# Top **5** Constraints with
On Premise Data Management
and how to overcome them...

Data has been proliferating the enterprise for years. More recently, with the advancement of Machine Learning, Artificial Intelligence and other automation technologies, there is more data being generated across the enterprise and it continues to grow. Data is useful when it can be harnessed and converted to useful insights, otherwise you will end up with dated, confusing and expensive management systems that don't help you accomplish the business outcomes.

According to the IDC, the sum of the world's data – the DataSphere — will grow from 33 zettabytes in 2018 to a mind-boggling 175ZB by 2025. The robust 61% compound annual growth rate for data.

Annual Size of the Global Datasphere
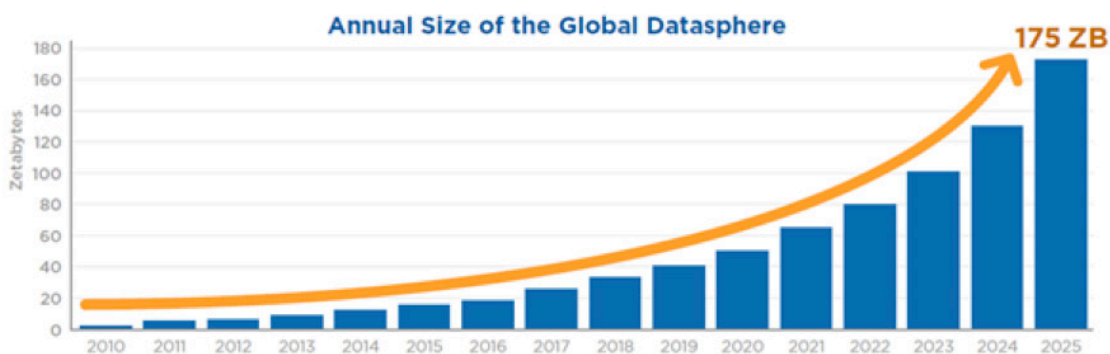
IDC mentioned - Organizations will need to embrace their role as data guardians, leverage the cloud, and take a global approach to their data. Different industries have different levels of data maturity, so companies should review where they stand relative to their industry index and what they need to do to not just survive – but more importantly to thrive – in their own Datasphere.

In this document, we attempt to highlight the common constraints most enterprises experience managing data with on premise infrastructure. This can be a combination of repositories including RDBMS, Data Marts, Application databases, offline repositories etc. In addition, we also provide some thoughts on what can be done to leverage a data cloud strategy to overcome some of these challenges and focus on the business outcomes by extracting intelligence from the vast amounts of data being generated across the enterprise.

1 Single source of truth

With data everywhere in an enterprise, getting access to the right data at the right time is one of the most critical challenges faced by data analysts and data engineers. It is highly likely that your organization has a massive number of redundant databases, multiple repositories with different views of similar data and multiple transaction systems with varied level of details. There is probably no single source that consolidates, rationalizes and optimizes access to entity related data. Many organizations, attempt to solve this issue by launching a cloud transformation initiative but that does not resolve the core issue of how the data is accessed differently by applications and users. In fact, a rapid cloud transformation would likely exacerbate the issue by introducing another layer of complexity. Prior to starting a cloud transformation initiative, enterprises should first rationalize their data management strategy. Identify the sources for data and design a schema to establish how the right data is transformed to a cloud infrastructure.

2 Ubiquitous access to promote sharing and collaboration

There is an increased need to democratize data among employees, customers, partners and prospective clients. Today enterprises are challenged with this due to several reasons including restrictions due data silos, security risks associated with relinquishing control, proliferation of multiple data sets and data marts that are dispersed across a few systems and infrastructure. Enterprises should strive to open their data into a real time data access framework to enable their customers, business partners and prospects. The ability for a customer to access real time data on a product support issue which can also give them visibility into the providers supply chain will result in superior service levels which in turn will be reflected in positive customer retention metrics and NPS data. Seamless collaboration and real time access for intra enterprise and inter-enterprise data are a couple of constraints faced by most enterprises.

NEXTPHASE.ai

3   Infrastructure focused vs Data Focused

With the growth in data across the business, enterprises are forced to invest heavily in maintaining the different elements of the technology stack for their on-premises infrastructure. As data volume continues to grow, enterprise IT spends the bulk of their time on re-sizing of the various tech elements including compute, databases, and storage. Infrastructure only grows and there is a constant guessing game of how much technology to buy based on projected data growth. The answer is it is never enough, and the technology vendors continue to feed the complexity by introducing new features and software updates to drive the footprint. The larger the technology footprint, the higher the number of monitoring tools, ITSM systems and processes are required to handle troubleshooting as well as advanced technology skills to maintain and manage it. Unfortunately, data management and most importantly the insights that can be delivered by the data is at the tail end of this process. The demands from data users are subject to critical path elements around infrastructure readiness vs the ability to prioritize data access, democratization and rationalization. For a data driven environment we currently live in it is critical to inverse this focus and drive decision making based on what needs to be achieved from the data to be analyzed.

