

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA AMBIENTAL

“CIRRUS: INFRAESTRUTURA DE *e-science* PARA
CONTEXTO AMBIENTAL”

César Eduardo Guarienti

Orientador: **Prof. Dr. Josiel Maimone de Figueiredo**

Cuiabá - MT
Fevereiro/2016

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA AMBIENTAL

**Cirrus: infraestrutura de *e-science* para contexto
ambiental**

César Eduardo Guarienti

Dissertação apresentada ao Programa de Pós Graduação em Física Ambiental da Universidade Federal de Mato Grosso, como parte dos requisitos para obtenção do título de Mestre em Física Ambiental.

Prof. Dr. Josiel Maimone de Figueiredo

Cuiabá, MT
Fevereiro/2016

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
Programa de Pós-Graduação em Física Ambiental

FOLHA DE APROVAÇÃO

TÍTULO: CIRRUS: INFRAESTRUTURA DE *e-SCIENCE* PARA
CONTEXTO AMBIENTAL

AUTOR: CÉSAR EDUARDO GUARIENTI

Dissertação de Mestrado defendida e aprovada em 18 de fevereiro de 2016, pela
comissão julgadora:


Prof. Dr. Raphael de Souza Rosa Gomes
Orientador
Instituto de Computação – UFMT


Prof. Dr. Josiel Maimone de Figueiredo
Examinador Interno
Instituto de Computação – UFMT


Prof. Dr. Allan Gonçalves de Oliveira
Examinador Interno
Instituto de Computação – UFMT


Prof. Dr. Jonathan Willian Zangeski Novais
Examinador Externo
Universidade de Cuiabá – UNIC/Cuiabá

DEDICATÓRIA

Aos meus pais, Osmar e Maria, à minha
esposa Gracyeli, à nossa filha que está
chegando: Marieli, aos meus sobrinhos
Leandro e Rafaela.

Dados Internacionais de Catalogação na Fonte.

G915c Guarienti, César Eduardo.
Cirrus: infraestrutura de e-science para contexto ambiental / César Eduardo Guarienti. -- 2016
vi, 60 f. : il. color. ; 30 cm.

Orientador: Raphael Souza de Rosa Gomes.
Dissertação (mestrado) - Universidade Federal de Mato Grosso, Instituto de Física, Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2016.
Inclui bibliografia.

1. Dados ambientais. 2. Softwares ambientais. 3. E-research. I. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.

Agradecimentos

A Deus pelo dom da vida e por ser meu amigo de todas as horas.

Aos meus pais eu agradeço pelo apoio irrestrito em todos os momentos, por cada luta enfrentada, pelas lições de vida e pelo exemplo.

A minha esposa, que "come sal"comigo, pelo apoio, carinho e cuidado, que fez e faz diferença na minha vida. Te Amo. Obrigado!

A minha irmã Rosângela por todo apoio em vários momentos da vida.

Ao Thiago Ventura pelos ensinamentos, exemplos e ajuda de sempre.

Ao meu sobrinho querido, Leandro, por me ajudar a renovar minhas forças e minha esperança na vida. Menino bonito do tio.

A minha sobrinha Rafaela, por injetar felicidade em forma de sorriso. Coisa linda do tio.

Aos meus padrinhos pelo apoio pela vida inteira.

Ao professor Josiel, orientador, pelos ensinamentos, paciência e parceria desde que cheguei na UFMT.

Ao professor Raphael por toda ajuda em vários trabalhos.

Aos meus amigos por me emprestarem casa quando faltava abrigo.

Ao Thiago Silva e Igor pelo esforço dobrado durante minha ausência.

Ao Paraná, Marta e Cesário pelo gentil acolhimento no PPGFA.

A Paloma, Safira, Nick e Laica pelos grandes momentos de alegria e lealdade.

Aos meus professores pela enorme contribuição na minha caminhada.

Aos alunos, professores, técnicos e estagiários do IC e PPGFA que, direta ou indiretamente, me ajudaram.

"Não te mandei eu? Esforça-te, e tem bom ânimo; não temas, nem te espantes; porque o Senhor teu Deus é contigo, por onde quer que andares."

Josué 1:9

SUMÁRIO

LISTA DE FIGURAS	I
LISTA DE TABELAS	III
LISTA DE ABREVIATURAS	IV
RESUMO	V
ABSTRACT	VI
1 Introdução	1
1.1 Objetivo Geral	2
1.2 Objetivos Específicos	2
2 Fundamentação Teórica	4
2.1 Monitoramento Ambiental	4
2.2 Computação Científica	8
2.3 E-science	10
2.4 Computação em Nuvem	12
3 Materiais e Métodos	17
3.1 Tecnologia	17
3.2 Testes	18
4 Apresentação e Análise dos Resultados	20
4.1 Infraestrutura Básica	20
4.2 Instâncias	24

4.3	Serviços	25
4.4	Gerenciamento	29
4.5	Exemplos de uso na área ambiental	32
5	Considerações Finais	37
5.1	Contribuições	37
5.2	Trabalhos Futuros	38
	REFERÊNCIAS	40

LISTA DE FIGURAS

1	Modelo CRISP-DM. Fonte: Chapman et al. (2000)	7
2	Influência do Entendimento do Negócio em outras etapas do modelo de processos CRISP-DM. Adaptado de Oliveira (2015).	8
3	Exemplo de sistema de computação em <i>cluster</i> . Fonte: Adaptado de Tanenbaum e Steen (2007).	9
4	Exemplo de arquitetura em camada de computação em grade. Fonte: Adaptado de Tanenbaum e Steen (2007).	9
5	Camadas da Computação em Nuvem. Fonte: Adaptado de Slabela et al (2009).	15
6	Arranjo da infraestrutura <i>e-science</i>	22
7	Tabela de consumo de CPU por VM.	23
8	Gráfico de consumo de CPU de um determinado usuário.	23
9	Espaço para download de imagens.	24
10	Desenho da estrutura de cópia dos dados na infraestrutura	27
11	OpenNebula protegido por camada adicional de segurança (<i>proxy</i> reverso).	28
12	Instâncias protegidas por <i>firewall</i>	29
13	Resumo utilizando gráficos com informações sobre memória e CPU.	30
14	Definição de permissão dos grupos.	30
15	Ambiente de edição das instâncias.	31
16	Configuração do cluster.	31
17	Instância com SO Ubuntu.	31
18	Teste do <i>software</i> Matlab no SO Windows 7.	32

19	Teste do Matlab no Windows 8.1.	33
20	Simulação com Octave.	33
21	Criação de gráfico com R e Rstudio.	34
22	Exemplo de uso do QuantumGIS.	34
23	Código python executado via terminal.	35
24	<i>Software</i> ENVI-met em execução em VM Windows.	35

LISTA DE TABELAS

1	Características de pesquisa tradicional e de <i>e-science</i>	11
2	Requisitos mínimos de cada infraestrutura	20
3	Avaliação do consumo de memoria RAM utilizando quatro ambientes gráficos	26
4	Plano de cópia dos dados	27
5	Exemplo de permissão para acesso.	30

LISTA DE ABREVIATURAS

CN	Computação em Nuvem	2
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>	6
LAN	<i>Local Area Network</i>	8
LXDE	<i>Lightweight X11 Desktop Environment</i>	25
PPGFA	Programa de Pós Graduação em Física Ambiental	3
SO	Sistema Operacional	12
SSL	<i>Secure Socket Layer</i>	27
UFMT	Universidade Federal de Mato Grosso	2
VM	<i>Virtual Machines</i>	12

RESUMO

GUARIENTI, C. E. Cirrus: infraestrutura de *e-science* para contexto ambiental. Cuiabá, 2016, 60f. Dissertação (Mestrado em Física Ambiental), Instituto de Física, Universidade Federal de Mato Grosso.

