

Data Management In The Modern World

Arun Uniyal

Maharaja Agrasen Himalayan Garhwal University;
Associate Director (Accenture) Gurugram, Harayana, India

Nidhi W/O Arun Uniyal,
Gurugram, Haryana, India

Abstract

- i) Organizations or enterprises are gathering and organizing data for their needs to meet transactional or analytical needs. While it is one of the most challenging tasks given the everchanging needs of the firm, but it brings additional difficulties when associated with regulatory or compliance requirements in the modern world. The amount of data generated post digital disruption is humongous which further increases the burden of gathering, securing, handling, storing, and organizing business information according to its needs, and/or in compliance to any applicable compliance / regulations in respective regions.
- ii) Emerging technologies pose additional threat to data given the end users are unaware of policies which inhibits the exposure of data beyond boundaries. By the time the policies or procedures are defined, it is already too late to act on such slippages.
- iii) This paper is an attempt to introduce the possible data management challenges and how we can address them given the technology is changing at a rapid pace.

Keywords: Data Management, Data Modeling, Emerging Technology, Data Governance, Data Architecture, Data Catalog, Big Data, Data Democratization, Polyglot model

Date of Submission: 10-08-2023

Date of Acceptance: 20-08-2023

I. INTRODUCTION

The digital economy has provided additional access to data which was not possible in the past. The data world is growing exponentially since the introduction of commodity-based hardware with big data ecosystem. The speed with which data is being produced today has opened avenues to explore data in many new ways. Use of techniques to exploit data and decisions in real time, strategic changes to businesses to adapt analytics or data as their key decision processes, proactive insight-based corrections in decision making, predicting, and adapting to market drift etc. are some of the avenues which are possible through rightful use of data.

There have been significant developments in the field of data where organizations are trying to reinvent themselves by driving key business decisions keeping data at its core. Data Driven or Analytics Driven organizations driving business value are far more realistic than the past. Using the data in a manner which should drive strategic needs is top agenda for majority of the players. In fact, data democratization and breaking siloed usage has opened many new avenues and business opportunities which were untapped in past.

Data Management allows organizations or enterprises to make best use of this data though collection, protection, processing, and storage so that it can help in business decision making. The emerging and evolving technologies has helped to generate and consume data at precedential high speeds, hence providing solutions for data management remained an essential task to make sense out of this vast quantity of data. A good data management solution should ensure access to reliable, up-to-date data to drive organizational decisions.

II. DATA MANAGEMENT: TYPES

Data Management as stated earlier ensures that all aspects of data, assuming it as one of the key assets, are met in alignment to internal or external needs securely. Data Management ensures that the data is managed throughout its lifecycle in an organization. The categories or types of data management which are mandatory for any enterprise are listed below and should be taken seriously while considering any aspect in respect to enterprise data management.

1. **Data Architecture:** The data architectures represent end to end flow of data in an enterprise. The data architectures are meant to describe the sources of data, its form and storage in the data supply chain, who does what, security or management constraints till the disbursement of data to various data streams or end consumers. The data architecture helps or is an outcome of an organizational data strategy which may possibly highlight challenges in managing data in an enterprise.

