

Análise de big data (Big data analytics)

De acordo com o SAS (2022), técnicas de *big data analytics* (análise de *big data*) examinam grandes quantidades de dados na tentativa de descobrir padrões ocultos, correlações e outros *insights*.

Para Gandomi e Haider (2015), o potencial do *big data* surge apenas quando aproveitado para apoiar na tomada de decisões. Para isso, são necessários processos eficientes para transformar grandes volumes de dados, diversos e rápidos, em *insights* significativos.

O processo geral de extração de *insights* de *big data* pode ser dividido em cinco etapas, como pode ser visto na Figura 1.

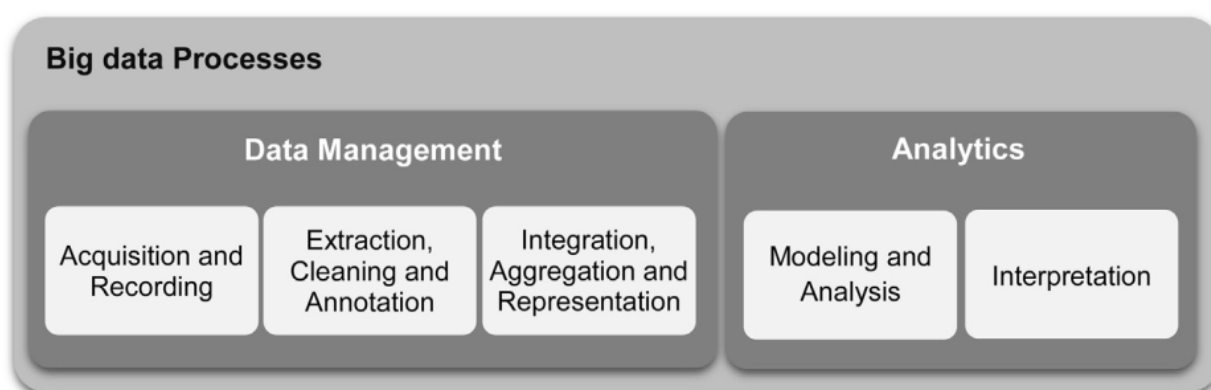


Figura 1: Processo geral de *big data*. Fonte: GANDOMI e HAIDER (2015).

Esses cinco estágios formam os dois principais subprocessos: gerenciamento de dados (*Data Management*) e análise (*Analytics*). O gerenciamento de dados envolve processos e tecnologias para adquirir, armazenar e preparar os dados para análise. A análise (*Analytics*), por outro lado, refere-se às técnicas usadas para analisar e adquirir inteligência a partir de *big data*. Deste modo, a análise de *big data* (*Big data analytics*) pode ser vista como um subprocesso no processo geral de “extração de insights” de *big data* (GANDOMI e HAIDER, 2015).

De acordo com o SAS (2022), não há uma tecnologia única que englobe a análise de *big data*. Existem análises avançadas que podem ser aplicadas a *big data*, mas, na realidade, vários tipos de tecnologia e técnicas precisam trabalhar juntas para a máxima obtenção de valor sobre os dados. SAS (2022) menciona os seguintes componentes relacionados à análise de *big data*:

- **Aprendizado de máquina:** subconjunto específico de Inteligência Artificial, que possibilita a produção de modelos que podem analisar dados maiores e mais complexos e fornecer resultados mais rápidos e precisos – mesmo em uma escala muito grande. Ao construir modelos precisos, uma organização terá mais chances de identificar oportunidades lucrativas – ou evitar riscos desconhecidos;

- **Gestão de dados:** os dados precisam ser de alta qualidade e bem governados antes que possam ser analisados de forma confiável. Com os dados entrando e saindo constantemente de uma organização, é importante estabelecer processos repetíveis para criar e manter padrões de qualidade de dados. Uma vez que os dados sejam confiáveis, as organizações devem estabelecer um programa de gerenciamento de dados mestres que coloque toda a empresa na mesma página;
- **Mineração de dados:** tecnologias de mineração de dados ajudam a examinar grandes quantidades de dados para descobrir padrões – e essas informações podem ser usadas para análises adicionais para ajudar a responder a questões comerciais complexas. Com o software de mineração de dados, pode-se filtrar todo o ruído caótico e repetitivo nos dados, identificar o que é relevante, e usar essas informações para avaliar resultados prováveis e acelerar o ritmo de tomada de decisões;
- **Hadoop:** este framework de código aberto pode armazenar grandes quantidades de dados e executar aplicativos em clusters de hardware comuns. Tornou-se uma tecnologia chave em *big data*;
- **Análise in-memory:** Ao analisar os dados da memória do sistema (em vez da unidade de disco rígido), você pode obter *insights* imediatos de seus dados e agir rapidamente. Essa tecnologia é capaz de remover latências de preparação de dados e processamento analítico para testar novos cenários e criar modelos; não é apenas uma maneira fácil para as organizações permanecerem ágeis e tomarem melhores decisões de negócios, mas também permite que elas executem cenários de análise iterativos e interativos;
- **Análise preditiva:** a tecnologia de análise preditiva usa dados, algoritmos estatísticos e técnicas de aprendizado de máquina para identificar a probabilidade de resultados futuros com base em dados históricos. Trata-se de fornecer uma melhor avaliação sobre o que acontecerá no futuro, para que as organizações possam se sentir mais confiantes de que estão tomando a melhor decisão de negócios possível;
- **Mineração de texto:** com a tecnologia de mineração de texto, pode-se analisar dados de texto da web, campos de comentários, livros e outras fontes baseadas em texto para descobrir insights, antes não notados. A mineração de texto usa aprendizado de máquina ou tecnologia de processamento de linguagem natural para vasculhar documentos – e-mails, blogs, feeds do Twitter, pesquisas, inteligência competitiva e muito mais – para auxiliar na análise de grandes quantidades de informações e descobrir novos tópicos e relacionamentos de termos.

Gandomi e Haider (2015) descrevem as cinco técnicas de análise de *big data* mais relevantes, de acordo com os autores. As técnicas descritas são listadas abaixo. Para mais detalhes, consultar o artigo referenciado:

- **Análise de texto:** também chamado de mineração de texto, esta técnica refere-se à extração de informações a partir de dados textuais, como feeds de notícias, e-mails, blogs, fóruns, notícias, questionários, logs de call center, etc;

- **Análise de áudio:** extração de informações de dados de áudio não estruturados. Quando aplicado à análise da linguagem falada, é também chamado de análise de fala;
- **Análise de vídeo:** também conhecido como *video content analysis* (VCA), ou análise de conteúdo de vídeo, envolve uma variedade de técnicas para monitorar, analisar e extrair informações a partir de *streams* de vídeo;
- **Análise de mídias sociais:** análise de dados estruturados e não estruturados provenientes de canais de mídias sociais (facebook, linkedin, twitter, instagram, tumblr, etc.);
- **Análises preditivas:** análise preditiva compreende uma variedade de técnicas que preveem resultados futuros com base em dados históricos e atuais. Na prática, a análise preditiva pode ser aplicada a quase todas as disciplinas – desde prever a falha de motores a jato com base no fluxo de dados de vários milhares de sensores, até prever os próximos movimentos dos clientes com base no que compram, quando compram e até mesmo no que dizem nas redes sociais. Em sua essência, a análise preditiva busca descobrir padrões e capturar relacionamentos nos dados.

Revisão #6

Criado 2022-03-08 16:58:03 UTC por FLAVIO LOPES DE MORAIS

Atualizado: 2022-12-14 09:02:06 UTC por FLAVIO LOPES DE MORAIS