

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE E VISUALIZAÇÃO DE DADOS DE LOCALIZAÇÃO DE
DISPOSITIVOS MÓVEIS CONECTADOS À INTERNET**

Nome: Marcelo Augusto Cordeiro

R.A.: 121023265

Orientador: Prof. Dr. Eduardo Martins Morgado

BAURU - SP
Maio/2016

MARCELO AUGUSTO CORDEIRO

**ANÁLISE E VISUALIZAÇÃO DE DADOS DE LOCALIZAÇÃO DE
DISPOSITIVOS MÓVEIS CONECTADOS À INTERNET**

Anteprojeto de pesquisa para o Trabalho de
Conclusão de Curso do Curso de Bacharelado
em Ciência da Computação da Universidade
Estadual Paulista “Júlio de Mesquita Filho”,
Faculdade de Ciências, campus Bauru.

Orientador: Prof. Dr. Eduardo Martins Morgado

BAURU - SP

Maio/2016

SUMÁRIO

1 INTRODUÇÃO	2
2 PROBLEMA	4
3 JUSTIFICATIVA	5
4 OBJETIVOS	7
4.1 Objetivo geral	7
4.2 Objetivo específico	7
5 FUNDAMENTAÇÃO TEÓRICA.....	8
5.1 Internet das Coisas	8
5.2 Big Data.....	8
5.3 Amazon Web Services	14
6 MÉTODO DE PESQUISA	15
7 CRONOGRAMA.....	16
8 REFERÊNCIAS.....	17

1 INTRODUÇÃO

As técnicas de localização sem fio podem ser divididas em dois grupos principais: técnicas de auto posicionamento e técnicas de posicionamento remoto.

Na primeira técnica, o dispositivo móvel utiliza sinais transmitidos por antenas para calcular a sua própria posição. Na segunda técnica, a posição do dispositivo móvel é calculada por uma série de receptores que medem os sinais recebidos e enviados pelo dispositivo (ZEIMPEKIS, GIAGLIS e LEKAKOS, 2003).

Os tradicionais sistemas de GPS (Global Positioning System) utilizam a técnica de auto posicionamento para calcular sua posição no globo terrestre baseado nos sinais recebidos de 24 satélites posicionados na órbita terrestre com 20.200 quilômetros de distância entre cada um (DJUKNIC e RICHTON, 2001).

Entretanto, a força do sinal GPS não é suficiente para penetrar a maioria dos prédios. A reflexão do sinal muitas vezes permite a leitura em ambientes fechados, porém o cálculo da posição não será confiável (CHEN e KOTZ, 2000). Portanto, são necessárias soluções diferentes para se criar um sistema de geoposicionamento que funcione em ambientes fechados.

Uma das melhores maneiras de se abordar esse problema é através da Internet das Coisas (Internet of Things – IoT).

O termo IoT foi utilizado pela primeira vez por Kevin Ashton em 1999 (ASHTON, 2009), e pode ser definido como “A habilidade de comunicação, conectividade, e computação de dispositivos compartilhando dados via internet para ajudar a melhorar produtos, serviços, capacidade de resposta e qualidade de vida.” (SMITH, 2015, tradução nossa)¹.

Por exemplo, utilizando uma série de sensores wi-fi posicionados em pontos fixos dentro de um prédio, com a triangulação do sinal é possível calcular a posição de dispositivos conectados à rede wi-fi (BLECKY, 2016).

Para oferecer uma posição confiável, é necessário que estes sensores colem e transmitam a força do sinal wi-fi em cada dispositivo com uma alta frequência.

Utilizando como exemplo o prédio do Laboratório de Tecnologia da Informação Aplicada (LTIA) da Faculdade de Ciências da Unesp de Bauru, em um dia

¹ Texto original: “The communication, connectivity, and computing ability of devices sharing data via the Internet to help improve products, services, responsiveness, and quality of life.”

comum, é observado uma média de 30 dispositivos conectados à rede wi-fi. Considerando um sensor que a cada 30 segundos colete 1 kB de dados de cada dispositivo, por mês, seriam coletados mais de 2 GB de dados. Portanto, para garantir um sistema escalável, é necessário a utilização de técnicas de Big Data para armazenar e manipular esses dados.

O melhor modo de se definir Big Data ainda é discutido por pesquisadores, mas uma definição simples é a de que “se é necessário se preocupar com o tamanho dos dados, então é Big Data.” (ESPOSITO, 2015, tradução nossa)².

² Texto original: “If I have to worry about the size of the data, then the data is big.”

2 PROBLEMA

Os sistemas de geoposicionamento atuais necessitam estar constantemente conectados à satélites ou à internet, o que impossibilita o seu uso em ambientes fechados em que a conectividade é limitada.

