

IMPLEMENTANDO UMA PLATAFORMA BIG DATA PARA VISUALIZAÇÃO DE DADOS GERADOS POR DISPOSITIVO IOT

IMPLEMENTING A BIG DATA PLATFORM FOR VIEWING DATA GENERATED BY IOT DEVICE

Jair Campos de Oliveira JUNIOR

jair.ntj@hotmail.com

Pós-Graduação Lato Sensu em BI com Big Data,

Centro Universitário Padre Anchieta, Jundiaí/SP

Juliano SCHIMIGUEL (Orientador)

schimiguel@gmail.com

Coordenação Pós-Graduação Lato Sensu em BI com Big Data,

Centro Universitário Padre Anchieta, Jundiaí/SP

Resumo

O volume de dados vem crescendo a cada dia e são gerados de diferentes fontes, como redes sociais, mídia, Data Warehouse, sensores e de dispositivos *IoT*, tanto da indústria como de dispositivos portáteis. Diferentes plataformas de serviços *Cloud* vêm desenvolvendo métodos e ferramentas para extrair e transformar estes dados em algo de valor. A internet das coisas está em todos os setores e mercados, estando presente no setor automobilístico, na saúde, nas casas, em fabricas, em produtos, na segurança pública e privada, no meio ambiente, na agricultura e nos alimentos, e em outros seguimentos. Nesse artigo será desenvolvido um dispositivo *IoT* para coleta de dados gerados por sensores e a implementação de uma plataforma para extração e visualização dos dados.

Palavras-Chave

Dados; Internet das coisas; sensores; plataformas; visualização.

Abstract

The data volume is growing day by day and is generated from different sources such as social networks, media, data warehousing, sensors and *IoT* devices from both industry and handheld devices. Different cloud service platforms have been developing methods

and tools for extracting and transforming this data into something of value. The internet of things is in every sector and market, being present in the automotive sector, health, homes, factories, products, public and private safety, the environment, agriculture and food, and other segments. In this article, an IoT device will be developed for sensor-generated data collection and the implementation of a platform for data extraction and visualization.

Keywords

Data; Internet of things; sensors; platforms; visualization.

INTRODUÇÃO

Já se foi a época onde as empresas aguardavam horas ou dias para ver painéis com base em dados antigos e obsoletos. Estudos revelam que cerca 2,2 milhões de terabytes em novos dados estão sendo criados todos os dias no mundo e estima-se que no ano de 2020 serão gerados aproximadamente 1,7 megabytes por segundo de novas informações para cada habitante no mundo (ADNEWS, 2019).

Um outro fator que vem contribuindo para o aumento exponencial do tráfego de dados no mundo são os dispositivos *IoT*, segundo a Cisco (2019) o tráfego IP global anual atingirá 4,8 ZB em 2022. Para se ter uma ideia em 2017 o tráfego IP global foi de 1,5 ZB por ano ou 122 EB por mês.

O crescimento exponencial da quantidade de dados e o aumento de recursos cada vez mais avançados para capturar estes dados vem fazendo com que empresas enfrentem o desafio de dar algum sentido a esta enorme onda de dados brutos. O Hadoop veio como a principal estrutura para lidar com todos estes dados. Até pouco tempo não se tinha recursos ou estruturas para criar aplicativos para o processamento em fluxo. Estes aplicativos permitiram que as empresas descem respostas rápidas a eventos, exemplo disso são os monitoramentos de mídias sociais para saber a aceitação de um determinado produto, análise de sentimentos diante de uma determinada notícia, entre outros. (JAIN, 2017).

Segundo Shields (2018), o espaço público em nuvens é dominado por 5 grandes empresas e de acordo com as previsões da IDC, empresa de inteligência de mercado, espera-se que os gastos globais em nuvem pública aumentem 23,2% chegando a US \$ 160 bilhões em 2018 e nos próximos cinco anos acredita-se que o espaço público em nuvem terá uma taxa de crescimento de 21,9% e os gastos com serviços públicos em nuvem alcancem US \$ 277 bilhões em 2021.

Com o surgimento de várias ferramentas para *Big Data* é fundamental conhecer a aplicabilidade e suas funções, assim como empresas que disponibilizam algumas destas ferramentas, plataformas como AWS (Amazon Web Services) e Microsoft Azure, líderes de mercado, oferecem serviços em nuvem, com soluções e produtos onde é possível construir aplicações com escalabilidade e flexibilidade.

Esse artigo está estruturado por seções que abordam alguns conceitos de Big Data, ferramentas para visualização dos dados, plataformas cloud, internet das coisas, o desenvolvimento de um dispositivo IoT para coleta dos dados e conclusão.

BIG DATA

A etimologia do “*Big Data*” foi usada pela primeira vez por Mashey (1998), ex-cientista chefe aposentado da SGI, onde produziu um deck de slides da SGI intitulado “*Big Data and the Next Wave of InfraStress*”, nesse deck ele demonstrava as várias tendências tecnológicas que seriam esperadas para os próximos anos.

Essa nova era já vem surgindo a algum tempo e tem raízes profundas, o *Big Data* existe há décadas para alguns veteranos da indústria que ao longo dos anos já lidavam com uma grande quantidade de dados transacionais. Com o crescimento de novas tecnologias como a computação móvel, as redes sociais e a computação em nuvem houve um aumento no volume, na velocidade e a variedade de dados. (DHIRAJ; CHAMBERS; MINELLI, 2013).

Du (2018) resume o *Big Data* não se tratando apenas de muitos dados e sim uma prática de descobrir novas informações a partir dos dados existentes e orientar a análise de novos dados. Uma empresa orientada a *Big Data* será mais ágil e competitiva para superar desafios e vencer competições.

Hoje, a tecnologia nos permite coletar dados a uma velocidade impressionante, tanto em termos de volume quanto de variedade. Existem várias fontes que geram dados, mas no contexto de *Big Data*, as fontes principais são as redes sociais, mídia, data warehouse e sensores.

As redes sociais são indiscutivelmente, a principal fonte de todo o *Big Data* que conhecemos hoje e se que proliferaram nos últimos anos. Trata-se, em geral, de dados não estruturados, representados por milhões de publicações nas mídias sociais e outros dados gerados segundo a segundo por meio de interações do usuário na Web em todo o mundo. O aumento do acesso à Internet em todo o mundo tem sido um ato autorrealizável para o crescimento de dados nas redes sociais. (DASGUPTA, 2018)

A mídia é em grande parte o resultado do crescimento das redes sociais e representa os milhões, se não bilhões, de uploads de áudio e vídeos que ocorrem diariamente. Vídeos enviados no YouTube, gravações de música no SoundCloud e imagens postadas no Instagram são exemplos excelentes de mídia, cujo volume continua a crescer de maneira irrestrita. (DASGUPTA, 2018)

As empresas investem há muito tempo em instalações especializadas de armazenamento de dados, comumente conhecidas como *data warehouses*. Um DW é essencialmente uma coleção de dados históricos que as empresas desejam manter e catalogar para fácil recuperação, seja para uso interno ou para fins regulatórios. À medida que os setores passam gradualmente para a prática de armazenar dados em plataformas como Hadoop e NoSQL, mais e mais empresas estão transferindo dados de seus *data warehouses* pré-existentes para algumas das tecnologias mais recentes. E-mails da empresa, registros contábeis, bancos de dados e documentos internos são alguns exemplos de dados DW que agora estão sendo transferidos para plataformas Hadoop ou do tipo Hadoop que utilizam vários *nodes* para fornecer uma plataforma altamente disponível e tolerante a falhas. (DASGUPTA, 2018)

Um fenômeno mais recente no espaço do *Big Data* foi a coleta de dados de dispositivos *IoT* e sensores. Embora sempre existam sensores e indústrias como petróleo e gás usem

sensores de perfuração para medições em plataformas de petróleo há muitas décadas, o advento de dispositivos portáteis, também conhecidos como Internet das Coisas, como Fitbit e Apple Watch, significa que agora cada indivíduo poderia transmitir dados na mesma taxa em que algumas plataformas de petróleo costumavam fazer há alguns anos atrás. (DASGUPTA, 2018)

Tendo como principais características geralmente *Big Data* pode ser explicado de acordo com os três Vs, sendo volume, variedade e velocidade.

- Volume: os dados podem vir de diversas fontes diferentes, como registros de vendas de negócios, os resultados coletados de experimentos científicos ou sensores em tempo real usados na internet das coisas (*IoT*). Os dados podem estar brutos ou pré-processados usando ferramentas de software separadas antes de serem aplicadas as análises. (ROUSE, 2017).
- Variedade: os dados também podem existir em uma grande variedade de tipos de arquivos, sendo dados estruturados, quase estruturados, semiestruturados ou não estruturados. (DHIRAJ; CHAMBERS; MINELLI, 2013).
- Velocidade: refere-se à velocidade na qual os fluxos de dados devem ser analisados. Todo o grande projeto de análise de dados irá ingerir, correlacionar e analisar as fontes de dados e, em seguida, renderizar uma resposta ou resultado com base em uma consulta abrangente. A velocidade também é significativa, uma vez que a análise de dados se expande em campos como aprendizagem de máquinas e inteligência artificial, onde os processos analíticos imitam a percepção ao encontrar e usar padrões nos dados coletados. (ROUSE, 2017).