4   Insight: Time to access and Total Cost of Ownership

Developing insight from enterprise data using a traditional dedicated infrastructure approach is highly dependent on several technology barriers. Having the infrastructure ready to capture and store the data, establishing all the integration routines and technologies such as ETL, FTP, API require detail planning and design. This is followed by continuous testing to ensure the appropriate routines to extract the data from various sources are accurately configured. Since the data recipient can't be subject to delayed data or redundant data due to analysis risk, you need to ensure that timely updates from source systems are accurately reflected so that the data repositories are relevant for the analysis to be performed. Beyond the challenges with maintaining a complex web of integration from a technology and functional perspective, the costs associated with a dedicated data repository infrastructure can be significant. The infrastructure elements for database storage and temporary storage allocations, the compute power and memory allocations as well as the middleware software required to support the integration layer to transfer the data, require upfront capacity planning, procurement and configuration. As the data volumes grow, the configuration and technology spend needs to stay ahead of the data growth curve. The rigidness of this infrastructure does not allow for spending flex when or if the infrastructure is not being utilized, hence the scaling structure tends to be only upward. The advanced capabilities offered by cloud data solutions today enable a more seamless way to access intelligent repositories as well as provide consumption models that are aligned with infrastructure utilization to optimize the economics.

5   Security and Governance

Most enterprises have disparate data that is stored across multiple repositories that are not all governed under the same security infrastructure. This can result in challenges with enforcing controls to protect against privacy rules and data breaches. Especially with the current environment of remote working, global users and widespread connectivity of the end user community, threats from deep faking and phishing are prevalent across the enterprise. Establishing a governance policy is one aspect of managing this infrastructure but enforcing that governance with the right criteria that does not limit the usability and value of the data is a critical success factor. With data infrastructure that is self-managed, enterprises need to plan and maintain a security framework that handles both the preventative and detective aspects of control. This includes investments in threat intelligent detection technologies, processes to monitor and alert potential breaches and the advanced skills required to enforce these security controls. Alternative options as we progress into a data governed infrastructure, enterprises should be considering the ability to implement processes that govern with revocable access, enable data sharing across multiple repositories under the same security policy management framework with standards. Cloud data solutions are designed to inherently handle GDPR, PII, CCPA and other complex compliance issues using a singular management framework.
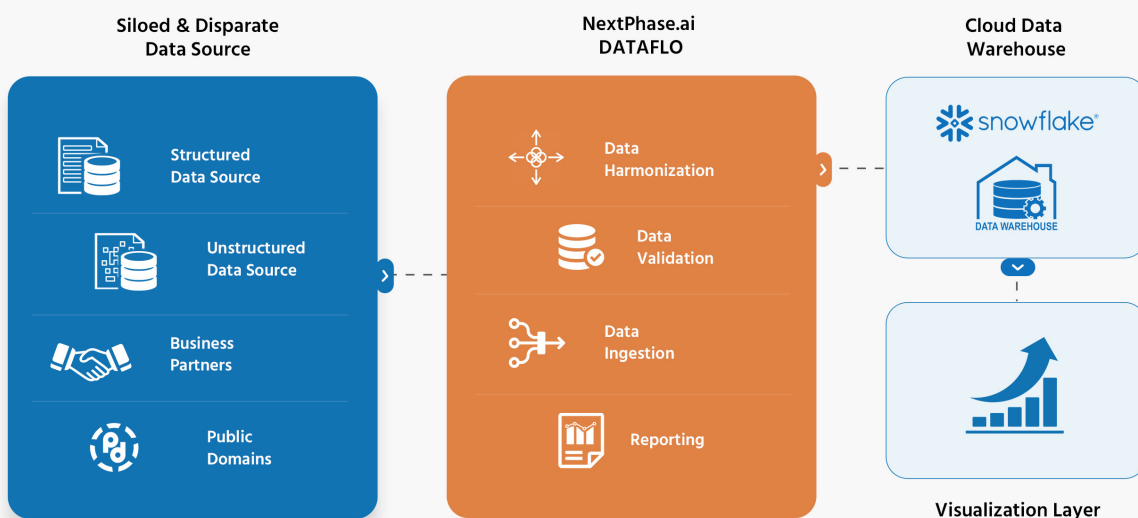
Conclusion

As you embark on your data management journey, consider starting with a Data Rationalization process where you identify the various sources of data across the enterprise. These should be organized using entity relationship mapping into logical repositories that enable a data harmonization effort including de-duplication, data validation and restructuring for extraction vs transaction processing. Before determining the future state for your data management, evaluate options for data cloud, understand Total Cost of Ownership, assess implications of global data security and the ability to seamlessly share data across your constituents to improve business actions for customer experience, user experience and employee experience. Leveraging a methodology to help you navigate the data management puzzle is more likely to result in attaining your expected outcomes.

About NextPhase.ai

Nextphase.ai is a data cloud services provider specializing in Snowflake, cloud data management and analytics technologies. We accelerate enterprise digital transformation initiatives by leveraging our innovative cloud data management technology, "NextPhase.ai DATAFLO" to optimize and rationalize disparate enterprise data into relevant insights. "NextPhase.ai DATAFLO" is designed to automate the lifecycle of data management transformation using AI and ML along with expeditious on-ramps to the Snowflake data cloud infrastructure. Nextphase.ai provides a range of technology consulting services for the Financial Services, Biotech and Technology industry sectors combining our platform-based services, seasoned talent, and industry proven methodology so our customers can harness more from their data. We are a Silicon Valley based company with global presence having delivered high value service engagements for numerous Global 2000 enterprises. Visit nextphase.ai

Siloed & Disparate Data Source

- Structured Data Source
- Unstructured Data Source
- Business Partners
- Public Domains

NextPhase.ai DATAFLO

- Data Harmonization
- Data Validation
- Data Ingestion
- Reporting

Cloud Data Warehouse

snowflake
DATA WAREHOUSE

Visualization Layer

NextPhase.ai
1710 S Amphlett Blvd #200
San Mateo, CA 94402
(888) 812-6087
hello@nextphase.ai

NEXTPHASE.ai