



**WHY**

**WHAT**



**WHO**



**HOW**



**WHEN**

# Data Lake

# Data Lake

## in a modern data WORLD



Is there a need for quick introductions to data in order for suitable actions to be performed in real time?

**Does your Bank (plan to) use Big Data in business?**

**What questions do we ask more and more frequently?**

Is there a need for the customer journey to be monitored on all channels, both online and offline?

Is there any data you are currently not collecting, but which you need or will need?

Is there a need to analyze calls from the call center?

### Some data issues in banking world

- It is estimated that 70% of the time for analytics projects is spent identifying, cleansing, and integrating data.
- Data is often difficult to locate because it is scattered among many business applications and business systems.
- Frequently the data needs reengineering and reformatting in order to make it easier to analyze.
- The data must be refreshed regularly to keep it up-to-date when it is in use by analytics.

### What's Changing in the industry?

- Chief Data Officer – new position in the banking industry
- The hopes and dreams around Big data are becoming a reality for neo banks and some fintech companies
- Regulation is pushing responsibility for governance on business executives resulting in
  - Business taking more control of information governance
  - Business wanting more access to data
  - Organization seeking more automation of governance actions

# Data Lake

## WHAT

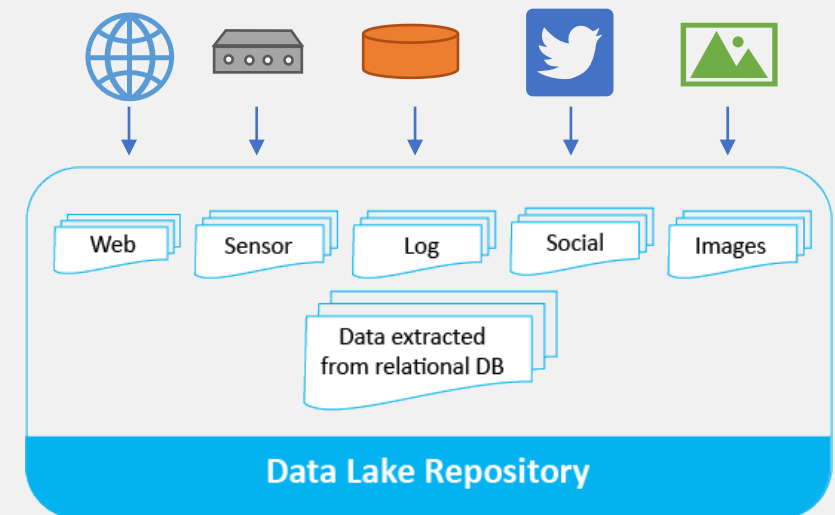
### What is Big Data

Everything we do in the modern world creates some form of data. With each purchase, visit to a website, our phones or other electronic devices leave behind some traces, i.e. data. It is nothing new that the basic force of some modern businesses lies in their ability to process huge amounts of data and make predictions, reach conclusions and propose actions based on such data. The expression Big Data is primarily used to emphasize the diversity in structures of processed data, and the classification of data into categories (in the table below) is the key to understanding how great a challenge it is to have disposal over all this data in a standardized manner in order to ensure results. Storage of all these data types so as to enable their optimal usage at the moment when there is a need for this is the first step in the exploitation of business value from the Big Data concept.

Type of data	DWH	DL
Structured (rows/columns)	✓	✓
Semi-structured (csv, logs...)	✓	✓
Unstructured (e-mails/docs)	✗	✓
Binary (photo, audio, video)	✗	✓

### What is Data Lake

Data Lake is a virtual place for collecting and keeping structured and non-structured data. Data can be stored in its initial form without the need to transform it in any way. After storage, it is possible to create various types of queries, searches and processing of data by using tools for analytics, real-time processing and machine learning algorithms. In this manner, companies can receive higher quality information from the data they already have but cannot use it in its initial form. All of the above makes **Data Lake a natural environment for Big Data and the basis of any bank initiatives in the field of artificial intelligence or machine learning.**



# Data Lake

## WHAT

### Data Lake or Data Warehouse?

**Data Lake is an expansion of the Data Warehouse concept** because, in addition to the structured data, it also stores non-structured data whose sources are mobile applications, IoT sensors or social media. They are searched in another way, using machine learning, text search algorithms, Big Data analytics.