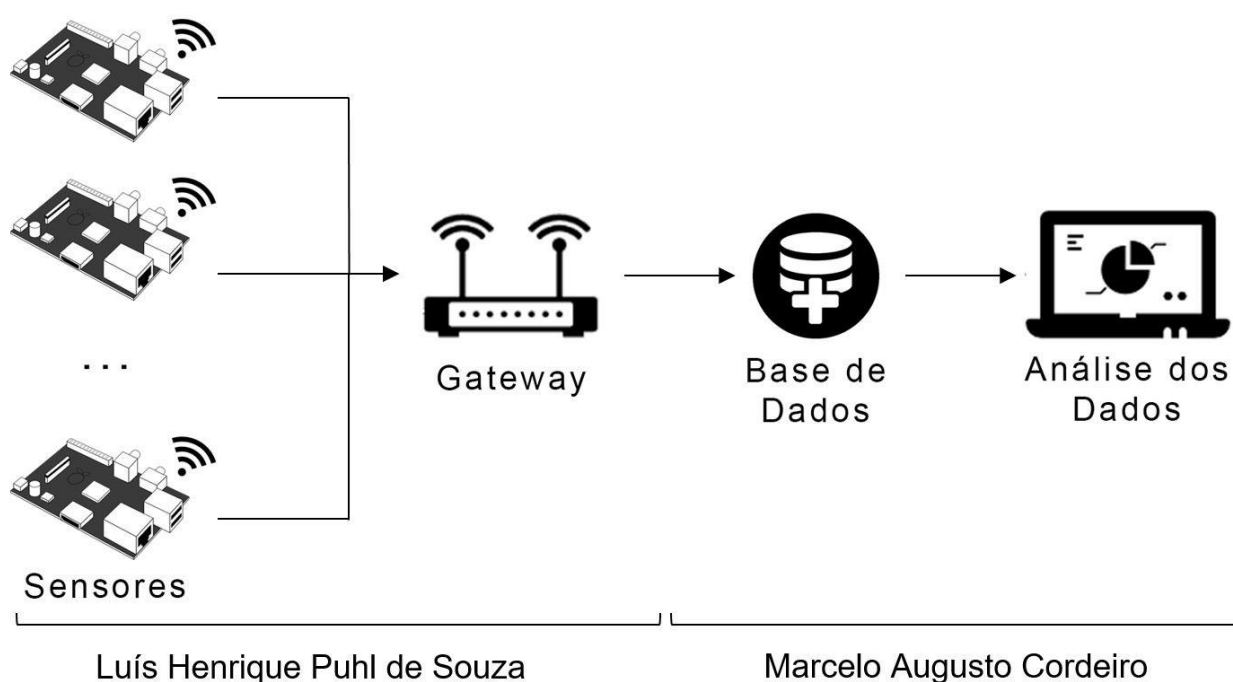
Adicionalmente, a constante monitoração da posição de dispositivos dentro de um prédio pode gerar informações valiosas que atualmente não são exploradas.

3 JUSTIFICATIVA

O projeto aqui proposto é sistema independente de geoposicionamento para ambientes fechados.

O aluno Luís Henrique Puhl de Souza também irá desenvolver parte deste projeto como seu Trabalho de Conclusão de Curso. Ele ficará responsável por todos os sensores, o que inclui sua construção, posicionamento, e configuração e programação dos gateways.

Figura 1: Divisão do Sistema



Fonte: elaborado pelo autor

Em um sistema de IoT, o gateway é o ponto de acesso único que conecta todos os sensores à internet. Deste modo, a base de dados não precisa se conectar com cada um dos sensores, todos os dados coletados já ficam agrupados e filtrados no gateway (KONSEK, 2015).

A estrutura do sistema aqui proposto pode ser facilmente escalada para diversas finalidades, desde o controle de funcionários dentro de uma empresa até um sistema completo de GPS para ambientes subterrâneos.

Por exemplo, considerando um ambiente onde a maior parte das pessoas está sempre carregando consigo pelo menos um dispositivo conectado à internet, monitorar a posição destes dispositivos significa também monitorar a posição das pessoas naquele ambiente.

Essa informação pode ser utilizada para as mais diversas finalidades, como para determinar as áreas do prédio onde se tem a maior concentração de pessoas durante um determinado período de tempo. Em ambientes comerciais, por exemplo, essa informação pode ser utilizada para melhor distribuição dos funcionários nas áreas com maior número de clientes.

Para possibilitar testes em um ambiente real, o projeto aqui proposto será instalado dentro do prédio do LTIA.

4 OBJETIVOS

4.1 Objetivo geral

Desenvolver uma aplicação web para processamento e visualização de dados gerados a partir de uma série de sensores de wi-fi e bluetooth fixados em pontos pré-determinados dentro de um prédio.

