



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# **Um Sistema Gerenciador de Workflows Científicos Para a Plataforma de Nuvens Federadas BioNimbuZ**

Vinícius de Almeida Ramos

Monografia apresentada como requisito parcial  
para conclusão do Bacharelado em Ciência da Computação

Orientador  
Prof. Dr. Aleteia Patricio Favacho

Brasília  
2016

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Ciência da Computação

Coordenador: Prof. Dr. Andre Drummond

Banca examinadora composta por:

Prof. Dr. Aleteia Patricio Favacho (Orientador) — CIC/UnB  
Prof. Dr. <sup>a</sup> Maria Emília Machado Telles Walter — CIC/UnB  
Dr. <sup>a</sup> Maria Inez Machado Telles Walter — DPO/UnB

### **CIP — Catalogação Internacional na Publicação**

Ramos, Vinícius de Almeida.

Um Sistema Gerenciador de Workflows Científicos Para a Plataforma de  
Nuvens Federadas BioNimbuZ / Vinícius de Almeida Ramos. Brasília :  
UnB, 2016.

45 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2016.

1. Computação em Nuvem, 2. Cluster, 3. Grid, 4. Workflow Científico,  
5. Sistema Gerenciador de Workflows Científicos, 6. Ciência da  
Computação, 7. Universidade de Brasília

CDU 004.4

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



# Um Sistema Gerenciador de Workflows Científicos Para a Plataforma de Nuvens Federadas BioNimbuZ

Monografia apresentada como requisito parcial  
para conclusão do Bacharelado em Ciência da Computação

Prof. Dr. <sup>a</sup> Maria Emília Machado Telles Walter      Dr. <sup>a</sup> Maria Inez Machado Telles Walter  
CIC/UnB      DPO/UnB

Prof. Dr. Andre Drummond  
Coordenador do Bacharelado em Ciência da Computação

Brasília, 2 de Fevereiro de 2016

# Dedicatória

"À minha família e amigos, que sempre estiveram comigo me apoiando e dando forças."  
- Vinícius de Almeida Ramos

# Agradecimentos

"Agradeço à Deus por tudo que Ele proporcionou, agradeço também aos meus lindos pais Luiz Carlos e Júlia Alencar que sempre tiveram paciência, dedicação e amor em todos os momentos de minha vida. Agradeço à minha querida e maravilhosa irmã Marina Alencar que me deu suporte não somente durante minha graduação, mas sim durante meus vinte e três anos e também ao meu cunhado Leon Sólon. Agradeço imensamente ao meu atual noivo e futuro marido Vinícius Ramos que ajudou em todos os momentos difíceis e comemorou vitórias ao meu lado.

Agradeço à todos meus colegas de graduação, em especial meu amigo Yago Santos que esteve comigo desde o início da graduação e que realizou esse trabalho ao meu lado, Guilherme Resende, José Fonseca, Érico Vinícius, Jéssica Naiara e Luísa Behrens que estiveram por perto durante toda essa jornada. Agradeço também aos meus amigos externos à UnB, que entenderam os momentos de ausência.

Agradeço aos professores que me apoiaram e principalmente aos meus queridos orientadores Jan Mendonça Correa e Maria Emilia Machado Telles Walter que não mediram esforços para nos ajudar."

- Luiza Alencar Azevedo

# Resumo

A necessidade por maior poder de processamento e armazenamento, consequência da complexidade das atuais aplicações e sistemas, tem dado espaço para o desenvolvimento de novos paradigmas na Computação. Com isso, criou-se o conceito de Computação em Nuvem. Essa nova forma de se prover serviços computacionais tem possibilitado o desenvolvimento e a criação de diversas aplicações que compartilham diferentes tecnologias e provedores de serviços. Neste cenário, aplicações em Bioinformática tem se beneficiado dessa nova plataforma, devido à exigência de quantidades cada vez maiores de processamento. O BioNimbuZ, plataforma de execução de workflows em Bioinformática desenvolvido na Universidade de Brasília pelo aluno Hugo Saldanha, utiliza o paradigma de Computação em Nuvem para processar fluxos de aplicações em diferentes provedores de serviços computacionais, como Microsoft Azure e Amazon EC2. Dessa forma, faz-se necessário o gerenciamento da execução desses fluxos (workflows) desde sua submissão ao sistema até sua completude, tal como o provimento de uma interface para que o usuário possa ter acesso a esses serviços. Este trabalho propõe melhorias no gerenciamento e controle dos workflows submetidos à plataforma BioNimbuZ e também o desenvolvimento de uma aplicação baseada em tecnologias Web e Webservices para prover uma nova interface de acesso ao BioNimbuZ.

