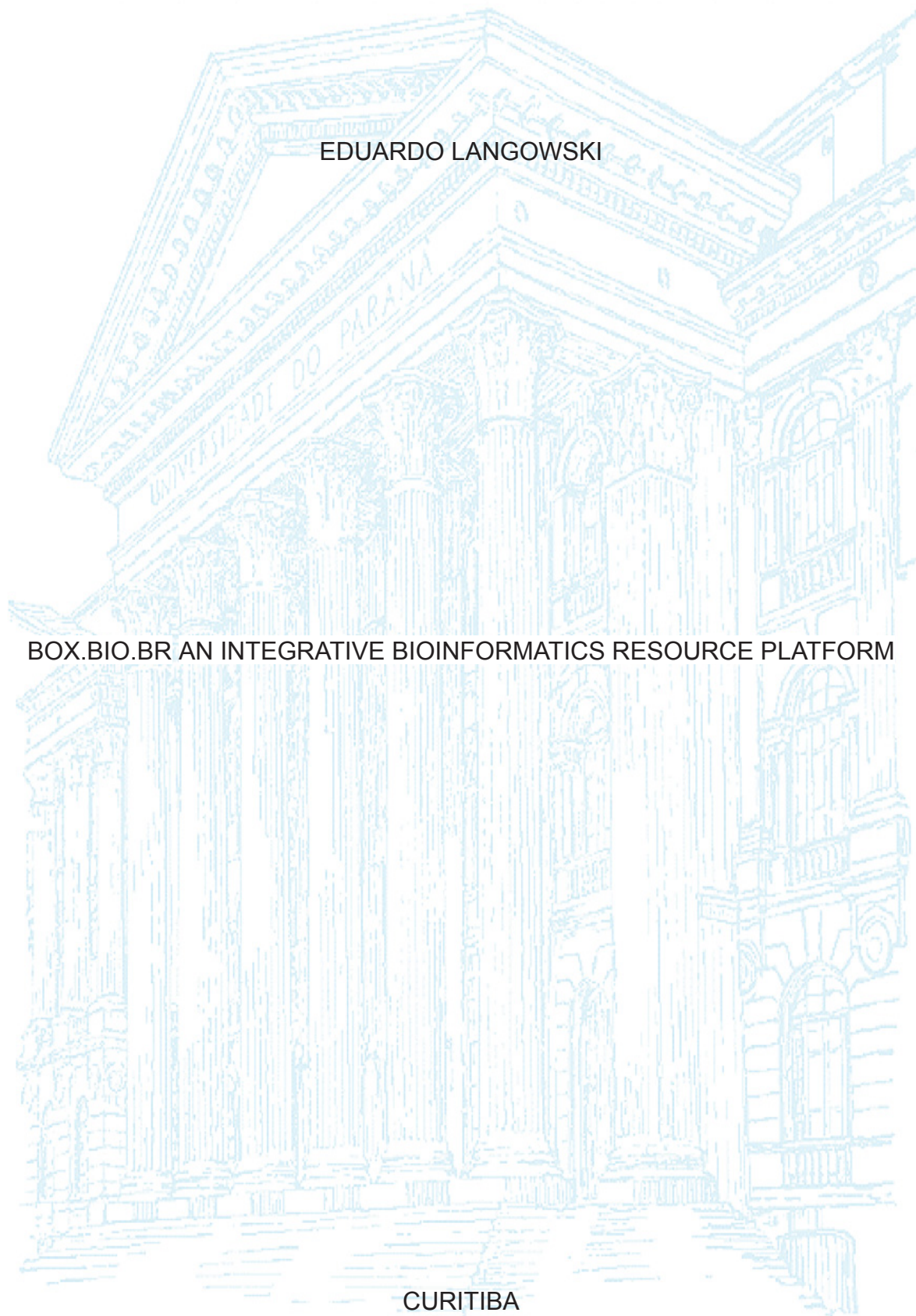


UNIVERSIDADE FEDERAL DO PARANÁ

EDUARDO LANGOWSKI

BOX.BIO.BR AN INTEGRATIVE BIOINFORMATICS RESOURCE PLATFORM

CURITIBA
2015



EDUARDO LANGOWSKI

BOX.BIO.BR AN INTEGRATIVE BIOINFORMATICS RESOURCE PLATFORM

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós Graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná, área de concentração Bioinformática.

Orientador: Professor Dr. Alessandro Brawerman

Coorientador: Professor Dr. Emanuel Maltempi de Souza

CURITIBA
2015

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA

L284b Langowski, Eduardo

BOX.BIO.BR an integrative Bioinformatics resource platform / Eduardo
Langowski. – dados eletrônicos. – Curitiba, 2015.
1 arquivo (33 f.) : PDF.

Requisitos do Sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Orientador: Prof. Dr. Alessandro Brawerman.
Co-orientador: Prof. Dr. Emanuel Maltempi de Souza.

Dissertação (mestrado) – Universidade Federal do Paraná, Setor de
Educação Profissional e Tecnológica, Programa de Pós-graduação em
Bioinformática.

1. Bioinformática. 2. Computação em nuvem. 3. Projetos de pesquisa.
4. Produção científica. 5. Software - Desenvolvimento. I. Brawerman,
Alessandro. II. Souza, Emanuel Maltempi de. III. Universidade Federal do
Paraná. Programa de Pós-Graduação em Bioinformática. IV. Título.

CDD : 570.285

Bibliotecária: Thays Luciana Barbosa de Farias CRB-9/1995

TERMO DE APROVAÇÃO

EDUARDO LANGOWSKI

"BOX.BIO.BR AN INTEGRATIVE BIOINFORMATICS RESOURCE PLATFORM"

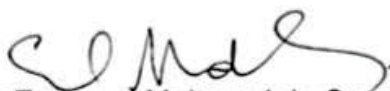
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:

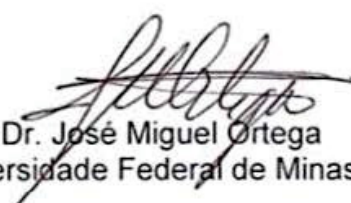


Prof Dr Alessandro Brawerman


Coorientador:



Prof Dr Emanuel Maltempi de Souza



Prof. Dr. José Miguel Ortega
Universidade Federal de Minas Gerais - UFMG



Prof. Dr. Jeroniza Nunes Marchaukoski
Universidade Federal do Paraná - UFPR

Curitiba, 09 de janeiro de 2015

*Aos amados:
Esposa Priscila
Pais Aristeu e Helen
Irmão Rodrigo
Avó Cacilda
e preciosa filha Sophia*

AGRADECIMENTOS

Meus desejos de gratidão são à Deus, por minha existência e por me tornar capaz, à todos os Mestres pela iluminação que sempre me indicam o caminho, aos familiares que me acolhem e me dão suporte nos mais diversos momentos da vida. Aos poucos amigos que como irmãos fazem parte de toda a história que como forja me molda para o homem que sou hoje. Agradeço também aos Professores que de alguma forma insistiram na transmissão de todo o conhecimento adquirido, pois é sabido que por mais nobre e honrada seja esta profissão, poucos são capazes de o fazer como os meus o fizeram. Aos Professores da vida, a quem chamo de Pai e Mãe, minha eterna gratidão por tudo. Ao meu irmão, pela confiança e paciência. Agradeço à minha Esposa, Filha e Parceirinha, por fazer parte do meu Lar e tornar meus dias alegres, embelezando-os com suas presenças, fisicamente e em pensamento. Ao meu Orientador e Amigo, que me mostrou as possibilidades do mundo acadêmico, além de depositar sua confiança em minha capacidade me convidando para as experiências que me fazem hoje ser um novo profissional. Aos colegas, colaboradores e profissionais do setor de Bioinformática da Universidade Federal do Paraná e da Universidade Federal de Minas Gerais, que possibilitaram os trâmites deste meu trabalho. Aos fundos de fomento que tornaram possível esta pesquisa.

