

BitBang

empowering data driven
decision-making

Digital Transformation in Luxury Industry

12/12/2022

Agenda

- Brief introduction
- Digital Transformation Initiative
- Databricks Overview
- Product Recommendation Use Case

Introductions



Erica Maccaferri

Head of Data Engineering & Machine Learning

BitBang



Who we are

BitBang was founded in 2003, with the vision that data is the most powerful asset of a business.

For over a decade, we have been providing data management consulting services and strategies to empower **Data Driven Decision-Making at Scale.**

Our services and solutions will help you transform your business through orchestrated executable data strategies that achieve desired business outcomes.

A photograph of a mountain climber in a red jacket and black pants standing on a dark, rocky ledge. They are pointing towards a vast, white, snow-covered mountain range. In the background, a large glacier with deep blue ice is visible, and a long line of climbers is seen walking across a snowy slope.

Mission

Empowering **business transformation**
and **data-driven decision making**
through **data strategy, data execution**
and **insights delivery** practices **at scale**

Digital Transformation Initiative

The Company

- A group build in a century
- Different brand with different consumer target
- Global presence with 700 boutique
- Grown thanks to acquisition
- Market differentiation

2016

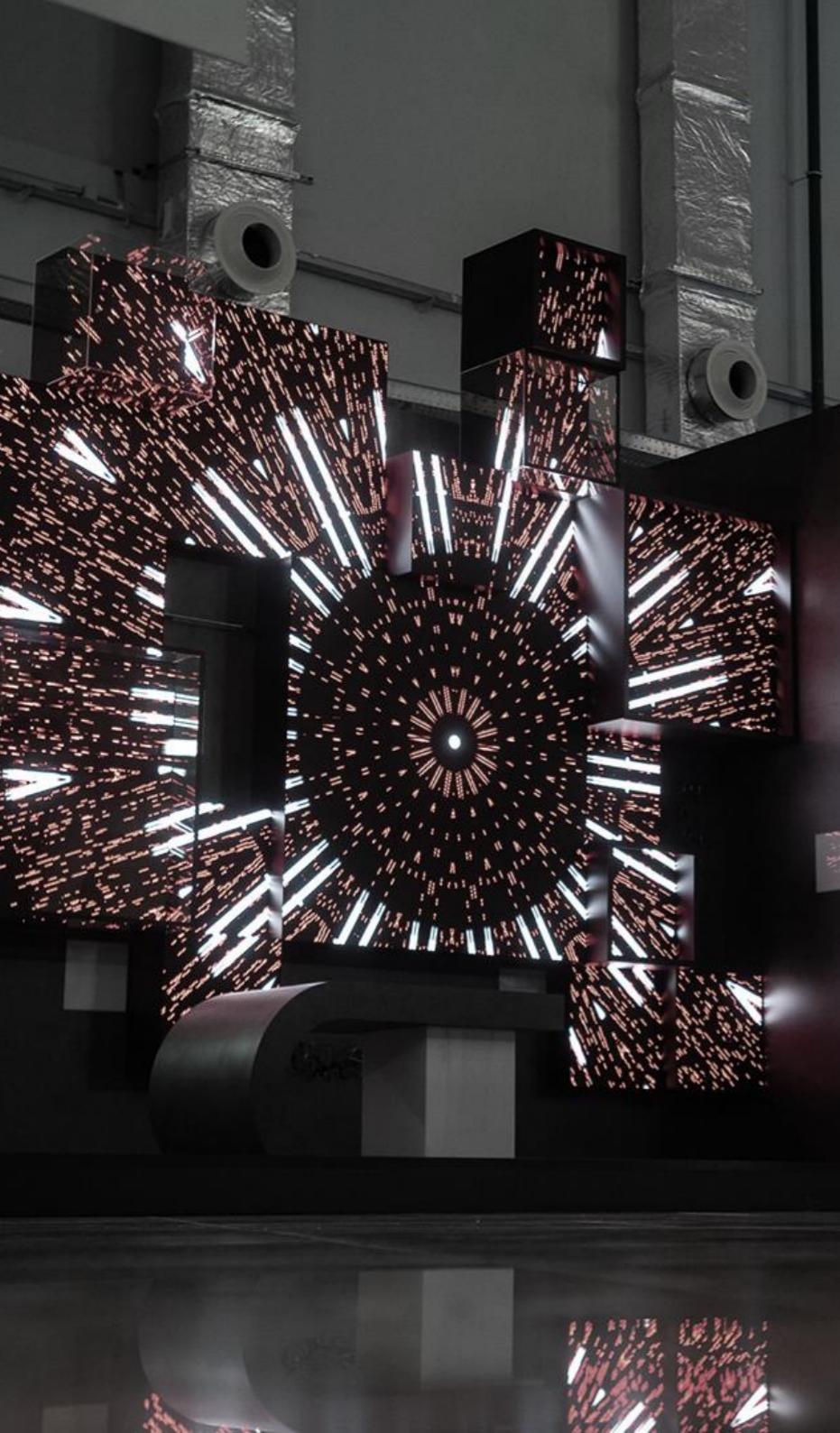
is the first time in which
customer can find company
products online

Web Impact on Retail



FASHION AND LUXURY

Brands need to provide the same special experience customers get in boutique in the digital world



People buy Experiences Not products



Every **touchpoint** matters and generates data



Measure the **experience** using data



Create a unique experience, **beyond people's expectations**, starting from data and analytics



Leverage Machine Learning and Artificial Intelligence for automation and **real time personalization**

**Every digital transformation is going
to begin and end with the customer,
and I can see that in the minds of every CEO I talk to.**

MARC BENIOFF, CHAIRMAN AND CO-CEO, SALESFORCE

Objective

The CIO has a good reputation with the company owners and decided to face the new era of retail investing in a digital transformation program.

The idea was to build a wide Digital & Customer Intelligence practice to guide the company digital strategy on every channel:

- | | |
|--|--|
| ■ Cross-Channel Campaign Management | Adobe Campaign |
| ■ Marketing Data Lake | Azure Data Platform |
| ■ Integrated Data Collection | Tealium (Enterprise Tag Manager) |
| ■ Digital Analytics | Adobe Analytics |
| ■ Marketing Personalization | Adobe Audience Manager (DMP) |
| ■ On-site Personalization | Adobe Target & usage with DMP Audiences |
| ■ Voice of Customer | Qualtrics |
| ■ Identity Management | Gigya & In-store Recognition with Cloud4Wi |
| ■ Advanced Attribution & Optimization | New Custom Attribution Project |

Abundance of variables





CUSTOMER VIEW



DATA



INFORMATION



KNOWLEDGE

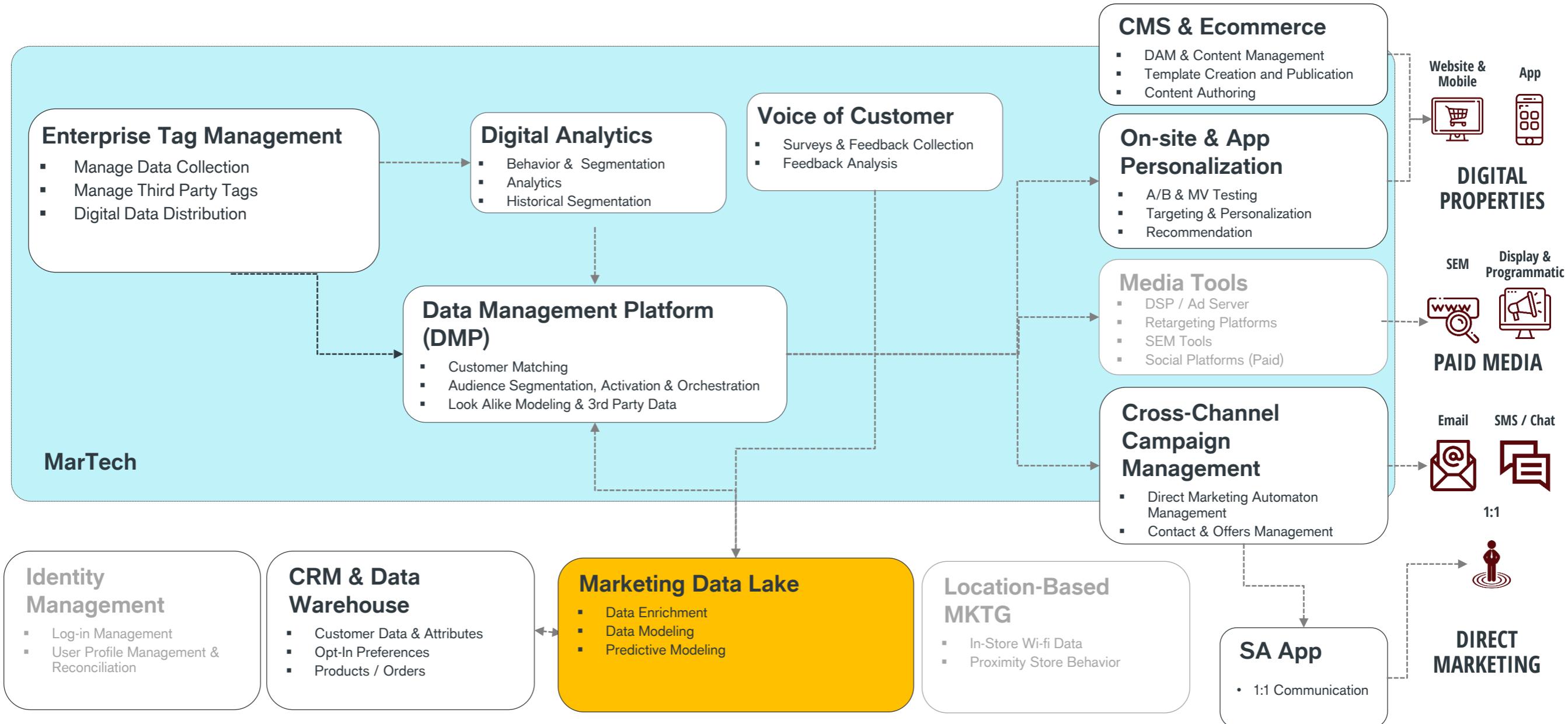


INSIGHTS



360°
CUSTOMER VIEW

Digital & Customer Intelligence Tech Blueprint



MarTech Glossary

Tag Management System

A tag, sometimes called a pixel, is a piece of JavaScript code used to **collect first-party data**. A TMS makes it simple for users to implement, manage, and maintain tags on their digital properties with an easy to use web interface.

Data Management Platform (DMP)

A DMP collects, organizes, and activates first-, second-, and third-party data from online and offline sources. It then uses that data to build detailed anonymized customer profiles that drive targeted advertising and personalization initiatives.

Cross-Channel Campaign Management

Enterprise marketing technology that supports customer data management, analytics, segmentation, and workflow tools for designing, executing, and measuring campaigns for digital and offline channels.

Organization and Technology

Technological point of view...

integrate touchpoint data into a single data platform, ideally in real time



Organizational point of view...

demolish the classic "silos" structures, which still characterizes many companies today



Ensure KPIs and internal verification processes are **consistent** with the Omnichannel Customer Experience Objectives and Strategies



Create a robust strategy that matches KPIs and budget / targets and implement a single point of reference across the business

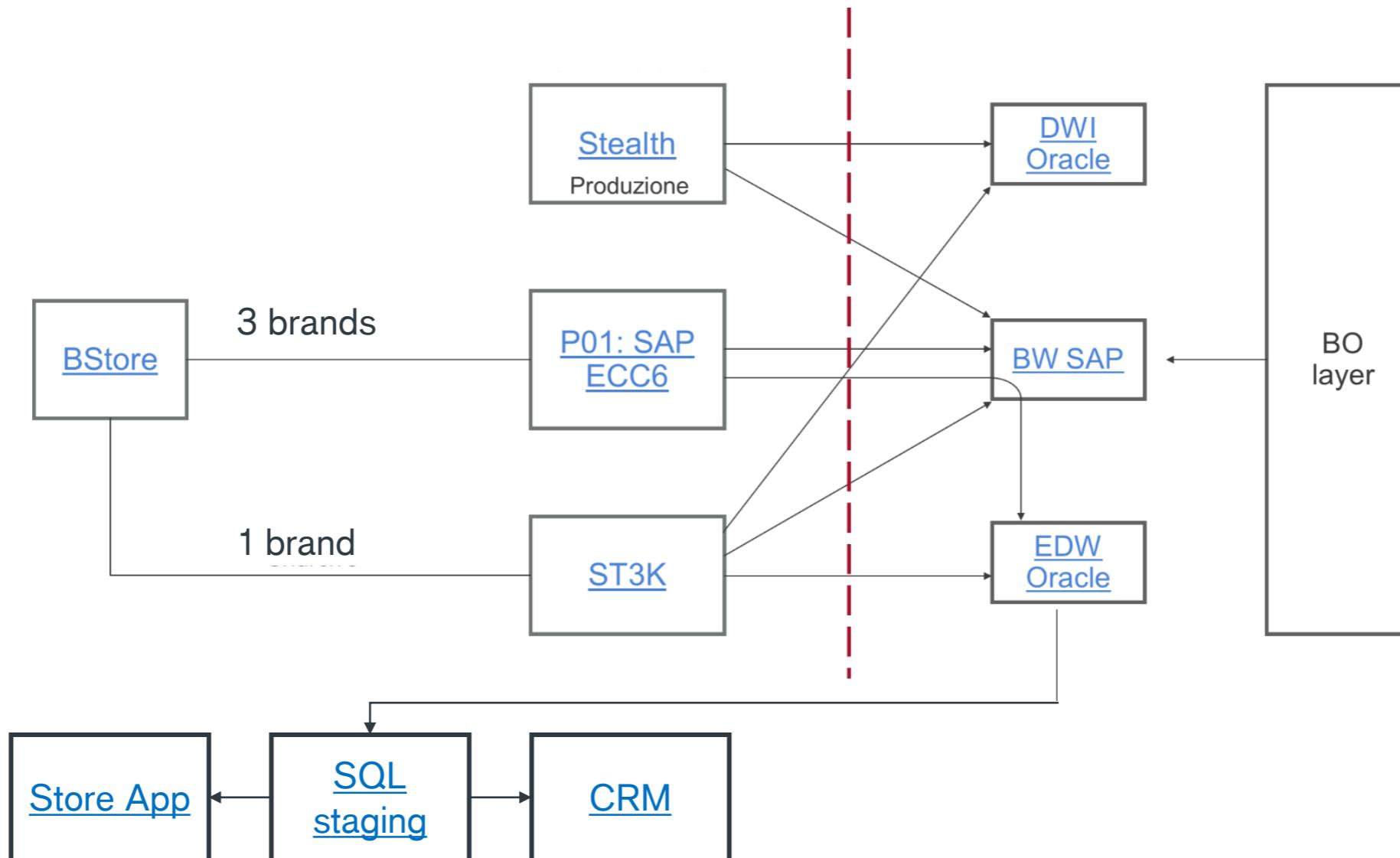


2018

Pain-points

- Business and marketing suffered of IT and BI bottleneck
- Data scientist had no environment to access data and build model (just their own notebook and manual extraction from BI tool)
- Direct marketing campaign were not orchestrated
- Marketing campaigns could not be personalized based on retail data (the most important for the business)

