

A Utilização de Sistemas Distribuídos no Desenvolvimento de Material Instrucional para a Inclusão Digital do Projeto *Conhecendo o Linux*

The use of Distributed Systems on the Development of Instructional Material for Digital Inclusion of the Project *Knowing Linux*

Wesley Guilherme Lage^{a*}; Israel Nogueira Alves^a

^aFaculdade Anhanguera de Belo Horizonte, MG, Brasil

*E-mail: feniuxnb@outlook.com

Resumo

O artigo aborda o processo de desenvolvimento de material instrucional utilizando técnicas de elaboração de material didático no projeto *Conhecendo o Linux*, por meio de soluções distribuídas, bem como os conceitos e paradigmas que permeiam a doutrina dos sistemas distribuídos, seus objetivos e meios utilizados para alcançar expressividade em novas tecnologias e soluções escaláveis que viabilizam o desenvolvimento de material de forma colaborativa e interativa. Este trabalho esclarece pontos importantes sobre os sistemas distribuídos, sua principal característica e funcionalidades, os aspectos que tornam viável sua utilização e suas vantagens. Esclarece os trâmites intermediários de desenvolvimento utilizados mediante ferramentas e funções específicas do sistema, disponibilidade de acesso, portabilidade e versatilidade apresentada pelo mesmo e exprime os resultados atingidos no decorrer do desenvolvimento do material, bem como os reflexos propagados no público-alvo. A conclusão a que se chega é a de que a utilização de soluções distribuídas para o desenvolvimento de materiais instrucionais ou para diversos outros trabalhos é, de fato, recomendada. Sistemas distribuídos como um todo têm uma forma de comunicação complexa, mas que proporciona uma sincronia e a replicação dos dados trabalhados de tal forma a permitir que diversas pessoas trabalhem em conjunto na manipulação das informações, de diversos locais distintos e simultaneamente, com grande interatividade. Tendem também a fornecer maior segurança quanto aos dados armazenados em nuvem.

Palavras-chave: Google Drive. Desenvolvimento de Apostilas. Google File System. Bigtable. Single-Sign On.

Abstract

The paper comprises the development of instructional material using techniques of didactic material elaboration at the Project Knowing Linux, through distributed solutions, as well as the concepts and paradigms which permeate the distributed systems doctrines, their goals and means used to reach expressiveness in new Technologies and scalable solutions which make it available the development of material in a collaborative and interactive way. This paper clarifies important points on distributed systems, their main characteristics and functionalities, the aspects which make it available its using and advantages. It clarifies the intermediate development handlings used by means of tools and specific functions of the system, access availability, portability and versatility showed by the same and it also presents the results reached throughout the material development, as well as the reflections propagated at the target-public. The conclusion reached is that the using of distributed solutions for the instructional materials development or for several other papers is, in fact, recommended. Distributed systems on the whole have a complex communication way, but they also offer a synchronicity and the worked data replication in a such way that it allows that several people work together at the information manipulation, from a number of distinct places and simultaneously, with a great level of interactivity. They also tend to provide a better safety concerning the data stored in a cloud.

Keywords: Google Drive. Handouts development. Google File System. Bigtable. Single-Sign On.

1 Introdução

Um sistema distribuído, segundo Tanenbaum (2007), define-se como “[...] um conjunto de computadores independentes que se apresentam aos seus usuários como um sistema único e coerente”. Ainda podemos considerá-lo como um conjunto de componentes de *hardware* ou *software*, localizados em computadores interligados em rede, que se comunicam e coordenam suas ações apenas enviando mensagens entre si (COULOURIS; DOLLIMORE; KINDBERG, 2007).

Este conceito é muito interessante para a aplicação em desenvolvimento colaborativo de material para quaisquer fins, uma vez que proporciona colaboratividade de forma escalável. A tecnologia empregada em soluções distribuídas permite aos desenvolvedores engajados em um projeto desenvolver

os materiais desejados em qualquer lugar, simultaneamente, a qualquer momento e o mais importante: proporcionando aos desenvolvedores maior mobilidade, portabilidade, interatividade e um padrão de desenvolvimento a ser seguido.