Meus sinceros agradecimentos.

"No man is an Iland, intire of itselfe; every man is a peece of the Continent, a part of the maine; if a Clod bee washed away by the Sea, Europe is the lesse, as well as if a Promontorie were, as well as if a Manor of thy friends or of thine owne were; any mans death diminishes me, because I am involved in Mankinde; And therefore never send to know for whom the bell tolls; It tolls for thee."

MEDITATION XVII

Devotions upon Emergent Occasions

John Donne

RESUMO

Na prática da Bioinformática é natural que existam três tipos de pesquisadores envolvidos, o especialista em Biologia, o especialista em Informática e o pesquisador híbrido, com conhecimentos avançados nas duas áreas. Como este último é um espécime raro, os demais tendem à uma dificuldade comum, na qual a demasiada aplicação de seu intelecto em uma das duas grandes áreas (Informática e Biologia), resulta em um baixo aproveitamento quando deve-se integrar os assuntos para um bom produto. Visando auxiliar na aproximação das duas áreas e encurtar o caminho entre a pesquisa e o resultado em Bioinformática, propõe-se o box.bio.br, também chamado de 3B. Uma plataforma em nuvem responsável por integrar ambas as áreas a partir de uma rede social, ofertando ferramentas voltadas para a produção científica, possibilitando desde o desenvolvimento teórico da solução, passando pela oferta de um sistema de arquivos, pela disponibilidade de nós de processamento, controle de acesso à informação e possibilidade de um gerenciamento de projeto ativo, com distribuição de tarefas, especificação de prazos, e construção de uma *timeline*. Desta forma, a plataforma 3B visa fornecer uma solução completa para gerenciamento de recursos computacionais, recursos humanos e projetos, auxiliando o progresso de pesquisas na área.

Palavras-chave: Infraestrutura, Gerenciamento de Projetos, Computação em Nuvem, Computação Distribuída, Redes Sociais.

ABSTRACT

In the practice of Bioinformatics is natural that there are three types of researchers involved, the expert in Biology, the specialist in Computer Science and the hybrid researcher, with advanced knowledge in both areas. As the latter is a rare specimen, others tend to a common difficulty in which the excessive application of their intellect in one of the two large areas (Computer Science and Biology), results in a low utilization instead of integrating knowledge for a good product. Aiming to assist in the two areas approximation and shorten the path between research and results in Bioinformatics, the `box.bio.br`, also called 3B, is proposed. It is the agent responsible for connecting researchers, through aspects of a social networking, integrating targeted scientific tools and providing the necessary infrastructure, serving from the starting point of a problem to its solution development, including the provision of a file system and the availability of distributed processing nodes.

Keywords: Infrastructure, Project Management, Cloud Computing, Distributed Computing, Social Network.

LISTA DE FIGURAS

- Fig. 1. 3B Platform architecture and its relation with 3BStorage and 3BProcess nodes. 22
- Figura 2 - Visão Geral da Plataforma 3B. 25

LISTA DE TABELAS

Tabela 1 - Bancos de dados acessados por webservice TogoWS. 20

LISTA DE ABREVIATURAS

3B	box.bio.br - AN INTEGRATIVE BIOINFORMATICS RESOURCE PLATFORM
3BProcess	Nó de processamento da plataforma 3B
3BRepo	Repositório da plataforma 3B
3BStorage	Nó de armazenamento da plataforma 3B
API	Application programming interface
App	Application
box.bio.br	URL http://box.bio.br
COMM	Communication
EBI	European Bioinformatics Institute
EC2	Elastic Compute Cloud
GNU	GNU's NOT UNIX (Recursive acronym)
IP	Internet Protocol
JSON	Javascript Object Notation
KEGG	Kyoto Encyclopedia of Genes and Genomes
MVC	Model View Controller
NAT	Network Address Translation
NCBI	National Center for Biotechnology Information
NGINX	Web Server NGINX
PDBJ	Protein Data Bank Japan
PHP	Hypertext Preprocessor (Recursive Backronym)
SQL	Structured Query Language
TOGOWS	Togo Web Services
XML	Extensible Markup Language

SUMÁRIO

1. INTRODUÇÃO	13
2. FUNDAMENTAÇÃO TEÓRICA	14
2.1. PROCESSAMENTO DISTRIBUÍDO	14
2.2. COMPUTAÇÃO EM NUVEM	15
2.3. REDE SOCIAL	16
3. TRABALHOS RELACIONADOS	18
3.1. GALAXY	18
3.2. BIOLINUX	18
3.3. CLOUDMAN	19
3.4. TOGOWS	19
4. ARTIGO CIENTÍFICO	21
5. PLATAFORMA 3B - MATERIAL EXTRA	24
5.1. 3B - UMA VISÃO GERAL	24
5.2. 3B - RECURSOS	24
5.2. 3B - INFRAESTRUTURA	25
6. CONCLUSÃO	29
7. REFERÊNCIAS BIBLIOGRÁFICAS	31

1. INTRODUÇÃO

A Bioinformática é composta por duas grandes áreas de pesquisa: a biologia e a computação. Bioinformatas necessitam de uma capacidade avançada de desenvolvimento de soluções computacionais, bem como o entendimento de assuntos biológicos. Caso exista qualquer dificuldade na aplicação destes conhecimentos, podemos ter uma excelente solução computacional aplicada de forma que não retorne resultados biológicos significativos. Como também podemos ter resultados importantes, porém não identificados por uma falta de capacidade de interpretação (GIBAS, C. 2001).

A plataforma box.bio.br, também chamada de 3B, foi concebida para dar suporte às pesquisas em Bioinformática, focando na integração entre suas duas grandes áreas, unindo pesquisadores, desenvolvedores e instituições através de uma rede social para a troca de conhecimento e solicitações de novas soluções. Além da exposição das necessidades e das soluções para os mais diversos problemas da área, a plataforma oferece um sistema de criação de infraestrutura automatizado que entrega ao usuário, de forma rápida e fácil, sistemas de arquivos na nuvem e centros de processamento, formando assim uma solução centralizadora de ferramentas para o desenvolvimento e execução de pesquisas.

Executada em navegadores Web e alocada em nuvem, a plataforma funciona de qualquer lugar e horário em dispositivos conectados à Internet. Seus conceitos de rede social conectam instituições, grupos e pessoas para a realização de pesquisas e análises em Bioinformática a partir de sua arquitetura de sistemas para armazenamento e processamento distribuídos, exigindo muito pouco do usuário final para o uso de uma robusta infraestrutura.

A rede social intrínseca à plataforma 3B permite a comunicação entre diferentes pessoas conectadas em instituições e grupos distintos, possibilitando a colaboração entre pesquisadores de diversas partes do mundo. Os relacionamentos entre pessoas, pesquisas, grupos e instituições são gerenciados através de um filtro de privacidade, que fornece controle de acesso e indica o que deve ser de visualização privada ou pública nos mais diferentes níveis da rede. Isto garante o sigilo de pesquisas em andamento e interfaces para colaboração entre os mais variados usuários da plataforma.

A arquitetura da plataforma permite ao usuário adicionar nós de processamento e nós de armazenamento, que ficam disponíveis em um grande sistema distribuído, gerenciado pelas políticas da 3B. Desta forma, é possível que a plataforma cresça facilmente em capacidade de execução de processos e armazenamento de dados, desde que haja colaboração dos membros da rede social.

