

## Origens do processo de análise de dados

Quando você decidiu aderir a este programa, provou ser uma pessoa curiosa. Portanto, vamos explorar sua curiosidade e falar das origens da análise de dados. Não sabemos totalmente quando ou por que a primeira pessoa decidiu registrar dados sobre pessoas e coisas. Mas sabemos que isso foi útil, porque a ideia existe ainda hoje.



Também sabemos que as raízes da análise de dados estão na estatística, que também têm uma história longa. Os arqueólogos consideram a construção das pirâmides, no antigo Egito, como o início da estatística. Os egípcios antigos eram mestres na organização de dados. Eles documentaram os cálculos e as teorias em papiros (materiais semelhantes a papel), que agora são vistos como os primeiros exemplos de planilhas e listas de verificação. Os analistas de dados atuais devem muito a esses escribas brilhantes, que ajudaram a criar um processo mais técnico e eficiente.

É hora de entrar no **ciclo de vida da análise de dados**, o processo de ir dos dados à decisão. Os dados passam por várias fases conforme são criados, consumidos, testados, processados e reutilizados. Com um modelo de ciclo de vida, todos os principais membros da equipe podem alcançar o sucesso, planejando o trabalho tanto no início quanto no fim do processo de análise de dados. O ciclo de vida da análise de dados é bem conhecido entre os especialistas, mas não existe uma única estrutura definida dessas fases. Pode não haver uma única arquitetura uniformemente seguida por todos os especialistas em análise de dados, mas há alguns fundamentos compartilhados em todos os processos de análise de dados. Neste texto, fornecemos uma visão geral de vários deles, começando com o processo que constitui a base do Certificado de Análise de Dados do Google.

O processo apresentado como parte do Certificado de Análise de Dados do Google é valioso à medida que você avança na carreira:

1. **Perguntar:** pergunta/objetivo/desafio de negócios
2. **Preparar:** geração, coleta, armazenamento e gerenciamento de dados
3. **Processar:** limpeza/integridade de dados
4. **Analisar:** exploração, visualização e análise de dados
5. **Compartilhar:** comunicação e interpretação de resultados
6. **Agir:** aplicação dos insights com o objetivo de resolver o problema

Compreender esse processo (e todas as iterações que ajudaram a torná-lo conhecido) será uma grande parte na orientação de sua própria análise e de seu trabalho neste programa. Vamos examinar algumas outras variações do ciclo de vida da análise de dados.

## **CICLO DE VIDA DA ANÁLISE DE DADOS DA EMC**

O ciclo de vida da análise de dados da EMC Corporation é cíclico, com seis etapas:

1. **Descoberta**
2. **Pré-processamento de dados**
3. **Planejamento de modelo**
4. **Criação de modelos**
5. **Comunicação dos resultados**
6. **Operacionalização**

A EMC Corporation agora é **Dell EMC**. Este modelo, criado por David Dietrich, reflete a natureza cíclica dos projetos reais. As fases não são marcos estáticos: cada etapa se conecta e leva à próxima e, no fim das contas, se repete. As perguntas-chave ajudam os analistas a testar se realizaram o suficiente para seguir em frente e garantir que as equipes tenham passado tempo suficiente em cada uma das fases e não comecem a modelar antes de os dados ficarem prontos. É um pouco diferente do ciclo de vida da análise de dados em que este programa se baseia, mas tem algumas ideias centrais em comum: a primeira fase está interessada em descobrir e fazer perguntas, os dados precisam ser preparados antes de serem analisados e usados, e então as descobertas precisam ser compartilhadas e colocadas em prática.

Para mais informações, consulte

[A gênese do ciclo de vida da análise de dados da EMC](#)

## CICLO DE VIDA ITERATIVO DA SAS

Um ciclo de vida iterativo foi criado por uma empresa chamada **SAS**, uma provedora de soluções de análise de dados líder no mercado. Ele pode ser usado para produzir resultados repetíveis, confiáveis e preditivos:

1. **Perguntar**
2. **Preparar**
3. **Explorar**
4. **Modelar**
5. **Implementar**
6. **Agir**
7. **Avaliar**

O **modelo SAS** enfatiza a natureza cíclica do modelo, visualizando-o como um símbolo de infinito. O ciclo de vida dele tem sete etapas, muitas das quais vimos em outros modelos, como **Perguntar, Preparar, Modelar e Agir**. Mas esse ciclo de vida também é um pouco diferente. Ele inclui uma etapa após a fase de ação, desenvolvida para ajudar os analistas a avaliar as soluções e, potencialmente, retornar de novo à fase das perguntas.

Para mais informações, consulte

[Como gerenciar o ciclo de vida da análise para decisões em escala](#)

.

## CICLO DE VIDA DA ANÁLISE DE DADOS COM BASE EM PROJETO

Um ciclo de vida da análise de dados com base em projeto tem cinco etapas simples:

1. **Identificação do problema**
2. **Desenvolvimento dos requisitos de dados**
3. **Pré-processamento de dados**
4. **Execução da análise de dados**
5. **Visualização dos dados**

Esse ciclo de vida do projeto de análise de dados foi desenvolvido por Vignesh Prajapati. Não inclui a sexta fase, ou o que chamamos fase da ação (**Agir**). No entanto, ainda aborda muitas das mesmas etapas dos ciclos de vida que já descrevemos. Ele começa com a identificação do problema e a preparação e o processamento dos dados antes da análise e termina com a visualização dos dados.

Para mais informações, consulte

[Como compreender o ciclo de vida do projeto da análise de dados](#)

.

## CICLO DE VIDA DA ANÁLISE DE BIG DATA

Os autores Thomas Erl, Wajid Khattak e Paul Buhler propuseram um ciclo de vida da análise de big data no livro **Big Data Fundamentals: Concepts, Drivers & Techniques**.

O ciclo de vida deles sugere fases divididas em nove etapas:

1. **Avaliação do caso de negócios**
2. **Identificação dos dados**
3. **Aquisição e filtragem dos dados**
4. **Extração dos dados**
5. **Validação e limpeza dos dados**
6. **Agregação e representação dos dados**
7. **Análise dos dados**
8. **Visualização dos dados**
9. **Utilização dos resultados da análise**

Esse ciclo de vida parece ter três ou quatro etapas a mais do que os modelos de ciclo de vida anteriores. Mas, na realidade, eles apenas dividiram o que chamamos de fases “Preparar” e “Processar” em etapas menores. Isso enfatiza as tarefas individuais necessárias para coletar, preparar e limpar os dados antes da fase de análise.

Para mais informações, consulte

[Considerações sobre adoção e planejamento de Big Data](#)

### Principal conclusão

De nossa jornada até as pirâmides e os dados do antigo Egito até agora, a forma como analisamos os dados evoluiu (e continua evoluindo). O processo de análise de dados é como a arquitetura da vida real: há diferentes maneiras de realizar as tarefas, mas as mesmas ideias centrais ainda aparecem em cada modelo do processo. Quer você use a estrutura deste Certificado de Análise de Dados do Google, quer use uma das muitas outras iterações que aprendeu, estamos aqui para ajudar a continuar em sua jornada de dados.