No entanto existem outras características adicionais como:

- Veracidade: é de grande importância que os dados coletados sejam condizentes com a realidade daquele momento, quando se trata de análise real-time, dados passados não podem ser considerados verídicos no momento que estão sendo analisados. A precisão dentro de um contexto irá impactar os resultados. (HURWITZ et al., 2016). Segundo Rijmenam (2019), a geração dados falhos são inúteis quando coletados, os dados podem causar problemas significativos para as organizações quando incorretos. Garantir que os dados e análises estejam corretos é de extrema importância, principalmente quando a tomada de decisão é de forma automatizada onde nenhum ser humano está mais envolvido. Sistemas de irrigação, por exemplo, falhas nos sensores de humidade de solo podem gerar dados falhos fazendo com que o sistema irrigue de forma incorreta o solo comprometendo assim o plantio e gerando prejuízos aos agricultores.
- Valor: no início do processo de análise saber as perguntas certas é importantíssimo, resultados só terão valor se tiverem sido tratados, analisados e usados para chegar a uma decisão. (TAURION, 2013). Os dados em si não são valiosos. O valor está quando os dados analisados são transformados em informações, nos permitindo adquirir conhecimento e sabedoria. (RIJMENAM, 2014).
- Variabilidade: *Big Data* é extremamente variável, isso significa que uma mesma palavra tem diferentes significados dependendo do contexto e,

portanto, é muito relevante na realização de análises de sentimento. Variabilidade significa que a definição está mudando e rapidamente. Um exemplo são tweets semelhantes, uma palavra pode ter um significado totalmente diferente. Para realizar uma análise de sentimento adequada, os algoritmos precisam decifrar o significado exato de uma palavra em seu contexto. Isso ainda é muito difícil. (RIJMENAM, 2014)

- **Visualização:** Essa é a parte difícil do *Big Data*. Significa tornar essa vasta quantidade de dados compreensível de uma maneira fácil de ler e entender. Com as visualizações corretas, os dados brutos podem ser colocados em uso. As visualizações, claro, não são gráficos comuns ou gráficos de pizza. São gráficos complexos que podem incluir muitas variáveis de dados, permanecendo legíveis e compreensíveis. A visualização pode não ser a tarefa tecnologicamente mais difícil, mas certamente é a mais desafiadora. O uso de um gráfico para contar uma história complexa é muito difícil, mas também extremamente crucial. Felizmente, há mais e mais startups de *Big Data* que se concentram nesse aspecto do desafio. No final, as visualizações farão a diferença. (RIJMENAM, 2014)
- **Volatilidade:** refere-se a quanto tempo os dados são válidos e armazenados. Isso é particularmente importante para análises em tempo real. Ele exige que uma janela de tempo alvo seja determinada para que os analistas possam se concentrar em questões específicas e obter um bom desempenho com a análise. (DU, 2018)

FERRAMENTAS DE VISUALIZAÇÃO

O Tableau é uma ferramenta de BI que possibilita a análise dos dados visualmente. Permite que os usuários criem e distribuam painéis interativos e compartilháveis, descrevendo na forma de gráficos e tabelas as tendências, variações e densidade dos dados. Usado por empresas de diversos setores, pesquisadores acadêmicos e organizações governamentais para análise visual de dados, o Tableau pode conectar-se a arquivos, fontes relacionais e de *Big Data*. Possuindo diversos recursos e um aplicativo de descoberta e exploração de dados permite que o tableau traga resultados em segundos. Alguns recursos exclusivos lhe permitem lidar com diversos cenários, fornecendo soluções para todos os tipos de setores, departamentos e ambientes de dados. Como o Tableau não requer um alto nível de conhecimento em programação, qualquer pessoa que tenha acesso aos dados pode começar a usá-lo para obter valor dos dados, o que é uma vantagem quando se tratando em velocidade de análise. Um outro ponto positivo é de não precisar de uma configuração complexa de software, sua versão para desktop usada pela maioria dos usuários é facilmente instalada e contém todos os recursos necessários para iniciar e concluir análise de dados, tornando-o autossuficiente. (TUTORIALSPPOINT, 2016).

O Power BI é uma ferramenta de visualização de dados e BI que converte dados de diferentes fontes em Dashboards e relatórios de BI. O Power BI suíte fornece vários softwares, conectores e serviços, o Power BI Desktop, o Power BI servisse baseado em SaaS e o Power BI mobile estão disponíveis para diversas plataformas. Esse conjunto de serviços podem ser usados pelos usuários para consumir dados e criar relatórios. O Power BI inclui componentes como Power BI Desktop, usado para criar relatórios e visualizações de dados, enquanto os Serviços de BI (Software como Serviço - SaaS) são usados para publicar os relatórios, o Gateway do Power BI pode ser usado para manter

seus dados atualizados e conectados às fontes de dados locais sem que haja a necessidade de mover os dados, o aplicativo mobile é possível conectar e visualizar os dados de qualquer lugar. (TUTORIALSPPOINT, 2017)

O Power BI suporta uma variedade de fontes de dados muito grande, é possível se conectar a diferentes arquivos simples, banco de dados SQL e nuvem do Azure ou plataformas da web, como Facebook, Salesforce e Google Analytics. É possível também incluir conexão ODBC para conectar a outras fontes de dados. (TUTORIALSPPOINT, 2017)

O WEKA é um software *open source* que fornece ferramentas para pré-processamento de dados, implementação de vários algoritmos de Machine Learning e visualização para que você possa desenvolver técnicas de aprendizado de máquina e aplicá-las a problemas reais de mineração de dados. O WEKA oferece muitos estágios ao lidar com o Big Data, começando com os dados brutos coletados em campo que podem conter valores nulos e campos irrelevantes. Com a ferramenta de pré-processamento é possível limpar estes dados e salvá-los localmente para aplicar os algoritmos. (TUTORIALSPPOINT, 2019)

Dependendo do tipo de modelo que será desenvolvido o WEKA oferece opções de classificação, cluster e associação, cada categoria do WEKA fornece a implementação de vários algoritmos onde você seleciona um algoritmo, define os parâmetros e o executa o conjunto de dados. Vários modelos podem ser aplicados no mesmo conjunto de dados, podendo comparar as saídas de diferentes modelos e selecionar o melhor que atenda aos seus objetivos. (TUTORIALSPPOINT, 2019)

COMPUTAÇÃO EM NUVEM

A computação em nuvem já faz parte da maioria das empresas, independente do seu porte, pois incluem as tecnologias como soluções estratégicas em seus negócios. Vários recursos estão disponibilizados na internet, por exemplo, armazenamento em nuvem. Usuários podem armazenar diversos arquivos como em seus discos locais, porém este armazenamento é fornecido como um serviço de algum provedor de nuvem. Um provedor de nuvem é uma empresa ou grupo de empresas que fornecem serviços em nuvem para pessoas físicas ou jurídicas. Estas empresas hospedam e gerenciam os serviços contratados pelos usuários e são responsáveis por manter este serviço. (MODI, 2019)

Modi (2019) e Wali (2018), destacam algumas vantagens na computação em nuvem, como pagamento conforme o uso, acesso ou presença global o que permite ser acessado de qualquer lugar via internet, recursos ilimitados ou escalabilidade, serviços gerenciados e hardware. Wali (2018) ainda cita outras vantagens sendo economia de custos, economia de espaço, não tem perda de tempo com inatividades causadas por falhas de hardware serão reduzidas.

Existem três principais áreas da computação em nuvem, sendo software como Serviço (SaaS), plataforma como Serviço (PaaS) e infraestrutura como Serviço (IaaS). (SHAH; SARKAR, 2018)

A figura 1 ilustra os níveis de controle que o usuário tem sobre sistemas operacionais, armazenamento, aplicativos, entre outros serviços. Se observamos da esquerda para direita o *User Application* é totalmente controlado pelo usuário e este controle vai

diminuindo gradativamente até que na área SaaS o usuário não tem nenhum tipo de gerenciamento.

O mercado de infraestrutura como serviço (IaaS) vem amadurecendo juntamente com a receita que cresce sem parar. O Gartner, Inc (2019) fez uma projeção em que a receita no mercado de infraestrutura na nuvem deverá aumentar para US \$ 81,5 bilhões até 2022. Para suportar uma ampla variedade de cargas de trabalho o mercado enxerga a AWS e a Microsoft como os dois provedores mais estruturados, o Google vem fazendo progresso constantemente em termos de adoção corporativa, mas está distante em termos de receita anual e de interesse entre os clientes corporativos do Gartner, permanecendo em terceiro lugar.

Figura 1. Nível de Gerenciamento do Usuário

User Application	IaaS Model	PaaS Model	SaaS Model
Application	Application	Application	Application
Data	Data	Data	Data
Runtime (Libraries)	Runtime (Libraries)	Runtime (Libraries)	Runtime (Libraries)
Operating System	Operating System	Operating System	Operating System
Virtualization	Virtualization	Virtualization	Virtualization
Server	Server	Server	Server
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

(SHAH; SARKAR, 2018)

O Amazon Web Services, popularmente conhecido como AWS surgiu em 2006 oferecendo como seu primeiro serviço o *Simple Storage Service* (Serviço de armazenamento simples), abreviado para S3. A ideia era oferecer um conceito de armazenamento de objetos pela web onde qualquer pessoa poderia colocar o objeto, essencialmente qualquer grupo de bytes, desde fotos digitalizadas, backups de arquivos, vídeos, áudios, ou seja, todo e qualquer arquivo que pudesse ser armazenado em um HD local. No começo o S3 era limitado e embora os objetos fossem lidos ou escritos de qualquer lugar, eles eram armazenados apenas na região dos Estados Unidos e não podiam ter mais que 5 gigabytes. (GOLDEN, 2013)

Uma grande parte dos serviços que a AWS oferece se encaixa na definição de infraestrutura como serviço (IaaS), este modelo mais maduro é certamente o mais ofertado, servidores virtualizados e matrizes de armazenamento virtualizado são hospedados em uma rede definida por softwares e com uma infraestrutura completamente isolada para cada cliente como um recurso privado. A criação de recursos na AWS na maioria das vezes começa com a criação do que é chamado de virtual private cloud (VPC). Por sua rede privada ser isolada pode-se hospedar servidores virtuais, HDs virtuais e serviços e produtos completamente gerenciados. A flexibilidade da AWS permite criar um monte de arquiteturas usando vários serviços e utilitários contidos na caixa de ferramentas IaaS. (WILKINS, 2019)

Segundo Wilkins (2019), os serviços de nuvem IaaS são fornecidos com serviços gerenciados que se baseia em serviços de computação, softwares personalizados, armazenamento e rede, fornecendo opções para que a Amazon gerencie e mantenha, assim não precisará ter todo o trabalho.