4.2 Objetivo específico

- Estudar ferramentas para armazenamento e processamento de Big Data;
- Definir a melhor ferramenta de Big Data para tratar a quantidade de dados gerados pelos sensores;
- Construir uma base de dados apropriada para a quantidade de dados do projeto;
- Planejar a estrutura do aplicativo web;
- Estudar as tecnologias web mais atuais;
- Definir as tecnologias web que serão usadas para o desenvolvimento do aplicativo;
- Implementar o aplicativo web;
- Testar a integridade do aplicativo perante situações não-ideais;
- Realizar os ajustes necessários para garantir um aplicativo estável e seguro;

5 FUNDAMENTAÇÃO TEÓRICA

“A próxima era da computação será fora do ambiente de computadores tradicionais. Com o paradigma da internet das coisas, muitos dos objetos que nos rodeiam estarão conectados de uma forma ou de outra. Identificação por Rádio Frequência (RFID) e as tecnologias de redes de sensores irão evoluir para satisfazer este novo desafio, no qual informações e sistemas de comunicação estarão invisivelmente integrados ao ambiente à nossa volta. Isso irá resultar na geração de quantidades enormes de dados, que precisam ser armazenados, processados e apresentados de uma forma eficiente e facilmente interpretável.”³ (GUBBI, BUYYA, *et al.*, 2013, tradução nossa)

5.1 Internet das Coisas

Internet das coisas é atualmente a tecnologia emergente mais empolgante entre desenvolvedores do mundo inteiro (GARTNER, 2015). Segundo Esposito (2015), uma pesquisa realizada em 2015 com mais de 500 profissionais da área de tecnologia da informação mostrou que IoT já é relevante para 58% do mercado de tecnologia, e 87% acreditam que será relevante no futuro de suas empresas.

Um dos maiores problemas da computação tradicional é a forma de se criar e capturar dados. A grande maioria de dados disponíveis na internet atualmente foram criados ou capturados por seres humanos, que possuem atenção, tempo e precisão limitados. Como consequência, conseguimos processar apenas uma pequena fração de toda a informação disponível à nossa volta (ASHTON, 2009).

Para explorar essa quantidade enorme de dados, é necessário computadores e redes de sensores que se encaixem no meio ambiente humano, ao invés de forçar humanos aos ambientes deles, o que torna todo o processo de captura e processamento de dados imperceptível e independente da destreza humana (WEISER, GOLD e BROWN, 1999).

5.2 Big Data

Acredita-se que o conceito de big data tenha se originado do advento de empresas de pesquisas que precisavam agregar e processar enormes quantidades de dados não estruturados (AMAZON WEB SERVICES, 2016).

³ Texto original: “The next wave in the era of computing will be outside the realm of the traditional desktop. In the Internet of Things (IoT) paradigm, many of the objects that surround us will be on the network in one form or another. Radio Frequency IDentification (RFID) and sensor network technologies will rise to meet this new challenge, in which information and communication systems are invisibly embedded in the environment around us. This results in the generation of enormous amounts of data which have to be stored, processed and presented in a seamless, efficient, and easily interpretable form.”

Até poucos anos atrás, poucas empresas possuíam os recursos para investir em tecnologias capazes de processar grandes quantidades de dados. Porém, em 2009, The Apache Software Foundation liberou o código do framework Hadoop, em desenvolvimento desde 2006, divulgando o modelo de programação MapReduce, capaz de armazenar e processar enormes quantidades de dados através de um sistema de arquivos distribuídos e tolerante a falhas (chamado de HDFS - Hadoop Distributed File System).

Como resultado, Hadoop difundiu o conceito e uso de big data no mercado de tecnologia e se tornou o framework mais utilizado até hoje (KUNTAMUKKALA, 2014).

Desde então, tecnologias de big data são utilizadas quando a quantidade de dados é grande, complexa e variada e difícil de ser adequadamente armazenada, administrada e processada através de bases de dados e programas tradicionais.

