



FiTecuSERV



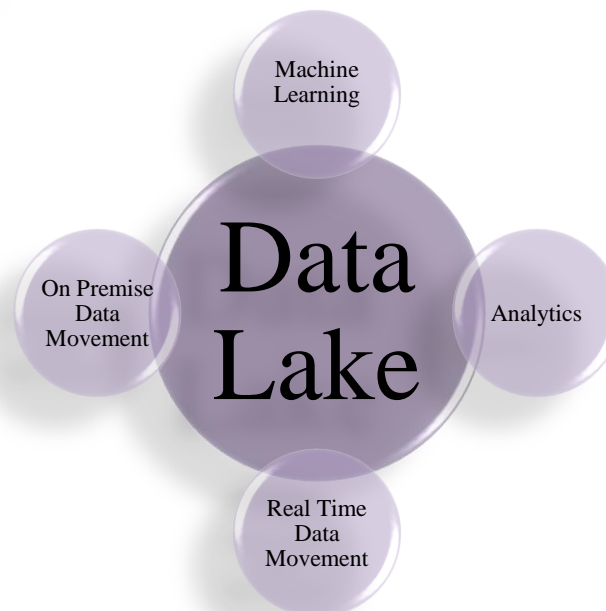
| Data Lakes

Introduction

Data lake concept allows us to safeguard enormous amount of raw data under a flat architecture in its native format till it is required.

This service focuses on the development of models and extraction processes for a centralized data area of structured and unstructured data that allows both the execution of the dashboard's analytical processes and visualizations, processing of large volumes of data (Big Data), real-time analytics, artificial intelligence processes, and machine learning.

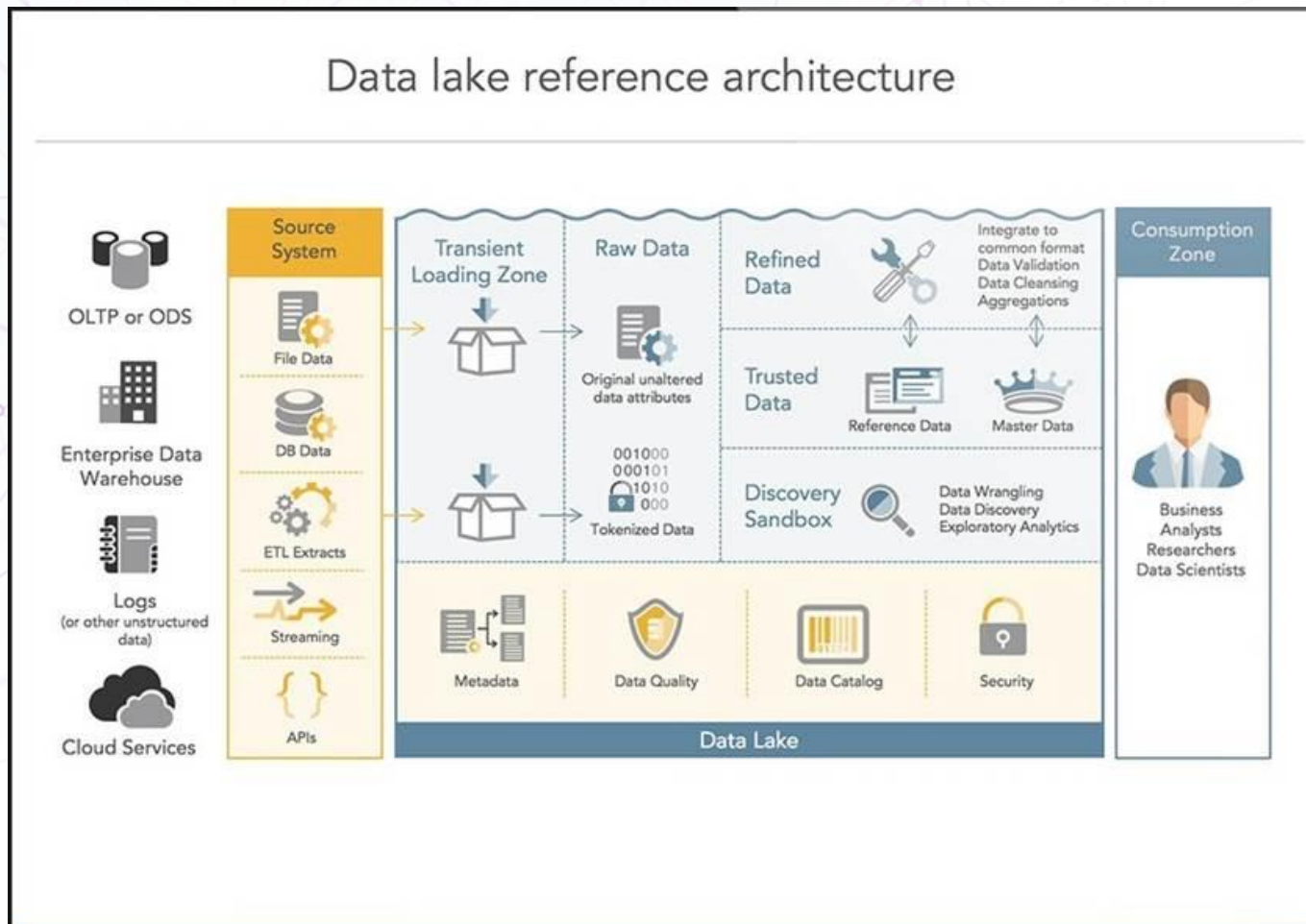
Above mentioned solutions empower the organization for IoT solutions. If cases that are no requirements or needs for unstructured data of high importance and volume, you can consider using Datawarehouse services.



