architecture domain, the data architecture is the interconnect which guarantees successful deployment of architecture by deriving right value if done well.

III. DATA ARCHITECTURE: APPROACHES

The implementation of data architecture is dependent on many factors. There are many approaches for deployment of a data architecture which is dependent on the strategic needs of the organization. These factors which influence the way a data architecture is to be deployed in an enterprise are:

1. Architecture Vision
2. Technical Architecture: specifically, the underlying infrastructure
3. Nature of data consumptions: OLTP vs OLAP
4. Volume of data
5. Business Case
6. Data Strategy

There are many approaches for implementation of a data architecture, but the following ones will highlight the key ones which has been in use or are successful from small to large scale implementations:

1. **Transactional data focused data architecture**: The data architecture mainly focusses on capturing key transactional data to meet day to day operational needs. A visual representation of this data architecture is shown below:
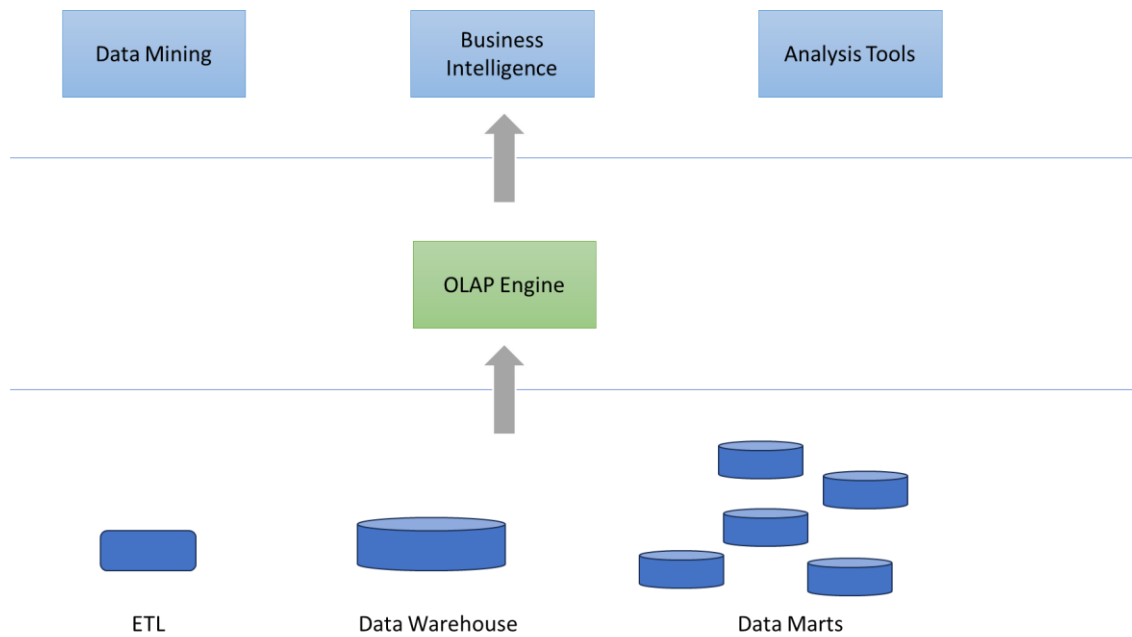
Figure 1: Traditional 3-tier architecture

2. **Analytics based data architecture**: There are specialized data architectures which are highly focused on driving analytical use cases in an enterprise. A generic representation of the same is shown below:
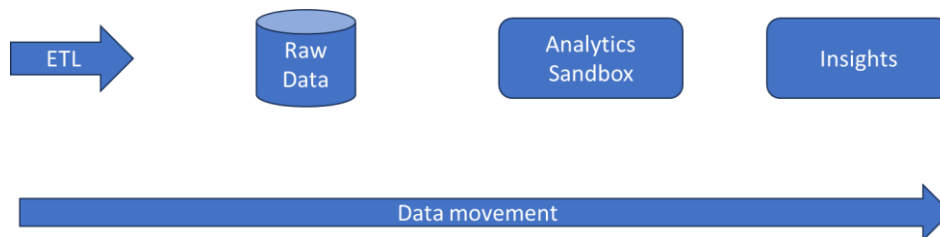