Data Warehouse

Data Lake

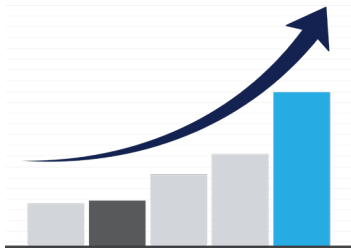
**Data Warehouse** is a base optimized for analysis of relation data coming from transaction systems and a range of business applications. The data structure is defined in advance and optimized for searching using SQL queries.

	Processing	Storage	Agility	Security	Users
DWH	Shape and structure before loading	High cost	Highly structured (time consuming to change)	<b>Mature and safe security options</b>	<b>BI for everyone... but not in reality</b>
Data lake	<b>Load raw data then shape when using</b>	<b>Low cost</b>	<b>Very flexible (easily change configuration)</b>	Security solutions in development phase	Data scientists



# Data Lake

## WHY



### Why is Data Lake necessary

#### “Why Banks Should Bank on Data Lakes”

Organizations that manage to obtain high-quality information and conclusions from all its data achieve a significant advantage over their competition. They are more successful because they achieve income growth by analyzing data from logs, website clicks, social media, etc., since in that way they can detect trends and react to changes faster, but also by saving on resources when this manner of business becomes a standard.

### General benefits



**Cost efficiency** – captures and stores data in a single data warehouse, making it cost-effective.



**Optimization** - As structured and semi-structured data is stored and managed in a single repository, data processing activities are optimized. Workloads such as data transformation and integration are performed relatively faster with this solution.



**Efficiency** - makes Extract, Transform, Load (ETL) faster and more efficient



**Data-based innovation** - allows analytics tools to work across data that may not have been associated before, generating new insights for businesses



# Data Lake

## WHY



## Examples of using Data Lake in banks:

### CRM and personalised marketing

- Customer relationship management consists of Customer profiling and knowledge, segmentation, measurement of satisfaction, retention and acquisition strategies and all of them can be enhanced with the use of big data and predictive analytics and to have such data, an organization has to deal with its storage and Data lake is a natural place for it to be stored - The key to success in marketing is to make a customized offer that suits the particular client's needs and preferences - meaning offering the right product to the right person at the right time on the right device

### Customer journey, CLV and smart pricing

- Outstanding customer support service is the key to keep a productive long-term relationship with your customers. Data science makes this process better automated, more accurate, personal, direct, productive and less costly concerning employee time. Customer lifetime value (CLV) is a prediction of all the value a business will derive from their entire relationship with a customer and it is also based on big data analytics and it is a foundation for smart pricing.

### Fraud prevention and AML

- An example of efficient fraud detection is when some unusually high transactions occur and the bank's fraud prevention system is set up to put them on hold until the account holder confirms the deal. For new accounts, fraud detection algorithms can investigate unusually high purchases of popular items, or multiple accounts opened in a short period. Similarly, predictive model can be trained with scenarios as well as from real time cases to predict AML watch cases.

### Risk modeling and credit scoring

- Modeling a credit scoring system is getting a new meaning with all the different types of data that can be stored and analyzed with Big data technology as all the soft behavior data is more available than ever and it opens one whole new world of opportunities to create a model for financing segments that were not eligible for loans within standard risk policies based on information that was very structured and limited before Big data

### IT development in the era of CX/EX

- By extracting, storing and analyzing logs of all activities clients perform on the bank's digital channels, it is possible to get insight into the real user experience of those channels and make improvements that are data-based and most likely to add value to the customer and boost their satisfaction as well as to solve bottlenecks of certain processes