Por fim, com a sua organização hierárquica, na qual pessoas fazem parte de um grupo, lideradas por um agente pesquisador, e este grupo faz parte de uma instituição, a plataforma 3B possibilita o gerenciamento mais ativo de projetos. É possível a distribuição de tarefas, especificação de prazos limites, criação de *timelines* e inserção de comentários. Assim, a plataforma visa reunir, em um único local, pessoas, ferramentas e infraestrutura, possibilitando que sejam criados, gerenciados e executados os mais diversos projetos computacionais de Bioinformática.

2. FUNDAMENTAÇÃO TEÓRICA

Esta seção visa apresentar conceitos importantes para a um melhor entendimento deste projeto por parte do leitor. São discutidos tópicos como processamento distribuído, computação em nuvem e aspectos de redes sociais.

2.1. PROCESSAMENTO DISTRIBUÍDO

Um sistema de processamento distribuído é uma arquitetura que interliga computadores (nós) através de um gerenciador, que além de controlar o consumo de processamento destes computadores, realiza uma distribuição de carga para um nó em condições favoráveis para executar aquela tarefa, dividindo o mesmo processo em vários membros do sistema (BORGES, L. E., 2010).

Seu uso soluciona problemas de falta de processamento que necessita muitas vezes de uma máquina dedicada à tarefa (BORGES, L. E. 2010). Dividindo a tarefa em sub tarefas e distribuindo ao longo da rede em que o sistema está montado o resultado tende a ser mais rápido, sem utilizar todos os recurso de uma máquina, deixando-a livre para executar outras tarefas ao mesmo tempo.

Pode se classificar um sistema distribuído, de acordo com sua formação, de duas formas, cluster ou grid. A diferença das duas se dá no planejamento e organização. Enquanto o cluster é mais estruturado, com condições e ambientes controlados, o grid é composto por computadores em lugares adversos e sem controle.

Os modelos mais comuns de nós de processamento são: computacional, de recursos e o híbrido.

- Computacional: compartilha o processador e a memória dos computadores a fim de potencializar o processamento das tarefas.
- Recursos: serve para o armazenamento de informações, funcionando para ter mais performance de armazenamento e segurança dos dados.
- Híbrido: disponibiliza o seu hardware para processamento e armazenamento, se tornando integralmente parte da rede.

Sendo assim, os sistemas distribuídos devem conter quatro itens para ter seu funcionamento garantido (JR. BARROS, S. J., 2012):

- Comunicação: os nós envolvidos devem estar conectados através de uma rede de qualidade para ter o tráfego de informações.
- Metadados: dados de processamento que devem ser mantidos de forma adequada para sua utilização no momento do processamento, como por exemplo os parâmetros de execução.
- Controle: gerenciamento dos processos, análise de nós livres, menor latência, processadores livres.

- Segurança: proteção dos dados, por exemplo: conexões criptografadas, uso de chaves para autenticação, etc.

As vantagens de seu uso estão na economia em termos de custo e benefícios, onde alocar processamento ou solicitar armazenamento em *datacenters* de terceiros remove este investimento de infraestrutura que pode ser utilizado em outras áreas. O custo na operação de *datacenters* externos é mais barato do que o investimento em infraestrutura local. A segurança é outro ponto positivo para a tecnologia pois o sistema todo depende do funcionamento de uma máquina ligada a rede, já que o mesmo distribui apenas requisições para nós aptos. Outra vantagem é a facilidade de expansão, onde agregando uma nova máquina na rede se obtém todos os benefícios e se soma com a rede o potencial que o novo hardware proporciona (WASLAWICK, E S. 2004).

Para os sistemas distribuídos existem também algumas desvantagens, pois como se trata de uma rede, o sistema depende da conexão física entre as máquinas, que se não estiver em boas condições, a performance da arquitetura passa a ser afetada.

Dentro da plataforma 3B é utilizado o conceito de processamento distribuído, porém não é a execução do processo em si que é distribuída, e sim as tarefas que os usuários demandam. Assim, a plataforma visa gerar um equilíbrio de processamento entre os mais diversos nós distribuídos na rede. Entretanto, a forma de gerenciamento é a mesma, há o controle de carga por nó, envio de tarefas e análises de disponibilidade.

2.2. COMPUTAÇÃO EM NUVEM

Computação em nuvem é composta de uma rede de servidores conectados à Internet e se define pelo uso de serviços como armazenamento e processamento de dados onde o usuário não utiliza o hardware local, realizando solicitações de serviços, através da Internet, a estes servidores de alto poder de desempenho (ALECRIM, E. 2013).

Esta tecnologia permite com que se faça requisições remotas a qualquer momento, de qualquer lugar. A ideia do termo nuvem deriva do conceito de que um aparelho, seja computador ou um *smartphone*, se conecte pela Internet a esses servidores e trafegue dados sem que o usuário visualize a unidade física (ALECRIM, E. 2013). Para se ter os benefícios da computação em nuvem, o aparelho, com qual o usuário irá conectar, é livre da instalação de softwares específicos e na maioria das vezes o sistema operacional é indiferente.

Além de se ter as vantagens de não requisitar a instalação de softwares específicos para o uso de serviços na nuvem e isenção de preocupações quanto a sistema operacional e infraestrutura, há outros motivos para aderir ao uso dos serviços ofertados pela computação em nuvem, sendo os mais importantes: a centralização dos dados, que facilita a organização nas informações de uma organização ou instituição tornando mais ágil a comunicação e acelerando resultados; e a economia com hardware e aplicativos para gestão dos dados; já que muitas soluções existentes são gratuitas ou cobram pela quantidade

consumida, não havendo desperdícios caso o serviço seja parado ou não estiver utilizando toda a sua capacidade (SISNEMA, 2009).

O gerenciamento dos recursos disponíveis é feito de maneira simples e amigável ao usuário, que visualiza todas as informações necessárias permitindo uma tomada de ações sobre aumentar ou diminuir a carga do serviço, seja de processamento ou de espaço de armazenamento. Esta solicitação de alteração tem resposta imediata não dependendo de uma compra de novo servidor ou da alteração de hardware. Atualmente, os serviços prestados de computação em nuvem já automatizam essa tomada de decisão da manutenção de recursos, em caso de acréscimo ou decréscimo de recursos empregados. Entretanto, a computação em nuvem conta com uma desvantagem que impossibilita seu uso, que é o acesso a Internet. Se o usuário tiver por algum motivo a perda da conexão, seus dados na nuvem são inacessíveis para o uso. Outro fator impactante é uma possível instabilidade ou baixa velocidade no acesso à Internet, afetando assim a experiência do usuário com a nuvem. Por fim, um outro problema que esta tecnologia enfrenta é uma desconfiança em termos de segurança de dados. Há um sentimento de desconfiança que é gerado sobre o fato de que os dados de um usuário podem estar alocados sobre posse de um terceiro.

É importante mencionar que soluções já estão sendo desenvolvidas para solucionar o problema de conectividade, criando formas de sincronismo *offline* (STEPANSKI & SANDO, 2011), deixando crer em uma expansão e maior dependência de serviços em nuvem, migrando não só aplicações específicas, mas o sistema operacional como um todo, fazendo com que o hardware seja cada vez mais descartável.

Na plataforma 3B, toda a infraestrutura necessária para a execução das atividades é em nuvem, o que faz do dispositivo do usuário para o acesso à ferramenta, uma interface que exige o mínimo de processamento possível. Por mais que um usuário decida aplicar sua infraestrutura na plataforma 3B, seu gerenciamento e uso será feito através da nuvem, possibilitando seu compartilhamento para os mais diversos usuários e instituições ao redor do mundo, caso seja de desejo e interesse de seu detentor.

