

# Modern Analytics Platform

Assess and Manage AI and ML Data Platforms



# AGENDA

1. REPLY & DATA REPLY
2. ANALYTICS PLATFORM ASSESSMENT
3. DATA LAKEHOUSE
4. PLANNING A DATA PROJECT



**Fabio Limardo**  
*Senior Consultant*  
[f.limardo@reply.it](mailto:f.limardo@reply.it)

[Info.data.it@reply.it](mailto:Info.data.it@reply.it)

[www.reply.com/data-reply/it/](http://www.reply.com/data-reply/it/)



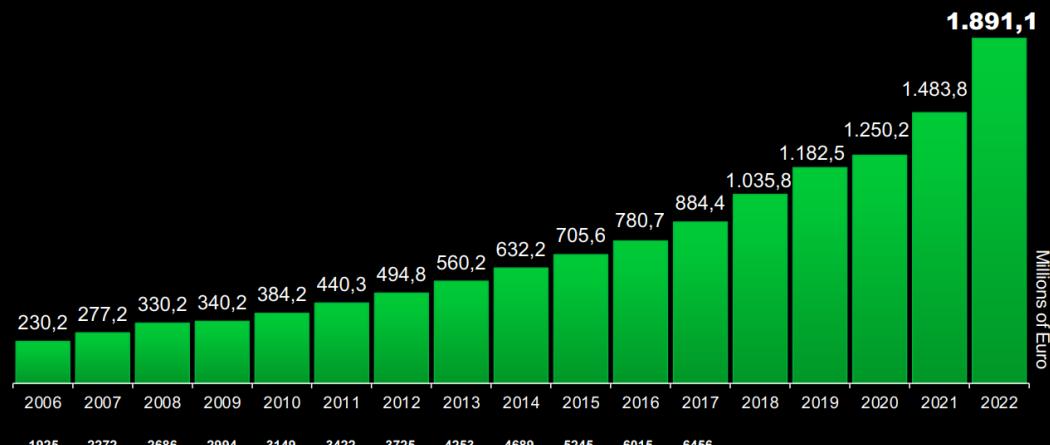
# WHERE WE ARE



Reply is a network of companies highly specialized in technologies and industry sectors. Our services include **Consulting, System Integration** and **Digital Services**, and we cover three main areas of competence: **Processes, Applications and Technologies.**



# REVENUE & PEOPLE



# REPLY SERVICES

## INTERNET OF THINGS

-  AUTONOMOUS VEHICLES
-  INDUSTRIAL SYSTEMS
-  CONNECTED PRODUCTS
-  ENERGY ECOSYSTEMS
-  HEALTHCARE

## INDUSTRY PLATFORMS

-  RETAIL & CPG
-  ENERGY
-  HEALTHCARE
-  TELECOM & MEDIA
-  MANUFACTURING & LOGISTICS
-  FINANCIAL SERVICES

## CUSTOMER EXPERIENCE

-  IMMERSIVE EXPERIENCE
-  DESIGN & UX
-  VIDEO
-  SOCIAL MEDIA & STORYTELLING
-  DIGITAL ECOSYSTEM

INTELLIGENT  
AUTOMATION

## ARTIFICIAL INTELLIGENCE

MACHINE  
LEARNING

DEVELOPMENT  
& OPERATIONS

## CLOUD PLATFORMS

DATA

SECURITY  
OPERATION CENTER

## CYBERSECURITY

SECURITY  
CONSULTING



# MAIN AWARDS & ACHIEVEMENTS

2023



IBEXA PARTNER  
OF THE YEAR  
Comwrap Reply

2023



SAP PARTNER OF THE YEAR  
& SAP ENERGY PARTNER  
OF THE YEAR 2023  
Reply

2023



ADODE PLATINUM  
SOLUTION PARTNER  
Reply

2023



ORACLE CLOUD  
SOLUTION PROVIDER  
Red Reply

2023



LEADER IN THE 'SALESFORCE  
ECOSYSTEM PARTNERS 2023'  
ISG PROVIDER LENS™ STUDY  
FOR GERMANY AND UK  
Reply

2023



NEW PRODUCT  
INNOVATION  
AWARD 2023  
Axulus Reply

2023



SOLUTIONS PARTNER  
FOR MICROSOFT  
CLOUD  
Reply

2022



BEST WORKPLACE  
FOR WELLBEING (UK)  
Airwalk Reply

2022



IMPLEMENTATION  
PARTNER & COMMUNITY  
IMPACT PARTNER  
OF THE YEAR  
Arlanis Reply

2022



SYSTEM INTEGRATION  
PARTNER OF THE YEAR  
& SECURITY PARTNER  
OF THE YEAR – EMEA  
Reply

2022



BEST USE OF  
TECHNOLOGY FOR  
ENGAGEMENT &  
INTERACTION (B2B)  
Avvio Reply

2022



SERVICES PARTNER  
OF THE YEAR 2022 –  
SOUTH EUROPE  
Communication Valley Reply