# Data Lake

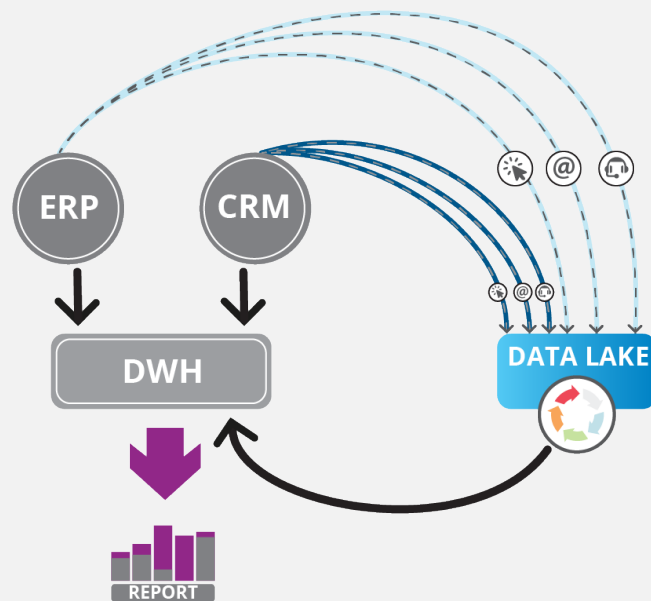
## HOW

**Ibis Data Lake** is a software solution for implementation of Data Lake environments according to the reference architecture and the best practices developed by the development center of Ibis Group that has been dealing in Big Data technologies since 2014. The best industry practice has been incorporated into the Ibis Data Lake, as well as the best experiences gained in many production Data Lake implementations carried out by Ibis Group.

The Ibis product is independent and is implemented on any of the leading Hadoop distributions.



\*besplatna ili komercijalna verzija



### Ibis Data Lake includes:



**Data Lake implementation** according to the reference architecture and best practices



**System configuration** and tuning according to the best practices



**Training** for system usage

... and with clients it is implemented in three predefined variations which include different licenses:

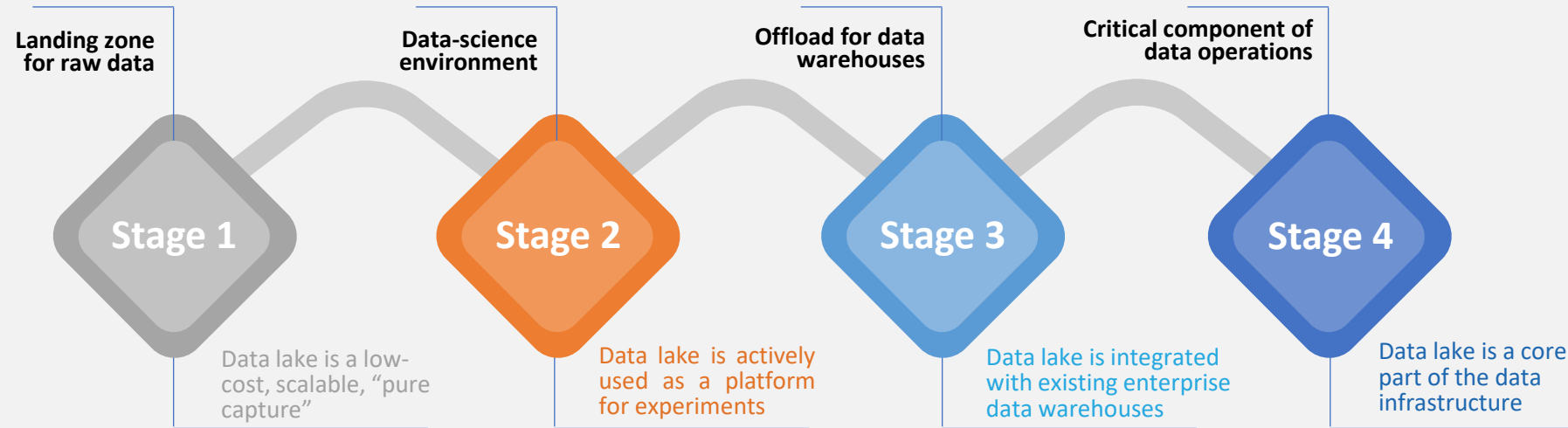
Edition	Performances	Components
Standard	Data storage centralization	Hadoop stack
Data Science	Machine learning/ Data mining	R-studio, Pyton, Knime, Jupiter
Full environment	Real-time stream processing	Apache Flink, Kafka