Computadores autônomos contribuem em cada parcela do sistema, pois as aplicações são tratadas em camadas, seguindo a característica de distribuição de dados atrelada à tecnologia.

Contextualmente apropriou-se a utilização de um sistema distribuído da empresa Google, o *Google Drive*, no desenvolvimento do material instrucional para a inclusão digital promovida pelo projeto *Conhecendo o Linux*, que visa promover a inclusão digital das pessoas nas escolas públicas de Belo Horizonte, difundindo o conhecimento sobre o sistema operacional Linux (RIBEIRO; MUNIZ, 2012). Segundo Cura (2012), o sistema *Google Drive* segue o princípio de

colaboratividade e computação em nuvem.

Baseando-se no conceito de distribuição de tarefas na nuvem, podem-se ramificar as funções através de vários componentes, ou seja, distribuir as aplicações por diversos computadores e assim seguir um cronograma que permita a realização satisfatória da maioria dos projetos adotados, aninhando elaboração simultânea com integridade de conteúdo. O sistema *Google Drive* ramifica suas funcionalidades em aplicações distintas, mas as apresenta em conjunto num sistema único. Não somente os computadores pessoais atrelados ou autenticados à rede, mas também os computadores servidores são suscetíveis a essa premissa.

Para as soluções distribuídas é implementada uma tecnologia de comunicação entre os processos na rede, a fim de que um único domínio seja apresentado como um servidor, mas opere buscando informações em várias partes atreladas à sua rede configurada. Ao mesmo tempo, preocupa-se com a transparência da comunicação das informações entre computadores ou usuários que as estão solicitando dos servidores.

De forma geral, independente de como o servidor funcione, a resposta de um sistema distribuído deve sempre ser homogênea e íntegra. A fim de esclarecer como o processo de desenvolvimento do material descrito no início deste artigo decorre, serão abordados aspectos da solução distribuída em questão.

Este trabalho tem por objetivo estudar novas tecnologias e soluções distribuídas para viabilizar o desenvolvimento de material didático de forma colaborativa.

Os objetivos específicos contemplam alcançar a medição distribuída do status do desenvolvimento de projetos e materiais instrucionais, bem como seguir um controle de demanda de fluxo, otimizar a produtividade e o tempo hábil de desenvolvimento, propondo e utilizando novas tecnologias para produzir o material e documentação e mantê-los.

2 Material e Métodos

2.1 Metodologia

No primeiro tópico do desenvolvimento, será discutido como os sistemas distribuídos se comportam na transmissão de dados ao se comunicarem e como se dispõe o comportamento do *Google Drive*; no segundo tópico do desenvolvimento será retratada a forma de armazenamento e replicação que o sistema *Google Drive* adota; nos resultados será abordada a experiência de utilização do *Google Drive* e os impactos causados ao público-alvo pelos trabalhos realizados; e, após serem expressos tais resultados, concluiu-se que, no decorrer da produção deste trabalho, o próprio sistema distribuído em questão foi utilizado pelos alunos desenvolvedores da

pesquisa. Todos os acessos às obras consultadas foram mantidos no *Google Drive* e todo o material deste trabalho foi desenvolvido utilizando-se as ferramentas do *Google Drive*.

2.2 Comunicação e transmissão de dados

A comunicação entre sistemas distribuídos acontece seguindo várias premissas, desde níveis mais baixos de protocolação para transferência de dados. Segundo Tanenbaum (2007): “Antes de trocar dados, o remetente e o receptor primeiro estabelecem explicitamente uma conexão e possivelmente negociam o protocolo que usarão. Após concluírem, devem liberar a conexão”.

Em suma, protocolos de rede formam toda a base para comunicação entre sistemas distribuídos. Esta comunicação acontece em baixo nível, acompanhando uma espécie de “cronograma” de envio e recebimento de mensagens de acordo com a programação feita para o sistema distribuído em questão sobre os protocolos estabelecidos na rede. Utilizam-se algoritmos de troca de mensagens; processos montam solicitações e enviam estas solicitações a outros processos.