Em 2008 a Microsoft anuncia o Azure com o nome de *Red Dog* e lançado oficialmente em 2010 com o nome de *Windows Azure* oferecendo funções da Web e banco de dados SQL. Em 2014 a Microsoft alterou o nome de sua plataforma em nuvem para Microsoft Azure. (WALI, 2018)

O Azure oferece suporte a sistemas operacionais como Linux e Windows, linguagens como C#, Python, Java, Node.js e outras, várias ferramentas, plataformas, estruturas e utilitários. A ideia é que o usuário tenha total liberdade em suas escolhas, seja na linguagem, plataforma, sistema operacional, armazenamento, ferramentas e utilitários. O Azure oferece uma série de tecnologias a nível mundial. O Azure possui data centers espalhados pelo mundo todo e estão combinados por regiões, cada região possui vários data centers para garantir uma recuperação mais rápida e eficiente em caso de desastre. (MODI, 2019)

O Azure fornece infraestrutura e serviços para ingerir uma imensa quantidade de transações com processamento, também fornece uma enorme capacidade de armazenamento de dados e uma série de serviços interconectados que podem transmitir dados entre si. Todos esses recursos fazem com que os dados possam ser processados gerando conhecimento e insights significativos. (MODI, 2019)

Em 2008 uma equipe de desenvolvedores do Google uma oferta de plataforma como serviço, o Google App Engine. Nos anos seguintes, o Google lançou vários serviços e recursos como o Google Cloud Storage em 2010, Compute Engine em 2013, Cloud SQL em 2014 e Kubernetes Engine em 2015. Durante esse período, o Google buscou expandir seu domínio em diversas áreas, como análise de dados, gerenciamento de infraestrutura, *IoT* e aprendizado de máquina. Em 2017, o Google havia estabelecido data centers em 39 zonas em 13 regiões. (HUNTER; PORTER; PS, 2019)

Google Cloud Platform possui produtos e serviços que atendem a várias demandas e serviços. Os principais serviços, como o *Compute Engine* e o *Cloud Storage*, permitem praticamente qualquer solução, enquanto muitos serviços especializados, como a *Cloud Vision API*, reduzem possibilitam as equipes enfrentarem espaços de problemas mais específicos. À medida que o Google avança no espaço da nuvem pública, a quantidade de produtos e serviços essenciais e especializados continuam a crescer rapidamente. (HUNTER; PORTER; PS, 2019)

Além das empresas já citadas, existem muitos outros fornecedores de plataforma em nuvem que oferecem serviços voltados para dispositivos *IoT*, quando fomos escolher devemos levar em conta o tipo de projeto, sua aplicabilidade e a complexidade no processo de implantação.

IBM BlueMix – O Bluemix é uma implementação focada em *IoT* da Open Cloud Architecture da IBM, com o Bluemix é possível criar, implementar e gerenciar aplicativos em nuvem voltados para internet das coisas. (SHOVIE, 2016).

IBM Watson – A plataforma IBM Watson é um serviço cloud que pode conectar

dispositivos *IoT* residenciais ou industriais e controlar sensores fornecendo um amplo conjunto de ferramentas integradas e complementares. Com elas é possível processar dados de dispositivos realizando análises em tempo real ou em lote. (IBM CORPORATION, 2019)

ThingSpeak – O ThingSpeak é uma plataforma como serviço (PaaS) de análise de *IoT* que permite agregar, visualizar e analisar fluxos de dados em tempo real. É possível enviar dados dos dispositivos e criar visualizações, bem como enviar alertas. (THINGSPEAK, 2019)

Thingworx – O Thingworx é uma interface do usuário baseada em nuvem para coleta e análise de dados sendo usado, por exemplo, para controle de fábrica. A proprietária da Thingworx em 2015 adquiriu o KepWare, que essencialmente é uma coleção de protocolos e middleware conectados às instalações da produção. Através da combinação dos sistemas os clientes têm acesso aos dados mais recentes da fábrica, assim como tendências e alarmes "outliers". (ELK, 2018)

Dweet – O Dweet.io é basicamente uma API simples que pode ser chamada de qualquer navegador ou aplicativo da Web e pode ser usado sem problemas por qualquer pessoa, graças a sua simplicidade. É realmente fácil armazenar dados de um projeto do Arduino e recuperar esses dados usando outros aplicativos. (SCHWARTZ, 2016)

IoT – INTERNET OF THINGS

O decreto número 9.854, de 25 junho de 2019 art. 2º considera-se:

I – Internet das Coisas – IoT – a infraestrutura que integra a prestação de serviços de valor adicionado com capacidades de conexão física ou virtual de coisas com dispositivos baseados em tecnologias da informação e comunicação existentes e nas suas evoluções, com interoperabilidade; II – Coisas – objetos no mundo físico ou no mundo digital, capazes de serem identificados e integrados pelas redes de comunicação; III – dispositivos – equipamentos ou subconjuntos de equipamentos com capacidade mandatória de comunicação e capacidade opcional de sensoriamento, de atuação, de coleta, de armazenamento e de processamento de dados; e IV – serviço de valor adicionado – atividade que acrescenta a um serviço de telecomunicações que lhe dá suporte e com o qual não se confunde novas utilidades relacionadas ao acesso, ao armazenamento, à apresentação, à movimentação ou à recuperação de informações, nos termos do disposto no art. 61 da Lei nº 9.472, de 16 de julho de 1997. (DIÁRIO OFICIAL DA UNIÃO, 2019, p. 10).

As vantagens da *IoT* abrangem todas as áreas desde estilo de vida à negócios. Aqui está uma lista de algumas das vantagens que a *IoT* tem a oferecer:

- Melhoria do envolvimento do cliente - A análise atual sofre de pontos cegos e falhas significativas na precisão; e como observado, o engajamento permanece passivo. A *IoT* transforma completamente isso para conseguir um engajamento mais rico e eficaz com o público. (TUTORIALSPPOINT,

2016)

- Otimização de Tecnologia - As mesmas tecnologias e dados que melhoram a experiência do cliente também melhoram o uso do dispositivo e auxiliam em melhorias mais potentes para a tecnologia. A *IoT* desbloqueia um mundo de dados funcionais e de campo críticos. (TUTORIALSPPOINT, 2016)
- Resíduos Reduzidos - *IoT* tornam claras as áreas de melhoria. A análise atual nos fornece uma percepção superficial, mas a *IoT* fornece informações reais que levam a um gerenciamento mais eficaz dos recursos. (TUTORIALSPPOINT, 2016)
- Coleta de dados aprimorada - A coleta de dados moderna sofre com suas limitações e seu design para uso passivo. *IoT* separa esses espaços e coloca exatamente onde os humanos realmente querem ir para analisar nosso mundo. Permite uma imagem precisa de tudo. (TUTORIALSPPOINT, 2016)
- Desvantagens, embora a *IoT* ofereça um impressionante conjunto de benefícios, ela também apresenta um conjunto significativo de desafios.
- Segurança - *IoT* cria um ecossistema de dispositivos conectados constantemente, que se comunicam através de redes. O sistema oferece pouco controle apesar de quaisquer medidas de segurança. Isso deixa os usuários expostos a vários tipos de invasores. (TUTORIALSPPOINT, 2016)
- Privacidade - A sofisticação da *IoT* fornece dados pessoais substanciais em detalhes extremos, sem a participação ativa do usuário. (TUTORIALSPPOINT, 2016)
- Complexidade - Alguns acham os sistemas *IoT* complicados em termos de design, implantação e manutenção, devido ao uso de várias tecnologias e um grande conjunto de novas tecnologias de ativação. (TUTORIALSPPOINT, 2016)
- Flexibilidade - Muitos estão preocupados com a flexibilidade de um sistema *IoT* para se integrar facilmente com outro. Eles se preocupam em encontrar-se com vários sistemas conflitantes ou bloqueados. (TUTORIALSPPOINT, 2016)
- Conformidade - A *IoT*, como qualquer outra tecnologia no campo dos negócios, deve obedecer aos regulamentos. Sua complexidade faz com que a questão da conformidade pareça incrivelmente desafiadora quando muitos consideram a conformidade com o software padrão uma batalha. (TUTORIALSPPOINT, 2016)

Existem alguns hardwares de baixo custo disponíveis no mercado para quem queira realizar algum projeto de *IoT*. Dispositivos como Arduino ou Raspberry PI 3 possibilitam a realização de vários projetos. No site dos fabricantes existem vários projetos disponíveis para experimento.

A *IoT* tem aplicativos em todos os setores e mercados. Ele abrange grupos de usuários daqueles que desejam reduzir o uso de energia em suas casas para grandes organizações que desejam otimizar suas operações. Isso não é apenas útil, mas quase crítico em muitos setores, à medida que a tecnologia avança e nos movemos em direção à automação avançada imaginada em um futuro distante. (TUTORIALSPPOINTS, 2016)

DISPOSITIVO IOT

Um dispositivo IoT que está sendo muito utilizado em estudos é um sistema de horta, uma horta para muitas pessoas é algo muito agradável, imagine poder estar em sua casa ou apartamento e enquanto cozinha colher hortaliças fresquinhas, temperos como manjeriço, alecrim, salsinha, cebolinha, entre outros, tudo ali a sua mão fresquinho, porém muitas vezes não temos tempo para regar ou não sabemos quando regar. Um sistema que possa monitorar uma pequena horta e ao mesmo tempo coletar dados para futuras análises seria algo bem interessante. Foi pensando nisso que estamos desenvolvendo o “*Less Water*”, um sistema de irrigação inteligente e sustentável, pois utiliza painéis solares como fonte energia renovável. Neste estudo de caso daremos início ao projeto e iremos abordar alguns aspectos como funcionamento, plataforma, sensores utilizados, montagem e extração dos dados para visualização. O *Less Water* é sistema automatizado e pode ser usado para auxiliar no cultivo de hortaliças, permitindo gerenciar e monitorar a temperatura do ar, umidade do solo e do ar. Geralmente, pequenas hortas são irrigadas de forma manual e sem nenhum parâmetro o que muitas vezes acaba causando a perda das hortaliças.