2.3. REDE SOCIAL

As redes sociais são estruturas formadas por atores que podem ser representados por indivíduos ou organizações. Através de conexões envolvendo estes atores, são formadas as estruturas sociais. As conexões também podem ser chamadas de interações (WASSERMAN & FAUST, 1994).

O crescimento e forma da rede se dá a partir das interações criadas individualmente por cada ator, fornecendo assim dados passíveis de análise para a identificação de padrões de desenvolvimento da mesma (WASSERMAN & FAUST, 1994).

Estudos levando em consideração redes sociais com pesquisadores gerando colaborações mostram que a sua exploração e aplicação no meio científico é relevante, pois a colaboração intelectual em um meio com interações

confiáveis, garantido pela estrutura da rede social, apresenta resultados positivos para o desenvolvimento científico (BARABÁSI A.L., *et. al.* 2002).

A utilização deste conceito no trabalho proposto é decorrente das necessidades já apresentadas pelos cientistas, da facilidade na intercomunicação entre diversos colaboradores e do uso frequente das comunidades ao Facebook, rede social digital inspiradora de vários projetos, tal qual o myExperiment (GLOBE, C.A., *et. al.* 2010) e o box.bio.br.

Na plataforma 3B a rede social é composta pelos atores Usuário, Grupo e Instituição. A hierarquia presente nesta estrutura relaciona os usuários aos grupos de pesquisa e estes a sua *Alma Mater*. Os vínculos criados pelas interligações da rede exercem um peso de co-responsabilidade nas atuações dos membros em projetos e colaborações, aumentando a confiabilidade das informações cadastradas no sistema.

3. TRABALHOS RELACIONADOS

Esta seção visa apresentar os trabalhos relacionados mais relevantes à este projeto. Eles se relacionam através de estratégia de uso de computação em nuvem, da oferta de uma plataforma que apresenta serviços de Bioinformática, pelo uso de serviços Web ou até mesma pela aplicação direta do trabalho neste projeto.

3.1. GALAXY

O projeto Galaxy (GIARDINE, B., *et al.*, 2005) é a união de diversos aplicativos de Bioinformática, disponibilizados através de um portal Web, que por sua arquitetura permite acesso remoto aos recursos existentes na ferramenta. Criado para permitir pesquisas independentes e análises através das combinações de seus resultados, a ferramenta é muito utilizada e já existem projetos que automatizam a reprodução de instâncias em infraestruturas em nuvem como a Amazon. Além de possibilitar o acesso remoto à interface de uso, o projeto também é disponibilizado para download e instalação completa em máquinas locais.

A diferença entre o projeto 3B e o Galaxy é que ao invés de ser uma suite de ferramentas distribuída em diversos laboratórios e centros de pesquisas, como acontece com o Galaxy, a plataforma 3B é centralizada em um servidor na nuvem, que ao crescer com uma nova funcionalidade, beneficia todos os seus usuários conectados, além de possibilitar que cada usuário faça o gerenciamento de sua infraestrutura, escolhendo exatamente o que vai ser utilizado e instalado em seu hardware. A descentralização da plataforma 3B ocorre somente em questões relacionadas a processamento e armazenamento, permitindo neste caso o crescimento rápido e fácil destes recursos. A configuração de execução de novas ferramentas e funcionalidades da rede social acontece em um único lugar, em nuvem.

3.2. BIOLINUX

BioLinux (BIOLINUX, 2014) é uma distribuição do sistema operacional Linux previamente preparada para ser instanciada em máquinas virtuais, provendo mais de 135 pacotes de aplicações com soluções envolvendo alinhamento de sequências, clusterização, montagem, exibição de dados, edição de arquivos e filogenia. A solução visa a rápida escalabilidade e criação de ambientes de análise e trabalho para a reprodução de resultados, sendo indicada para infraestruturas em nuvem para Bioinformática (KRAMPIS, *et al.*, 2012).

A imagem (BIOLINUX, 2014) para novas instâncias está disponível para a criação de ambientes virtuais em laboratórios privados, porém é possível a sua instalação e configuração em estrutura fornecida pela EC2 da amazon.com. Utilizando o serviço da Amazon, as opções de poder de processamento são grandes, possibilitando que seja criado o ambiente mais favorável para a necessidade do usuário do BioLinux.

Como a plataforma 3B necessita de um sistema operacional para o gerenciamento dos nós de processamento e nós de armazenamento, recomendamos o uso do BioLinux, pois o potencial de uso de soluções aumenta em sua configuração padrão, já que o usuário não precisa instalar manualmente pacotes essenciais e conhecidos para os mais diversos tipos de processamentos desejados. Após o uso da imagem BioLinux em nós de processamento, todos os aplicativos presentes no novo nó que estiverem configurados na plataforma já estão disponíveis para seu uso, reduzindo consideravelmente o tempo e esforço para colocar uma nova infraestrutura em produção.

3.3. CLOUDMAN

Diante da adoção de sequenciamento com grande geração de dados, a necessidade de uma tecnologia de computação sofisticada e escalonável se fez presente, demonstrando que as alternativas as infraestruturas locais passassem a ser cada vez mais atraentes. Assim a computação em nuvem começou a fazer parte da realidade das soluções de grandes estruturas computacionais.

Mesmo com a solução de computação em nuvem à disposição, ainda assim não se tornou simples a criação de uma infraestrutura apta para a execução de análises e geração de resultados. Pensando nisto, foi criado o CloudMan (AFGAN, E, *et al.*, 2010), uma ferramenta WEB que facilita a instancia de novas máquinas virtuais na plataforma EC2 da Amazon com o sistema Galaxy instalado, fornecendo as suas soluções para o trabalho do biólogo experimental.

Através de um painel acessível de navegadores web, o usuário pode controlar todas as instâncias criadas através de uma conta no serviço EC2. Informações de status e opções para o gerenciamento destas instâncias são facilmente administradas pelo usuário. Através de exibições com cores correspondentes com os status das instâncias, é fácil a interpretação do uso e dimensionamento da infraestrutura ativa.

Diferente da estrutura do CloudMan, a plataforma 3B fornece a criação de instâncias de nós de processamento, chamados 3BProcess, em qualquer hardware ou serviço existente, não sendo especificamente desenvolvido para a estrutura da Amazon. Além disto, por ter uma plataforma de credenciamento de soluções central, os aplicativos registrados são identificados em qualquer um dos nós 3BProcess, o que não acontece no CloudMan, já que a mesma coleção de soluções, no caso o Galaxy, que é replicado em diversas instâncias no EC2 da Amazon.

3.4. TOGOWS

Com o crescente número de soluções desenvolvidas e disponibilizadas na WEB, o uso de *WEB services* se tornou uma maneira de integrar as mais variadas aplicações através de um padrão de comunicação. O TogoWS (KATAYAMA, T., 2010) é resultado do trabalho de unificação da comunicação entre os mais diversos bancos de dados biológicos, no qual através de uma API, o desenvolvedor centraliza e obtém de uma única conexão, dados de diversas origens, garantindo uma maior abrangência nas pesquisas desejadas.

Na plataforma 3B, o TogoWS é utilizado para a interface de pesquisa em bancos de dados biológicos, que possibilita ao usuário requisitar informações dos bancos de dados listados na Tabela 1.

A plataforma recebe o resultado da requisição no formato JSON, fornecido pelo TogoWS, e os formata para exibição na tela.