Considerando que estamos abordando os padrões de um sistema distribuído, é bem aceitável que a comunicação entre os processos não seja a mesma como em uma única máquina. Entretanto, os processos se comunicam de forma parecida, não obstante seguindo uma tecnologia que coordena a transmissão de dados orientada a mensagens, segundo Tanenbaum (2007). O autor menciona que a transmissão orientada a mensagens pode ocorrer de forma persistente, quando a mensagem é armazenada no sistema e entregue, independente da aplicação *send e receiver* estar rodando; ou transiente, quando a mensagem só é armazenada e enviada caso o algoritmo *send e receiver* esteja rodando - entretanto, é possível a comunicação através de repassadores do tipo *armazena-e-reenvia*. Caso um repassador não consiga, por algum motivo, entregar a mensagem ao próximo, a mesma simplesmente é destruída; torna-se síncrona, quando o processo fica bloqueado até ter a confirmação, positiva ou não, da entrega da mensagem; ou assíncrona, quando o remetente continua sua execução imediatamente após ter apresentado sua mensagem para entrega, esta ficando temporariamente armazenada no *middleware*. Esta orientação acontece com embasamento em chamadas de procedimento remoto, ou RPC¹ (Remote Procedure Call), que são utilizadas em soluções distribuídas para atingir alto nível de transparência na comunicação entre cliente-servidor (TANENBAUM, 2007).

O computador cliente e o servidor se comunicam como em uma chamada de procedimento local, entretanto utilizando apêndices de cliente/servidor. Uma rotina é extraída da biblioteca local e inserida no programa objeto através da pilha

¹ Remote Procedure Call - São procedimentos chamados remotamente de forma específica a permitir que o computador requisitante do serviço (computador cliente) e o computador fornecedor (servidor) continuem a execução instantânea do programa mesmo sem a confirmação de entrega da resposta ou requisição.

de processos em execução, a fim de conduzir o sistema a agir como se a chamada ao referido procedimento obtivesse sua resposta localmente (TANENBAUM, 2007).

A empresa Google implementa em seus sistemas um sistema de arquivos chamado *Google File System* (GHEMAWAT; GOBIOFF; LEUNG, 2003), projetado para armazenamento em blocos de 64 MB, especificamente para alta performance e escalabilidade, a fim de promover o suporte adequado aos seus aplicativos na nuvem, fornecendo recursos indispensáveis para o funcionamento colaborativo, escalável e otimizado destes.

O *Drive* permite que diversos usuários acessem o mesmo documento de qualquer lugar, a qualquer momento, e fornece uma plataforma adequada para que estes usuários possam simultaneamente trabalhar no documento ou projeto em questão. Ao editar um documento, o *Drive* permite ao usuário a possibilidade de ver outros usuários online no mesmo documento, e segundo Ghemawat, Gobioff e Leung (2003), através do registro de acréscimo, o *Google File System* permite que um usuário edite o documento, sem bloqueio do mesmo, envie mensagens privadas a outro usuário e as modificações são visualizadas simultaneamente por todos logados. O envio e recebimento de informações, tanto diretamente ao documento quanto por mensagens privadas, é íntegro. Arquiteturalmente, essa interatividade é possível devido à atividade e boa comunicação de quatro camadas dos protocolos de rede (baixo nível) - física, enlace, rede e transporte - e aplicações de middleware, que complementam a função de comunicação com API específicas (TANENBAUM, 2007).

O referido registro de acréscimo é um controle de versão das edições realizadas anteriormente, chamado “histórico de revisões”. A cada alteração realizada, o documento é salvo por uma *thread* distinta, um trecho ou parte de código que roda dentro do contexto do processo em execução (MACHADO; MAIA, 2007). Segundo Machado e Maia (2007), “um ambiente *multithread* permite a execução concorrente de sub-rotinas dentro de um mesmo processo”. Segundo Tanenbaum (2007), a principal utilização da prática *multithreading* é no lado dos servidores, que precisam executar tarefas concorrentes. Neste contexto, o *Google File System* salva metadados com as informações da versão anterior à edição realizada. Se a edição é feita ininterruptamente, o histórico terá versões salvas a cada minuto. Caso seja uma edição paulatina, o histórico apresentará versões por período de tempo.

