

# Processamento de Dados Massivos em Nuvem

## Relatório Computação em Nuvem

*Estudo de Caso:*

*LinkSaúde Tecnologia da Informação S/A*

Ráfagan Soares C. S. Mariz  
Vinícius de Oliveira Silva

Matrícula: 2012049936  
Matrícula: 2013007820

Data de entrega: 11/12/2017

## Introdução

O trabalho a seguir consiste em exemplificar uma aplicação de alguns dos conceitos de computação em nuvem vistos em sala durante as aulas da disciplina “Processamento de Dados Massivos em Nuvem” ofertada no segundo semestre de 2017 pelo DCC-UFMG.

A proposta de elaborada pelos autores consistia em descrever o funcionamento e a infraestrutura por trás de uma aplicação em nuvem real e funcional que impacta diretamente na vida de seus usuários, oferecendo comodidades pelas quais a maioria de seus clientes está satisfeita em pagar.

A aplicação escolhida como objeto de estudo deste trabalho é a ferramenta “LinkSaúde - Office” que é uma solução online para médicos fazerem cadastro de seus pacientes e emitirem receitas, prontuários, pedidos de exames, etc. A vantagem dessa solução com relação ao cenário corrente na indústria é que o LinkSaúde Office armazena todos os dados inseridos na nuvem, oferecendo interfaces web e mobile para acesso simples a tais informações.

Através do uso da plataforma, um médico pode, por exemplo, cadastrar um paciente em um hospital em que atende e depois acessar o mesmo cadastro em sua clínica particular. Para o paciente, também é vantajoso, já que, se ele já foi cadastrado por um médico que utiliza o LinkSaúde, não precisará cadastrar-se novamente em um outro consultório que também faz uso da plataforma. O paciente tem um cadastro único com todas as suas ocorrências de saúde, o que leva a uma centralização das informações, gerando assim mais praticidade e eficiência nas consultas e procedimentos ligados ao bem estar das pessoas.

Essa ferramenta foi escolhida porque o aluno Vinícius Silva estagiou por 4 meses na empresa responsável pelo desenvolvimento da plataforma, tendo contribuído ativamente na sua construção. Acredita-se que a ferramenta seja uma representação típica dos casos de uso de uma plataforma em nuvem, tendo enfrentado durante toda a sua construção os principais desafios na área, como balanceamento de carga, redundância dos dados, garantia de disponibilidade, etc. Este trabalho tem, portanto, o objetivo de mostrar como tais desafios são superados e quais as vantagens que têm sido obtidas em decorrência da escolha de manter um serviço como esse na nuvem.

## Desafios

### **Redundância:**

A empresa LinkSaúde Tecnologia da Informação S/A presta serviços para profissionais da área da saúde, e por isso deve submeter-se a legislação específica para esse fim. A lei obriga que todos aqueles que armazenam informações de prontuários médicos mantenham backup dos dados, e por isso, a LinkSaúde deve se assegurar que todos os dados inseridos diariamente na plataforma possuam cópias disponíveis para acesso 24h por dia.