**Palavras-chave:** Computação em Nuvem, Cluster, Grid, Workflow Científico, Sistema Gerenciador de Workflows Científicos, Ciência da Computação, Universidade de Brasília

# Abstract

The need for greater processing power and storage, caused by the complexity of today's applications and systems, has given space to the development of new paradigms in computing. Thus, it created the concept of Cloud Computing. This new way of providing computing services has enabled the development and creation of various applications that share different technologies and service providers. In this scenario, applications in Bioinformatics has benefited from this new platform due to demand increasing amounts of processing. The BioNimbuZ, a Bioinformatics workflow execution platform developed at the University of Brasilia by the student Hugo Saldanha uses the Cloud Computing paradigm to process application flows in different computer services providers, such as Microsoft Azure and Amazon EC2. Thus, it is necessary to manage the execution of these flows (workflows) since its submission to the system until their completion, such as provide an interface for the user to have access to these services. This paper proposes improvements in the management and control of workflows undergoing BioNimbuZ platform and also the development of an application based on Web technologies and Web services to provide a new interface to access BioNimbuZ.

**Keywords:** Cloud Computing, Cluster, Grid, Scientific Workflow, Scientific Workflow Management System, Computer Science, University of Brasilia

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	2
1.2	Problema . . . . .	2
1.3	Objetivo . . . . .	3
1.3.1	Principal . . . . .	3
1.3.2	Específicos . . . . .	3
<b>2</b>	<b>Computação em Nuvem</b>	<b>4</b>
2.1	Sistemas Distribuídos . . . . .	4
2.1.1	Cluster Computacional . . . . .	5
2.1.2	Grid Computacional . . . . .	6
2.2	Nuvem Computacional . . . . .	7
2.2.1	Conceitos Básicos . . . . .	7
2.2.2	Arquitetura . . . . .	9
2.2.3	Tipos de Nuvens . . . . .	9
2.2.4	Modelos de Serviço . . . . .	10
2.3	Federação de Nuvens . . . . .	10
2.4	Workflow Científico . . . . .	10
2.5	Sistema Gerenciador de Workflows Científicos . . . . .	10
<b>3</b>	<b>BioNimbuZ</b>	<b>11</b>
<b>4</b>	<b>Proposta de Nova Arquitetura e Mudanças Realizadas</b>	<b>12</b>
<b>5</b>	<b>Resultados Obtidos com as Alterações</b>	<b>13</b>
<b>6</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>14</b>
	<b>Referências</b>	<b>15</b>



# Lista de Figuras

2.1	Arquitetura de um Cluster . . . . .	5
-----	-------------------------------------	---

# Lista de Tabelas

# Capítulo 1

## Introdução

A busca por cada vez mais poder de processamento tem desenvolvido novos estudos e paradigmas na Computação. Grids computacionais, clusters e, mais atualmente, nuvens computacionais têm tentado suprir essa necessidade de maneiras distintas. Em 1969, Leonard Kleinrock, um dos cientistas que chefiou o projeto ARPANET (o qual se tornou a base da Internet) disse que a computação funcionaria a partir de um modelo como vemos atualmente na telefonia e na eletricidade, como um serviço. Neste modelo, usuários acessam esses serviços independentemente de onde os mesmos estão localizados ou de qual maneira são entregues, abstraindo diversas características do ambiente envolvido. Esse modelo se aproxima do conceito de Computação em Nuvem.

Entretanto, não existe na literatura consenso para a definição de computação em nuvem, porém algumas características fundamentais [2] estão presentes na maioria delas, como: virtualização, escalabilidade, interoperabilidade, qualidade de serviço (QoS), gerenciamento de falhas, transparência, elasticidade. Esses aspectos fundamentais constituem o modelo de nuvens computacionais como serviços, em que o sistema passa a ilusão ao usuário de que este possui acesso à recursos ilimitados de software e hardware. Nesse contexto, surgiram os tipos de nuvens: privada, pública, híbrida e, mais atualmente, a federação de nuvens.