A área ambiental utiliza de recursos computacionais para dar suporte às suas pesquisas, principalmente na automatização de atividades como: comunicação, proposição, teste de hipóteses e simulação de modelos. Nos últimos anos a quantidade de dados gerados tem atingido proporções gigantescas, com isso a manipulação e interpretação dessas informações demanda o uso de novas técnicas e infraestrutura que atendam minimamente essas necessidades. Dessa forma, o presente trabalho teve por objetivo criar e implantar o ambiente Cirrus, que é uma infraestrutura *e-science* baseada em tecnologias de computação em nuvem que favorece o uso de computação científica com acesso unificado via Web. Para validar o ambiente foram testados variados softwares relacionados a área ambiental, bem como a possibilidade de uso de sistemas operacionais variados. Os resultados evidenciam que é viável o uso de ambiente *e-science* para utilização científica propiciando uma infraestrutura que suporta diferentes necessidades do Programa de Pós-Graduação em Física Ambiental (PPGFA) da Universidade Federal de Mato Grosso (UFMT).

Palavras-chaves: dados ambientais; *softwares* ambientais; *e-research*.

ABSTRACT

GUARIENTI, C. E. Cirrus: infrastructure e-science to the environmental context. Cuiabá, 2016, 60f. Dissertation (Master in Environmental Physics), Institute of Physics, Federal University of Mato Grosso.

The environmental area uses computational resources to support their research, mainly in the automation of activities such as: communication, proposition, hypothesis testing and simulation models. In recent years the amount of data generated has reached gigantic proportions, thus the handling and interpretation of this information requires the use of new techniques and infrastructure that minimally meet those needs. Thus, this study aimed to build and deploy the Cirrus environment, which is an infrastructure it e-science based on cloud computing technologies that favors the use of scientific computing with a unified Web access. To validate the environment were tested various software related to the environmental area, as well as the possibility of use of various operating systems. The results show that it is viable the environment of use e-science for scientific use providing an infrastructure that supports the different needs of the Graduate Studies Program in Environmental Physics (PPGFA) of the Federal University of Mato Grosso (UFMT).

Keywords: environmental data; environmental software; e-research.

Capítulo 1

Introdução

"Cirrus é o nome de um tipo de nuvem constituída por cristais de gelo."

A difusão e o barateamento dos dispositivos de captação de dados contribuíram para que pesquisadores da área ambiental utilizassem recursos computacionais em suas pesquisas. Assim, a quantidade e variedade de dados gerados é cada vez maior, com isso a manipulação e interpretação dessas informações demandam uso de novas técnicas e infraestruturas que atendam minimamente essas características.

As tecnologias computacionais têm contribuído nos mais variados campos da atividade humana. A sua contribuição na produção de ciência tem ganhado destaque no cenário mundial. É possível citar, por exemplo, os benefícios conseguidos por meio do uso de ferramentas computacionais no auxílio de automação das atividades de pesquisa. Além disso, a computação facilita e auxilia os cientistas na proposição, teste de hipóteses, simulação de modelos. Outro ponto positivo é o favorecimento da comunicação entre cientistas e grupos de pesquisa distintos. É nesse contexto que o *e-science* está inserido.

Pesquisadores já utilizam recursos computacionais em suas atividades, mas novos desafios de pesquisa necessitam de novas soluções. Atualmente o método científico tem buscado apoio na área de Ciência da Computação para enfrentar desafios cada vez mais sofisticados. Nesse ponto a infraestrutura computacional precisa suportar diferentes necessidades dos pesquisadores como: trabalho colaborativo em ambientes virtuais na Internet, adição ou deleção de recursos computacionais instantaneamente, reuso de *hardware* e *software*, compartilhamento de diretórios e informações. Dessa forma, a ciência precisa receber

o suporte computacional para atender essas diferentes necessidades da pesquisa científica.

O *e-science* busca dar suporte computacional às atividades de pesquisa permitindo ao pesquisador o acesso a um conjunto de soluções que envolvem *software* e *hardware* (equipamento físico) sem que ele precise comprar ou configurar esse ambiente, assim, podendo se concentrar no desenvolvimento do seu estudo.

As tecnologias que ganharam maior destaque nos últimos anos são as relacionadas com a área de redes, que possibilitaram ao meio científico a disponibilização na Internet de uma infraestrutura completa, com capacidade de processamento e armazenamento. Essas novas tecnologias foram denominadas de Computação em Nuvem (CN), e permitem que a criação e configuração de máquinas fiquem ao alcance por meio de um simples *browser*, programa navegador de Internet.

As vantagens apresentadas pela Computação em Nuvem chamaram atenção do domínio científico, pois os experimentos necessitam de suporte computacional para desenvolver as atividades inerentes ao método científico. Essas características demonstram que um contexto de *e-science* é fortalecido com o uso de CN.

Este estudo visou a disponibilização de *e-science* aplicado à área ambiental utilizando tecnologia de Computação em Nuvem, sob demanda e acessível pela Internet, para utilização de pesquisadores que necessitem analisar, por exemplo, variáveis que interferem no clima.

1.1 Objetivo Geral

Implantar uma infraestrutura de *e-science* com o uso de Computação em Nuvem para um contexto ambiental.

1.2 Objetivos Específicos

Para atingir o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

- Determinar o tipo de nuvem computacional para criar ambiente *e-science* que melhor atenda as necessidades dos pesquisadores do Programa de Pós Graduação em Física Ambiental (PPGFA) da Universidade Federal de Mato Grosso (UFMT).

- Avaliar e implantar diferentes opções de nuvem computacional.
- Utilizar instâncias de diferentes Sistemas Operacionais.
- Disponibilizar recursos computacionais aos grupos de pesquisadores do PPGFA na forma de ambiente *e-science*.
- Compartilhar com a comunidade científica documentação necessária para implementar o cenário de *e-science* ambiental.
- Atender as características de segurança da informação, como confidencialidade e disponibilidade.
- Possibilitar o compartilhamento de recursos e informações entre grupos internos e externos de pesquisadores.

Capítulo 2

Fundamentação Teórica

2.1 Monitoramento Ambiental

Compreender a dinâmica do meio ambiente é fundamental para que a sociedade mantenha seu desenvolvimento e sobrevivência. Existem diversos estudos que buscam compreender a dinâmica do meio ambiente, em Curado et al. (2014) analisa-se padrões de comportamento dos fluxos de energia entre as estações seca e chuvosa, em Rodrigues et al. (2013) padrões do balanço de energia são estudados. Na defesa civil existem outros exemplos, na previsão de inundações em Estébanez (2012), manchas solares em Bolzan (2004), precipitação em Nicholson et al. (2003), análise espectral de radiação solar global, temperatura e umidade do ar (VILANI, 2011).