### **Disponibilidade:**

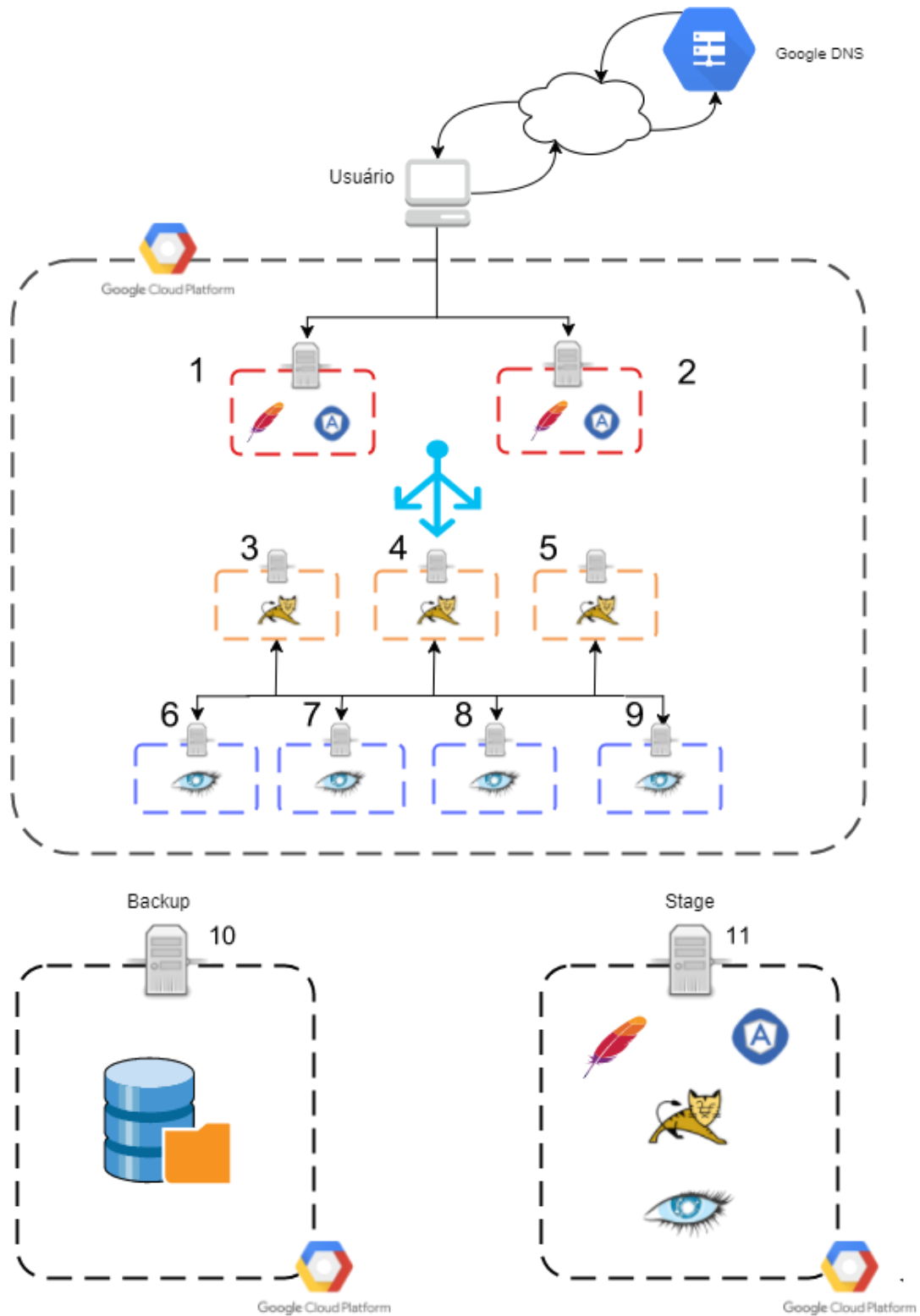
Uma eventual indisponibilidade do sistema da LinkSaúde significa também uma indisponibilidade de acesso a dados que podem salvar uma vida. Significa que médicos não contarão com todas as ferramentas uma vez disponíveis para atender um paciente da melhor maneira possível e isto é extremamente grave. Por esses motivos, a solução adotada pela empresa deve ficar disponível tanto tempo quanto for possível, servindo aos médicos que pagam por esse serviço.

### **Balanceamento de Carga:**

O LinkSaúde é um serviço por assinaturas desenvolvido e mantido por uma pequena startup de Belo Horizonte que, pelo menos a princípio, não conta com uma base muito extensa de clientes. Levando em conta apenas este fato, um indivíduo corre o risco de pensar que não há, portanto, necessidade de utilizar um balanceador de carga, uma vez que um único servidor simples é perfeitamente capaz de servir a aplicação para toda a clientela. Por outro lado, se voltarmos à questão da disponibilidade, podemos perceber que faz sentido termos diversas instâncias do servidor de aplicação rodando simultaneamente em servidores separados, de modo que, caso por alguma eventualidade algum deles falhe, os usuários simplesmente sejam redirecionados para os demais e sequer sejam impactados pelo problema.

## Arquitetura

A arquitetura projetada pela equipe de desenvolvedores da LinkSaúde é baseada em uma estrutura que pode ser representada da seguinte forma:



## Funcionamento

Conforme mostra a representação gráfica da arquitetura exibida na seção anterior, o cluster da LinkSaúde conta com 11 máquinas organizadas de forma a prover a eficiência e a robustez que uma aplicação para a área da saúde exige. Esta seção dedica-se a detalhar o funcionamento do cluster e como ele possibilita que a aplicação “LinkSaúde - Office” esteja sempre disponível para os mais de 50 clientes que fazem uso da plataforma.