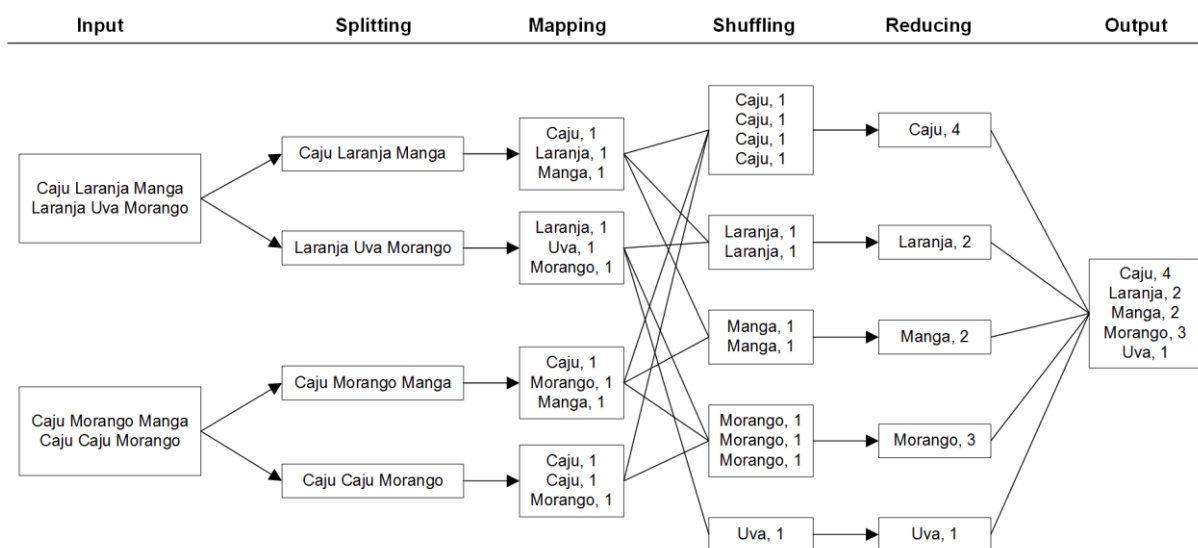
5.2.1 MapReduce

MapReduce é um modelo de programação criado para ser utilizado no Hadoop no processamento de enormes quantidades de dados. Recebe esse nome por se dividir em duas funções principais: *mapping* e *reducing*.

Para realizar o mapeamento, os arquivos desejados são recebidos do HDFS e então mapeados para pares do tipo <chave, valor> através de uma lógica de negócios definida pelo desenvolvedor. Estes resultados são então ordenados e reduzidos através de mais uma lógica de negócios para produzir o resultado final desejado.

A Figura 2 ilustra este processo para, como exemplo, realizar a contagem de palavras em dois arquivos:

Figura 2: Esquema de funcionamento do MapReduce para contagem de palavras



Fonte: elaborado pelo autor

Cada etapa do algoritmo possui uma função específica:

- Input: entrada dos dois arquivos, normalmente capturados de um HDFS.
- Splitting: as linhas dos arquivos são separadas em blocos independentes entre si.
- Mapping: mapeia cada palavra do bloco para uma chave do tipo <palavra, 1>.
- Suffling: ordena os elementos comuns entre os blocos para facilitar a redução.
- Reducing: reduz todos os blocos de acordo com a lógica de negócios desejado, neste caso, somar os valores que possuem a mesma chave.
- Output: arquivo com a contagem de palavras, normalmente enviado para ser armazenado em um HDFS.

5.2.2 Spark

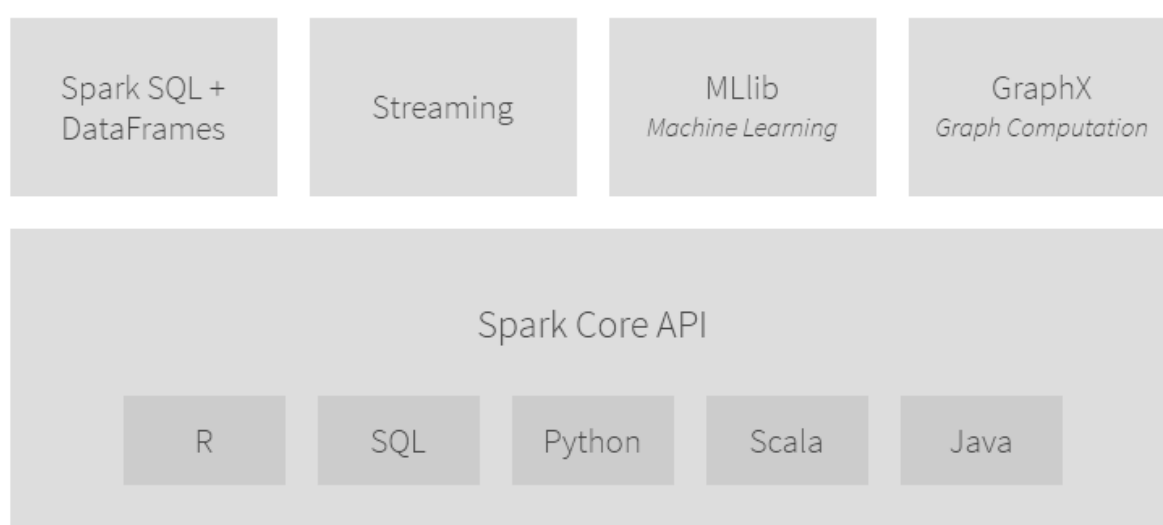
Spark (THE APACHE SOFTWARE FOUNDATION, 2016B) é um framework para processamento *in-memory* de big data em clusters desenvolvido em 2009 na Universidade da Califórnia, Berkeley, que doou o projeto para a The Apache Software Foundation em 2010, que o mantém desde então.

