

Aula 06

Internet das Coisas

Gerenciamento de Dados em IoT

Prof. Julio Cezar Estrella
jcezar@icmc.usp.br

Roteiro

- Desafios e Características dos Dados
- Metadados
- Gerenciamento dos Dados
 - Coleta
 - Modelagem
 - Pré-Processamento
 - Armazenamento
 - Extração do conhecimento

Desafios e Características dos dados

- Para extrair todo o potencial contido nos dados da IoT, é necessário que, primeiro, algumas medidas sejam tomadas:
 - Descrição do problema
 - Modelagem dos dados
 - Pré-processamento
 - Armazenamento dos dados

Desafios e Características dos dados

- A infraestrutura de comunicação e de processamento é um aspecto que influencia na qualidade dos dados coletados, processados e armazenados
- Há uma diferença entre os dados coletados pelos sensores e os metadados gerados e precisam ser também armazenados em projeto de IoT

Desafios e Características dos dados

- Uma adequada infraestrutura para o projeto pode:
 - Fornecer suporte a um ecossistema de software que é capaz de processar e gerar valor sobre os dados ali guardados e processados.
 - Fornecer base para algoritmos de IA, que são capazes de trazer insights fora do olhar comum (trazem padrões, correlações).

Desafios e Características dos dados

- Aumento da capacidade de processamento e avanço de técnicas de IA
 - Dedicar recursos para processar dados que hoje ainda não são coletados e que podem ser fundamentais para calcular métricas que impactam negócios e a vida das pessoas e organizações.

Desafios e Características dos dados

- Dados são o combustível para que novos resultados sejam atingidos
- Paradoxo
 - Hoje temos muita capacidade de processamento para os dados gerados pelos sensores
 - Temos também diversas formas de processar esses dados
 - Mas há pouca preocupação com a geração, coleta, propagação, armazenamento

Desafios e Características dos dados

- Dados gerados apresentam algumas características.
 - A forma como guardamos esses dados, precisam manter essas características.
 - Geração contínua dos dados
 - Grande quantidade dos dados
 - Formatos diferentes
 - Ausência de sincronização dos dados
 - Os dados coletados possuem pouca informação

Desafios e Características dos dados

- Geração contínua dos dados
 - Sensores que reportam ruído sonoro em uma edificação
 - Sensores que coletam temperatura, umidade do ar, gases, etc
- Não há prazo fixo para início e término de geração desses dados

Desafios e Características dos dados

- Grande quantidade dos dados
 - Trata-se de uma consequência da primeira característica
 - A quantidade de dados tende a crescer rapidamente
 - Isso impacta a forma como esses dados serão armazenados futuramente

Desafios e Características dos dados

- Formatos diferentes
 - Exemplo de sensoriamento em uma edificação
 - Dados de temperatura (número)
 - Câmera de monitoramento (áudio, vídeo)
 - Dados da localização do prédio (coordenada geográfica)

Desafios e Características dos dados

- **Ausência de sincronização dos dados**
 - Dispositivos utilizam diferentes tecnologias (Sigfox, Wifi, Bluetooth, LoRA, 3G, 4G, 5G).
 - Conexão não fica aberta o tempo todo
 - Formato e capacidade de envio (velocidade) são diferentes.
 - Por conta dessa heterogeneidade dentro do mesmo projeto, não podemos esperar que chegam no mesmo tempo e reportem no mesmo instante que chegaram
 - A falta de sincronizacao é afetada pela velocidade dos dispositivos (dois dispositivos com capacidade diferente de processamento)

Desafios e Características dos dados

- **Dados coletados com poucas informações**
 - Um dado coletado por um sensor não representa a situação geral de um ambiente
 - Exemplo: A temperatura de uma localização de uma sala não representa a temperatura global da sala
 - A precisão é afetada:
 - Pelo baixo custo de produção, que reduz a qualidade de sensores e dispositivos.
 - Os dados precisam ser trabalhados/tratados em conjunto para atestar se é ou não válido (se medidas foram feitas corretamente pelo sensor conectado ao dispositivo).

Metadados

- Os metadados nos trazem uma imagem completa da situação na qual os sensores estavam inseridos no momento em que coletaram os dados.
- Um sensor (marca, modelo, software) pode ser atualizado um dia/mês/ano depois
- Exemplos:
 - Identificação do sensor, classe ou tipo do dispositivo, modelo do dispositivo, data de fabricação, número de série do hardware

Metadados

- Importante
 - Eles são imutáveis ou raramente muda. O ID de um dispositivo não deveria mudar, a classe ou tipo também raramente muda.
 - Quando alterado, precisamos:
 - Guardar o histórico de alterações para recriar a situação na qual o dado original foi gerado/coletado.

