

Marlon Diego Casagrande França

Desempenho na comunicação de dados entre cliente WEB e API WEB

Araquari – SC

Julho de 2017

Marlon Diego Casagrande França

Desempenho na comunicação de dados entre cliente WEB e API WEB

Trabalho de conclusão de curso apresentado
como requisito parcial para a obtenção do
grau de bacharel em Sistemas de Informação
do Instituto Federal Catarinense.

Instituto Federal Catarinense – IFC

Câmpus Araquari

Bacharelado em Sistemas de Informação

Orientador: Prof. Dr. Eduardo da Silva

Coorientador: Prof. Dr. Eduardo da Silva

Araquari – SC

Julho de 2017

Marlon Diego Casagrande França

Desempenho na comunicação de dados
entre cliente WEB e API WEB/ Marlon Diego Casagrande França. – Araquari – SC,
Julho de 2017-

53 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Eduardo da Silva

Monografia (Graduação) – Instituto Federal Catarinense – IFC
Câmpus Araquari

Bacharelado em Sistemas de Informação, Julho de 2017.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Instituto
Federal Catarinense. III. Câmpus Araquari. IV. Título

Marlon Diego Casagrande França

Desempenho na comunicação de dados entre cliente WEB e API WEB

Trabalho de conclusão de curso apresentado
como requisito parcial para a obtenção do
grau de bacharel em Sistemas de Informação
do Instituto Federal Catarinense.

Trabalho aprovado. Araquari – SC, 24 de novembro de 2016:

Prof. Dr. Eduardo da Silva
Orientador

Professor
Convidado 1

Professor
Convidado 2

Araquari – SC
Julho de 2017

*Dedicatória do trabalho de conclusão que deve
ser algo breve e conciso.*

Agradecimentos

Página com os agradecimentos do autor à pessoas importantes para a realização do trabalho.

*“Colocar alguma frase importante que,
tenha motivado o autor no decorrer do desenvolvimento
do trabalho.”
(Referência de local da frase)*

Resumo

Segundo a [ABNT \(2003, 3.1-3.2\)](#), o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (...) As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Palavras-chave: latex. abntex. editoração de texto.

Abstract

This is the english abstract.

Keywords: latex. abntex. text editoration.

Lista de ilustrações

Figura 1 – Características dos mecanismos de comunicação	19
Figura 2 – RPC fluxo de dados. (refazer)	23
Figura 3 – SOAP fluxo de dados. (refazer)	25
Figura 4 – Modelo de maturidade de Richardson. (refazer)	29
Figura 5 – HTTP Caching. (refazer)	33
Figura 6 – Modelagem WMS	35
Figura 7 – Diagrama de dimensões do armazém (REFAZER)	36
Figura 8 – UML sequence REST (REFAZER)	41
Figura 9 – UML sequence GraphQL (REFAZER)	42
Figura 10 – Arquitetura das APIs e diferentes pontos de medidas (REFAZER) . . .	45
Figura 11 – Comparação do tempo de resposta	47
Figura 12 – Comparação do tamanho de resposta	47
Figura 13 – Comparação do tempo de resposta	48
Figura 14 – Comparação do tamanho de resposta	49

Lista de tabelas

Tabela 1 – Configuração do Ambiente	40
Tabela 2 – Servidor REST	40
Tabela 3 – Fluxo de dados para responder as questões	46

Lista de códigos

Código 1 – Controller para carregar um item	38
Código 2 – Controller para carregar um item	38
Código 3 – Consulta de itens	41
Código 4 – Listagem dos itens (Validar)	41

Lista de abreviaturas e siglas

API	Application Program Interface
REST	Representational State Transfer
HTTP	Hypertext Transfer Protocol
HTTPS	Hyper Text Transfer Protocol Secure
ERP	Enterprise Resource Planning

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
\in	Pertence

Sumário

1	INTRODUÇÃO	16
2	FUNDAMENTOS	18
2.1	IPC - Interprocess Communication	18
2.2	Cliente/Servidor	19
2.3	Aplicações Monolíticas	19
2.4	SOA - Service Oriented Architecture	20
3	PROTOCOLOS DE COMUNICAÇÃO	22
3.1	Remote Procedure Call	22
3.1.1	Limitações	23
3.2	Simple Object Access Protocol	24
3.2.1	Limitações	25
3.3	REpresentational State Transfer	26
3.3.1	Restrições	27
3.3.2	Modelo de Maturidade de Richardson	29
3.3.3	Limitações	30
3.4	Arquiteturas baseadas em JSON/Graphs	31
3.4.1	Operações	31
3.4.2	Schemas e Tipos	32
3.4.3	Autorização	33
3.4.4	Caching	33
3.4.5	Limitações	34
4	ESTUDO DE CASO	35
4.1	Hipóteses	36
4.2	Cenários	37
4.2.1	Ferramentas utilizadas	37
4.2.2	Ambiente	39
4.2.3	Detalhes implementação REST	39
4.2.4	Detalhes implementação GraphQL	41
4.3	Métricas	43
5	RESULTADOS	46
5.1	Questão 1	46
5.1.1	Utilização da CPU	46

5.1.2	Consumo de memória	46
5.1.3	Tempo de resposta	46
5.1.4	Tamanho da resposta	47
5.2	Questão 2	48
5.2.1	Utilização da CPU	48
5.2.2	Consumo de memória	48
5.2.3	Tempo de resposta	48
5.2.4	Tamanho da resposta	49
6	CONCLUSÃO	50
	REFERÊNCIAS	51

1 Introdução

Em um mundo cada vez mais interconectado, nós como usuários constantemente exigimos maior disponibilidade de informações através da Web. Essas informações não devem apenas estarem acessíveis, como também é crucial que seja rápido o tempo para acessá-las. Com a expansão da internet para dispositivos dos mais diversos tipos, tais como *smartphones*, *tables* e dispositivos de *IoT*, questões como tempo de carregamento e consumo de banda vem se tornando fatores cada mais mais debatidos.

O uso de dispositivos móveis para acessar a internet cresceu 63% somente em 2016. É o que aponta o relatório de Previsão Global de Tráfego de Dados Móveis, elaborada pela Cisco. O tráfego de dados móveis passou de 4.4 exabytes ¹ por mês em 2015, para uma média de 7.2 exabytes mensais em 2016. Esse aumento foi resultado de uma adição de cerca 429 milhões de novos dispositivos móveis a rede, sendo *smartphones* os responsáveis pela maior parte deste crescimento (VNI, 2017).

Conforme Work (2011), usuários tendem a dar mais importância à velocidade em que recebem a informação do que a estética que ela é apresentada. O tempo de carregamento é um fator decisivo para permanência em uma página Web, uma vez que a maioria dos usuários estão dispostos a aguardar de 6 à 10 segundos antes de abandonar a página. Cada segundo de *delay* pode resultar em uma redução de até 7% nas taxas de conversão, o que para um *e-commerce* que fatura R\$ 100.000,00 por mês, este segundo pode potencialmente custar R\$ 2.5 milhões em vendas não efetuadas por ano.

Nas últimas décadas, houve uma mudança para um modelo de computação chamado cliente/servidor, que aborda as falhas da computação centralizada. Claramente, o modelo de computação centralizado permanece válido em certos ambientes de negócios, no entanto, apesar de muitos benefícios, a computação centralizada é reconhecida como tendo promovido uma cultura de gerenciamento de informações que não conseguiu atender as necessidades de seus clientes.

Em 2010, ocorreu um grande avanço no número de APIs públicas impulsionado pela transição no modelo de comunicação entre aplicações distribuídas, onde estas passaram a utilizar amplamente o protocolo HTTP e o modelo cliente-servidor para a troca de informações na World Wide Web (MASO, 2016). A adoção do REST (REpresentational State Transfer) como o método predominante para construir APIs públicas tem ofuscado qualquer outra tecnologia ou abordagem nos últimos anos. Embora várias alternativas (principalmente SOAP) ainda estejam presentes no mercado, adeptos do modelo SOA para construção de aplicações tomaram uma posição definitiva contra eles e optaram por REST

¹ Um exabyte é equivalente a um bilhão de gigabytes e mil petabytes.

como modelo de comunicação e JSON como seu formato de mensagem (LENSMAR, 2013).

Segundo PALIARI (2012) REST é cada vez mais usado como alternativa ao “já antigo” SOAP em que a principal crítica a este é a burocracia, algo que o REST possui em uma escala muito menor. REST é baseado no *design* do protocolo HTTP, que já possui diversos mecanismos embutidos para representar recursos como código de *status*, representação de tipos de conteúdo, cabeçalhos, etc. O principal nesta arquitetura são as URLs do sistema e os *resources*², aproveitando os métodos HTTP para se comunicar.

Foram propostas múltiplas soluções para aumentar a eficiência na busca de dados, alguns em relação aos formatos de consulta e resposta das requisições, enquanto outros estão otimizando o número de solicitações na rede. Uma tendência recente envolve *frameworks* consultas declarativos de dados, onde as aplicações clientes especificam quais dados precisam em um campo por campo base, em vez de buscar tudo a partir de um local específico definido por um URL. Os *frameworks* então otimizaram a comunicação com os servidores para obter os dados de forma eficiente.

Esta é a proposta do GraphQL. Construído pelos desenvolvedores do Facebook para atender as necessidades internas da rede social em 2012¹, o GraphQL foi lançado ao público em geral em 2015, e já vem ganhando diversos adeptos. Com a promessa de mitigar alguns problemas crônicos do *design* do REST, como versionamento de APIs, múltiplas viagens de ida e volta e excesso de dados trafegados na rede, a abordagem do Facebook já vem sendo usada por diversas empresas.

REST é de fato, o modelo mais utilizado para comunicação entre cliente e servidor nas aplicações atuais. Este trabalho foca em identificar as diferenças em termos de tempo de carregamento, quantidade de dados trafegados e consumo de recursos entre aplicações REST e o ainda pouco conhecido GraphQL. Ambas as tecnologias são frequentemente usadas e discutidas como soluções para servir dados entre clientes Web e servidores, com cada vez mais desenvolvedores se interessando pela nova abordagem GraphQL

² resource é um recurso, entidade

2 Fundamentos

2.1 IPC - Interprocess Communication

Nem sempre um programa sequencial é a melhor solução para um determinado problema. Muitas vezes, as implementações são estruturadas na forma de várias tarefas interdependentes que cooperam entre si para atingir os objetivos da aplicação, como por exemplo em um navegador Web ([HTTP://WIKI.INF.UFPR.BR/MAZIERO/LIB/EXE/FETCH.PHP?MEDIA=LIVRO.PDF](http://wiki.inf.ufpr.br/maziro/lib/exe/fetch.php?media=livro.pdf), 2013).