Figure 2: Analytics Sandbox based architecture

3. **Data Lake focused data architecture**: With availability of commodity-based hardware, big data came into existence and with that new data architecture was introduced which was highly focused on lake. This data lake-based architecture has helped organizations to rethink how they should be ingesting data into data platforms and explore or exploit it for insights which are not possible with traditional platforms. A basic data lake architecture is shown below:
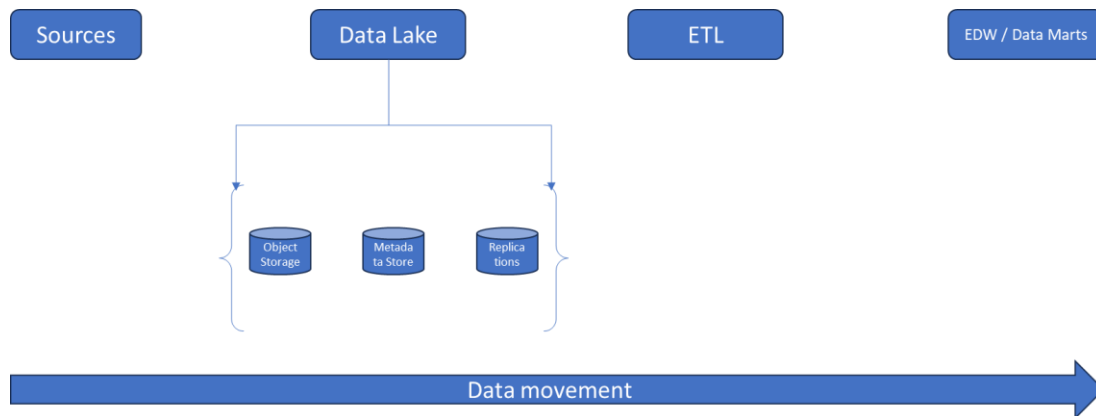
Figure 3: Data Lake based architecture

4. **Delta Lake centric data architecture:** A refined architecture which overcomes some of the challenges of data lake is centered around a snapshot of incremental data changes. This snapshot or delta build on top of existing lake provides a delta lake architecture as per following representation:
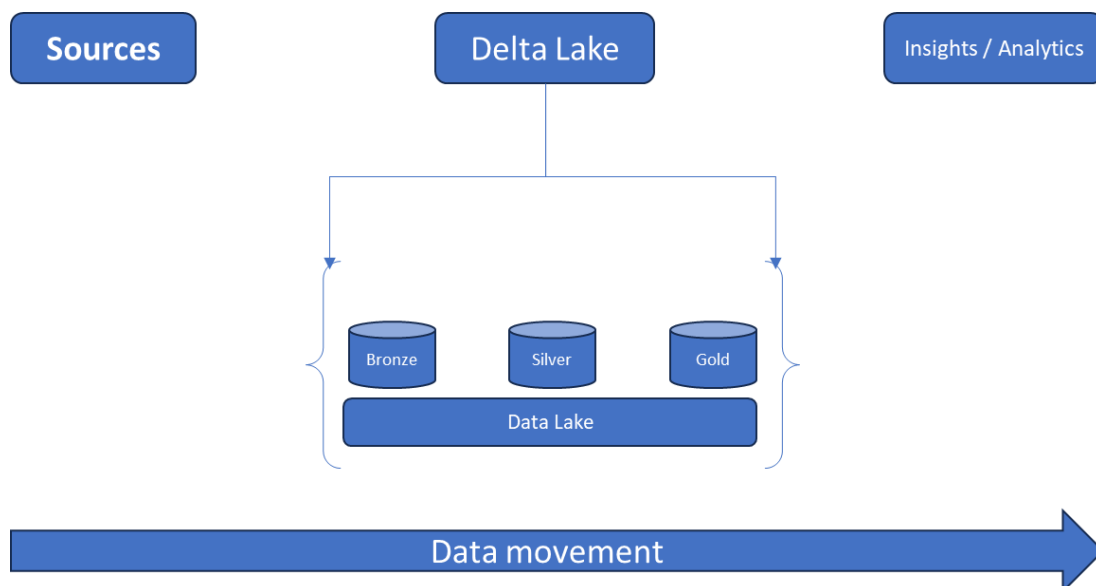

Figure 4: Delta Lake

5. **Modern data architecture**: The modern data architectures further expands the thoughts around strategic business needs. The new age data architectures are built on the thoughts of data or delta lake to provision data through mesh, data fabric or through a product centric approach. While the data architecture takes a leap in near future, monetization of well structured, healthy data through a service-based architecture will be of key importance.s

IV. DATA ARCHITECTURAL: CHALLENGES

The tremendous growth in businesses in the past has added more complexity to architectural deployments which inherently impacted the complexity of underlying data models. Data which has become pivotal in

enterprises, their survival is solely depends on a strong data architecture capability. Data Management is becoming difficult with expanding data volumes, velocity, and variety. More and more organizations want to build a data platform which can drive business transformations and hence positioning of data architecture plays an important role in defining the state of data in the organizations. Some of the common data architecture challenges faced by the industry are:

*1)***Data Silos**: Businesses are still trapped in data silos due their own inherent reasons. Traditional business were more reliant to siloed data stores to maintain privacy. The strategic data was trapped to specific databases for functional or departmental specific needs. These siloed data stores have created a lot of challenges to analytical industry by restricting the outcomes to available data from specific databases. It was hard to derive 360 degree view of data for business needs. As organizations and technology evolved and embarked on the digital journey, realizing the need for accessible and shareable data was the need of hour. With the modern data architectures which are more mature with integration of data science, Artificial Intelligence and lot many technological advancements, businesses are reforming them as data centric or data driven or data powered organizations. Hence the modern data architecture should be designed in a way to eliminate siloed data sets and use every bit of data, may it be internal or external to derive business outcomes.

