

ADR 2: Seleção de Provedor de Nuvem para Integração de Dados

Status

Proposto

Contexto

A decisão arquitetural anterior (ADR 1) estabeleceu a necessidade de uma arquitetura de integração de dados e sistemas baseada em nuvem para resolver os desafios da BioInnovate Corp., como silos de dados, formatos inconsistentes, qualidade deficiente, escalabilidade limitada, falta de insights em tempo real e conformidade¹. Agora, é fundamental selecionar o provedor de nuvem mais adequado (AWS ou Azure) para implementar essa solução.

Decisão

A decisão é implementar a arquitetura de integração de dados utilizando o Microsoft Azure como provedor de nuvem principal. Esta escolha se baseia na avaliação das capacidades de cada provedor em relação às necessidades da BioInnovate Corp., com ênfase na integração de dados, conformidade e facilidade de uso para equipes com perfis diversos.

Justificativas:

- **Integração e Ecossistema:** O Azure oferece um ecossistema robusto e bem integrado de serviços para dados e análises, como Azure Data Factory para ETL/ELT, Azure Synapse Analytics para *Data Lakehouse* e processamento analítico, Azure Databricks para processamento distribuído e inteligência artificial, e Azure Stream Analytics para processamento em tempo real. A integração nativa entre esses serviços simplifica o desenvolvimento e a manutenção dos pipelines de dados.
- **Conformidade e Segurança:** O Azure possui um forte portfólio de certificações de conformidade relevantes para o setor de biotecnologia (como HIPAA e GDPR ²), além de ferramentas avançadas de segurança e governança de dados (Azure Purview para governança de dados, Azure Security Center para monitoramento de segurança).

- **Produtividade do Desenvolvedor e Facilidade de Uso:** A interface do Azure e a integração com ferramentas Microsoft (como Visual Studio) podem acelerar o desenvolvimento e reduzir a curva de aprendizado para equipes já familiarizadas com o ambiente Microsoft.
- **Custos:** Embora ambos os provedores sejam competitivos em termos de custos, o Azure oferece modelos de preços flexíveis e opções de otimização que podem ser vantajosas para o crescimento da BioInnovate.

Implicações Imediatas:

- Foco na capacitação da equipe em tecnologias Azure.
- Planejamento detalhado da arquitetura de dados e migração para o Azure.
- Configuração inicial de redes, segurança e identidades no ambiente Azure.

Consequências

Positivas:

- **Aceleração da Implementação:** O ecossistema integrado do Azure pode agilizar o desenvolvimento e a implantação dos pipelines de dados e do *Data Lakehouse*.
- **Segurança e Conformidade Aprimoradas:** Facilidade na implementação de controles de segurança e na aderência às regulamentações específicas do setor.
- **Escalabilidade e Flexibilidade:** A capacidade de escalar recursos sob demanda no Azure atende ao crescimento do volume de dados da BioInnovate³.
- **Redução de Esforço de Manutenção:** Serviços gerenciados do Azure minimizam a sobrecarga de gerenciamento de infraestrutura.
- **Colaboração Aprimorada:** Ferramentas de colaboração e acesso a dados no Azure facilitam o compartilhamento de descobertas⁴.

Negativas:

- **Curva de Aprendizagem:** Embora potencialmente menor para equipes com familiaridade Microsoft, ainda haverá uma curva de aprendizado para as especificidades dos serviços de dados do Azure.

- **Dependência do Provedor:** A escolha do Azure estabelece uma dependência tecnológica do ecossistema Microsoft.

Requisitos Futuros:

- Estabelecimento de um centro de excelência em nuvem dentro da BioInnovate para gerenciar o ambiente Azure.
- Otimização contínua de custos e desempenho dos recursos do Azure.
- Exploração de serviços avançados de IA/ML no Azure para análise de dados genômicos e de ensaios clínicos.

Alternativas Consideradas

- **Amazon Web Services (AWS):**
 - **Vantagens:** Ampla gama de serviços, maturidade no mercado de nuvem, grande comunidade de usuários e vasta documentação. Serviços como S3 (Data Lake), Glue (ETL), Redshift (Data Warehouse) e EMR/Athena (Data Lakehouse/Querying). Possui forte capacidade de escalabilidade e segurança.
 - **Desvantagens:** Pode apresentar uma curva de aprendizado mais íngreme para equipes não familiarizadas com o ecossistema AWS. A integração entre alguns serviços pode exigir mais configuração manual em comparação com o Azure.
 - **Motivo da Descartagem:** Embora seja uma alternativa sólida, o Azure foi preferido devido à sua integração nativa de serviços de dados, ferramentas de produtividade para desenvolvedores e um alinhamento ligeiramente superior com as necessidades de conformidade e o possível *skillset* existente de algumas equipes com tecnologias Microsoft.

Referências

- BioInnovate Corp.pdf ⁵
- ADR 1: Integração de Dados e Sistemas Baseada em Nuvem