L'AKADÉMIE DU DIGITAL · PROGRAMA DE FORMAÇÃO Certificação Qualiopi em preparação

Data Engineer

FOR-DEV-M6-A · Desenvolvimento, DevOps e Dados

Conceber, construir e explorar os pipelines e as infraestruturas de dados

DURAÇÃO
35 horas — 5 dias

PREÇO
2?450 € s/ IVA

PÚBLICO
Escolas e centros de formação

PRAZO DE ACESSO
Acesso em 2 dias úteis após validação da necessidade e dos pré-requisitos.

Preço estabelecido com base em 35 h a 70 € s/ IVA por hora. Percurso dedicado a escolas e centros de formação, estabelecido por orçamento; preço indicativo e negociável consoante o volume e o contexto.

O módulo alterna sequências teóricas e oficinas práticas ao longo de cada dia. Abrange a modelização dos dados, o SQL avançado e os armazéns de dados, os pipelines ETL/ELT e a sua orquestração, o big data e o processamento distribuído, as arquiteturas modernas e a governação. O fio condutor das oficinas é SQL/Python/Airflow/Spark/Kafka, complementado por um panorama alargado do ecossistema.

Objetivos pedagógicos

- Compreender o papel do data engineer e a arquitetura de uma plataforma de dados
- Conceber e modelizar bases de dados (relacionais e NoSQL) e um armazém de dados
- Construir pipelines de dados (ETL/ELT) e orquestrá-los
- Compreender os desafios do big data e do processamento distribuído (batch/streaming)
- Conhecer as arquiteturas modernas (data warehouse, data lake, lakehouse) e o cloud de dados
- Implementar qualidade, governação, segurança e supervisão dos dados

Público-alvo

- Estudantes e formandos das escolas e centros de formação parceiros
- Programadores ou data analysts que queiram orientar-se para a engenharia de dados
- Perfis técnicos que visam funções de data engineering

Pré-requisitos

São exigidas bases de programação (Python recomendado) e de SQL, bem como à-vontade com a linha de comandos. Noções de bases de dados e de ciência de dados facilitam a progressão. Estes elementos são validados por um formulário online e / ou por uma entrevista prévia.

- Dominar as bases da programação (de preferência Python) e de SQL.

- Sentir-se à vontade com a linha de comandos e a manipulação de ficheiros.
- Ter noções de bases de dados e de dados (módulos Full-Stack e Data Scientist aconselhados).
- Ter autonomia técnica e rigor de organização.

Programa

Dia 1 — A profissão e a modelização dos dados

Profissão e arquitetura de dados	1h15	Data engineer vs data scientist; arquitetura de uma plataforma; panorama do ecossistema (bases de dados, ETL, distribuído, cloud). Exposição interativa, esquema de arquitetura, panorama de ferramentas.	Perguntas orais; questionário de escolha múltipla sobre o ecossistema.
Analisar uma arquitetura	2h00	Identificar os componentes de uma arquitetura de dados. Trabalho prático guiado, estudo de arquitetura, grelha de análise.	Análise de arquitetura validada.
Modelização dos dados	1h15	Relacional (PostgreSQL, MySQL), NoSQL (MongoDB, Cassandra, Redis), normalização, modelização decisional (estrela/floco de neve). Exposição, panorama das bases de dados, exemplos de esquemas.	Questionário de escolha múltipla sobre a modelização.
Modelizar um esquema	2h00	Conceber um esquema de dados adaptado a um caso. Trabalho prático, ferramenta de modelização, acompanhamento.	Esquema de dados validado.

Dia 2 — SQL avançado e armazéns de dados

SQL avançado	1h15	Junções complexas, funções de janela, otimização, índices. Exposição, exemplos de consultas, planos de execução.	Perguntas orais; questionário de escolha múltipla.
Consultas SQL avançadas	2h15	Escrever consultas avançadas sobre um conjunto de dados. Trabalho prático guiado, base de dados, consultas progressivas.	Consultas avançadas validadas.
Armazéns de dados	1h00	Data warehouse, modelização dimensional, panorama (BigQuery, Snowflake, Redshift). Exposição, esquemas dimensionais, panorama cloud.	Questionário de escolha múltipla sobre os armazéns de dados.
Mini-armazém	2h30	Construir e alimentar um mini-armazém de dados. Trabalho prático, modelização em estrela, carregamento, consultas.	Mini-armazém validado.