BANCOS DE DADOS ACESSADOS POR WEBSERVICE TOGOWS					
uniprotebi-ena	ebi-emblrelease_con	ebi-merops_family	ebi-sra-submission	ncbi-blastdbinfo	ncbi-nucest
ncbi-nucleotide	ebi-emblrelease_standard	ebi-merops_id	ebi-taxonomy	ncbi-books	ncbi-nucgss
ncbi-snp	ebi-emblrelease_tsa	ebi-nrn1	ebi-uniparc	ncbi-cdd	ncbi-nucleotide
ncbi-pubmed	ebi-emblrelease_wgs	ebi-nrn2	ebi-uniprot	ncbi-clone	ncbi-omia
ncbi-omim	ebi-ena	ebi-nrpl1	ebi-uniref100	ncbi-dbvar	ncbi-omim
pdj-pdb	ebi-ensemblGenomes_gene	ebi-nrpl2	ebi-uniref50	ncbi-epigenomics	ncbi-pcassay
ebi-arrayexpress-r	ebi-ensembl_gene	ebi-omim	ebi-uniref90	ncbi-gap	ncbi-pccompound
ebi-atlas-experiments	ebi-epo	ebi-patentFamilies	ebi-uspto	ncbi-gapplus	ncbi-pcsubstance
ebi-atlas-genes	ebi-genome_assembly	ebi-patentdb	ebi-wgs_masters	ncbi-gcassembly	ncbi-pmc
ebi-biomodels	ebi-go	ebi-pdbe	kegg-compound	ncbi-gds	ncbi-popset
ebi-chebi	ebi-gpcrdb	ebi-pdbechem	kegg-drug	ncbi-gencoll	ncbi-probe
ebi-chembl-activity	ebi-hgnc	ebi-pride	kegg-enzyme	ncbi-gene	ncbi-protein
ebi-chembl-assay	ebi-intact-experiments	ebi-project	kegg-genes	ncbi-genome	ncbi-proteinclusters
ebi-chembl-target	ebi-intact-interactions	ebi-reactome	kegg-glycan	ncbi-genomeprj	ncbi-pubmed
ebi-dgva	ebi-intact-interactors	ebi-resid	kegg-module	ncbi-geoprofiles	ncbi-pubmedhealth
ebi-efo	ebi-intenz	ebi-rhea	kegg-orthology	ncbi-homologene	ncbi-seqannot
ebi-ega	ebi-interpro	ebi-sbo	kegg-pathway	ncbi-journals	ncbi-snp
ebi-emblcids	ebi-jpo	ebi-sra-analysis	kegg-reaction	ncbi-medgen	ncbi-sra
ebi-emblnew_con	ebi-kipo	ebi-sra-experiment	ncbi-assembly	ncbi-mesh	ncbi-structure
ebi-emblnew	ebi-lrg	ebi-sra-run	ncbi-bioproject	ncbi-ncbisearch	ncbi-taxonomy
ebi-emblnew_tsa	ebi-medline	ebi-sra-sample	ncbi-biosample	ncbi-nlmcatalog	ncbi-toolkit
ebi-emblnew_wgs	ebi-merops_clan	ebi-sra-study	ncbi-biosystems	ncbi-nuccore	ncbi-toolkitall
ncbi-toolkitbook	ncbi-unigene	ncbi-unists	pdj-pdb		

Tabela 1 - Bancos de dados acessados por webservice TogoWS.
 FONTE: O AUTOR, 2014

4. ARTIGO CIENTÍFICO

Esta seção apresenta o artigo científico a ser submetido à Oxford *Bioinformatics* como um *Application Notes*, ou seja, um artigo de formato curto que apresenta uma solução em software, neste caso, a própria plataforma.

BOX.BIO.BR – INTEGRATIVE BIOINFORMATICS RESOURCE PLATFORM

Eduardo Langowski¹, Emanuel Maltempi de Souza¹ and Alessandro Brawerman^{1,**}

¹Bioinformatics Graduation Program, R. Alcides Vieira Arcoverde, 1225. Jd. das Américas.

Curitiba – PR – 81520-260 - Brazil

Received on XXXXX; revised on XXXXX; accepted on XXXXX

ABSTRACT

Motivation: In the practice of Bioinformatics is natural that there are three types of researchers involved, the expert in Biology, the computer expert and the hybrid researcher, with advanced understanding in both areas. As the latter is a rare specimen, others tend to a common difficulty, which is the integration of knowledge.

Aiming to assist in the two areas approximation and shorten the path between research and results in Bioinformatics, the box.bio.br, also called 3B, is proposed. It is the agent responsible for connecting researchers, through aspects of a social networking, integrating targeted scientific tools and providing the necessary infrastructure, serving from the starting point of a problem to its solution development, including the provision of a file system and the availability of distributed processing nodes.

Availability: The 3B platform is freely available at <http://box.bio.br>.

1 INTRODUCTION

Bioinformatics consists of two major areas of research, Biology and Computer Science. A Bioinformatician needs skill on how to develop computational solutions as well as the understanding of biological issues. If there is a lack of understanding on biological knowledge, we can have an excellent computational solution used in a way that does not return significant results. The same happens on the opposite, with a low experience on computational areas, important results can be unidentified with weak algorithms [1].

The box.bio.br (3B) platform is designed to support research in Bioinformatics, focusing on the integration of its two large areas, connecting researchers through a social network for the exchange of knowledge and new solution requests. Besides the exhibition of needs and solutions to various problems of the area, the platform provides an automated creation system, that delivers to the user, quickly and easily, file systems in the cloud and processing centers, thus forming a centralized platform for the development of research in the area. Running in Web browsers and located in the cloud, the platform works from any and everywhere on devices connected to the Internet. The social network core connects institutions, groups, individuals and computing centers, helping them to conduct research and analysis in Bioinformatics through a network that provides storage and distributed processing, requiring very little end-user skills to use a robust infrastructure. The relationships between people, resources, groups and

institutions are managed through access control functionalities, which indicates what should be private or public, in many different levels of the network. This ensures the confidentiality of ongoing research and interfaces for collaboration among the various users. The platform architecture allows the user to add processing and/or storage nodes, which are available in a large distributed system, managed by the 3B policies. With these features the platform brings together into one place, people, tools and infrastructure, providing a universe in which Bioinformatics projects are created, managed and executed.

2 METHODS

The platform is built to run in the cloud and its Web interface provides responsive design, which uses the javascript Bootstrap 3.0 framework to provide the best user experience. The view automatically adjusts to be accessed from mobile phones, tablets or any display resolution.

As depicted by Figure 1, the 3B software architecture, based on the MVC model, was developed with the PHP version 5.5.14 programming language, using the CodeIgniter framework version 2.2.0 and the user data and social network data are stored in a MySQL database server, version 5.1.73. The Web server is running on a NGINX version 1.7.7, chosen for its superior performance in concurrent access requests to the server.

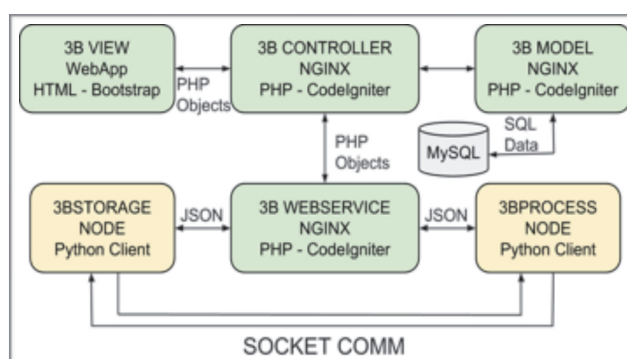


Fig. 1. 3B Platform architecture and its relation with 3BStorage and 3BProcess nodes.