Esses estudos são realizados na área de meteorologia, sendo um ramo da ciência que lida com a atmosfera e os processos físicos que ocorrem nele em macro, meso e microescala (AHRENS, 2012). Os fenômenos de macroescala constituem em movimentos de comprimento horizontal da ordem maior ou igual a 1000 km, sendo limitado pela circunferência da Terra, e tempo igual ou maior que um dia. Os movimentos de macroescala na vertical podem ser ampliados por toda a troposfera (BUSKE, 2008). Estes fenômenos de macroescala são investigados pela macrometeorologia.

De acordo com Atkinson (1989) o estudo dos sistemas atmosféricos de dimensões na escala média, isto é, entre as dos fenômenos observados às escalas “local” e “sinóptica” dá-se o nome de mesometeorologia, já o estudo do ambiente em que está relacionado a vida da humanidade é investigado pela micrometeorologia (FOKEN; NAPPO, 2008).

Dados micrometeorológicos são comumente coletados por meio de sensores em estações micrometeorológicas, e possuem valores com medidas de variáveis

climáticas como temperatura, precipitação, umidade relativa do ar, dentre outras (VENTURA et al., 2013). Para realizar a análise de dados micrometeorológicos é adequado um planejamento com etapas que podem auxiliar o pesquisador em cada uma das atividades que visam monitorar o meio ambiente.

O monitoramento ambiental pode ser realizado por meio do estudo de medidas das variáveis microclimáticas e sua evolução no tempo, podendo ser realizadas em diversas localidades por períodos curtos ou longos de tempo. É possível também mensurar as ações decorrentes da influência da sociedade na natureza. Segundo Moresi et al. (2005), as funções do monitoramento ambiental são:

- Verificar se impactos ambientais estão ocorrendo;
- Dimensionar sua magnitude;
- Avaliar se as medidas tomadas são eficazes;
- Propor, quando necessário, a adoção de medidas mitigadoras complementares.

Por meio do monitoramento ambiental é possível obter informações sobre os fatores que influenciam o estado de conservação, preservação, degradação e recuperação ambiental auxiliando em tomadas de decisão (MORESI et al., 2005).

Segundo Machado (1995), o monitoramento ambiental pode ser realizado de maneira isolada ou integrada, auxiliando na elaboração de documentos para tomada de decisão. Nesses casos, o monitoramento do ambiente é essencial para obter resultados confiáveis, pois, sem o registro de medições ou observações de períodos anteriores, as decisões ficam restritas apenas a uma avaliação da situação presente. Desta forma, o monitoramento ambiental, que consiste no processo de coleta de dados, estudo e acompanhamento sistemático dos recursos ambientais pode auxiliar em análises ambientais.

As etapas do monitoramento ambiental são utilizadas por pesquisadores em diversos estudos inclusive com dados micrometeorológicos. Segundo Moresi et al. (2005), o ciclo é composto por 5 etapas, assim descritas:

- **Planejamento e Direção:** Nesta fase os pesquisadores realizam o estudo sobre as necessidades da pesquisa de determinado recurso ambiental. Esse procedimento é importante, pois viabiliza o conhecimento do problema e lista as necessidades de monitorar determinado ambiente.

- **Armazenamento e Processamento de Informações:** então são utilizados ferramentas para armazenar e processar as informações. Essa fase consiste no do ciclo das etapas de monitoramento ambiental.
- **Coleta e Relatórios Adequados:** que de modo geral, corresponde à fase em que os pesquisadores buscam os dados referentes à pesquisa em questão, e produzem relatórios desses dados verificando sua relevância.
- **Análises e Produção:** é feito a avaliação das hipóteses por meio de simulações com os dados obtidos da fase anterior, para que na fase seguinte seja avaliada a decisão considerada necessária,
- **Disseminação:** com base nos processos anteriores os pesquisadores realizam a tomada de decisão.

Nos trabalhos de Thamada et al. (2013) e Dourado et al. (2013) foi utilizado um método semelhante ao monitoramento ambiental chamado *Cross Industry Standard Process for Data Mining* (CRISP-DM) que a princípio foi utilizado como método para elencar fases de um processo em indústrias, mas que atualmente tem sido utilizado na área ambiental. Desta forma, o CRISP-DM, pode ser utilizado em qualquer área inclusive ambiental, para planejamento e tomada de decisão. Proposto em 1996, o CRISP-DM foi exposto a comunidade científica somente no ano 2000 por Shearer (2000) e Wirth e Hipp (2000) que apresentaram oficialmente em seus trabalhos.

As etapas do monitoramento ambiental possibilitam a análise de dados micrometeorológicos semelhante à realizada em mineração de dados. A mineração de dados é utilizada na análise de dados como os da bolsa de valores, porém também pode ser utilizado na área ambiental. De modo geral, a mineração de dados é vista com um caráter exploratório, devido a grande quantidade de dados coletados e armazenados (GONÇALVES, 2000). Em mineração de dados existem várias atividades que devem ser executadas, e, algumas vezes são executadas mais de uma vez sendo um processo iterativo (CORTES et al., 2002).

Cada etapa do modelo de processos para mineração de dados, CRISP-DM, pode ser visto na Figura 1, e são explicados na sequência:

- **Entendimento do negócio:** onde os objetivos do projeto são listados e a definição do problema é feita. Nesta etapa, no caso dos dados micrometeorológicos é feita análise dos problemas relacionados a variável estudada.

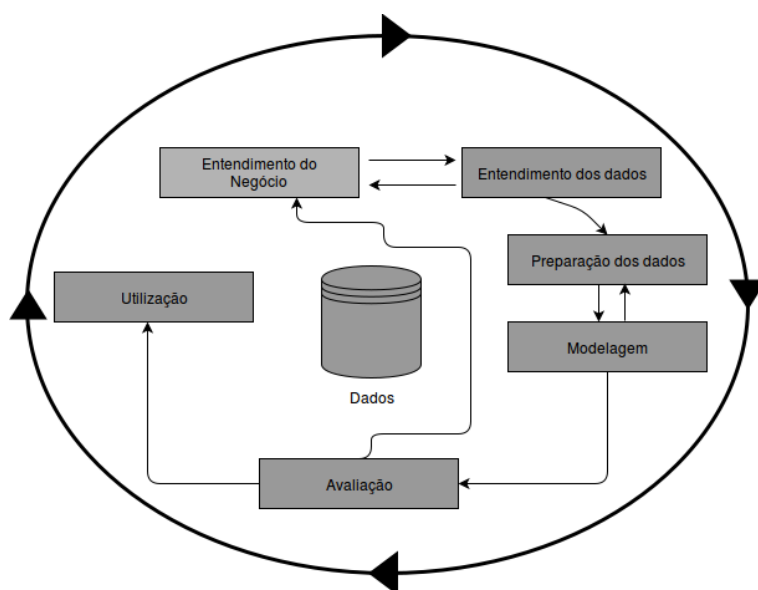


Figura 1: Modelo CRISP-DM. Fonte: Chapman et al. (2000)

- Entendimento dos dados: nesta etapa são executadas atividades que proporcionam entendimento prévio sobre o comportamento e os significados dos dados micrometeorológicos.
- Preparação dos dados: os dados são manipulados, e, no caso de dados micrometeorológicos que possuam falhas é realizado um trabalho de correção.
- Modelagem: nesta etapa várias técnicas como detecção de padrões e agrupamento, classificação são selecionadas e aplicadas nos dados micrometeorológicos.
- Avaliação: nesta etapa são avaliados os padrões encontrados e comportamento dos dados micrometeorológicos.
- Utilização: o conhecimento adquirido é consolidado e disseminado para a comunidade e utilizado para tomada de decisão.