2022



FRAUNHOFER VALIDATION  
FOR MATERIAL FLOW AND  
LOGISTICS (ILM)  
LEA Reply

2022



21 ORACLE SERVICE  
EXPERTISE  
CERTIFICATIONS  
Reply

2022



LEADER IN DIGITAL  
EXPERIENCE SERVICES  
Reply

2022



PARTNER OF THE YEAR  
IN THE EMA REGION  
Spike Reply, Live Reply,  
Glue Reply

2022



NIELSEN NORMAN  
INTRANET DESIGN  
ANNUAL AWARD 2022  
Bitmama Reply

2022



SAP® PINNACLE AWARD  
IN THE CUSTOMER  
EXCELLENCE CATEGORY  
Reply

2022



PREMIER PARTNER  
GOOGLE 2022  
Go Reply, Like Reply

2022



SAP RECOGNIZED  
EXPERTISE DESIGNATION  
FOR UTILITIES, CONSUMER  
PRODUCTS AND RETAIL  
Syskolan Reply, 4brands Reply,  
Power Reply, Portatech Reply

2022



INNOVATOR IN THE  
AVASANT RETAIL DIGITAL  
SERVICES 2022-2023  
RADARVIEW™  
Reply

2022



VISIONARY IN  
MAGIC QUADRANT  
FOR WMS  
Reply

2022



"BEST IN CLASS" PROVIDER IN PAC  
INNOVATION RADAR "SAP-RELATED  
SERVICES IN GERMANY 2022" &  
"LEADING SALESFORCE SERVICE  
PROVIDERS IN EUROPE 2022"  
Reply

2022



PREMIER CLOUD  
SOLUTION  
PARTNER  
Blue Reply

2022



SALESFORCE CONSULTING  
PARTNER AND EXPERT IN SALES  
& SERVICE CLOUD, AUTOMOTIVE  
& MANUFACTURING  
Arlanis Reply

2022



BVWD INTERNET  
AGENCY RANKING  
1st PLACE  
Reply Digital  
Experience

2022



GOOGLE CLOUD PARTNER  
ADVANTAGE SPECIALIZATIONS  
IN INFRASTRUCTURE, CLOUD  
MIGRATION AND MANAGED  
SERVICES PROVIDER  
Go Reply

2022



PLATINUM  
PARTNER  
Open Reply





**BIG DATA | ARTIFICIAL INTELLIGENCE | QUANTUM**  
WE SUPPORT C-LEVELS TO EXTRACT VALUE FROM DATA

In all industries and all business functions.  
With more than 100 projects already delivered in production.

## BIG DATA & VISUALIZATION

Group of Big Data architects and developers with extended experience in Hadoop & NoSQL solutions, both on-prem and in Cloud, and Data Visualization tools.

## DATA SCIENCE & AI

Specialized in designing and implementing Advanced Analytics solutions, Artificial Intelligence and Machine Learning models.