2.1 STORAGE NODE

A storage node is a computer added to the 3B platform whose goal is to offer storage space to a specific user, a group of users

* To whom correspondence should be addressed.

or all the other users. Any regular 3B platform user can add a computer and offer file storage space.

To become a storage node a specific piece of software, the 3BStorage App, must be installed and configured by its user. It is necessary to have administrator privileges to install and configure the 3BStorage in any given computer.

To get the 3BStorage App, one has two options, install it from the 3B Linux repository or download it on the following address <http://box.bio.br/3bstorage>. After downloading and installing it, one must edit the 3bstorage.ini configuration file to have his/her storage node available for the platform users.

The storage node works as a shared file repository. Any type of file is accepted, for instance, fasta pre-processed files may be shared among users to facilitate other processing steps; documents, images, or notes may be shared to speedup research.

Several storage and processing nodes may be created in the same local area network, which would greatly improve file transfer between those nodes fastening the entire process.

2.2 PROCESSING NODE

A processing node is responsible for executing jobs demanded by the 3B platform users. These nodes are added to the platform in a similar way as the storage nodes, however their goal is to offer processing power. To register a processing node, a user needs to install and configure the 3BProcess App, available at <http://box.bio.br/3bprocess> and at the 3B Linux repository.

For any service (application) offered by a processing node, the 3BProcess App checks if it is already registered within the platform, otherwise, the user needs to register it before its use. In the last case, the user needs to upload a XML file, which will provide the platform with the parameters required to communicate with this service. This ensures that the platform understands how to execute specific jobs on any given processing node.

The user who collaborates with the XML configuration file will be scored on the social network, for both, the creation of the file and its use by other users. This XML will be related to the corresponding version of the software and will be available to all other users. The operating system recommended for processing nodes is the BioLinux (2), however, the 3BProcess App works with any other operating system capable to compile and execute Python files.

3 DISCUSSION

This section briefly describes how the platform is employed to attend the needs of a researcher, taking into account that 3BProcess and 3BStorage nodes have been already configured in his/her account.

The platform contains a list of services (applications) registered, through XML files, and available to be used. Whenever a user needs to run a job, he/she selects the desired service; fills in the necessary fields, most applications will ask for an input file and at least a couple of parameters; and finally the platform allocates the task to a particular 3BProcess node

capable of performing the job. From the platform's point of view, this marks the beginning of an execution pipeline, starting with the input file compression step; its transmission directly to the 3BProcess node; the data extraction to a temporary folder; the execution of this job, according to the received parameters; the analyzes of this job status, to check the end of the process; and finally the result compression and sending it back to the 3BStorage, alerting the user interface that this particular task was performed.

As a social network, the success of the 3B platform depends on its users. It is expected that some users collaborate more than others, providing more solutions, participating in more projects and alerting misuse of the platform. These users are rewarded by receiving participation points, which are summed to their institution score, helping the institution to get a better global ranking position. For instance, a researcher may describe a problem and the need to have a tool developed to solve that problem. A developer may implement the computational solution of that problem. Besides working together to solve a specific problem, both are rewarded with points. Furthermore, their solution is also rewarded with points according to its use by others.

Finally, since the platform is organized in hierarchical groups, going on from institutions, to research groups and end users, a project leader can manage his/her project through the 3B platform. The project leader creates the project and may distribute tasks and deadlines to the project members. The platform creates a timeline helping the members to identify the most urgent and important tasks. Everyone registered to work on the project may add comments at any time.

4 CONCLUSION

With the provision of supporting tools for infrastructure and research management, a social network that connects experts in biological and computational areas, and the capacity to grow in file storage and processing power, the 3B platform provides a complete solution for storage, management and processing of various research projects in Bioinformatics. Its decentralized architecture allows the user to add storage and processing nodes, offering their capabilities to a specific user, a group of users or all the platform users. Access control functions ensure that in one hand sensitive research data remain only with their proprietary users and institutions, but in the other hand the core of the platform, i.e., the social network and the collaboration features to build new Bioinformatic tools, would not be affected. Since the 3B platform is available in cloud-based infrastructure, the users have their data available at anytime, anywhere, seamlessly, as long as there is an Internet connection.

Funding: Agencies CAPES and CNPq.

References

1. Gibas, C. and Jambeck, P. (2001) Developing Bioinformatics Computer Skills, O'Reilly Media
2. Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, Thurston M: (2006) Open software for biologists: from famine to feast. Nat Biotechnol

5. PLATAFORMA 3B - MATERIAL EXTRA

Esta seção apresenta detalhamentos que podem ajudar a esclarecer alguns pontos importantes da plataforma e suas funcionalidades. Começamos apresentando novamente uma visão geral da plataforma, seguindo de maiores detalhes de seus recursos e da infraestrutura oferecida e gerenciada por ela.

5.1. 3B - UMA VISÃO GERAL

A plataforma 3B visa gerenciar infraestrutura, projetos e pessoas. Neste sentido, uma de suas principais funcionalidades é o gerenciamento de nós de armazenamento, chamados 3BStorage e de nós de processamento, chamados 3BProcess. Além desta infraestrutura, a plataforma também oferece as conexões entre Instituições, Grupos, Usuários e suas Pesquisas. Estes relacionamentos acontecem dentro da rede social, que pontuam os Usuários através de suas colaborações com a comunidade, seja criando arquivos de configuração XML para a utilização de aplicações nos nós 3Bprocess ou colaborando em Pesquisas com resultados e conhecimentos. A plataforma contém um sistema de controle de permissões, no qual todo o conteúdo pode ser classificado como público ou privado, para todos os níveis hierárquicos da rede social.

Conforme demonstrado na Figura 2, há uma integração muito forte entre os diversos atores por intermédio da plataforma. Note que os nós de processamento e armazenamento são representados por retângulos, ilustrando que a plataforma gerencia inúmeras instâncias de nós 3BStorage e 3BProcess relacionados na sua estrutura. Para o gerenciamento e uso desta infraestrutura, representados por círculos, estão os agentes da rede social. Diversos usuários podem se conectar e usufruir da plataforma independente de um relacionamento com outros atores, como por exemplo Grupos, Pesquisas e Instituições. A hierarquia existente entre estes atores, partindo no nível superior é composta por: Instituições, Grupos, Pesquisas e Usuários.

A rede social motiva a integração e colaboração de seus atores atribuindo pontuações individuais. Por exemplo, solicitações de novas ferramentas, desenvolvimento destas ferramentas, o uso das mesmas e compartilhamento de infraestrutura, são ações que geram pontos para os atores envolvidos. Na hierarquia da rede social, atores superiores possuem pontuação de participação e influência na rede baseados na soma dos resultados de seus agentes inferiores. ou seja, pessoas geram pontos para grupos e estes para suas instituições. Com isto é possível quantificar e promover os agentes que realizam maior colaboração na comunidade 3B.

5.2. 3B - RECURSOS

Rede Social: Para realizar a integração entre pesquisadores, o recurso Connection disponibiliza cadastros de Instituição, Grupos, Projetos e Pessoas. O cadastro de Instituição é realizado por centros de pesquisas que possuem grupos e pessoas envolvidas. Neste cadastro é possível gerenciar e publicar os trabalhos realizados. Grupos, por sua vez, são concentrações de Pessoas, nos quais os projetos de pesquisa devem ser cadastrados. Cada projeto possui Pessoas

relacionadas, que são os membros do grupo responsáveis por realizar tarefas dos projetos. Cada Instituição e Grupo tem uma Pessoa relacionada como responsável, um líder de projetos.