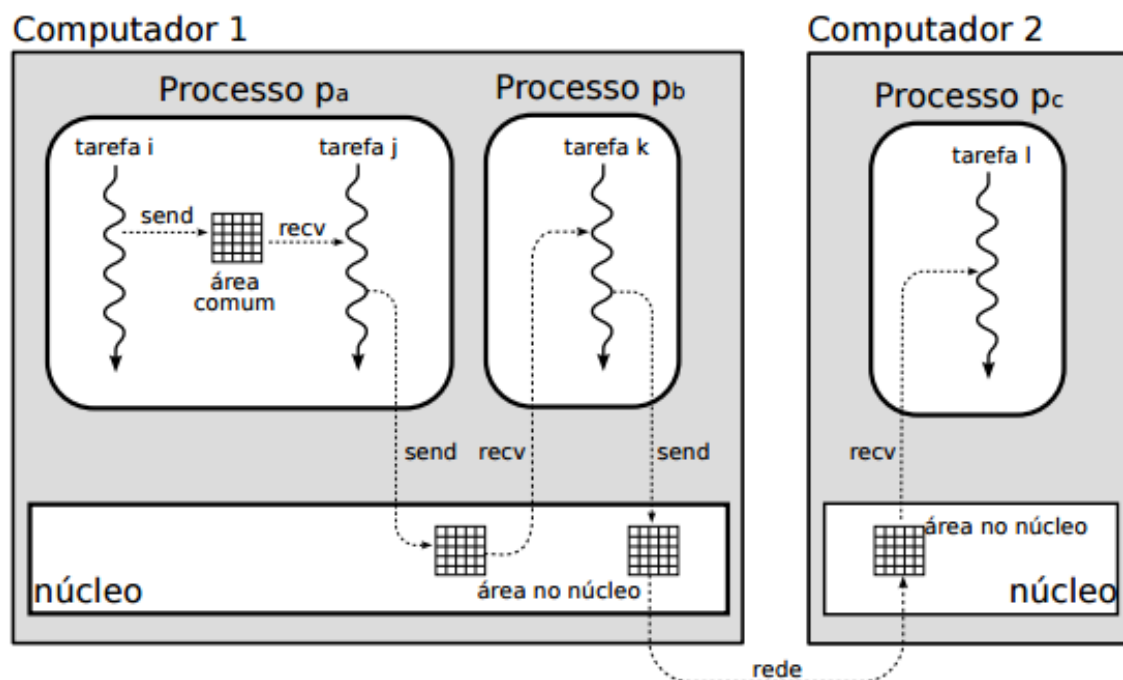
Em um sistema multitarefa, seja com um único processador ou com mais de um processador ou ainda com vários núcleos por processador, os processos alternam sua execução segundo critérios de escalonamento estabelecidos pelo sistema operacional. Processos de aplicações concorrentes compartilham diversos recursos do sistema, como arquivos, registros, dispositivos de Entrada/Saída e áreas da memória (KNOW, 2008).

Os mecanismos que garantem a comunicação entre processos concorrentes e os acessos aos recursos compartilhados são chamados *interprocess communication*. Algumas formas de IPC facilita a divisão de trabalho entre diversos processos especialistas. Outras formas facilitam a divisão de trabalho entre computadores dentro de uma rede.

Normalmente, os aplicativos podem usar IPC categorizados como clientes ou servidores. Um cliente é um aplicativo ou um processo que solicita um serviço de alguma outra aplicação ou processo. Um servidor é um aplicativo ou um processo que responde a uma solicitação de cliente. Muitas aplicações agem como um cliente e um servidor, dependendo da situação. (MICROSOFT, 2008)

A figura abaixo mostra como ocorre a comunicação entre tarefas de um processo, entre processos de um computador, e entre computadores na mesma rede.

Figura 1 – Características dos mecanismos de comunicação



Fonte – Características dos mecanismos de comunicação

<http://wiki.inf.ufpr.br/maziero/lib/exe/fetch.php?media=so:so-livro.pdf> (2013)

2.2 Cliente/Servidor

Em uma arquitetura cliente/servidor de duas camadas, *Remote Procedure Calls*(RPC) ou *Structured Query Language*(SQL) são normalmente utilizadas para a comunicação entre cliente e servidor. O servidor geralmente tem suporte para *Stored Procedures* e *Triggers*. Isso significa que o servidor pode ser programado para implementar regras de negócios que são mais adequadas para serem executadas no servidor ao invés de no cliente, resultando em um sistemas muito mais eficiente. (SCHUSSEL, 2001)

Desde 1992, fornecedores de software desenvolvem e trazem ao mercado muitas ferramentas para simplificar o desenvolvimento de aplicativos para a arquitetura cliente/servidor de 2 camadas. As mais conhecidas são: Microsoft Visual Basic, Delphi da Borland e PowerBuilder da Sybase. Essas ferramentas combinadas com milhões de desenvolvedores que sabem usá-las, significa que a abordagem de duas camadas de cliente/servidor é uma solução econômica para certas classes de problemas.

2.3 Aplicações Monolíticas

Em engenharia de software, uma aplicação monolítica descreve uma única aplicação de software em camadas no qual a interface de usuário e código de acesso aos dados são

combinados em um único programa a partir de uma única plataforma.

Uma aplicação monolítica é autônoma e independente de outras aplicações de computação. A filosofia do projeto consiste em um aplicativo que não é responsável apenas por uma determinada tarefa, mas que também pode executar todos os passos necessários para completar uma determinada função.

A arquitetura monolítica é um padrão comumente usado para o desenvolvimento de aplicações corporativas. Esse padrão funciona razoavelmente bem para pequenas aplicações, pois o desenvolvimento, testes e implantação de pequenas aplicações monolíticas é relativamente simples. No entanto, para aplicações grandes e complexas, a arquitetura monolítica torna-se um obstáculo ao desenvolvimento e implantação, dificulta a utilização de uma entrega contínua, além de limitar a adoção de novas tecnologias. Para grandes aplicações, faz mais sentido usar uma arquitetura de microservices, que divide a aplicação em um conjunto de serviços.

2.4 SOA - Service Oriented Architecture

Este documento e seu código-fonte são exemplos de referência de uso da classe `abntex2` e do pacote `abntex2cite`. O documento exemplifica a elaboração de trabalho acadêmico produzido conforme a ABNT NBR 14724:2011 *Informação e documentação - Trabalhos acadêmicos - Apresentação*.

O modelo apresentado é baseado no “Modelo Canônico” criado pela equipe do projeto `abnTEX2` e implementa os requisitos das normas da ABNT. Uma lista completa das normas observadas pelo `abnTEX2` é apresentada em [Araujo \(2015a\)](#). Aqui, está apresentada a forma que o modelo será utilizado no curso de Bacharelado em Sistemas de Informação do IFC - Araquari.

Este documento deve ser utilizado como complemento dos manuais do `abnTEX2` ([ARAÚJO, 2015a](#); [ARAÚJO, 2015b](#); [ARAÚJO, 2015c](#)) e da classe `memoir` ([WILSON; MADSEN, 2010](#)).

Na introdução o autor coloca o problema ou a indagação que o levou a escrever o texto. A introdução nos dá, então, uma idéia do assunto tratado. Além disso, nela o autor coloca também o ponto de vista ou o ângulo sob o qual ele vai abordar o assunto e, às vezes, o método, ou seja, o caminho que vai seguir (se vai apresentar casos para chegar a uma generalização, ou se vai partir de um princípio geral e deduzir suas consequências).

Também na introdução, o tema é apresentado e esclarecido aos leitores as indicações de leitura do trabalho. Deve-se utilizar o projeto do TCC para colocar na introdução o objetivo principal, os objetivos específicos, o problema e a hipótese.

A respeito de materiais e métodos, pode-se falar sobre a infra-estrutura necessária

para o trabalho, incluindo servidores, estações, equipamentos de rede, *softwares* com suas respectivas versões e tudo o mais que for necessário.

A introdução termina com a apresentação dos demais capítulos do trabalho. O capítulo 2 contém o referencial teórico deste trabalho. No capítulo 3 é apresentado o desenvolvimento, o cenário, os testes e a discussão dos resultados. Por fim, temos a conclusão, onde discutiremos os resultados, as dificuldades encontradas e faremos sugestões de trabalhos futuros.

3 Protocolos de Comunicação

Neste capítulo será apresentado alguns dos principais protocolos para a comunicação entre aplicações. Serão conduzidas breves abordagens dos protocolos RPC e SOAP, que hoje vem perdendo espaço no mercado, mas tiveram um papel muito importante para a evolução dos sistemas SOA. Em seguida, será mostrada uma descrição mais detalhada dos modelos de comunicação REST e GraphQL, uma vez que estes são os dois modelos a serem seguidos no estudo de caso.

Será apresentado as principais características desses protocolos, como também suas limitações. Nenhum dos protocolos podem ser descartado na hora de tomada de decisão sobre qual modelo utilizar para a comunicação entre clientes e serviços remotos ou APIs. Todos eles possuem limitações, porém cumprem bem o papel para que foram desenhados e é possível notar como os mesmos evoluíram a medida que novas demandas se tornarão necessária no desenvolvimento de software ao decorrer do tempo.

3.1 Remote Procedure Call

Remote Procedure Call (RPC), ou Chamada de Procedimento Remoto, é um mecanismo onde uma aplicação solicita o serviço de uma outra aplicação que se encontra em um outro endereço. Um RPC exige que uma aplicação envie uma ou mais mensagens para outra aplicação a fim de invocar um procedimento da outra aplicação. A aplicação destinatária responde enviando uma ou mais mensagens de volta para seu remetente. O termo "RPC" identifica todo o processo de comunicação entre as aplicações ([MERRICK; ALLEN; LAPP, 2006](#)).

O RPC é similar ao modelo de chamadas de procedimentos locais. Nas chamadas locais a aplicação remetente envia argumentos para um procedimento em uma localização bem definida e transfere o controle para o procedimento, posteriormente reassumindo o controle. Logo após, os resultados do procedimento são extraídos de uma localização também bem definida, e a aplicação remetente continua sua execução ([SUN MICROSYSTEMS, 1995](#)).

Ainda segunda a especificação do RPC ([SUN MICROSYSTEMS, 1995](#))(??), há alguns aspectos importantes onde as chamadas remotas se diferem das chamadas locais:

- Tratamento de erros: Falhas no servidor remoto ou da rede devem ser tratadas quando usando uma RPC.
- Variáveis globais e efeitos colaterais: Uma vez que o servidor não tem acesso ao

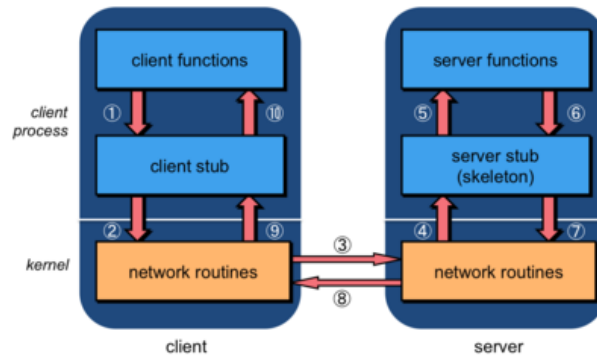


Figura 2 – RPC fluxo de dados. (refazer)

espaço de endereço do cliente, argumentos ocultos não podem ser passados como variáveis globais ou retornados como efeitos colaterais.

- Desempenho: Chamadas remotas geralmente operam uma ou mais ordens mais lentos do que chamadas de procedimentos locais
- Autenticação: Uma vez que as chamadas remotas pode ser transportadas através de redes inseguras, autenticação pode ser necessário. Autenticação previne uma entidade de se passar por outra e executar chamadas indevidas.

3.1.1 Limitações

Diferentes implementações de RPC são geralmente mutualmente incompatíveis. Assim sendo, o uso de uma implementação específica provavelmente resultará na dependência da implementação do fornecedor. Esse tipo de incompatibilidade entre diferentes implementações implica em um diverso número de funcionalidades, suportando inúmeros protocolos de rede e diferentes sistemas de arquivos.