Confidential

**OUR EXPERTS:**   
Big Data Engineers & Data Scientists

100+

Production projects, up and running

CROSS  
INDUSTRY  
experience



# REPLY INDUSTRIES



**Manufac**

Confidential

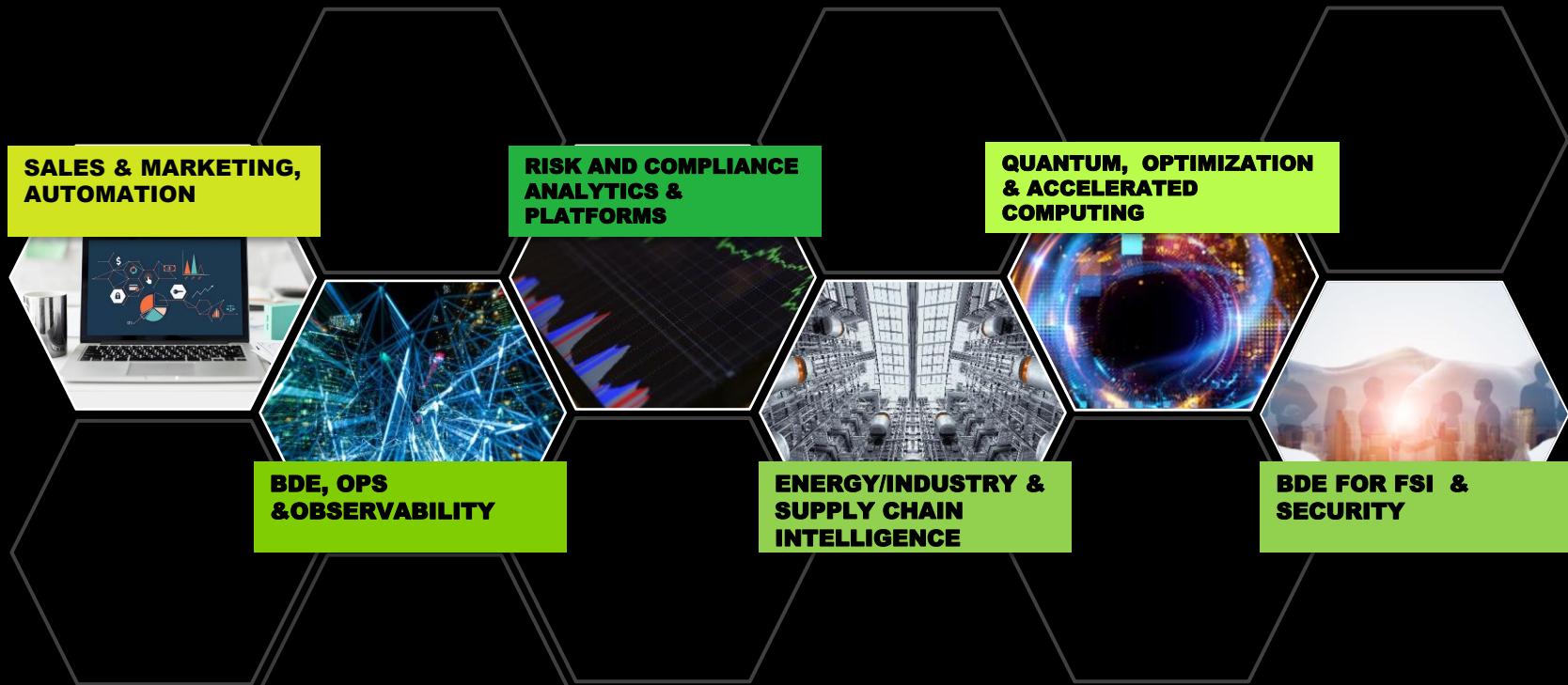
**Utilit**

**Telco/Media/Publisher**

**Pharma**

**Others**

# DATA REPLY - VERTICAL OFFERING



# DATA REPLY - PARTNERSHIPS



## AWS COMPETENCIES

Big Data	DevOps	Oracle	Migration
IoT	Industrial Software	SaaS	Machine Learning
Security	Financial	Retail	Energy



CLOUDERA



# COLLABORATIONS



DEVO LAB  
DIGITAL ENTERPRISE VALUE  
AND ORGANIZATION

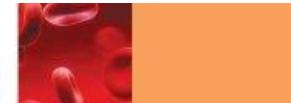


# SCIENTIFIC PUBLICATIONS ON PEER REVIEWED JOURNALS



**nature  
communications**

International Journal for  
**Numerical Methods in  
Biomedical Engineering**



**SCIENTIFIC  
REPORTS**



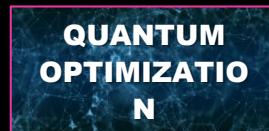
**CMBBE**

International Symposium  
on Computer Methods in Biomechanics  
and Biomedical Engineering



# DATA REPLY EXPERTISE

- SALES & MARKETING ANALYTICS
- BIGDATA ENGINEERING & SECURITY INTELLIGENCE
- ENTERPRISE ANALYTICS
- INDUSTRY & IOT ANALYTICS
- QUANTUM & GPU ACCELERATED COMPUTING



# DATA REPLY

## TECHNOLOGIES EXPERTISE

### DATABASE NO SQL



### INDEXING



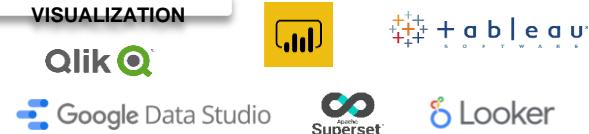
### CONTAINERIZATION



### DATA SCIENCE LIBRARIES



### DATA VISUALIZATION



### LANGUAGES E FRAMEWORK



### ACCELERATED COMPUTING



### BIG DATA PLATFORM



### CLOUD PLATFORM



### ETL TOOLS



### DS TOOLS



### GENERATIVE AI





Why does Data Reply  
give money to Fabio?