2.3 Replicação e armazenamento

Os sistemas Google são projetados para utilizar um banco de dados chamado *Berkeley DB* (BDB-HA) idealizado pela empresa Oracle (PERL; SELTZER, 2006). Segundo Perl e Seltzer (2006), este é um banco de dados integrado com um modelo de replicação que é compatível com o algoritmo de replicação *Paxos*, um algoritmo projetado para resistir a falhas, implementado em sistemas distribuídos como o *Google Drive*,

a fim de garantir que a replicação e sincronização dos dados seja realizada de forma adequada; funciona propondo valores para solicitações, valores estes que sejam aceitos pela maioria dos processos envolvidos. O algoritmo não considera as falhas, apenas inicia ciclos de transmissão de valores quando estas ocorrem, garantindo concordância e validade dos dados (PRISCO; LAMPSON; LYNCH, 2000), indispensáveis para a replicação. Por ser um sistema distribuído, é importante que o *Google Drive* se utilize de tal premissa, adequada ao ambiente de produção altamente distribuído e pouco convencional da Google.

O *Google Drive* conta com uma primorosa sincronia de dados, em grande parte devido ao serviço de replicação, sincronização e armazenamento e à utilização da *BigTable* (CHANG *et al.*, 2006). Segundo os autores, o *BigTable* é um sistema proprietário para direcionar armazenamento em massa projetado para gerenciar a manipulação de arquivos de grande porte de forma escalável e de alta disponibilidade, garantindo que os dados dos usuários na rede da Google sejam administrados e dimensionados de forma distribuída e confiável em milhares de máquinas servidoras espalhadas em todo o mundo.

Utilizando o conceito não relacional (NoSQL), é orientado a colunas (tuplas com nome, valor e *timestamp*). Este sistema é implementado em mais de 60 produtos da Google, como por exemplo, o *Google Drive*, *Google Analytics* e *Google Earth* (CHANG *et al.*, 2006).

O *BigTable* proporciona aos usuários um controle dinâmico de seus dados e até mesmo a possibilidade de executar serviços que consomem maior carga de processamento e transmissão em rede, sem que haja considerável perda de latência de dados na apresentação dos mesmos aos usuários finais.

A Google, de forma geral, se comporta com um único login, com o *Single Sign-On* (PERL; SELTZER, 2006), que integra todos os *apps* da empresa, a fim de evitar que o usuário tenha que reinserir seus dados de autenticação toda vez que necessite acessar uma aplicação. Segundo Mittal (2012), o *Single Sign-On* torna possível o controle de acesso a múltiplas aplicações de forma consistente, garantindo sincronia dos dados.

O *Google Drive* mantém alta sincronização com outros usuários mantidos na rede. Quando são realizadas gravações em um documento, as gravações são direcionadas para o servidor mestre. O mestre aplica as atualizações e, em seguida, propaga as atualizações para os registros do banco de dados que correspondem às réplicas, as quais se aplicam às mudanças, como se as atualizações estivessem literalmente “correndo” (PERL; SELTZER, 2006).

Em outras palavras, ao gravar um dado em um documento no *Drive*, este dado é encaminhado ao servidor mestre, que replica o dado através dos métodos do banco de dados, gravando em réplicas nos servidores através do mundo, por meio da tabulação e redimensionamento de dados realizado pela *BigTable*. Com isso, garante-se escalabilidade,

disponibilidade, estabilidade, usabilidade, sincronização, segurança dos dados, consistência e tolerância a falhas; esta última em grande parte devido à implementação de algoritmos de replicação como Paxos, tolerante a falhas.

3 Resultados e Discussão

Em outubro de 2012, surgiu o projeto chamado *Conhecendo o Linux* (RIBEIRO; MUNIZ, 2012), que visava disseminar o conhecimento do sistema operacional Linux e suas distribuições através de escolas públicas, introduzindo aos alunos que ainda não possuíam conhecimento do mundo digital uma nova experiência. O projeto foi incentivado e orientado por um professor da Faculdade Anhanguera de Belo Horizonte, onde o projeto foi desenvolvido.

Os organizadores do projeto tinham pela frente um grande desafio: realizar os preparativos para ministrar um curso de inclusão digital na Escola Municipal Professora Efigênia Vidigal. Grande parte deste desafio, de acordo com o cronograma estabelecido para tal projeto, consistia em construir uma equipe competente de voluntários à qual fosse possível delegar as diversas tarefas necessárias para o bom cumprimento dos requisitos. Formou-se a equipe de apenas dois voluntários, para a qual foi designada a tarefa de desenvolver as apostilas, elaborando todo o material instrucional que seria utilizado para ministrar as aulas na referida escola.