Em Oliveira (2015) é proposto um novo entendimento sobre o CRISP-DM, altera-se o fluxo original modificando a interação entre as fases, essa interpretação é baseada na ligação entre o conhecimento levantado na etapa de Entendimento do Negócio e Entendimento dos dados. Na Figura 2 é possível verificar a interação de fases em linhas azuis.

Todo o processo de manipulação e análise dos dados micrometeorológicos como mostrado no monitoramento ambiental e no modelo CRISP-DM, consistem em atividades que buscam o entendimento para determinado problema. Esse

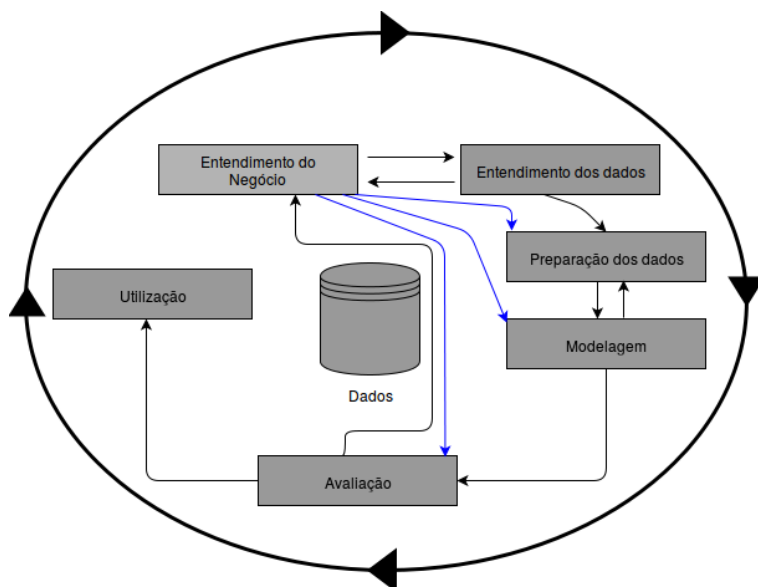


Figura 2: Influência do Entendimento do Negócio em outras etapas do modelo de processos CRISP-DM. Adaptado de Oliveira (2015).

processo é utilizado também na área da computação como pode ser observado nos trabalhos de Ventura (2015), Guarienti (2015) e Oliveira (2015).

Os modelos de monitoramento ambiental estão inseridos dentro de uma nova ordem do método científico. O uso de modelos e simulações grandes e complexas por partes dos cientistas impulsionou o desenvolvimento da computação científica (CORDEIRO et al., 2013).

2.2 Computação Científica

O avanço ocorrido nas últimas décadas na tecnologia de computadores cresceu em escala exponencial. Grandes inovações como microprocessadores de grande capacidade e a interconectividade das redes locais, *Local Area Network* (LAN) mudaram as perspectivas da computação. Por exemplo, nos últimos 50 anos o ganho de preço em função do desempenho foi superior a 10 elevado a décima potência, algo sem precedente em outros setores (TANENBAUM; STEEN, 2007).

Com o intenso desenvolvimento dos computadores, é cada vez mais difícil torná-los mais rápidos e transpor barreiras físicas como o aquecimento. Então, uma alternativa é usar computação distribuída, que pode ser dividida em dois grupos (TANENBAUM; STEEN, 2007). No grupo um existe a computação de *cluster*, com máquinas semelhantes, que usam o mesmo sistema operacional e estão na mesma LAN, como pode ser observado na Figura 3. No grupo dois

está a computação em *grid* (grade), que funciona como uma federação¹, possui ambiente heterogêneo de computadores, sistemas operacionais e redes, conforme pode ser visto na Figura 4.

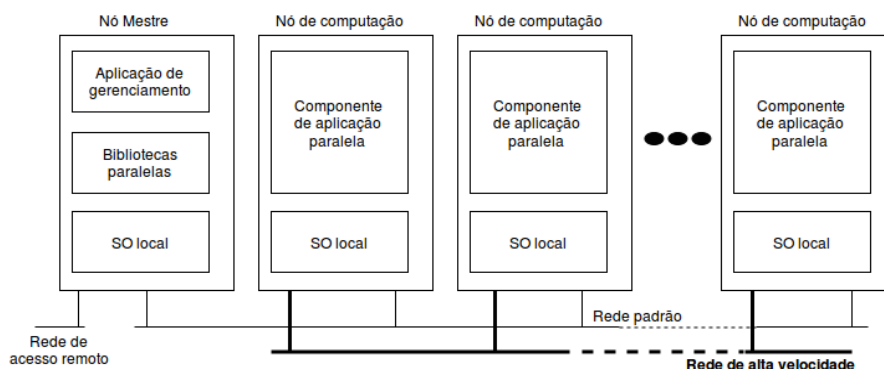


Figura 3: Exemplo de sistema de computação em *cluster*. Fonte: Adaptado de Tanenbaum e Steen (2007).

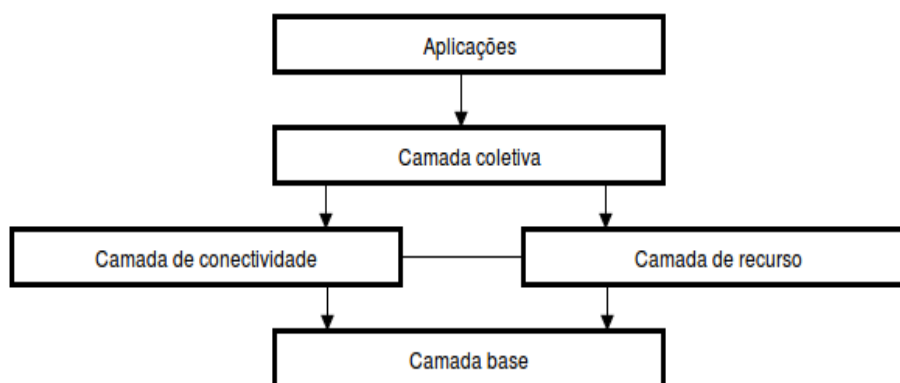


Figura 4: Exemplo de arquitetura em camada de computação em grade. Fonte: Adaptado de Tanenbaum e Steen (2007).

