



# FiTecuSERV



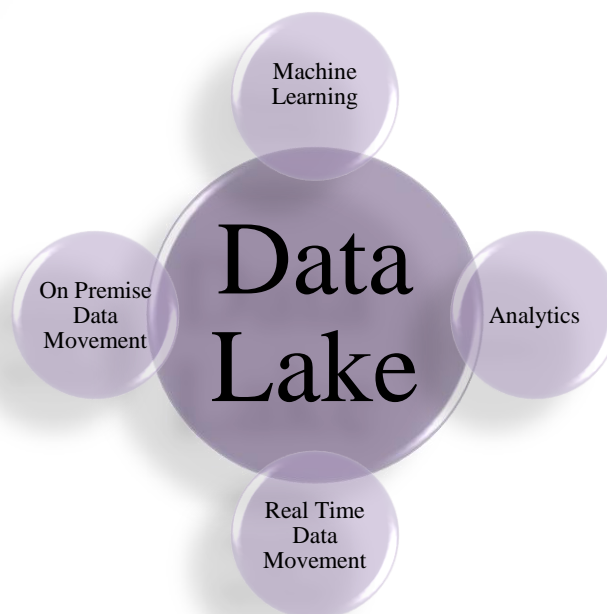
| Data Lakes

## Introdução

O conceito de data lake permite-nos guardar e proteger uma enorme quantidade de dados brutos numa arquitetura plana, no seu formato nativo e até que seja necessário.

Este serviço foca no desenvolvimento de modelos e processos de extração para uma área de dados centralizados, sejam estruturados e não estruturados que permitirá tanto a execução de processos analíticos e visualização, processamento de grandes volumes de dados (Big Data), análises em tempo real, processos de inteligência artificial e aprendizagem máquina.

As soluções mencionadas acima capacitam a organização para soluções de IoT. Nos casos em que não forem necessários considerar o tratamento de dados não estruturados de grande importância e volume, considere o uso de serviços Datawarehouse.