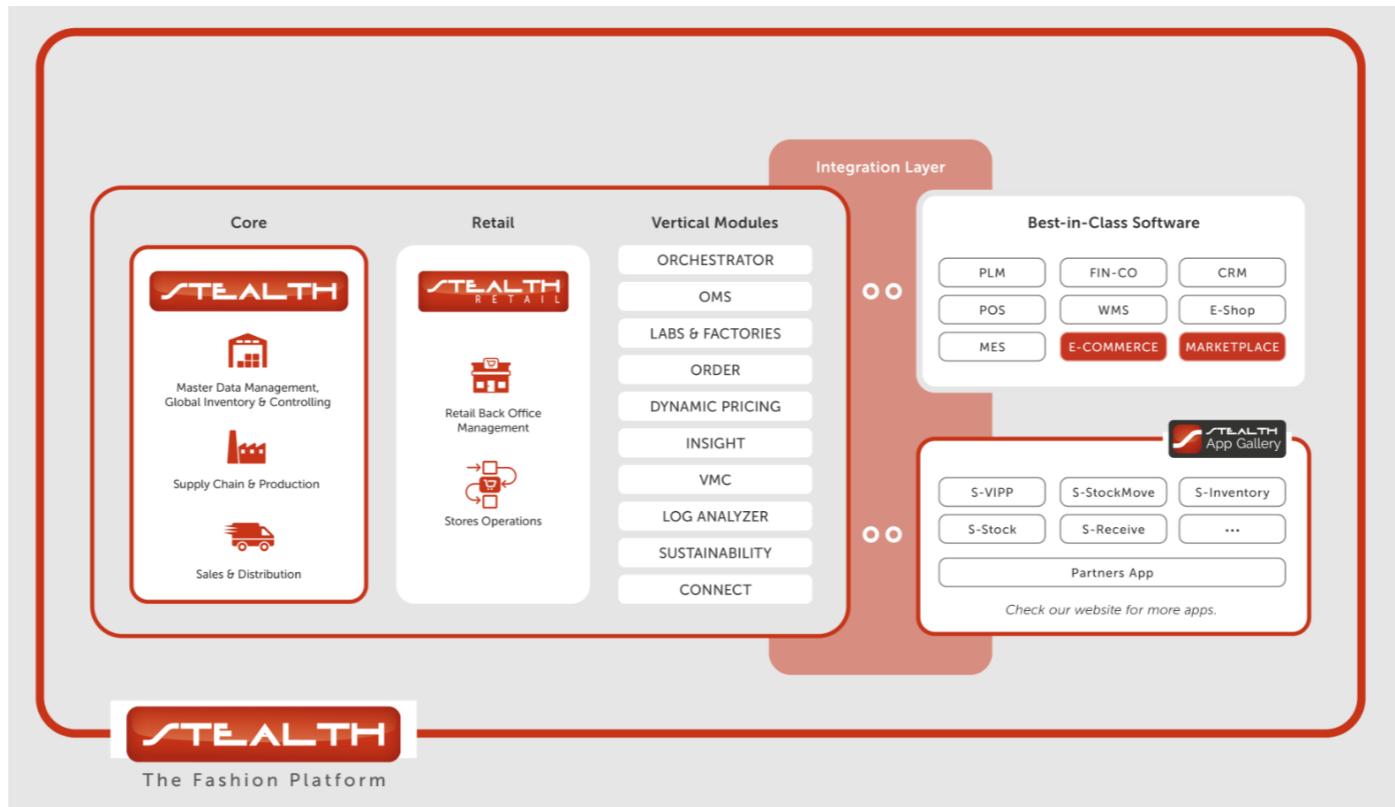
Simplified AS-IS Architecture



Stealth

Stealth®, the Fashion Platform, is the solution designed for managing all the typical processes of the Fashion and Luxury sector.

It checks on all the processes related to active and passive cycles of the Fashion Industry, such as manufacturing, Supply Chain, sales, and distribution. The internal Master Data eases the constant monitoring of the distributed manufacturing processes and the management of brand, language, and company trends.

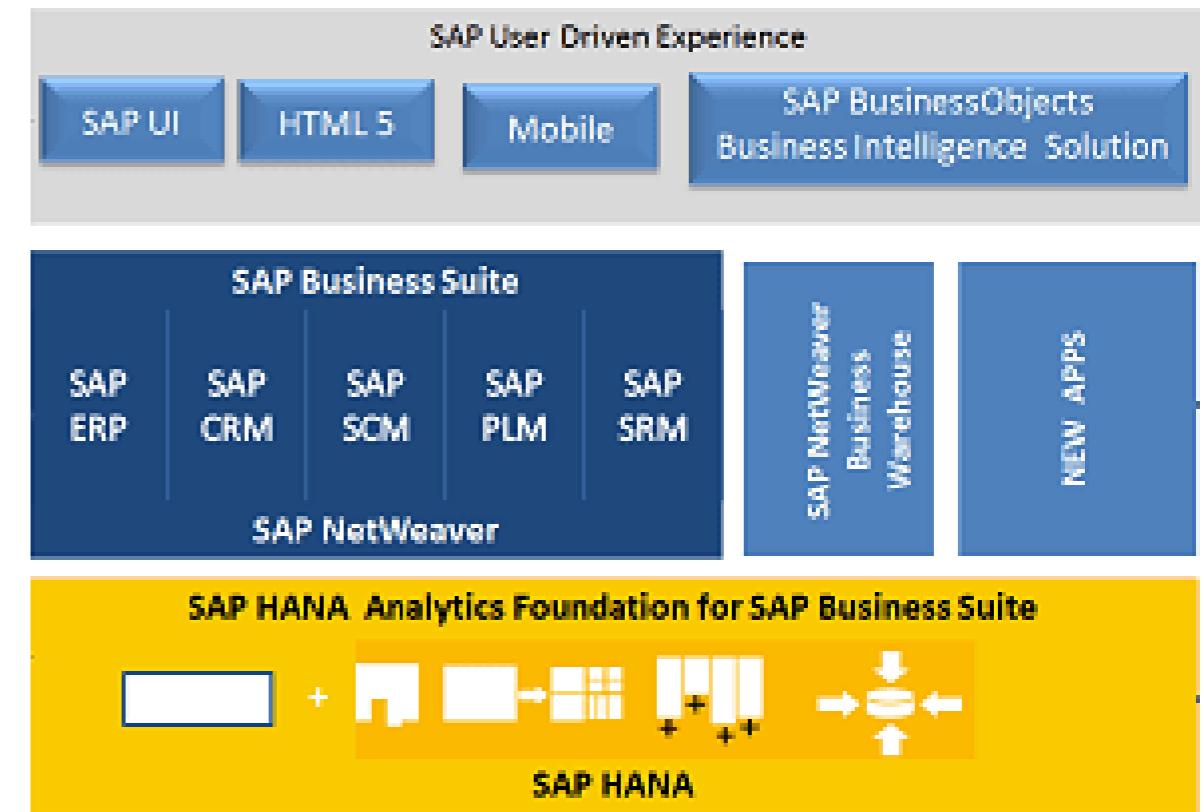


SAP ECC

SAP ECC is the largest system element of the Business Suite and stands for SAP enterprise resource planning (ERP) central component.

This central component integrates digital information from all of the various areas of an organization in real-time, so decision-makers can get a holistic view of the entire organization's operations and make decisions based on this information.

One of the brand do not use SAP ECC
but a module called ST3K

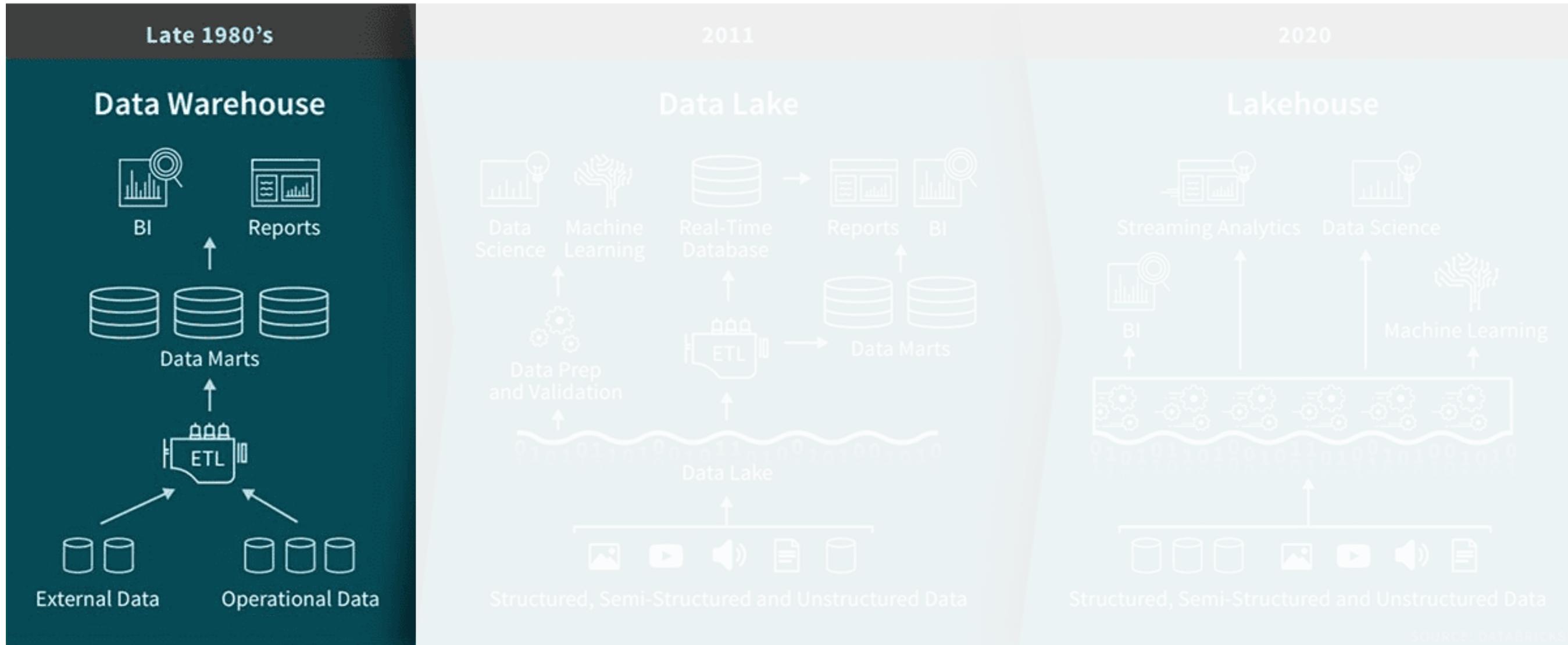


SAP business suite can operate on any database including HANA.

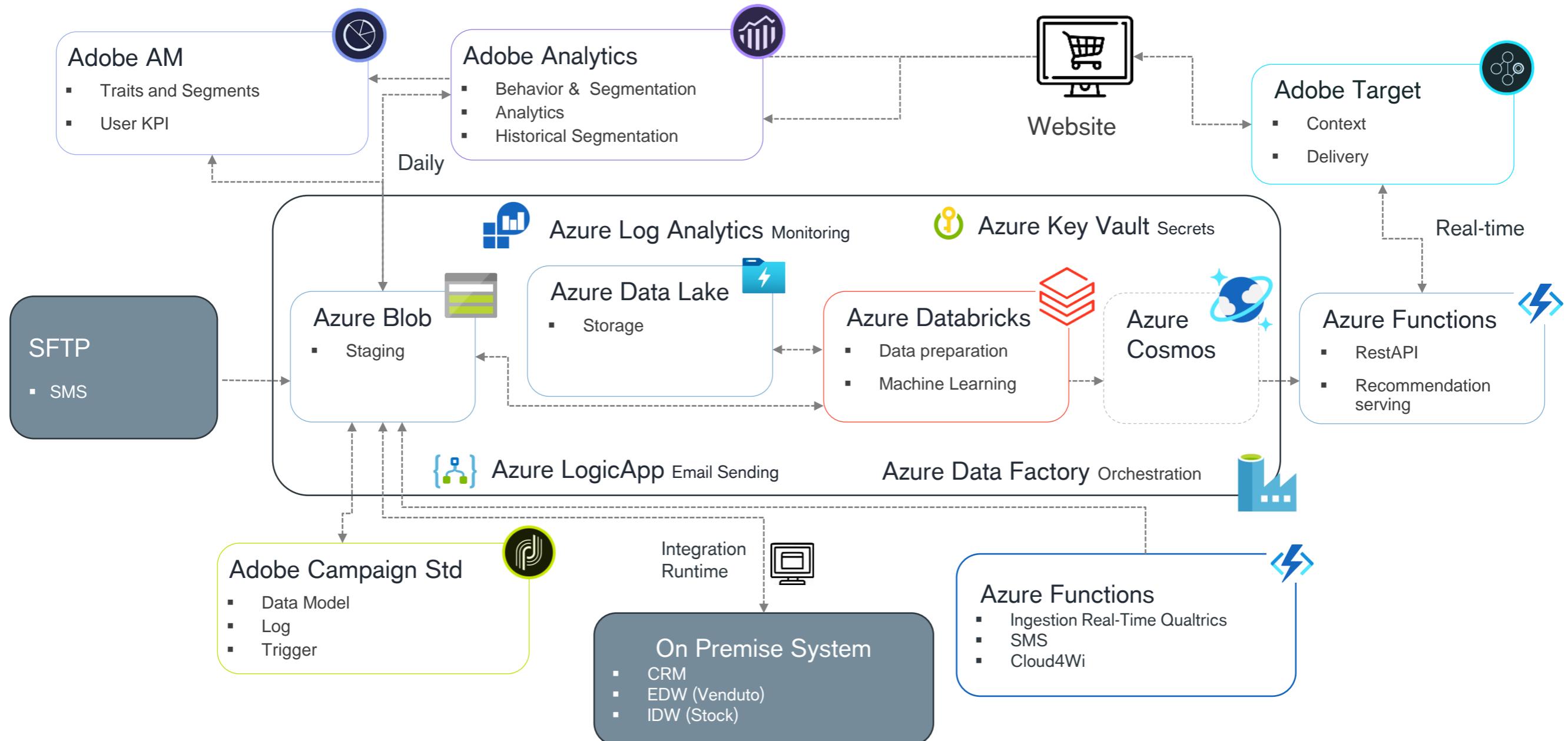
BStore

- Point-of-service platform
- Planned migration to XStore.
- All brands register sales in this system