Como todo serviço web, para que o cliente consiga ter acesso, é necessário que a sua máquina faça uma requisição (request) para um servidor na Internet que hospeda o conteúdo do serviço a ser acessado. No entanto, para que seja possível fazer a requisição para o servidor correto, o cliente precisa conhecer o endereço IP do mesmo, e como pode-se facilmente perceber, esperar que os usuários conheçam este endereço (composto por uma sequência de 16 algarismos arbitrários) é uma decisão pouco inteligente. Dessa forma, é extremamente comum que as empresas que operam na Internet utilizem aliases para seus endereços, requerendo que seus clientes memorizem um endereço muito mais amigável do que uma sequência numérica arbitrária.

Para fazer isso, as empresas contratam um serviço denominado DNS (Domain Name Server) que funciona como uma espécie de dicionário, traduzindo um nome (alias) para um endereço IP.

No caso da LinkSaúde, a empresa contratou o serviço de DNS do Google para associar o domínio “linksaude.com” ao endereço IP dos servidores numerados como 1 e 2 na figura. Ambas as máquinas estão configuradas exatamente da mesma forma e servem rigorosamente o mesmo conteúdo, fazendo com que, independentemente de qual servidor for acessado pelo cliente, a requisição será respondida exatamente da mesma forma.

Instâncias do servidor Apache estão instaladas nessas máquinas. Ambas estão configuradas com os módulos Proxy, Load Balancer e têm o objetivo de responder às requisições feitas pelo cliente com um pequeno aplicativo Angular que executará em seu browser. Além disso, essas máquinas também têm a função de receber requests HTTP provenientes das próprias aplicações Angular e redirecionar esses requests a um dos três servidores de aplicação (máquinas enumeradas de 3 a 5) disponíveis. Quando um desses requests chega a uma das máquinas balanceadoras, a política de distribuição Round Robin é utilizada para escolher qual servidor de aplicação atenderá àquele cliente solicitante. Uma vez que um servidor é escolhido, as máquinas de load balancing tentam fazer com que as próximas requisições provenientes daquele cliente sempre sejam atendidas pelo mesmo servidor de aplicação. Apesar deste comportamento não ser estritamente necessário (já que os servidores de aplicação executam uma aplicação REST), do ponto de vista de performance, existem vantagens que devem ser consideradas.

Os servidores de aplicação, que efetivamente respondem às requests dos clientes, executam instâncias do Apache Tomcat que servem de containers para uma aplicação Java que é responsável por implementar as regras de negócio do serviço.

Uma outra característica particularmente interessante das máquinas balanceadoras é o fato de que elas utilizam o esquema de *partition keys* do Banco de Dados Cassandra (instalado nas máquinas enumeradas de 6 a 9) que as permitem identificar prontamente qual destas máquinas contém a informação requisitada. Dessa forma, é possível minimizar o tempo gasto no atendimento à request e fornecer uma experiência de uso mais agradável ao cliente.

A solução Apache Cassandra foi escolhida para prover o serviço de persistência devido ao fato de que essa ferramenta foi proposta para conseguir atender um número elevado de usuários conectados e fornece portanto, um desempenho comparável às melhores alternativas do mercado. A opção de montar uma estrutura composta por 4 máquinas servindo o banco de dados foi adotada porque a Sociedade Brasileira de Informática para a Saúde exige que os dados de prontuários médicos sejam armazenados com redundância tripla, isto é, os dados devem estar guardados em pelo menos três máquinas. Desta forma, foi considerado que uma dessas máquinas poderia eventualmente falhar e uma extra foi adicionada por precaução, totalizando assim as 4 máquinas que vemos na figura.

A máquina de número 10 é utilizada para backup dos dados e está programada para executar uma rotina CRON a cada madrugada que faz, via rede, um requerimento de dump nos servidores de banco de dados e armazena um arquivo compactado contendo toda a base de dados do LinkSaúde. Esse procedimento é feito de forma a evitar uma perda completa dos dados armazenados nos servidores Cassandra e sua execução é realizada em horários de baixa utilização do sistema para que o workload extra nas máquinas não afete a experiência de uso dos clientes.

Já a máquina de número 11, chamada de “stage” é um servidor utilizado apenas para fins de desenvolvimento e possui uma cópia completa de todo o stack de funcionamento da aplicação. Em versões experimentais dos sites e aplicativos desenvolvidos pela empresa, é o “stage” quem responde às requisições feitas. Modificações realizadas na aplicação Java executada nos Tomcats também são testadas neste servidor antes de serem colocadas em produção. Essas medidas são adotadas para preservar os clientes de serem impactados por eventuais erros presentes em aplicações em desenvolvimento e também para manter os dados no banco a salvo de procedimentos errôneos que poderiam comprometer a sua consistência.