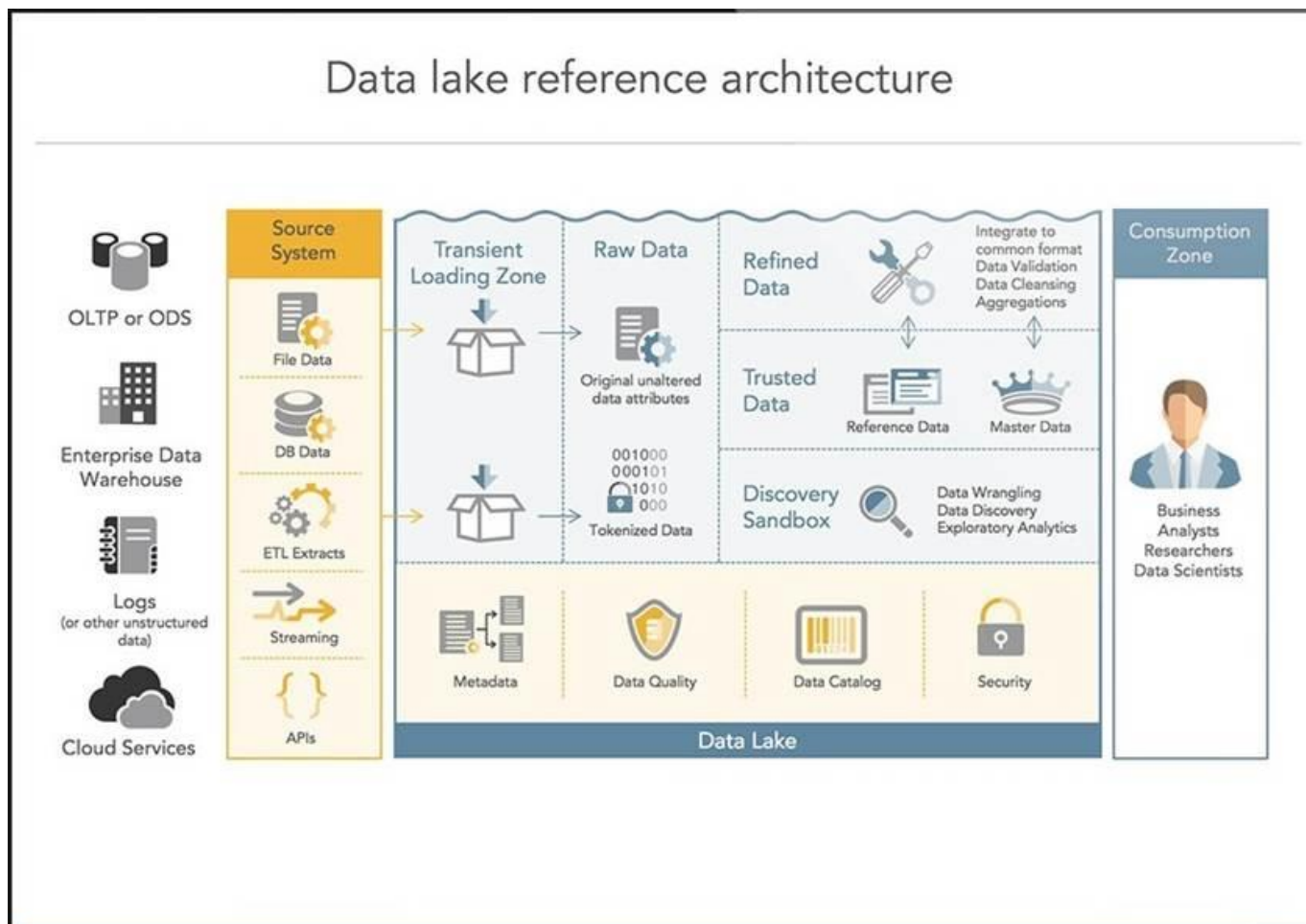


## Fatores principais para construir um Data Lake

- **Ingestão:** Obter dados de várias fontes de dados e carregá-los no *Data Lake* com a ajuda de conectores.
- **Retenção:** Dados não utilizados ou pouco utilizados do *Data Lake*. Deve-se tomar cuidado ao definir a política de retenção de dados considerando requisitos de uso, de conformidade e legais.
- **Qualidade:** Se tiver dados de qualidade no *Data Lake*, isso o ajudará a adquirir bons padrões do negócio.
- **Armazenamento:** Quando se trata de capacidade de armazenamento, os data lakes são facilmente escaláveis, pelo que requisitos de escalabilidade devem ser considerados, integrados e idealmente automatizados assim que os dados aumentam.
- **Análise Dados:** Os *Data Lakes* têm um repositório centralizado, o que nos permite obter melhores informações na execução de análises.
- **Auditoria:** Assim que os dados começam a ser carregados no *Data Lake*, eles precisam ser rastreados e as alterações precisam ser capturadas pelos usuários considerando evolução e tempo.
- **Governance:** *Data Lakes* tendem a ser altamente disponíveis e podem ser facilmente integrados em uma organização dentro de vários departamentos. Um modelo de *governance* deve ser definido e comunicado de forma clara a todos os potenciais utilizadores.
- **Segurança:** Os dados devem ser protegidos em todos os momentos e devem ser integrados apenas com os utilizadores autorizados.