# Data Lake

## HOW

While Data Lake offers enormous potential, building its architecture is not an easy job, particularly when integrating multiple legacy data platforms and applications. The key to building and scaling an actionable data lake that meets the needs of the business is to balance the agility and flexibility. We are using the McKinsey best practice following 4 stages:



## Data Lake in real time

In the phase of full implementation of all services and integration with DWH, Ibis Data Lake processes data in real time and by combining it with batch data it adds special value in different business segments:

- Real time overdrafts and credit cards offering
- AML and fraud prevention notifications
- Omnichannel customer support

### Batch sources

- DWH, CRM
- Web clicks
- NBS, APR
- Social networks
- Mails, documents

### Realtime sources

- Core Banking Syst
- Card transactions
- E/M -banking
- IOT
- LOGs

Data lake is implemented as an on-premise or as a Cloud/hybrid option, and it can also be implemented as a cloud-native solution by using a cloud service of one of the Cloud providers, such as: Microsoft Azure, IBM Softlayer, Oracle Cloud (in such case, cloud services are additionally licensed)

ORACLE

Partner

Platinum  
Business  
Partner

IBM



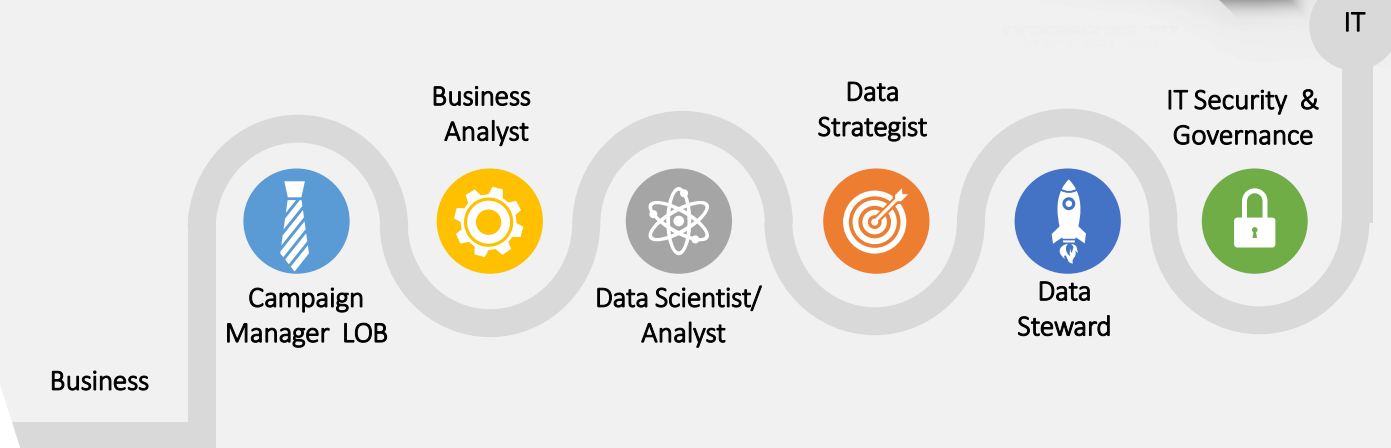
# Data Lake

## WHO



- **Business structures (depending on use case):** business line managers, risk and AML managers, bank CRM and CX experts
- **Outsourced expert support - Ibis Solutions team consisting of:**
  - ✓ Banking experts in the fields of business & risk
  - ✓ Tech experts in the field of software development & data science

