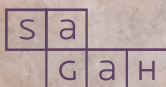


INTRODUÇÃO A BIG DATA E INTERNET DAS COISAS (IOT)

Ramiro Córdova Júnior



SOLUÇÕES
EDUCACIONAIS
INTEGRADAS



Big Data e Internet das Coisas (IoT)

Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Construir o conceito de Big Data e Internet das Coisas (IoT).
- Identificar elementos que contemplam uma solução de IoT.
- Definir possibilidades de uso de IoT para automação.

Introdução

Com o avanço da tecnologia e a utilização cada vez maior da internet em todo mundo, torna-se necessário aprimorar os serviços prestados pela tecnologia da informação (TI). Diante desse contexto, os conceitos conhecidos como Big Data e Internet das Coisas (IoT), ocupam um papel de destaque nas discussões relacionadas aos avanços em diversas áreas. Tanto o tratamento de grandes volumes de dados (Big Data) como a conectividade de dispositivos auxiliando no cotidiano das pessoas (IoT), são temas recentes e que têm muito a evoluir em termos de discussões das suas aplicações.

Neste capítulo, você aprenderá sobre as características das tecnologias chamadas de Big Data e IoT.

Conceito de Big Data

Atualmente, o tema Big Data desperta o interesse em todas as pessoas que têm algum envolvimento com atividades relacionadas à gestão da informação. Porém, quem não tem afinidade com a TI pode ter dificuldades para entender do que se trata. A princípio, podemos definir o conceito de Big Data como conjuntos de dados extremamente amplos e que, por esse motivo, necessitam de ferramentas preparadas para lidar com grandes volumes de dados, de forma

que toda e qualquer informação nesses meios possa ser encontrada, analisada e aproveitada em tempo hábil.

Com o aumento significativo da quantidade de dados gerados pela internet e com o surgimento das mídias sociais, é necessário gerenciar e armazenar as informações de maneira organizada. Esses dados podem ser classificados em estruturados, não estruturados e semiestruturados, com base no seu gerenciamento e armazenamento.

Dados estruturados

Dados estruturados são aqueles que possuem formato e comprimento definido, por exemplo, números, datas e grupos de palavras. Em geral, uma base de dados usada em sistemas informatizados convencionais é organizada de forma que se tenham dados armazenados em estruturas tubulares, nas quais as linhas armazenam uma ocorrência de um evento, caracterizado por um conjunto de colunas que representam características que descrevem um exemplar (instância) daquele evento.

Na maioria dos casos, os dados estruturados são resultados de processos de geração de dados inerente a sistemas transacionais ou resultantes de observações e processos de medição. Esses dados, geralmente, são armazenados em um conjunto de tabelas relacionadas entre si.

Dados não estruturados

Dados não estruturados são aqueles que não seguem um formato específico, por exemplo, imagens de satélite, dados científicos, fotografias e vídeos, texto próprio de empresas e dados de mídias sociais. Esse tipo de dado requer dispositivos de armazenamento e processamento, que suportem seu formato e garantam melhor eficiência em suas análises.

Dados semiestruturados

Pode-se definir dados semiestruturados como um meio termo entre a estruturação e total falta de estruturação. Como exemplo de dados semiestruturados, temos os conteúdos disponibilizados na Web acompanhados de tags e arquivos XML.

Análise de dados

A coleta e o armazenamento de dados têm como finalidade a extração de informações que possam gerar vantagens competitivas para as organizações, bem como auxiliar nas tomadas de decisões. Os diversos dispositivos conectados à internet produzem dados das mais variadas formas. Porém, dados, mesmo que em um grande volume, não passam de dados. É necessário gerar informação e conhecimento para explorar os benefícios desses dados brutos, ou seja, os dados necessitam ser analisados.

Os sistemas gerenciais têm como característica o fato de apresentarem as informações, mas a inteligência nos negócios converge para a análise detalhada dos dados, a procura de padrões, modelos ou repetições. Se não forem encontrados padrões, fica complicado afirmar quais eventos geram determinadas consequências. Isso gera uma confusão de dados desordenados e sem explicações. Diante desse contexto, existem três tipos de análises que podem ser aplicadas aos dados, conforme você verá a seguir.

- **Análise exploratória:** permite a compreensão de como os dados estão distribuídos e qual o formato em que se encontram. A informação adquirida nas análises exploratórias serve de apoio na tomada de decisão sobre o tipo de tarefa de mineração dos dados e na escolha do algoritmo.
- **Análise preditiva:** prevê resultados futuros com base em dados do passado. Pode ser entendida como um processo que permite descobrir o relacionamento existente entre os exemplares de um conjunto de dados, descritos por uma série de características e rótulos atribuídos a esses dados.
- **Análise de agrupamento:** consiste no agrupamento de objetos de acordo com suas características. Também é conhecida como *clustering* e permite, com base em um conjunto de características, a descoberta de relações existentes entre exemplares de conjuntos de dados.
- **Regras de associação:** procurar encontrar padrões ou relacionamentos que aparecem com frequência em bases de dados.