## Arquitetura de referência de *Data Lakes*



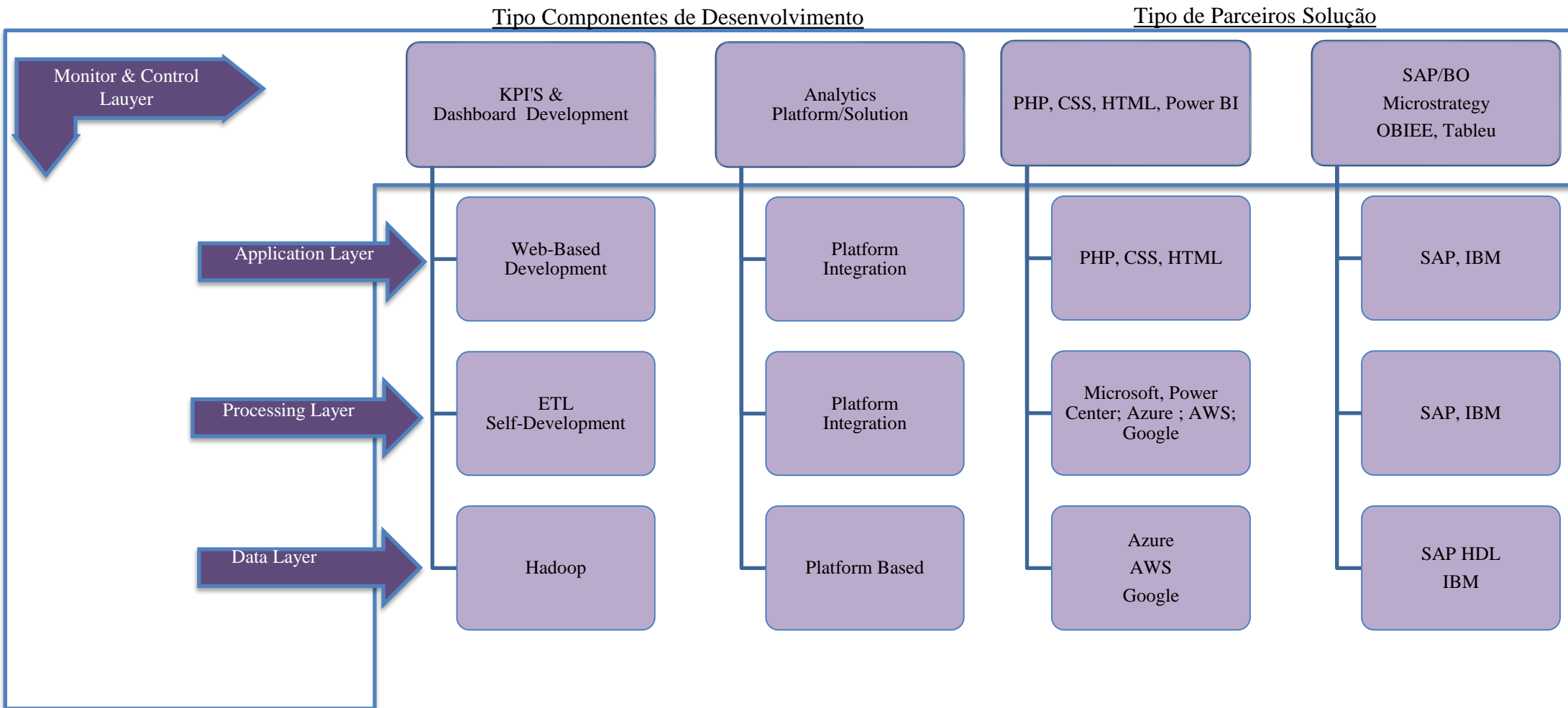
O *Data Lake* tem a capacidade de armazenar dados ilimitados, no seu formato original, pelo tempo que precisar. Ao contrário das soluções de armazenamento tradicionais, o *data lake* é um sistema online onde todos os dados estão disponíveis para consulta pelos utilizadores finais, conforme se pode ver aqui na representação gráfica.

Uma arquitetura de *Data Lake* consiste em diversos componentes que capturam os dados, processam e os guardam assegurando qualidade, uniformidade e segurança. Detém ainda outros componentes bastante importantes para a análise exploratória e a descoberta de padrões.

