Make your analytics more agile

Choose data lake in the cloud

A data lake is a central repository for an enterprise to store all data in its native format including structured, semi-structured and unstructured data.

Through its ability to store huge amounts of varied information, data lake can be used for flexible analytics that supports smart decision making. Data from the various sources can be used for many different applications and analyses, including real time analytics and machine learning. The aim is to be as agile as possible in achieving optimal results and responding to new business opportunities.

One of the biggest benefits of data lakes is the flexibility to drive your business forward through agile analytics that can measure performance and improve productivity by making informed judgements, for example. This is done by taking advantage of consistent big data and deep learning algorithms to deliver real-time decision analytics.

Although data lakes can be built on premise, they lend themselves to being created in the cloud. Cloud offers the performance and scalability that data lake requires, together with economies of scale and access to a range of analytics engines. Enterprises also benefit from enhanced user availability.

Why build a data lake?

Schema-based traditional data warehouses are not optimized for the variety and sheer volume of big data. Data lakes, on the other hand, can store data from diverse sources in their native format, including multimedia, social and XML. Data lakes are infinitely scalable, designed for rapid data ingestion and support the demands of the emerging Internet of Things (IoT), all of which make them a perfect partner for big data.

42%

of enterprises implement data lakes to improve operational efficiency.[1]

Key features of a data lake:

- **Ability to store all native data types including data gathered from mobile apps, social media, IoT devices and enterprise applications**

- **Highly flexible**

- **Unlimited ways to query data**

- **Elimination of data silos**

- **Schema on-read offering unparalleled data ingestion speeds**

- **Highly extensible at lower cost**

In addition, data stored in data lakes is easily accessible. Subsets can be made available to different groups in an enterprise, for example, whether it is for self-service or delivered to data scientists for analysis.

Data lakes come with the promise of more visibility into your data and breaking down silos around your enterprise by storing all data in one central repository. But a large number of data lake initiatives end up failing due to the organizational and cultural changes required to create and operate business projects on top of data lakes.

This is where Orange Business Services comes in. We help you assess if you have the business cases necessary for building a data lake and help you through the complexities of creating it. Enterprises, for example, often forget that data lakes come with no structure, so they need to construct a catalog integrated with governance. Without this, the data has no context and data scientists will waste time trawling through undefined data sets for data sources.

Business Services

orange™

1. Aberdeen Research – Angling for insight into today's data lake 2017.

Building your data lake in the cloud

According to research by Aberdeen Research, enterprises are looking to implement data lakes for two key reasons: they want to take advantage of more sophisticated and advanced analytical tools and techniques, and they want their data to perform more efficiently. This includes everyday functions such as data access and retrieval.

70%

of mature organizations will have more data flowing from data lakes to data warehouses than vice versa by 2021.[2]

Moving or building your data lake in the cloud has several advantages including agile pay-for-use, on-demand infrastructure, frequent feature updates, enhanced security and geographical coverage. Creating a properly architected and governed data lake in the cloud, however, isn't quite as simple as it sounds. Moving your data to a cloud data lake, for example, can't be done in one shift. It needs to be done over time, choosing the business cases best suited to migrate first.

Securing your data lake

Securing the data in your data lake is imperative. This requires a holistic view of the data in your data lake, how you plan to use it, the governance requirements, authorized access and planned applications.

Data lakes do not come with the governance and compliance policies you get with traditional database management. This is something you need to put in place. Data can be flagged, for example, to specify access. Orange Business Services can help you with these data lake security issues, leveraging the expertise of Orange Cyberdefense.

The importance of a data catalog

Creating a data catalog combined with governance is crucial in understanding the data in your data lake and ensuring its trustworthiness. The data catalog is designed to provide a single source of truth about the contents of the data lake and helps you to understand the sources as well as the transformations of the data. In addition, a data catalog helps data analysts confirm they are using the right data and that it conforms to both organizational policies and regulations such as GDPR.

Our data lake consultants can map a data lake for you that will address your specific business requirements, providing you with a secure, flexible, cost-effective way of storing, processing and analyzing your data.

Six steps to a building a data lake in the cloud

A successful data lake is far more than a simple data repository, it is a dynamic tool that can provide your enterprise with valuable insight essential to business growth and digital transformation. Adhering to these six steps will help you to build a productive data lake.

1 **The scoping phase is very important.** If the data lake is to work effectively you should always start with concrete use cases and continuously examine their business value. At the beginning it is a good idea even to choose high quality data sources.

2 **Carefully plan your data lake build and deployment,** outlining the skills that will be necessary to make it a success.

3 **Storage costs to load data into the cloud may be relatively inexpensive.** But, if you want to get data back into your own systems or replicate it over availability zones it can be very expensive. Consider if this is something you may need to do in the future.

4 **Include the active and secure management of data** in your data lake plan.

5 **Ensure you have data governance and a data catalog implemented** in your data lake from the beginning so you know what data is in the data lake, why it is there and who is using it – or you will end up with an unusable data swamp.

6 **Load data in stages, perform data quality checks and manage it.** Your results will only be as good as the quality of your data.

Click here to find out more about how Orange Business Services can help you set up a data lake in the cloud.

orange™ **Business Services**