5V's do Big Data

O principal objetivo do desenvolvimento de soluções de Big Data é oferecer uma abordagem consistente no tratamento do constante crescimento e da

complexidade dos dados. Para isso, você precisa considerar o conceito dos 5 V's do Big Data: o volume, a velocidade, a variedade, a veracidade e o valor.

- **Volume:** trata-se do volume diário de troca de e-mails, transações bancárias, interações em redes sociais, registro de chamadas e tráfego de dados em linhas telefônicas. Essas situações servem de ponto de partida para a compreensão do volume de dados presentes no mundo atual. É importante também compreender que o conceito de volume é relativo à variável tempo, ou seja, o que é grande hoje, pode não ser nada amanhã.
- **Velocidade:** as empresas necessitam de dados atuais sobre seus negócios, por isso a importância da velocidade é tamanha e, em algum momento, deverá existir uma ferramenta capaz de analisar os dados em tempo real. Atualmente, os dados são analisados somente após serem armazenados, mas o tempo gasto para o armazenamento em si já desclassifica esse tipo de análise como uma análise 100% em tempo real. A velocidade com a qual você obtém essa informação é uma vantagem competitiva das empresas. A velocidade pode limitar a operação de muitos negócios, quando utilizamos o cartão de crédito, por exemplo, se não obtivermos uma aprovação da compra em alguns segundos, pensamos em utilizar outro método de pagamento. É a operadora perdendo uma oportunidade de negócios pela falha na velocidade de transmissão e análise dos dados do comprador.
- **Variedade:** diferentemente do passado, hoje em dia os dados não são estruturados e armazenados em tabelas relacionais, pois uma grande variedade de dados como mensagens, fotos, vídeos e sons torna mais complexa a análise desses dados. Os dados não estruturados podem ser administrados juntamente aos dados tradicionais.
- **Veracidade:** para colher bons frutos do processo do Big Data é necessário obter dados verídicos, de acordo com a realidade. O conceito de velocidade, já descrito, é bem alinhado ao conceito de veracidade pela necessidade constante de análise em tempo real, isso porque dados passados não podem ser considerados dados verídicos para o momento da análise.
- **Valor:** quanto maior for a riqueza de dados, mais importante será saber realizar as perguntas certas no início de todo o processo de análise. É necessário estar focado para a orientação do negócio, o valor que a coleta e a análise dos dados trarão para o negócio. Não é viável realizar todo o processo de Big Data se não se tem questionamentos que ajudem o negócio de modo realístico.

Tecnologias para o Big Data

As tecnologias que sustentam o Big Data podem ser categorizadas sob duas perspectivas, as ferramentas de análise (*analytics*) e as tecnologias de infraestrutura, que servem para processar e armazenar os grandes volumes de dados. O Big Data não se resume a um grande volume de dados não estruturados, ele também inclui as tecnologias que possibilitam o processamento e a análise dos dados. A tecnologia associada ao Big Data possibilita a criação de modelos estatísticos que servem para otimizar e prever os dados. A seguir, você verá uma relação de algumas tecnologias referentes ao Big Data.

- **Hadoop:** plataforma de software em Java de computação distribuída voltada para *clusters* e processamento de grandes volumes de dados, com suporte a tolerância a falhas.
- **Map Reduce:** é um *framework* desenvolvido pela Google para suportar computações paralelas em grandes coleções de dados em *clusters* de computadores.
- **Linguagens de Script:** linguagens de programação adequadas ao Big Data, como Python.
- **Visual Analytics:** método de análises em grandes volumes de dados com saída em formato visual ou gráfico.
- **Processamento de Linguagem Natural (PLN):** conceito de inteligência artificial que permite a análise de textos.
- **In-memory analytics:** processamento de Big Data realizado na memória do computador com o objetivo de aumentar a velocidade das análises.

No que se refere à infraestrutura, as arquiteturas de hardware atuais auxiliam no processamento de dados com serviços especializados para Big Data. O advento da tecnologia In-memory analytics colabora muito com a velocidade de processamento, pois evita que os dados sejam armazenados em disco.