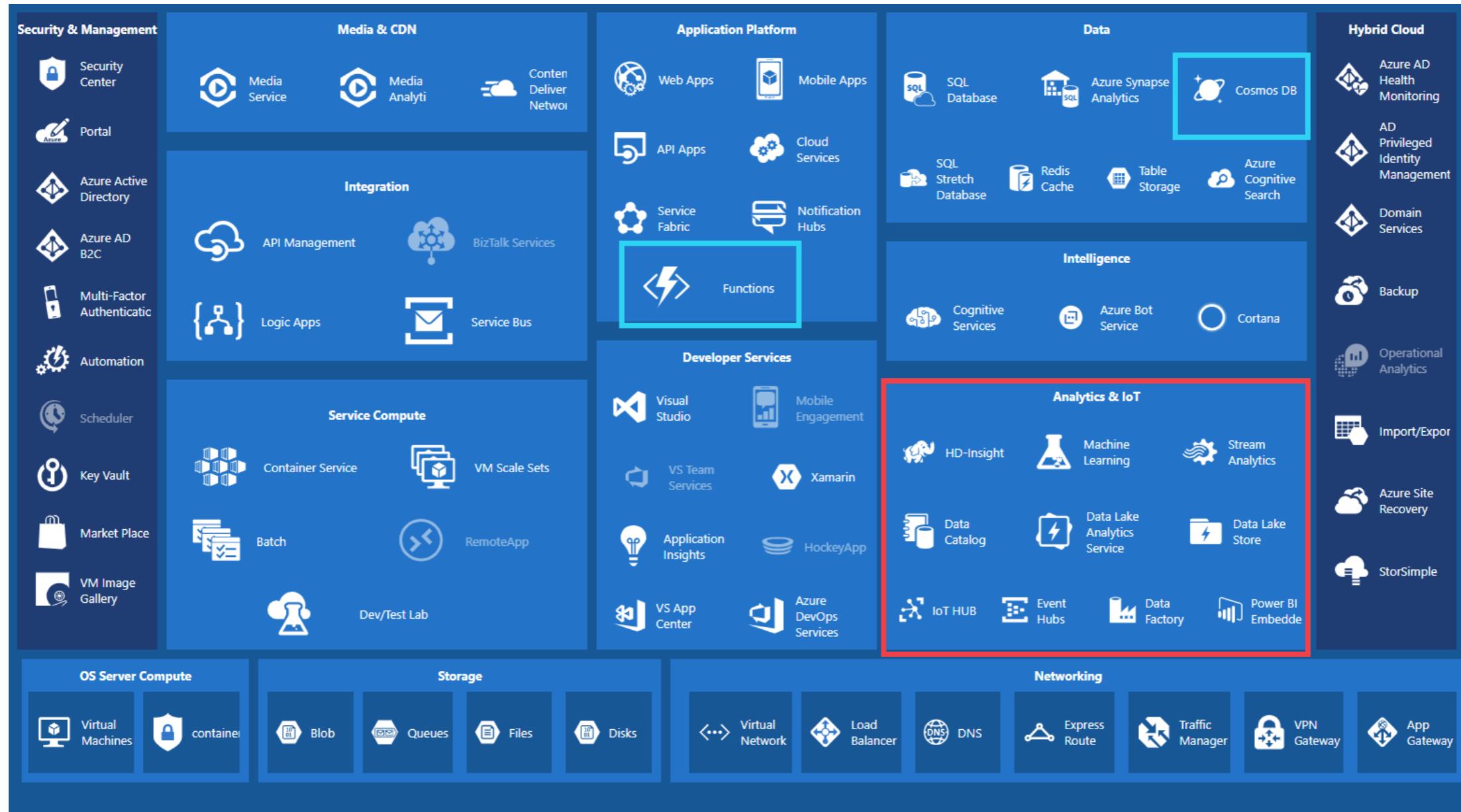
AS-IS



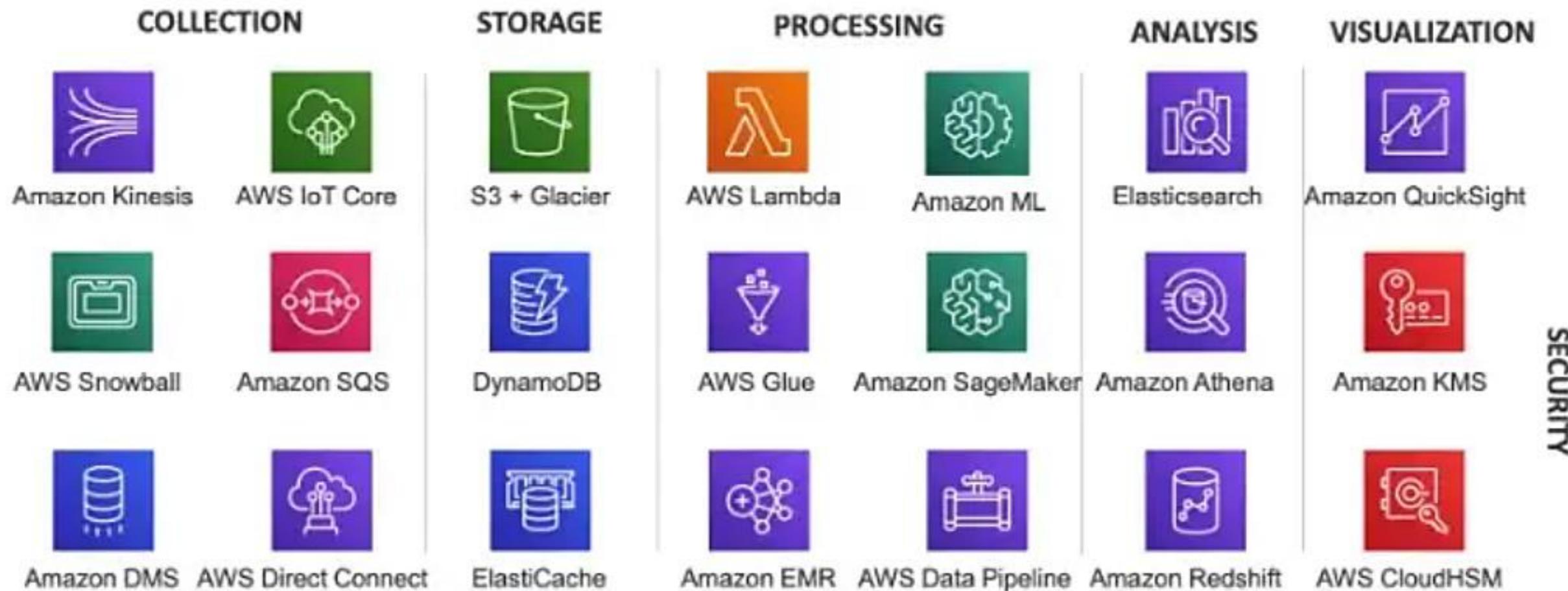
Data Platform focus on Azure



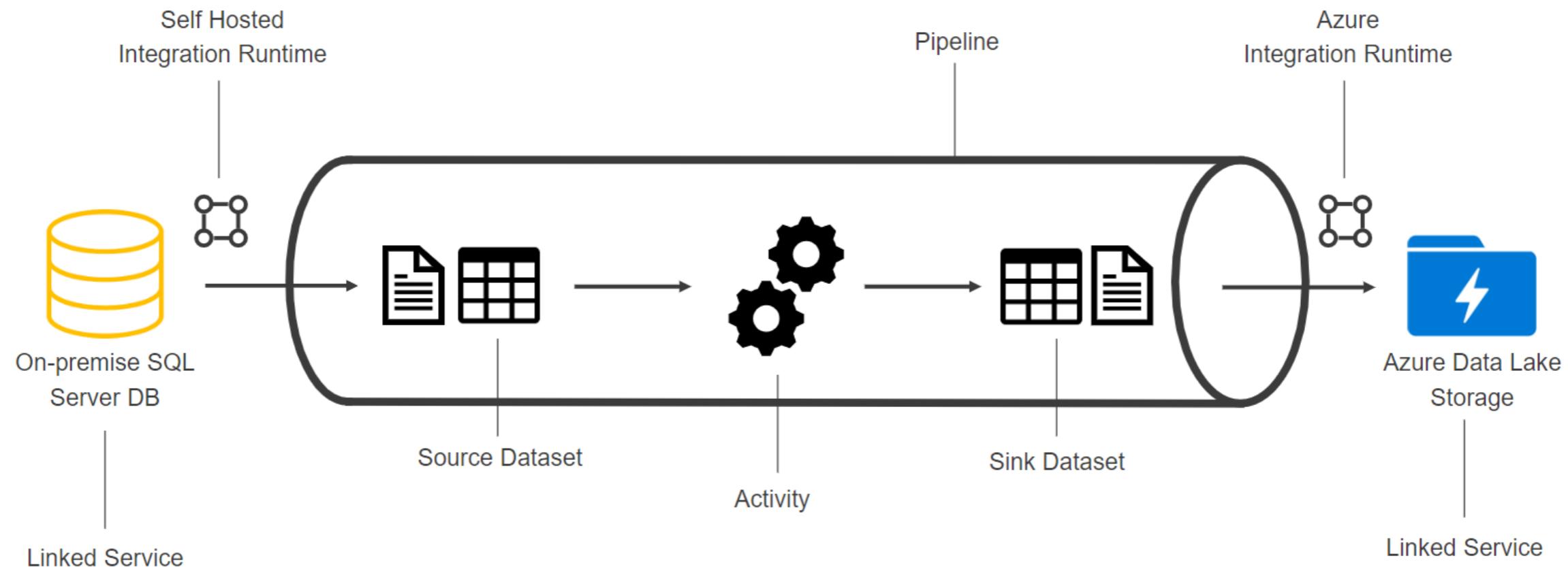
Azure Services



AWS – Data Services



Data Factory – Data Ingetion



Storage Account and Data Lake Gen 1

Storage Account is a low cost, distributed storage (as S3) that provides different format such as:

- Object
- Files
- Queue
- Table (key value)

Data Lake Gen 1 (retirement in 2024) is the Microsoft version of HDFS

Why Azure Databricks and not HDInsight

Databricks was the most popular commercial services around the Apache Spark data analytics platform

Microsoft and Databricks have actually worked on this integration and Databricks a first-party service on Azure. It means that users will be able to spin up Databricks in the Azure Portal just like they would any other native Azure service and benefit of deep integrations with all of the other Azure services, Azure SLAs and MS support.

HDInsight required a deeper knowledge of DevOps to keep the cluster up and running and its Spark is less optimized.

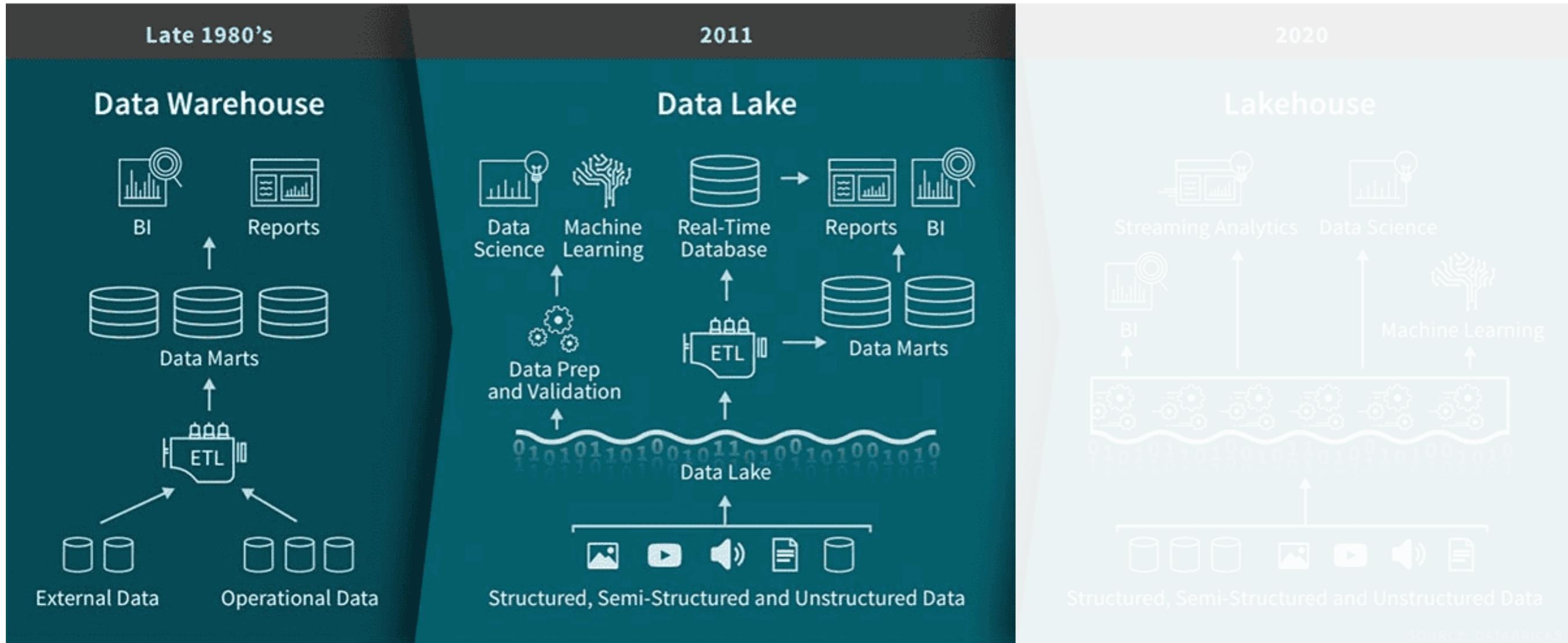
Data Lake Organization

Raw data layer – also called the Ingestion Layer. The main objective is to ingest data into Raw as quickly and as efficiently as possible. To do so, data should remain in its native format. We don't allow any transformations at this stage. With Raw, we can get back to a point in time, since the archive is maintained.