## Histórico e Decisões Estratégicas

Ao projetar e conceber uma empresa cujo ramo de atuação é prover um serviço online, vários fatores devem ser considerados do ponto de vista estratégico, e nesta seção pretende-se abordar alguns deles.

De modo a entender as tomadas de decisão da empresa, é preciso, inicialmente, contextualizar o desenvolvimento do sistema. A LinkSaúde Tecnologia da Informação S/A é uma pequena startup de Belo Horizonte fundada por Thiago Madureira Braga, um empreendedor mineiro egresso do curso de Engenharia de Controle e Automação da UFMG e do Mestrado em Ciência da Computação da UFRJ.

Durante sua pós graduação, Thiago desenvolveu um estudo em Mineração de Dados em textos aplicado dados médicos e percebeu que seria possível criar uma solução empreendedora a partir de seus esforços na área. O objetivo seria cruzar e minerar os dados de vários prontuários e laudos de um paciente e fornecer insights para os médicos no sentido de diagnosticar e tratar as enfermidades dos pacientes.

Ao concluir seu curso no Rio de Janeiro e voltar para Belo Horizonte, Thiago reuniu recursos e deu início às atividades da empresa objeto de estudo deste trabalho. A empresa começou oferecendo serviços básicos de cadastro de pacientes, fichas médicas, prontuários, receitas e outras informações relevantes para a área da saúde. O objetivo era conquistar clientes para sustentar o desenvolvimento da empresa e agregar uma sólida base de dados que passaria por técnicas avançadas de processamento e análise, com o objetivo final de fornecer inteligência computacional para ajudar os médicos a exercer sua função com eficiência e qualidade.

Tendo esses objetivos em mente, considerando a dimensão da empresa e a formação de seu fundador, a tarefa de entender os motivos por trás de cada decisão da empresa torna-se mais simples. Como a LinkSaúde é uma microempresa com recursos capitais e humanos limitados, foi planejado utilizar uma plataforma de software que tivesse fácil desenvolvimento e pudesse ser mantida e evoluída por poucas pessoas. Dessa maneira, a tecnologia Java foi escolhida para ser a base de sustentação do stack da empresa, uma vez que é uma plataforma gratuita, com grande participação no mercado e cuja oferta de profissionais é ampla.

Tendo essa escolha em mente, a decisão das demais soluções tecnológicas que fariam parte do sistema foi natural. Optou-se por utilizar soluções Apache (como HTTP Server, Tomcat e Cassandra) por esta ser uma organização respeitada e cuja política de licenças é levada extremamente a sério. Esta

questão é de fundamental importância para a empresa, uma vez que a sobrevivência da mesma no mercado depende de sua habilidade de utilizar os softwares que necessita, sem mudanças repentinas no licenciamento.

Uma outra questão relevante que foi considerada na escolha das tecnologias a serem utilizadas é a sua compatibilidade e aceitação dentro do contexto Linux / Open Source. O sistema operacional adotado em todas as máquinas da empresa (tanto do cluster quanto as de desenvolvimento) é o Debian, uma distribuição GNU / Linux bastante sólida e consolidada e que foi escolhida devido ao seu destaque dentre as alternativas de sistemas operacionais gratuitos no mercado.

## Análise de Custo

A plataforma do LinkSaúde foi pensada para ter um custo competitivo, mas sem sacrificar a qualidade dos serviços prestados para o cliente. Desta forma, os projetistas da aplicação compararam a relação custo-benefício de diversas soluções de hospedagem e chegaram à conclusão que o Google Cloud Platform era a alternativa que melhor atenderia às necessidades da empresa.

A arquitetura foi implementada através do uso de máquinas virtuais Debian contratadas diretamente do Google dispondo apenas do sistema operacional instalado. Todas as tecnologias necessárias para fazer com que o sistema funcione tiveram que ser manualmente instaladas em seus respectivos servidores e nenhum serviço específico da plataforma do Google foi utilizado. Essa decisão foi tomada em nome da flexibilidade, uma vez que o fundador da empresa optou por manter a possibilidade de migrar da solução do Google para outras plataformas concorrentes (Amazon WS, por exemplo), caso lhe fosse conveniente no futuro.

As máquinas foram contratadas com um único núcleo de processamento, 1TB de armazenamento de dados (compartilhado por todo o cluster) e 200GB alocados para a máquina de stage.