Graças à sua facilidade de uso e velocidade, o Spark se tornou a ferramenta para big data que mais cresce no mercado. Se comparado com Hadoop, Spark

executa até 100 vezes mais rápido em memória e até 10 vezes mais rápido em disco. Ademais, segundo Esposito (2015), 39% dos usuários de Hadoop reportarem já estarem, também, utilizando Spark.

O *Spark Ecosystem* inclui ferramentas avançadas para as mais diversas finalidades, como aprendizado de máquina e banco de dados de grafos. Além disso, um dos maiores diferenciais do Spark é o módulo *Spark SQL*, que, mesmo se tratando de uma base de dados não relacional, permite o uso de SQL de forma extremamente rápida em suas consultas.

Figura 3: Spark Ecosystem



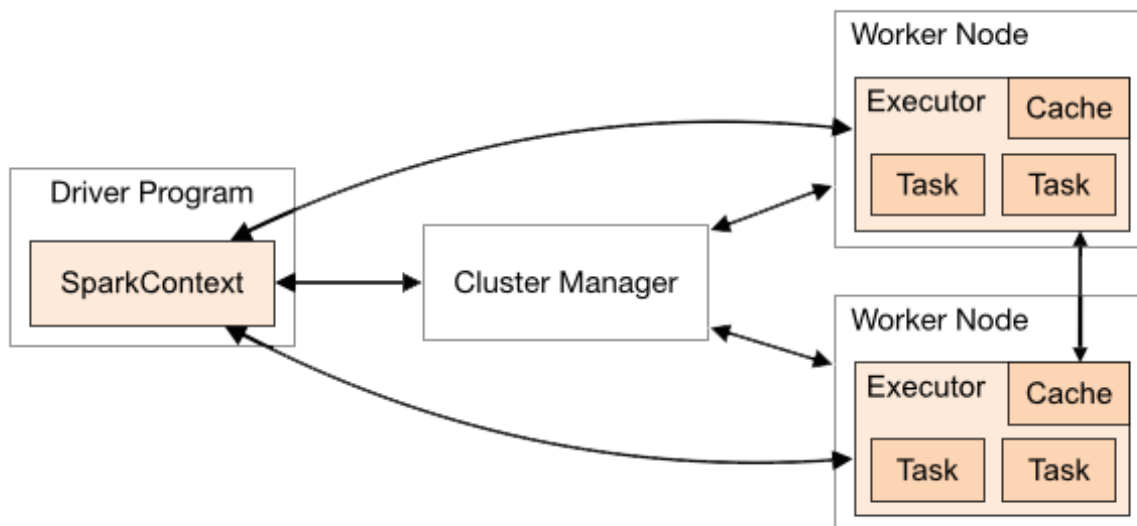
Fonte: <https://databricks.com/spark/about>

Spark, assim como Hadoop, trabalha com o conceito de uma máquina mestre, onde está o *Driver Program*, e máquinas escravas, chamadas de *Worker Nodes*, onde a informação é de fato armazenada. A máquina mestre se comunica principalmente com o *Cluster Manager*, um programa responsável por manter os endereços de cada máquina escrava no cluster, assim como o endereço de suas cópias. Por padrão, cada *Worker Node* será sempre replicado três vezes, assim, se uma das máquinas escravas falhar, o *Cluster Manager* automaticamente irá buscar os dados desejados em uma de suas duas cópias (KUNTAMUKKALA, 2014). Estas cópias também serão utilizadas caso uma máquina escrava pare de enviar o *heartbeat* para a máquina mestre, um sinal enviado periodicamente para indicar seu funcionamento.

Cada tarefa solicitada pela máquina mestre é enviada para o *Cluster Manager*, que então a encaminha para a máquina escrava com o dado solicitado. A execução

da tarefa é feita na máquina escrava e então seu resultado é enviado de volta para o *Cluster Manager*, que o repassa para a máquina mestre.