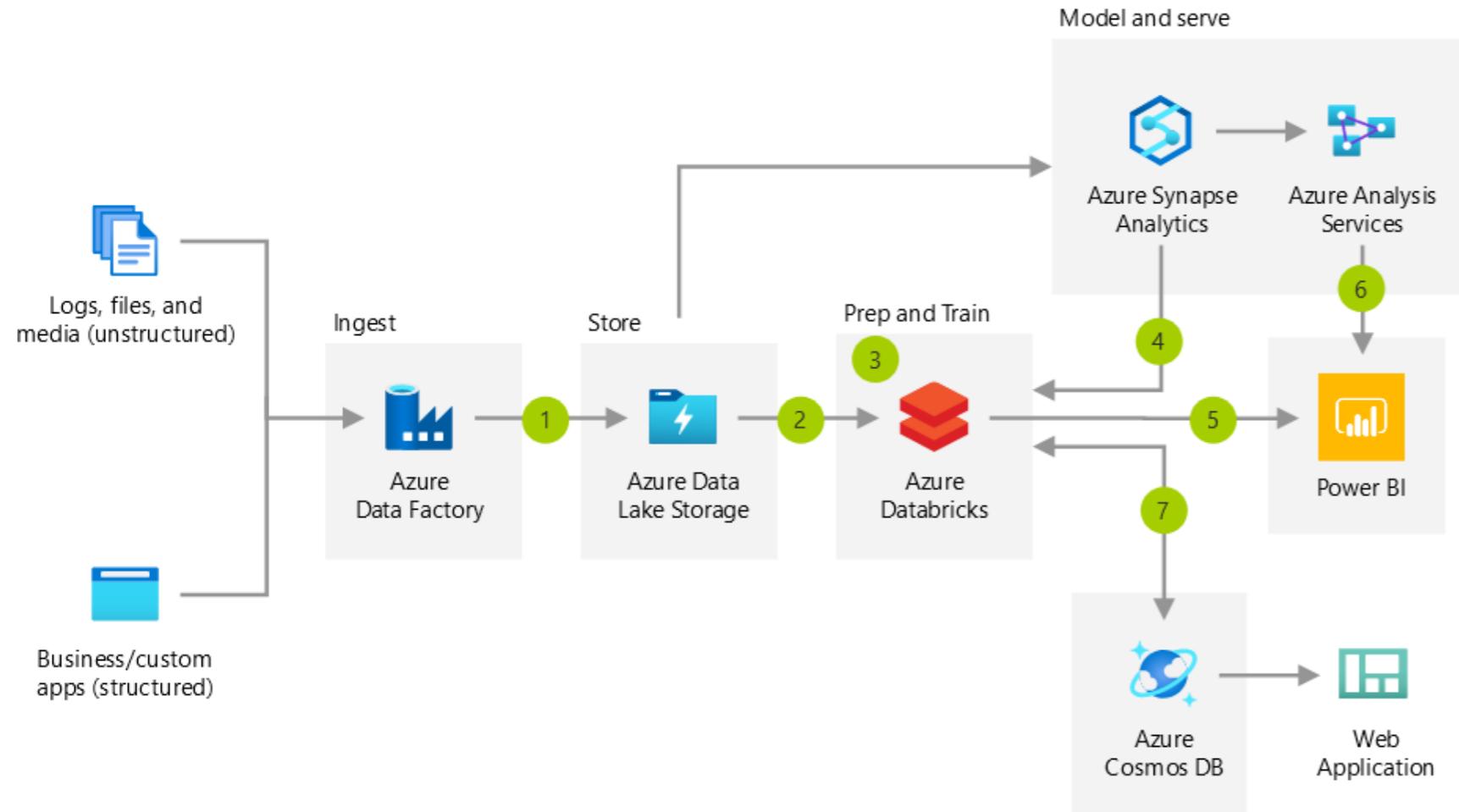
Cleansed data layer – also called Curated Layer/Conformed Layer. Data is transformed into consumable data sets and it may be stored in files or tables. The purpose of the data, as well as its structure at this stage is already known. You should expect cleansing and transformations before this layer. Also, denormalization and consolidation of different objects is common.

Application data layer – also called the Trusted Layer/Secure Layer/Production Layer, sourced from Cleansed and enforced with any needed business logic. If any of your applications use machine learning models that are calculated on your Data Lake, you will also get them from here. The structure of the data will remain the same, as in Cleansed.

Data Lake



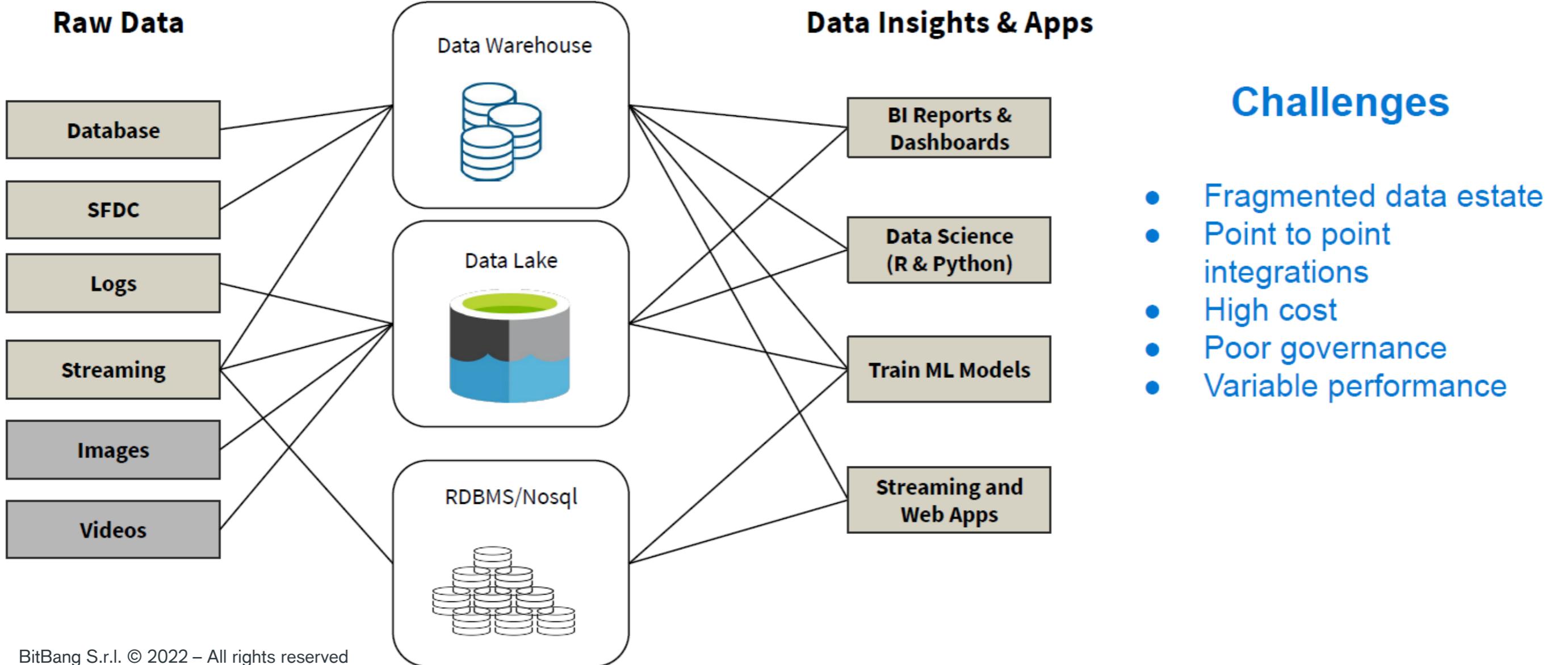
Azure Architecture Blueprint



The CIO would like to protect BI role in the company and decided not to reach this design:

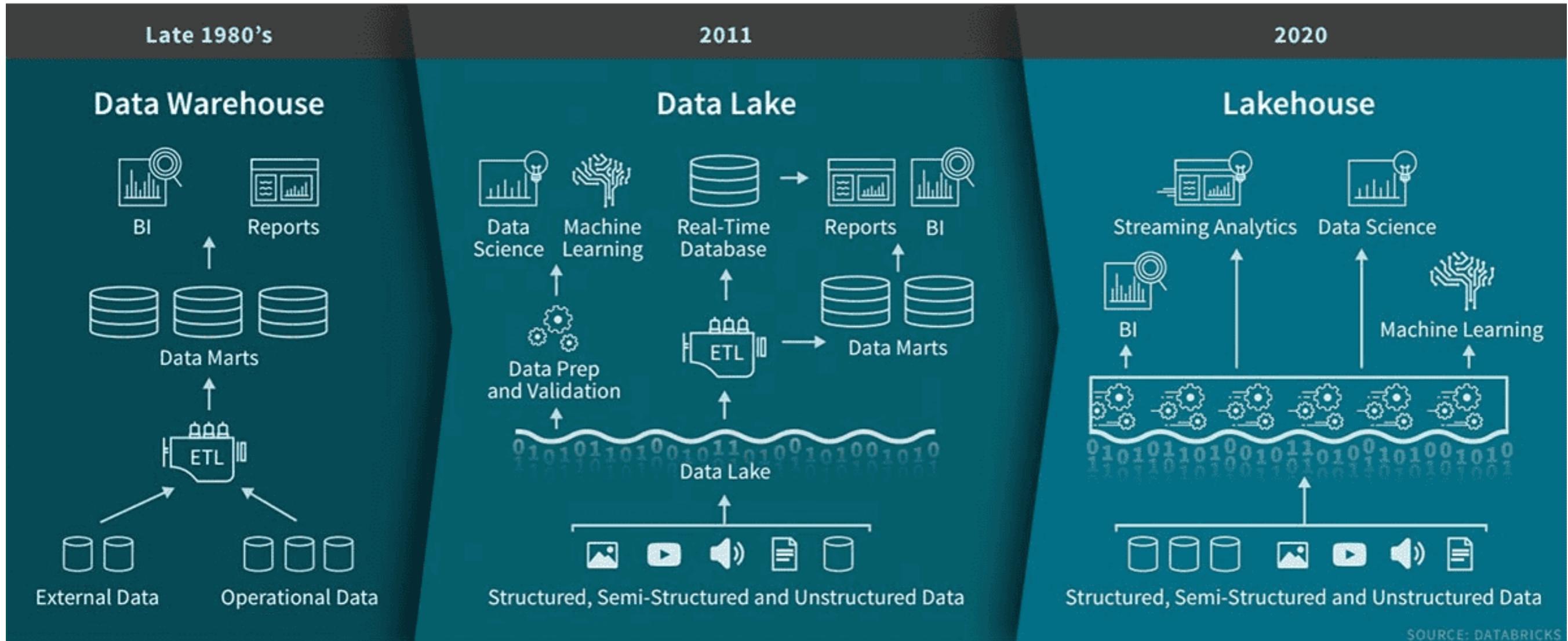
- BI rely on file moved from data lake to a shared folder
- Databricks could not compute or update KPI already used in reports
- There are no pipeline to the DWH

Disjoint Data Landscape



2019-2022

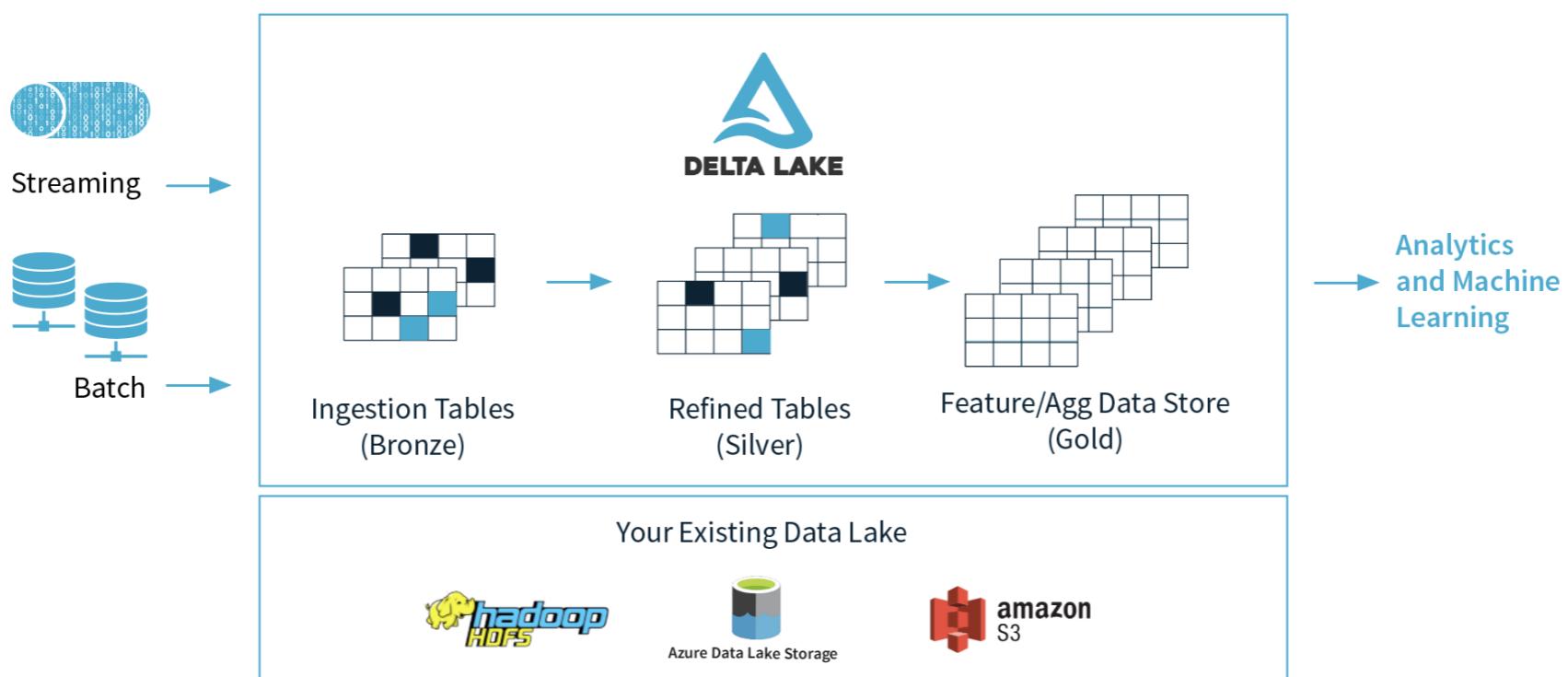
Data Lakehouse



Delta Lake

Introduced in April 2019, [Databricks Delta Lake](#) is, in short, a *transactional storage layer* that runs on top of cloud storage such as [Azure Data Lake Storage](#) (ADLS) Gen2 and adds a layer of reliability to organizational data lakes by enabling many features such as ACID transactions, data versioning and rollback.