Cadastros de Instituição e Pessoas podem ter 3BStorage e 3BProcess relacionados. Estes nós de infraestrutura podem ser compartilhados por projetos, disponibilizando os nós para que determinadas pesquisas realizem seus resultados.

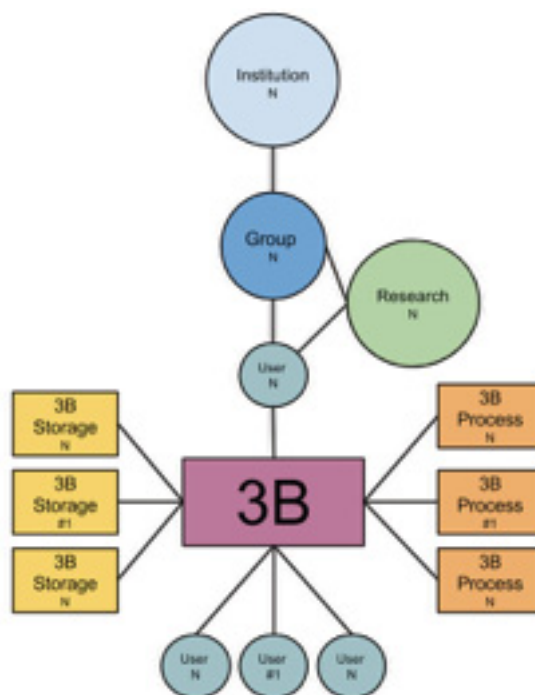


Figura 2 - Visão Geral da Plataforma 3B.
FONTE: O AUTOR, 2014

Gerenciamento do projeto: Durante a evolução do projeto, é possível que sejam controladas as ideias e colaborações envolvidas através de distribuição de tarefas, especificação de prazos para cumprimento das mesmas, geração de uma *timeline*, e inclusão de comentários. Permitindo assim, o registro em um único local de todas as interações das Pessoas com a projeto.

Acesso a bases biológicas: Utilizando o sistema de web services TogoWS, a plataforma 3B concebe uma integração com 136 bancos de dados biológicos. Através desta integração é possível localizar sequências genéticas direto dos bancos de dados e armazená-las em um 3BStorage previamente configurado em sua conta na plataforma 3B. Posteriormente é possível utilizar o arquivo salvo como parâmetro de um processamento com uma ferramenta.

5.2. 3B - INFRAESTRUTURA

3BStorage

O sistema de armazenamento de arquivos consiste no mapeamento de espaço em disco rígido de um servidor, chamado 3BStorage, para dentro da

plataforma 3B. Esta estrutura possibilita a descentralização dos dados em diversos 3BStorage, em diferentes localidades. Os arquivos ficam disponíveis online na plataforma 3B, possibilitando sua manipulação através da interface Web, que funciona em qualquer dispositivo que contenha um navegador Web.

O software de gerenciamento que é executado no hardware do 3BStorage foi escrito em Python, acompanhado por um arquivo de configuração. Sua instalação pode ser feita manualmente, baixando o software da própria página 3B, ou automaticamente, adicionando o repositório 3BRepo, em seu sistema operacional Linux, e instalando-o com o comando `apt-get install 3BStorage`.

A sua configuração é feita através de um arquivo chamado `3BStorage.ini`. Abaixo a estrutura do arquivo e seus elementos:

```
[Data]
Public_IP: 200.222.3.21
Send_Timer: 1
Port: 5005
Root: /home/user/folder/
```

Em `Public_IP` deve ser informado o endereço IP público que será utilizado para o acesso ao storage. Caso o 3BStorage não possua endereço IP público em sua interface de rede para acesso direto, seu valor deve ser `NO_ACCESS`. Esta configuração fará com que 3BStorage não atue como um servidor, passando a enviar os dados ativamente para a plataforma 3B. Este envio para a plataforma deve ser configurado em minutos no campo `Send_Timer`. Outra possibilidade é configurar um NAT e fazer o redirecionamento de portas no roteador da rede apontando para a máquina 3BStorage, na porta configurada no campo `Port`, que identifica ao sistema a porta que vai fornecer a conexão. O campo `Root` identifica qual a pasta raiz para exibir dentro do sistema 3B os diretórios, sub-diretórios e arquivos.

3BProcess

O processamento dos arquivos encontrados dentro do 3BStorage acontece nos nós de processamento 3BProcess. Sua configuração é realizada através do arquivo `3BProcess.ini` e sua instalação pode ocorrer baixando o cliente na seção de download da plataforma 3B ou instalado por via do 3BRepo, devidamente configurado pelo comando `apt-get install 3BProcess` no caso de Linux.

```
[Data]
Public_IP: 200.222.3.21
Port: 6006
Temp_Folder: /home/user/tmp
```

A instalação do cliente de processamento 3BProcess é sugerida em um sistema operacional Linux Distribuição BioLinux. Esta distribuição já contém diversas soluções para Bioinformática, porém seu uso não é obrigatório. Qualquer sistema operacional com ferramental da área instalado pode servir como nó de processamento.

3BRepo

O repositório oficial da plataforma é o responsável por distribuir as últimas versões dos aplicativos que gerenciam os nós de processamento e armazenamento. Seu endereço está disponível na URL <http://box.bio.br/3brepo>. Para seu uso, deve-se adicionar o 3BRepo à lista de repositórios do sistema operacional.

Para a versão homologada BioLinux e Ubuntu, o comando para adicionar o repositório é

```
# sudo add-apt-repository ppa:<repository-name>
```

Instruções de uso

Considerando um usuário que não é cadastrado na plataforma, para configurar um ambiente de trabalho deve-se executar os seguintes passos:

1 - Cadastro de usuário para acesso à plataforma (obrigatório). Na página inicial existe um botão chamado “Sign Up”, que fornece o formulário de cadastro.

2 - Baixar, instalar e configurar o aplicativo 3BStorage, para o armazenamento de seus arquivos (opcional; somente em caso de oferta de serviço de armazenamentos). Esta estrutura possibilita a descentralização dos dados em diversos 3BStorage, em diferentes localidades. Os arquivos ficam disponíveis online na plataforma 3B, possibilitando sua manipulação através da interface Web, que funciona em qualquer dispositivo que contenha um navegador Web, inclusive dispositivos móveis.

O software de gerenciamento que é executado no hardware de armazenamento foi escrito em Python, e acompanha um arquivo de configuração. Sua instalação pode ser feita manualmente, baixando o software da própria página 3B (menu downloads, 3BStorage), ou automaticamente, adicionando o repositório 3BRepo em seu sistema operacional linux e instalando com o comando `apt-get install 3BStorage`.

A sua configuração é feita através de um arquivo chamado 3BStorage.ini. Os dados presentes no arquivo são: Public_IP, Send_Timer, Port e Root.

Em Public_IP deve ser informado o endereço IP público que será utilizado para o acesso ao servidor de armazenamento. Caso o 3BStorage não possua endereço IP público ou válido, seu valor deve ser NO_ACCESS. Esta configuração fará com que 3BStorage atue como um servidor ativo, passando a enviar os dados para a plataforma 3B. Este envio para a plataforma deve ser configurado em minutos no campo Send_Timer. Uma outra possibilidade caso o usuário não queira uma máquina como servidor ativo é configurar um NAT no roteador da rede apontando para a máquina 3BStorage, na porta informada no campo Port, que identifica ao sistema a porta que vai escutar a conexão. O campo Root identifica qual a pasta raiz para exibir dentro do sistema 3B os diretórios, sub-diretórios e arquivos.

