

# Data Lakes: Driving the Digital Transformation Journey



# Introduction

## Importance of Robust Data Management in Digital Transformation

Digital transformation is ushering in a new wave for businesses that are aiming to gain a position in today's competitive landscape. But unlike general belief, digital transformation isn't just about implementing advanced tools and technologies for enhancing business operations. It marks a radical rethinking of how an organization uses technology, people and processes to fundamentally change business performance.

So, how an organization can ensure that its digital transformation initiative is successful?

One of the solutions is to have an effective data management strategy in place as what an organization does with its critical data determines how it operates and delivers value to its customers. To accomplish this, organizations should be able to access, trust and use its data at all times. This puts data management at the core of digital transformation.

Businesses today need to develop a powerful data architecture that has the potential to transform by providing a single repository of all the organization's data, and leveraging the economics of big data in a consistent way with algorithms to drive real-time analytics.

To achieve this, most of the organizations opt for a Data lake.

An abstract graphic featuring concentric circles and radial lines in various shades of blue and white, creating a sense of depth and data flow. The lines are of varying thickness and opacity, some solid and some dashed, giving it a complex, layered appearance.

## Data Lakes:

# Data Management Strategies for an Efficient Digital Transformation

An essential component of an organization's digital transformation journey is agile data management strategy. When aligned with the company's transformation efforts, the data management strategy holds high significance and it acts as a key component for developing a roadmap for successful digital transformation journey. An organization should focus on the below areas to ensure a successful and effective data management strategy:

## 1 Availability of Data

If data is not available at the time it is needed the most, then that data turns insignificant.

For a company, this could not only lead to financial losses but also affect the reputation and trust factor of the organisation. Data availability, in this digitally driven economy should be intertwined to database systems in a way that it promises an all-time data disposal facility, notwithstanding the IT infrastructure crisis whatsoever. Organizations valuing integrity should use technology that values availability of data.



## 2 Focus on Data Integration

Data integration is the process of combining data from disparate sources using data pipelines and a variety of integrations so that users can query and manipulate it as needed.

As organizations aim to create single source of truth, the demand for data integration will surely witness a steady rise. This will make the data more comprehensive and usable for solving complex business problems and making smarter business decisions



### 3 Utilize both Structured & Unstructured Data

Organizations that wish to make the most of their data use tools & technologies that are capable of utilizing the benefits of both structured and unstructured data. Insights gained from unstructured data can be used to support the business decisions. By analyzing social media content such as tweets and Facebook posts, companies can discover the trends and behavior of customers towards their products/services. This will help them to understand the customer's viewpoint and take corrective measures in order to enhance customer satisfaction. Combining both structured and unstructured data stores enable organizations to acquire complete intelligence, leading to exponentially more powerful insights.



### 4 Ensure your Data is Easily Accessible

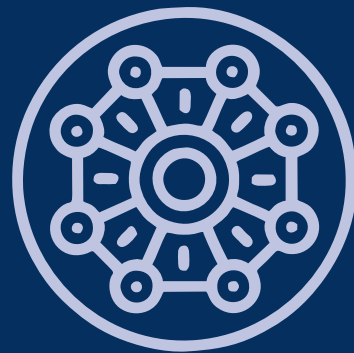
While proper data storage is essential, it is also important for organizations to provide ease-of-access to those who need to use it. The data should be stored in such a way that people can quickly find and use the information across multiple devices and systems. However, organizations should also take care that it does not create security concerns. Right data should be accessible only to the right users while denying access to those who doesn't have the access rights.





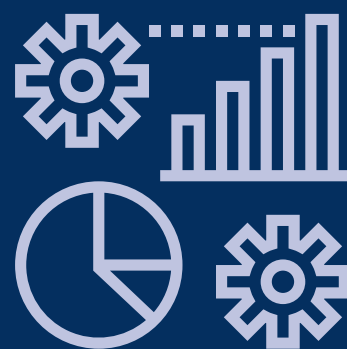
## 5 Scalability of Data Systems

In the current business scenario, data collected is increasing manifold every second. This presses the need for a system that could handle this growing amount of work, or even better, has the potential to perform additional work in the same elapsed time. For this an increased storage capacity and better processing power is needed. An effective database system should provide unlimited scalability as and when required. Scalability for a company should be linked to its growth and should happen only at the precise moment the organization feels the need for it. If not done, it would increase costs due the additional resources.



## 6 Implement Data Governance

Data Governance is driven by a governing council which has defined a set of procedures around the overall management of data while orchestrating people, processes, and technology to ensure proper data effectiveness. It includes policy and rules for data quality management, master data management, data availability & usability, data integrity, data security, and compliance. Data strategy and governance should be closely aligned to efficiently manage data for delivering timely, trustworthy and relevant information to make informed business decisions.