# DATA REPLY - VERTICA



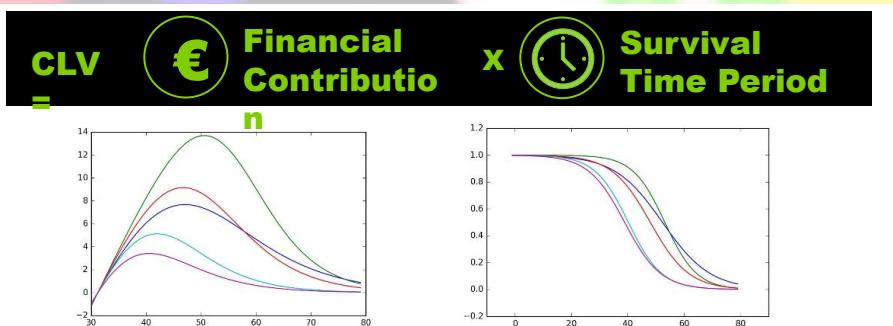
**SALES & MARKETING,  
AUTOMATION**

**RISK AND COMPLIANCE  
ANALYTICS &  
PLATFORMS**



# CLV FORECASTING

- The customer is a global leader media company
- They need a new model that integrates the **churn risk** knowledge, as survival probability, with the prediction of the **future net margin** bring by the customer in order to forecast the **Customer Lifetime Value** for each client
- The desired model time horizon is **three years**



## CHALLENGE

- Need of collection and ingestion of **cost data** in Data Lake in a single stable source
- Extend an existing **churn model** to a longer time horizon from 3 to 12 months and projecting the result to 3 years
- Predict the future value, in terms of **revenues**, of a customer in a high **variable and dynamic context** and extend it to 3 years
- Integrate the **future business actions** and decisions in order to improve the final model performances



## THE SOLUTION

Churn Model and Net Revenue model have been developed in python and later combined in Google Big Query:

- Churn model** is trained on segments of the customer base, one model for 3,6,9 and 12 months. The algorithm used is a **LightGBM Classifier**.
- Net revenue model** is trained to all the customer base and it aims to predict the entity of the difference in the future revenue in time horizons of 3,6,9 and 12 months. The algorithm used is a **LightGBM Tweedie Regressor**.
- The prediction at 3 years is projected with interpolation of the predictions obtained with the previous models.
- CLV value** is calculated as a combination of the financial contribution and survival probability with advanced mathematical logics.



## BUSINESS VALUE

- Provide **CLV** as a **new metric** in order to understand, segment and classify the customers based on their predicted future monetary value
- Monitoring overall customer base value trends
- Timely forecast results on **3 years**
- Provide a new KPI that can be exploit for **Marketing** and **Campaign** actions

Client: **Fashion**

# PROMO OPTIMIZATION

- The client is a American brand of fashion that sells all over the world
- The client is one of the most known brand for bottom wear
- Multiple promotions are planned during the year, both **online** and in the **stores**.
- Due to COVID emergency the main goal is to maximise revenues and sell-through.



## CHALLENGE

- Deal with heterogeneous sales behaviors (recurrent, seasonal, sporadic..)
- Not all products had price changes in the past.
- Take into account different constraints for the different channels.



## THE SOLUTION

- Retrieve the data stored in AWS S3 through **Spark**
- Data preparation and pre-processing on Dataiku.
- Hierarchical approach for building **elasticity curves**
- **Optimization** model to identify the best discount for each product, given a set of constraints.



## BUSINESS VALUE

- Increase in **revenues** (2 times more in e-comm, 1 time more in Retail)
- Increase in **sell-trough** (5 times more in e-comm, 2 times more in Retail)
- Specific level of **margin** preserved