Uma Federação de Nuvens compreende o uso de diversos provedores em um único serviço. Isso provê características adicionais aos modelos anteriores de nuvem pública, privada e híbrida, como: migração de recursos (como imagens de máquinas virtuais), redundância de dados, processamento paralelo, replicação de recursos, combinação de serviços complementares, fragmentação de dados (ex: itens do tipo 1 são armazenados no provedor A enquanto itens do tipo 2 são armazenados no provedor B. Isso se torna útil quando requisitos funcionais e não funcionais diferem para tipos de dados diferentes). Adicionalmente, uma federação de nuvens possibilita o desenvolvimento de sistemas flexíveis e interoperáveis, o que diminui os custos de desenvolvimento e facilita sua expansão, com o custo de se adicionar complexidade ao sistema.

Neste cenário, a Bioinformática tem se beneficiado com esse conceito de nuvens computacionais pela sua característica de tratar grandes quantidades de dados, produzidas pelas modernas máquinas que executam algoritmos de sequenciamento genômico. Dessa forma, diversas ferramentas foram projetadas e implantadas tirando proveito dos recursos disponibilizados pela computação em nuvens. O BioNimbuZ, projeto desenvolvido por Hugo Saldanha [3], faz uso da infraestrutura de uma federação de nuvens para execu-

tar workflows em Bioinformática de maneira transparente, flexível, eficiente e tolerante à falhas, com acesso à grande poder de processamento e armazenamento.

Na Bioinformática, um workflow é um conjunto de diversas fases em que análises computacionais são executadas a partir de dados obtidos por meio de sequenciadores automáticos. Cada pesquisa implica em uma combinação de diferentes ferramentas já existentes ou a serem desenvolvidas, o que adiciona complexidade ao sistema, pois torna-o mutável a cada nova pesquisa. Sistemas científicos que gerenciam workflows [4] devem automatizar a execução de workflows científicos, suportando usuários na montagem, composição e verificação da execução do workflow gerado pelo usuário.

Este trabalho propõe um modelo de sistema gerenciador de workflows científicos que trata o problema de gerenciamento dos workflows submetidos ao ambiente de nuvem federada BioNimbuZ, além de implementar uma interface baseada em tecnologias Web para que usuários possam acessar os serviços disponibilizados pelo BioNimbuZ através da Internet. Também trata do problema de comunicação entre este sistema e o núcleo do BioNimbuZ, utilizando Webseices.

## 1.1 Motivação

Com o número crescente de sistemas computacionais utilizados na Bioinformática para execução de workflows científicos, percebeu-se a necessidade de criação de sistemas de que gerenciem o ciclo de vida destes workflows. Este ciclo de vida é composto por diversas fases, como [4]: Criação e Composição, Planejamento de Recursos, Execução, Análise da Execução, Compartilhamento de Resultados.

Diante deste contexto, este trabalho trata da criação de um sistema de gerenciamento de workflows científicos para o ambiente de nuvem federada BioNimbuZ para que seja possível a criação, manutenção, execução e análise dos workflows submetidos à esta plataforma tal como trata do problema de comunicação entre este sistema e o núcleo do BioNimbuZ.

## 1.2 Problema

Atualmente, a execução de um dado software no BioNimbuZ é realizado via linha de comando (terminal) de maneira sequencial, isto é, o usuário deve, primeiramente, fornecer os arquivos à plataforma, depois iniciar um dado serviço que consumirá esses arquivos enviados, resultando em uma saída. Caso haja mais passos no workflow, o usuário deve executá-los manualmente e assim sucessivamente. Ou seja, não há suporte à criação e gerenciamento de workflows, pois o usuário deve intervir em todos os passos da execução. Assim, este trabalho deve:

- Prover meios para que o usuário possa gerenciar seus workflows
- Implementar uma interface composta por tecnologias Web para facilitar o controle do ciclo de vida de um workflow pelo usuário.
- Garantir que estes usuários tenham acesso posterior aos workflows criados e executados.

- Tratar da comunicação entre este sistema e o BioNimbuZ a partir de Webservices.
- Garantir o controle de acesso de um dado usuário à somente seus workflows.
- Garantir a visualização posterior e devolver o resultado íntegro ao usuário.