3 - Baixar, instalar e configurar o cliente 3BProcess, para a execução dos processos solicitados (opcional; somente em caso de oferta de serviço de processamento). O processamento dos arquivos encontrados dentro do 3BStorage acontece nos nós de processamento 3BProcess. Sua configuração é realizada através do arquivo 3BProcess.ini e sua instalação pode ocorrer baixando o cliente na seção de download da plataforma 3B ou instalado via 3BRepo, devidamente configurado pelo comando `apt-get install 3BProcess`, no caso de Linux.

Os campos presentes neste arquivo de configuração são: `Public_IP`, `Port` e `Temp_Folder`.

O campo `Public_IP` deve receber o endereço público para acesso à máquina. `Port` é a porta que deve ser aberta para receber as conexões. `Temp_Folder` é a pasta com permissão de escrita para que os arquivos de saída e processamento possam manuseados pelo sistema.

A instalação do cliente de processamento 3BProcess é sugerida em um sistema operacional Linux Distribuição BioLinux. Esta distribuição já contém diversas soluções para bioinformática. Porém seu uso não é obrigatório. Qualquer sistema operacional com ferramentas de bioinformática instaladas pode servir como nó de processamento.

Quando a plataforma recebe uma requisição de processamento, a mesma deve consultar se há a possibilidade de executar no nó desejado aquele processo. Isso deve ser feito rodando a chamada do processo com opção `version`, pelo 3BProcess, de maneira transparente ao usuário. Caso positivo, a requisição de processamento é enviada, caso negativo, o sistema procura um próximo 3BProcess para executar este processo.

6. CONCLUSÃO

Através das tecnologias utilizadas para o desenvolvimento da plataforma 3B, o produto final fornece ao usuário a facilidade na criação de infraestrutura para o uso de soluções em Bioinformática, uma rede social de relacionamento entre pessoas, grupos e instituições, e possibilidades de gerenciamento de projetos de forma mais ativa pelo líderes de grupos.

Para que a solução tenha uma adesão casa vez maior, seu núcleo funciona em nuvem, trabalhando independente da infraestrutura (3BProcess e 3BStorages) criadas por seus usuários. Grupos de pesquisa que contém hardware próprio para seus processos e armazenamento, podem utilizar a plataforma para gerir tarefas. Já usuários que não tenham infraestrutura própria, podem usufruir de máquinas compartilhadas dentro da rede social aplicada na plataforma.

Devido ao seu caráter hierárquico é possível também o gerenciamento de projetos por líderes de grupos de pesquisa. Estes líderes podem distribuir tarefas para os membros dos grupos e indicar prazos limites para a conclusão das tarefas. Além disto, a plataforma, neste caso, oferece também uma *timeline* para indicar de maneira visual as etapas cumpridas no projeto e permite que qualquer membro do grupo inclua seus comentários, propiciando um ambiente de discussão para melhora do projeto.

As principais vantagens no uso da plataforma 3B são a velocidade e simplicidade para a configuração de uma infraestrutura robusta para o processamento de dados em Bioinformática e a centralização de informações das pesquisas em desenvolvimento, usando-se computação em nuvem, sem exigir do usuário muito tempo e conhecimento em Informática. Com o uso da plataforma, naturalmente a gama de serviços funcionais amplia, devido a colaboração possível entre os usuários da plataforma.

Com a concepção deste projeto, novas implementações foram identificadas como potenciais trabalhos futuros, sendo possíveis extensões e funcionalidades da plataforma 3B. As principais são identificadas abaixo:

- Para a explorar os arquivos de sequência genética, a integração ou desenvolvimento de um sistema de visualização e anotação, desenvolvido com tecnologias para execução em navegadores web, se faz necessária.
- Melhoria da implementação de divisão tarefas para controle e divisão de processamento no gerenciamento de distribuição de carga entre nós 3BProcess. A ferramenta atualmente é própria e desenvolvida utilizando a linguagem de programação Python. Em pesquisas de soluções para esta necessidade, foi levantada a consideração do uso da ferramenta de paralelismo chamada GNU Parallel para o trabalho de processamento distribuído.
- Gerenciamento bibliográfico e de conteúdo relacionado ao desenvolvimento da pesquisa para futura produção científica, criando um local de armazenamento de conteúdo em anexo à timeline. Esta solução centralizada em nuvem provê segurança e organização ao usuário.

- Integração de modelos LaTeX para a formatação automática de conteúdo armazenado no gerenciamento de pesquisa, baseado em tags. Assim, de uma forma semi-automática, a plataforma pode gerar modelos de diversos formatos, preenchidos com conteúdos organizados previamente através destas tags.
- Criação de edição dos modelos utilizados na formatação dos resultados JSON produzidos pelos web services do TogoWS, disponibilizando para a comunidade de desenvolvedores a possibilidade de criação de plugins para os mais diversos tipos de funcionalidades baseadas no retorno de informação gerada pelo sistema TogoWS.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AFGAN E.; et al. Galaxy CloudMan: delivering cloud compute clusters, **BMC Bioinformatics**, v. 11, suppl 12, s4, 2010.

ALECRIM, E. **O que é cloud computing? (computação nas nuvens)** Disponível em <<http://www.infowester.com/cloudcomputing.php>> Acesso em: 26/07/2014.

Amazon EC2 cloud service Disponível em <<http://aws.amazon.com/ec2>> Acesso em 08/10/2013.

BARABÁSI, A.L.; et al. Evolution of the social network of scientific collaborations. **Physica A**, v. 311, 590-614, 2002.

Bootstrap Framework Disponível em <<http://getbootstrap.com>> Acesso em 15/06/2013.

BORGES, L. E. **Python para Desenvolvedores**, 2. ed. Novatec Editora, 2010.

Cloud BioLinux Disponível em <<http://www.cloudbiolinux.org>> Acesso em: 19/09/2013.

GIARDINE, B.; et al. Galaxy: A platform for interactive large-scale genome analysis. **Genome Res.**, v. 15, n. 10, 1451-5, 2005.

GIBAS, C. **Computers + Biology = Bioinformatics** Disponível em <http://archive.oreilly.com/pub/a/oreilly/news/bioinformatics_0401.html> Acesso 12/11/2013.

GLOBE, C.A.; et al. myExperiment: a repository and social network for the sharing of bioinformatics workflows. **Nucl. Acid Res.**, v. 38, suppl 2, W677-W682, 2010.

JR BARROS, S. J. **Implementação de um cluster híbrido experimental para processamento de aplicações distribuídas**, Diss. UNIVERSIDADE FEDERAL DE PERNAMBUCO, 2012.

KATAYAMA, T.; et al. TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services. **Nucleic Acids Res**, v. 38, W706-W711, 2010.

MySQL Database Disponível em <<http://www.mysql.com>> Acesso em: 28/08/2013.

NGINX Web Server Disponível em <<http://nginx.org>> Acesso em 28/11/2013.

PHP: Hypertext Preprocessor Disponível em <<http://php.net>> Acesso em: 24/09/2013.

Phyton Software Foundation Disponível em <<https://www.python.org>> Acesso em: 10/11/2014.

SISNEMA CLOUD COMPUTING - NOVO MODELO DE COMPUTAÇÃO

Disponível em <<http://sisnema.com.br/Materias/idmat019433.htm>> Acesso em: 10/11/2014.

STEPANSKI, R. & SANDO, H. G. **Cloud Computing SAAS (Software-as-a-service)**. Universidade do Sul de Santa Catarina, Palhoça, 2011.

WASLAWICK, E. S. **Análise e projetos de sistema de informação**. Rio de Janeiro, v. 2. Elsevier Brasil, 2004.

WASSERMAN & FAUST. **Social Network Analysis in the Social and Behavioral Sciences. Social Network Analysis: Methods and Applications**. Cambridge: Cambridge University Press, 1994.