Dia 3 — Pipelines ETL/ELT e orquestração

Ingestão e ETL/ELT	1h15	Ingestão, ETL vs ELT, transformação, idempotência; panorama (dbt, Talend, Fivetran, Airbyte, NiFi). Exposição, esquemas de pipelines, panorama de ferramentas.	Perguntas orais; questionário de escolha múltipla.
Pipeline ETL	2h15	Construir um pipeline ETL (extração ? transformação ? carregamento). Trabalho prático guiado em Python, fontes e destino, acompanhamento.	Pipeline ETL funcional.
Orquestração de pipelines	1h00	Airflow (fio condutor), planeamento, dependências; alternativas Dagster, Prefect. Exposição, noção de DAG, panorama dos orquestradores.	Questionário de escolha múltipla sobre a orquestração.
Orquestrar com um DAG	2h30	Orquestrar um pipeline através de um DAG (Airflow). Trabalho prático guiado em Airflow, planeamento, gestão de dependências.	DAG orquestrado validado.

Dia 4 — Big data e processamento distribuído

Desafios do big data	1h15	Volume, velocidade, variedade; processamento distribuído (Spark); legado Hadoop/MapReduce, alternativas Flink, Dask. Exposição, esquemas distribuídos, panorama dos motores.	Perguntas orais; questionário de escolha múltipla.
Processamento distribuído (Spark)	2h15	Tratar um grande volume de forma distribuída (iniciação ao Spark). Trabalho prático guiado em Spark, conjunto de dados volumoso, acompanhamento.	Processamento distribuído validado.
Batch vs streaming	1h00	Batch vs tempo real; Kafka (fio condutor), alternativas Pulsar, Kinesis, RabbitMQ. Exposição, esquema de streaming, panorama dos barramentos de mensagens.	Questionário de escolha múltipla sobre batch/streaming.
Processamento batch e demonstração de streaming	2h30	Implementar um processamento batch; demonstração de streaming. Trabalho prático de batch, demonstração Kafka, revisão coletiva.	Processamento batch validado.

Dia 5 — Arquiteturas modernas, governação e projeto

Arquiteturas modernas e cloud	1h00	Data warehouse / data lake / lakehouse; formatos abertos (Delta Lake, Iceberg, Hudi); cloud (BigQuery, Snowflake, Redshift, Databricks, Synapse, Dataflow). Exposição, esquemas de arquiteturas, panorama cloud.	Questionário de escolha múltipla sobre as arquiteturas.
Explorar uma arquitetura cloud	2h00	Explorar uma arquitetura cloud de dados (iniciação). Trabalho prático guiado, ambiente cloud de iniciação, acompanhamento.	Exploração cloud validada.
Qualidade e governação	1h00	Qualidade, governação, segurança, RGPD, supervisão dos dados. Exposição, boas práticas, exemplos de supervisão.	Questionário de escolha múltipla sobre a governação.
Pipeline de dados de ponta a ponta e apresentação de resultados	2h30	Conceber e apresentar um pipeline de dados completo. Trabalho prático de projeto, apresentações, revisão coletiva, questionário final de escolha múltipla, inquérito de satisfação.	Avaliação sumativa: pipeline de dados + questionário final de escolha múltipla.

Modalidades e meios pedagógicos

Formação presencial nas nossas instalações ou nas suas, em grupo reduzido. Alternância de contributos teóricos, oficinas práticas e simulações.

Material necessário :

16 GB de RAM recomendados

- Os formandos devem dispor de um PC com Wi-Fi e 16 GB de RAM recomendados.
- Efetivo adaptado ao formato escola / centro de formação.
- Percurso dedicado a escolas e centros de formação, estabelecido por orçamento.

Modalidades de avaliação

Avaliação das aprendizagens durante e no final da formação; certificado de conclusão entregue a cada formando.

Acessibilidade

Formação acessível a pessoas com deficiência. Contacte-nos antes da inscrição para estudarmos as adaptações necessárias.

Financiamento

Orçamento ou convenção de formação a pedido. A certificação Qualiopi está em preparação: a elegibilidade para financiamentos OPCO será aberta após a sua obtenção.

Auditoria de certificação prevista para T1 2027. Sem elegibilidade OPCO antes da sua obtenção.