# MY STORY (LONG STORY SHORT)

2017 - 2018

**Data Engineer**

Spark, ETL Pipeline, AWS



2020 - Now

**Project Manager**

Assess and Manage Data Platform



**Data and  
Business Analyst**

Business Process, Data Mapping, Data Analysis

2018 - 2020

**Bonus**



Earth Observation  
via  
Space Technology



# Analytics Platform Assessment

From the Datawarehouse to the Data Lakehouse



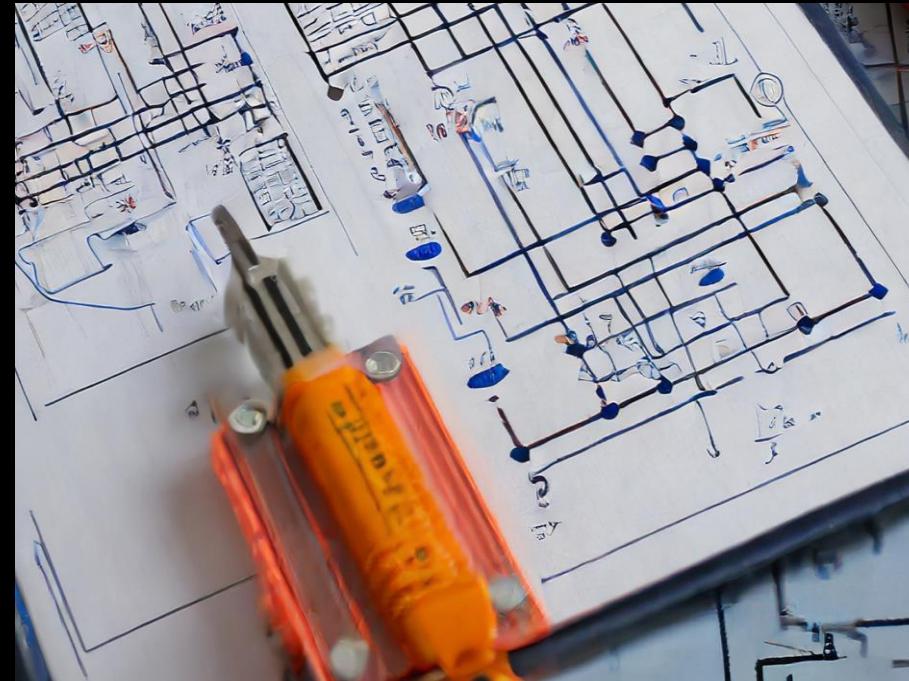
# WHAT IS AN ASSESSMENT

Cambridge Def: *the act of **judging** or deciding the amount, value, quality, or **importance of something**, or the judgment or decision that is made*



# ANALYSIS AS IS SOLUTION

- **Functional organisation** of the data ecosystem
- Logical and technical **architecture**
- Data ecosystem **dimensioning** (cardinality of technological components, data volume, userstiming and SLA/CutOff)
- Detail of the **technology components** (data flows, data source, ETL and transformations, data layerdata, reporting components)