Figura 4: Esquema de funcionamento do Spark

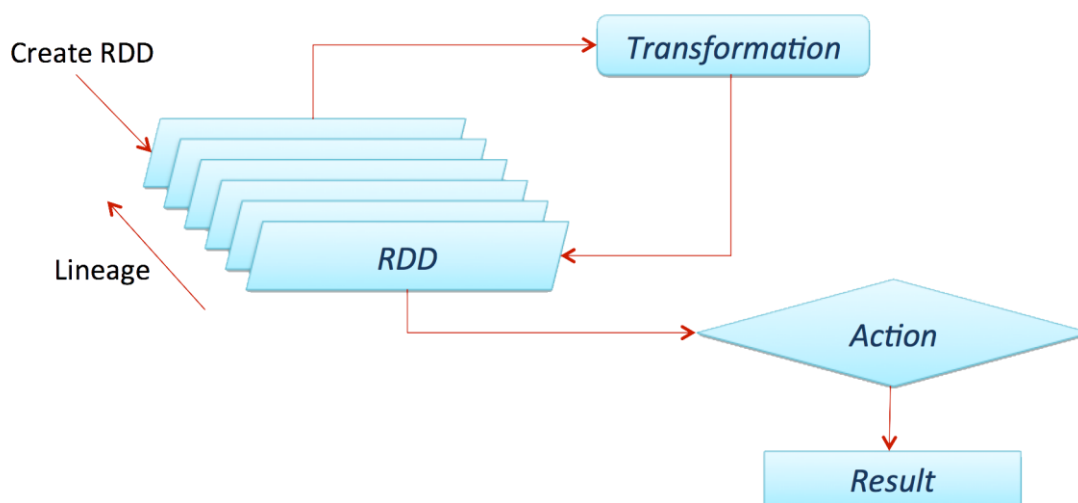


Fonte: <http://spark.apache.org/docs/1.3.0/cluster-overview.html>

Além disso, Spark introduz o uso de RDDs (Resilient Distributed Dataset), que funcionam como uma coleção imutável e distribuída de dados, particionados entre várias máquinas em um cluster. Opera com base em duas funções principais: transformações e ações.

Cada transformação aplicada em um RDD gera outro RDD, e então cada ação aplicada sobre esse novo RDD irá gerar um resultado. Porém, por implementar avaliação preguiçosa, as transformações são executadas apenas quando uma ação é aplicada sobre o RDD gerado por ela.

Figura 5: Esquema de funcionamento de RDDs no Spark

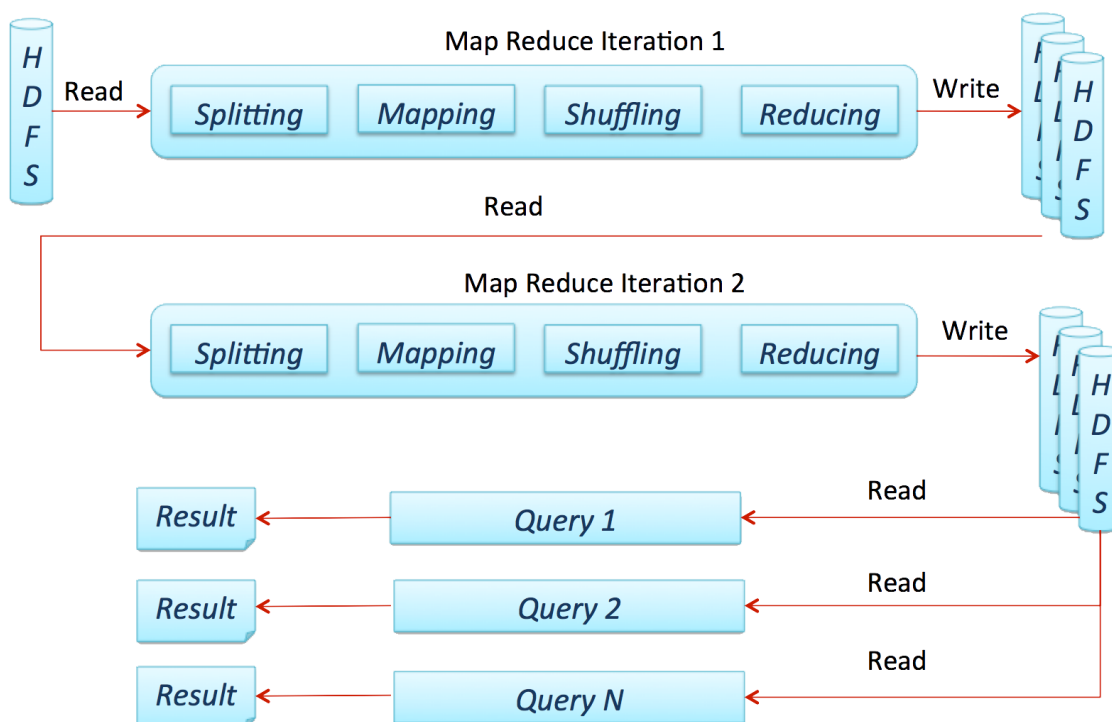


Fonte: <https://dzone.com/refcardz/apache-spark>

O uso de RDDs permite que Spark faça a maior parte do processamento das informações na memória principal das máquinas, pois elas são de fato executadas apenas quando o resultado final é solicitado.

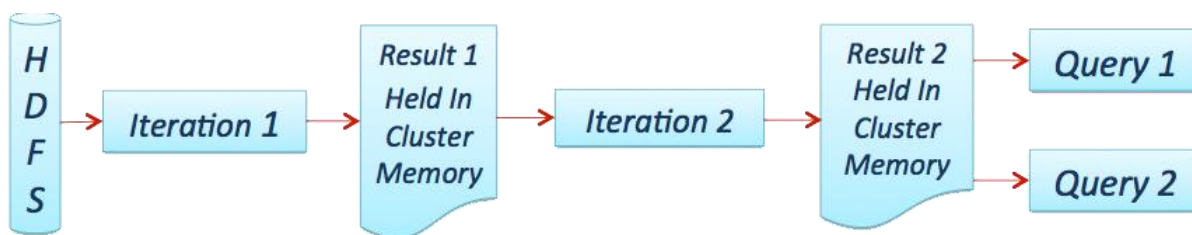
As figuras 6 e 7 mostram a comparação entre o MapReduce realizado no Hadoop e realizado no Spark.