## **1.3 Objetivo**

### **1.3.1 Principal**

Propor e implementar um Sistema Gerenciador de Workflows Científicos para que o usuário possa: compor um workflow de maneira gráfica, enviar arquivos necessários à sua execução, salvar o estado do workflow para que o usuário tenha acesso posterior, enviar o workflow para ser executado pelo núcleo do BioNimbuZ e, ao término da execução, enviar o resultado e os arquivos de saída de volta ao usuário.

### **1.3.2 Específicos**

# Capítulo 2

## Computação em Nuvem

O objetivo deste capítulo é mostrar os conceitos envolvidos no paradigma de computação distribuída, mais especificamente os de Computação em Nuvem, abordando suas características e suas particularidades. Também descreve as chamadas Federações de Nuvens Computacionais e como este trabalho pretende utilizá-las para integrar um Sistema Gerenciador de Workflows Científicos à atual plataforma do BioNimbuZ. Para isso, a Seção 2.1 mostra o histórico da computação distribuída e seus principais elementos (como Grids e Clusters), a Seção 2.2 apresenta os conceitos básicos que norteiam o entendimento sobre nuvens computacionais, mostrando como diversas pesquisas contribuíram para o conceito atual desta tecnologia. A Seção 2.3 mostra os detalhes da chamada Federação de Nuvens Computacionais, quais seus objetivos, suas características e o que a distingue dos outros modelos de computação distribuídas. A Seção 2.4 mostra o conceito de um Workflow Científico e a Seção 2.5 descreve o que é um Sistema Gerenciador de Workflows Científicos.

### 2.1 Sistemas Distribuídos

A Internet é utilizada por bilhões de pessoas com vários propósitos diferentes, como ler e-mails, visualizar conteúdo multimídia, fazer compras ou apenas realizar uma busca por um assunto de interesse. Isso passa ao usuário a ilusão de que a informação e o sistema que a provê se encontram localmente em sua máquina. Mas, a internet representa um enorme sistema distribuído que se parece como um recurso único disponível em com um conjunto mínimo de configuração de conexão e em apenas poucos cliques [1].

O conceito de sistema distribuído possui diversas definições e pontos de vista. Colouris define um sistema distribuído como “um sistema em que componentes de hardware e software comunicam e coordenam suas ações apenas trocando mensagens” [7]; Tanenbaum o define como “uma coleção de computadores independentes que aparecem ao usuário do sistema como um único computador” [8]; Keith Marzullo [9] define um sistema distribuído da seguinte maneira: “Um sistema distribuído é uma coleção de processos sequenciais  $P_1, P_2, \dots, P_n$  e uma rede capaz de implementar canais de comunicação unidirecionais entre pares de processos para troca de mensagem”.

Um dos aspectos mais importantes de sistemas distribuídos é a transparência. Esse conceito Pelas definições, é possível elencar alguns aspectos comuns e essenciais que nos ajuda a distinguir sistemas distribuídos:

- São formados por um conjunto de computadores (ou unidades de processamento);
- São conectados por uma rede, sua comunicação é feita através de troca de mensagens, portanto não compartilham memória;
- São vistos pelo usuários como um recurso único (transparência [8]);
- Facilitam a utilização de recursos por usuários (ou aplicações).
- São escaláveis, isto é, possibilitam o aumento ou a diminuição de usuários ou recursos de maneira facilitada.
- Possuem desafios de temporização e sincronização. Assim, embora seja possível perceber os desafios inerentes à implementação deste tipo de sistema, o ganho de poder computacional vale à pena esse esforço.

### 2.1.1 Cluster Computacional

A primeira iniciativa na construção do conceito e implementação de Clusters Computacionais foi realizada pela IBM em meados dos anos 60 como alternativa de interligar grandes Mainframes para prover aos seus usuários uma forma comercial mais eficiente de paralelismo. Contudo, o conceito de Clusterização não ascendeu como previsto até o surgimento de outras três tecnologias: microprocessadores de alta-performance, redes de alta velocidade e protocolos para computação distribuída de alta-performance [13]. Os recentes avanços nessas tecnologias, somadas à sua acessibilidade (baixo preço decorrente da alta demanda) facilitaram a implementação de clusters computacionais como solução do antigo problema da eficiência de custos na busca de sistemas paralelos de alta performance.