[Kukreja \(2014\)](#) aponta diversas limitações técnicas e de desempenho encontradas geralmente em implementações de RPC's. Dentre estas limitações, se destacam:

- Organização dos parâmetros *Encontrar outro*: Para organizar os parâmetros, o cliente precisa conhecer o número e o tipo dos parâmetros exatos. Como não existe uma padrão bem definido, a estruturação dos parâmetros podem variar entre diferentes implementações, o cliente não tem facilidade de constatar o número de parâmetros que existem ou os tipos desses parâmetros.
- Falta de paralelismo: Com RPC, em um determinado momento, ou o servidor ou o cliente está ativo. Como cliente e servidor são co-rotinas, a prática de paralelismo presente em outros modelos de comunicação, não é possível de ser implementada.
- Falta de transmissão: Em banco de dados, um cliente solicita a um servidor que execute a operação para procurar dados no banco de dados que atendam a algumas

condições. Usando o modelo do RPC, um servidor deve esperar até que todas as informações sejam encontradas antes de responder. Caso a operação de busca em todos os dados seja demorada, o cliente permanece ocioso por algum tempo, esperando que a última tupla seja pesquisada.

3.2 Simple Object Access Protocol

Simple Object Access Protocol (SOAP) é um protocolo para troca de mensagens em ambientes descentralizados e distribuídos. [Box et al. \(2000\)](#) define o SOAP como um protocolo baseado em XML que consiste em três partes: um envelope que defina um *framework* para descrever o que há mensagem e como processar seu conteúdo, um conjunto de regras de codificação para expressar instâncias de tipos de dados definidos por uma aplicação, e uma convenção para representar chamadas remotas e suas respostas.

Para [Mueller \(2013\)](#), o SOAP é um protocolo de acesso a serviços baseados em padrões que existe por um tempo e utiliza-se de todos os benefícios do uso a longo prazo. Originalmente desenvolvido pela Microsoft, ele pode não ser tão simples quanto o acrônimo sugere.

[Rouse \(2014\)](#) argumenta que o SOAP pode ser considerado parecido com as Chamadas de Procedimento Remoto (RPC), porém eliminando algumas das complexidades frequentemente encontradas nas implementações de RPC's. Utilizando o SOAP, é possível chamar serviços de outras aplicações, que estão sendo executadas em qualquer *hardware*, independente do sistemas operacional ou linguagem de programação.

Embora a especificação do SOAP tenha evoluído para longe da necessidade de acessar objetos, como ocorre com as Chamadas de Procedimentos Remoto, existe ainda convenções para o encapsulamento e envio de chamadas RPC utilizando uma representação uniforme de chamadas e repostas. Definindo um padrão para o mapeamento das chamadas RPC para SOAP torna-se possível para infraestrutura traduzir as invocações de serviços em mensagens com o protocolo SOAP em tempo de execução, sem a necessidade de redesenhar todo o serviço Web presente na plataforma ([SKONNARD, 2003](#)).

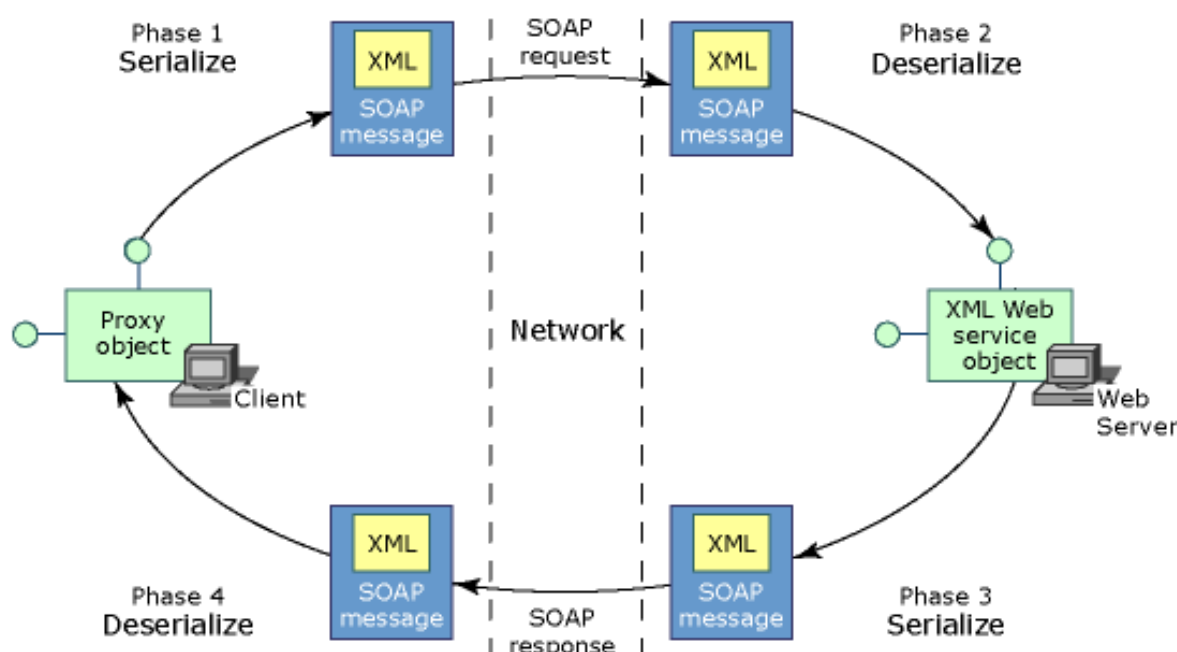


Figura 3 – SOAP fluxo de dados. (refazer)

Leopoldo (2007) assinala que SOAP é um protocolo que define uma gramática XML especializada, que padroniza o formato das estruturas das mensagens. As mensagens são, por outro lado, o método fundamental de troca de informações entre os Web Services e os seus consumidores. Ao utilizar XML para codificar mensagens o SOAP nos dá alguns benefícios:

- Permite a comunicação entre sistemas protegidos por *firewalls*, sem precisar abrir portas adicionais e possivelmente não seguras
- XML pode ser facilmente lido por humanos, sendo portanto, mais fácil de entender e eliminar erros.
- Os dados do SOAP são estruturados usando XML. Portanto, as mensagens podem ser compreendidas por quase todas as plataformas de hardware, sistemas operacionais e linguagens de programação.
- Pode ser usado, potencialmente, em combinação com vários protocolos de transporte de dados, como HTTP, SMTP e FTP.
- O SOAP mapeia satisfatoriamente para o padrão de solicitação / resposta do HTTP.

3.2.1 Limitações

Kohlhoff (2003) explica que tecnologias de metadados abertas, como o XML, podem fornecer um grande ganho de usabilidade, mas o sucesso dessas tecnologias exige que seu

uso não degrada o desempenho de forma irracional. O XML é extremamente robusto, no entanto, seu uso pode afetar negativamente o desempenho do SOAP nas seguintes áreas:

Velocidade de codificação e decodificação. A conversão de dados de binário para ASCII e vice-versa é o principal custo de desempenho do XML. O uso de um protocolo baseado em texto também impede a aplicação de otimizações disponíveis para protocolos binários quando a comunicação ocorre entre sistemas homogêneos.

Tamanho da mensagem. Para XML, um fator de expansão de 6-8 vezes em relação aos dados binários originais não é incomum. Mediu a expansão em 4 a 10 vezes e descobriu que o tamanho da representação de dados do SOAP é tipicamente cerca de 10 vezes o tamanho da representação binária equivalente. Este tamanho substancialmente maior pode resultar em maiores custos de transmissão de rede e latência aumentada.

Como é possível notar, as maiores limitações encontradas na utilização do SOAP como modelo de comunicação estão relacionadas a estruturação usando XML. Embora esse formato padrão de mensagens tenha trazido vantagens em relação a tecnologias anteriores ao SOAP, ele também trouxe uma série de barreiras conforme sua popularidade foi crescendo. *A necessidade buscar alternativas para superar esses problemas foi crescente, e o mercado encontrou novos formatos foram aparecendo, tendo como destaque o JSON*

3.3 REpresentational State Transfer

Representational State Transfer (REST) é um *design* de arquitetura baseado em um conjunto de princípios que descrevem como os recursos em rede são definidos e abordados. Estes princípios foram descritos pela primeira vez em 2000 por Roy Fielding como parte de sua dissertação de doutorado ([BARRY, 2003](#)).

[Lensmar \(2013\)](#) assinala que a adoção do REST (REpresentational State Transfer) como o método predominante para construir APIs públicas tem ofuscado qualquer outra tecnologia ou abordagem nos últimos anos. Embora várias alternativas (principalmente SOAP) ainda estejam presentes no mercado, adeptos do modelo SOA, descrito na seção 2.4, para construção de aplicações tomaram uma posição definitiva contra eles e optaram por REST como sua abordagem e JSON como seu formato de mensagem.

Segundo [Battle \(2008\)](#), REST é um padrão de operações de recursos que emergiu como a principal alternativa ao já antigo SOAP para o design de serviços em aplicativos Web 2.0. Considerando que a abordagem tradicional baseada em SOAP para Web Services usa objetos remotos completos com invocação de método remoto e funcionalidade encapsulada, o REST trata apenas de estruturas de dados e da transferência de seu estado. A simplicidade do REST, juntamente com seu ajuste natural sobre o HTTP, contribuiu para o seu status como um método de escolha para aplicativos da Web 2.0 para expor seus dados.

3.3.1 Restrições

As restrições do REST são regras desenhadas para estabelecer as características distintas da arquitetura REST. Cada restrição é uma decisão de projeto pré-determinada que pode ter impactos positivos e negativos. A intenção é que os aspectos positivos de cada restrição equilibrem os negativos para produzir uma arquitetura geral que se assemelhe à Web.

Uma arquitetura que elimina uma restrição REST, geralmente é considerada como não mais conforme ao REST. Isso exige que as decisões sejam tomadas para entender os potenciais *trade-offs* quando se desviam deliberadamente da aplicação de restrições REST.

A primeira restrição estabelecida é a da arquitetura **cliente-servidor**, descrita na Seção 2.2. A separação de responsabilidades é o princípio por trás das restrições cliente-servidor. Ao separar as responsabilidades da interface do usuário com as responsabilidades de armazenamento de dados, melhoramos a portabilidade da interface do usuário em várias plataformas e melhoramos a escalabilidade, simplificando os componentes do servidor. Talvez o mais importante para a Web, no entanto, é que a separação permite que os componentes evoluam de forma independente, apoiando assim o requisito de escalabilidade da Internet de múltiplos domínios organizacionais ([FIELDING, 2000](#))

Em seguida, adicionamos uma restrição à interação cliente-servidor: a comunicação deve ser totalmente **stateless**, ou seja, independente de estado, de modo que cada solicitação do cliente para O servidor deve conter todas as informações necessárias para entender o pedido e não pode tirar proveito de nenhum contexto armazenado no servidor. O estado da sessão é, portanto, mantido inteiramente no cliente ([FIELDING, 2000](#)).

Essa restrição induz as propriedades de visibilidade, confiabilidade e escalabilidade. A visibilidade é melhorada porque um sistema de monitoramento não precisa olhar além de um único dado de solicitação para determinar a natureza completa da solicitação. A confiabilidade é melhorada porque facilita a tarefa de recuperação de falhas parciais. A escalabilidade é melhorada porque não ter que armazenar o estado entre solicitações permite que o componente do servidor rapidamente libere recursos e simplifica ainda mais a implementação porque o servidor não precisa gerenciar o uso de recursos em todos os pedidos.

A restrição de **cache** foi criada para melhorar a eficiência da rede. Ela exige que os dados dentro de uma resposta a uma solicitação sejam rotulados de forma implícita ou explícita como cacheáveis ou não armazenáveis em cache. Se uma resposta for armazenada em cache, um cache do cliente terá o direito de reutilizar esses dados de resposta para solicitações equivalentes posteriores.

A vantagem de adicionar restrições de cache é que eles têm o potencial de eliminar parcial ou totalmente algumas interações, melhorar a eficiência, escalabilidade e desempenho

percebido pelo usuário, reduzindo a latência média de uma série de interações. O *trade-off*, no entanto, é que um cache pode diminuir a confiabilidade se os dados obsoletos dentro do cache diferirem significativamente dos dados que teriam sido obtidos se a solicitação fosse enviada diretamente para o servidor.