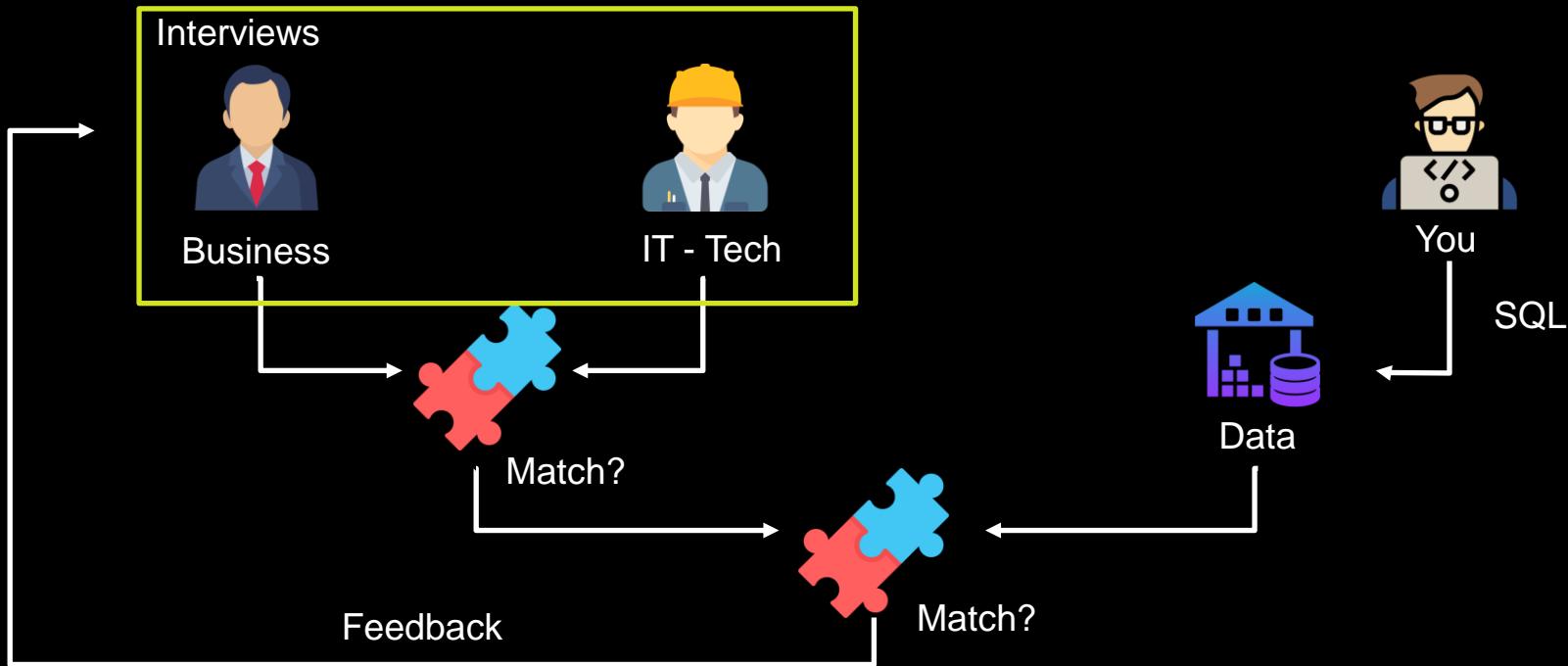




# Functional organization as Silos



# BUSINESS PROCESS AND DATA MAPPING



# ANALYSIS TO-BE SOLUTION

- **Limitations of the current solution** with respect to modern data centric infrastructures (enabling factors to the advanced analytics)
- **Drivers for evaluating** the new solution
- **New use cases** to be enabled



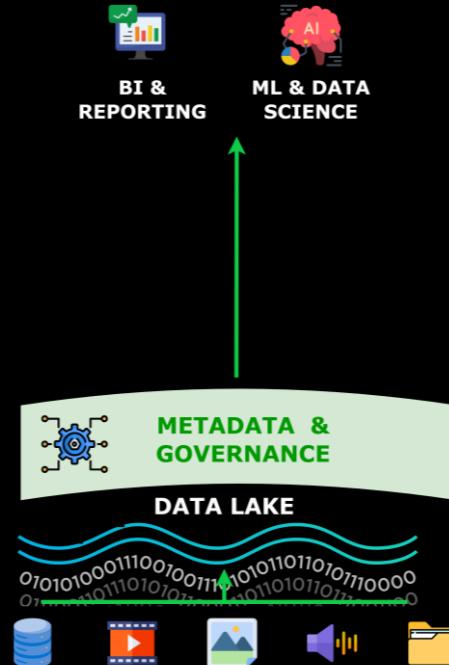
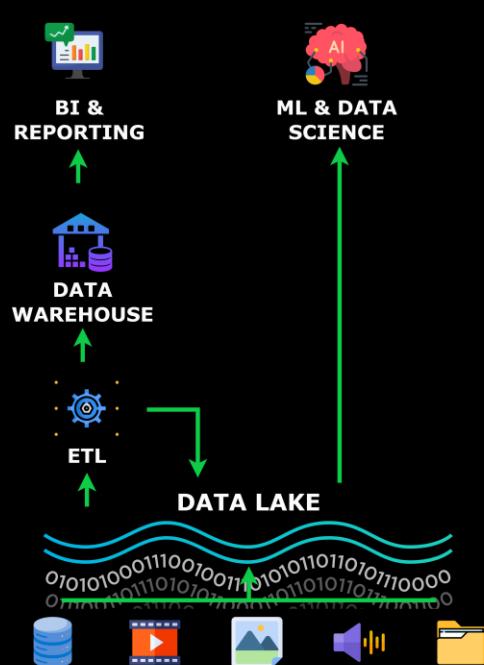
# DESIGN TO-BE SOLUTION

- **Vendor** selection
- Logical and technical **architecture** of the To-Be solution
- **Dimensioning** of To-Be solution components



# DESIGN TO-BE SOLUTION

## SOLUTION DESIGN



First Gen Data Platform

Second Gen Data Platform

Modern Data Platform



# DATA WAREHOUSE VS DATA LAKE

## FEATURES

Data Warehouse	Data Lake
MPP OLAP architecture	Centralized repository
Structured data	Structured and unstructured data
ETL processes feed the centralized dwh with raw data coming from different sources	Store data at one location in an open format ready to be read
BI and batch analytics use cases	BI, batch and real-time analytics, ML use cases
Schema on write	Schema on read
Coupled storage and computation	Decoupled storage and computation
Costly data storage	Cost-effective data storage