Key factors to Build a Data Lake

- **Data Ingestion:** Pulling data from various data sources and loading it into the data lake with the help of connectors.
- **Data Retention:** Unused or low-use of data from the data lake. Data retention policy should be taken carefully to be clear defined considering user, compliance and legal requirements.
- **Data Quality:** If you have quality data in the data lake it will help to acquire good insights for business.
- **Data Storage:** When it comes to storage capacity, data lakes are easily scalable, so scalable requirements should be consider, integrated and ideally automatized when the data increases.
- **Data Discovery:** Data lakes have a centralized repository, which allows us to get better information on self-service analytics.
- **Data Auditing:** As soon as the data begins to load into the data lake, it needs to be tracked and changes need to be captured by the users considering evolution and time.
- **Data Governance:** Data lakes tend to be highly available and can easily be integrated into an organization within various departments. A governance model should be clear defined and clear communicate to all potential users.
- **Security:** The data must be protected at all times and has to be integrated with only the authorized users.

Data Lake Reference Architecture



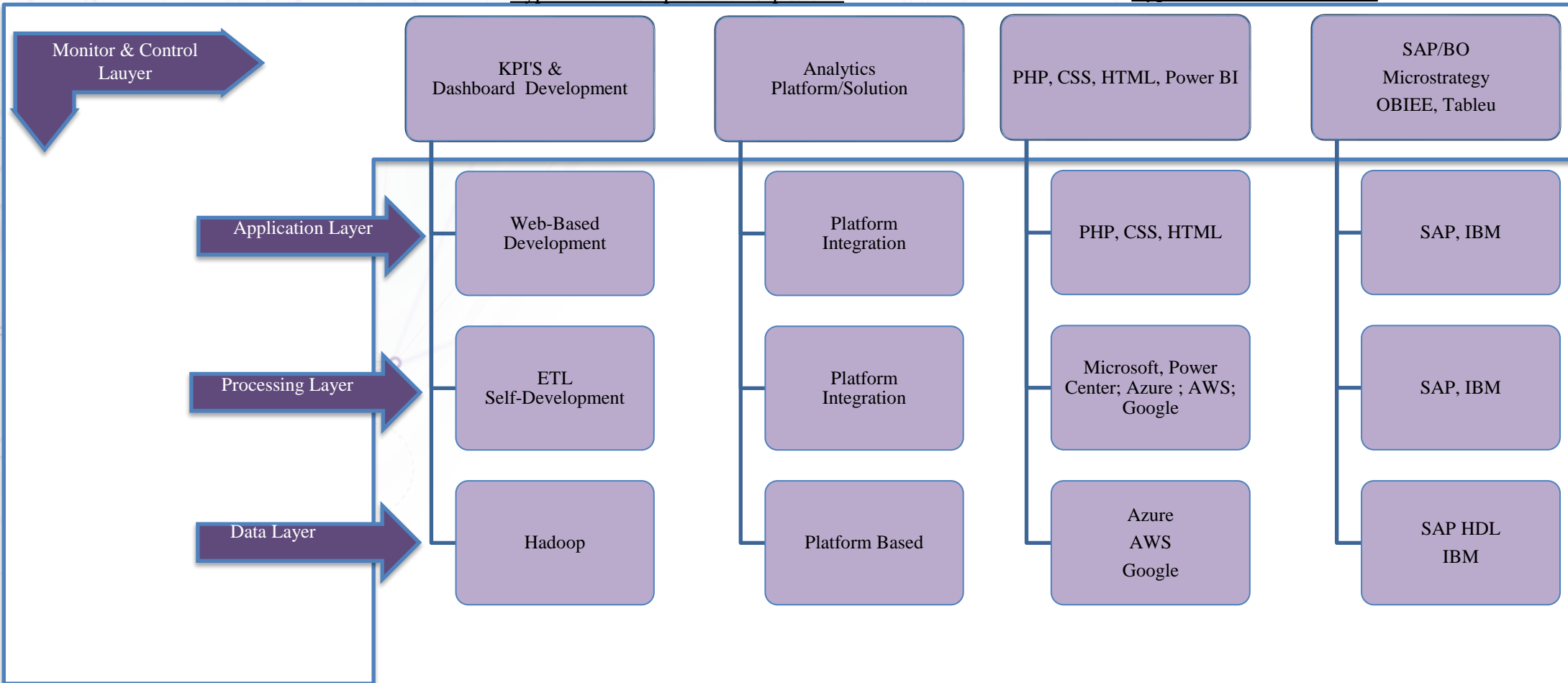
The data lake has the capability of storing unlimited data, in its original format for as long as you need. Unlike traditional storage solutions, the data lake is an online system where all data is available for query by end-users as shown here in the pictorial representation.

A Data Lake architecture consists in several components that capture data, processes it and stores, ensuring quality, uniformity and security but also core components for exploratory analyses and pattern discoveries.