A princípio os desenvolvedores realizavam suas tarefas comunicando-se por e-mails, enviando as últimas atualizações para cada componente do grupo ou até mesmo utilizando mídias removíveis para o armazenamento e transporte dos dados. O método inicialmente utilizado tornava o trabalho fático e demandava demasiado tempo, uma vez que eram necessárias reuniões quase diárias no mesmo local com todos os integrantes da equipe de desenvolvimento e coordenação, a fim de discutir os parâmetros de elaboração a serem seguidos para a criação das apostilas.

Haviam gastos elevados, com transporte e mesmo alimentação, além do problema com a disponibilidade de horários dos integrantes para presença às reuniões simultaneamente. Também conviviam com o constante perigo da falta de segurança no armazenamento das informações reunidas e desenvolvidas, devido ao fato de utilizarem DVD e pen-drives, frágeis e sujeitos a avarias em transporte.

A equipe de desenvolvimento do projeto, tendo em vista tal situação, se engajou a buscar soluções para suplantar os problemas contraídos através das situações citadas anteriormente. Adotou a utilização do *Google Drive*, abordado neste artigo; solução esta que, para o desenvolvimento adequado e em tempo hábil dos materiais instrucionais, os desenvolvedores acreditam ser uma boa opção. Como mencionou um deles:

Nem sempre temos a disponibilidade de locomoção ou mesmo de tempo para realizar contínuas reuniões para prosseguir com o desenvolvimento. O *Google Drive* permite reuniões virtuais e possibilita que possamos, em lugares distintos, realizar nosso trabalho a qualquer hora, sem perder a linha de raciocínio.

Com a utilização do *Google Drive*, além das vantagens de se utilizar dos recursos variados que o sistema provê e da versatilidade que se obtém, a experiência é única para se realizar um trabalho extenso dentro do pouco tempo disponível, sem perda de conteúdo. O *Google Drive* também garante segurança e acessibilidade adequadas aos dados armazenados, além da possibilidade de *linkar* os dados com outros sistemas, como, por exemplo, o *Trello*², um sistema de gerenciamento de trabalhos e projetos.

Com o controle de versão, ou histórico de revisões, é possível acompanhar o trabalho realizado e verificar todas as edições realizadas anteriormente, com a possibilidade de reverter o quadro do material a qualquer momento sem dificuldades, como mencionado anteriormente. Existe um chat para diálogo ativo simultâneo à edição, onde, junto ao documento, podem-se realizar reuniões e discussões sobre o material em desenvolvimento; e, por permitir que isso seja realizado on-line, no mesmo ambiente de desenvolvimento, aproveita-se melhor o tempo e os recursos para a interação de componentes da equipe para um melhor desenvolvimento e correções. Podem ser inseridos comentários em qualquer parte do documento, “marcando” qualquer palavra ou mesmo objeto com observações, solicitações ou soluções para a correção ou adequação de um determinado fator. Então, mesmo que os outros usuários não estejam on-line, estes poderão, ao ver os comentários posteriormente, realizar as modificações ou mesmo responder com variáveis a serem observadas.

Existem níveis de compartilhamento dos trabalhos com outros usuários. Funciona com permissões. É possível compartilhar um trabalho para outros com permissão de acesso total ao documento. O usuário possuirá a permissão para edição do mesmo. Porém, pode-se compartilhar o documento com permissão apenas para visualização, não permitindo quaisquer alterações no trabalho e, em última instância, compartilhar com permissão para comentar. Esta última permite que o usuário efetue comentários na obra, mas não permite que o mesmo realize quaisquer modificações no trabalho.

A pesquisa causou crescente impacto positivo na comunidade. Os alunos da Escola Municipal Professora Efigênia Vidigal ficaram satisfeitos por poderem receber o material didático preparado pelos alunos voluntários e ficaram empolgados com as aulas. A disponibilização do material desenvolvido, por ser feita via *web*, permite que os alunos possam acessar as últimas alterações realizadas no material, além de permitir também que façam o *download* do mesmo.