Gerenciamento dos Dados

- Para usuários e aplicações a razão de ser da IoT é extrair o conhecimento a partir de dados coletados pelos sensores
- Extração de conhecimento
 - Modelagem e análise de dados, com semântica para posteriormente tomar decisões adequadas para prover informações ou novos serviços

Gerenciamento dos Dados

- Etapas ideais da extração do conhecimento
 - Coleta dos dados
 - Adequação ao modelo
 - Pré-processamento
 - Armazenamento
 - Extração do conhecimento

Gerenciamento dos Dados

- Coleta dos dados
 - Quantidade de dispositivos e de dados
 - Frequência da coleta
 - Metadados dos dispositivos
 - Comunicação

Gerenciamento dos Dados

- Modelagem
 - Considerando os dados e metadados, como armazená-los?
 - A modelagem é imprescindível
- A criação de um repositório de dados precisa respeitar as características de dados e metadados de IoT. Como as que são destacadas a seguir

Gerenciamento dos Dados

- Modelagem
 - Ser escalável
 - Processar muitos dados, rapidamente com baixo custo financeiro e computacional
 - Armazenar dados não estruturados
 - Os repositórios de dados precisam fornecer possibilidades para a fácil extração, que não sobrecarregue a infraestrutura e capaz de suportar diferentes formatos que variam de um dispositivo para outro
 - Armazenar metadados
 - Guarda informações de dispositivos que fizeram a geração dos dados

Gerenciamento dos Dados

- Modelagem

- Dados brutos obtidos pelos sensores dificilmente possuem explicitamente uma organização hierárquica, relacionamentos ou mesmo um formato padrão para manipulação.
- Esta etapa de modelagem refere-se em definir uma representação para manipular e trabalhar com esses dados visando uma interoperabilidade e formatos padrões interpretáveis.

Gerenciamento dos Dados

- Modelagem

- A representação conceitual da informação varia de contexto para contexto e pode ser feita por exemplo com base:
 - Em chave-valor (arq. texto)
 - Esquema em linguagem de marcação (XML, JSON)
 - Em objetos
 - Em lógica
 - Em ontologia (RDF, OWL)

Gerenciamento dos Dados

- Pré-processamento

- Para que os dados possam ser processados adequadamente, eles devem possuir qualidade suficiente para serem utilizados
- Evita imperfeições
- Utilizar o pré-processamento antes da utilização para a extração do conhecimento

Gerenciamento dos Dados

- Armazenamento

- Essencial para que uma grande quantidade de dados de sensores possam ser armazenados
- Em geral o acesso é feito com base em servidores de armazenamento
 - Muitos disponíveis sob plataformas computacionais específicas para prover serviços em IoT (Vamos tratar disso na próxima aula)

Gerenciamento dos Dados

- Extração do conhecimento
 - A modelagem explora dados brutos de modo a estruturá-los em uma representação específica
- **Contexto**: informação para ser utilizada para caracterizar a situação de uma entidade (objeto, pessoa, lugar → tudo que é relevante para a situação)
- **Situação**: representa a interpretação semântica de contextos (oriunda da combinação de muitas informações → proporciona conhecimento do mundo físico)

Gerenciamento dos Dados

- A **inferência** é uma estratégia para a extração do conhecimento (semântica de uma situação), com base nos dados coletados
- A inferência está atrelada a etapa de modelagem
 - Técnica de modelagem/Técnica de inferência
- Algumas técnicas de inferências:
 - **Aprendizado supervisionado**
 - **Aprendizado não-supervisionado**
 - **Aprendizado baseado em regras**
 - **Ontologias**
 - **Lógica Probabilística**

Conclusão

- Dados não dizem nada e não implicam em valor
- É preciso projetar aplicações e serviços que gerem valor a partir dos dados obtidos
- Considerar características dos dados, metadados, processo de coleta, pré-processamento, armazenamento e extração do conhecimento
- Considerar plataformas computacionais que darão apoio a toda a cadeia de valores gerados pela IoT.

Referências

- <https://homepages.dcc.ufmg.br/~mmvieira/cc/papers/internet-das-coisas.pdf>
- Dobson, S. and Ye, J. (2006). Using fibrations for situation identification. In Pervasive 2006 workshop proceedings – pag. 645–651
- <https://www.aliger.com.br/blog/iot-e-ciencia-de-dados>

Atividade

- Disponível no Moodle conforme consta no cronograma da disciplina

Próxima Aula

- Plataformas para gerenciamento de dados em IoT