

# **“Introdução aos fundamentos de Business Intelligence e Engenharia dos Dados”**

“Data is the new oil”, “Data is the new sun”, “Data is the new gold”, “Data is the new currency”. Provavelmente já ouviu algumas destas expressões. Elas ilustram a importância e omnipresença dos dados nas nossas vidas diárias. De facto, os dados são o rasto da nossa presença digital. Eles dão corpo às mensagens que constantemente produzimos e partilhamos com o mundo (muitas vezes sem nos darmos conta). Não admira que a maioria dos negócios procurem formas de potencializar o valor dos dados.

Os dados são factuais mas primários e tipicamente surgem-nos sujos de ruído. Os dados são como a água: proveniente de várias fontes (chuva, rios, lagos, fábricas, etc), parte carrega químicos industriais consigo, outras partes, outrora límpidas e translúcidas, ficam contaminadas... parte é fresca e pura, outras partes revelam odores desagradáveis.

O trabalho do Engenheiro de Dados consiste em extrair água de várias fontes distintas, testá-la, descontaminá-la de forma a que seja própria para consumo, e determinar a melhor forma de a distribuir para que outros possam usá-la para fins úteis e interessantes. Esta viagem da água é feita de várias etapas, cada uma com os respectivos desafios e armadilhas.

Neste curso vamos descobrir “a história da água e do engenheiro”.

O objetivo ‘número 1’ do curso é atingir uma visão clara de como as diferentes partes que constituem uma solução completa, “end-to-end”, de Business Intelligence, se coordenam e ligam entre si. Seguiremos uma abordagem “hands on” e construiremos uma pequena solução de Business Intelligence com base numa amostra de dados. Faremos trabalho de implementação, mas não sem antes estabelecermos um referencial conceptual sólido. Iremos aprender não só “como” o engenheiro dos dados faz o seu trabalho, mas também o “porquê” de o fazer dessa forma. Iremos compreender as suas metodologias e aprender a usar algumas das suas ferramentas mais importantes. Em determinadas alturas mergulharemos em detalhes técnicos, mas sempre com a orientação necessária de forma a nunca nos perdermos da visão e objectivos globais.

## **Ao terminar o curso, os formandos terão aprendido:**

- O que é “Business Intelligence”;
- Porque precisamos de dar uma nova estrutura aos dados e construir uma base de dados dedicada, quando queremos fazer “Business Analytics”;
- As metodologias e melhores práticas na reestruturação dos dados para a Análise do Negócio: a Modelação Dimensional;
- Uma metodologia para colaborar com os “business owners” e recolher os requisitos para os modelos de dados do negócio;
- Como contruir a base de dados para a análise: a “Data Warehouse”;
- Como “Extrair” dados (da sua origem), “Transformá-los” (para a sua nova estrutura) e “Carregá-los” (na Data Warehouse): o Processo ETL (Extract, Transform and Load);
- As possíveis vantagens de ter uma etapa adicional na “pipeline de dados”, onde estes são mapeados para uma espécie de cubos multidimensionais: os “cubos OLAP”;
- Como a última etapa (onde os dados são apresentados aos utilizadores) se encaixa no ciclo de vida do projecto de BI e algumas recomendações para o desenho de relatórios e dashboards eficazes.

## **E serão capazes de:**

- Contextualizar o vasto tema do “Business Intelligence”;
- Usar bases de dados relacionais (numa perspectiva de utilizador);
- Participar em sessões de brainstorm com business owners, para documentar os processos de negócio e desenvolver os novos modelos de dados;
- Tirar partido das boas práticas de modelação dimensional de forma que os seus modelos de dados tenham bom desempenho e sejam fáceis de compreender e usar;
- Desenhar e construir data warehouses simples e participar de forma ativa em projetos de sistemas de informação maiores e mais complexos;
- Desenhar e implementar processos de ETL simples e passar para desafios de Integração de Dados mais avançados;
- Criar cubos OLAP com base em data warehouses previamente construídas;
- Compreender e participar em discussões técnicas de projectos de engenharia de dados de nível “enterprise”.

**Notas:**

- Antes de iniciarmos a viagem pelo Business Analytics faremos uma breve revisão sobre os conhecimentos essenciais para o uso de bases de dados relacionais.
- Esta formação não cobre: os tópicos específicos de Ciência dos Dados (“Data Mining”, “Prediction” e “Statistical Learning”), os tópicos específicos de Big Data (desafios característicos de Big Data, desenho de soluções e tecnologias); devido ao limite de tempo, apenas faremos um olhar superficial sobre a etapa de “Apresentação” na pipeline dos dados (onde estes, agora transformados em informação, são apresentados ao utilizador).
- Tecnologia e ferramentas: no decorrer da formação usaremos software open source. Em particular faremos uso intensivo do Pentaho (<https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform.html>);

**Pré-requisitos:**

A experiência prévia com alguma linguagem de programação é recomendável (mas não absolutamente necessária);

A Compreensão, a um nível básico, das operações mais simples de manipulação de conjuntos (se o formando compreender o que é um conjunto e, perante um conjunto de valores, for capaz de o ordenar, identificar o seu valor máximo, o seu valor mínimo e calcular o seu valor médio, então estará em condições de participar na formação);