O custo mensal de se manter uma solução como essa no ar, 24h por dia, conforme o modelo adotado pelo LinkSaúde, gira em torno de R\$100,00 por máquina, o que totaliza aproximadamente R\$1100,00 por mês para manter todo o sistema funcionando. Um outro parâmetro que o Google Cloud permite que o contratante especifique é a região geográfica do Data Center que abrigará as máquinas contratadas. A região escolhida pelo LinkSaúde foi a Central-US, na região central do território dos Estados Unidos.



## Prova de conceito

Como o “LinkSaúde Office” é uma aplicação funcional e de código fechado, apesar do aluno Vinícius Silva ter participado ativamente de sua construção, não é possível obter seu código fonte e entregá-lo como uma demonstração. Desta forma, para fornecer um exemplo básico de funcionamento da aplicação no contexto deste trabalho, uma conta de teste gratuita foi criada e disponibilizada para que o leitor possa explorar, por si, as funcionalidades implementadas até o momento.

Para que seja possível fazer isso, o leitor deve visitar a página <http://office.linksaude.com> e fazer login utilizando as credenciais abaixo:

E-mail: [pdmn.nuvem@gmail.com](mailto:pdmn.nuvem@gmail.com)

Senha: [tppdmn20172](#)

O leitor é encorajado a navegar pelos menus, simular alguns dos casos de uso e, se possível, fornecer feedback para o administrador do site, que pode ser contatado no endereço [thiago@linksaude.com](mailto:thiago@linksaude.com).

Apenas para efeito de completude, alguns screenshots da aplicação são fornecidos abaixo:

The screenshot displays the LinkSaúde application interface. At the top, a dark blue header bar contains the LinkSaúde logo on the left and the user name 'Rafagan-Vinicius' with a profile icon on the right. Below the header, the interface is divided into three main sections. On the left is a sidebar with a green 'CADASTRAR PACIENTE' button at the top. Below it is a search bar labeled 'Procurar pelo nome' with the text 'teste' entered. A list of patients follows, including Rosilene Arantes Teixeira, Gislene Arates Teixeira, Teste, Mônica Tostes Teixeira da Silva, Paciente Teste1, Paciente Teste2, Paciente de Teste, and Meu Usuario Teste. The main area on the right features a large profile picture of a woman, a 'REMOVER' button, and a section for 'Informações essenciais' (Essential Information) which includes fields for Name, Birthdate, E-mail, Celular, Sexo, Sangue, CPF, and Endereço. A 'PRONTUÁRIO' (Medical History) button and a 'SALVAR' (Save) button are also visible. At the bottom right, there is a red circular button with a white camera icon.

Tela de Cadastro de Pacientes

Link Saúde		Rafagan-Vinicius	
Consultório		Rafagan-Vinicius	13/12/2017
Página Inicial		08:00	08:30
Cadastro de pacientes		09:00 - Jessica Costa Arantes	09:30
Agenda de marcações		10:00	10:30
Tarefas e lembretes		11:00	11:30 - Ismael Fernando Poli Villas Boas Consulta simples (Consulta de rotina)
		12:00	12:30
		13:00	13:30

*Agenda de Marcações*

## Conclusão

O objetivo deste trabalho era que os alunos propusessem uma aplicação em nuvem fictícia e descrevessem seu funcionamento básico, enunciasssem os desafios que teriam de ser resolvidos para manter o correto funcionamento da aplicação e analisar os custos decorrentes dessa operação. Essa era uma proposta bastante interessante, mas por ser puramente teórica, era difícil ter uma noção realista do cenário vivido por quem realmente implementa um serviço em nuvem na prática. Por esse motivo, decidiu-se por abordar o desafio desta proposta de uma maneira um pouco mais aprofundada e entrar em contato com pessoas da indústria que sustentam seus negócios através do desenvolvimento de software para a web.

Uma entrevista por videoconferência foi feita com o empresário Thiago Braga, fundador da LinkSaúde Tecnologia da Informação S/A, e, através deste contato, foi possível extrair uma quantidade razoável de informação sobre o que tem sido feito na área nos últimos anos. Acredita-se que houve um aprendizado diferenciado e bastante considerável na área, já que o escopo deste trabalho permitiu obter-se uma maior integração Universidade-Empresa.

## Referências

LinkSaúde Tecnologia da Informação S/A

<http://linksaude.com>

Acessado em 08/11/2017

LinkSaúde Office

<http://office.linksaude.com>

Acessado em 08/11/2017

Google Cloud Platform

<https://cloud.google.com/products/calculator/>

Acessado em 06/11/2017