

Plantão de Dúvidas 14/05/2025

Aula 3

Big Data e Infraestrutura de Dados

1. Por que a infraestrutura de dados é considerada uma das bases mais importantes para aplicações de IA?

Resposta:

Sem uma base sólida de dados — ou seja, dados bem coletados, organizados e armazenados — nenhum modelo de IA funciona de forma eficiente. A qualidade da informação é o que alimenta qualquer processo analítico ou preditivo.

2. O que é um pipeline de dados e qual seu papel no contexto de Big Data financeiro?

Resposta:

Um pipeline de dados é uma sequência automatizada de etapas que coletam, processam e transportam dados de uma fonte até seu destino (como um banco ou ferramenta analítica). Ele garante que os dados cheguem prontos para uso em tempo real ou em batch, mantendo a consistência e a integridade.

3. Quais são os principais desafios no armazenamento de grandes volumes de dados financeiros?

Resposta:

- Os maiores desafios são:
- Escalabilidade (lidar com o crescimento de dados);
- Segurança e conformidade regulatória (LGPD, Bacen);
- Organização eficiente (dados estruturados vs não estruturados);
- Custo de armazenamento em nuvem.

4. Que tipo de arquitetura de dados é mais comum em aplicações financeiras modernas?

Resposta:

Muitas instituições usam arquiteturas em nuvem, especialmente data lakes e data warehouses, combinados com tecnologias como Spark, Kafka e ferramentas da AWS, GCP ou Azure. Essas arquiteturas são escaláveis e permitem processamento distribuído, essencial para grandes volumes e decisões rápidas.

5. Por que o processamento em tempo real tem ganhado importância no mercado financeiro? Cite um exemplo prático.

Resposta:

Porque decisões precisam ser tomadas em segundos — seja para identificar uma transação suspeita, monitorar riscos ou ajustar limites de crédito.

Exemplo:

Um sistema antifraude que detecta padrões atípicos assim que uma transação acontece, bloqueando automaticamente ou alertando o usuário.