# Why Data Lake?

Implementing the right data platform is crucial to shape the foundation of data transformation within an enterprise. Out of all existing data platforms, Data Warehouse still continues to be the most preferred platform to provide organizations with cleansed, organized and governed data for business purposes. It is a highly transformed and structured platform that stores current and historical data which is used by business executives to make business decisions. However, certain challenges associated with Data Warehouse, such as requirement of the data to be structured before analysis has led to the advent of centralized data repository that provides quick access to ever-changing types, volume and velocity of data i.e. Data Lake.

## What is a Data Lake?

Data Lake is a centralized repository that allows an organization to store structured as well as unstructured data at any scale in its native format, process and analyze to derive new business insight. It helps in transforming information management into a proactive and real-time practice by enabling organizations to swiftly react when new business challenges are identified as they can make use of data throughout its entire life cycle, thus drastically decreasing the time to insight creation.

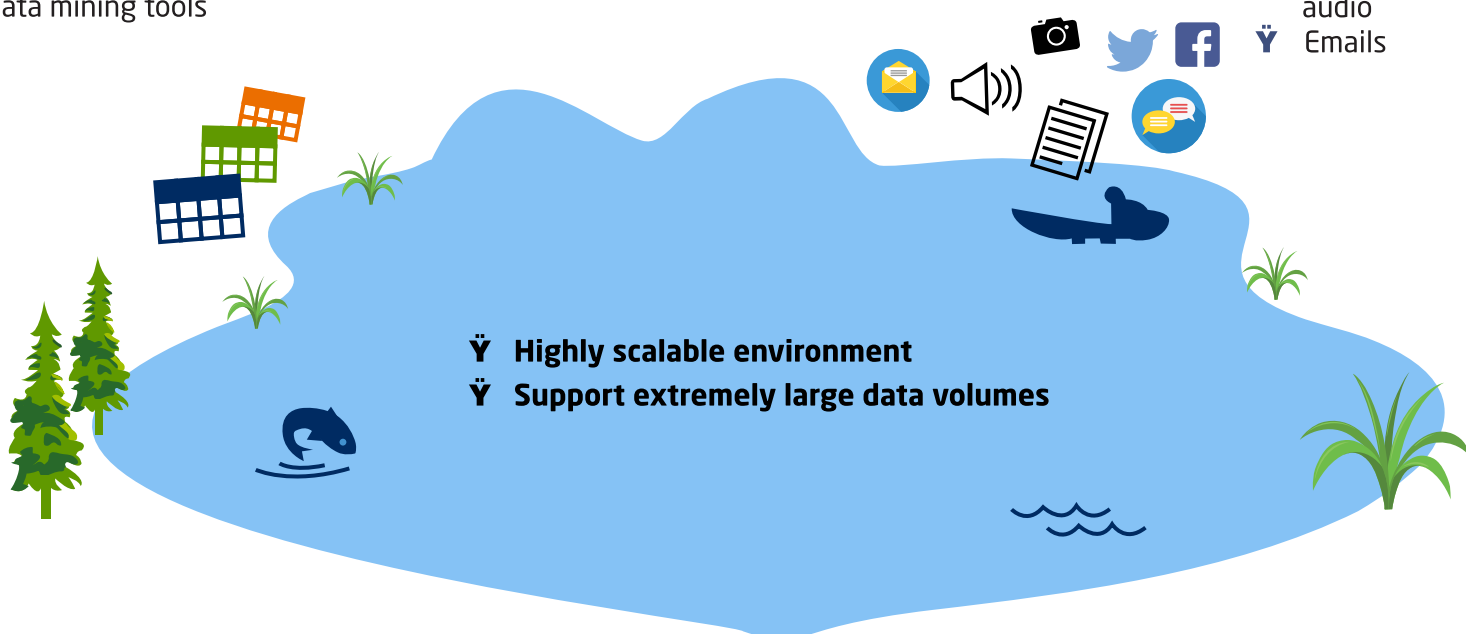
### Accept data in its native format from a wide variety of data sources

#### STRUCTURED DATA

- Information in rows and columns
- Easily ordered and processed with data mining tools