Para Beserra et al. (2015) apesar do barateamento dos *clusters* a sua utilização de forma isolada não é capaz de atender as necessidades científicas, por exemplo, na questão de integração para cooperação em grandes projetos.

Segundo Cordeiro et al. (2013) a metodologia de trabalho científico, que foi iniciada de forma empírica descrevendo fenômenos naturais, já sofreu várias mudanças, surgiram novas ferramentas, novos instrumentos e paradigmas, as mudanças que tiveram maior impacto estavam ligadas a novas ferramentas conceituais ou tecnológicas que criaram novos tipos de ciência.

O uso de ferramentas computacionais contribui para alterar a forma de se fazer ciência em várias áreas, por exemplo, na meteorologia. O uso de computadores possibilitou novos tipos de pesquisa produzindo avanços e criando novos

¹Vários conjuntos de computadores.

desafios para a própria ciência da computação. Essa nova realidade científica vem alterando o papel na ciência da computação, afastando-se o papel de ferramenta apoiadora para alcançar o patamar de parte fundamental do método científico das outras áreas (CORDEIRO et al., 2013).

A grande quantidade de dados gerados por cientistas de várias áreas requer o uso de novas técnicas computacionais para tratar essas informações. O avanço na computação paralela/distribuída e o desenvolvimento de computadores cada vez mais rápidos permitiu um crescente uso de simulações numéricas por parte de várias ciências. Isso tornou possível simular modelos mais complexos que analisam e produzem grande quantidade de dados (CORDEIRO et al., 2013). Além dos avanços tecnológicos é necessário ampliar os horizontes com o trabalho colaborativo, é possível citar exemplos positivos como o Projeto Genoma Humano, que contou com a participação de cientistas do mundo todo.

A ciência da computação, em muitas situações, deixou de ser um ponto de apoio e passou a ser base para a criação de conhecimento. Vários experimentos puderam ser feitos utilizando modelos usando simulações virtuais. Cientistas têm usado com frequência *workflows*, isto é, conjunto de programas organizados em fluxo que possuem objetivo de executar alguma atividade (SILVA, 2011). Dessa forma, com toda a participação da computação como um dos pilares do método científico, a ciência cada vez mais passa a ser, na verdade, e-ciência (*e-science*) (CORDEIRO et al., 2013).

2.3 E-science

E-science, também conhecido como *e-research*, é o nome do campo multidisciplinar de pesquisa que propõe novas infraestruturas para armazenamento, acesso e análise de grandes conjuntos de dados e se apresenta como novo paradigma na atividade científica a partir dos investimentos em computação de larga escala e repositório de dados. Nesse novo momento as práticas de atividade científica diferem das formas tradicionais, por exemplo, com o uso intensivo de poderosas infraestruturas computacionais, compartilhamento e projetos conjuntos em nível global e de produção, armazenamento e compartilhamento de repositórios em larga escala.

Com o *e-science* muda-se a forma como os cientistas fazem seu trabalho, as ferramentas que utilizam, os tipos de problemas que são tratados e, até mesmo, os resultados que são publicados. Agora, o pesquisador pode utilizar uma infraestrutura capaz de subsidiar seu estudo por meio do compartilhamento e reuso

de dados já coletados (MEDEIROS; CAREGNATO, 2012).

Por meio do *e-science* os cientistas se depararam com novas barreiras, por exemplo, ter que lidar com dados que não conheciam, visto que é uma nova realidade que apresenta características como: crescimento do volume, complexidade e heterogeneidade de dados (CHEN et al., 2010). O uso de do *e-science* é uma tendência para os pesquisadores, pois podem trabalhar nas suas aplicações em ambientes que possuem escalabilidade e alta disponibilidade, além de não precisarem se preocupar com a manutenção e o tratamento de falhas (SAAD; CERIN, 2015).

Existe um esforço para tornar acessíveis pela Internet dados científicos, em especial nas ciências físicas (NAROCK; FOX, 2012). Nessa linha, é possível citar, por exemplo, o projeto EU VERCE (*VirtualEarthquake and seismology Research Community e-science environment in Europe*), que desenvolveu um VRE (*Virtual Research Environment*) para computação intensiva no tratamento dos dados de sismologia, utilizando modelos e exemplos (ATKINSON et al., 2015). Armazenar e compartilhar dados científicos é um serviço essencial que pode ajudar pesquisadores em várias perspectivas na descoberta de dados de pesquisa (ARDESTANI et al., 2015). A ciência continua aumentando sua demanda de processamento, no entanto, o desafio passa a ser a capacidade de interpretar e produzir informação a partir dessa intensa carga de dados (MARTINEZ-ORTIZ et al., 2015).

É um cenário novo em que as práticas são bem diferenciadas em toda a cadeia de produção do conhecimento científico, conforme Appelbe e Bannon (2007) na Tabela 1 as diferenças entre pesquisa tradicional *e-science*:

Tabela 1: Características de pesquisa tradicional e de *e-science*

Características	Pesquisa tradicional	<i>E-science</i>
Participantes	Grupo de pesquisa local	Em diferentes localidades
Dados	Somente acesso local	Distribuídos
Computação	Computador pessoal	Em larga escala
Rede	Não é distribuída pela Internet	Infraestrutura e Internet
Divulgação	Impressa, digital e conferências	Portais especializados

Esse novo método de se fazer ciência possui vários aspectos que diferem do modelo de pesquisa tradicional (APPELBE; BANNON, 2007). É possível destacar os seguintes pontos:

- As equipes de pesquisas são distribuídas com diferentes habilidades.

- Os dados gerados são armazenados e acessíveis de forma distribuída.
- Usa computação em larga escala ou sob demanda ou possibilita o acesso a instrumentos computacionais compartilhados.
- Uso de Internet.
- Resultados são disseminados em portais especializados.

Esse novo modelo de ciência enfrenta novas questões como alocação de recursos computacionais em um cenário de uso irregular. Para tanto, o objetivo é prover serviços de acordo com a demanda, dessa forma, otimizando gastos e tempo de alocação (VUKOJEVIC-HAUPT; GOMEZ-SAEZ, 2015).

Relacionado ao *e-science* está o conceito de *virtual workspaces*, provisionamento automático para atender necessidades de usuários, observando questões correlatas como alocação dinâmica e elasticidade² (KEAHEY et al., 2005). É uma temática que já possui iniciativas no mundo acadêmico como criação de um ambiente virtual de trabalho com foco em unidades de pesquisa (FERNANDES et al., 2011).

A falta de condições mínimas para criar e manter um ambiente computacional prejudica a prática de *e-science*, mas esse cenário pode ser alterado com o uso de uma outra tecnologia: Computação em Nuvem (CORDEIRO et al., 2013).

2.4 Computação em Nuvem

A partir da segunda metade do século XX o avanço dos processadores e do poder computacional fez com que os computadores passassem a não usar toda sua capacidade, então, nesse contexto parte dos recursos estavam sendo desperdiçados. A virtualização aparece como uma possível solução para essa questão, pois, além de proporcionar um maior uso dos recursos, com a criação de *Virtual Machine* (VM), ou seja, máquinas virtuais, permite a expansão do ambiente sem exigir aquisição de novas máquinas físicas. Máquina virtual é uma camada de *software* que pode simular uma máquina física, possuindo seu próprio Sistema Operacional (SO) e bibliotecas. O gerenciamento e alocação de recursos são feitos pelo *hypervisor*, que é uma camada de *software* entre *hardware* e SO. Essa tecnologia oferece portabilidade e flexibilidade, visto que diferentes SOs podem executar sobre o mesmo equipamento (CARISSIMI, 2008).

