

Governança e Catálogo de Dados no Âmbito de uma Estratégia de Dados

Proposta de Bolsa/Estágio na Altice Lab



ID do Projeto	2024/25_N28
Departamento	SSO53
Proponente	Luis Cortesão
Data de validade	31-Dec-2024

IDI - Áreas Chave

Big Data & Analytics

Tema / Título

Governança e Catálogo de Dados no Âmbito de uma Estratégia de Dados

Contexto

As empresas sabem que os seus dados são um bem estratégico e querem utilizá-los para tomar decisões mais inteligentes mas têm de enfrentar problemas muito complexos para alcançar esse desiderato. Muitas vezes os dados estão dispersos em silos, presos em sistemas departamentais que não falam bem uns com os outros, a qualidade dos dados é má, e os custos associados para resolver esses problemas são muito elevados. Em conjunto com a resposta às pressões do mercado, a maioria das empresas vai dar prioridade às necessidades urgentes, táticas e quotidianas em detrimento das iniciativas estratégicas a longo prazo.

Avançar para uma cultura mais orientada para os dados é certamente alcançável, e começa com uma estratégia de dados. Uma estratégia de dados é muitas vezes vista como um exercício técnico, mas uma estratégia de dados moderna e abrangente aborda mais do que os dados; é um roteiro que define pessoas, processos e tecnologia.

Do conjunto de múltiplas iniciativas conducentes a uma estratégia de dados, a definição de políticas de governança de dados e a sua articulação com um catálogo de dados é de extrema importância.

A governança de dados refere-se ao conjunto de processos, normas, políticas e tecnologias necessárias para gerir, manter e explorar os dados como um recurso organizacional. É essencial para garantir a qualidade, integridade, segurança e conformidade dos dados.

A utilização de catálogos de dados fortalece sua governança, proporcionando uma visão abrangente dos ativos de informação e promovendo a transparência na gestão dos dados organizacionais.

Os catálogos de dados não só fornecem contexto aos principais interessados para encontrar e compreender dados, como também automatizam a gestão de metadados e tornam-na colaborativa. Desta forma, os intervenientes externos não só o compreenderão, como também atuarão sobre ele e o curarão para que possam aproveitar o catálogo de dados para uma utilização mais alargada. Consequentemente, a dimensão da automatização e da colaboração de dados é da maior importância no suporte às políticas de governança que sustentam uma estratégia de dados.

Objetivos do Projeto

No âmbito deste estágio, o bolsheiro será integrado num projeto de identificação e definição de boas práticas para a construção de uma estratégia de dados, com particular enfoque na implementação de um catálogo de dados adequado às políticas de governança de dados da organização.

Aspetos Inovadores

Exploração de estratégias, conceitos e implementação de soluções no âmbito:

- Data Governance/Strategy;
- Information Architecture;
- Big Data & Analytics.

Ferramentas a utilizar

- Programação Fullstack;
- Software development process tools : Git, Wiki, etc.

Referências Bibliográficas

- <https://www.cio.com/article/3521011/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html>
- <https://hbr.org/2017/05/whats-your-data-strategy>
- <https://www.cio.com/article/3612316/creating-a-data-driven-culture.html>
- <https://data.world/blog/what-is-a-data-catalog/>

Atividades

1. Estado da Arte de Estratégia de Dados e Governança de dados e Catálogos de Dados;
2. Identificação de boas práticas na construção de uma estratégia de dados e políticas de governança de dados associadas;
3. Identificação e desenho de cenário de integração de um Catálogo de Dados no Data Lake ALB;
4. Relatório Intermédio;
5. Implementação e validação de Catálogo de Dados no âmbito de Data Lake ALB;
6. Relatório Final.

Competências Chave Requeridas

- Pessoa dinâmica e autónoma, com vontade de aprender;
- Comunicativo e com espírito de equipa;
- Bons conhecimentos de bases de dados e modelização de dados, desenvolvimento em sistemas Linux e experiência em linguagens de programação (Java, Python, ...);
- Capacidade e vontade de aprender tecnologias Big Data (Hadoop, PySpark ou similar) e Data Science (pandas, scikit-learn, etc).

Orientador (nome e e-mail)

Luis Cortesão - luis-m-cortesao@alticelabs.com

Para concorrer podes enviar a tua candidatura, envia e-mail para o Programa GENIUS: genius@inova-ria.pt