Driven by GDPR we migrated from simple parquet file to delta lake tables.



The New Databricks

Platform for BI, Reporting, Data Science, ML

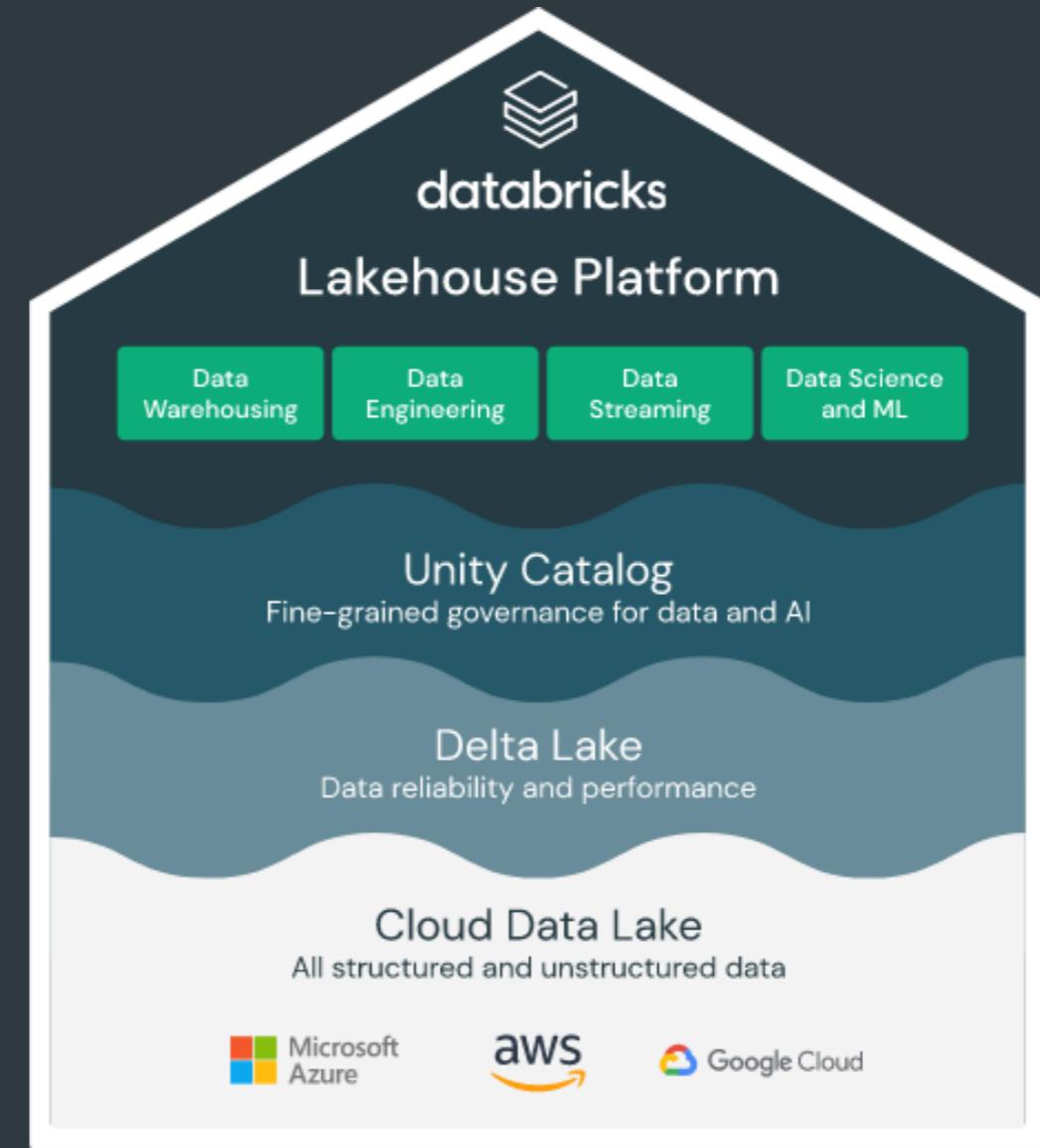
- Collaborative environment for Analysts, Data Scientists and Data Engineers
- Support for leading languages and frameworks

Structured transactional layer on the lake

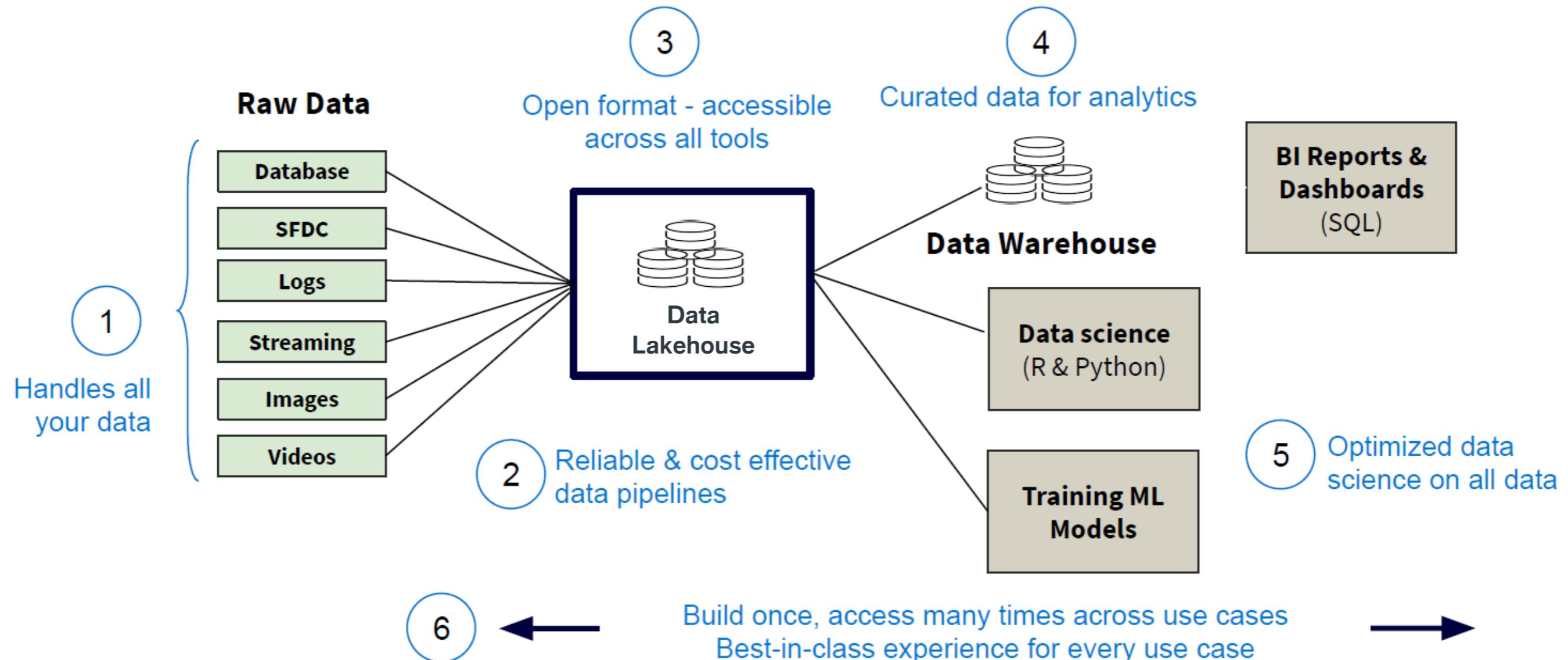
- Support for batch and streaming
- Open format on top of data lakes with no lock-in

Your data lake for all your data

- Cheap durable cloud storage (S3, ADLS)
- Support for all data types



Integrated Data Landscape



We are still adopting the news framework

With the arrival of a new CxO the CIO was not leading the Digital Transformation Program

Budget was siloed and few people could work on analysis

Project were not planned but developed with business urgency and PoC were kept in production widening technical depth

CIO resigned and the head of BI took his role in the project, avoiding decisions and to conflict with BI project

The new CIO now is aware of the missed opportunity they have due to a slow adoption of the technology

Databricks Overview

Databricks Community Edition

Community Edition Terms of Service

Welcome to Databricks Community Edition! We are pleased to provide [Databricks Community Edition](#) (the “Community Edition Services”) at no charge to those interested in learning and exploring the use of Databricks’ cloud-based data analytics platform, which enables data analysts and others to easily tap the power of Apache Spark and Databricks’ other proprietary functionality.

There Are Strict Limits On What Your Data Can Include. In order for us to provide the Community Edition Services to you at no charge, we have implemented certain cost saving elements within the architecture of the Community Edition Services including, among other things, the use of a multi-tenant environment with limited data security protections. In addition, Databricks personnel have generally unrestricted access to your account (“**Your Account**”) and any data used or exposed to the Community Edition Services for the purposes of monitoring and improving the quality of the service. Therefore, you should have no expectation of privacy regarding the data you submit or otherwise make available in any way to the Community Edition Services (collectively, “**Your Data**”)



Databricks

One open, simple platform to store and manage all of your data for all of your analytics workloads

Who they are

Vision

Accelerate innovation by unifying data science, engineering and business

Solution

Unified Analytics Platform

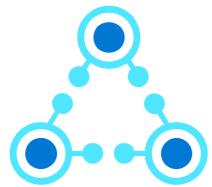
Original creators of:



5000+ global companies use the platform across big data and machine learning lifecycle

Databricks business value at a glance

Data engineering, management and transformation



Lakes



Warehouses



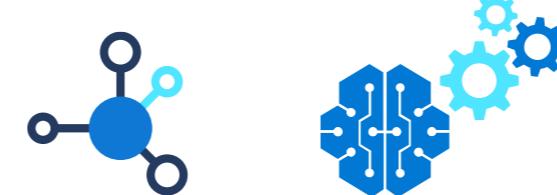
Streams



NoSQL



Data Science, Machine Learning and Deep Learning



- Recommendation and personalization engines
- Risk, fraud, intrusion detection and prevention
- Customer 360 engagements, ad targeting
- Inventory asset optimization and allocation
- Genomics and DNA sequencing
- Predictive maintenance, SCM seasonal costing
- Sentiment and customer churn analysis
- Security compliance and intelligence

Business Intelligence



- Enable SQL analyst to access data lake for exploration and report design
- Modern BI tool get access to a single source of truth



Azure
Databricks



DS Workspace

Collaborative notebooks,
production jobs



SQL Analytics

Collaborative BI

Databricks Runtime



Databricks Cloud Service

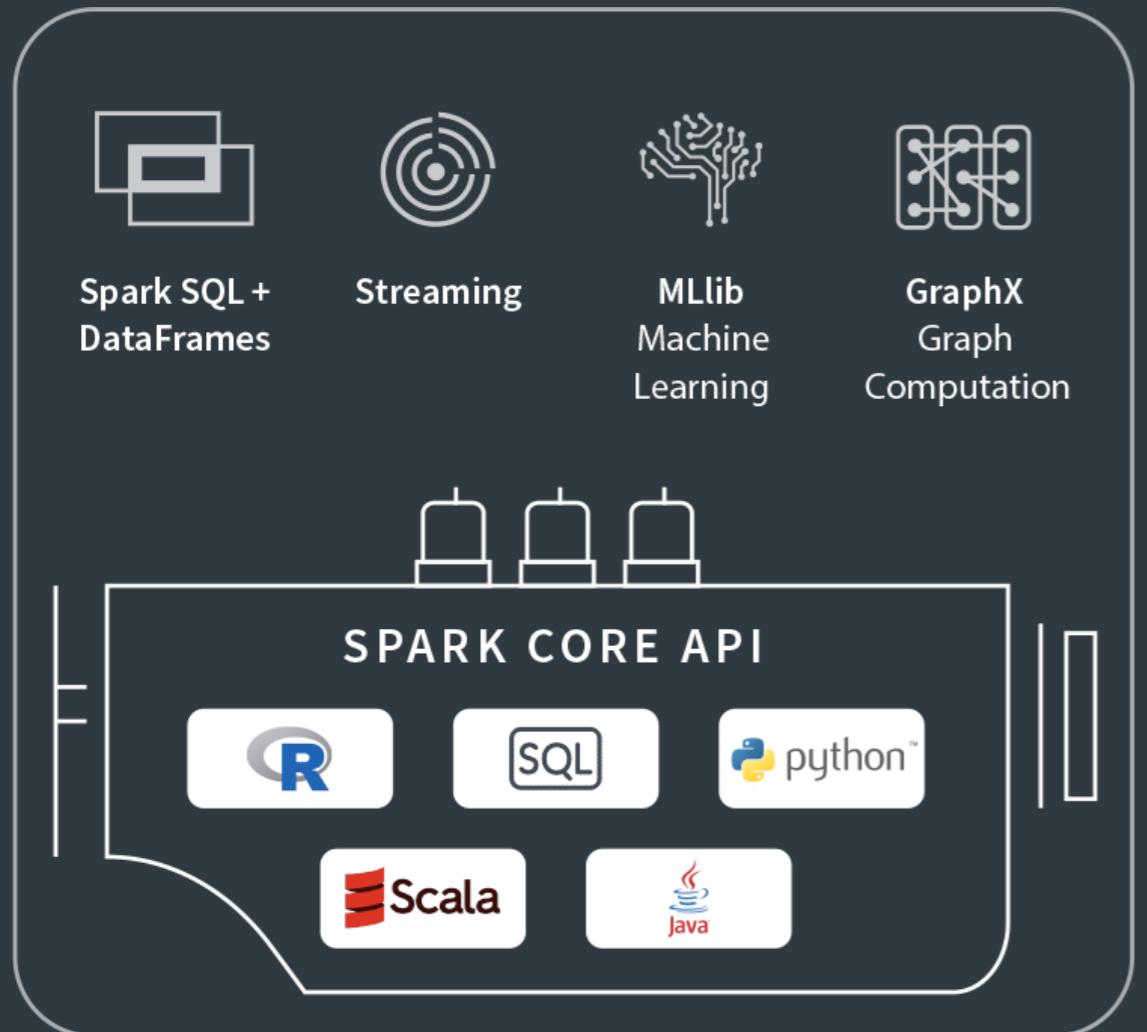


Microsoft Azure



* Microsoft Azure First Party Services

Databricks Runtime



Data processing engine built by the **best Spark developer** (75% of the open source code):

- A highly optimized version of Apache Spark: it means that jobs are **more likely to successfully give results**, even written by not very skilled Spark users
- It always **up to date with last release**
- Auto-scaling infrastructure for **easy self-service** without DevOps
- Better management of data (es. simplifies **GDPR Compliance**) thank to Delta Lake, that provides update, deletes and transactions
- Optimized for **Business Intelligence**
- **Time travel** on Delta Tables

Features

Caching: Copies of remote files are cached in local storage using a fast intermediate data format, which results in improved successive read speeds of the same data.