²Capacidade de adicionar ou remover recursos computacionais em tempo de execução, sem interrupções.

Com o tempo outras opções de virtualização foram aparecendo, uma das mais novas é o Docker³, criado em 2013, plataforma de código aberto para desenvolvedores e administradores de sistema, que disponibiliza armazenamento e compartilhamento de imagens no repositório oficial (YU; HUANG, 2015). Essa tecnologia não necessita da camada de virtualização sobre o *hardware*. Isto porque tira proveito de características do *kernel* (núcleo) Linux, por exemplo, capacidade de isolamento dos ambientes no mesmo equipamento físico. Mas também possui desvantagens como a falta de um gerenciamento gráfico e dificuldades de assegurar bom nível de segurança (BACIS et al., 2015). A partir das tecnologias de virtualização, redes, *grids* e *clusters* surge uma nova tecnologia: a Computação em Nuvem (CHANDRASEKARAN, 2014).

A CN oculta toda a complexidade de uma infraestrutura e coloca vários serviços na Internet à disposição de um gerenciamento unificado, que normalmente envolve um programa navegador, por exemplo: Firefox ou Chrome. Por meio de uma interface estão disponíveis: armazenamento, processamento e até *software*, que podem ser entregues em tempo de execução como serviço e sob demanda. Trata-se de uma tecnologia que tem ganhado sucessiva atenção do meio acadêmico (BUYA et al., 2013).

A ideia central da CN é oferecer serviços de computacionais por meio da *web*. Com essa tecnologia é possível aumentar ou diminuir máquinas virtuais e armazenamento instantaneamente, com mínimo esforço gerencial segundo Mell e Grance (2011), ocultando toda a complexidade da execução em camadas inferiores mais próximas ao *hardware*.

A migração de serviços do ambiente local para Internet representa uma mudança significativa e um avanço sobre tecnologias como a virtualização, que, por padrão, só pode ser gerenciada localmente. Essa oferta de serviços na rede mundial de computadores pode transformar equipamento pessoais, como celular e microcomputador, em meios de acesso e não mais em ambiente próprio de trabalho.

A CN consegue oferecer componentes básicos de uma infraestrutura computacional como armazenamento, processamento e largura de banda sob demanda, isto é, um usuário desses serviços transfere para o provedor toda a responsabilidade de manutenção, atualização e *backup*, *software*. Dessa forma, existe uma espécie de terceirização dos serviços de Tecnologia da Informação, que de certa forma passam a ser um serviço de utilidade pública (CARR, 2009).

Com essa nova visão de alugar uma infraestrutura alheia por determinado

³<https://www.docker.com>

ponto os investimentos em computação podem ser reduzidos significativamente. Já não é necessário renovar parques computacionais e possuir uma equipe pra manter uma estrutura física. Uma empresa, por exemplo, pode contratar os recursos que tem necessidade e pagar exatamente pelo que consome. No cenário convencional o empresário teria que pagar por equipe e recurso mesmo nos momentos de ociosidade.

A tecnologia de Computação em Nuvem trás características como a elasticidade, um grande volume de recursos/serviços podem ser criados ou deletados instantaneamente. Isso quer dizer que o processamento de alto desempenho vai ao encontro de CN, tanto que provedores de serviço como a Amazon⁴ já oferecem soluções específicas para computação de alta performance (CORDEIRO et al., 2013). Em um mês a necessidade pode ser alto poder de processamento, no mês seguinte pode ser disco, um provedor de CN pode se adaptar a essas diferentes necessidades. Assim, por exemplo, evita-se que um cliente precise comprar vários discos, pois precisa desse recurso por um tempo específico. É mais vantajoso contratar esse recurso. Se no meio do processo esse cliente perceber que a quantidade contratada não vai ser suficiente basta solicitar mais recurso, algo que seria mais complicado em um cenário convencional.

Para Neto e Faxina (2015) mais que economizar dinheiro o uso de Computação em Nuvem alinha-se com práticas de TI⁵ Verde, uso de recursos que minimizem agressões ao meio ambiente, pois possibilita economia com energia e refrigeração, além de evitar novas aquisições de equipamentos físicos (ARAÚJO; CAVALCANTE, 2015). É uma tecnologia que tem características de sustentabilidade e que facilita a troca de informações.

A CN favorece o compartilhamento de dados com pessoas de qualquer lugar do mundo, visto que as informações já estão na Internet, em tese basta configurar as permissões de acesso. Essa característica permite uma experiência em ambiente de trabalho compartilhado.

A Computação em Nuvem é dividida basicamente em três tipos:

- SaaS (*Software as a service*), ou seja, *software* como serviço. São programas colocados na Internet para atender objetivos específicos, por exemplo, o *Google Docs*, editor de texto e planilha (CIURANA, 2009).
- PaaS (*Platform as a service*), isto é, plataforma como serviço. Disponibiliza um ambiente de desenvolvimento e teste na nuvem, com essas carac-

⁴<https://aws.amazon.com/pt/ec2/>

⁵Tecnologia da Informação

terísticas é possível citar a ferramenta Aneka, que possui bibliotecas que possibilitam a criação de aplicativos (CHU et al., 2007).

- IaaS (*Infrastructure as a service*), que significa, infraestrutura como serviço. Neste tipo de nuvem armazenamento, processamento são oferecidos com serviço sob demanda. É possível citar como exemplos de IaaS: Amazon, Linode e Rackspace (MARINO et al., 2015).

Na Figura 5 é possível visualizar a arquitetura da Computação em Nuvem dividida por camadas, com um exemplo em cada divisão.

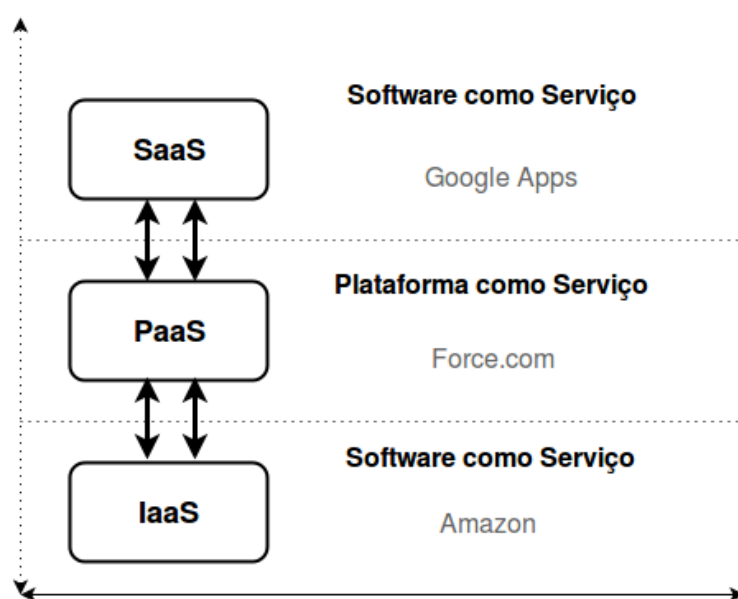


Figura 5: Camadas da Computação em Nuvem. Fonte: Adaptado de Slabela et al (2009).