#### UNSTRUCTURED DATA

- Social media tools
- Raw, unorganized data
- Images, video and audio
- Emails





# How a Data Lake Differs from a Data Warehouse

	Data Lake	Data Warehouse
Data Types	Structured, Semi-structured, Unstructured, Raw	Structured, Cleaned, Processed
Data Capture	Captures all forms of data for future usage	Captures only structured data and organizes them in schemas
Purpose	Undefined, can be used for big data analytics, ML, predictive analytics and so on	Specific pre-defined purpose, mostly used for BI
Data Source	Native raw data from any source including SNS, IOT, ERP/CRM systems and so on	Structured data (historical and relational) typically extracted from transactional systems, operational databases and applications
Data Quality	Raw data that may or may not be curated for use	Centralized curated data which is ready for use in BI and analytics
Data Retention	Retains all the data for an unlimited amount of time	Does not retain data forever, data is purged periodically
Data Processing	Schema-on-read, raw data only transformed when put to use	Schema-on-write, structured and cleansed data
Agility	Highly agile, can be configured and reconfigured as and when required	Less Agile, Pre-configured
Users	Data Scientists and Data Engineers	Analysts and Business Users
Storage Costs	Relatively Less Expensive	Expensive and Time-consuming
Top Use Cases	Building data pipelines, stream processing, machine learning, real-time analysis	Batch processing, BI/reporting
Best Suited for	Exploration, Discovering Patterns, Innovation and Flexibility	Repeatable Processes, Ongoing Analyses, Constant Operational Use such as generating business reports and dashboards

“As per a survey conducted by Aberdeen, the organizations that implemented data lakes outperformed similar companies by almost 9% in organic revenue growth.”<sup>1</sup>

Data lake enables organizations to run new form of analytics such as machine learning over log files, social media, and internet-connected devices, thus enabling them to identify new opportunities, enhance customer retention by uncovering their preferences, make meaningful business decisions and ensure growth.

An isometric illustration of a data lake analytics environment. It features several floating rectangular panels displaying various data visualizations: a world map with regional highlights, a donut chart, a line graph with multiple data series, and a table of data. Below these panels, two stylized human figures are shown interacting with the data. The entire scene is set against a light blue background with a network of lines and nodes, suggesting a complex data ecosystem.

# Benefits of Data Lake:

## Simplified Data Access

All the data can be accessed from one single place. Data Lake also provide role-based access control.

## Reduced Costs

Highly scalable solution which significantly minimizes capital expenditures. Direct access to data to multiple users reduces IT costs.

## Advocate Single Source of Truth

It helps in consolidating all the data and provides a single source of truth, both for a company's operational systems and it's analytics.

## Diverse Data Structures

It has the ability to capture and handle widely diverse data structures and file types, including machine data from IoT, robots, sensors etc.

## Modernize Data Warehouse (DW)

Complementing existing DW with data lake gives more flexibility, speeds up data processing and allows to capture all forms of data, freeing up bandwidth of DW for business intelligence analytics.

## Increased Agility for Data Users

Ability to handle ad hoc queries and conduct real-time analysis while getting rid of the time and costs involved with IT assistance.

## Enhance Predictive Analytics with Machine Learning

It allows machine learning to complement the growing work in the predictive analytics space by ensuring outcomes and recommendations are more accurate and highly personalized.

## Advanced Analytics

It offers broader range of analytics such as discovery-oriented and predictive analytics, based on advanced technologies such as AI/ML etc, to gain competitive advantage and improve customer retention.



# Starting the Journey of Successful Data Lake Implementation

In order to effectively design a data lake, organizations must apply an agile approach piloting with prioritized use cases, testing and refining them, in contrast to an extensive one-time project to connect all data to the data lake. The design should depend on the organization's business goals, priorities, and selection of use cases. The data lake journey should begin with proper alignment of an organization's IT team and other business units. They should work in conjugation to develop an agile approach for building a data lake and must have the same outlook while answering questions such as-How do our existing capabilities of data management look like? How do we tackle the complicated streaming data? How complex is our data acquisition process? What type of skills, tools, and technologies are available in our organization?

Walking through all these questions and design principle would enable businesses to build an agile development model, thereby helping them realize the business benefits of data lakes quickly and placing a limit on future reworks and iterations.