Internet das Coisas (IoT)

O termo IoT, do inglês *Internet of Things* e que é traduzido como Internet das Coisas, é utilizado para transmitir a ideia de que a internet pode estar presente em todas as coisas. A ideia por trás do conceito é que todos equipamentos podem estar conectados a internet e, assim, facilitar a vida das pessoas no seu dia a dia.

O conceito de IoT é baseado na ideia de fusão do mundo real com o mundo digital, fazendo os indivíduos estarem em constante comunicação e interação com outras pessoas e objetos. A IoT possui funções de reconhecimento inteligente, localização, rastreamento e gerenciamento dos diversos dispositivos, trocando informações a todo o momento.

A IoT deve permitir uma conexão autônoma e segura para troca de dados entre os dispositivos e aplicações do mundo real, utilizando a internet para comunicação, troca de informações, tomada de decisões e invocação de ações. A tendência é que sejam oferecidos serviços surpreendentes, baseados no conceito de IoT. Assim, esse tema se torna muito relevante em instituições acadêmicas, na indústria e para as organizações governamentais, pois a IoT tem um potencial significativo de proporcionar melhorias em questões sociais, profissionais e econômicas.

Arquitetura da IoT

Ainda não foi definida uma arquitetura universal para IoT, devido ao fato de que as aplicações desenvolvidas funcionam em uma arquitetura de rede, quase que exclusivamente orientada à aplicação. As aplicações de IoT precisarão se integrar com outros sistemas e aí é que surgem os desafios relacionados à falta de interoperabilidade, ou seja, as aplicações possuem arquiteturas distintas. Para se atingir o nível de integração que a IoT irá demandar, será necessário uma base comum para construção de subarquiteturas, ou seja, um conjunto comum de blocos de construção, para o desenvolvimento da arquitetura do sistema, que garantirá interoperabilidade em vários níveis.

Diante desse cenário, diversas empresas que compõem a Standards Development Organizations (SDO), que atua no desenvolvimento de padrões, se uniram para formar uma arquitetura genérica para IoT, a oneM2M, cujo principal objetivo é desenvolver especificações para garantir a funcionalidade global dos sistemas.

A arquitetura básica M2M possui três tipos de camadas:

- camada de aplicação;
- camada de serviços comuns;
- camada de serviços de rede.

A camada de aplicação engloba os elementos de ponta como sensores, geolocalizadores, smartphones, e formam a chamada entidade de aplicação. Essa entidade, implementa uma ou mais lógicas de serviços, que são reconhecidos por uma identificação exclusiva, o que permite que uma solicitação de uma lógica de serviço faça uso do aplicativo desenvolvido para se comunicar.

Na camada de serviços comuns, existe uma outra entidade chamada de entidade de serviços comuns. Essa entidade representa uma solicitação de um conjunto de funções comuns, utilizando uma arquitetura funcional. A camada de aplicação interage com a camada de serviços comuns por meio de bibliotecas de desenvolvimento (API).

As especificações desenvolvidas pela M2M serviram como um projeto estrutural comum para redes e sistemas para executar um amplo conjunto de aplicações e pretendiam desenvolver uma camada de serviço comum, que facilitasse a interoperabilidade entre redes de sensores, dispositivos e aplicativos baseados em nuvem.

Elementos da IoT

A implementação de soluções para IoT é constituída por redes heterogêneas e amplamente distribuídas. Uma importante característica dessas redes é a identificação por radiofrequência, sensores, tecnologias inteligentes e nanotecnologias, que são importantes para que exista uma grande variedade de serviços. A rede utilizada em IoT abrange uma infraestrutura que inclui softwares, hardware e serviços, que serão utilizados para dar suporte às redes de comunicação. Os elementos físicos podem trocar dados sobre as suas propriedades devido à integração física.

A identificação de dispositivos pode ser realizada por meio de tecnologias de identificação, como *radio-frequency identification* (RFID). O RFID é um método que permite a identificação automática e a captura de dados utilizando campos eletromagnéticos que identificam e rastreiam etiquetas anexadas aos objetos. A tecnologia RFID utiliza um chipset que armazena os dados exclusivos e os recursos de segurança disponíveis. O chip possui uma antena que o permite receber energia e que sejam transmitidas informações. As etiquetas contêm informações eletrônicas e não necessitam estar visualmente alinhadas com um equipamento leitor, além de poderem ser divididas em etiquetas passivas, que recebem energia a partir das ondas de rádio de um leitor RFID próximo. As etiquetas ativas possuem fonte de energia local, como uma bateria, e podem operar a centenas de metros do leitor RFID.

Equipamentos como sensores e atuadores são extremamente importantes para a coleta e saída de dados, e têm a capacidade de permitir que os dispositivos conectados se comportem de forma autônoma. Atualmente, esses equipamentos são capazes de permitir a tomada de decisões inteligentes.