A camada mais alta, chamada de SaaS, é a mais próxima do usuário (cliente do serviço), segundo Vaquero et al. (2008) é a camada onde podem ser oferecidos serviços *online*, como edição de texto. Para Chen e Jin (2012) a camada intermediária, chamada PaaS, é onde os fabricantes desenvolvem um ambiente que fica disponível aos usuários fornecendo a plataforma como serviço. Vaquero et al. (2008) afirma que a camada mais baixa, isto é, a IaaS, é usada para gerenciar recursos computacionais como armazenamento e processamento utilizando virtualização para oferecer serviço sob demanda.

Além dos tipos, também existe a classificação das nuvens computacionais em relação ao modelo de implantação, pois existem diferentes aspectos de acesso e disponibilidade (SOUSA et al., 2010). Dessa forma, são organizadas em:

- Nuvem pública: a infraestrutura de CN é disponibilizada para o público em geral.
- Nuvem privada: o ambiente computacional pode ser local ou remoto, e é usado exclusivamente por uma organização. A administração pode ser feita pela equipe interna ou pode ser terceirizada.
- Nuvem híbrida: infraestrutura formada por duas ou mais nuvens interligadas permitindo a troca de dados e aplicativos.
- Nuvem comunidade: fornece ambiente computacional em nuvem para um conjunto de organizações com interesse comum.

Com o intenso uso de CN também cresce a necessidade de assegurar níveis mínimos de segurança da informação, pois essa tecnologia enfrenta o desafio de avançar em campos como segurança e privacidade, pois muitas vezes dados sensíveis estão sendo tratados nesses ambientes (HASHIZUME et al., 2013). Cada IaaS tem suas especificidades pra tratar questões como segurança.

Para implementar a Computação em Nuvem existem várias soluções disponíveis, é possível listar três *softwares* de código aberto com esse propósito: Eucalyptus⁶, OpenNebula⁷ e OpenStack⁸. Cada opção possui características próprias. Os três possuem comunidades que possibilitam interação entre desenvolvedores e usuários no intuito de ajudar a resolver problemas técnicos. O Eucalyptus permite uso de imagens, arquivos base para criação de VMs, com SO Linux e Windows. OpenNebula e OpenStack possibilitam a migração de VMs entre máquinas físicas diferentes (YADAV, 2013).

A Computação em Nuvem consegue ir além da virtualização oferecendo gerenciamento detalhado de grupos e usuários, integração com outras tecnologias, escalabilidade, facilidade de compartilhamento das informações, *backup*, e consegue entregar recursos, por exemplo: memória, disco, processamento, *link* como serviço. E o principal: proporciona o acesso à toda uma infraestrutura em forma de serviço pela Internet.

⁶<http://www8.hp.com/br/pt/cloud/helion-eucalyptus-overview.html>

⁷opennebula.org

⁸www.openstack.org

Capítulo 3

Materiais e Métodos

3.1 Tecnologia

Este estudo utilizou Computação em Nuvem, focado na camada IaaS, mais próxima ao *hardware*, pois trata-se do gerenciamento de recursos computacionais como armazenamento e processamento, com foco na disponibilização de ambiente *e-science* voltado para atender demandas da área ambiental. A transformação da nuvem computacional em ambiente *e-science* envolveu a disponibilização de *softwares* e serviços relacionados a área ambiental.

As informações necessárias para implementar uma infraestrutura de *e-science* ambiental, por meio de nuvem científica, que atendesse efetivamente a necessidade dos pesquisadores foi alcançada por meio de pesquisa bibliográfica, baseada em produção científica e na leitura dos materiais disponibilizadas pelas soluções existentes.

Três opções de Computação em Nuvem foram escolhidas para teste. As seguintes características foram verificadas: exigência de *hardware*, integração com outras tecnologias, controle detalhado de usuários¹, grupos e controles de acesso, capacidade de migrar VMs entre *hosts*, gerenciamento de energia e capacidade de trabalhar em forma de *cluster*.

Após a revisão bibliográfica foi feita a implementação da infraestrutura *e-science* utilizando Computação em Nuvem usando recursos (processador, memória, rede, Internet) do parque computacional da própria UFMT. O conjunto de produtos utilizados envolvem a integração e adaptação de ferramentas baseadas na filosofia de *software* livre. Com isso, o domínio da tecnologia é favorecido pelo acesso ao código e a facilidade na difusão das soluções geradas é maior. A revisão do estado da arte contribuiu para a escolha do OpenNebula.

¹Por exemplo: cotas de memória, disco e processador.

Esse IaaS possui um gerenciamento detalhado de permissões a respeito da segurança, mas que basicamente são definidos assim:

- Usuário - permissão de acesso somente a sua VM ou ambiente de compartilhamento definido.
- Administrador - controle total, pode criar ou deletar: redes, *clusters*, VMs, *templates*², imagens, usuários e grupos.

Usuário é o cliente do serviço, então pode acessar estritamente a sua VM, salvo em casos expressos de algum compartilhamento, como conteúdo ou máquina virtual.

Administrador é o responsável pela infraestrutura, então, pode criar ou excluir serviços e delimitar os níveis de segurança, por exemplo, no acesso de diretórios e VMs. Dessa forma, pode permitir que diferentes pesquisadores possam acessar o mesmo ambiente para realizar trabalho colaborativo. Para assegurar maior nível de segurança no ambiente *e-science* ele foi implementado em um contexto de outros serviços.

Uma das questões que vem sendo debatidas juntamente com a CN é sua capacidade de tratar problemas na área de segurança (VOGEL et al., 2015). Nesse quesito é possível destacar que o IaaS implementado foi configurado em uma LAN que possui: *firewall*, *proxy reverso* e *backup*.

Firewall é o *hardware* ou *software* que define se determinado conteúdo é permitido ou não no tráfego de dados com destino ou origem da rede local. *Proxy reverso* é um recurso de segurança que impede um usuário de acessar diretamente um servidor de páginas da Internet, criando uma camada adicional nessa conexão. *Backup* é a criação de cópias sistemáticas dos dados, impede, por exemplo, que em um cenário de desastre dos todos dados sejam perdidos. Além desse serviços citados, a CN oferece possibilidade de reuso, seja de *software* ou *hardware*.

3.2 Testes

Existem várias pesquisas em andamento, cada uma com suas próprias necessidades e duração. A Computação em Nuvem permite que ao fim de um estudo, esses recursos usados, seja *software* ou *hardware*, possam ser reusados em novas pesquisas, reduzindo assim a necessidade de comprar e configurar mais equipamentos. Desse modo, a CN favorece o reuso da infraestrutura.

²Modelo de máquinas virtual que possui algumas configurações básicas, por exemplo, tamanho do disco.