# DATA WAREHOUSE VS DATA LAKE

## ISSUES

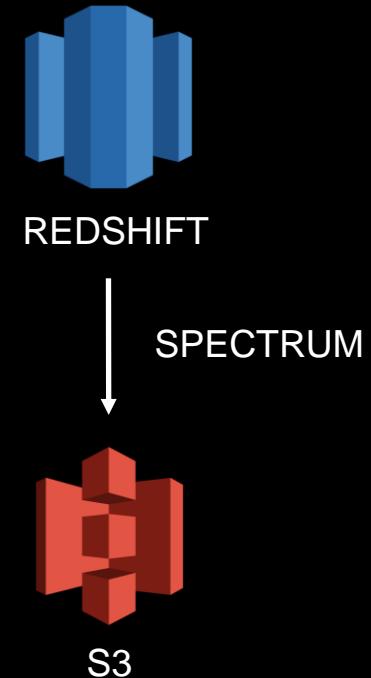
Feature	Data Warehouse	Data Lake
<b>Input data</b>	Not suited for handling unstructured, semi-structured, high variety or high velocity data	Complex ETL pipelines to build the data lake
<b>Storage</b>	Too expensive	Since data is stored in original format, its analyzation requires high technical expertise
<b>Architecture</b>	Tightly coupled storage and processing	Query performance may be degraded
<b>Flexibility</b>	Lack in flexibility: difficult to adapt to new requirements Lots of information not included in the main question difficult to access	Issues with data security and governance Issues with data quality
<b>Use cases</b>	Not the best solution for ML use cases	Risk of dealing with data swamps



# MODERN DATA PLATFORM

## STEPS TOWARDS LAKEHOUSE

- Support for external tables in Parquet and ORC formats
  - ETL complexity, staleness and advanced analytics challenges still remains
  - **Poor performance** because the SQL engine is often optimized to work with its internal data format
- Run SQL queries directly on the data lake storage
  - **DWH are not replaced**: ACID transactions and efficient access methods still missing



# DATA LAKEHOUSE

## FEATURES

- ACID transactions
  - Storage systems used in **Lakehouses does not provide atomic APIs**, but
  - The **Metadata layer** helps manage ACID transactions keeping track of which data in the storage systems are part of a table version
- Cost reduction
  - Data stored in **cheap storage**, such as Amazon S3, Azure Data Blob, etc



# DATA LAKEHOUSE

## FEATURES

- Elimination of simple ETL jobs
  - Query engine **directly connected** to the Data Lake
- Reduced data redundancy
  - Using a single tool to process raw data removes the problems of having to clean and manage data spread on many platforms
- Schema enforcement and **Data Governance**
  - Thanks to a Metadata layer
- Direct connection to BI tools



# DATA LAKEHOUSE

## FEATURES

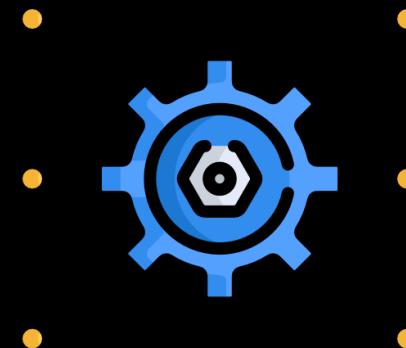
- Unstructured and streaming data support
- Open format support
  - e.g. Parquet, ORC
  - Usable directly by many clients
- **Dataframe** API support
- Storage and processing decoupled
  - **Separation of concern**
  - Scalability



# DATA LAKEHOUSE

## INGESTION AND STORAGE

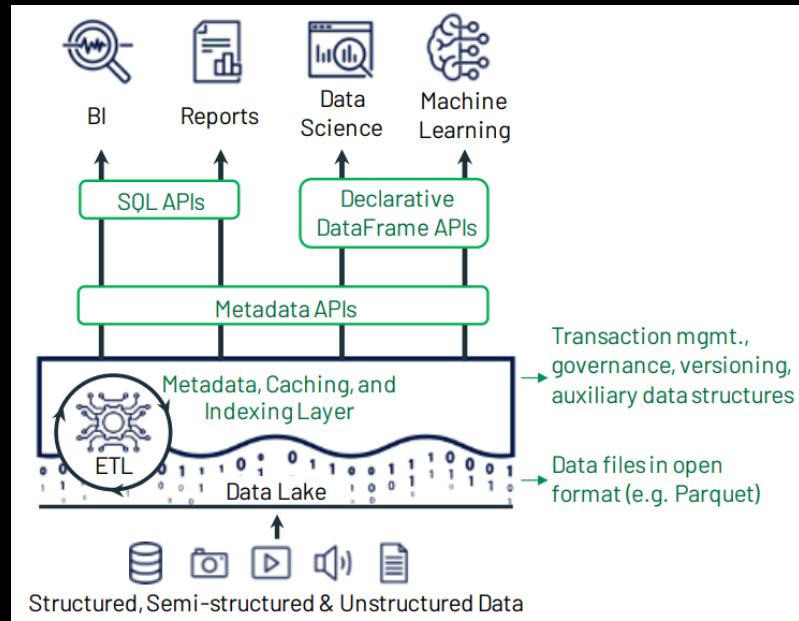
- Ingestion
  - Gets data from various sources and delivers it to the **Data Lake storage**
  - Batch and streaming data processing
  - Connects to internal and external sources
- Storage
  - Low-cost object storage (e.g. AWS S3)
  - Clients can read directly using **open file formats**
  - Multiple clients use the same data
  - Schemas of structured and semi-structured data are **kept in a metadata layer**