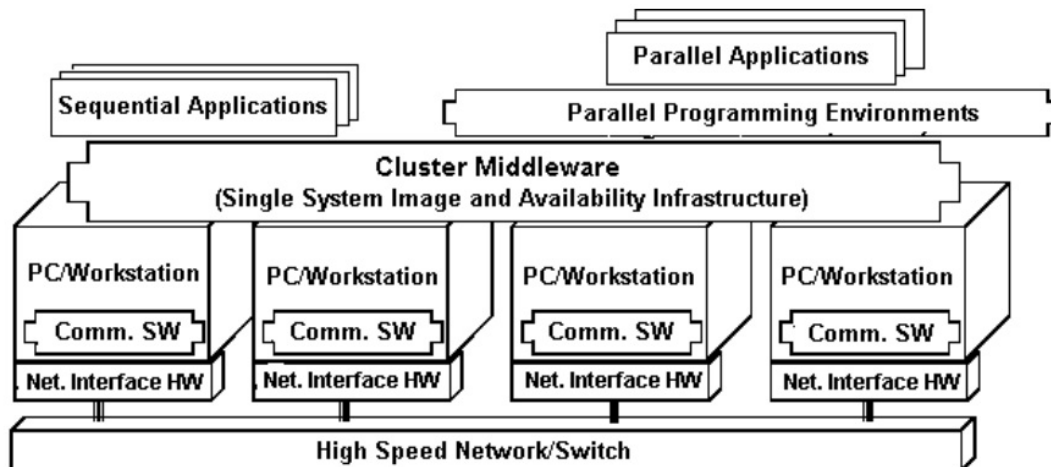


Figura 2.1: Arquitetura de um Cluster

Abaixo estão listados alguns componentes essenciais em computadores utilizados em clusters [14]:

- Vários computadores de alta performance (PCs, Workstations, Mainframes).
- Sistemas Operacionais.

- Conectados por uma rede de alta performance (como Gigabit Ethernet).
- Middleware de gerenciamento de Clusters (serviços e abstrações que facilitam o desenvolvimento de aplicações distribuídas. O middleware permite ao cluster manter a uniformidade na presença de diferentes hardwares e SOs).
- Ambiente de computação paralela.
- Aplicações.

Diversos projetos acadêmicos surgiram para dar uma face ao conceito de clusters, como o Beowulf [14], Berkeley NOW [15] e HPVM [16]. O objetivo destes projetos foi provar o ganho de performance de clusters sobre as plataformas tradicionais de sistemas distribuídos.

Beowulf foi um cluster construído em 1994 por Thomas Sterling e Don Becker e consistia de 16 processadores DX4 conectados por um canal Ethernet, dedicados à programação paralela [14]. O cluster criado foi muito bem visto pela área acadêmica e comercial, tanto que a NASA se interessou por este novo modelo. Em pouco tempo, o cluster Beowulf foi bastante difundido, tornando-se “Projeto Beowulf” e passou a ser visto como gênero dentro da comunidade de Computação de Alta Performance (High Performance Computing).

Berkeley NOW (Network of Workstations) difere do cluster Beowulf pois seus nós computacionais são computadores completos - PCs com teclado, mouse e som - conectados pela internet (os nós de um Cluster Beowulf são nós especializados, não servindo portanto como um computador pessoal). Na maioria dos casos, nós NOW são utilizados à noite ou nos fins de semana, quando não estão sendo utilizados. Outro modelo de utilização destes nós ocorre quando o servidor utiliza parte do poder de processamento, utilizando ciclos ociosos de CPU, agregando-os pela internet.

Esse modelo de computação distribuída tinha diversas limitações quanto ao tipo de aplicação que seria executado no cluster NOW. Nesse contexto, o cluster Beowulf tinha um melhor desempenho pelos seguintes aspectos:

- Possui processadores dedicados.
- Utilizava redes privadas de alta performance.
- Seu software era customizável.
- Softwares podiam ser clonados e enviados pela internet

### 2.1.2 Grid Computacional

Em meados dos anos 90 o termo Grid Computacional foi criado para descrever tecnologias que possibilitariam usuários obterem poder computacional sobre demanda, assim como a energia é disponibilizada pela rede elétrica. Se distingue do modelo tradicional de Computação Distribuída pelo foco em compartilhamento de grandes quantidades de recursos e, geralmente, por ser orientada à computação de alta performance. É uma forma de computação que envolve coordenar e compartilhar recursos computacionais, aplicações, dados, armazenamento e/ou recursos de rede através de sistemas dinâmicos e geograficamente dispersos [11] de maneira flexível, segura e coordenada servindo à chamadas Organizações Virtuais [12]. Uma Organização Virtual é um conjunto de indivíduos e/ou



instituições interessados em acessar diretamente máquinas, software, dados e outros recursos. Esse compartilhamento é extremamente controlado, com regras definindo o que está, com quem e as condições daquilo que está sendo compartilhado.