Stakeholders for this topic and resources required for setting up a Data Lake infrastructure and practicing in using its potentials:



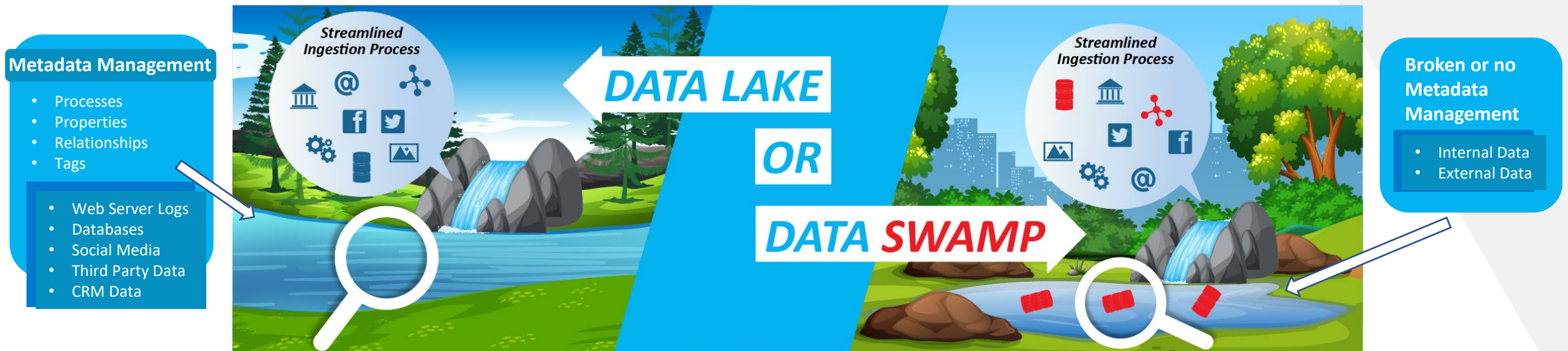
## Who benefits from a Data Lake?

1. Users, business analysts and data scientists can easily find the information they need without extensive IT involvement
2. Data strategists and data stewards can make information available to users in an organized and well-governed manner
3. IT security and governance teams can be assured that information is governed according to well-defined organizational and regulatory policies

# Data Lake

## WHEN

- ASAP – in order for the data to be used for any purpose, modelling is required, models need to be tested, and testing requires series of historical data – series length is also necessary depending on the model type, but most require at least 6 to 12 months. Therefore, in order to implement some of the data-based models, it is necessary to think about this much sooner and start collecting data as soon as possible. A relatively cheap option of storing data in a Data Lake in banks gives the option of also storing some data that still has no defined exact use value, but that at the moment when it appears and is crystallized, there are already relevant series of data that can be quickly put to use by creating predictive models.
- The above must not be taken lightly and Data Lake is not a place that is filled without a strategic approach and expert and responsible coordination of all needs of the bank in a single place. What is important in the decision and realization of the Data Lake system for storing data is to pay attention that the lake does not turn into a swamp (Data Swamp), which can occur if the entire process of setting up and using such infrastructure is led without experience – that is why it is necessary to engage both internal and experienced external resources on such a complex project, so that everyone together could use the advantages of their previous experiences.
- Data Lake is an investment into the future – NOW is the best time



**For more info**  
**please contact us:**



Nikola Nikačević  
Key Account Manager  
[nikola.nikacevic@ibis-solutions.rs](mailto:nikola.nikacevic@ibis-solutions.rs)  
Mobile: +381 64 64 50 437



Vladimir Rakić  
Director of Software Development  
[vladimir.rakic@ibis-solutions.rs](mailto:vladimir.rakic@ibis-solutions.rs)  
Mobile: +381 64 8278 395



Adress Tošin bunar 272, 11070 Belgrade

Phone: 381 11715 22 00

Fax: 381 11715 22 01

E-mail: [info@ibis-solutions.rs](mailto:info@ibis-solutions.rs)

[www.ibis-solutions.rs](http://www.ibis-solutions.rs)