Z-Order Clustering: The colocation of related information in the same set of files dramatically reduces the amount of data that needs to be read, resulting in faster query responses.

Join Optimizations: Significant performance gains are possible with range join and skew join optimizations through different query patterns and skew hints.

Data Skipping: Statistical information on minimum and maximum values automatically collected when data is written is used at query time to provide faster queries.

Easy-to-Use Cluster Management: User-friendly user interface simplifying the creation, restarting, and termination of clusters, providing increased visibility to your clusters for easier manageability and cost control.

High Availability: The Databricks cluster manager transparently relaunches any worker instance that is revoked or crashes, ensuring your service is always up and running without the need to manage it yourself.

Elastic On-Demand Clusters: Build on-demand clusters in minutes with a few clicks, and scale up or down based on your current needs. Reconfigure or reuse resources as needs change for your team or service.

Backward Compatibility with Automatic Upgrades: Choose the version of Spark you want to use, ensuring legacy jobs can continue to run on previous versions, while getting the latest version of Spark hassle free.

Flexible Scheduler: Execute jobs for production pipelines on a specified schedule, from minute to monthly intervals in different time zones, including cron syntax and relaunch policies.

Notifications: Notify a set of users whenever a production job starts, fails, and/or completes with zero human intervention, through email or third party production pager integration, for peace of mind.

Flexible Job Types: Run different types of jobs to meet your different use cases, including notebooks, Spark JARs, custom Spark libraries and applications.

Optimized Data Sources: Central repository for your Spark Data Sources, with broad support including SQL, NoSQL, Columnar, Document, UDFs, File stores, File formats, Search engines, and more.

Databricks Runtime - Types

Focus on the data science, not the infrastructure. With just few click anyone can start pre-configured and optimized Spark clusters, no specific operational skills are needed.

It can be used with external developer tool (Eclipse, Intellij, other notebook)

Databricks Runtime

Databricks Runtime includes Apache Spark but also adds a number of components and updates that substantially improve the usability, performance, and security of big data analytics.

Databricks Runtime for Machine Learning

Databricks Runtime ML is a variant of Databricks Runtime that adds multiple popular machine learning libraries, including TensorFlow, Keras, PyTorch, and XGBoost.

Databricks Runtime for Genomics

Databricks Runtime for Genomics is a variant of Databricks Runtime optimized for working with genomic and biomedical data.

Databricks Light

Databricks Light provides a runtime option for jobs that don't need the advanced performance, reliability, or autoscaling benefits provided by Databricks Runtime.

Delta Engine

Databricks optimized engine for both SQL and data frame workloads:

- It enables **BI use cases** with high concurrent user, providing dedicated endpoints to BI clients.
- It reduces costs thank to better performance query execution times up to **10x faster than Apache Spark 3.0**.

Databricks Runtime - Benefits

PERFORMANCE

The Databricks Runtime has been highly optimized by the original creators of Apache Spark. The significant increase in performance enables new use cases not previously possible for data processing and pipelines and improves data team productivity.

COST EFFECTIVE

The runtime leverages auto-scaling compute and storage to manage infrastructure costs. Clusters intelligently start and terminate, and the high cost-to-performance efficiency reduces infrastructure spend.

SIMPLICITY

Databricks has wrapped Spark with a suite of integrated services for automation and management to make it easier for data teams to build and manage pipelines, while giving IT teams administrative control.

Data Science Workspace

The Data Science Workspace is a collaborative environment for practitioners to run all analytic processes in one place, and manage ML models across the full lifecycle.

It is composed by:

- Notebook
- Managed Mlflow

And provides different features to simplify data science and data transformation development in an interactive environment:

- Access to any data source via Spark (BigQuery, Databases, Document DB, Kafka, ...)
- Git and Automatic Notebook versioning
- Model versioning
- Real-Time Coauthoring and Comments
- Dashboards

Data Science Workspace - Benefits

WORK TOGETHER

Share notebooks and work with peers in multiple languages (R, Python, SQL and Scala) and libraries of choice. Real-time coauthoring, commenting, and automated versioning simplify collaboration while staying in control.

SHARE INSIGHTS

Quickly discover new insights with built-in interactive visualizations or any library like matplotlib or ggplot. Export results and notebooks in html or ipynb format, or build and share dashboards that always stay up to date.

OPERATIONALIZE AT SCALE

Schedule notebooks to automatically run Machine Learning and data pipelines at scale, and create multi-stage pipelines using notebooks workflows. Set up alerts and quickly access audit logs for easy monitoring and troubleshooting.

COMPLETE MODEL LIFE-CYCLE

Track and manage machine learning models

Notebook

Microsoft Azure | Databricks

Portal emaccaferri@bitbang.com

r_kmeans_singlenode (R)

Detached File Edit View: Standard Permissions Run All Clear

Schedule Comments Experiment Revision history

You are viewing a notebook revision from Jun 18 2018, 12:05 PM CEST. [Exit](#)

Command took 0.33 seconds -- by fforde@bitbang.com at 20/3/2018, 16:07:56 on unknown cluster

Cmd 4

Data Visualization

Relazione tra la lunghezza e la larghezza del petalo e la specie di Iris

L'obiettivo è vedere se specie diverse hanno petali diversi

Cmd 5

```
1 data(iris)
2 ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) + geom_point()
3
4 - Credit Rating
5 Value Customer scegliamo MEDIUM
6 §- Value Customer scegliamo MEDIUM
7 - Discount shopper
```

Jun 18 2018, 12:05 PM CEST Luca Guerra All changes saved [Save now](#)

Jun 18 2018, 12:05 PM CEST Luca Guerra

Jun 18 2018, 12:01 PM CEST Luca Guerra

Jun 18 2018, 11:41 AM CEST Luca Guerra

May 4 2018, 17:23 PM CEST Federico Forte

Home Workspace Recents Data Clusters Jobs Models Search

MLFlow

Microsoft Azure | Databricks

Portal emaccaferri@bitbang.com

databricks_v2

Registered Models > CLV_model ▾

Details Serving

Created Time: 2020-07-31 15:00:40 Last Modified: 2020-07-31 17:13:02

▼ Description 

Model to predict the CLV

▼ Tags

Name	Value	Actions
No tags found.		

Add Tag

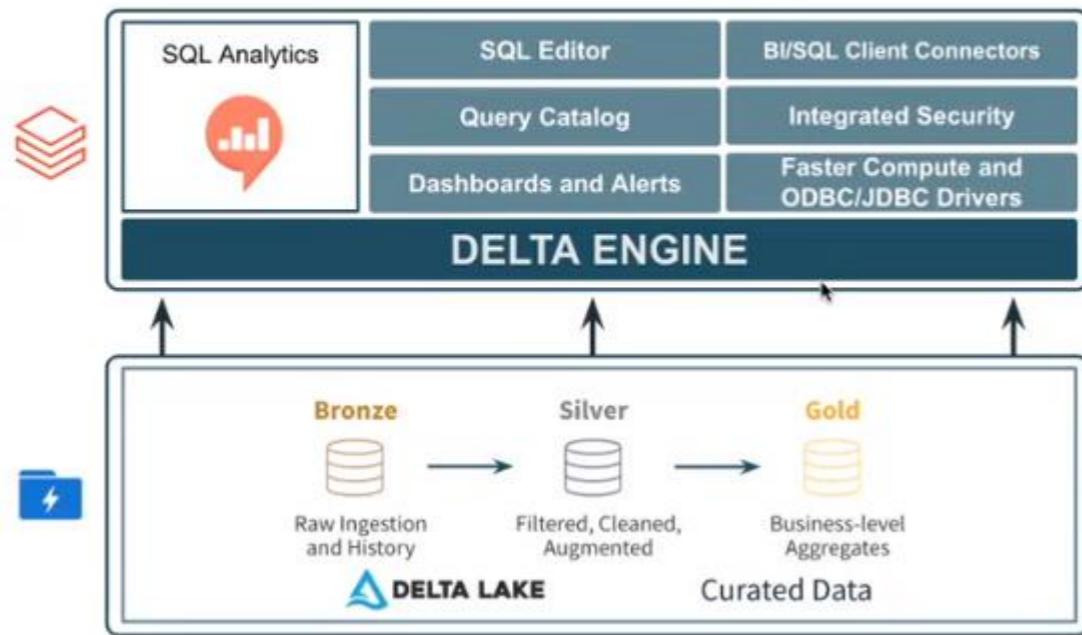
Name Value Add

▼ Versions All Active(2) Compare

<input type="checkbox"/>	Version	Registered at ▾	Created by	Stage	Pending Requests	Description
<input type="checkbox"/>	Version 3	2020-07-31 15:24:49	asilvi@bitbang.com	Production	-	
<input type="checkbox"/>	Version 2	2020-07-31 15:22:42	asilvi@bitbang.com	Staging	-	
<input type="checkbox"/>	Version 1	2020-07-31 15:00:41	asilvi@bitbang.com	Archived	-	1° version of the model. The ...

SQL Analytics

SQL Analytics, BI, and Reporting



SQL Analytics allows customers to perform **BI and SQL workloads on a multi-cloud lakehouse architecture**:

- A dedicated SQL-native workspace
- built-in connectors to common BI tools
- query performance innovations, and governance and administration capabilities.

This dedicated workspace for data analysts that uses a **familiar SQL-based environment to query Delta Lake tables** on data lakes.

However, since the data analysts and data scientists are both working from the same data source, the **overall infrastructure is greatly simplified** and a single source of truth is maintained.

SQL Analytics - Benefits

SQL-native Interface

Full-featured **SQL-native query editor**. Regularly used SQL code can be saved as snippets for quick reuse, and query results can be cached to keep run times short.

Easily Create Visualizations and Share Dashboards

Create **visualizations** that can be quickly organized into dashboards with an intuitive **drag-and-drop** interface, **automatically refresh**, alert.

Data warehouse performance at data lake economics.

SQL Analytics automatically and transparently load balances queries across multiple clusters to provide **high concurrency** and low latency without disruption.

Run BI and SQL workloads directly on your data lake up to 9x better price/performance than traditional cloud data warehouses through the advantages of lakehouse architecture.

30TB TPC-DS Price/Performance
Lower is better



Average Rating by Company Location

```

1 SELECT
2   Company_Location,
3   avg(Rating) AS avgRating
4 FROM chocolate.reviews
5 WHERE Review_Date = {{Year}}
6 GROUP BY Company_Location
7 ORDER BY avgRating ASC
    
```

Shared Endpoint (1)

chocolate (1) Filter tables & columns...

chocolate.flavors_of_cacao

Company (Maker-if known) STRING
Specific Bean Origin or Bar Name STRING
REF INT
Review Date INT
Cocoa Percent STRING
Company Location STRING
Rating DOUBLE
Bean Type STRING
Broad Bean Origin STRING

Year (1)
2019

Table Map (Choropleth) (1) Add Visualization

Company_Location	avgRating
India	2.50
Israel	2.75
Puerto Rico	2.75
South Africa	2.75
Portugal	
Venezuela	

Add description

Brooke Wenig created 19 days ago updated 4 days ago

alan.choi@databricks.com Refresh Schedule Never

Sales

Cash Generated – Sales - Payments Summary

\$16,008,872

Cash Generated

6 days ago

Customer Summary – Sales - Customer Summary

41,746.00 31,321.00 20,896.00 10,471.00 46.00

22 days ago

Number of Sales – Sales - Payments Summary

103,886

Number of Sales

6 days ago

Amount per Day – Sales - Amount per Day

0 50k 100k 150k

Jan 2017 Jul 2017 Jan 2018 Jul 2018

amount

22 days ago

Product Category Distribution – Sales - Product Category Distribution

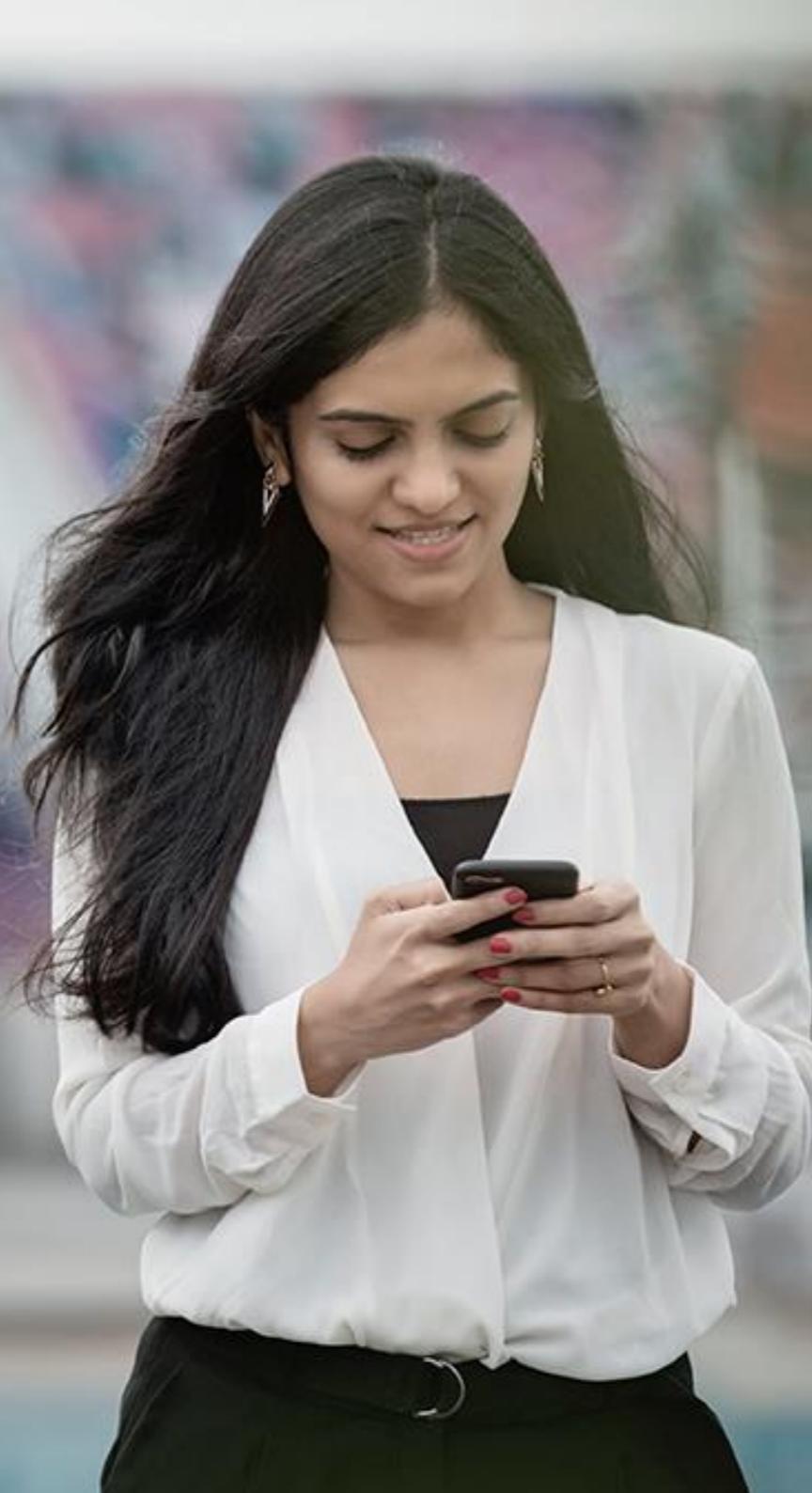
Category	Percentage
beleza_saude	22.3%
cama_mesa_banho	23.3%
informatica_acessorios	19.2%
esporte_lazer	16.9%
relogios_presentes	18.3%