2. **Data Pipelines:** ETL (Extract, Transform & Load) programs in the traditional world or data pipelines in the modern world are ways to transport data from one point to another. Data pipelines are essential to get information in the desired location once data is exposed by the source systems. With the current data platforms, the data pipelines are meant to transport variety of data at speed from sources to final location. May it be structured, unstructured or semi-structured dataset received or generated within the organization, the current big-data ecosystem is able to handle and make sense out of it.
3. **Data Cataloging:** Data in an enterprise can be used in different context or ways which makes it difficult to manage it. Data Catalog helps to capture metadata information of the information to paint a holistic picture of data by providing lowest level of details like changes, locations, quality, etc. by treating data as a core asset to organization. This further helps consumers to search right data within a common metadata repository to derive unified outcomes for business benefits.
4. **Data Security or Access Controls:** One of the most important aspects of data management in an organization is to help answer the question “Who”, “What”, “How” and “How much” data assets are accessible by various actors in an organization. Data Security & Access controls defined in the data management policies through implementation of data governance policies will help to answer all these queries, hence securing the data the way it is expected. Furthermore, there might be external factors like compliance or regulatory requirements which mandates additional security protocols to be implemented on organizational data assets, the data security pillar of data management will ensure they are adhered and audited from time to time in order to meet these requirements.
5. **Data Processing or Transformations:** Data sourced from various systems may not be useful to consumers unless it is transformed or converted into a form which can derive some value to the business. Data Processing or data transformation is a process which helps to convert raw data into usable KPIs or information. There are many ways to convert raw data into useful KPIs that may include use of automated tools, scripting languages, ETL tools, COTS product etc. which not only transforms this data but also automate execution of data pipelines through native capabilities or via integration with third party tools.
6. **Data Modeling:** Data Modeling is a process to architect a storage to keep specific information of the organization. With increasing and emerging data in the real world, the data models are not limited to store structured data but expands to help store unstructured or semi-structured data through use of polyglot models. There are systems available in the market which can store data in any form may it be structured, semi-structured or unstructured in nature. They further help to exploit this data for meaningful insights.
7. **Data Quality & Cleansing:** Also known as data preparation, data quality and cleansing exercise ensures that any inconsistencies or errors in data are rectified before use. This at times can be contextual based on use cases or generic to ensure health of data in an enterprise.
8. **Governing the Data:** Data Governance is that aspect of data which guarantees rightful use of data if implemented well. The core principles on which any data governance body relies are:
 - a. **Process:** The processes or standards which should be followed based on identity of data in an enterprise. This also cuts across the security of data assets in the organization.
 - b. **People:** Identifying right resources for consumption of data with their responsibilities, and level of information which is best according to their role in an enterprise. This also help to answer few important aspects of data assets like what information, who can access it, to what level and how.
 - c. **Technology:** Selection of right tools or technologies which can help to govern the data through integration with various other technologies available in the enterprise.
 - d. **Value:** Any data governance process is redundant unless it helps to deliver positive value to business through adoption of right process, standards, controls and tools.
9. **Analytical Sandbox or Models:** Any organization can't derive significant value through basic consumption of data. Through meaningful implementation of analytical use cases, business can be benefitted if they are aligned to data strategy or business architecture through use of right data. Analytical Sandbox & Models help to set the right architecture to leverage data which can be exploited and then can be modeled to generate insights which can help business to unfold many unknown areas.
10. **Data Consumption:** Data processing, acquisition or preparation can only help to convert the data into right shape, but consumption of data through meaningful resources may it be human or technological, will continually track the progress. The insights, KPI reporting or past trends can be analyzed, and corrective actions can be taken if the businesses are enabled with right consumption tools.

III. DATA MANAGEMENT: IMPORTANCE & ADVANTAGES

- **Importance:** Management of data will be the initial step for effective data analysis which can generate important data insights in-turn adding value to customers & deliver business outcomes. If data is managed effectively, the data trust in the organization will improve significantly. This in-turn will guarantee secure data access. Some of the inherent benefits of successful data management are:

- **Maintainability:** The data management guarantees optimal management of data if implemented well. In the world, where data is growing exponentially and enterprises are relying heavily on data driven decisions, it is important that the total cost of ownership should be kept optimal while maintain the data ecosystem. The data management processes help to keep an eye on the investments and can flag anomalies during the data operations.
- **Veracity or Data Trust:** Effective data management in an organization helps to build a reliable ecosystem. The potential data access is trusted given the errors in minimized by implementing standards, processes, and policies which can be automated via tools. This processes further helps the organizations to respond to changes immediately, due to availability of up-to-date trusted data.
- **Visible Unified Data Assets:** Data Management can help to improve visibility of data assets across the enterprise which in turn can help to leverage use of common data repository for key assets uniformly. This improves the ease of access to end consumers quickly, and confidently for analyzing data. The productively will be uplifted by proper management of data due to access to data which is of importance to employees.
- **Data Security:** One of the important dimensions of data management is securing organizational data assets. The data management processes help to protect enterprises from data losses, threats or breaches with controls which inhibits external parties to use this data without proper authorization. The vital information of an organization is accessible and retrievable in adverse situations. Critical data assets like PII or PCI data, should further be managed in accordance with regulatory or compliance requirements.
- **Scalability:** Successful implementation of Data management enables organizations to effectively scale data platform. The processes can be repeated to keep latest version of data and metadata. This in turn will help to reduce operational overheads by repeatable nature of processes, and hence reduce the manpower to generate outcomes.

