

A aula interativa do **Bootcamp Engenheiro de Dados** começará em breve!

Atenção:

- 1) Você entrará na aula com o microfone e o vídeo DESABILITADOS.
- 2) Apenas a nossa equipe poderá habilitar seu microfone e seu vídeo em momentos de interatividade, indicados pelo professor.
- 3) Utilize o recurso Q&A para dúvidas técnicas. Nossos tutores e monitores estarão prontos para te responder e as perguntas não se perderão no chat.
- 4) Para garantir a pontuação da aula, no momento em que o professor sinalizar, você deverá ir até o ambiente de aprendizagem e responder a enquete de presença. Não é necessário encerrar a reunião do Zoom, apenas minimize a janela.



Nesta aula



- □ Apresentação do curso e dos professores.
- ☐ Visão geral do pipeline de dados.
- □ Papel do profissional de Engenharia de Dados.
- → APIs de Coleta de Dados.
- Exemplo de coleta de dados via API.

Módulos



Fundamentos

- Pipeline de dados e tecnologias.
- Coleta e preparação.
- Web crawler.
- Pentaho Data Integration e Apache Airflow



Armazenamento de dados

- Data warehouse e data lake.
- SGBDs SQL e NoSQL
- Sistemas de arquivo SGBD em nuvem.
- MySQL, MongoDB, Cassandra, Redis, Neo4j
- ETL/ ELT (Extração, Transformação e Carga).



Ecossistema Hadoop

- Ingestão de dados (Sqoope flume)
- Armazenamento (HDFS e HBase)
- Processamento (Mapreduce e Spark)
- Data lake.



Infraestrutura e Arquitetura

- Virtualização, containers, Docker, Kubernetes.
- Recursos para escalabilidade, elasticidade e processamento distribuído.
- Azure, AWS,



Professores





Neylson Crepalde

- Graduado em Regência UFMG.
- Especialista em Gestão Cultural UNA.
- Mestre e Doutor em Sociologia UFMG.
- Head of MLOps na A3Data
- AWS Certified Solutions Architect and Accredited Educator



João Paulo Barbosa Nascimento

- Graduado em Sistemas de Informação PUCMinas
- Mestre e Doutor em Modelagem Matemática e Computacional - CEFET-MG
- Atua há 6 anos como professor em cursos de graduação e pós-graduação
- Atua por 16 anos como Analista de Sistemas



Ricardo Brito Alves

- Graduado em Ciência da Computação PUC-Minas
- MBA em Gestão Estratégica de Projetos Una.
- Mestre em Engenharia Elétrica PUC-Minas
- Atua a mais de 20 anos em projetos de software, data mining, BI, e IA.



Gustavo Aguilar

- Graduado em Ciência da Computação PUC-Minas
- Especialista em Banco de Dados Newton Paiva
- MBA em Ciência de Dados IGTI
- Atua a mais de 20 anos na área de banco de dados, arquitetura e engenharia de dados, cloud computing DevOps.

Pipeline de dados







Área responsável por planejar e preparar toda a infraestrutura e arquitetura de dados.

Engenharia de Dados

Propõe soluções que envolve quase todo pipeline de dados.

Foco em transformar dados em um formato útil para análise.



Projetar e implementar o pipeline de dados

Garantir a qualidade dos dados

Atividades do profissional de engenharia de dados

Modelar e implementar esquema de dados

Definir soluções de integração de dados Definir as estruturas de armazenamen to de Dados

Implementar rotinas de ETL

O que são API's?

- API Application Programming Interface.
 - Interface de Programação de Aplicações.
- Pacote que consiste em um conjunto de definições e protocolos usado no desenvolvimento e na integração de software de aplicações.







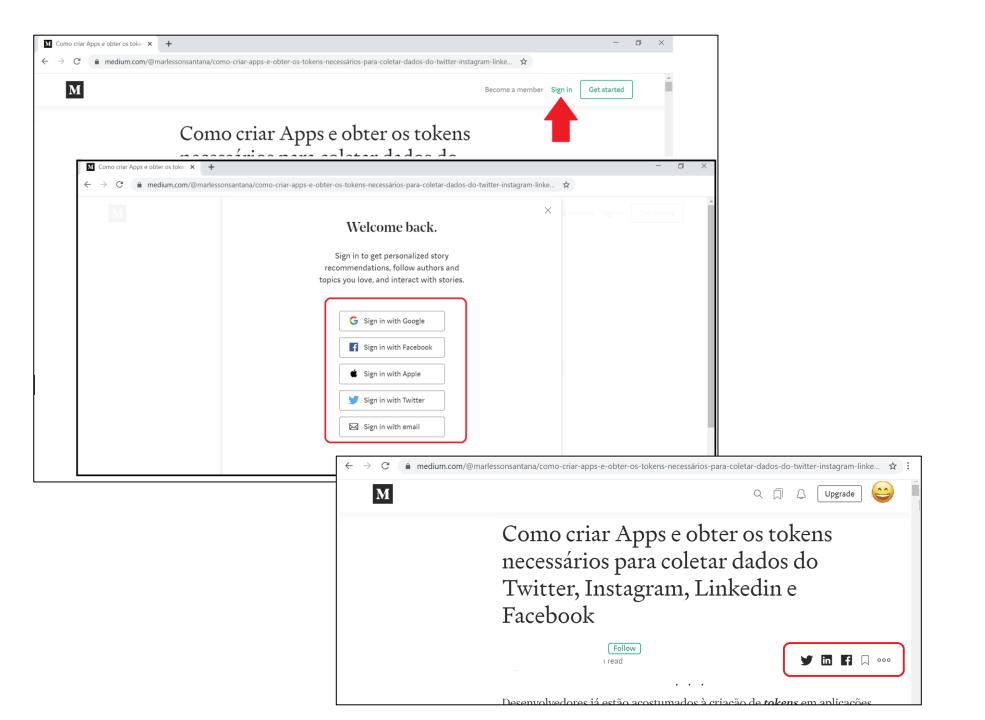




Sistema/Serviço









Principais API's para redes sociais



• Twitter - https://developer.twitter.com/en



Linkedin - https://www.linkedin.com/developers/apps



• Facebook - https://developers.facebook.com/docs/graph-api?locale=pt_BR



Instagram - https://www.instagram.com/developer/register/



Necessário ter uma conta para criar aplicativos e chaves de acesso.

Exemplo Twitter API



- Necessário ter uma conta no Twitter.
 - https://apps.twitter.com/
- Cadastrar como desenvolvedor (estudante).
 - https://developer.twitter.com/en
- Dashboard Twitter Dev
 - https://developer.twitter.com/en/portal/dashboard