A característica central que distingue a arquitetura REST de outros estilos baseados em rede é a ênfase em uma **interface uniforme** entre os componentes. Ao aplicar o princípio de engenharia de software de especialização/generalização para a interface do componente, a arquitetura geral do sistema é simplificada e a visibilidade das interações é melhorada. As implementações são dissociadas dos serviços que eles fornecem, o que incentiva a evolução independente. O trade-off, no entanto, é que uma interface uniforme degrada a eficiência, uma vez que a informação é transferida em uma forma padronizada e não específica para as necessidades de uma aplicação.

Para obter esta interface uniforme, são necessárias várias restrições arquitetônicas para orientar o comportamento dos componentes. REST é definido por quatro restrições de interface: identificação de recursos; Manipulação de recursos através de representações; Mensagens auto-descritivas; E, hipermídia como motor do estado da aplicação.

A fim de melhorar o comportamento dos requisitos de escalamento, o REST também implementa restrições de **sistema em camadas**. O sistema em camadas permite que uma arquitetura seja composta de camadas hierárquicas ao restringir o comportamento dos componentes, de modo que cada componente não possa "ver" além da camada imediata com a qual eles estão interagindo. Ao restringir o conhecimento do sistema a uma única camada, colocamos um limite na complexidade geral do sistema e promovemos a independência do componentes. As camadas podem ser usadas para encapsular serviços legados e para proteger novos serviços de clientes legados, simplificando componentes e movendo funcionalidades raramente usadas para um intermediário compartilhado. Os intermediários também podem ser usados para melhorar a escalabilidade do sistema ao permitir o balanceamento de carga de serviços em várias redes e processadores.

A principal desvantagem dos sistemas em camadas é que eles adicionam sobrecarga e latência ao processamento de dados, reduzindo o desempenho perceptível pelo usuário. Para um sistema baseado em rede que suporta restrições de cache, isso pode ser compensado pelos benefícios do armazenamento em cache compartilhado em intermediários.

A última restrição que o REST implementa é a restrição de **código por demanda**. O REST permite que a funcionalidade do cliente seja estendida baixando e executando o código na forma de *applets* ou *scripts*. Isso simplifica os clientes, reduzindo o número de recursos necessários para serem pré-implementados. Permitir que os recursos sejam baixados após a implantação, melhora a extensibilidade do sistema, no entanto, também reduz a visibilidade e, portanto, é apenas uma restrição opcional dentro do REST.

3.3.2 Modelo de Maturidade de Richardson

O Modelo de Maturidade de Richardson, é um modelo criado por Leonard Richardson que descreve os requisitos necessários para o desenvolvimento de uma API REST bem estruturada e compatível com as restrições definidas pela arquitetura REST. Quanto melhor a API adere as restrições citadas na sessão anterior, melhor será pontuada. O modelo de Richardson descreve 4 níveis (0-3), onde o nível 3 designa uma API verdadeiramente *RESTful*. Richardson usou três fatores para decidir a maturidade de uma API: URI (Uniform Resource Identifiers), métodos HTTP e HATEOAS (Hypermedia). Quanto mais uma API emprega essas tecnologias, mais madura ela é categorizada.

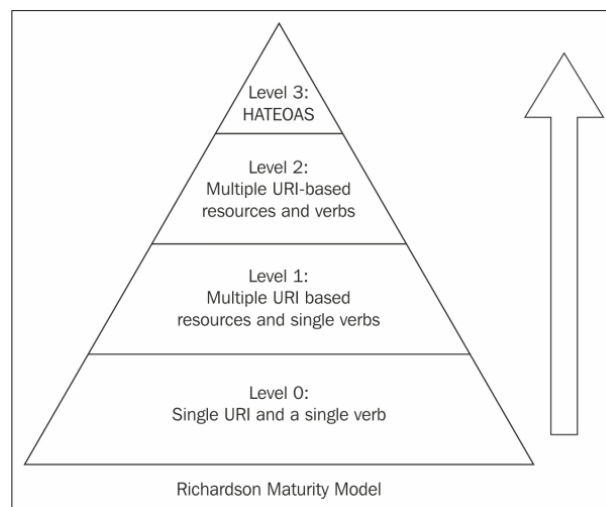


Figura 4 – Modelo de maturidade de Richardson. (refazer)

Autor....

- Nível 0: O ponto de partida para o modelo é usar HTTP como um sistema de transporte para interações remotas, mas sem usar nenhum dos mecanismos da web. Essencialmente, o que você está sendo feito é utilizar o HTTP como um mecanismo de tunelamento para seu próprio mecanismo de interação remota, geralmente com base em Chamadas de Procedimento Remotas.

É como se estivesse sendo chamado funções, porém através do uso do HTTP. Todos serviços são centralizados em um único *endpoint*, ou seja, todas as solicitações são feitas em uma única URI.

- Nível 1: Quando uma API faz a distinção entre recursos diferentes, pode se considerar que atingiu o nível 1. Esse nível usa vários URIs, onde cada URI é o ponto de entrada para um recurso específico. Ainda assim, esse nível usa apenas um único método, como o POST.

- Nível 2: Esse nível indica que a API deve usar as propriedades do protocolo para lidar com escalabilidade e falhas. Não é recomendado que se use um único método POST para todas as chamadas, mas faça uso do GET quando estiver solicitando recursos e use o método DELETE quando desejar excluir recursos. Além disso, o uso dos códigos de resposta do protocolo de aplicação também é recomendado. Não deve ser usando, por exemplo, o código 200 (OK) quando algo der errado.
- Nível 3: O nível três de maturidade faz uso de todos os três fatores, isto é, URIs, métodos HTTP e HATEOAS. Esse é o nível em que a maioria das APIs menos implementam e pois não seguem o princípio de HATEOAS.

HATEOAS (Hypermedia as the Engine of Application State) é uma abordagem para a construção de serviços REST, onde o cliente pode descobrir dinamicamente as ações disponíveis em tempo de execução. Todo o cliente deve exigir para começar é um URI inicial e um conjunto de tipos de mídia padronizados. Uma vez que carregou o URI inicial, todas as futuras transições de estado da aplicação serão conduzidas pelo cliente selecionando as escolhas fornecidas pelo servidor.

Não seguir essa abordagem não é necessariamente ruim. Há alguns bons argumentos a serem feitos a favor e contra a utilização de HATEOAS. Enquanto, por um lado, torna as APIs fáceis de descobrir e usar, que geralmente vem ao custo de mais tempo e esforço de desenvolvimento.

3.3.3 Limitações

Contudo, com a evolução e o aumento da complexidade das APIs, a comunicação via o protocolo REST vem se mostrando muitas vezes inviável pois sua implementação trás como consequências alguns problemas estruturais como:

- A necessidade de executar múltiplas requisições entre o cliente e o servidor a fim de obter objetos complexos e com atributos aninhados Para aplicações móveis operando em condições variáveis de rede, essas múltiplas viagens de ida e volta são indesejáveis, pois geram atrasos e maior tráfego na rede ([SCHROCK, 2015a](#)).
- A prática de *over-fetching*, ou seja, quando o cliente busca alguma informação do servidor e a resposta contém mais informação que o cliente precisa. Esse tipo de problema acarreta no uso desnecessário de recursos de comunicação ([GUSTAVSSON, 2016](#)).
- O versionamento da API, que ocorre quando há alterações significativas na API, sujeitas a quebra de código nos clientes consumidores. Essa prática pode ser extremamente penosa se a API é usada por uma grande massa de clientes que não são facilmente atualizados ([KROHN, 2013](#)).

3.4 Arquiteturas baseadas em JSON/Graphs

Recentemente, um novo design de arquitetura vem ganhando espaço, preenchendo algumas lacunas que arquiteturas anteriores deixaram. Lideradas pelo Facebook com seu GraphQL e Netflix com o Falcor, esta nova arquitetura dá um passo para trás comparando-se ao REST, atingindo apenas o nível 0 no Modelo de Maturidade de Richardson.

Os principais problemas que essa arquitetura ajuda a resolver são: A dependência das aplicações cliente nos servidores, eliminando a necessidade do servidor de ter que manipular as informações ou o tamanho da resposta; A necessidade de executar diversas requisições para acessar os dados exigidos por uma *view*; Melhorar a experiência de desenvolvimento *frontend*, pois existe uma relação próxima entre os dados necessários à interface da aplicação e a forma como um desenvolvedor pode expressar uma descrição desses dados para a API. O presente trabalho irá apenas focar no GraphQL como representante das arquiteturas baseadas em JSON/Graph.

GraphQL é uma linguagem de consulta criada pelo Facebook em 2012, que fornece uma interface comum entre o cliente e o servidor para manipulação e busca de dados. GraphQL utiliza de um sistema chamado *cliente-specified queries*, onde o formato de resposta de uma requisição é definida pelo cliente. [Schrock \(2015b\)](#) afirma que uma vez que a estrutura de dados não é codificada, como nas APIs tradicionais, a consulta de dados do servidor se torna mais eficiente para o cliente.

Consultas utilizando GraphQL sempre retornam apenas o que foi pré definido pela requisição, fazendo suas respostas sempre serem previsíveis. As consultas com GraphQL acessam não apenas as propriedades de um único recurso, mas também seguem as referências entre eles. Enquanto as APIs REST típicas exigem o carregamento de múltiplas URLs, as APIs GraphQL obtêm todos os dados que precisam em uma única requisição.

Adicionar novos campos ou tipos a uma API GraphQL não afeta nenhuma consulta ou funcionalidade já existente. Campos não mais utilizados podem ser obsoletos e ocultos de ferramentas de mapeamento. Ao usar uma única versão em evolução, as APIs GraphQL dão as aplicações acesso contínuo a novos recursos e encorajam um código de servidor mais limpo e mais sustentável.

3.4.1 Operações

Existem três tipos de operações modeláveis com GraphQL:

- Query: Uma consulta de somente leitura
- Mutation: Uma escrita seguida por consulta

- Subscription – Uma requisição de longa duração que obtém dados em resposta a eventos disparados pelo servidor.

Uma Query em GraphQL é uma maneira de obter dados de uma maneira somente de leitura em uma API GraphQL. De uma maneira geral, GraphQL é se baseia em consultar campos específicos em objetos. Isso significa que a consulta tem exatamente o mesmo formato que a resposta. Isso é essencial para GraphQL, porque a resposta é sempre previsível, e o servidor sabe exatamente quais os campos que o cliente está pedindo.

Como GraphQL não se limita apenas em consultas de dados, as APIs também podem implementar operações para criar, atualizar e destruir dados. Para esses tipos de operações, o GraphQL usa o termo Mutations. As Mutations são uma maneira de alterar dados em seu servidor. É importante notar que as *mutations* consistem em uma alteração seguida de uma busca do dado que acabou de ser alterado, tudo em uma única operação.

Os aplicativos em tempo real precisam de uma maneira de enviar dados do servidor. As subscriptions permitem que aplicações publiquem eventos em tempo real através de um servidor de assinaturas GraphQL. Com o modelo de subscriptions baseado em eventos no GraphQL - muito parecido com as queries e mutations - um cliente pode dizer ao servidor exatamente quais dados devem ser enviados e como esses dados devem ser encontrados. Isso leva a menos eventos rastreados no servidor e notificados para o cliente.