# DATA LAKEHOUSE

## METADATA LAYER

- On top of open file formats
- Track which files are part of different table versions
- Offer **ACID-compliant** transactions
- **Time travel feature**
- **Data quality** enforcement features (schema enforcement, custom constraints API)
  - Data that does not match this criteria could be rejected or moved to a dedicated location (Delta)
- **Governance** features (access control, audit logging)

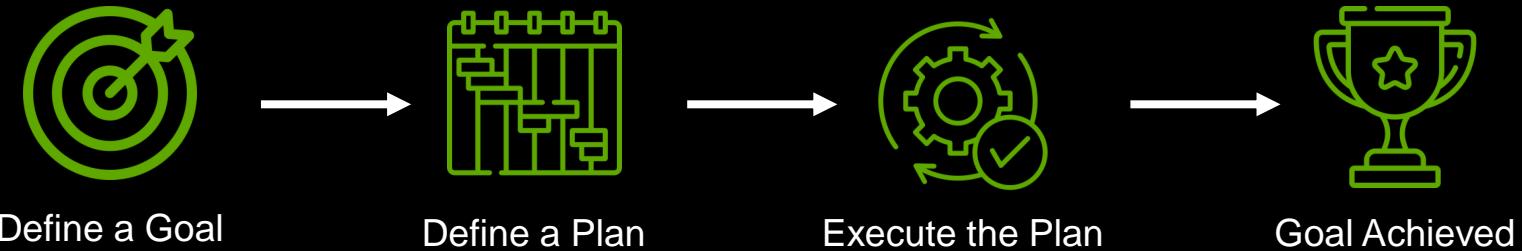


# PROJECT PLAN

- Roadmap of implementation of the To-Be solution
- Project phases
- Milestones and releases
- Project Medotology
- Project team



# PROJECT MANAGEMENT



Define a Goal

Define a Plan

Execute the Plan

Goal Achieved



# PROJECT MANAGEMENT

## IN REAL CASES



Define a Goal

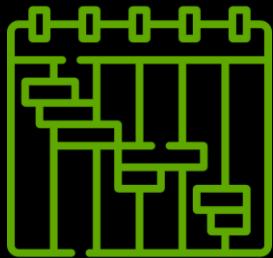
*Are you sure there is a goal?*

- AI and ML subject to often **overblown hype**, no silver bullet to tackle lame processes or lack of data
- **Setting expectations**, correcting them if necessary, helping to address
- AP enabling use case, analysing the GAP
- Focus on **business value**, who pays you if not
- **Challenging but realistic targets**, cost, time and technology constraints



# PROJECT MANAGEMENT

## IN REAL CASES



Define a Plan

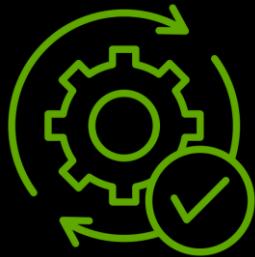
*I have **no idea what we have to do** but we have to say **when we do it***

- Waterfall (**does not work technically**) or Agile (complex - no long-term visibility)? In the reality mix of the two, what changes (or should change) is the scope -> difficult realization
- Define priorities, business value? Yes but given technical platform issue, **promote E2E flows**
- Delivery VS Demand



# PROJECT MANAGEMENT

## IN REAL CASES



Execute the Plan

*Try to deliver something  
helpful*

- **Change** more or less everything compared to original plan
- **Change** Scope
- **Change** Priority
- **Change** Timeline
- **Change** resources (Senior Dev who knows everything quit)
- ...



# PROJECT MANAGEMENT

## IN REAL CASES



Goal Achieved

*Check if really helpful*

- Just the first (or one of) iteration
- Define next steps
- Application Maintenance
- Tech Debt Analysis
- Retrospective (why we have achieved some results and others have not)



# THANK YOU

[www.reply.com](http://www.reply.com)