

Integração, Gerenciamento e Análise de Big Data

Módulo 8: Sistemas de informação

- RESUMO EXECUTIVO

Este módulo capacita estudantes a construir pipelines de dados completos para ambientes de grande volume, desde a coleta e tratamento até a geração de informações estratégicas por meio de *dashboards* e análises automatizadas.

Durante 10 semanas, uma turma dedicada de alunos e seus professores irá entender o seu desafio e propor soluções utilizando metodologias ágeis. Ao final do cronograma, sua empresa receberá até 5 protótipos funcionais, permitindo que você acelere e valide ideias.

- O DESAFIO

Organizações geram volumes massivos de dados a cada dia, mas a maioria não consegue transformar esses dados em informações úteis. Dados espalhados em fontes diferentes, sem padronização e sem estrutura adequada, impedem análises confiáveis.

A incapacidade de processar e visualizar esses dados em tempo hábil resulta em decisões baseadas em intuição, e não em evidências.

- OBJETIVO DO MÓDULO

Construir um pipeline de dados completo em ambiente de nuvem, integrando múltiplas fontes de dados, aplicando técnicas de ETL e *machine learning*, e entregando painéis de controle gerencial (*dashboards*) que suportem a tomada de decisão estratégica da organização parceira.

- PROBLEMAS QUE PODEMOS RESOLVER

1. Dados Fragmentados e Sem Padronização

- Integração de múltiplas fontes de dados (bancos, APIs, arquivos, sensores).
- Padronização e limpeza de dados para análise confiável.
- Construção de data lakes e data warehouses estruturados.

2. Análise e Predição em Grande Volume

- Aplicação de algoritmos de *machine learning* para análise preditiva.
- Detecção de padrões e anomalias em grandes conjuntos de dados.
- Análise prescritiva e diagnóstica para suporte à decisão.

3. Visualização e Comunicação de Dados

- Criação de dashboards executivos com métricas e KPIs.
- Relatórios automatizados para diferentes públicos.
- Monitoramento em tempo real de indicadores de desempenho.

- CASOS DE APLICAÇÃO

1. Casos de Sucesso Reais (O que já entregamos)

- **Mobilidade e Transporte Público:** Construímos com a **CPTM** um pipeline de dados para análise de fluxo de passageiros, otimização de rotas e monitoramento de indicadores operacionais em tempo real.
- **Consultoria e Serviços Empresariais:** Desenvolvemos com a **Integration Consulting** a integração de dados de múltiplos clientes para geração de relatórios consolidados, benchmarking e análise de tendências de mercado.
- **Indústria e Manufatura:** Estruturamos para a **SKZ Oberle** o monitoramento de dados de produção, aplicando manutenção preditiva e otimização de processos fabris a partir da integração de dados de sensores e sistemas de controle.

2. Outras Aplicações Possíveis (Onde mais podemos gerar valor)

Organizações geram volumes massivos de dados diariamente, mas a falta de estrutura impede análises confiáveis. Nossos alunos podem construir pipelines completos na nuvem (do armazenamento à visualização) para organizar a "casa de dados" da sua empresa e resolver desafios em:

- **Varejo e E-commerce:** Integração e análise do comportamento do consumidor digital para permitir a personalização de ofertas e a previsão de demanda a partir de dados de navegação e histórico de compras.
- **Saúde e Gestão Hospitalar:** Consolidação de dados clínicos e administrativos fragmentados para análise epidemiológica, controle de custos operacionais e melhoria na eficiência do atendimento ao paciente.
- **Agronegócio:** Construção de Data Lakes para integrar grandes volumes de dados de safra, clima, tendências de mercado e logística, oferecendo suporte robusto à decisão e otimização de operações agrícolas.

- POSSÍVEIS ENTREGÁVEIS

Os grupos desenvolvem soluções validadas e funcionais, podendo incluir:

- Pipeline de dados completo (coleta, tratamento, armazenamento e análise) em ambiente de nuvem
- Data warehouse ou data lake estruturado com dados integrados.
- Dashboard gerencial interativo com métricas e KPIs.
- Modelos de machine learning para análise preditiva ou diagnóstica.
- Relatório técnico com arquitetura de dados, ETL e recomendações.

- PREMISSAS E REQUISITOS

1. Da Empresa Parceira:

- Disponibilização de fontes de dados para integração (bancos de dados, APIs, arquivos, planilhas).
- Acesso a especialistas do negócio e, caso necessário, especialistas técnicos para reuniões quinzenais.
- Definição clara das perguntas de negócio que os dados devem responder.
- Abertura para testar e validar as soluções propostas.

2. Do Projeto:

- Foco em problemas que envolvam múltiplas fontes de dados ou grandes volumes.
- Existência de dados disponíveis para análise (estruturados ou não).
- Potencial de impacto mensurável no negócio por meio de análise de dados.
- Alinhamento com práticas éticas e LGPD.

- RESTRIÇÕES E ESCOPO

1. Escopo Incluído:

- Construção de pipeline de dados (ETL) em ambiente de nuvem.
- Modelagem e estruturação de data warehouse/data lake.
- Desenvolvimento de dashboards e relatórios.
- Aplicação de modelos de machine learning.
- Documentação técnica e de negócios.

2. Fora do Escopo:

- Implementação em produção (deployment final).
- Migração completa de dados legados.
- Integração com sistemas internos da empresa.
- Manutenção e operação pós-projeto.

- POSSÍVEIS TECNOLOGIAS / ARQUITETURA

- Tecnologias que podem ser utilizadas:
 - Python,
 - SQL,
 - AWS/GCP/Azure,
 - Apache Spark,
 - bancos NoSQL,
 - ferramentas de BI (Power BI, Tableau, Metabase)