NEC \Orchestrating a brighter world

Data Lakes

Driving the digital business transformation journey

Whitepaper

nec.com.au

About NEC

NEC has delivered world-class technology solutions and services to customers across the globe, for more than a century. For over 50 years in Australia, NEC has built a sophisticated technology and anything-as-a-service company which brings together the best technology and the best people to ensure our customers capture maximum value from their IT and networking investments.

NEC connects people through reliable communication infrastructure while also helping to keep communities safe and secure with intelligent surveillance systems and the world's leading biometrics identification technologies.

NEC \Orchestrating a brighter world

Data Lakes

Driving the digital business transformation journey

Contents

NEC \Orchestrating a brighter world

Introduction

The importance of robust data management in Digital Transformation

Digital Transformation has set new business rules and has made it inevitable for organisations to revamp their business models. Data driven digital transformation has proved to be a game changer in industries across the globe.

Organisations are increasingly making use of data analysis in order to unlock fresh opportunities, gain new insights, enhance customer experience, boost operational efficiency and improve their bottom line.

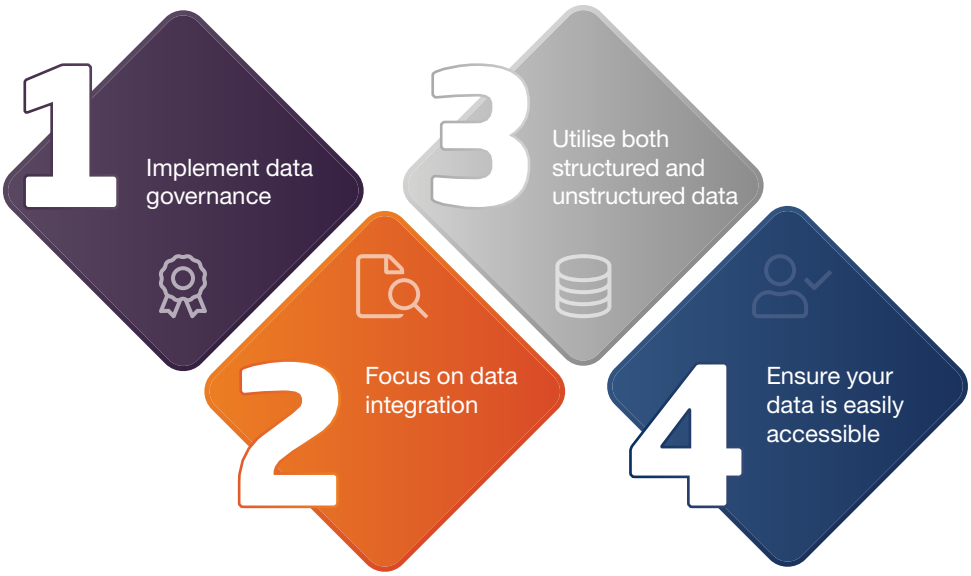
A major challenge in leveraging data to drive benefits is the lack of ability of organisations to implement an effcient data management strategy courtesy the large volume, variety and fragmented data. Now is the time to make data management and governance a priority and to develop optimised data management strategies. Lack of a robust data management strategy increases the risk of data leaks, unnecessary data storage, data management cost and poor data governance.

Setting up of a centralised Data Lake is one option to mitigate some of these risks. It consolidates a large volume of different data types from disparate sources on to a common platform in a cost-effective way.

Key data management strategies

An essential component of an organisation's digital transformation journey is an agile data management strategy. When aligned with the company's digitisation efforts, the data management strategy holds high significance and is a key component in developing a roadmap for a successful digital transformation journey.

The key activities to be performed in line with the data management strategy are:



Implement data governance

Data Governance is driven by a governing council which defines a set of policies and 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 effciently manage data for delivering timely, trusted and relevant information to make informed business decisions.

Focus on data integration

Data integration is the process of combining data from disparate sources using data pipelines and a variety of integrations, to enable users to query and manipulate it as needed. As organisations aim to create a single source of truth, the demand for data integration will witness a steady rise. This enables analysis to be more comprehensive in order to solve complex business problems and making smarter business decisions.