Um sistema gerenciador de um Grid Computacional não está somente voltado ao processamento de dados, mas também ao gerenciamento de recursos de todo o sistema, seja ele software (aplicações, protocolos, sistemas operacionais) ou hardware (armazenamento, consumo de CPU, utilização de memória) e está dividido em 5 camadas [1]:

- **Camada de Rede:**
- **Camada de Conectividade:** Define o método de comunicação e protocolos de autenticação para transações específicas da rede interna do Grid. Provê mecanismos de segurança criptográfica para verificar a identidade de usuários e recursos.
- **Camada de Recursos:** Define protocolos para transações, inicialização, monitoração, controle e custos de operações que utilizam recursos individuais. Esta camada se preocupa apenas com o seu conjunto de recursos, não se importando com o estado global dos recursos compartilhados.
- **Camada de Serviços Coletivos:** Enquanto a Camada de Recursos está ligada à interações com recursos únicos, a Camada de Serviços Coletivos tem o objetivo de capturar e gerenciar transações entre conjuntos de recursos.
- **Camada de Aplicação:** A última camada na arquitetura de Grids abrange as aplicações que operam dentro do ambiente de uma Organização Virtual, ou seja, esta camada está preocupada nas aplicações clientes que são executadas dentro de um grupo que compartilha o mesmo recurso.

### **Imagem da arquitetura de um Grid Computacional**

Um Grid Computacional emprega múltiplos clusters que são fracamente acoplados (), heterogêneos e geograficamente dispersos [1]. Sua arquitetura está dividida nas seguintes 5 camadas

Apesar de terem características similares, alguns aspectos diferem o entendimento sobre Nuvens Computacionais e Grids, como: - Modelo de Negócio: Grids seguem o modelo tradicional de negócio, ou seja, geralmente

## **2.2 Nuvem Computacional**

### **2.2.1 Conceitos Básicos**

Com o rápido desenvolvimento de tecnologias de processamento e armazenamento de dados e o crescimento da Internet, recursos de computação tornaram-se mais acessíveis, mais poderosos e mais disponíveis do que nunca. Essa tendência tecnológica permitiu a realização de um novo modelo de computação chamada computação em nuvem, em que os recursos computacionais (CPU, rede, memória, armazenamento) são fornecidos como serviços que podem ser alugados e liberados pelos usuários através da Internet em um modelo sob-demanda. Esse conjunto de tecnologias provê aos usuários uma gama enorme de opções de usos dessa nova tecnologia. Com ela, os serviços são disponibilizados de maneira transparente, representando uma nova maneira de se utilizar recursos computacionais.

A ideia principal por trás da Computação em Nuvem não é exatamente nova. Em meados dos anos 60, John McCarthy já havia previsto que o modelo de computação seria provido ao público de uma maneira similar como acontece em redes elétricas, por exemplo. Isso aconteceu quando grandes empresas capazes de implantar enormes *datacenters* a custos competitivos resolveram adotar esse novo paradigma da computação. Em 2006, o termo "nuvem" ganhou popularidade quando o CEO da Google, Eric Schmidt, o utilizou para descrever o modelo de negócios para prestação de serviços pela internet. Desse momento em diante, o termo "nuvem" realmente começou a ficar popular com ações de marketing e com o nascimento de diversas empresas especializadas.

A partir deste ponto, diversas definições surgiram em busca de um padrão para o conceito de Computação em Nuvem. Foster *et al.* a definem [1] como um paradigma de computação distribuída em larga escala movida pela economia da indústria, na qual um grande conjunto de recursos computacionais são providos sob-demanda à usuários externos pela Internet. Isso reforça a ideia de McCarthy de computação provido como serviços. Armbrust *et al.* trazem a definição de Computação em Nuvem [17] como a união das aplicações disponibilizadas como serviços pela Internet e o hardware nos *datacenters* que as provém.