22 days ago

Product Recommendation Use Case

Business Requirements in Luxury for eCommerce Product Pages

- ✓ Improve **Product Visibility**, **Cross-Sell** and **Up-Sell** through dedicated product carousels
- ✓ Evolve from the «one fits all» approach and harness more advanced **dynamic logics**



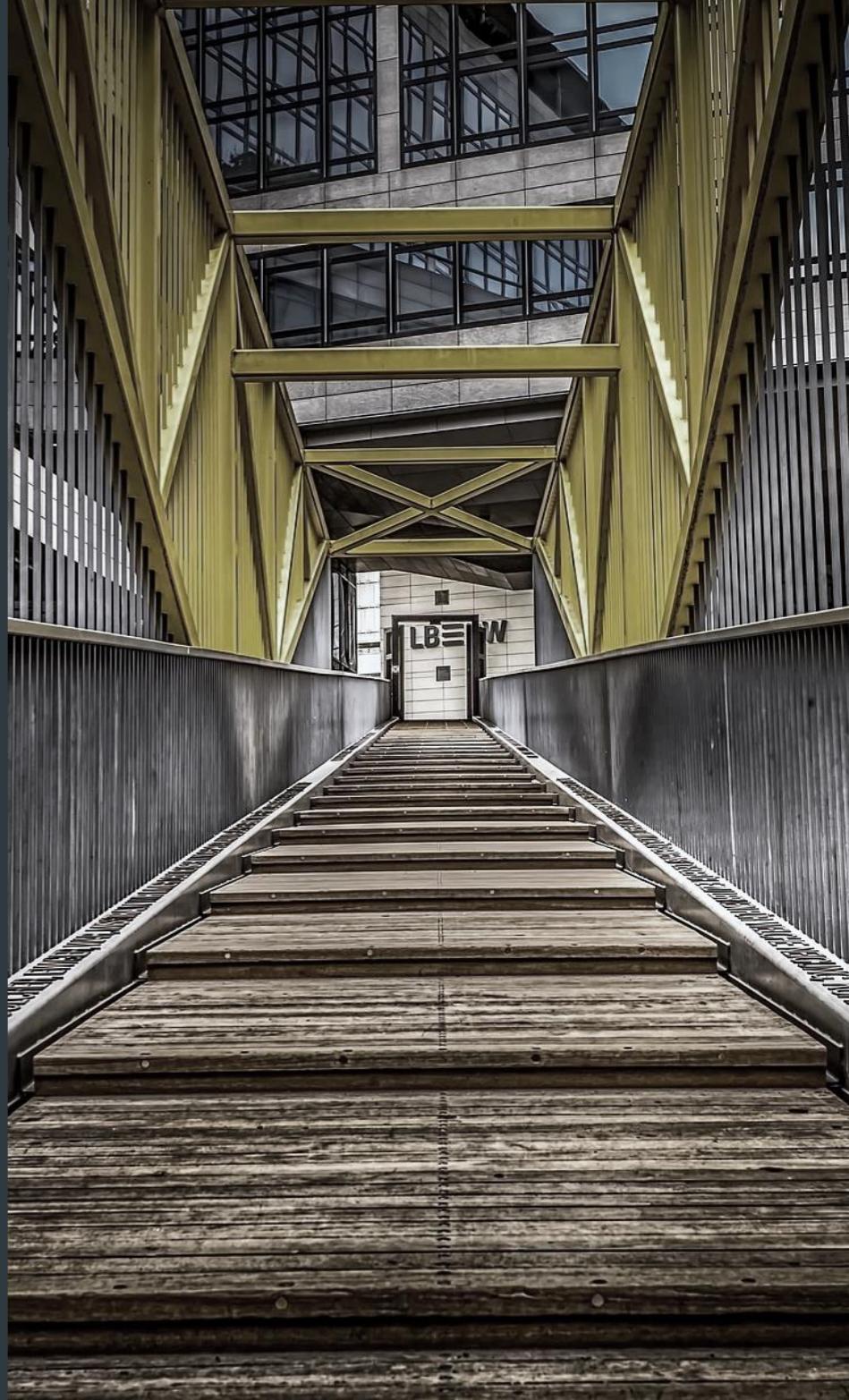
- ✓ Adopt **dedicated layouts** for website section
- ✓ Use custom **product-centric** algorithmic rules to adapt the results to the specific contexts of each product category and country

The Solution

Aim

Attain a **world-wide** and **cross-brand** management of data driven recommendation carousels, with a scalable solution.

- For each product (up to the third level of the category), provide an algorithm that recommends products related to the current one in PDP, based on the similarity between products.
- For fastened production deployment, the first release was based on the Databricks Accelerator.



Adobe Target Experiment Output

Adobe Target Recommendation Module

Adobe Target Recommendations activities automatically display products, services, or content that might interest your visitors based on previous user activity, preferences, or other criteria. Target Recommendations helps direct visitor to relevant items they might otherwise not know about. Recommendations lets you provide your visitors with relevant content at the right time and in the right place.

Recommendations helps you **optimize and customize real-time** suggestions **across channels**, apps, pages, email messages, and other delivery options to increase engagement and conversion while reducing management effort.

Adobe Target Recommendations are rules that determine which products or content to recommend based on a **predetermined set of visitor behaviors**. Criteria can be based on:

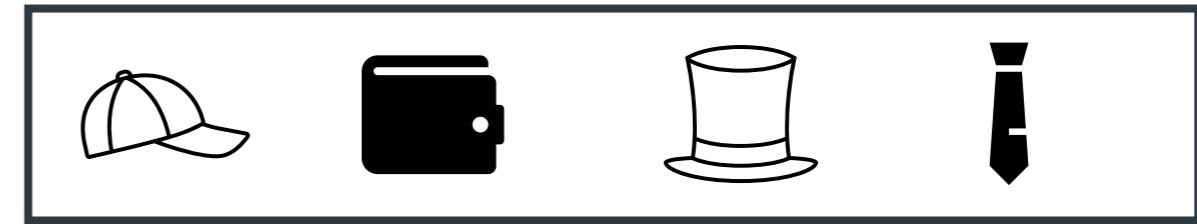
- popular trends
- a visitor's current and past behaviors
- similar products and content

Adobe Target

Page	Product Page
Metrics	<ul style="list-style-type: none">Carousel InteractionsAdd to cartConversion funnel
Solution	
A/B Test Recomm.	PDP standard without carousel PDP with Most viewed in Accessories and static for other categories
Outcome	Engagement improvement and increased conversion

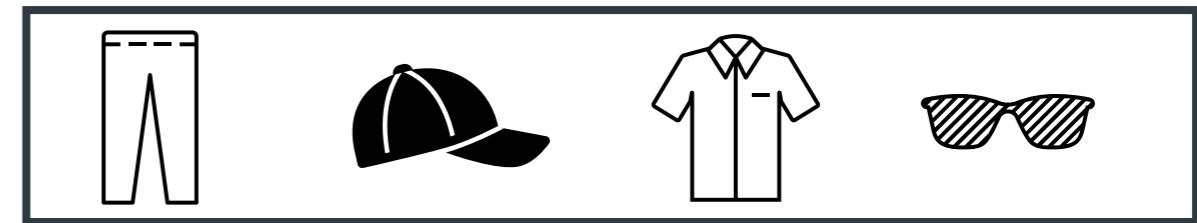


ADD TO CART



Others

ADD TO CART





Critical points analysis

- Limited scalability due to the high level of customization required for each individual activity
- High efforts and long elapsed time to cover the entire project span
- Unavoidable step-by-step approach with gradual releases for different categories and different countries

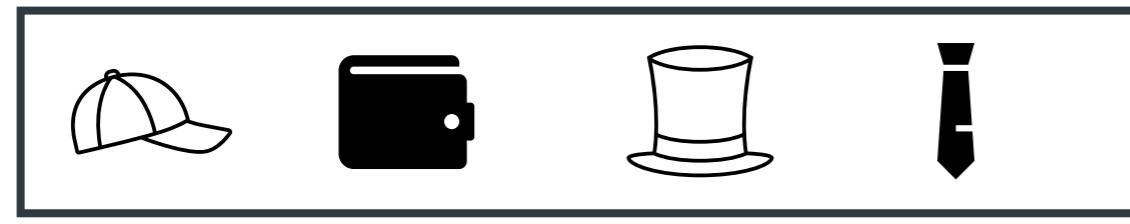
«You may also like» Solution

Azure Data Lake & Adobe Target

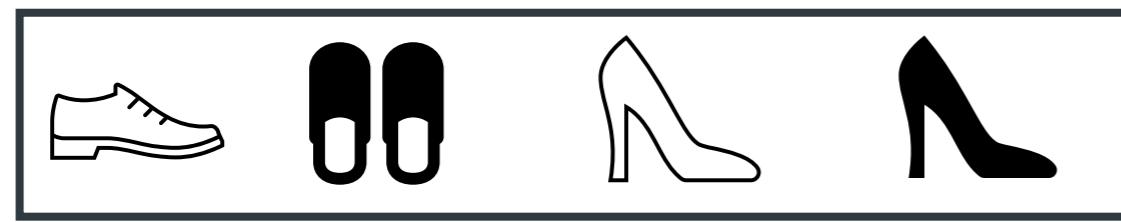
Custom solution



ADD TO CART



ADD TO CART



Due to lack of real-time website tracking, we will provide an algorithm that recommends **products related** to the current one in **PDP**, (without considering for the moment specific user characteristics).

The algorithm approach will be based on the **similarity** between products:

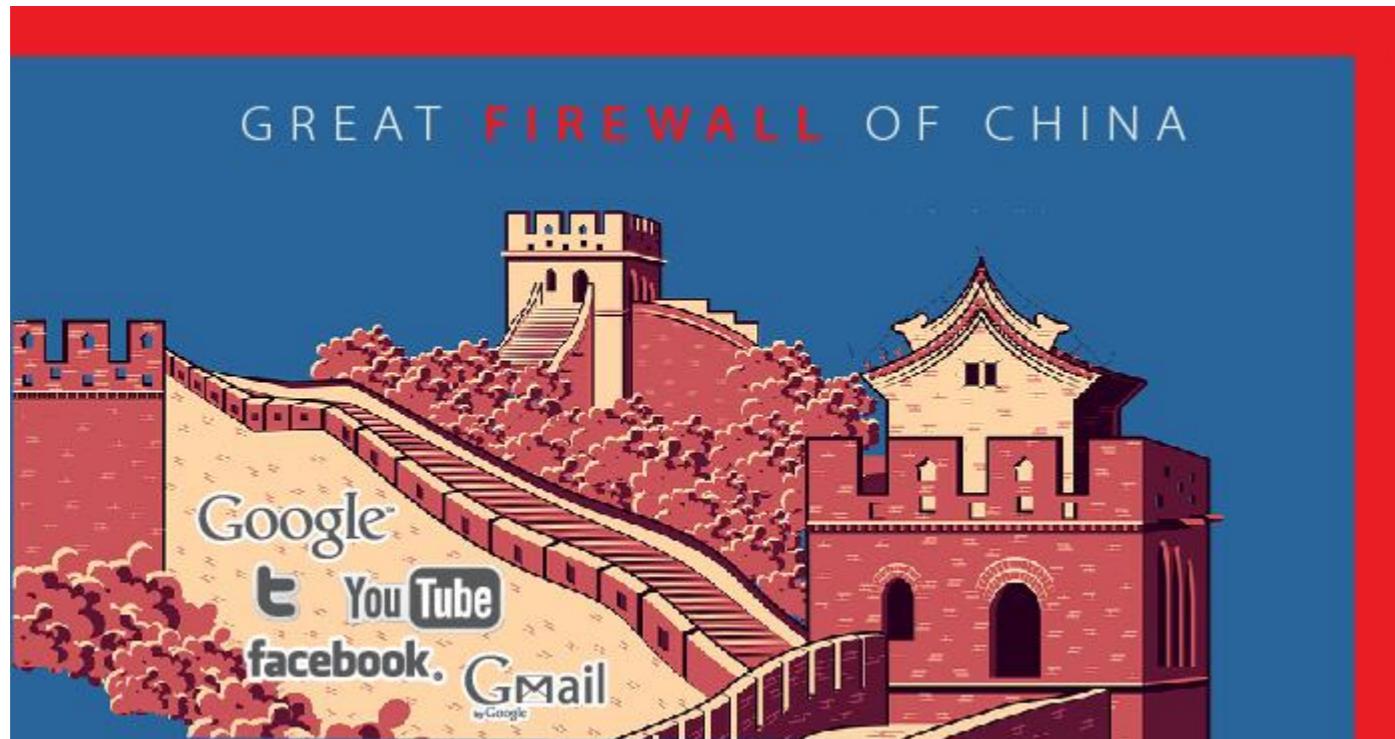
- considering the number of common users **who view them**: "Who viewed this also viewed these other products".
- Considering similar characteristic from metadata or images

To have **significant** results, a relatively "large" time window is required and data change slowly, so the computing can take place **once per day**.

Fashion & Luxury for Chinese Market

Chinese customer are high spender on Luxury Goods

E-commerce in China are protected by the Great Firewall, so during the architecture design for the recommendation engine we had to consider to replicate part of the solution in a far-east zone on the cloud.