According to projections from IDC, 80% of worldwide data will be unstructured by 2025. For many large companies, it has reached that critical mass already.

Utilise both structured and unstructured data

Organisations that wish to leverage their data, use technology tools that are capable of utilising the benefits of both structured and unstructured data. Insights gained from unstructured data can be used to support the business decisions. By analysing social media content such as tweets and Facebook posts, companies can discover the trends and behaviour 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 organisations to acquire complete intelligence, leading to exponentially more powerful insights.

Ensure your data is easily accessible

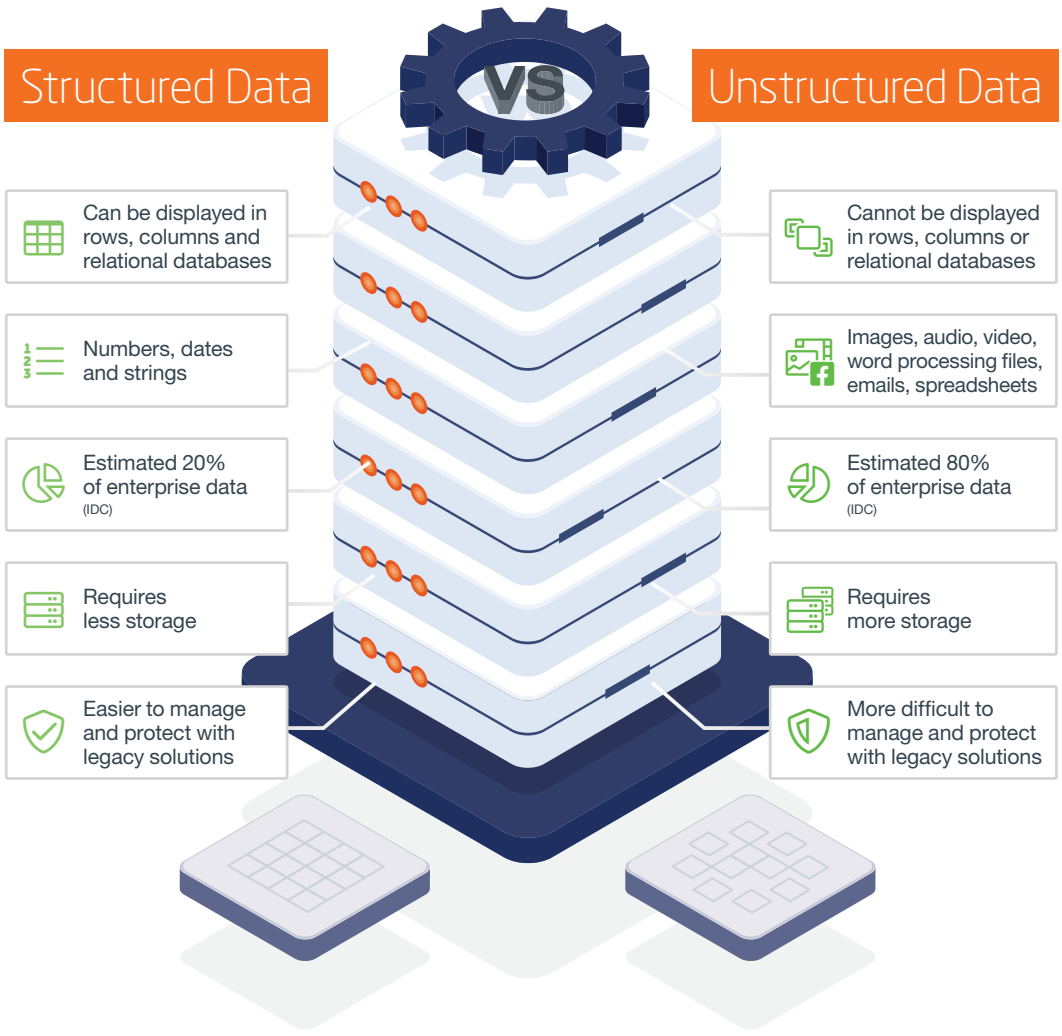
While proper data storage is essential, it is also important for organisations to provide ease-of-access for those who need to use it. The data should be stored in such a way that people can quickly find and access the information. However, organisations should also ensure that the right data is accessible to the right users while denying access to those who doesn't have the access rights.