3.4.2 Schemas e Tipos

O sistema de tipos do GraphQL descreve as capacidades de um servidor GraphQL e é usado para determinar se uma consulta é válida. O sistema de tipos também descreve os formatos de entrada de variáveis de consulta para determinar se os valores fornecidos em tempo de execução são válidos. As capacidades do servidor GraphQL são referidas como *schema* do servidor. Um *schema* é definido baseado nos tipos que ele suporta.

Cada servidor GraphQL define um conjunto de tipos que descrevem completamente o conjunto de dados possíveis que você pode consultar nesse servidor. Então, quando as consultas chegam, elas são validadas e executadas contra os *schemas*. O *schema* descreverá quais campos o servidor pode responder e quais tipos de objetos as respostas irão conter. A informação de tipo é muito importante para o GraphQL e o cliente pode assumir de forma segura que o servidor retornará tipos consistentes de objetos para o mesmo campo.

A unidade fundamental de qualquer *schema* GraphQL é o tipo. Existem oito formatos de tipos suportados no GraphQL. O tipo mais básico é o **Scalar**. Um campo do tipo *Scalar* representa um valor primitivo, como uma *string* ou um número inteiro. Muitas vezes, as respostas possíveis para um campo do tipo *Scalar* são enumeráveis. O GraphQL oferece um tipo **Enum** nesses casos, onde o tipo especifica o espaço de respostas válidas. Campos do tipo **Object** definem um conjunto de campos, onde cada campo é

outro formato, permitindo a definição de hierarquias de tipos arbitrários. O GraphQL suporta dois tipos abstratos: **Interfaces** e **Unions**.

Todos os tipos até agora são considerados nulos e singulares, por exemplo, uma *string* retorna um valor nulo ou singular. O sistema de tipo pode querer definir que ele retorna uma lista de outros tipos; O tipo **List** é fornecido por esse motivo e envolve outro tipo. Da mesma forma, o tipo **Non-Null** envolve outro tipo e indica que o resultado nunca será nulo.

Finalmente, muitas vezes é útil fornecer estruturas complexas como entradas para consultas GraphQL; o tipo **Input Object** permite que o *schema* defina exatamente quais dados são esperados do cliente nessas consultas.

3.4.3 Autorização

Segundo [Glass et al. \(2000\)](#), Autorização é o ato de determinar se um direito particular, como acesso a algum recurso, pode ser concedido à uma determinada credencial. Nos sistemas de informação multiusuários, o administrador do sistema define quais usuários podem ter acesso ao sistema e quais são os privilégios de uso.

Normalmente o GraphQL não controla diretamente o controle de acesso, em vez disso, delega essa responsabilidade para a camada de lógica de negocio da aplicação.

3.4.4 Caching

De acordo com [Grigoriuk \(2017\)](#), consultas de dados através da internet podem ser lentas e custosas, e por esse motivo, a capacidade de armazenar em cache e reutilizar os recursos obtidos anteriormente é um aspecto crítico da otimização para o desempenho.

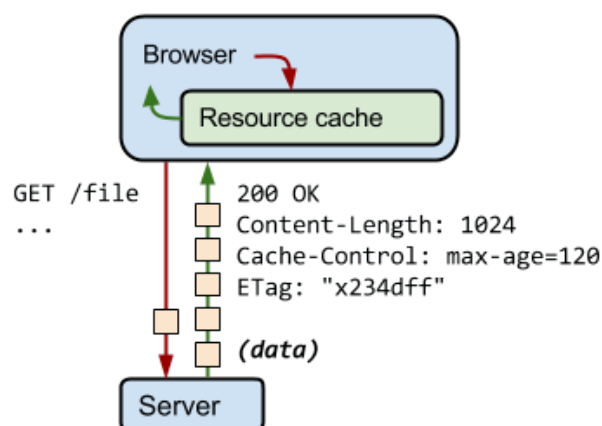


Figura 5 – HTTP Caching. (refazer)

Em uma API baseada em *end-points*, os clientes podem usar o armazenamento em cache do protocolo HTTP para evitar buscas desnecessárias identificando quando dois

recursos são iguais. A URL nessas APIs é o identificador globalmente exclusivo que o cliente pode aproveitar para criar um cache. No GraphQL, porém, não existe uma primitiva semelhante a uma URL que forneça esse identificador globalmente exclusivo para um determinado objeto. Entretanto, existem várias práticas que podem ser aplicadas a um servidor GraphQL para atingir o mesmo objetivo.

Um padrão possível para isso é reservar um campo, como `id`, para ser o identificador globalmente exclusivo. Da mesma forma que os URLs de uma API baseada em recursos fornecem uma chave globalmente exclusiva, o campo `id` neste sistema fornece uma chave globalmente única. Outra abordagem possível funcionaria de forma semelhante à ao padrão usada nas APIs baseadas em *end-points*. O texto da consulta em si pode ser usado como identificador globalmente exclusivo.

3.4.5 Limitações

Segundo aponta [Buna \(2017\)](#), uma ameaça importante que o GraphQL *facilita* são os ataques de negação de serviço. Um servidor GraphQL pode ser atacado com consultas excessivamente complexas que irão consumir todos os recursos do servidor. Esse tipo de ataque não é específico do GraphQL, mas é preciso um cuidado redobrado para evitar-los.

Há entretanto, alguns procedimentos que se implementados, podem mitigar a ameaça de negação de serviço. É possível fazer uma análise de custos na consulta com antecedência e impor *um certo tipo* de limites na quantidade de dados que uma requisição pode consumir. Também é possível implementar um tempo limite para que requisições que levam muito tempo para serem resolvidas, sejam excluídas da fila de execução .

4 Estudo de Caso

Para obter um exemplo do mundo real, o estudo de caso foi projetado com base em uma aplicação de WMS (*Warehouse Management System*). Rouse (2009) explica que um sistema de WMS é um software que auxilia as operações do dia-a-dia em um armazém. Os sistemas WMS permitem o gerenciamento centralizado de tarefas, como o rastreamento de níveis de inventário e locais de estoque. Os sistemas WMS podem ser um aplicativo independente ou parte de um sistema de ERP.

Com o intuito de realizar consultas relevantes, foi modelado um esquema contendo seis entidades, capazes de representar consultas reais, e ainda produzirem informações pertinentes para a comparação de desempenho. A construção das APIs seguiram os princípios do ciclo tradicional do desenvolvimento de software, portanto a modelagem foi elaborada antes do início da implementação das APIs. Isso assegura que nenhuma decisão de tecnologia afetou a modelagem das entidades, que pode ser exemplificado na figura 6.

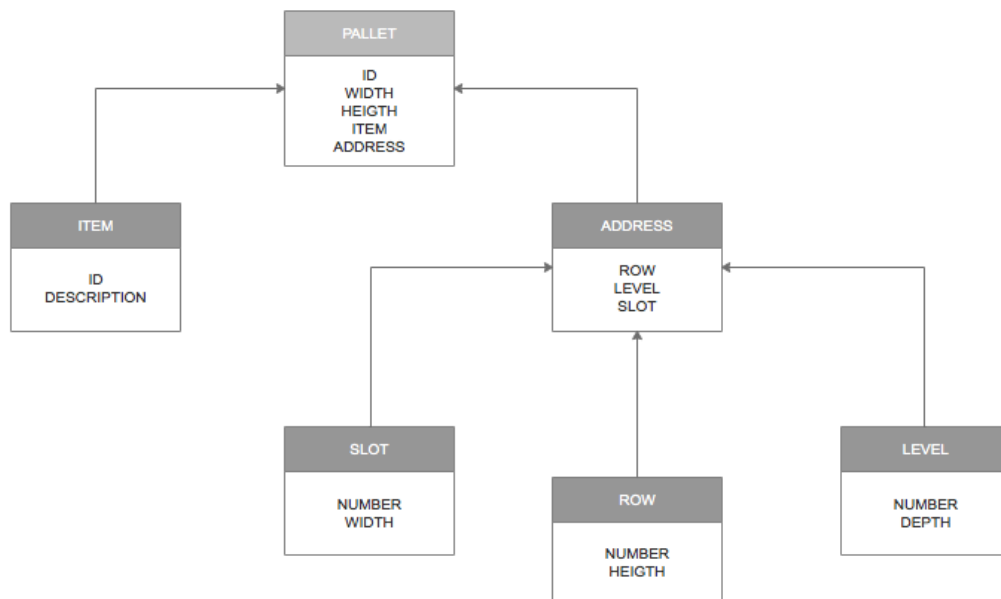


Figura 6 – Modelagem WMS

A modelagem acima, representa o gerenciamento de um armazém, com capacidade de armazenar diversos itens. Os itens são dispostos em pallets, e alocados em endereços dentro do armazém. Como pode ser observado na figura 7, os endereços do armazém são formados a partir de uma combinação de três *dimensões*: Prateleira, Linha e Nível. Cada combinação dessas três propriedades é capaz de alocar um pallet, que por sua vez pode conter diversas unidades de um mesmo item.

Imagine um item de código 22B12, por exemplo, que representa um produto X. Este produto é disposto em um pallet com capacidade de armazenar 30 unidades do item 22B12. O armazém é composto por 26 prateleiras, sequenciadas de 'A' a 'Z'. Cada prateleira possui 2 linhas de profundidade e 3 níveis de altura. O pallet com o código 001 contém 30 unidades do item 22B12, e precisa ser alocado dentro armazém, para isso é utilizado um sistema de códigos envolvendo as 3 dimensões do armazém.

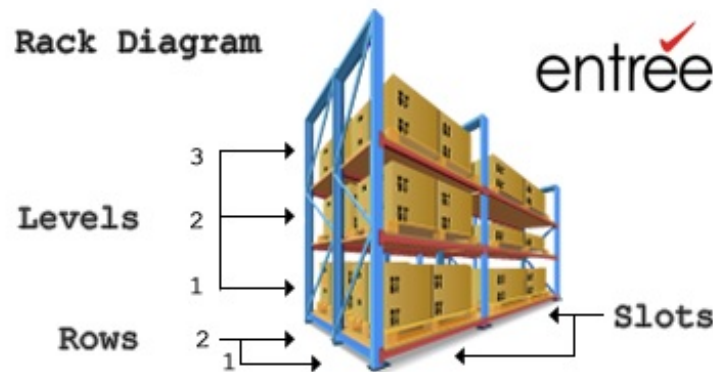


Figura 7 – Diagrama de dimensões do armazém (REFAZER)

60 unidades do item 22B12 estão dispostos em dois pallets(001, 002) existentes no armazém e cada um dos pallets será destinado a um endereço. O pallet 001 será alocado na terceira prateleira, no segundo nível e na primeira linha. Após a formação do endereço, o pallet 001 estará localizado no endereço C0101.

4.1 Hipóteses

Hipóteses na diferença de desempenho entre as APIs derivaram dos fundamentos teóricos. Elas baseiam-se no entendimento de que os protocolos utilizados possibilitam a implementação de uma combinação de técnicas afim de afetar positivamente o desempenho da API. Portanto, ambas as implementações devem ter as mesmas propriedades, seguindo suas melhores práticas, modelos de maturidade e documentação.

As hipóteses deste trabalho são:

- Utilizando GraphQL irá reduzir o tamanho da resposta;
- Utilizando GraphQL irá reduzir o tempo de resposta;
- Utilizando REST irá reduzir a utilização de consumo de CPU;
- Utilizando REST irá reduzir o consumo de Memória;

Com o propósito de validar as hipóteses definidas, foram determinadas duas perguntas que envolvessem todas as entidades. Para cada pergunta existe apenas uma resposta correta e sua lógica é baseada em campos das estruturas de dados de retorno.

Questão 1: Qual item ocupa a maior quantidade de pallets alocados no armazém?

Questão 2: O item 22B12 está armazenado em quais endereços?