As redes de comunicação utilizadas para IoT são comuns e conhecidas: Wi-Fi, Bluetooth e near field communication (NFC). Porém, essas redes possuem alcance limitado e determinadas aplicações dependem de redes móveis como 3G e 4G. As redes móveis atuais (2G, 3G e 4G) são direcionadas a dispositivos como smartphones, tablets e laptops. O foco está sobre aplicações de texto, voz, imagem e vídeo. Esse aspecto não impede as redes atuais de serem utilizadas para IoT, mas uma otimização para dispositivos variados é necessária, principalmente para garantir o baixo consumo de energia e de recursos de processamento. Isso deve vir com a próxima onda de redes móveis, o 5G (quinta geração).

Com a eminente difusão da IoT, temos sensores, chips e dispositivos relacionados por todos os lados. Todos esses itens precisam estar conectados. Com a tecnologia de endereçamento de rede IPv6, que oferece um número extremamente elevado de endereços para os dispositivos, conectar esses dispositivos não será problema. A limitação vem das tecnologias de comunicação: as redes atuais não foram projetadas para permitir tantas conexões de dispositivos tão distintos. Por isso, a rede 5G terá um papel importante para IoT, pois, além de proporcionar uma velocidade bastante elevada para transmissão de dados, permitirá, por exemplo, que cada dispositivo baseado em IoT utilize apenas os recursos necessários, sempre na medida exata. Isso evitará gargalos na rede, assim como desperdício de energia (um grave problema em dispositivos que usam bateria).

Riscos associados à IoT

A IoT possibilita inúmeras oportunidades e conexões, muitas das quais não conseguimos imaginar nem entender seu impacto nos dias de hoje. É fácil perceber por que esse assunto tem sido tão comentado atualmente, ele certamente abre portas para muitas oportunidades, mas também muitos desafios.

Com a grande quantidade de dispositivos conectados surge um cenário que impõe alguns riscos. Parâmetros preventivos e corretivos já estão sendo estudados, e o foco principal são as ameaças relacionadas à segurança e à privacidade. Os riscos não são apenas individuais, problemas de ordem coletiva também podem ocorrer. Um exemplo do cotidiano é a conectividade

dos semáforos, que, quando interconectados, podem causar problemas de congestionamentos e caos nas cidades em caso de falhas.

A indústria tem um papel importante na definição de critérios que garantam a disponibilidade dos serviços, proteção das comunicações, normas de privacidade, confidencialidade e integridade dos dados. Todos esses aspectos tornam a tarefa mais complexa do que o trivial. Além das questões tecnológicas, ainda existem desafios relacionados às convenções globais e legislação dos países.

Outro aspecto que não pode ser esquecido no contexto de IoT é a transparência, ou seja, deixar claro para empresas e usuários os riscos associados à IoT. Além disso, orientações relacionadas às questões de segurança também devem ser realizadas.



Leituras recomendadas

ALECRIM, E. *O que é Internet das Coisas (Internet of Things)?* [S.l.]: Info Wester, 2017. Disponível em: <<https://www.infowester.com/iot.php>>. Acesso em: 04 mar. 2018.

ANSELMO, F.; SILVA, M. P. Big Data: tornando dados não estruturados em informações legíveis. *Revista UNIPLAC*, Lages, v. 4, n. 1, 2016.

HEITLINGER, P. *O guia prático da XML*. Porto: Centro Atlântico, 2001. Disponível em: <<http://www.centroatl.pt/titulos/tecnologias/imagens/oguiapraticoda-xml-excerto.pdf>>. Acesso em: 04 mar. 2018.

LACERDA, F.; LIMA-MARQUES, M. Da necessidade de princípios de Arquitetura da Informação para a Internet das Coisas. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 20, n. 2, p. 158-171, abr./jun. 2015.

MENESES, C. A. A. *Big Data nas empresas*: o uso de dados como diferencial estratégico do setor de programação em uma empresa de televisão. Palhoça: Unisul, 2017.

ROSE, K.; ELDRIDGE, S.; CHAPIN, L. *The internet of things*: an overview. Geneva: ISOC, 2015.

SILVA, L. J. *Internet das coisas*. 2017. 49 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Elétrica) – Curso de Engenharia Elétrica e Telemática, Universidade do Sul de Santa Catarina, Palhoça, 2017.

XML: introdução e conceitos básicos. [S.l.]: Macoratti, [2017?]. Disponível em: <<http://www.macoratti.net/xml.htm>>. Acesso em: 04 mar. 2018.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.

Conteúdo:



SOLUÇÕES
EDUCACIONAIS
INTEGRADAS