NIST (*National Institute of Standards and Technology*) a define como "um modelo que possibilita o acesso ubíquo, conveniente, sob-demanda à um conjunto configurável de recursos computacionais (redes, servidores, armazenamento, aplicações, e serviços) que podem ser rapidamente provisionados e liberados com um esforço mínimo de gerenciamento ou interação do provedor de serviço"[16].

Na indústria de Tecnologia da Informação (TI), o surgimento do modelo de Computação em Nuvem tem causado um tremendo impacto nos últimos alguns anos, onde grandes empresas como Google, Amazon e Microsoft se esforçam para fornecer plataformas de nuvem cada vez mais poderosas, disponíveis, de confiança e com melhor custo-benefício. Ao mesmo tempo, empresas procuram reformular seus modelos de negócios para se utilizarem dos benefícios deste novo paradigma. Dessa forma, a computação em nuvem fornece vários recursos interessantes que a torna atraente para empresas, como [5]:

- **Sem investimento inicial:** A computação em nuvem usa um modelo de precificação em que o usuário paga apenas pelo recurso computacional utilizado. Um prestador de serviço não precisa investir na infra-estrutura para começar a se utilizar dos ganhos tecnológicos de uma nuvem computacional. Ele simplesmente aluga os recursos de acordo com suas próprias necessidades e pagar por aquilo que utilizar.
- **Reduz o custo operacional:** Recursos em um ambiente de nuvem podem ser rapidamente alocados e desalocados sob-demanda. Assim, um prestador de serviços não precisa mensurar sua capacidade de acordo com a carga de pico (carga máxima). Isso proporciona uma enorme economia uma vez que os recursos podem ser liberados para economizar custos de serviço quando a demanda é baixa.
- **Altamente escalável:** Provedores de infra-estrutura possuem grandes quantidades de recursos a partir de *datacenters* e os tornam facilmente acessíveis. Um provedor pode facilmente expandir sua capacidade, a fim de lidar com um possível rápido aumento em exigência de serviço.

- **Acessibilidade:** Serviços hospedados na nuvem são geralmente baseado na web. Assim, são facilmente acessíveis através de uma grande variedade de dispositivos com conexões de Internet. Estes dispositivos não só incluem computadores desktop e notebooks, mas também celulares e tablets.
- **Reduz os riscos de negócios e despesas de manutenção:** Por terceirizarem os serviços de infra-estrutura para as nuvens, prestadores de serviços transferem os riscos empresariais (tais como falhas de hardware) para os provedores de infraestrutura, que muitas vezes têm um maior conhecimento e estão melhor equipados para o gerenciamento desses riscos.

No entanto, embora a computação em nuvem tem mostrado consideráveis oportunidades para a indústria de TI, também traz muitos desafios únicos que precisam ser cuidadosamente abordados.

### 2.2.2 Arquitetura

### 2.2.3 Tipos de Nuvens

As nuvens computacionais podem ser divididas em quatro tipos diferentes de implantação: nuvens públicas, privadas, comunitárias ou híbridas [16].

- **Nuvens Públicas:** Sua infraestrutura é disponibilizada para o público em geral, mantida e gerenciada por uma instituição acadêmica, comercial ou governamental e não impõe condições para sua utilização. Seus recursos computacionais são provisionados dinamicamente e são acessíveis através da internet (geralmente com a utilização de *webservices*).
- **Nuvens Privadas:** A utilização da infraestrutura desse tipo de nuvem é exclusivo de uma única companhia ou grupo de empresas, compreendendo múltiplos usuários. Seu gerenciamento pode ser feito pela própria instituição, por outra empresa ou pela junção de ambas as partes.
- **Nuvens Comunitárias:** Esse tipo de nuvem é provisionada para a utilização exclusiva de uma comunidade específica de consumidores de uma organização com interesses em comum (por exemplo: organizações com o mesmo requisito em segurança, mesmos requisitos em performance). Sua implantação pode ocorrer dentro ou fora da organização.
- **Nuvens Híbridas:** Esse tipo de infraestrutura é uma composição de dois ou mais tipos de nuvens (públicas, privadas ou comunitárias) em que o provedor de serviços disponibiliza tecnologias proprietárias que permitem portabilidade de dados e aplicações entre as nuvens que a compõe. Essa infraestrutura é muito útil quando, por exemplo, se quer trafegar dados sensíveis e sigilosos em um ambiente controlado e gerenciado internamente (utilização de uma nuvem privada) enquanto outros dados e aplicações podem trafegar em nuvens públicas ou comunitárias.