Figura 6: MapReduce realizado no Hadoop



Fonte: <https://dzone.com/refcardz/apache-spark>

Figura 7: MapReduce realizado no Spark



Fonte: <https://dzone.com/refcardz/apache-spark>

Com Hadoop o resultado de cada iteração do MapReduce precisa ser gravado em disco, enquanto no Spark o resultado de cada iteração é diretamente utilizado como entrada para a próxima iteração, sem deixar a memória principal das máquinas.

5.3 Amazon Web Services

Amazon Web Services (AWS) é uma suíte de serviços para computação em nuvem oferecidos pela Amazon.com. Os serviços são oferecidos através de 33 zonas dentro de 12 diferentes regiões geográficas no mundo (AMAZON WEB SERVICES, 2016).

De acordo com Leong (2015), é estimado que os consumidores da AWS estão implementando 10 vezes mais servidores do que as próximas 14 empresas mais populares no segmento combinadas. Consumidores incluem NASA, Pinterest, Netflix e CIA (Agência Central de Inteligência do governo dos Estados Unidos).

Os serviços mais difundidos da empresa são o Amazon Elastic Compute Cloud (Amazon EC2) e Amazon Simple Storage Service (Amazon S3), ambos combinados no serviço Amazon Elastic MapReduce (Amazon EMR), feito especialmente para servir aplicações de big data.

O Amazon EC2 são servidores web com uma interface intuitiva e simples, o que permite sua configuração e uso com extrema facilidade, especialmente se comparados com a configuração e uso de servidores privados.

O Amazon S3 é um serviço para armazenamento de objetos na nuvem, ou seja, é um servidor otimizado para armazenamento de dados. Assim como o EC2, também possui uma interface intuitiva e simples.

Por fim, o Amazon EMR une estes dois serviços, integrados com frameworks para processamento de big data, como Hadoop e Spark, para oferecer uma solução rápida e econômica para a criação e gerenciamento de clusters de big data.

6 MÉTODO DE PESQUISA

Considerando que os sensores já estarão construídos e conectados à internet, a primeira fase deste projeto será analisar a quantidade de dados que será gerada pelos sensores e então, através da análise das principais tecnologias para Big Data, como Cassandra (THE APACHE SOFTWARE FOUNDATION, 2016A) e Storm (THE APACHE SOFTWARE FOUNDATION, 2016C), será planejado uma base de dados adequada para armazenar e manipular esta quantidade de dados.

A segunda fase do projeto será planejar e desenvolver o back-end do aplicativo que será responsável por interceptar estes dados no gateway e agrupá-los na base de dados. Para essa fase do projeto será necessário o estudo da linguagem Node.js (NODE.JS FOUNDATION, 2016).

Na terceira fase do projeto, será desenvolvido o aplicativo web para analisar essa base de dados e inferir informações úteis sobre a movimentação dos dispositivos conectados à rede wi-fi do prédio, como, por exemplo, a última posição conhecida de cada dispositivo. Nesta fase do projeto serão utilizadas ferramentas como Bootstrap (BOOTSTRAP, 2016).

A quarta e última fase do projeto terá como foco testes para garantir a integridade do aplicativo, principalmente em relação à segurança dos dados armazenados (CLOUD SECURITY ALLIANCE, 2015).

7 CRONOGRAMA

O projeto será dividido em sete atividades:

- a) Revisão Bibliográfica;
- b) Estudo das Tecnologias de Big Data;
- c) Implementação da Base de Dados;
- d) Estudo das Tecnologias Web;
- e) Implementação do Aplicativo;
- f) Ajustes do Aplicativo;
- g) Documentação.

Tabela 1: Cronograma de Atividades

Atividade	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out
A	X	X	X					
B		X	X					
C			X	X	X			
D				X	X			
E					X	X	X	X
F							X	X
G							X	X

Fonte: elaborado pelo autor

8 REFERÊNCIAS

AMAZON WEB SERVICES. AWS Global Infrastructure, 2016. Disponível em: <<https://aws.amazon.com/about-aws/global-infrastructure/>>. Acesso em: 10 Maio 2016.

AMAZON WEB SERVICES. Big Data Technology Fundamentals. **Amazon Web Services - Training and Certification**, 2016. Disponível em: <<http://aws.amazon.com/training/course-descriptions/bigdata-fundamentals/>>. Acesso em: 10 Maio 2016.