4.2 Cenários

A escolha das ferramentas a serem usadas na implementação foi uma das partes mais importantes no planejamento do estudo de caso. Foi necessário pensar em uma especificação que apresentasse uma curva rápida de aprendizagem, um fluxo de execução replicável e agnóstico à plataforma, e uma documentação madura em relação a implementação das APIs.

As APIs foram escritas na linguagem de programação JavaScript, utilizando a especificação do ECMAScript 5. Foi escolhido esta linguagem com vista ao fácil acesso a bons frameworks para construção de serviços WEB, como o Node.js. Node.js é uma plataforma para desenvolvimento de servidores web baseadas em rede utilizando JavaScript e o V8 JavaScript Engine, ou seja, com Node.js pode-se criar uma variedade de aplicações Web utilizando código em JavaScript [Lopes \(2014\)](#).

4.2.1 Ferramentas utilizadas

Node.js dispõe de inúmeros recursos e ferramentas que possibilitam a construção de APIs. Mesmo assim, o ecossistema de bibliotecas no Node.js conta com ferramentas que simplificam ainda mais a construção de aplicações para servidores Web. Nesta sessão serão citados algumas nessas bibliotecas que foram utilizadas para desenvolver tanto a API REST quando a API GraphQL.

Express.JS

Express é um *framework* para Node.js extremamente flexível, que forcene um conjunto robusto de feramentas para a construção de aplicações Web e Móveis. Conta com um robusto sistema de roteamento, facilitando o desenvolvimento de APIs. O Express forcene uma fina camada de abstração nas principais funcionalidades do Node.js, sem sobrepor os recursos no Node.

Nos protótipos desenvolvidos para executar o experimento, o Express teve como papel de *middleware* gerenciando as rotas, e delegando a responsabilidade de interpretação das requisições para os *Controllers* na API REST e para os *Resolvers* na API GraphQL.

O código 1 ilustra como os *Controllers* enviam as informações da requisição para o model, que é o responsável na API REST em executar as consultas no banco de dados. No exemplo abaixo, o trecho de código é o responsável por retornar um Item baseado no id recebido como parâmetro da requisição.

```
1 //item.controller.js
2 import Item from '../models/item.model';
3
4 /**
5  * Load item and append to req.
6  */
7 function load(req, res, next, id) {
8   Item.get(id)
9     .then((item) => {
10       req.item = item;
11       return next();
12     })
13     .finally(e => next(e));
14 }
```

Código 1 – Controller para carregar um item

Já o código 2 mostra como o Express.JS gerencia a requisição recebida via método GET, e delega a responsabilidade para o schema.js, onde está contida a lógica para a interpretação dos parâmetros da requisição e retornar um objeto JSON com a devida resposta.

```
1 //server.js
2 import express from 'express';
3 let app = express();
4
5 import schema from './schema.js';
6
7 app.get('/', (req, res) => {
8   graphql(schema, req.query.query)
9     .then((result) => {
10       res.send(result);
11     });
12 });
```

Código 2 – Controller para carregar um item

MongoDB

MongoDB é um banco de dados *open source* que utiliza um modelo de dados orientado a objetos. O mesmo tem como característica conter todas as informações importantes em um único documento, possuir identificadores únicos universais (UUID),

possibilitar a consulta de documentos através de métodos avançados de agrupamento e filtragem, também conhecido como *MapReducers*. Ao invés de usar tabelas e linhas, como os bancos de dados relacionais, o MongoDB usa uma arquitetura baseada em coleções e documentos.

Ferramentas clientes

Serão utilizadas ferramentas que se comportaram como cliente das APIs construídas. O foco deste trabalho não é em cima das aplicação clientes, entretanto é importante conhece-las pois são elas que invocarão as consultas como também é através delas que algumas métricas serão extraídas. O software Postman será usado fundamentalmente para a execução das buscas nas APIs.

O Postman é uma cadeia completa de ferramentas para desenvolvedores de APIs, usadas empresas em todo o mundo. O Postman é uma ferramenta elegante e flexível usada para construir softwares conectados via APIs - de forma rápida, fácil e precisa. Fundada em 2014, a empresa escalou grandes alturas e hoje, ela é utilizada por mais de 3 milhões usuários ao redor do mundo.

Ele funciona como uma emulador para execução de consultas em APIs. Com ele é possível executar buscas utilizando qualquer dos métodos do protocolo HTTP, possibilitando customização tando do corpo quando do cabeçalho das requisições se preciso. Junto com a resposta da consulta, o Postman traz também informações extremamente relevantes, como tempo que a API levou para responder e o tamanho em bytes contido na resposta. Essas funcionalidade alidas com a opção de executar um numero customizavel de iterações para cada requisição, será a base para avaliar o desempenho das APIs REST e GraphQL.

4.2.2 Ambiente

A configuração da máquina utilizada para o teste de desempenho local é descrita na Tabela 1. As APIs REST e GraphQL foram construídas para responder as requisições recebidas retornando respostas no formato JSON, a fim de gerar um fator de medição de desempenho da execução dos testes. Para a execução do experimento, apenas os softwares necessários estavam ativos, portanto, ao executar as buscas, o servidor REST ou GraphQL estará ativo, além do servidor de banco, as aplicações clientes, e as aplicações de coleta das métricas.

4.2.3 Detalhes implementação REST

O servidor REST consiste em um servidor HTTP escrito em Node.js que recebe as requisições HTTP recebidas. Dependendo do método e URL da requisição, roteia-o para

Item	Descrição
Marca/Modelo	Mac
Processador	Intel
Memória	X GB
Disco rígido	SSD 128
Quantidade núcleos	4

Tabela 1 – Configuração do Ambiente

URI	Descrição
/item	Consulta lista de itens
/item/:id	Consulta item pelo id
/pallet	Consulta lista de pallets
/pallet/:id	Consulta pallet pelo id
/address	Consulta lista de endereços
/address/:id	Consulta endereço pelo id
/slot	Consulta lista de prateleiras
/slot/:id	Consulta prateleira pelo id
/row	Consulta lista de linhas
/row/:id	Consulta linha pelo id
/level	Consulta lista de nível
/level/:id	Consulta nível pelo id

Tabela 2 – Servidor REST

o *controller* correspondente. O *controller* faz a consulta na base de dados do MongoDB. Ao realizar o gerenciamento das rotas, o servidor realiza os devidos logs de latência. Após a consulta, a resposta desejada é enviada de volta para o cliente. Esse fluxo acontece em qualquer cenário, independente do número de requisições desejadas.

No presente trabalho, como o objetivo é apenas medir o desempenho de consultas utilizando REST, todas as requisições utilizarão o método HTTP GET. Para as medições da API REST, a aplicação cliente irá enviar por exemplo, uma requisição para recuperar todos os itens cadastrados. Como pode ser observado da tabela 2, é necessário executar uma busca do tipo localhost:8080/item, que será interpretada pelo servidor REST, identificando qual é a rota responsável por essa requisição. O servidor consultará a base de dados retornando uma resposta no formato JSON, com todos os itens cadastrados na API. Esse fluxo é ilustrado na figura 8

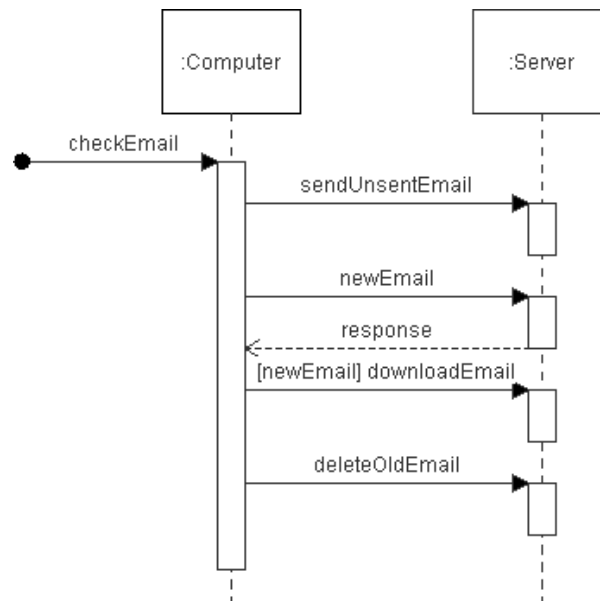


Figura 8 – UML sequence REST (REFAZER)

4.2.4 Detalhes implementação GraphQL

O servidor GraphQL foi implementado também usando Node.js. A diferença em comparação com a implementação do servidor REST é que o servidor GraphQL envia todos os pedidos ao núcleo GraphQL ao invés de rotear as requisições recebidas para vários *controllers* diferentes. O GraphQL analisa a consulta e envia os parâmetros para os *resolvers* responsáveis, localizados nos *schemas* GraphQL. *Resolvers* são funções definidas para todos os campos no *schema* GraphQL, cada um retorna os dados para o campo específico. Estas funções são executadas quando os campos correspondentes são consultados e os resultados são retornados na resposta.

```

1 query RootQuery {
2   items {
3     id
4     description
5   }
6 }
  
```

Código 3 – Consulta de itens

A consulta acima retorna como resposta lista a de todos os itens cadastrados na base de dados. Note que a resposta contém apenas os atributos que a consulta requisitou:

```

1 {
2   "data": [
3     {
4       "id": 22B12
5       "description": "Flat screen",
6     },
  
```

```
7      {  
8          "id": 21C44  
9          "description": "Computer screen",  
10     },  
11     {  
12         "id": 43F12  
13         "description": "Smartphone screen",  
14     },  
15 ]  
16 }
```

Código 4 – Listagem dos itens (Validar)

Aqui tenho que explicar o diagrama de sequência para o GraphQL

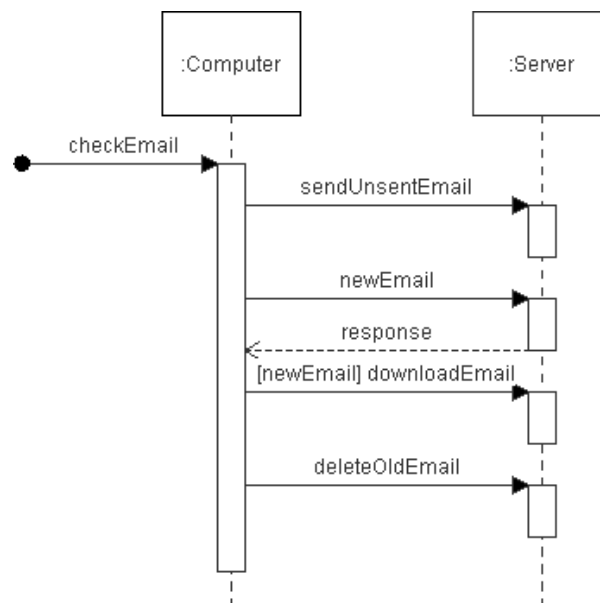


Figura 9 – UML sequence GraphQL (REFAZER)

4.3 Métricas

Afim de comparar as medidas de desempenho de APIs desenvolvidas em REST com APIs desenvolvidas em GraphQL, algumas métricas precisam ser definidas. O desempenho de cada API vai depender de sua implementação, porém, escolhendo as métricas corretas, o efeito da implementação pode ser reduzido. Focando em medições corretas, as diferenças relevantes das APIs podem ser destacadas e melhor ponderadas.

Para o presente trabalho, serão utilizadas quatro métricas diferentes: Utilização da CPU, Consumo de memória, Tempo de resposta e o Tamanho da resposta. É importante ressaltar que cada métrica foi medida separadamente, para que *logs* e *outputs* pertinentes a uma métrica específica não interfira no resultado das demais.