Data Lake as a data management technology

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 continues to be the most preferred platform to provide organisations with cleansed, organised 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 the requirement of data to be structured before analysis has led to the advent of centralised data repositories; known as data lakes, that provide quick access to ever-changing types, volume and velocity of data.

What is a Data Lake?

Data Lake is a centralised repository that allows an organisation to store structured as well as unstructured data at any scale in its native format, process and analyse to derive new business insight. It helps in transforming information management into a proactive and real-time practice by enabling organisations 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.



Data Lakes vs Data Warehouses

Differences in managing data

Data Lake			Data Warehose
Structured, semi-structured, unstructured, raw.	Data Types	Structured, cleaned, processed.	
Native raw data from any source including SNS, IOT, ERP / CRM systems.	Data Sources	Structured data (historical and relational) typically extracted from transactional systems, operational databases and applications.	
Retains all the data for an unlimited amount of time.	Data Processing	Does not retain data forever, data is purged periodically.	
Schema-on-read, raw data only transformed when put to use.	Data Retention	Schema-on-write, structured and cleansed data.	

Differences in data use and application

	Data Lakes	Data Warehouses
PURPOSE	Application of data has not yet been determined	Pre-defined, usually for business decisions
CONFIGURABILITY	Can be configured and reconfigured as and when required	More complicated and costly because of their structured nature
USE CASES	Building data pipelines, machine learning and predictive analytics	Batch processing, business intelligence and reporting
BEST SUITED FOR	Experimental analyses, pattern recognition and prediction	Repeatable processes, ongoing analyses e.g. business reports
USERS	Data scientists, Big data developers, machine learning engineers	Analysts and business users
STORAGE COST	Lower cost storage compared to data warehouses	Relatively higher cost to data lakes with faster query results

Organisations that implemented Data Lakes outperformed similar companies by almost 9% in organic revenue growth.

Aberdeen: Angling for insight in today's Data Lake

Data Lakes enable companies to run new forms of analytics such as machine learning over log files, social media, and internet connected devices; identifying new opportunities, enhancing customer retention by uncovering their preferences, and aiding in meaningful business decisions to ensure growth.

Benefits of a Data Lake

Advanced Analytics

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

Simplified data access

User-defined access to multiple users enabling them to work with data from multiple sources across the organisation.

Reduced costs

Highly scalable solution which significantly minimises capital expenditures. Direct access to data to multiple users reduces IoT assistance cost.

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.

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.

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 its analytics.

Modernise data warehouse

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.

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 personalised.