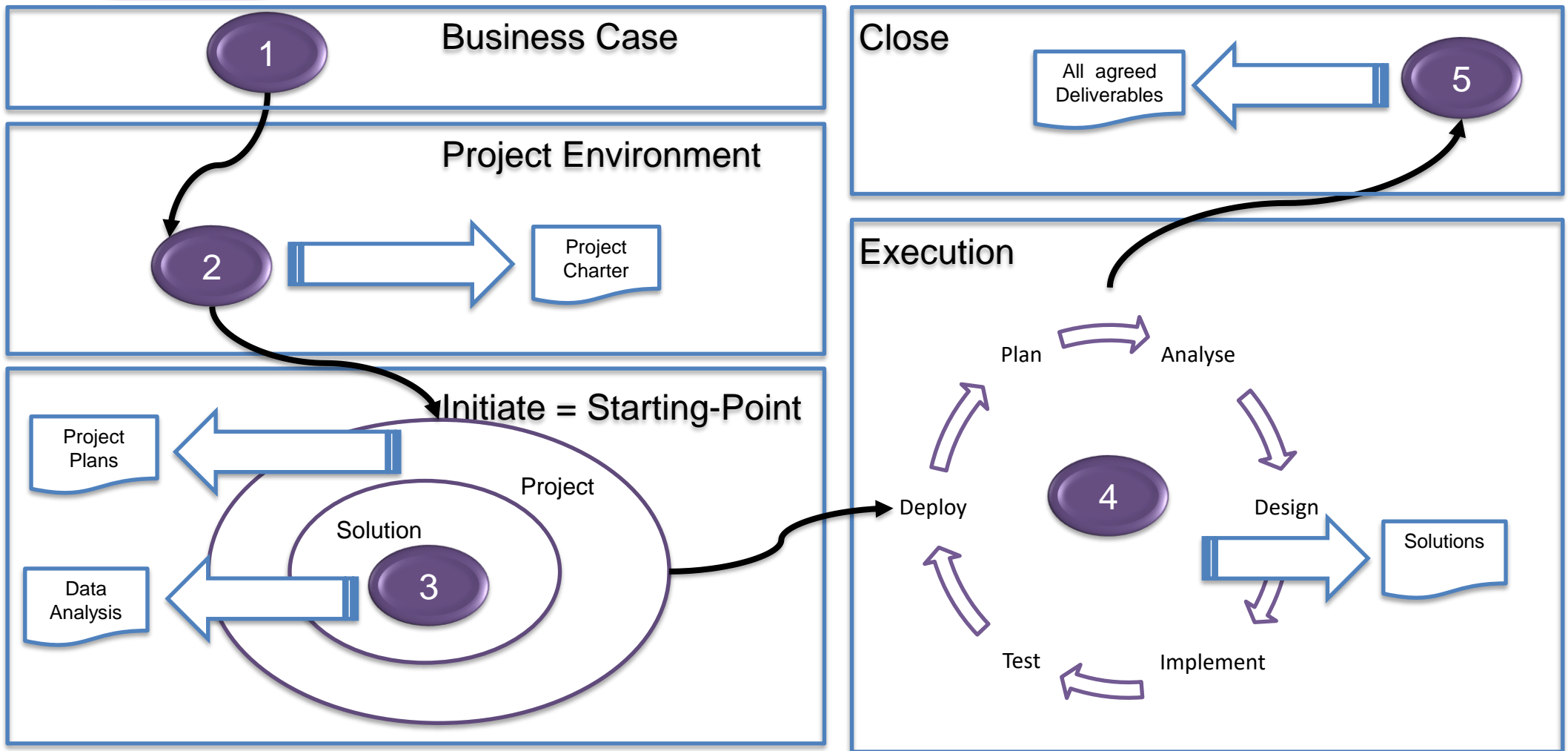




## Diferença entre Desenvolvimento à medida vs Plataformas

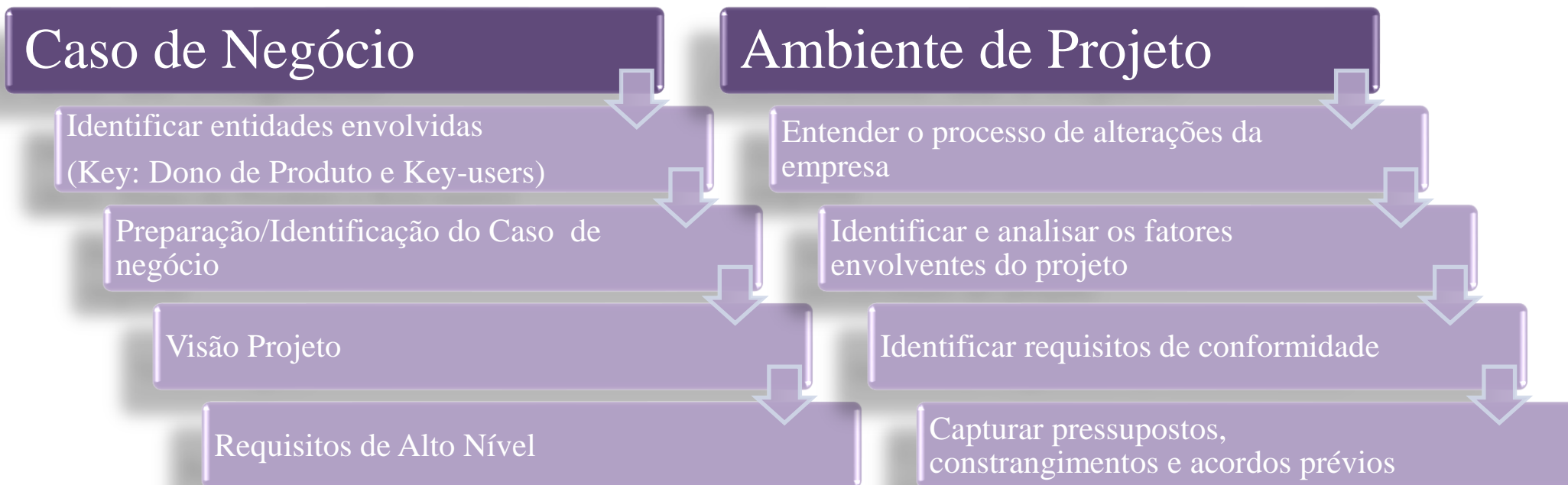


## Modelo Implementação AitecServ (AIM)





## AIM – Caso de Negócio & Ambiente Projeto



## AIM – Starting Point

### Gestão:

Neste fase a principal preocupação é de elaborar os diferentes planos do projeto. Estes projetos requerem normalmente **planeamento claro e detalhado** de forma a garantir o seu sucesso.

Este tipo de projeto caracterize-se por equipas multidisciplinares e de requisitos de gestão de maior complexidade.

### Solução:

Nesta fase focamo-nos em avaliar o processo de captura de alterações, aferindo quando e como os dados são alterados e eliminados nos sistemas fonte potenciais.

Iremos ainda identificar com os principais utilizadores de negócio as necessidades de informação mais pertinentes bem como indicadores e características que os contextualizam e por fim iremos avaliar as necessidade de categorização e etiquetagem dos dados para dados não-estruturados e **identificar requisitos para automação e aplicação de técnicas de AI + machine learning**.

### Gestão

*Identificar principais entregáveis*

*Identificar Riscos*

*Preparação inicial dos Planos Projeto*

*Linhas orientadoras de optimização & governance*

### Solução

*Avaliar Processo de “Change Data Capture”*

*Identify data availability & Quality factors*

*Determinar Req. Reporting*

*Identificar necessidades de categorização e etiquetagem*





## | AIM – Execução

### Gestão

*Monitorizar e  
Controlar os  
objetivos Projeto*

*Gerir os diversos  
planos projeto*

*Gerir o Backlog de  
Produto e dos  
Sprints*

### Solução

*Implementação  
segundo princípios  
AGILE*

*Primeira entrega de  
piloto com “Go-No  
Go”*

*Incremental e  
Iterativo*

#### **Gestão:**

A atividade principal será de **monitorizar e controlar o desempenho e os objetivos do projeto**, efetuando os ajustamentos necessários de forma a manter o projeto em linha com o planeado.

#### **Solução:**

Os nossos serviços seguem preferencialmente uma abordagem AGILE, **entregando numa primeira Release da solução as características de valor mínimas de produto**, permitindo uma avaliação da solução, endereçar os objetivos pretendidos e refinar o backlog de produto.



# Atecserv

Obrigado!

Para mais informação ou para marcar  
uma reunião contacte-nos em:



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