### 2.2.4 Modelos de Serviço

A Computação em Nuvem emprega um modelo de negócios orientado a serviços, ou seja, recursos de *hardware* e *software* são disponibilizados sob-demanda. Conceitualmente, cada camada da arquitetura descrita na subseção anterior pode ser implementada como um serviço para a camada acima. Em outras palavras, a camada acima será a consumidora dos serviços providos pela camada imediatamente abaixo. Esses serviços são categorizados em três modelos, conforme abaixo:

- ***Infrastructure-as-a-Service*** (IaaS): provê recursos computacionais ao usuários, tais como poder computacional, armazenamento de dados e redes virtuais para que possam implantar e executar qualquer tipo de software. A tecnologia de virtualização é essencial para este modelo, pois dá ao provedor deste serviço a habilidade de que vários usuários possam compartilhar recursos de uma mesma máquina física. Exemplos de provedores deste modelo: Amazon EC2, Google Compute Engine e GoGrid.
- ***Platform-as-a-Service*** (PaaS): Dá ao consumidor deste serviço a capacidade de implantar aplicações criadas pelo usuário ou adquiridas de terceiros, criadas a partir de linguagens de programação, bibliotecas de software, *APIs* e ferramentas suportadas pelo provedor. Nesse modelo, o usuário não gerencia aspectos do hardware e do software necessários para execução da infraestrutura da nuvem.
- ***Software-as-a-Service*** (SaaS): Provê aplicações que estão sendo executadas em algum ponto da Internet por um provedor IaaS, eliminando a necessidade de instalar e executar a aplicação no computador do usuário. São independentes de plataforma e geralmente provêm uma interface para que o usuário possa acessar e utilizar esse serviço.

## 2.3 Federação de Nuvens

Federação de Nuvens

## 2.4 Workflow Científico

Workflow Científico

## 2.5 Sistema Gerenciador de Workflows Científicos

Sistema Gerenciador de Workflows Científicos

# Capítulo 3

## BioNimbuZ

Introduzir e descrever o BioNimbuZ em seu estado atual (anterior ao desenvolvimento desta Monografia) - Camadas - Módulos - Provedores (Amazon, Microsoft, UnB)

## Capítulo 4

# Proposta de Nova Arquitetura e Mudanças Realizadas

Descrever a arquitetura do BioNimbuZ (Linha de comando, módulos, provedores, camadas...) - Demonstrar o funcionamento anterior (linha de comando) e propor uma nova interface que dê suporte ao desenvolvimento do novo módulo controlador de Jobs. - Propor uma nova arquitetura baseada em aplicação WEB e comunicação com o núcleo via webservices (Protocolo de troca de mensagens, tratamento de erros, \*\* solução de envio de arquivos \*\*). - Propor uma nova política de upload de arquivos baseado em paralelização. - Descrever as mudanças propostas (Protocolo de comunicação via REST, nova interface, gerenciamento da execução de Jobs, nova política de envio de arquivos...) - Explicar os temas: webservices, REST, servidor de aplicação, aplicação WEB, gerenciamento de Jobs. - Quais as vantagens e desvantagens das soluções propostas

Descrever mudanças realizadas no BioNimbuZ - Aplicação WEB - Protocolo de Comunicação via WebServices - Interface de Comunicação "Aplicação - Núcleo- Gerenciador de Estado dos Jobs (JobController)

## Capítulo 5

# Resultados Obtidos com as Alterações

Descrever estudos de caso com a nova plataforma do BioNimbuZ (login, envio de arquivos, execução de Jobs, gerenciamento da execução de uma tarefa, estatísticas de execução, histórico de execuções...). - Realizar testes de envio de arquivo com a nova política de upload. - Realizar testes de execução de Jobs.

## Capítulo 6

# Conclusão e Trabalhos Futuros

Descrever o que pode ser melhorado na solução proposta e quais os próximos passos.



## Referências