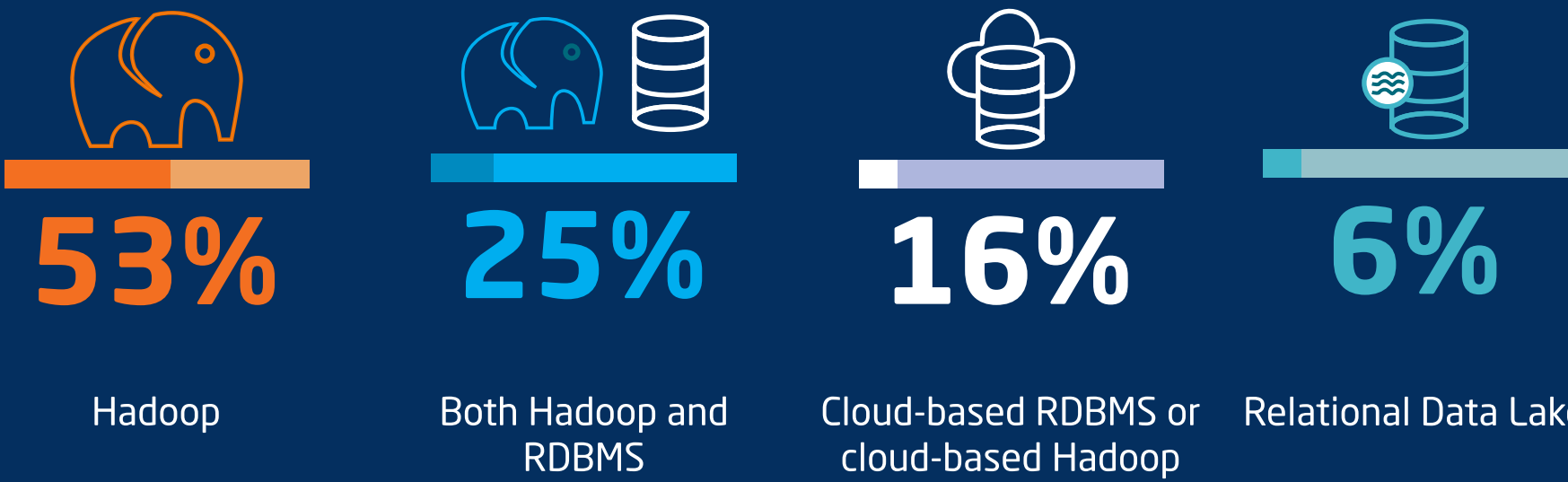
## An organization should follow the below mentioned key guiding principles to design a data lake:

- ① Foster data-driven innovation by making raw and modeled data open to capable data scientists
- ② Favour Open Standard and Technology Independent Solutions in developing the data lake
- ③ Avoid duplication of investment and functionalities in an organization's data systems by providing data warehousing and self-service BI services to different divisions within the organization
- ④ Centralize organization's data and enable timely information extraction
- ⑤ Ensure adaptability to a changing customer landscape, and support for variety, volumes, and velocity of data sources by design
- ⑥ Adopt proactive and strict information security measures to mitigate unauthorized accesses, undesired private information disclosures, and cybersecurity threats while conserving information value and utility.

# Key steps that an organization should follow to build its data lake to gain maximum business impact:

- 1 Create business goals and prioritize at least 2-3 use case
- 2 Select data platform based on organizations business requirement
- 3 Choose the tools & technologies that are easy to operate and satisfy the user’s requirement
- 4 Supplement your existing talent resources with specialized data lakes consultants to leverage their experience. Also, train the existing staff for Hadoop, analytics, lakes, etc.
- 5 Have a clear data governance strategy in place. Avoid dumping everything into the data lake and ensure that all the data residing in the data lake is properly cleansed, classified and protected or it will ultimately get clogged-up and become a data swamp which is nothing but a murky business liability.

## Preferred Platform for Data Lake



The best practice would be to narrow down on specific use cases around themes such as predictive analytics, omnichannel marketing, customer engagement and so on depending on the business requirements, and enjoy the agility of data lakes by accessing and analyzing any data in its native format to get deeper business insights.



# Conclusion

As more than 80% of the data generated across the world is unstructured, businesses have acknowledged the requirement of big data architecture for uncovering fresh opportunities and boosting growth. Looking at the current scenario, adoption of Data Lakes will surely witness an upswing as organizations will start reaping benefits from data lake implementations.


Organizations might have varied requirements that will form basis of the datalake adoption- whether for expanding analytics programs, extracting business value from new data assets or extending data warehouses- it's definitely worth the effort in this era of evolving data.


# Why NEC?


Core business experience combined with domain and technical expertise with evolving tools and technologies.


### Skilled Workforce


In-house business & domain experts

 Data Stewards

 Data Scientist

 Big Data Engineers

 Architects

 Consultants & Data Analysts

### Domain Expertise

Solving Complex Data Challenges  
Myriad of Successful Analytics Projects

### Dedicated Analytics Research Labs




Work for solutions for societies


### Robust Partner Ecosystem













Strong partnerships & alliances across data platforms and analytics solutions

# Technology Partners









**For further Queries & Demonstration, please contact:**  
**Noida Office - NEC Technologies India Pvt. Ltd.:**  
18<sup>th</sup> & 19<sup>th</sup> Floor, Tower C, Advant Navis Business Park,  
Plot 7, Sector 142, Noida, Uttar Pradesh 201305  
Tel: +91 120 683 2710

**Head Office - NEC Technologies India Pvt. Ltd.:**  
101 to 116, 1st Floor Splendor Forum, 3 District Center, Jasola  
New Delhi - 110025 Tel: +91 11 61101001  
Email: [enquiries@india.nec.com](mailto:enquiries@india.nec.com)  
Website: <https://in.nec.com>