*2)***Data Cleansing and Preparation***: Data cleansing and preparation is one the biggest hurdle in any organization due to either lack of infrastructure or talent. With modern data architectures and accessibility to infrastructure over cloud at a lower TCO, the businesses are rethinking their investment in this area to derive meaningful insights from their underlying data assets.

*3)***Data Strategy**: A data strategy aligned to business case and long term organizational objectives is the initial step in setting up the right data architecture. Businesses which are driving value from their data investments relies on a strong data straetgy and its alignment to strategic business or consumer needs.

*4)***Lack of Governance**: Ownership of data is one of the biggest challenge in any enterprise due to everchaging requirements or market needs. A strong governance council will emphasize the need for right people, processes, technology and literacy to uplift business value.

*5)***Total Cost of Ownership (TCO)***: Majority of the businesses are focussed on business use cases without thinking about the underlying data platform. Traditional data platforms have grown in size with increasing data volumes. The businesses themselves have grown multi-fold which are causing problems in the existing data platforms to scale and own it. There is a need to relook enterprise data architectures with the changing needs and apetite which can help to bring down costs significantly.

*6)***Lack of vision / Peer Pressure**: Many organizations are trying to setup a data ecosystem without thinking about their business needs or strategic business case due to peer pressure. This is causing lot of challenges in evolving and setting the right foundational data architecture due to lack of vision. At times businesses are experimenting with data just to reap immediate benefits bypassing undepinning principles of core architecture. This is setting their landscape for disasters.

*7)***Organizational Talent**: Lack of skills within the enterprises also plays significant role in adoption of new technologies, applications or architectures. While data architecture is a specialized skill, supply of this special skill within or in the market is very thin. This is biggest cause why organizations have a dependency to scale their data architectures.

*8)***Time to Market or Scale**: At times, the overall implementation time for data architectures spans across monhts or years. Though there is interest in setting up data ecosystems by enterprises, the ROI when related to time to implementation plays a vital role to decide whether to delay or rethink data strategy to meet immediate goals.

There are many other factors which plays a role in successful data strategy and architecture given the challenges, but capturing everything will not make sense in this paper. In summary, all the data architecture decisions are dependent on a successful/impactful data strategy, vision, TCO, data literacy, and a strong governance under data management which allows organizations to implement platforms which drives transformations.

V. DATA ARCHITECTURE: WHAT'S NEXT

The data architecture field is expanding with ever-growing data in the enterprises. Data has found its place from traditional OLTP applications to latest Cloud ecosystems. With the industry moving into age of distributed computing with introduction of blockchain, Artificial Intelligence, Cloud Computing, Machine Learning etc. the emphasis to have a strong data foundation is the need of the hour. Some of the latest developments in the field of data architectures are:

1) Data Mesh

2) Data Fabric

3) Data Products

4) Domain driven data stores

5) AutoML

6) DataOps

7) Sustainable data platforms

The industry is changing at a rapid pace. Technology shift happens every 3-5 years now compared to decades. With increasing interest into this field some of the trend's notices in this field may include:

1) Automation

2) Self-healing data stores

3) Data supply driven via AI

4) Large scale computing

5) Analytics driven business

6) System driven governance etc.

CONCLUSION

This paper is an attempt to understand the principles of data architecture, its challenges and what is in store as this field grows soon. There is a need to rethink how architecture is deployed in organizations in context of data. The current limitations for a successful data architecture are more driven by motivation or strategy which should be overcomed to reap benefits from available data assets within or outside enterprises. This will not only benefit businesses but governments to strategically position policies, processes or controls for betterment within or outside.

REFERENCES

[1] https://pages.dataiku.com/hubfs/Architecture%20Basics%20Guide%20Dataiku.pdf

[2] Humam Khalid Yaseen & Ahmed Mahdi Obaid; Big Data: Definition, Architecture & Applications; https://www.researchgate.net/publication/339419749_Big_Data_Definition_Architecture_Applications

[3] Nandani Pandey, Priyanka et al; A Review on Big Data Architecture; http://www.ijtrd.com/papers/IJTRD4253.pdf

[4] Donna Burbank, Charles Roe; Trends in Data Architecture; https://content.dataversity.net/rs/656-WMW-918/images/6_67501_TrendsinDataArchitecture_Final.pdf

[5] https://opengroup.com TOGAF® standard 9.2

[6] https://en.wikipedia.org/wiki/The_Open_Group_Architecture_Framework

[7] Rick Sherman; Business Intelligence Guidebook; https://www.sciencedirect.com/

[8] www.dataversity.com/