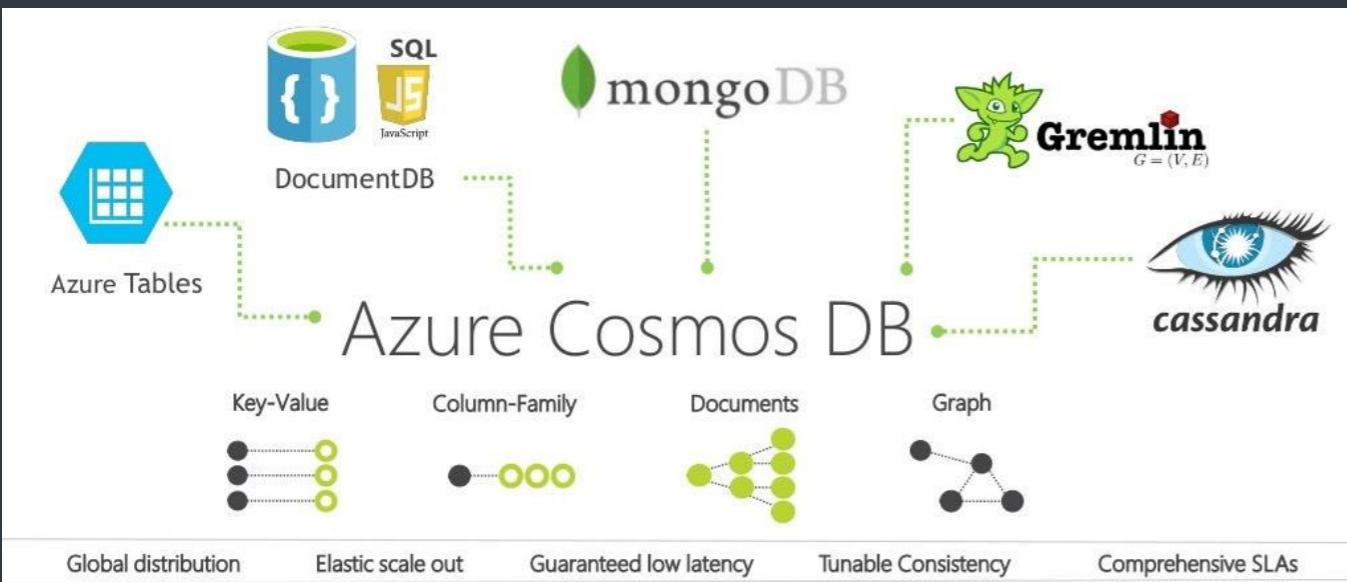


Great Firewall aka Golden Shield Project

The combination of legislative actions and technologies enforced by the People's Republic of China to regulate the Internet domestically.

Brand face it's effects because it slows down website performance outside China.

Azure Cosmos DB



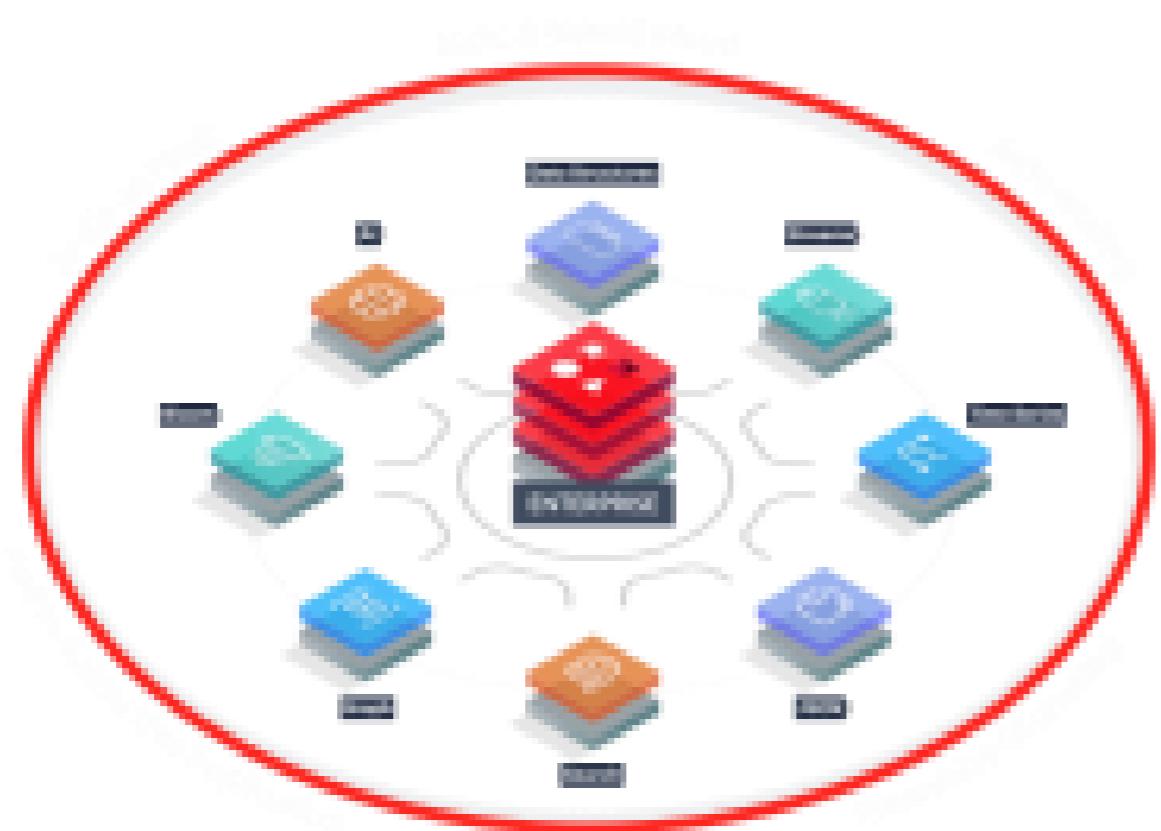
Azure Cosmos DB is a fully managed NoSQL and relational database for modern app development. Azure Cosmos DB offers single-digit millisecond response times, automatic and instant scalability, along with guarantee speed at any scale. Business continuity is assured with SLA-backed availability and enterprise-grade security.

App development is faster and more productive thanks to:

- Turnkey multi region data distribution anywhere in the world
- Open source APIs
- SDKs for popular languages.
- As a fully managed service, Azure Cosmos DB takes database administration off your hands with automatic management, updates and patching. It also handles capacity management with cost-effective serverless and automatic scaling options that respond to application needs to match capacity with demand.

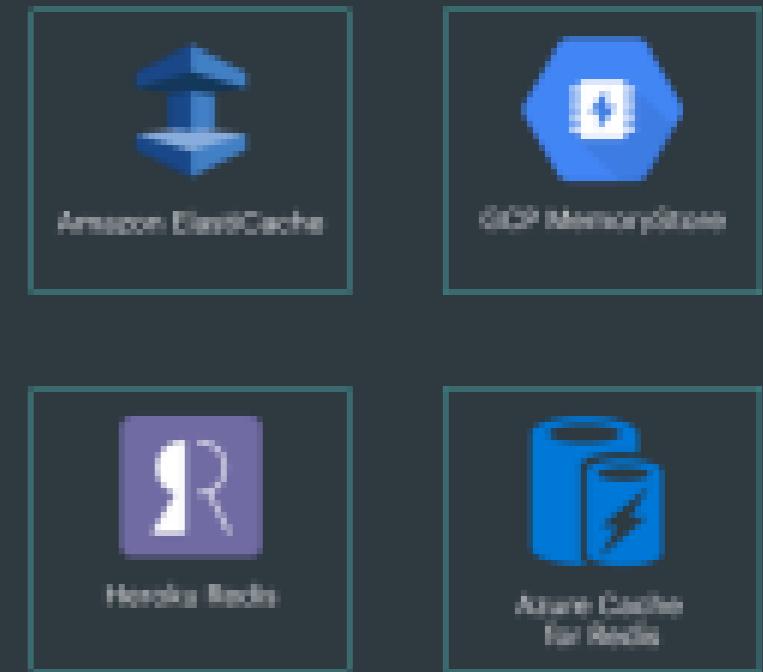
Redis

Redis is database that supports both enterprise caching and a multi-model database for building modern applications and low latency microservices architectures. It provides an in-memory data store



Azure Cache for Redis

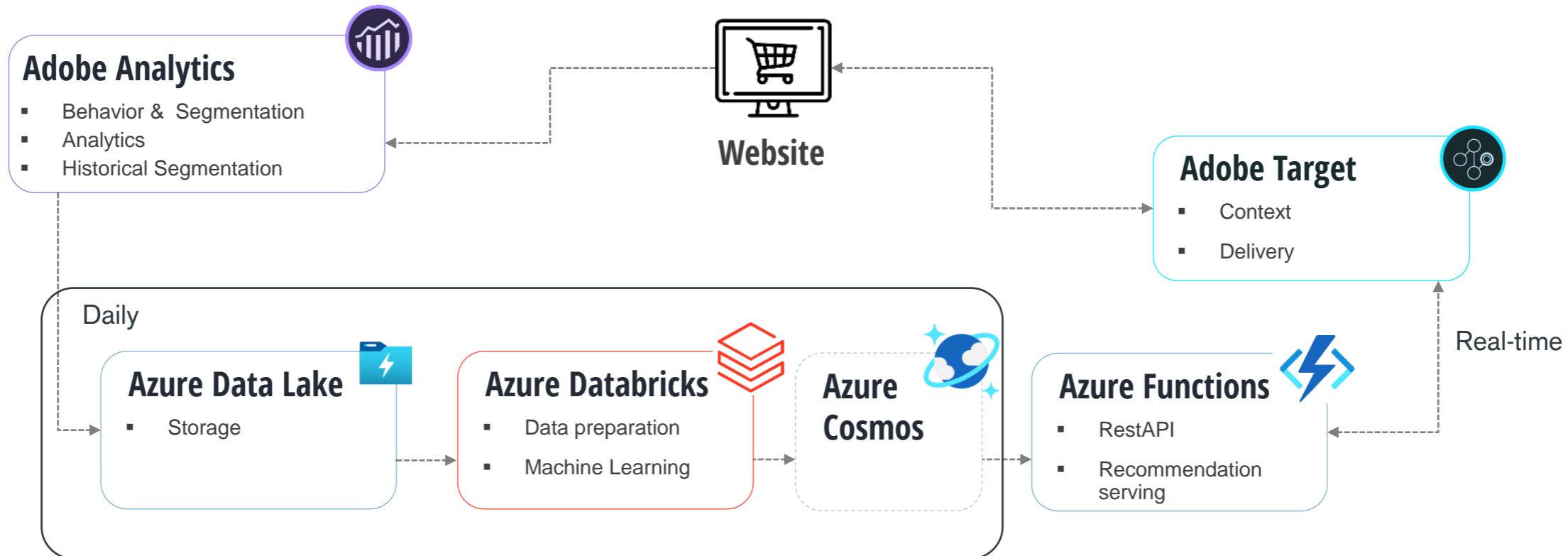
With other Redis providers, you are limited to only using Redis as a cache



Technical solution

The synergy between Azure and Target allows to provide Item-Item recommendations promptly:

- Target activates in real-time invoking the Data Lake providing the context of the request (ex. SKU 123, website)
- Data Lake calculates recommendations according to requirements (country, subcategory)
- Target receives the result and manages the delivery populating the carousel



Why still use Target

Even if the Rest API of our recommendation engine could be called for any system, in particular server side, Target remains the tool serving the carousel to avoid website development.

This choice was made to be independent from the planned roadmap and to release the project faster.

With this configuration the API has to accept unauthenticated call from every IP, so we used Akamai algorithm to protect it from excessive call made by both or similar.

Activity details

- Exclusion of purchased items: this could be doable providing the current visitor ID (both if the visitor is currently logged-in or has logged-in in the past with the current browser) and could include even offline purchases, but will not be working if the visitor is using a new browser never used before for logging-in
- Product category level recognition: for every product fully navigated through the standard menu, each category will be recognized, until the 3rd level, allowing us to filter the output of the algorithm for each of them. In case of “particular” navigations (i.e. internal search or collections lists) the categorization will follow a predetermined static schema
- Fallback algorithm: in case the algorithm will provide fewer results than expected it's possibile to rely on an alternative logic less tight (i.e. considering items from the 2nd level and not just from the 3rd category level) or to use a static list of predefined items
- Failure management: in case for any reason the results of the algorithm would require too much time to be available, we will manage to show the original version of the component (static carousel if available, no carousel in case the element was not originally present)

Potential evolutions of the PDP recommendation

Customize

recommendations:

- Logged in / recognized / anonymous
- Deduct information from users behavior

Test different algorithms

and recommendation logics

Promote products just added to the catalog in the recommendations

Consider the **similarities between products** based on their characteristics

It's possible to do this:

- From structured catalog data
- From the free text of the editorial descriptions
- From the pictures



Use Case Extension

Sales Assistant App

- Recommendations based on user similarity of particular purchase preferences

Commercial DEM

- Communications with product recommendations based on abandoned carts and new product announcements

Logged Area

- Recommendations based on user similarity of particular purchase preferences on the website

Logged Area Architecture Issue

To reduce costs we choose consumption plan for all Functions serving recommendation, but the release plan of the website put us in front of an unexpected phenomenon:

1. the new area of the site was released first in a small country
2. This country does not generate many visits on the target page for recommendation
3. There are not enough calls to the API to keep the service running
4. Cold start problem and response even after 30 seconds



100%
Increase
in engagement with carousel

The Results



Unified experience
through channels and
next-best product

«Recently Viewed» Solution

Adobe Target

Scenario

MY ACCOUNT

Name
John

Surname
Doe

Contact email
johndoe@gmail.com

Credit Card
XXXX-XXXX-XXXX-234



In this scenario logic recommendation is focused on users/**visitors** and not products.

Carousel filling strategy is to show **real-time** an ordered list of the last products visualized by users.

Usually this kind of activity is effective in touchpoint not bounded to a single product such as **HomePage** or logged area; in general where to **returning user** focus is on quickly find **already visited item**.

Technical solution

The focus is on response speed and to show User-Item recommendation updated in real-time, the architecture contains less services:

- Adobe Target identifies each visitors and contains the product catalogue
- Real-time specific Adobe Target tracking collects the visited SKU lists for each visitor
- The combination of Visitor ID –SKU lists – product information enable Adobe Target to build the carousel



Recap



Customer-scalable modern data platform



Meant to power-up Analytics



Low time to market of new project



BitBang

Info@bitbang.com

Bologna

Via E. Mattei 102,
40138, Bologna, BO
+39 051 58 75 314

P.IVA 02329121202
Capitale Soc: € 60.000
R.E.A. BO 431512

Milano

Via S. Marco 21,
20121, Milano, MI
+39 02 12 41 23 352

www.bitbang.com