ASHTON, K. That 'Internet of Things' Thing. **RFID Journal**, 2009. Disponível em: <<http://www.rfidjournal.com/articles/view?4986>>. Acesso em: 09 Março 2016.

BLECKY. SubPos Positioning System. **Hackaday.io**, 2016. Disponível em: <<https://hackaday.io/project/4872-subpos-positioning-system>>. Acesso em: 21 Março 2016.

BOOTSTRAP. Bootstrap, 2016. Disponível em: <<http://getbootstrap.com/>>. Acesso em: 10 Março 2016.

CHEN, G.; KOTZ, D. A Survey of Context-Aware Mobile Computing Research. **Dartmouth Computer Science Technical Report TR2000-381**, Dartmouth College, 2000. Disponível em: <<http://www.cs.dartmouth.edu/reports/TR2000-381.pdf>>. Acesso em: 22 Março 2016.

CLOUD SECURITY ALLIANCE. Security Guidance for Early Adopters of the Internet of Things (IoT), 2015. Disponível em: <https://downloads.cloudsecurityalliance.org/whitepapers/Security_Guidance_for_Early_Adopters_of_the_Internet_of_Things.pdf>. Acesso em: 18 Março 2016.

DJUKNIC, G. M.; RICHTON, R. E. Geolocation and Assisted GPS. **IEEE Computer**, Bell Laboratories - Lucent Technologies, 2001. 123-125. Disponível em: <http://www.cs.columbia.edu/~drexel/CandExam/Geolocation_assistedGPS.pdf>. Acesso em: 25 Março 2016.

ESPOSITO, J. **D'Zones 2015 Guide to Big Data, Business Intelligence, and Analytics**, 2015. Disponível em: <<https://dzone.com/storage/assets/332483-dzone-guidetobigdata-2015.pdf>>. Acesso em: 22 Março 2016.

GARTNER. Hype Cycle for Emerging Technologies, 2015. Disponível em: <<http://www.gartner.com/smarterwithgartner/whats-new-in-gartners-hype-cycle-for-emerging-technologies-2015/>>. Acesso em: 01 Maio 2016.

GUBBI, J. et al. Internet of Things (IoT): A vision, architectural elements, and future directions. **Future Generation Computer Systems** **29**, 2013. 1645-1660. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167739X13000241>>. Acesso em: 02 Maio 2016.

KONSEK, H. IoT Gateways and Architecture. **D'Zones 2015 Guide to The Internet of Things**, 2015. Disponível em: <<https://dzone.com/storage/assets/162677-dzone-2015-iot-2.pdf>>. Acesso em: 15 Março 2016.

KUNTAMUKKALA, A. DZone - Apache Spark - An Engine for Large-Scale Data Processing, 2014. Disponível em: <<https://dzone.com/refcardz/apache-spark>>. Acesso em: 10 Maio 2016.

LEONG, L.; TOOMBS, D.; GILL, B. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide, 2015. Disponível em: <<https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519>>. Acesso em: 10 Maio 2016.

NODE.JS FOUNDATION. Node.js, 2016. Disponível em: <<https://nodejs.org/>>. Acesso em: 17 Março 2016.

SMITH, T. Executive Insights on The Internet of Things. **D'Zones 2015 Guide to The Internet of Things**, 2015. Disponível em: <<https://dzone.com/storage/assets/162677-dzone-2015-iot-2.pdf>>. Acesso em: 15 Março 2016.

THE APACHE SOFTWARE FOUNDATION. Cassandra, 2016A. Disponível em: <<http://cassandra.apache.org/>>. Acesso em: 24 Março 2016.

THE APACHE SOFTWARE FOUNDATION. Spark, 2016B. Disponível em: <<http://spark.apache.org/>>.

THE APACHE SOFTWARE FOUNDATION. Storm, 2016C. Disponível em: <<http://storm.apache.org/>>. Acesso em: 24 Março 2016.

WEISER, M. The Computer for the 21st Century. **Scientific American Ubicomp**, 1991. Disponível em: <<http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>>. Acesso em: 12 Maio 2016.

WEISER, M.; GOLD, R.; BROWN, S. J. The origins of ubiquitous computing research at PARC in the late 1980s. **IBM Systems Journal**, 38, n. 4, 1999. Disponível em: <<http://www.cs.cmu.edu/~jasonh/courses/ubicomp-sp2007/papers/03-weiser-origins.pdf>>. Acesso em: 03 Maio 2016.

ZEIMPEKIS, V.; GIAGLIS, G. M.; LEKAKOS, G. A Taxonomy of Indoor and Outdoor Positioning. **Newsletter ACM SIGecom Exchanges**, Athens University of Economics and Business - Department of Management Science & Technology, Volume 3, n. Issue 4, 2003. 19-27. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.2253&rep=rep1&type=pdf>>. Acesso em: 22 Março 2016.