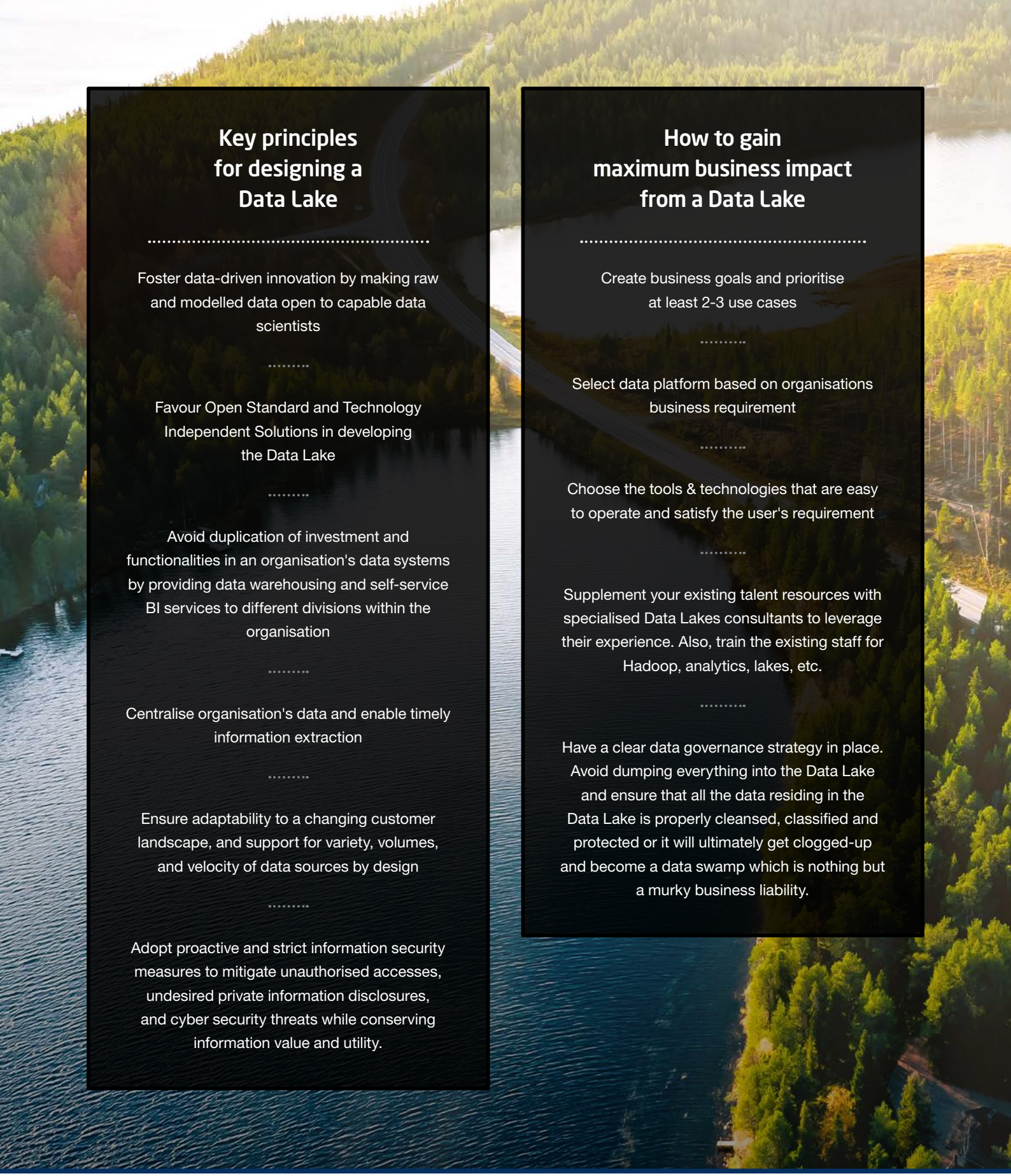
How to start the journey of successful Data Lake implementation

In order to effectively design a Data Lake, organisations must apply an agile approach piloting with prioritised 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 organisation's business goals, priorities and selection of use cases.

The Data Lake journey should begin with proper alignment of an organisation'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:

1. What do our existing capabilities of data management look like?
2. How do we tackle the complicated streaming data?
3. How complex is our data acquisition process?
4. What type of skills, tools, and technologies are available in our organisation?

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



Key principles for designing a Data Lake

Foster data-driven innovation by making raw and modelled data open to capable data scientists

Favour Open Standard and Technology Independent Solutions in developing the Data Lake

Avoid duplication of investment and functionalities in an organisation's data systems by providing data warehousing and self-service BI services to different divisions within the organisation

Centralise organisation's data and enable timely information extraction

Ensure adaptability to a changing customer landscape, and support for variety, volumes, and velocity of data sources by design

Adopt proactive and strict information security measures to mitigate unauthorised accesses, undesired private information disclosures, and cyber security threats while conserving information value and utility.

How to gain maximum business impact from a Data Lake

Create business goals and prioritise at least 2-3 use cases

Select data platform based on organisations business requirement

Choose the tools & technologies that are easy to operate and satisfy the user's requirement

Supplement your existing talent resources with specialised Data Lakes consultants to leverage their experience. Also, train the existing staff for Hadoop, analytics, lakes, etc.

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.

The best practice would be to narrow down on specific use cases around themes such as predictive analytics, omni-channel marketing, customer engagement and so on depending on the business requirements, and enjoy the agility of Data Lakes by accessing and analysing 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 adoption rate, application of Data Lakes will surely witness an upswing as organisations will start reaping benefits from Data Lake implementations.

Organisations might have varied requirements that will form basis of the Data Lake 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.


nec.com.au
NEC
\Orchestrating a brighter world