Os principais componentes usados na montagem do dispositivo foram um Adaptador RoboCore ESP8266 Adapter, uma placa Arduino Uno (Rev3), um sensor de umidade e temperatura DHT22, um módulo WiFi ESP8266 atualizado com ESP-01, um Resistor 10k Ω , um sensor de Umidade de Solo e uma fonte de alimentação para protoboard. A seguir uma breve explicação dos componentes utilizados:

- Sensores – Um dos primeiros parâmetros para a horta é a umidade do solo, Lucietti (2014) destaca a água como o principal componente e que sem ela não ocorre nenhum processo de transformação nos vegetais que vai desde a germinação até todo seu desenvolvimento, sendo assim devemos medir a umidade do solo para que as plantas cresçam bem. No projeto utilizamos um sensor de umidade capacitivo, este sensor pode ser alimentado com tensões que variam de 3,3V à 5,5V e sua tensão de saída vai de 0 à 3V. Uma das vantagens deste sensor está na sua durabilidade, pois por não haver condutores metálicos que ficam em contato direto com o solo não estão sujeitos a corrosão, possui um custo baixo e pode ser facilmente encontrado.

Outro sensor utilizado foi o DHT22, um sensor de umidade e temperatura do ar com tensão de entrada de 3.3V até 6V, com um range de umidade de 0 a 100% e de temperatura de -40 a 80°C e possui uma ótima precisão, este sensor é muito conhecido e possui muitos recursos disponíveis para desenvolvimento.

A temperatura e umidade têm impacto significativo no crescimento das plantas e manter a temperatura e a umidade dentro de certos valores também mantém as hortaliças saudáveis.

O consumo total de água é influenciado pelo tipo de cultura e estágio de desenvolvimento, pelo clima (temperatura, umidade relativa do ar, radiação solar, ventos), pelo tipo de solo, pela cobertura e coloração do solo. O consumo de água pelas plantas pode ser considerado como evapotranspiração = evaporação + transpiração. (LUCIETTI, 2014, on-line)

- Placa MCU e Microcontrolador – para o desenvolvimento do *Less Water* optamos pela placa Arduino UNO R3 por possuir uma plataforma de desenvolvimento simples e de fácil entendimento, uma outra vantagem está na quantidade de material disponível para pesquisa e aplicação.
- Os *protoboard*, também conhecidos como placas de ensaio sem solda nos permiti fazer conexões e alterações sem que haja a necessidade de fazer soldas permanentes em componentes.

Além dos componentes já citados, serão necessários cabos para ligação dos circuitos e sensores, fonte de alimentação, cabo USB para conectar o microcontrolador, alguns adaptadores, caixa para acomodar os circuitos etc.

Antes de iniciarmos a montagem dos circuitos devemos instalar o IDE do Arduino no computador, existe uma outra opção que é o Arduino Web Editor um IDE que faz parte do Arduino Create, mas infelizmente até a data de desenvolvimento do dispositivo este produto não estava disponível no Brasil. Para fazer o download acesse o site arduino.cc. (ARDUINO, 2019)

Após instalado, a tela de desenvolvimento do Arduino é muito simples e é chamada de *sketch* (esboço), na tela de esboço em ferramentas você pode gerenciar as bibliotecas, ver os dados dos sensores pelo monitor serial e configurar no gerenciador de placas o modelo do Arduino ou placas de terceiros. Mais informações podem ser consultadas na aba de ajuda ou pelo site Arduino.cc

Ligação dos sensores

O esquema de ligação do sensor de solo está sendo da seguinte forma:

- O VCC conecta ao 5Volts do Arduino.
- O GND conecta ao GND do Arduino.
- O Sinal conecta ao pino A0 do Arduino.

Como já citado anteriormente para nosso projeto utilizamos o sensor capacitivo, na figura 2 temos o código-fonte usado para testar o funcionamento do sensor de solo.

Figura 2. Código para teste do sensor de solo

```
/*Codigo-fonte para funcionamento do sensor de umidade de solo*/
void setup(){
  // Comunicação serial
  Serial.begin(9600);
}
void loop(){
  int valSolo; // Declaração da variavel
  valSolo = analogRead(A0); // Define a I/O
  // Imprime os valores dos sensores na serial
  Serial.print("Umidade_do_Solo: ");
  Serial.println(valSolo);
  // define tempo entre cada medição
  delay(2000);
}
```

Fonte: (KURNIAWAN, 2017)

O esquema de ligação do sensor DHT22 está sendo da seguinte forma:

- O VCC (Pino 1) conecta ao 5Volts do Arduino.
- O GND (Pino 4) conecta ao GND do Arduino.
- O Sinal (Pino 2) conecta ao D9 do Arduino.

É preciso colocar um resistor de 10K Ω entre o VCC e a linha de dados (Pino2) para manter uma comunicação adequada entre o sensor e o microcontrolador.

Na figura 3 o exemplo do código-fonte para leitura do sensor DHT22, diferente do código anterior, foi necessário incluir a biblioteca de sensores DHT. Essa biblioteca pode ser instalada a partir do software Arduino. Vá para o Sketch | Incluir Biblioteca | gerenciar bibliotecas e você receberá uma caixa de diálogo.

Figura 3. Código-fonte para teste do sensor DHT22

```
#include "DHT.h" //inclui a biblioteca DHT
#define DHTTYPE DHT22 // define o tipo de sensor
#define DHTPin 9 // DHT Sensor
DHT dht(DHTPin, DHTTYPE); // Inicializa DHT sensor
//declaração das variáveis
float temperatura;
float umidade;
void setup() {
  Serial.begin(9600);
  pinMode(DHTPin, INPUT);
  dht.begin();
}
void loop() {
  // Obtém os valores da temperatura e imprime
  temperatura = dht.readTemperature();
  Serial.print("Temp: ");
  Serial.print(temperatura);
  Serial.print("°C");
  Serial.print(" | ");
  // Obtém os valores da umidade e imprime
  umidade = dht.readHumidity();
  Serial.print("Umid: ");
  Serial.print(umidade);
  Serial.println("%");
  //Imprime os dados a cada 1 segundo
  delay(1000);
}
```

Fonte: (KURNIAWAN, 2017)

Conectividade

Para o envio dos dados via Wi-Fi utilizamos o módulo transceptor Wi-Fi ESP8266-01. O

módulo possui oito pinos disponíveis para conexão com dispositivos de microcontroladores externos. Esse chip pode ser usado como um conector Wi-Fi da placa Arduino, mas também pode ser usado como um NodeMCU independente, usando o IDE do Arduino é possível carregar diretamente um *sketch* na placa ESP8266.

O esquema de ligação modulo ESP-01 está sendo da seguinte forma:

- VCC está conectado no 3.3V do Arduino.
- GND está conectado no GND do Arduino
- RX está conectado no D7 do Arduino.
- TX está conectado no D6 do Arduino.
- Ch-Pd está conectado no 3.3V do Arduino.
- Reset está conectado ao D8 do Arduino.

Na figura 4 temos o código-fonte para teste de conexão da placa ESP-01, da mesma forma que foi feito no DHT é necessário instalar a biblioteca SoftwareSerial para o ESP8622.

Figura 4. Código-fonte para teste de conexão do ESP-01

```
#include <SoftwareSerial.h>
SoftwareSerial EspSerial(6, 7); // Rx, Tx
#define HARDWARE_RESET 8
#define DEBUG true
}
void setup() {
  EspSerial.begin(9600);
  Serial.begin(9600);
  Serial.println("ESP8266 Setup test - use AT coomands");
  connectWiFi();
}
void connectWiFi(void) {
  sendData("AT+RST\r\n", 2000, 0); // reset
  sendData("AT+CWJAP=\"Nome da rede\", \"senha\"\r\n", 2000, 0); //Connect network
  delay(3000);
  sendData("AT+CWMODE=1\r\n", 1000, 0);
  sendData("AT+CIFSR\r\n", 1000, 0); // Show IP Adress
  serial.println("Dispositivo Conectado");
}

String sendData(String command, const int timeout, boolean debug) {
  String response = "";
  EspSerial.print(command);
  long int time = millis();
  while ( (time + timeout) > millis() ) {
    while (EspSerial.available()) {
      // The esp has data so display its output to the serial window
      char c = EspSerial.read(); // read the next character.
      response += c;
    }
  }
  if (debug) {
    Serial.print(response);
  }
  return response;
}
void loop() {
  while (EspSerial.available()) {
    Serial.write(EspSerial.read());
  }
  while (Serial.available()) {
    EspSerial.write(Serial.read());
  }
}
```

Fonte: (ROVAI, 2017)

Fonte energia renovável

Em resumo, a tecnologia fotovoltaica transforma a luz solar em eletricidade. Um sistema fotovoltaico, ou sistema elétrico solar, é composto de várias células solares fotovoltaicas. Uma célula fotovoltaica individual é geralmente pequena, produzindo tipicamente cerca de 1 ou 2 W de energia. Para aumentar a potência das células fotovoltaicas, elas são conectadas para formar unidades maiores chamados módulos. Os módulos, por sua vez, podem ser conectados para formar unidades ainda maiores, chamadas de matrizes, que podem ser interconectadas para produzir mais energia, e assim por diante. (PLANTE, 2014)

Para este projeto utilizado dois painéis fotovoltaico, sendo cada painel com 50W de potência, corrente máxima de 2,76 amperes e voltagem máxima de aproximadamente 18,1 volts. A ligação dos dois painéis foi feita em paralelo para que some as correntes e assim obter aproximadamente 5,5 amperes.

Foi instalado um regenerador de tensão para diminuir a tensão que sai dos painéis para as baterias e circuito, na saída do regenerador para o circuito foi instalado um regulador de tensão de aproximadamente 12 volts, evitando assim possíveis oscilações ou picos de tensão. O sistema é alimentado por duas baterias de 12 volts por 7 amperes ligadas em paralelo para que também somem as correntes e de mais autonomia para a alimentação.