Utilização da CPU

A utilização da CPU é uma medida do percentual de ciclos que representa uma porcentagem em que as unidades de processamento ficam dedicadas a executar um processo em particular. No contexto do presente trabalho, é a porcentagem de processamento dentro da CPU utilizada em cada requisição.

Esta métrica será extraída através do módulo *OS* presente no core do Node.js. A função *os.loadavg()* trás como resultado uma medida da atividade do sistema, calculada pelo sistema operacional e expressa como um número fracionário. A média será a medida de comparação entre as duas abordagens. A seguinte fórmula será usada para calcular esta métrica:

$$S = mrs1 + mrz2.. + mrsn$$

Onde *MRS* é o tempo total para obtido em cada requisição necessária, e *S* representa a soma dos tempos, resultando do

Consumo de memória

O consumo de memória, junto com a utilização da CPU, é uma das medidas mais importantes pois é nesses pontos que observamos o verdadeiro custo por trás da escolha da ferramenta. O consumo de memória é a quantidade de em bytes utilizado pela API, e será medido através da soma da utilização de memória em cada consulta necessária para atender os senários propostos. A seguinte fórmula será usada para calcular esta métrica:

$$S = mrs1 + mrz2.. + mrsn$$

Onde MRS é o tempo total para obtido em cada requisição necessária, e S representa a soma dos tempos, resultando do

Ferramenta X utilizada

Tempo de resposta

Tempo que cada requisição levou para realizar a consulta. Esse tempo é calculado a partir do início da requisição, até o retorno da resposta completa. No caso da API REST, esta métrica será acumulativa, ou seja, a soma do tempo de todas as requisições necessárias. A seguinte fórmula será usada para calcular esta métrica:

$$S = mrs1 + mrs2.. + mrsn$$

Onde MRS é o tempo total para obtido em cada requisição necessária, e S representa a soma dos tempos, resultando do

Tamanho da resposta

O tamanho da resposta será calculada baseado no tamanho em textitbytes da resposta. Essa métrica será calculada a partir da média de 12 consultas. Novamente, para a API REST será usada a média das somas de todas as consultas necessárias para obter a resposta desejada. A seguinte fórmula será usada para calcular esta métrica:

$$S = mrs1 + mrs2.. + mrsn$$

Onde MRS é o tempo total para obtido em cada requisição necessária, e S representa a soma dos tempos, resultando do

Conclusão

A figura 10 mostra como as métricas serão extraídas. A utilização do CPU e o consumo de memória serão medidos utilizando ferramentas do próprio Node.js. Ambos serão medidos via *logs* no código fonte dos protótipos implementados. O tamanho da resposta e o tempo de resposta serão extraídos utilizando a ferramenta Postman, ao final de cada consulta.

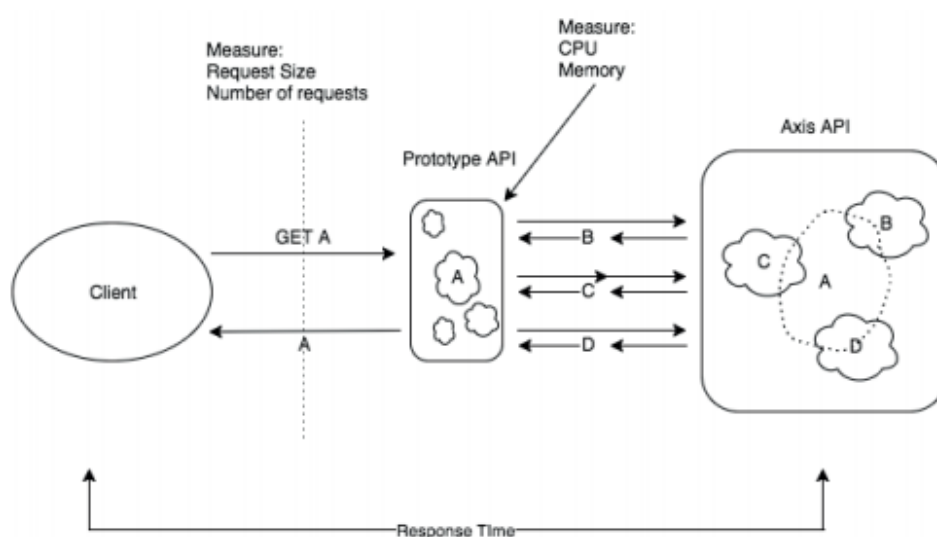


Figure 3.1: System model with the different measurements pointed out

Figura 10 – Arquitetura das APIs e diferentes pontos de medidas (REFAZER)

5 Resultados

Neste capítulo, os resultados e a análise dos experimentos realizados durante o trabalho são reportados. Os resultados obtidos nos testes de validação apresentam a média dos valores coletados após diversas execuções dos cenários de forma sequencial.

Foram executadas 30 interações para cada requisição necessária a fim de obter as respostas das questões levadas no capítulo anterior. Os dados necessários puderam ser buscados em uma única requisição à API GraphQL, enquanto foram precisos à API REST duas requisições para chegar a resposta da primeira questão, e vinte e sete para obter a resposta da segunda. Os detalhes podem ser constatados na tabela 3.

Requisição	Resultado	Número de requisições
/items	ID do item 22B12	1x
/items/:id	Detalhes do item 22B12	1x
/pallets	Pallets contendo item 22B12	1x
/pallets/:id	Detalhes do Pallet contendo o item 22B12	5x
/addresses/:id	Detalhes do Endereço contendo o item 22B12	5x
/levels/:id	Nível contendo o item 22B12	5x
/slots/:id	Prateleira contendo o item 22B12	5x
/rows/:id	Linha contendo o item 22B12	5x

Tabela 3 – Fluxo de dados para responder as questões

5.1 Questão 1

5.1.1 Utilização da CPU

5.1.2 Consumo de memória

5.1.3 Tempo de resposta

Como esperado, a API escrita com GraphQL tem realmente um tempo de resposta menor. A figura 5.1.3 mostra a diferença do desempenho das APIs para encontrar a resposta da primeira questão.

Com mil pallets cadastrados, a API REST teve como resultado um tempo de resposta de 147.3 ms, enquanto a API GraphQL respondeu a consulta em 115.63 ms, representando uma diferença de X por cento. Quando aumentamos o numero de pallets para 10 mil, a API REST respondeu as consultas em 1108.13 ms e a API GraphQL devolveu os resultados em 925.63 ms, uma diferença de Y por cento. Por último, as consultas com 30

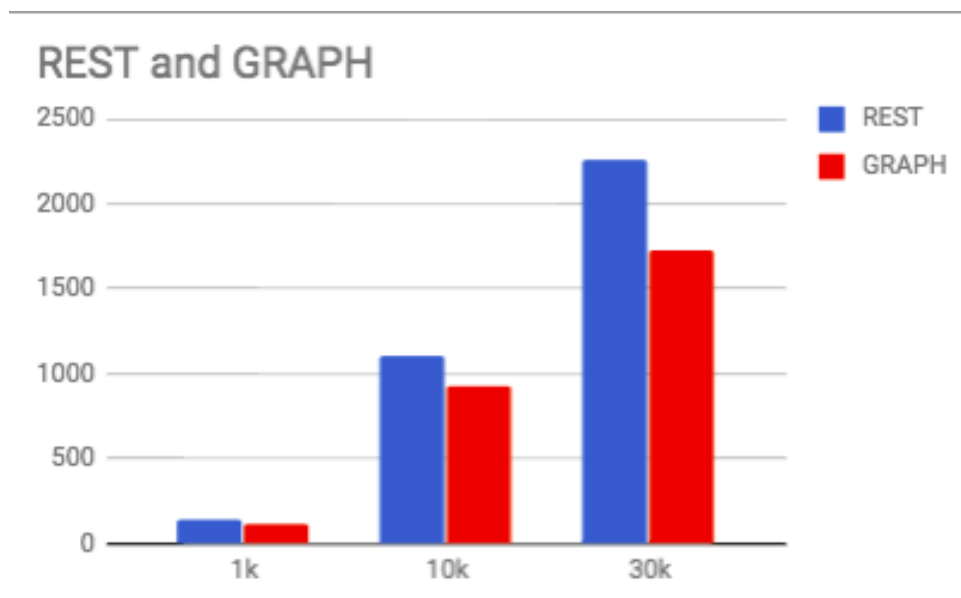


Figura 11 – Comparação do tempo de resposta

mil pallets foram respondidas em 2261.1 ms na API REST e 1725.7 ms na API GraphQL, o que representa uma diferença de Z por cento.

5.1.4 Tamanho da resposta

Outro resultado esperado era que o tamanho da resposta da API GraphQL fosse menos do que o tamanho da resposta da API REST. Essa hipótese se confirmou como pode ser visto da figura 5.1.4.

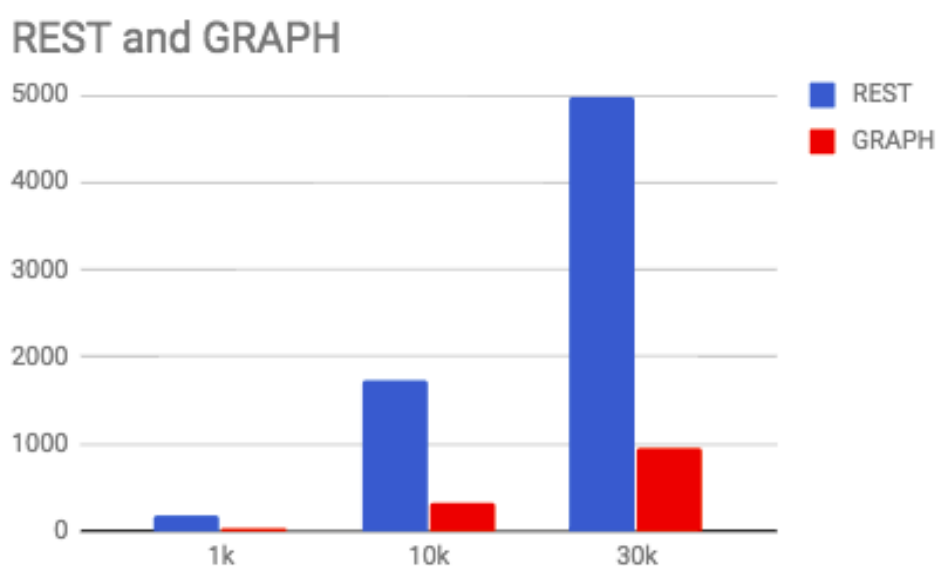


Figura 12 – Comparação do tamanho de resposta

A API REST respondeu as requisições da questão 1 com um tamanho de resposta

de 174.04 Kb, 1740.17 Kb e 4980 Kb para mil pallets, 10 mil pallets e 30 mil pallets respectivamente. Da mesma maneira, a API GraphQL teve como resultado respostas com 31.68 Kb, 322.53Kb e 967.35Kb.

5.2 Questão 2

Para as repostas da questão 2, as consultas foram mais complexas na API GraphQL, e mais numerosas na API REST.

5.2.1 Utilização da CPU

5.2.2 Consumo de memória

5.2.3 Tempo de resposta

É possível identificar com clareza, como mostra a figura 5.2.3, que a API REST leva um tempo consideravelmente maior para responder todas as requisições do que a consulta na API GraphQL.