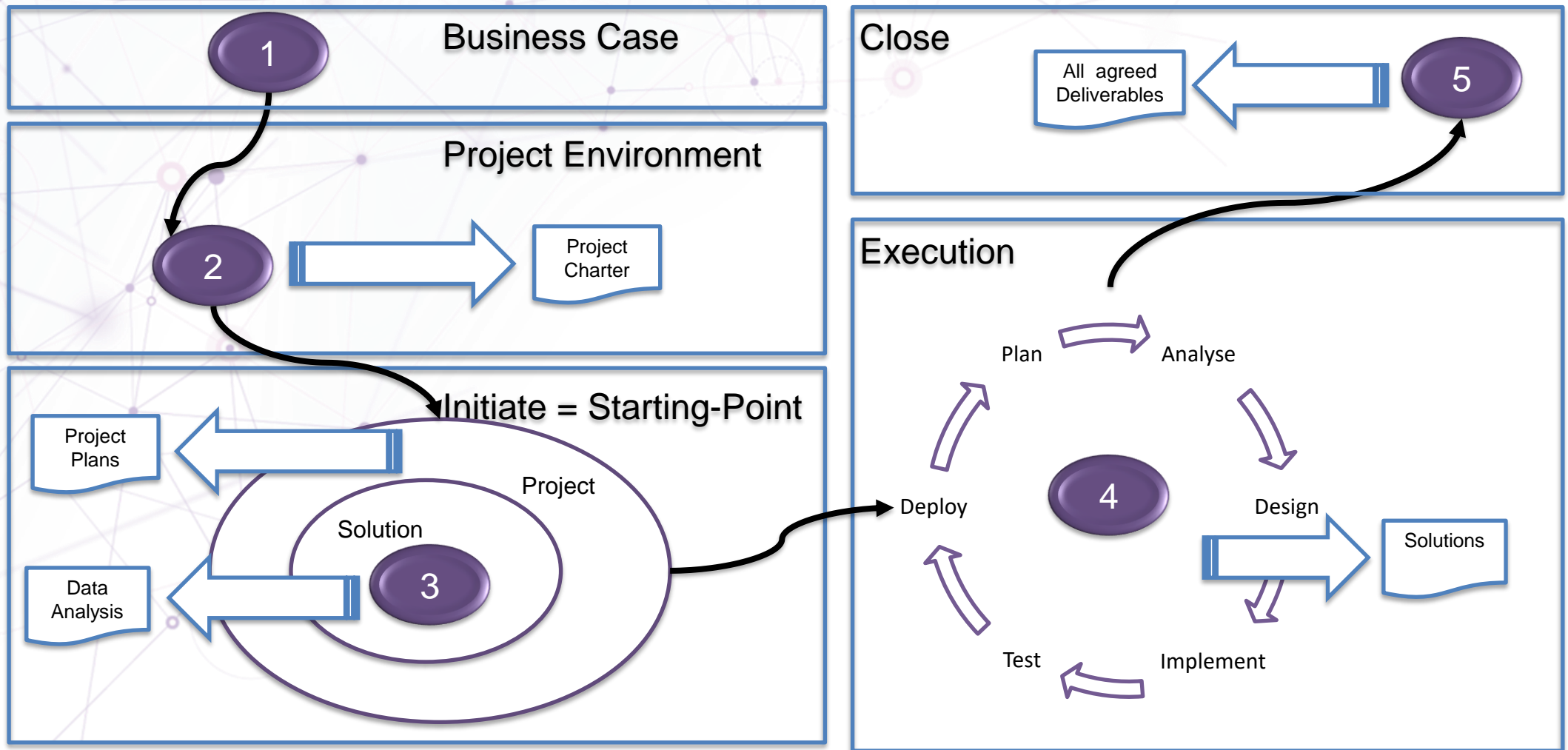
Difference between Self-Development vs Platform Solution

Types of Development Components

Types of Solution Partners



AitecServ Implementation Model (AIM)



Business Case + Project Environment

Business Case

Identify Stakeholders (Identify Product Owner and Major Business Users).

User engagement and commitment strategy.

Business Case Preparation/Identification

Project Vision / High Level Requirements

Project Environment

Understand the organization's change management process.

Identify and analyze the factors surrounding the project.

Identify compliance needs.

Capture assumptions, constraints and prior agreements.

AIM – Starting Point

Management:

In this stage the main focus is to elaborate the several management plans. This projects normally require **a very clear and detailed plans** to ensure project success.

This kind of service is characterized by multidisciplinary teams and complex project management requirements.

Solution:

In this stage we evaluate the “Change Data Capture” process, by assessing how data is changed and deleted in the potential source systems.

We also identify the main reporting needs and the major KPI’s (Key Performance Indicators) and related contextualize characteristics.

And last, we will evaluate labeling and categorization needs for unstructured data and identify **requirements for automation and/or AI + machine learning techniques**

Management

Identify Major Deliverables

Identify Risks

Initial preparation of project plans

Define Guidelines for optimization & governance

Solution

Evaluate “Change Data Capture” process

Identify data availability & Quality factors

Identify reporting needs

Identify labelling & Categorization needs

AIM – Execution

Management

Monitor and control project objectives

Manage the various project plans

Manage Product and Sprint Backlog

Solution

Implementation following AGILE principles

first release of a pilot with a “Go-No Go”

Incremental and interactive

Management:

The main task are to **monitor and control the project objectives and performance**, making the necessary adjustments to keep the project on track.

Solution:

Our services follow preferentially an agile approach, **providing in the first release a pilot version with the minimum valuable product** for solution evaluation, address the expected objectives and to a refine the product backlog.



Atecserv

Thank
You!

For more information or to Schedule a
meeting please contact us on:



<http://www.atecserv.pt/en/contact>
<mailto:geral@atecserv.pt>