- **Advantages:**
 - **Regulatory compliance:** Every industry has their own compliance and regulatory requirements. At times, management of data plays a bug role avoiding fins from these central authorities by proper management and handling of data. Sensitive data i.e., gender sensitive, patient records, credit card information, personal data etc. are some of the examples which require special handing. In the recent times, many enterprises faced penalties from authorities due to mishandling of sensitive data.
 - **Protection of data assets:** Data management enables organizations to implement right controls and backup policies which have proved to be of key significance while handling data. Right controls on the one side enables organizations to implement role-based access control depending on the nature of individual responsibilities, on the other hand any adverse data loss can be prevented dur to strategic processes which allow high availability and backup of critical data as per agreed standards. The purpose of data is known to enterprise by enabling enterprises through data literacy programs and / or data catalog, the mishandling of data takes a significant drop even if the reasons are unknown.
 - **Uniformity in data assets & sharing:** Implementation of right controls, processes and standards in an organization allows uniformity of information for data assets. This uniformity in turns guarantees repeatable outcomes for same analysis by different stakeholders hence improving the adoption of data management processes. If implemented across the organization, the data assets can be shared across the organization to enable better decision making and informed decision making. The overall cost for data operations can be optimized by proper implementation of these processes, controls, or standards.

Some of the other benefits of data management includes:

- Eradicate data redundancy.
- Create real-time data consistency.
- Improved data strategy
- Increased data adoption across the enterprise
- Better decision making through data or analytics driven processes.
- Data Democratization
- Improved governance of data
- Better audit-controls

IV. DATA MANAGEMENT CHALLENGES

Organizations and enterprises are investing significant amount and time to develop reliable systems to support their decision making. This will help them to make timely business decisions and deliver business value. To accomplish business outcomes and drive business intelligence, frequent confrontation with challenges while working with data are described below:

1. **Volume of data:** Tremendous amount of data is being generated everyday by enterprises. To obtain right data, maintaining it, generating value from data became a challenge to most of these enterprises. As more and more data is being collected, it needs increased monitoring and validation. The management of entire data lifecycle becomes more challenging. If the data is explored, managed, and governed well, enterprises may use it to gain better understanding depending on the nature of data. This in turn will allow them to take informed decisions, enhance business processes & optimize offered services. Large businesses may have many business solutions, but each solution stores data in its own data repository. Evaluating and managing large data storage is a one of the biggest obstacles in data management.
2. **Variety of data storages:** Adoption of technology has helped organizations to look beyond traditional ways to store data. While majority of the businesses were focused to store only structured datasets in the past, big data technology has enabled enterprises to experiment with unstructured as well as semi-structured datasets. Storing such a variety of data was a barrier in the past which resulted in siloed data stores which were un-exploited to derive any value out of it. Dark data stores were part and parcel of data journeys, and it continues to challenge enterprises today as well to make best use of available data within any enterprise. While there are available platforms available in the market which can help to store significant volume and variety of data, whether entire dataset will find its place in such a unified data platform is purely dependent on an enterprises data strategy and how the data is being used in the enterprise.
3. **Data Health:** One of the biggest obstacles in deriving value from data is its health. May it be quality, quantity or nature which can help businesses to define use cases as per strategic needs. Data accuracy and removing noise from data is one of the most challenging tasks while creating any useful information or deriving insights from it. At times, similar data is stored in various form, shape or formats which require a lot of effort to convert or standardize it in a format which can be utilized for business purposes.
4. **Lack of data governance:** Like data management, data governance is one of the preliminary steps for any organization while defining a data strategy to adequately governing all data assets in the organization. The governance will define what is important and what not. Who should have access to what data assets to what level. If someone has access to data assets than what actions are allowed based on specific role, responsibilities, or alignment of an individual in respect to data asset in question. In simpler terms, the required processes, policies, standards, tools and associated business outcomes which the underlying data asset can tap are unknown in absence of adequate data governance.
5. **Data Synchronization and integration:** The main goal of any data management process is to ensure that required data is available in right, form, shape, and format for end consumption across the organization as per defined data strategy. Integration of data from various sources and synchronizing unified or decentralized data stores is one of the most important aspects of data management. Unless defined in the data strategy, all data assets wherever they reside across the organization should reflect the same health, or metadata always. This is becoming a nightmare with increasing data volume or ingestion rate with limited compute or stores available in an organization.
6. **Data Strategy:** Any business should have a well-defined, detailed data strategy available to manage or govern their data assets in alignment to technical, business or application architectures. It is observed that organizations lacking strong data governance fail to meet the business objectives due to untapped potential of data driven decisions. A data strategy will not only define the critical data elements for the organization but will also lay down a plan to use this data for strategic business purposes.
7. **Security:** Data has become very important asset to organizations in the modern age. While the data strategy and governance provide sufficient controls to secure all organizational data assets, sensitive information which could harm a company and/or individuals should be managed in accordance with severity or sensitivity. Utmost care should be taken while storing, transporting, or consuming data to ensure it is not exposed to unwanted parties in a form or shape which could cause any threat to respective parties. As integration with third party external system comes under the lens of regulations, point-to-point or system to system integrations should be implemented in adherence to applicable local or country laws to avoid any penalties. Wherever possible encryption or masking of data will help to secure the information for any misuse.
8. **Automation & Data Ops:** Manual management of data with increasing volume or velocity will require massive manpower. For enterprises where this volume is humongous, it will be merely impossible to manage or govern the data with human resources. Need of automated tools which can help to govern or automate this human effort will not only reduce the manpower but the cost of ownership end to end. Data Ops, a new standard helps to look at automating the process of maintenance of data assets which are crucial to any organization.
9. **Data Analysis:** Any data in its raw format can't deliver any value to organization even if it is of best quality. There are tools available in the market which can help to exploit the data through exploratory data analysis. The industry is also offering technology which is useful in analysis and extraction of large amounts of data