Ligando os componentes

Após testados todos os componentes vamos unificar tudo, na tabela 1 temos uma sequência de imagens do processo de instalação dos painéis, baterias e dispositivo, abaixo também é possível conferir o esquema de ligação dos sensores no Arduino.

No Arduino:

- O Sinal do sensor de solo conecta ao pino A0 do Arduino
- O Sinal do sensor DHT22 (Pino 2) conecta ao D9 do Arduino.
- No ESP-01 o RX está conectado no D7 do Arduino.
- No ESP-01 o TX está conectado no D6 do Arduino.
- No ESP-01 o Reset está conectado ao D8 do Arduino.

Na fonte de alimentação:

- O VCC do sensor de solo conecta ao 5Volts da protoboard.
- O GND do sensor de solo conecta ao GND da protoboard.
- O VCC (Pino 1) do DHT22 conecta ao 5Volts da protoboard.
- O GND (Pino 4) do DHT22 conecta ao GND da protoboard.

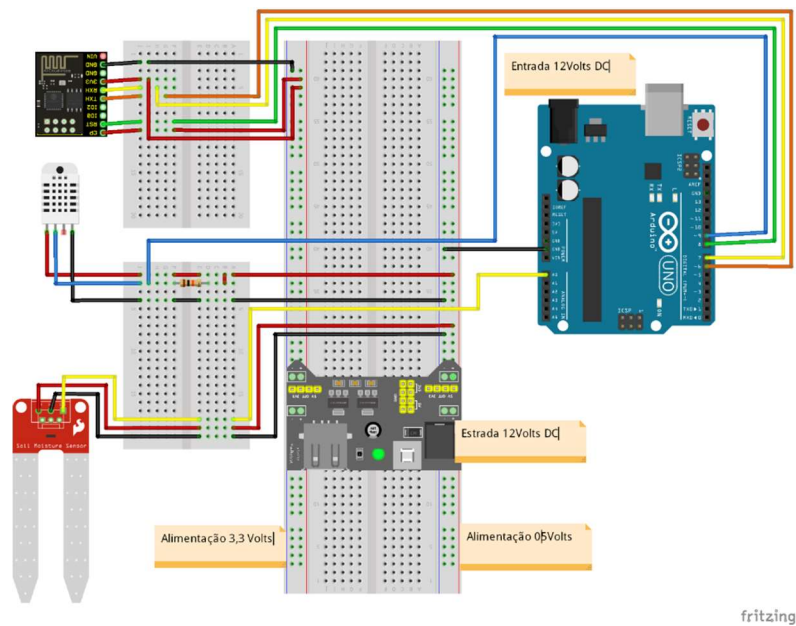
É preciso colocar um resistor de $10K\Omega$ entre o VCC da protoboard e a linha de dados (Pino2) do DHT22 para manter uma comunicação adequada entre o sensor e o microcontrolador.

- VCC do ESP-01 está conectado no 3.3V da protoboard.
- GND do ESP-01 está conectado no GND da protoboard.
- Ch-Pd do ESP-01 está conectado no 3.3V da protoboard.
- O GND do Arduino deve ser ligado do GND do protoboard.

Para o desenho do esquema utilizamos o aplicativo Fritzing, um diagrama de Fritzing é uma representação gráfica da aparência do circuito montado, uma das vantagens é que possui um símbolo semelhante ao componente físico o que facilita sua montagem.

Na figura 5 temos o diagrama completo do nosso projeto IoT.

Figura 5. Esquema Inicial do Projeto



Definindo uma plataforma big data para IoT

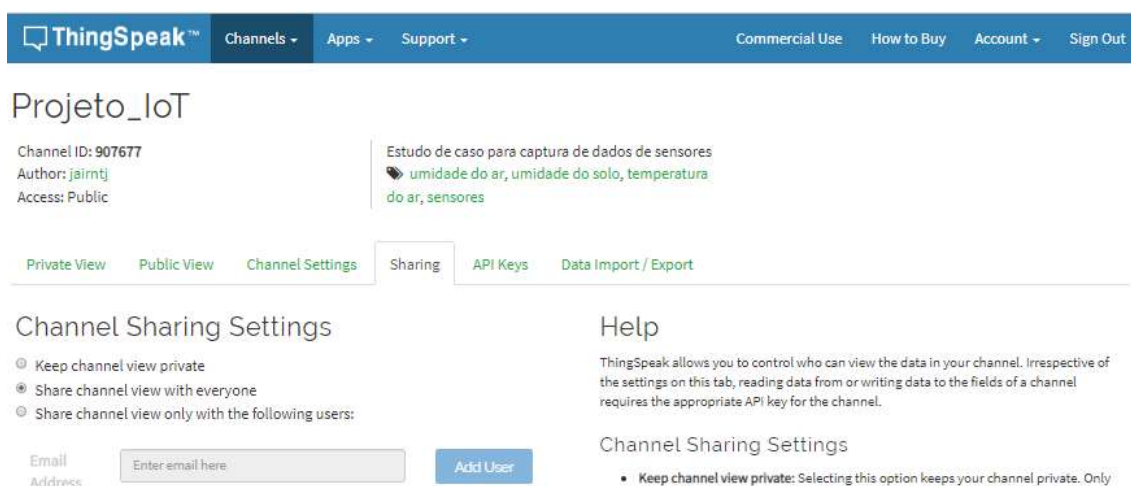
Após o dispositivo montado e testado vamos dar sequência ao nosso projeto implementando uma plataforma Big Data. Foi escolhida uma Plataforma como Serviço (PaaS) para este projeto, pois desta forma podemos nos concentrar muito mais na funcionalidade de seus aplicativos. Entre as empresas citadas anteriormente optamos pela plataforma cloud ThingSpeak para projetos IoT pela sua facilidade de uso com Arduino, por possuir vários recursos on-line, além de não requerer cartão de crédito para registro gratuito e uso acadêmico. Como certeza é a melhor maneira de começar a explorar o mundo da internet das coisas.

As etapas básicas consistem em:

- Abrir o site <https://thingspeak.com> para registro e inscrição.
- Criar um canal selecionando Channels.

Após concluída as etapas acima devemos anotar o Write API Key, pois esta chave é necessária para gravar os dados no canal e deverá ser inserida em statusChWhiteKey conforme mostrado na figura 6.

Figura 8. Tela de opção para canal privado ou público



Na plataforma em channels é possível visualizar vários canais públicos ou buscar um canal específico através do ID, que neste caso cadastramos como jairmtj e Projeto_IoT. Abaixo tela de visualização do canal público, na opção *Export recent data* o usuário pode exportar os 100 últimos dados armazenados.

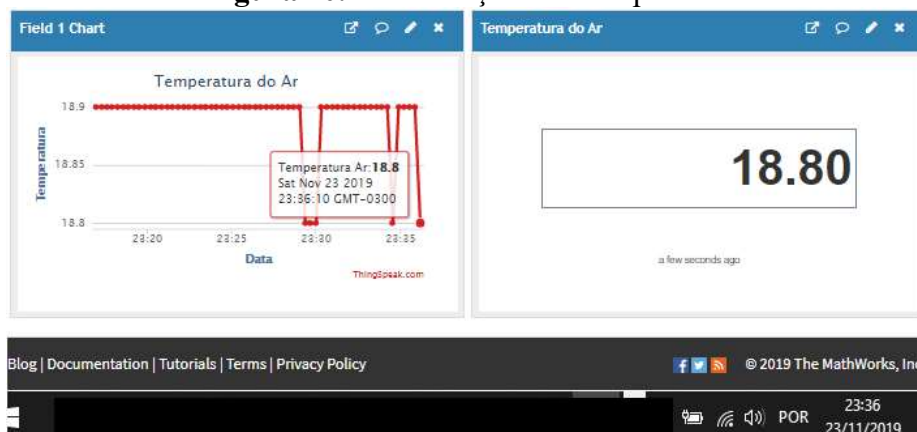
Figura 9. Tela visualização pública



Extraindo os dados para visualização

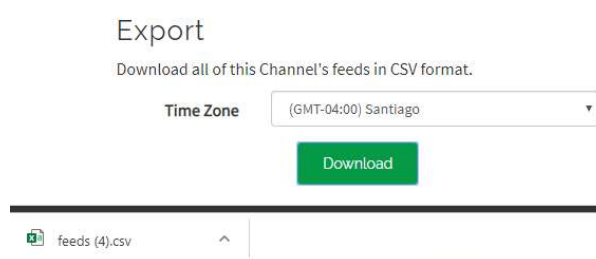
Existe duas maneiras de visualização dos dados, sendo uma em tempo real e outra em lote. Nas visualizações em tempo real os dados são gerados pelos sensores instalados, coletados pelo microcontrolador e transmitidos para a plataforma ThingSpeak pelo módulo Wi-Fi ESP-01. A plataforma por sua vez, transmite os dados através de suas API para visualização e ao mesmo tempo armazena esses dados. Na figura 10 temos um exemplo de dados sendo visualizados em tempo real.

Figura 10. Visualização em Tempo Real



Para extrair os dados em lote, devemos ir na aba *Data Import/Export* e em *Export* devemos escolher o fuso horário, quando houver e clicar em *download*. Devido a não mudança do horário de verão neste ano de 2019, escolhemos o fuso horário de (GMT-04:00) Santiago para que as datas e horários correspondam à realidade. Na figura 11, exemplo da exportação dos dados onde é gerado um arquivo com nome feeds.csv.

Figura 11. Exportação dos Dados



Durante o período de 19/11/2019 às 00:00:10 até o dia 23/11/2019 às 23:59:58, foram geradas 21688 linhas de dados, cada linha possui data/hora, Id e valores dos sensores (figura 12). Dados como *latitude*, *longitude*, *elevation* e *status* não foram validados nesta coleta e deverão ser descartados no processo de transformação.