² Disponível em: <https://trello.com>

4 Conclusão

O *Google Drive* é uma ferramenta muito útil para o desenvolvimento de projetos e pesquisas, e a elaboração de todo o material didático utilizado para ministrar as aulas do projeto *Conhecendo o Linux* na Escola Municipal Professora Efigênia Vidigal foi viável devido ao fato de o *Google Drive* fornecer boa colaboratividade entre os desenvolvedores.

Quando se trata de projetos a serem desenvolvidos em curto período de tempo e por diversas pessoas interativamente, sistemas distribuídos contribuem em muito para o bom andamento do cumprimento das tarefas delegadas a cada departamento do time engajado no projeto em questão.

O *Google Drive* é um sistema adequado, pois apresenta alta sincronia e replicação dos dados entre os usuários e serviços na nuvem; proporciona também que, simultaneamente, sejam realizadas e visualizadas alterações ou mesmo comentários a respeito da obra realizada.

O sistema da Google proporciona um bom nível de segurança dos dados, utiliza algoritmos tolerantes a erros e se comunica de forma distribuída, de forma a evitar um congestionamento de tráfego da rede em seus servidores, espalhados em todo o mundo.

O controle de versão é especialmente adequado para trabalhos em que se necessita de revisões da obra. O sistema mantém uma “imagem” da edição realizada anteriormente; caso o usuário deseje retornar à versão anterior, não encontrará problemas para fazê-lo.

Apesar de se apresentar de forma convenientemente acessível, o sistema demanda de uma boa conexão com a internet. Problemas de conexão com a rede mundial de computadores podem gerar transtornos ao se trabalhar com sistemas distribuídos, uma vez que os mesmos utilizam o tráfego de rede para replicar as informações aos servidores; especialmente o *Drive*, que replica as informações de

atualizações e também, mediante funcionamento da *Bigtable*, dimensiona os dados de forma homogênea nos servidores da rede Google.

De forma geral, o incentivo à utilização de sistemas distribuídos é pertinente, pois traz benefícios tanto para os desenvolvedores quanto para a própria obra realizada, devido aos diversos recursos disponíveis à aplicação na mesma.

Referências

- CHANG, F. *et al.* *Bigtable: a distributed storage system for structured data*. 2006. Disponível em: <http://static.googleusercontent.com/media/research.google.com/pt-BR//archive/bigtable-osdi06.pdf>. Acesso em: 5 dez. 2014.
- COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T. *Sistemas distribuídos: conceitos e projetos*. São Paulo: Bookman, 2007.
- CURA, H. *Google lança serviço cloud: Google Drive*. 2012. Disponível em: <http://www.pplware.sapo.pt/internet/google-lanca-servico-cloud-google-drive/>. Acesso em: 5 dez. 2014.
- GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S. *The Google file system*. 2003. Disponível em: <http://static.googleusercontent.com/media/research.google.com/pt-BR//archive/gfs-sosp2003.pdf>. Acesso em: 19 jun. 2014.
- KUNAL MITTAL. *Extend Single Sign-On to the cloud*. 2012. Disponível em: <http://www.ibm.com/developerworks/cloud/library/cl-singlesignoncloud/>. Acesso em: 19 jun. 2014.
- MACHADO, F.B.; MAIA, L.P. *Arquitetura de sistemas operacionais*. São Paulo: LTC, 2007.
- PERL, S.E.; SELTZER, M. Harvard: Harvard University & Oracle Corporation. *Data Management for Internet-Scale Single-Sign-On*, 2006.
- PRISCO, R.; LAMPSON, B.; LYNCH, N. *Theoretical computer science. Revisiting the Paxos algorithm*, 2000.
- RIBEIRO, J.A.S.; MUNIZ, J.D. *Implantação do curso de Linux em escolas públicas: projeto conhecendo o Linux*. 2012. Monografia (Bacharelado em Sistemas de Informação) - Faculdade Anhanguera de Belo Horizonte, Belo Horizonte, 2012.
- TANNENBAUM, A.S; STEEN, M.V. *Sistemas distribuídos: princípios e paradigmas*. São Paulo: Pearson/Prentice Hall, 2007.