but there are still challenges around adequate and proper use of these tools, extracting of right data, building the pipeline which should use minimal resources etc. While the industry will emerge, these tools will mature and may offer these capabilities out of the box to solve for them.

10. **Data Observability:** Data observability refers to an organization's comprehensive understanding of the health and performance of the data within their systems. Data observability tools employ automated monitoring, root cause analysis, data lineage, and data health insights to proactively detect, resolve, and prevent data anomalies.
11. **Organizational Talent:** The current market is short of data management professionals. The available talent is being paid higher perks because they are vital to companies. They help to maintain data assets which are critical to internal or external use. These experts further help businesses to comply to regulatory or compliance requirements, which otherwise may not be possible due to large data volumes in the enterprise. There is an immediate need to uplift the talent to help manage these data assets and need significant effort while maintaining existing data platforms.

V. DATA MANAGEMENT: WHAT IS IN STORE IN THE FUTURE?

The data is available on all kinds of platforms today. Starting from traditional datacenters to on-premises private cloud base, hybrid infrastructure to pure cloud ecosystems. The data is being generated in all forms and shapes and should be managed, moved across systems securely. At the same time, the data should be made available for consumption to everyone in an enterprise depending on need or as defined in the data governance or management process in alignment to strategy. As per one of the latest industry report the data industry will grow to ~200 Zettabytes by 2025 where data will be sourced from all kinds of systems like IoT (Internet of things) devices, social media feeds, manufacturing systems, traditional systems, etc. The potential exploitation of unstructured data will bring new nuances to business decisions for mission critical applications. To keep pace with these ever-increasing enormous demands of managing high velocity voluminous and varied data every day, automation tools must be adopted in standard data management processes. Newer technologies like data observability will grow in importance, especially when citizen user and / or localized data use expand with time. These factors call for careful consideration of an enterprises data management roadmap in accordance with emerging trends.

- 1) **Emerging Data Management Framework beyond traditional systems:** With businesses adopting hybrid, multi-cloud or multi-tenant systems, there is a need of new data management framework which should be able to address growing need of data across the platforms. Siloed processes in one or independent systems will only increase the challenge of data management and finding an interlock with the business & data architecture will become difficult with increasing data footprints.
- 2) **Data Observability:** Applications now use multi-cloud, hybrid and / or on-premises platforms to access and process data. Data Observability is an ability of the system to track data and events across multiple platform. It is one of the key focus for enterprises looking to monitor end-to-end movement of data across applications. The issue with organizations that have adopted observability tools is that either they are using too many tools to monitor end-to-end data movement. There are OEM vendors which are recognizing this problem and are focused on delivering integrated, observability tool set. These tools enable enterprises to reduce the footprint of observability tools being used in an enterprise into a single toolset that's able to monitor data and its movements across multiple platforms.
- 3) **Upkeeping the legac applications while moving to new:** The businesses are moving to new platforms and systems with maturing technology market. Biggest hurdle in this shift is to maintain and manage traditional systems alongside new platforms which are hard to replace due to internal or external factors. There is a need to offer capabilities within these tools which can help to observe traditional systems with these new age tech to deliver required outcomes.
- 4) **Use of AI/ML for management of data:** The use of matured algorithms in the area of AI/ML can offer a lot of benefits to businesses. The organizations are using these algorithms since a long time but with maturing technology and advanced tools, the need to use human resources can be reduced by implementing AI/ML programs which can offer potential uplift to many data management processes. The AI/ML programs can help in many ways like:
 - a) Data Catalogs
 - b) Metadata Management processes
 - c) Ingestion Frameworks
 - d) Data Pipelines and data mappings
 - e) Anamoly detection
 - f) Data Discovery
 - g) Data Operations
 - h) Observability