Figura 12. Arquivo feeds.csv

	A
1	created_at,entry_id,field1,field2,field3,latitude,longitude,elevation,status
2	2019-11-19 00:00:10 -03,1,19.30,92.20,272,,,,
3	2019-11-19 00:00:49 -03,2,19.30,92.30,272,,,,
...	
21686	2019-11-23 23:59:00 -03,21685,18.80,99.90,270,,,,
21687	2019-11-23 23:59:20 -03,21686,18.80,99.90,270,,,,
21688	2019-11-23 23:59:39 -03,21687,18.80,99.90,270,,,,
21689	2019-11-23 23:59:58 -03,21688,18.80,99.90,271,,,,

Carregando os Dados

Para gerar as visualizações dos dados extraídos iremos a utilizar o Power BI e o WEKA para crias as visualizações, primeiramente iremos carregar os dados. No Power BI em obter dados, selecione a opção desejada, no nosso caso selecionaremos texto/csv.

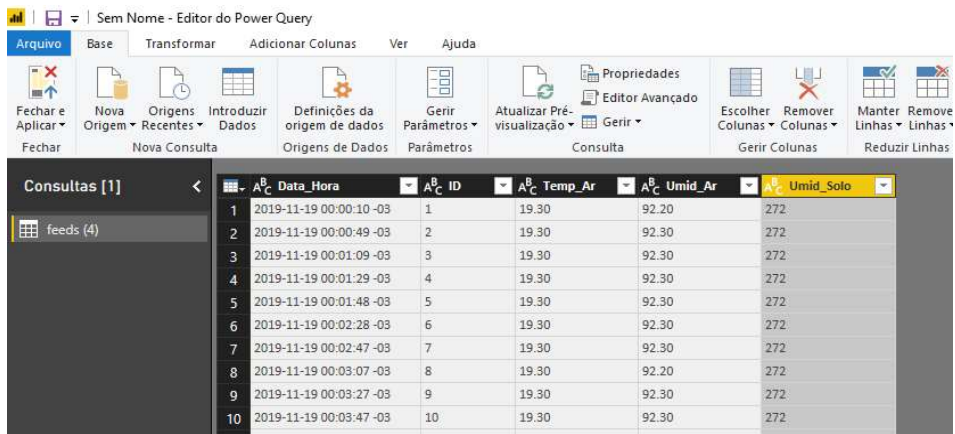
Na figura 13 temos uma pré-visualização dos dados obtidos.

Figura 13. Dados Obtidos para Análise

created_at	entry_id	field1	field2	field3	latitude	longitude	elevation	status
19/11/2019 00:00:10 -03:00	1	1930	9220	272				
19/11/2019 00:00:49 -03:00	2	1930	9230	272				
19/11/2019 00:01:09 -03:00	3	1930	9230	272				
19/11/2019 00:01:29 -03:00	4	1930	9230	272				
19/11/2019 00:05:05 -03:00	14	1930	9230	272				
19/11/2019 00:05:25 -03:00	15	1930	9220	272				
19/11/2019 00:05:45 -03:00	16	1940	9230	272				
19/11/2019 00:06:05 -03:00	17	1940	9230	272				
19/11/2019 00:06:25 -03:00	18	1930	9220	272				
19/11/2019 00:06:44 -03:00	19	1930	9220	272				
19/11/2019 00:07:04 -03:00	20	1940	9220	271				

Antes de carregarmos os dados para gerar as visualizações devemos editar descartando os dados nulos ou em branco. Na tela de pré-visualização clique em editar e faça as modificações necessárias. Na figura 14 temos as colunas latitude, longitude, elevation e status removidas. Também renomeamos as colunas dos dados que vamos utilizar.

Figura 14. Dados Editados

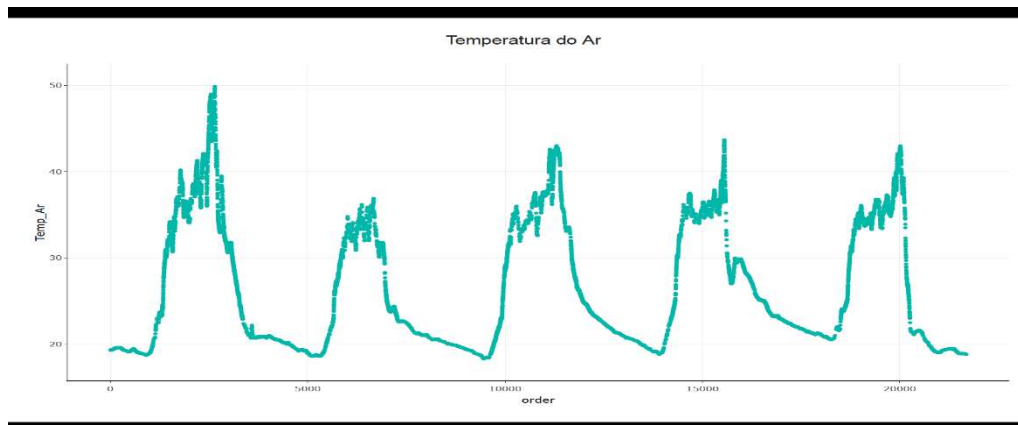


	A ^B _C Data_Hora	A ^B _C ID	A ^B _C Temp_Ar	A ^B _C Umid_Ar	A ^B _C Umid_Solo
1	2019-11-19 00:00:10-03	1	19.30	92.20	272
2	2019-11-19 00:00:49-03	2	19.30	92.30	272
3	2019-11-19 00:01:09-03	3	19.30	92.30	272
4	2019-11-19 00:01:29-03	4	19.30	92.30	272
5	2019-11-19 00:01:48-03	5	19.30	92.30	272
6	2019-11-19 00:02:28-03	6	19.30	92.30	272
7	2019-11-19 00:02:47-03	7	19.30	92.30	272
8	2019-11-19 00:03:07-03	8	19.30	92.20	272
9	2019-11-19 00:03:27-03	9	19.30	92.30	272
10	2019-11-19 00:03:47-03	10	19.30	92.30	272

Gerando gráficos

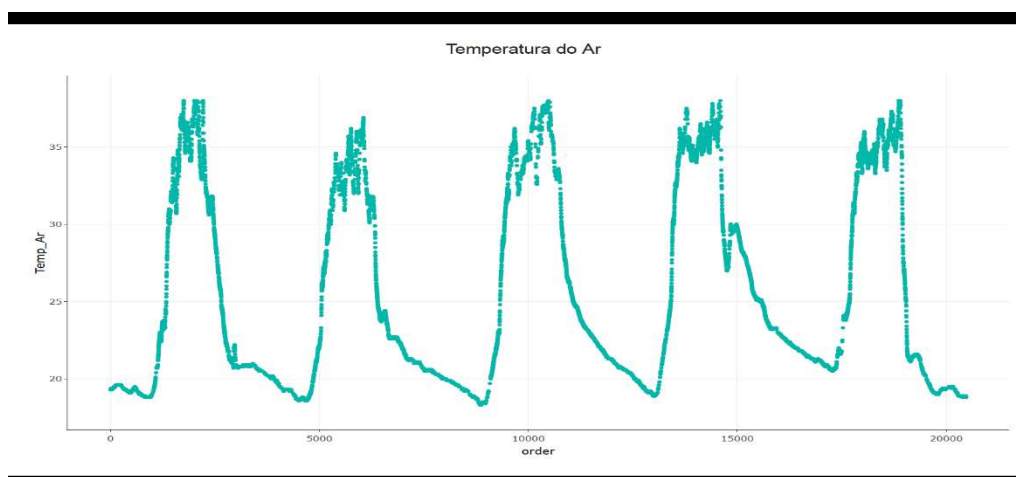
A primeira visualização gerada foi a temperatura do ar e já podemos ver que existe algo de errado com os valores, pois segundo Pegorim (2019) a maior temperatura registrada na cidade de São Paulo deste 1961 até 2018 foi de 37.8°C.

Gráfico 1. Temperatura do Ar



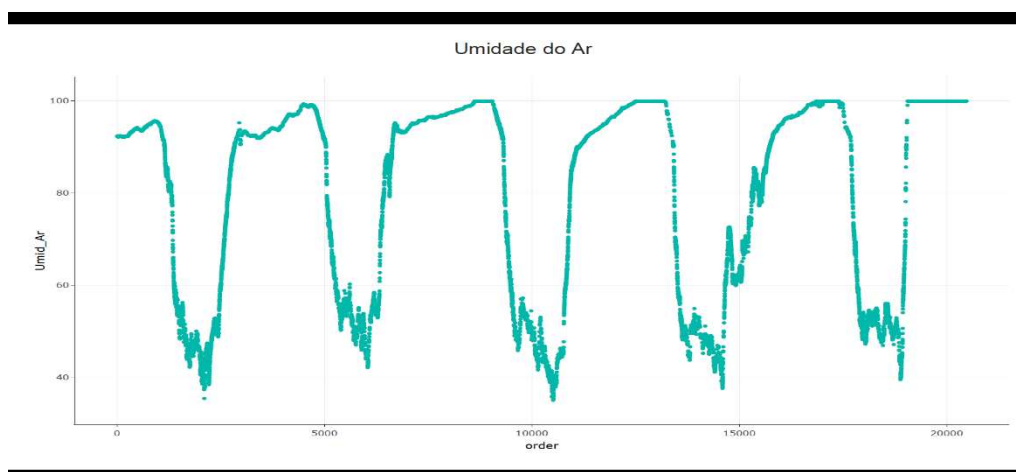
Essas temperaturas foram vistas também em tempo real, percebemos que durante a noite os valores correspondiam ao Climatempo e durante o dia os valores eram bem superiores aos do site, pois o sol estava batendo diretamente na caixa do circuito o que causava o aquecimento dela interferindo assim nos valores gerados pelo sensor que estava dentro da caixa. Para corrigir esta falha retiramos o sensor de temperatura e umidade de dentro da caixa o que causou uma redução dos valores obtidos. Com base nas informações do Climatempo, descartamos os valores acima de 38°C, gerando o gráfico com valores mais próximos da realidade.

Gráfico 2. Valores excluídos



No gráfico 3 podemos visualizar os dados da umidade do ar.

Gráfico 3. Umidade do Ar



Ao juntarmos os dois valores notamos que um é o contrário do outro, ou seja, quanto maior a temperatura do ar, menor será a umidade do ar ou vice-versa.