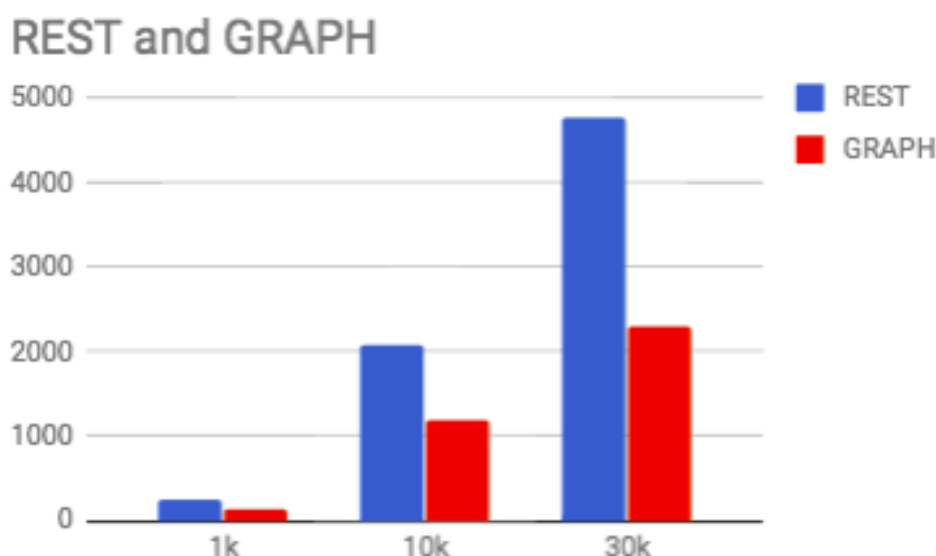


Figura 13 – Comparação do tempo de resposta

Com mil pallets cadastrados, a API REST teve como resultado um tempo de resposta de 254.56 ms, enquanto a API GraphQL respondeu a consulta em 148.5 ms, representando uma diferença de X por cento. Quando aumentamos o numero de pallets para 10 mil, a API REST respondeu as consultas em 2072.03 ms e a API GraphQL devolveu os resultados em 1201.3 ms, uma diferença de Y por cento. Por último, as consultas com 30 mil pallets foram respondidas em 4770.2 ms na API REST e 2291.1 ms na API GraphQL, o que representa uma diferença de Z por cento.

5.2.4 Tamanho da resposta

A API GraphQL também mostrou-se mais vantajosa em termos de tamanho de resposta na Questão 2. A figura 5.2.4 mostra a comparação entre os protótipos quando comparados o tamanho da resposta.

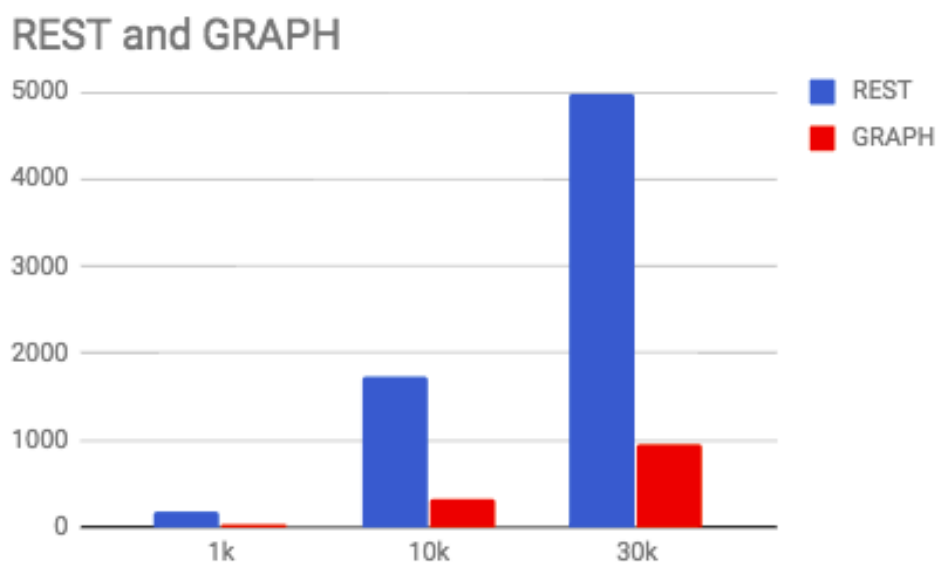


Figura 14 – Comparação do tamanho de resposta

A API REST respondeu as requisições da questão 1 com um tamanho de resposta de 259.53 Kb, 2522.17 Kb e 7221 Kb para mil pallets, 10 mil pallets e 30 mil pallets respectivamente. Da mesma maneira, a API GraphQL teve como resultado respostas com 101.73 Kb, 1005.23 Kb e 2850 Kb.

6 Conclusão

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.

Referências

- ARAUJO, L. C. *A classe abntex2: Modelo canônico de trabalhos acadêmicos brasileiros compatível com as normas ABNT NBR 14724:2011, ABNT NBR 6024:2012 e outras*. [S.l.], 2015. Disponível em: <<http://www.abntex.net.br/>>.
- ARAUJO, L. C. *O pacote abntex2cite: Estilos bibliográficos compatíveis com a ABNT NBR 6023*. [S.l.], 2015. Disponível em: <<http://www.abntex.net.br/>>.
- ARAUJO, L. C. *O pacote abntex2cite: tópicos específicos da ABNT NBR 10520:2002 e o estilo bibliográfico alfabético (sistema autor-data)*. [S.l.], 2015. Disponível em: <<http://www.abntex.net.br/>>.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 6028: Resumo - apresentação*. Rio de Janeiro, 2003. 2 p.
- BARRY, D. K. *Service Architecture - Representational State Transfer (REST)*. 2003. Disponível em: <<https://goo.gl/KB2CGw>>. Acesso em: 17 mai 2017.
- BATTLE, E. B. R. *Web Semantics: Science, Services and Agents on the World Wide Web*. 1300 North 17th Street, Suite 400 Arlington, VA 22209, United States: BBN Technologies, 2008.
- BOX, D. et al. Simple Object Access Protocol (SOAP) 1.1. [Http://www.w3.org/TR/soap](http://www.w3.org/TR/soap). 2000.
- BUNA, S. *REST APIs are REST-in-Peace APIs. Long Live GraphQL*. 2017. Disponível em: <<https://goo.gl/3NDGrq>>. Acesso em: 15 abr 2017.
- FIELDING, R. T. *Architectural Styles and the Design of Network-based Software Architectures*. Tese (Doutorado) — University of California, Irvine, 2000. AAI9980887.
- GLASS, S. et al. *Mobile IP Authentication, Authorization, and Accounting Requirements*. RFC Editor, 2000. RFC 2977. (Request for Comments: 2977, 2977). Disponível em: <<https://rfc-editor.org/rfc/rfc792.txt>>.
- GRIGORIK, I. *GraphQL Overview*. 2017. Disponível em: <<https://goo.gl/nGfQ4x>>. Acesso em: 15 abr 2017.
- GUSTAVSSON, K. *Efficient data communication between a webclient and a cloud environment*. 2016. Disponível em: <<https://goo.gl/Z07LrK>>. Acesso em: 15 abr 2017.
- [HTTP://WIKI.INF.UFPR.BR/MAZIERO/LIB/EXE/FETCH.PHP?MEDIA=SO:SO-LIVRO.PDF](http://wiki.inf.ufpr.br/maziro/lib/exe/fetch.php?media=so:so-livro.pdf). *Sistemas Operacionais: Conceitos e Mecanismos*. [S.l.], 2013. Acesso em: 01 mai 2017.
- KNOW, I. dont. *I dont know*. 2008. Disponível em: <http://www.facol.br/sophia/2741/apostila09_sincronizacao_comunicacao_entre_processos.pdf>.
- KOHLHOFF, R. S. C. *Evaluating SOAP for High Performance Business Applications: Real-Time Trading Systems*. 2003. Disponível em: <<https://goo.gl/2w8a41>>. Acesso em: 17 mai 2017.

- KROHN, A. *Do you really need API Versioning?* 2013. Disponível em: <<https://goo.gl/lhvEph>>. Acesso em: 15 abr 2017.
- KUKREJA, N. G. R. Remote procedure call: Limitations and drawbacks. *International Journal of Research (IJR)*, v. 1, n. 10, p. 914–917, 11 2014.
- LENSMAR, O. *Is REST losing its flair - REST API Alternatives*. 2013. Disponível em: <<https://goo.gl/bPm1ez>>. Acesso em: 15 abr 2017.
- LEOPOLDO, M. R. B. *Simple Object Access Protocol Entendendo o Simple Object Access Protocol (SOAP)*. 2007. Disponível em: <<https://goo.gl/ATS7QB>>. Acesso em: 17 mai 2017.
- LOPES, C. *O que é Node.js e saiba os primeiros passos*. 2014. Disponível em: <<https://goo.gl/SGQuT2>>. Acesso em: 15 abr 2017.
- MASO, M. *Um Modelo de Comunicação para Automação na Execução de Consultas de Dados sobre APIs Web*. 2016. Disponível em: <<https://goo.gl/mlMYso>>. Acesso em: 15 abr 2017.
- MERRICK, P.; ALLEN, S.; LAPP, J. *XML remote procedure call (XML-RPC)*. Google Patents, 2006. US Patent 7,028,312. Disponível em: <<https://www.google.com/patents/US7028312>>.
- MICROSOFT. *Interprocess Communications*. 2008. Disponível em: <[https://msdn.microsoft.com/en-us/library/windows/desktop/aa365574\(v=vs.85\).aspx#base_using_rpc_for_ipc](https://msdn.microsoft.com/en-us/library/windows/desktop/aa365574(v=vs.85).aspx#base_using_rpc_for_ipc)>.
- MUELLER, J. *Understanding SOAP and REST Basics And Differences*. 2013. Disponível em: <<https://goo.gl/EZWLg1>>. Acesso em: 02 ago 2017.
- PALIARI, M. *COMO FUNCIONA UM WEBSERVICE REST*. 2012. Disponível em: <<https://goo.gl/lJgyrd>>. Acesso em: 15 abr 2017.
- ROUSE, M. *warehouse management system (WMS)*. 2009. Disponível em: <<https://goo.gl/sgh88Y>>. Acesso em: 15 abr 2017.
- ROUSE, M. *SOAP (Simple Object Access Protocol)*. 2014. Disponível em: <<https://goo.gl/CYyXtD>>. Acesso em: 17 mai 2017.
- SCHROCK, N. *GraphQL Introduction*. 2015. Disponível em: <<https://goo.gl/OYsnm0>>. Acesso em: 15 abr 2017.
- SCHROCK, N. *GraphQL Overview*. 2015. Disponível em: <<https://goo.gl/wX2VJ8>>. Acesso em: 15 abr 2017.
- SCHUSSEL, G. *Client/Server: Past, Present and Future*. 2001. Disponível em: <<http://ciains.info/elearning/Solutions/Architecture/ClientServer/CS-past,presentFuture.pdf>>. Acesso em: 02 mai 2017.
- SKONNARD, A. *Understanding SOAP*. 2003. Disponível em: <<https://goo.gl/1G2yiz>>. Acesso em: 17 mai 2017.
- SUN MICROSYSTEMS. *RPC: Remote Procedure Call Protocol Specification Version 2*. [S.l.], 1995. Rev. 3.

VNI, C. M. *Cisco Visual Networking Index*. 2017. Disponível em: <<https://goo.gl/MHsDgQ>>. Acesso em: 15 out 2017.

WILSON, P.; MADSEN, L. *The Memoir Class for Configurable Typesetting - User Guide*. Normandy Park, WA, 2010. Disponível em: <<http://mirrors.ctan.org/macros/latex/contrib/memoir/memman.pdf>>. Acesso em: 19 dez. 2012.

WORK, S. *How Loading Time Affects Your Bottom Line*. 2011. Disponível em: <<https://goo.gl/xUpdJ8>>. Acesso em: 15 out 2017.