- i) Data Governance & Controls
 - j) Auditability
 - k) Data encryption & masking
 - l) Automation / Orchestration program / processes etc.
- 5) **Data Preparation:** Businesses spend most of their time to prepare data for analytical or AI/ML processes to drive outcomes or insights through data science professionals. The limited pool of available talent pool is the biggest hurdle with growing data ecosystems for enterprises. Mature tools which can help to explore data and suggest recommendations will help to identify key health parameters in the available data pool which will reduce overall time to market for implementation of identified use cases.

VI. CONCLUSION

This paper is an attempt to understand the “*data management principles and challenges in the real world*”. This is just an attempt to cover major categories, potential challenges, and future of data management due to ever changing technological advancements. The need to implement data management principles across various data platforms is the need of the hour. The extension of data management to address common data challenges, exponential growth of data, real time data ingestion at scale, architectural considerations are some of the factors which are necessary for strategic data needs. Thus data management is the first and one of the important aspects of data strategy to categorize data as an asset and manage it to drive business outcomes.

REFERENCES

- [1] Devin Partida, Data Management: Types And Challenges At <https://www.datamation.com/big-data/data-management-types-and-challenges/#:~:Text=Data%20Management%20Challenges&Text=A%202022%20Deloitte%20study%20revealed,Technologies%20can%20present%20another%20challenge>.
- [2] <https://www.tableau.com/learn/articles/what-is-data-management>
- [3] Mary Shacklett, 7 Data Management Trends: The Future Of Data Management, <https://www.datamation.com/trends/data-management-trends/>
- [4] Maxim Drobot, How To Mitigate The Challenges Of Data Growth, <https://www.red-gate.com/blog/database-development/how-to-mitigate-the-challenges-of-data-growth>
- [5] Lisa Swartz, What Is Data Management? Benefits, Types And Best Practices, <https://www.netsuite.com/portal/resource/articles/erp/data-management-benefits.shtml>
- [6] Haisam A Malak, 11 Data Management Challenges And Solutions, <https://theecmconsultant.com/data-management-challenges/>
- [7] 8 Data Management Challenges And How To Solve Them, <https://www.piloggroup.com/data-management-challenges.php>
- [8] Nick Ismail, Data Management Challenges And Opportunities In 2018, <https://www.information-age.com/data-management-challenges-opportunities-9260/>
- [9] Nick Ismail, 5 Technology Trends Transforming Data Management, <https://www.information-age.com/5-technology-trends-transforming-data-management-10488/>
- [10] Peter Radanliev, New And Emerging Forms Of Data And Technologies: Literature And Bibliometric Review, <https://link.springer.com/article/10.1007/S11042-022-13451-5>
- [11] Garath Macdonald, Data Management In The Digital Age, <https://www.genengnews.com/topics/bioprocessing/data-management-in-the-digital-age/>
- [12] Vivek Singh, New-Age Tech And Culture Driving Data Democratization, <https://devops.com/new-age-tech-and-culture-driving-data-democratization/>
- [13] Data Management Architectures, <https://www.gavstech.com/data-management-architectures/>
- [14] Box Communications, The Importance Of Data Management In Digital Transformation, <https://blog.box.com/importance-of-data-management-in-digital-transformation>
- [15] Wipro Technologies, <https://www.wipro.com/business-process/why-data-management-and-data-governance-matter/>
- [16] Craig Stedman, What Is Data Management And Why Is It Important?, <https://www.techtarget.com/searchdatamanagement/definition/data-management>
- [17] Kevin Bartley, Data Management 2021, <https://rivery.io/blog/data-management-2021-trends-technology-that-will-define-the-year/>
- [18] Andrzej Szajna Et Al, The Management Of Digital Data Using Innovative Technologies, <https://www.sciencedirect.com/>
- [19] Eran Levy, Streaming Data Architecture In 2022, <https://www.upsolver.com/blog/streaming-data-architecture-key-components>