Gráfico 4. Comparação entre temperatura e umidade

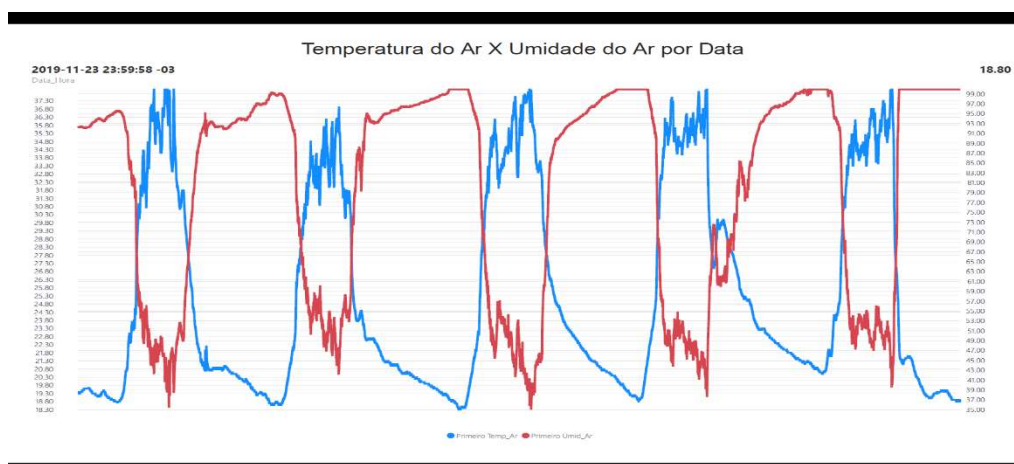
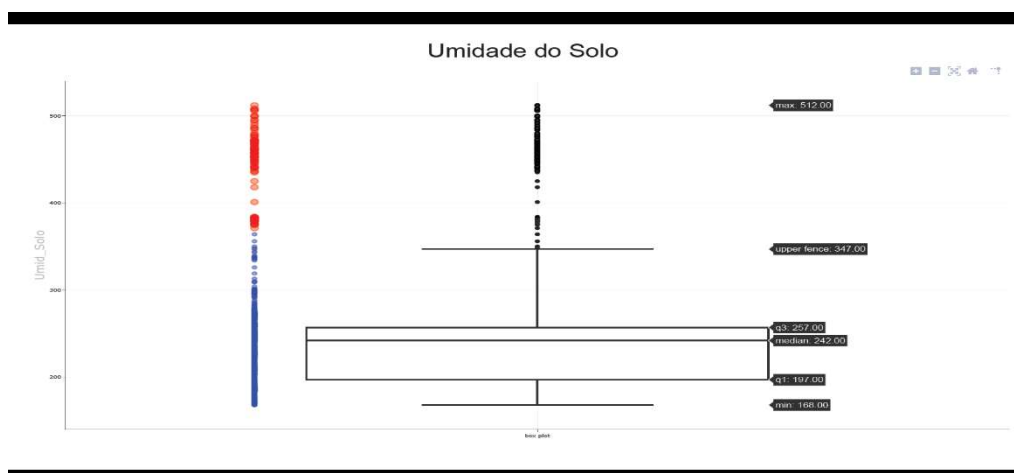


Gráfico 5. Umidade do solo



O gráfico 5 geramos um box plot e foi possível identificar alguns *Outliers* que estão acima do limite superior, os valores obtidos são:

- Limite Inferior – 168.00
- Primeiro Quartil Q_1 – 197.00
- Segundo Quartil ou Mediana Q_2 – 242.00
- Terceiro Quartil Q_3 – 257.00
- Limite Superior – 347.00

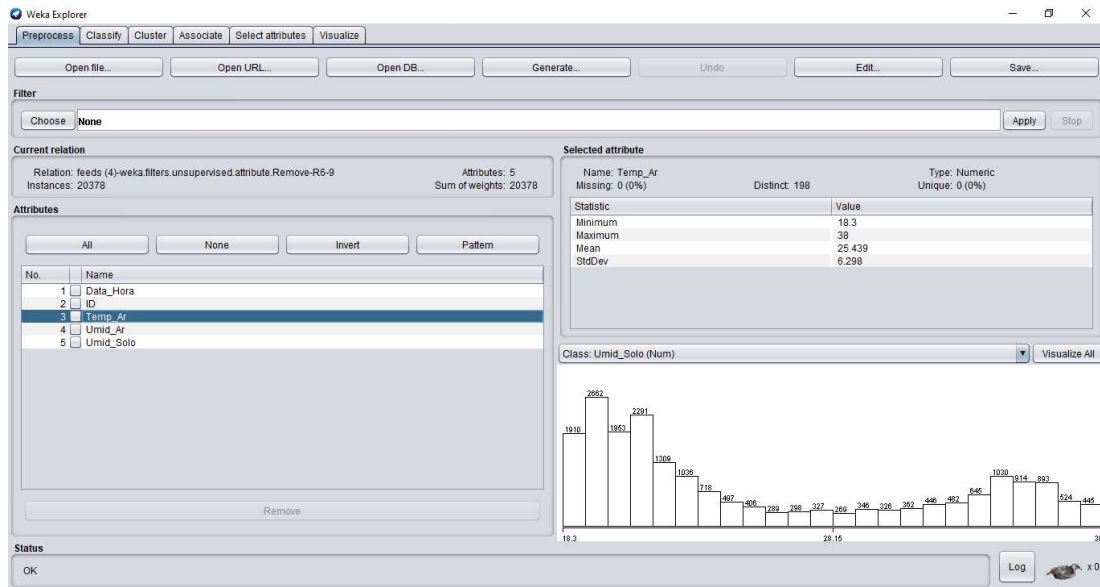
Ainda é cedo para estipular um padrão para analisar a umidade do solo, caso tivéssemos pelo menos 30 dias de dados coletados, poderíamos criar um padrão para solo seco, úmido e molhado. Após removermos os *Outliers* abaixo os novos valores:

- Limite Inferior – 167.00
- Primeiro Quartil Q_1 – 197.00
- Segundo Quartil ou Mediana Q_2 – 234.00
- Terceiro Quartil Q_3 – 250.00

- Limite Superior – 326.00

Para gerar as visualizações no WEKA devemos pré-processar os dados da mesma forma que fizemos no Power BI. Na figura 19 os dados já foram tratados, foram excluídas as colunas em branco, as temperaturas acima de 38°C, os *Outliers* do sensor de umidade de solo e renomeado as colunas.

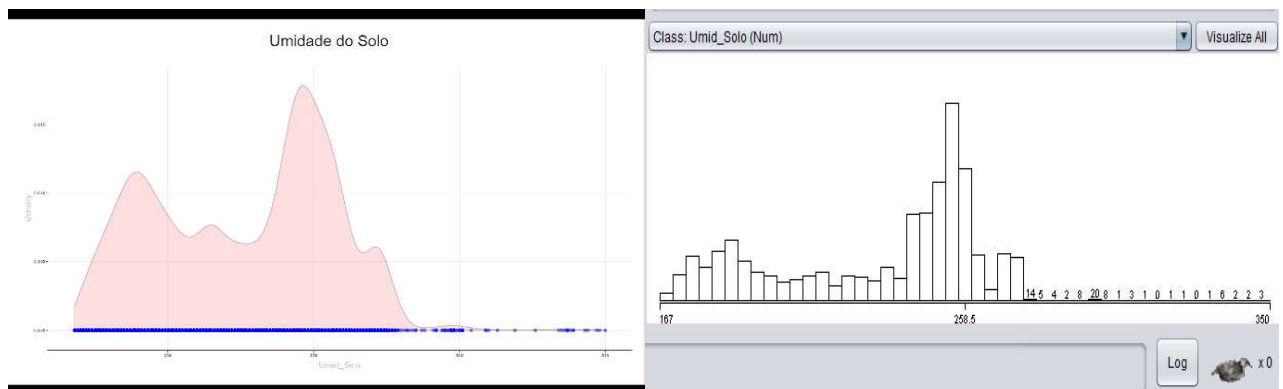
Figura 15. Tela de pré-processamento com dados tratados



Uma das vantagens do WEKA é que na tela de pré-processamento ele já nos traz informações como mínima, máxima, média, desvio padrão e a visualização de um histograma.

Segundo Wilke (2019), os histogramas são uma opção popular de visualização desde pelo menos o século 18, em parte porque eles são facilmente gerados à mão. Com o crescimento das tecnologias e o poder computacional os histogramas vêm sendo substituídos cada vez mais por gráficos de densidade em que tentamos visualizar a distribuição de probabilidade subjacente dos dados desenhando uma curva contínua. Na figura 20 comparamos um gráfico de densidade gerado pelo Power BI a partir dos dados coletados dos sensores e um histograma gerado pelo WEKA com os mesmos dados e podemos notar a semelhança entre eles.

Figura 16. Densidade Power BI e Histograma WEKA



CONCLUSÕES

A presente pesquisa se propôs a desenvolver um dispositivo IoT com o objetivo de coletar dados gerados por sensores, a implementação de arquitetura *big data* ou plataformas cloud como serviços (PaaS) e gerar visualizações em tempo real ou em lote para análises.

No processo de desenvolvimento do *Less Water*, pesquisei sobre alguns microcontroladores e dispositivos para conectividade, pois para ser considerado um dispositivo IoT ele precisa estar conectado. O que pude perceber foi que um simples dispositivo gera uma quantidade de dados sem precedentes, como já dito anteriormente em apenas 4 dias foram geradas 21688 linhas de dados contendo data/hora, ID e dados de 3 sensores, uma média de 5422 dados por dia ou para ser mais exato uma linha de dados a cada 16 segundos. Em um ano este único dispositivo pode gerar até 1.979.000 linhas de dados ou 78,8Mbytes por ano. A questão é, haverá infraestrutura para atender toda esta demanda?

Isso irá depender de como e o que queremos consumir como dados. Várias plataformas como serviços (PaaS) oferecem serviços voltados para o universo IoT, escolhi a Plataforma ThingSpeak por ter sido citada em diversas bibliografias, pela facilidade de implantação e para estudos acadêmicos se torna atrativa por não exigir cartão de crédito. Nesse projeto de pesquisa ela atendeu bem aos meus objetivos, disponibilizando visualizações dos dados em tempo real e armazenando ao mesmo tempo.

Com relação as ferramentas de visualização e análise, foi possível detectar algumas falhas nos sensores de temperatura e de umidade de solo, no caso da temperatura os valores estavam bem acima do normal, causando assim uma ação corretiva, já no sensor de umidade do solo o gráfico do box plot nos permitiu criar uma escala para as condições do solo e isso poderá ajudar em trabalhos futuros quando for implantado um sistema de irrigação em que com base nos dados dos sensores decida quando irrigar.

Em outra oportunidade de pesquisa verificar se é possível incluir algum algoritmo de aprendizado de máquina em que através de dados coletado possa decidir se deve irrigar ou não.

REFERÊNCIAS BIBLIOGRÁFICAS

ADNEWS, Redação. **5 plataformas de gestão de dados que todo profissional de marketing deveria conhecer**. Disponível em: <<https://adnews.com.br/adtech/5-plataformas-de-gestao-de-dados-que-todo-profissional-de-marketing-deveria-conhecer/>>. Acesso em: 04 jul. 2019.

ARDUINO. **ARDUINO UNO REV3**. [S. l.], 6 nov. 2019. Disponível em: <https://store.arduino.cc/usa/arduino-uno-rev3>. Acesso em: 27 out. 2019.

CISCO, Cisco. **Cisco Visual Networking Index: Forecast and Methodology, 2017-2022**. 2019. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>>. Acesso em: 04 julho.2019.

DASGUPTA, Nataraj. **Practical Big Data Analytics: Hands-on techniques to implement enterprise analytics and machine learning using Hadoop, Spark, NoSQL and R**. United Kingdom: Packt Publishing Limited, 2018. 412 p. ISBN 9781783554393. Disponível em:

<https://learning.oreilly.com/library/view/practical-big-data/9781783554393/>. Acesso em: 17 ago. 2019.

DHIRAJ, Ambiga; CHAMBERS, Michele; MINELLI, Michael. **Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses**. New Jersey: John Wiley & Sons, Inc., 2013. 224 p. (ISBN: 9781118147603). Disponível em: <<https://learning.oreilly.com/library/view/big-data-big/9781118239155/>>. Acesso em: 19 jul. 2019.

DIÁRIO OFICIAL DA UNIÃO. DECRETO nº 9.854, de 25 de junho de 2019. Art. 84, caput, inciso VI, alínea “a”, da Constituição. **Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas.**, Brasília, p. 10-11, 26 jun. 2019. Disponível em: <http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=26/06/2019&jornal=515&pagina=10&totalArquivos=149>. Acesso em: 26 nov. 2019.

DU, Dayong. **Apache Hive Essentials**. 2. ed. rev. Birmingham, UK.: Packt Publishing Ltd., 2018. 210 p. ISBN 9781788995092. Disponível em: <https://learning.oreilly.com/library/view/apache-hive-essentials/9781788995092/>. Acesso em: 31 ago. 2019.

ELK, Klaus. **Embedded Software for the IoT**. 3ª. Ed. Berlim, Alemanha: Walter de Gruyter Inc., 2018. ISBN 978-1-5474-0104-8. *E-book*.

GARTNER, INC. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide. **Magic Quadrant for Cloud Infrastructure as a Service, Worldwide**, [s. l.], 16 jul. 2019. Disponível em: <https://www.gartner.com/doc/reprints?id=1-10BCYZ24&ct=190725&st=sb>. Acesso em: 11 out. 2019.

GOLDEN, Bernard. **Amazon Web Services™ For Dummies®**. NJ, USA: John Wiley & Sons, Inc., 2013. 346 p. ISBN 978-1-118-65198-8. *E-book*.

HUNTER, Ted; PORTER, Steven; PS, Legorie Rajan. **Building Google Cloud Platform Solutions**. Birmingham, UK.: Packt Publishing Ltd., 2019. ISBN 978-1-83864-743-8. *E-book*.

HURWITZ, Judith et al. **Big Data para leigos**. 1ª. ed. Rio de Janeiro: Alta Books, 2016. 302 p.

IBM CORPORATION. **IBM Watson IoT Platform**. [S. l.], 2019. Disponível em: https://www.ibm.com/support/knowledgecenter/en/SSQP8H/IoT/kc_welcome.htm. Acesso em: 12 nov. 2019.

JAIN, Ankit. **Mastering Apache Storm**. Birmingham, UK.: Packt Publishing Ltd., 2017. ISBN 978-1-78712-563-6. Disponível em: <https://learning.oreilly.com/library/view/mastering-apache-storm/9781787125636>. Acesso em: 19 set. 2019.

KURNIAWAN, Agus. **Intelligent IoT Projects in 7 Days**. Birmingham, UK.: Packt Publishing Ltd., 2017. ISBN 978-1-78728-642-9. Disponível em: <https://learning.oreilly.com/library/view/intelligent-IoT-projects/9781787286429/>. Acesso em: 25 out. 2019.

LUCIETTI, Donato. **Irrigação das Hortaliças**. Santa Catarina, 2014. Disponível em: <http://cultivehortaorganica.blogspot.com/2014/01/irrigacao-das-hortalicas.html>. Acesso em: 25 out. 2019.

MASHEY, John. **Big Data and the Next Wave of InfraStress**. 1998. Disponível em: <http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf>. Acesso em: 08 jul. 2019

MODI, Ritesh. **Azure for Architects: Implementing cloud design, DevOps, containers, IoT, and serverless solutions on your public cloud**. 2ª. ed. Birmingham, UK.: Packt

Publishing Ltd., 2019. 536 p. ISBN 9781789614503. Disponível em: <https://learning.oreilly.com/library/view/azure-for-architects/9781789614503/>. Acesso em: 14 out. 2019.

PEGORIM, Josélia. **Novo recorde de calor em São Paulo**. [S. l.]: ClimaTempo, 9 jan. 2019. Disponível em: <https://www.climatempo.com.br/noticia/2019/01/09/novo-recorde-de-calor-em-sao-paulo-0898>. Acesso em: 23 nov. 2019.

PLANTE, Russell H. **Solar Energy, Photovoltaics, and Domestic Hot Water**. Califórnia, EUA: Elsevier Editora LTDA, 2014. ISBN 978-0-12-420155-2. Disponível em: <https://learning.oreilly.com/library/view/solar-energy-photovoltaics/9780124201552/>. Acesso em: 19 nov. 2019.

RIJMENAM, Mark Van. **Think Bigger: Developing a Successful Big Data Strategy for Your Business**. New York: Amacom, 2014. 288 p. (ISBN-13: 978-0-8144-3415-4). Disponível em: <https://learning.oreilly.com/library/view/think-bigger/9780814434154/>. Acesso em: 02 jul. 2019.

ROUSE, Margaret. **What is big data?** Disponível em: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>>. Acesso em: 20 ago. 2019.

ROVAI, Marcelo José. **“IoT feito fácil”**: Conectando coisas desde qualquer canto do mundo! [S. l.], 2017. Disponível em: <https://mjrobot.org/2017/12/21/IoT-feito-facil-conectando-coisas-desde-qualquer-canto-do-mundo/>. Acesso em: 6 nov. 2019.

SARKAR, Aurobindo. **Learning Spark SQL: Architect streaming analytics and machine learning solutions**. Birmingham - Mumbai: Packt Publishing, 2017. 452 p. Disponível em: <https://www.safaribooksonline.com/library/view/learning-spark-sql/9781785888359/7b6bdf28-8f58-47f5-a94c-1fb7d7261973.xhtml>>. Acesso em: 26 fev. 2018.

SCHWARTZ, Marco. **Internet of Things with Arduino Cookbook**. Birmingham, UK.: Packt Publishing Ltd., 2016. ISBN 978-1-78528-658-2. Disponível em: <https://learning.oreilly.com/library/view/internet-of-things/9781785286582/>. Acesso em: 13 nov. 2019.

SHAH, Amit; SARKAR, Aurobindo. **Learning AWS: Design, build, and deploy responsive applications using AWS Cloud components**. 2ª. ed. Birmingham, UK.: Packt Publishing Ltd., 2018. ISBN 978-1-78728-106-6. Disponível em: <https://learning.oreilly.com/library/view/learning-aws-/9781787281066>. Acesso em: 15 out. 2019.

SHIELDS, Anne. **Is Azure Responsible for Microsoft's All-Time Highs?: Amazon and Microsoft are leading the public cloud space**. 2018. Disponível em: <https://marketrealist.com/2018/03/is-azure-responsible-for-microsofts-all-time-highs>>. Acesso em: 09 mar. 2019.

SHOVIE, John C. **Raspberry Pi IoT Projects: Prototyping Experiments for Makers**. Washington, USA: Apress, 2016. ISBN 978-1-4842-1377-3. *E-book*.

TAURION, Cezar. **Big Data**. Rio de Janeiro: Brasport, 2013. 102 p.

THINGSPEAK. **ThingSpeak for IoT Projects**. [S. l.], 2019. Disponível em: <https://thingspeak.com>. Acesso em: 12 nov. 2019.

TUTORIALSPPOINT (org.). **Internet of Things (IoT) Tutorial**. 2016. Disponível em: https://www.tutorialspoint.com/internet_of_things/index.htm>. Acesso em: 07 maio 2018.

TUTORIALSPPOINT (org.). **Power BI Tutorial**. [S. l.], 2017. Disponível em: https://www.tutorialspoint.com/power_bi/index.htm. Acesso em: 6 out. 2019.

TUTORIALSPPOINT (org.). **Tableau Tutorial**. [S. l.], 2016. Disponível em: <https://www.tutorialspoint.com/tableau/index.htm>. Acesso em: 6 out. 2019.