Para testar o ambiente Cirrus foi convidado um grupo de cinco alunos do PPGFA. Dessa forma, foi explicado aos participantes instruções básicas: como acessar a infraestrutura na Internet, como instalar ou desinstalar um programa, como acessar sites e como fazer *download* de informações.

Em resumo a metodologia do ambiente *e-science* compreende o gerenciamento pela *web* de recursos computacionais, *software* e *hardware*, para tratamento de dados ambientais com foco em atingir o nível de serviços demandados por um contexto de *e-science* ambiental. A meta foi criar e disponibilizar ambiente *e-science* para o desenvolvimento das pesquisas ambientais propiciando uma infraestrutura adequada às necessidades de diferentes pesquisas do PPGFA.

Capítulo 4

Apresentação e Análise dos Resultados

O capítulo de resultados foi dividido em cinco partes, a primeira parte é chamada de infraestrutura básica, pois explica-se como foi feito o arranjo dos servidores e máquina virtual que compõe o ambiente *e-science*, além de citar *softwares* e SO usados.

4.1 Infraestrutura Básica

Este trabalho tratou da implantação de uma camada da Computação em Nuvem, chamada IaaS (*Infrastructure as a Service*), ou seja, infraestrutura como serviço. Foi utilizado a ferramenta de código aberto OpenNebula para o gerenciamento da nuvem computacional, que possui características de nuvem pública, disponível na Internet, e privada, pois utiliza recursos locais. Para transformar o nuvem computacional em ambiente *e-science* foram disponibilizados vários *softwares* relacionados a área ambiental e serviços que auxiliam o método científico. Na implantação da infraestrutura foram consideradas três opções de IaaS: CloudStack, OpenStack e OpenNebula.

Na tabela 2 estão listadas as plataformas e seus requisitos computacionais mínimos:

Tabela 2: Requisitos mínimos de cada infraestrutura

Plataforma	Número mínimo de computadores
CloudStack	7
OpenNebula	2
OpenStack	5

Para implementar o ambiente *e-science* para tratamento de dados ambientais no presente trabalho, foram escolhidas ferramentas *open source*. A infraestrutura disponível possibilitou o teste apenas do OpenNebula, pois as demais opções exigiam um número de máquinas que não era possível implementar. Além do OpenNebula outras soluções *open source* foram estudadas:

- OpenStack - esta solução foi preterida por sua arquitetura mínima exigir maior necessidade de *hardware*. Além disso, possui uma estruturação fragmentada e nomenclatura complexa de seus serviços. No teste perdeu-se muito tempo tentando entender a nomenclatura usada para nomear os serviços.
- CloudStack - possui a maior exigência de equipamentos físicos das soluções estudadas. E por não ter disponível no parque computacional os elementos mínimos (indicados pelos desenvolvedores) não apresentou resultados satisfatórios.

Não foi possível testar o OpenStack e o Cloudstack. Os recursos disponíveis para este estudo não atendiam os requisitos mínimos informados na documentação oficial desses IaaSs. Dessa forma, somente foi possível testar o OpenNebula.

O ambiente *e-science* Cirrus foi composto por uma máquina virtual que faz a função de gerenciamento, chamada de *frontend*, responsável pelos processos próprios do OpenNebula segundo Beserra et al. (2015), e dois *hosts*¹ configurados para trabalhar em forma de *cluster*² (DEITEL et al., 2005).

Os equipamentos que compõe a infraestrutura foram configurados dentro da mesma *Local Area Network* (LAN), ou seja, rede local, que possui velocidade de um (1) Giga bit por segundo (Gbps). Esse ambiente está protegido por *nobreak*, equipamento capaz de suprir a falta de energia elétrica por determinado tempo, *firewall* e *proxy* reverso. Na Figura 6 é possível visualizar a organização:

O Ubuntu foi escolhido por ser um Sistema Operacional (SO) recomendado pelos próprios desenvolvedores do OpenNebula conforme informações do site oficial. O Windows 7 e Windows 8.1 foram testados por sua popularidade.

O ambiente *e-science* foi configurado de modo que os servidores trabalhassem em forma de *cluster*, isto se deve ao fato de que se um nó não consiga ser ligado normalmente outro nó que esteja disponível pode assumir as VMs que

¹Também chamados de servidores, computadores com maior poder computacional com função de prover recursos

²Associação de vários computadores interligados por uma rede de alta velocidade com objetivo de executar tarefas como um único sistema

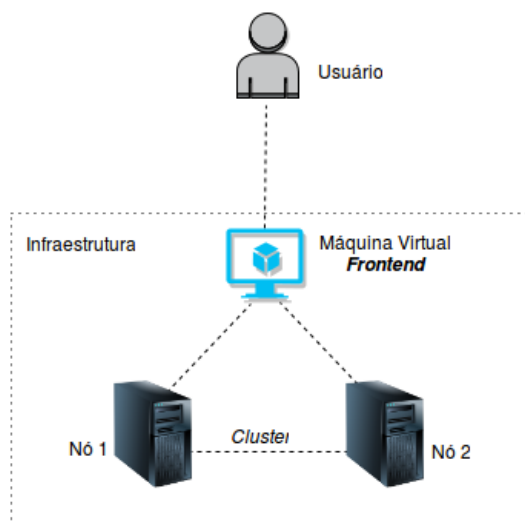


Figura 6: Arranjo da infraestrutura *e-science*.

estavam designadas para o *host* que ficou indisponível. Esse tipo de problema pode ocorrer, por exemplo, em função de problemas com instabilidade elétrica.

A documentação oficial do IaaS escolhido exige que o nome dos *hosts* e seus respectivos IPs (*Internet Protocols*) estejam devidamente configurados no arquivo `"/etc/hosts"`, isso para forçar que os servidores conheçam IP e nome dos demais componentes da solução.

Na parte de segurança as definições foram feitas usando grupos, usuários, permissões específicas baseadas em lista com controle de acesso. Dessa forma foi possível assegurar que o acesso a determina VM fosse feito apenas pelo usuário ou grupo permitido. Para atingir um nível maior de segurança um usuário precisa se autenticar no ambiente e também na máquina virtual, isto é, precisa passar por duas validações de acesso.

Durante a instalação do OpenNebula é gerada uma senha aleatória para o usuário administrativo `oneadmin`, pois é com esse perfil que ações administrativas devem ser executadas. Outro item relacionado a segurança pode ser configurado na criação de uma nova VM: inserir a chave SSH pública do *frontend* para que possa se conectar sem senha com as VMs, que por padrão não permitem acesso via *secure shell*. A comunicação entre *frontend* e os demais nós acontece o tempo todo pois trocam informações sobre o estado de seus serviços, a segurança é baseada em chave SSH compartilhada.

Para acompanhar os *logs*, informações sobre serviços ativos, que são gerados referentes ao OpenNebula deve-se observar as informações geradas na pasta: `"/var/log/one"`. Através dessas mensagens é possível verificar se uma máquina virtual ficou indisponível, ou até se houve alguma falha de comunicação entre os

nós.

Depois do SO instalado em todos os servidores integrantes da infraestrutura *e-science* foram feitas todas as atualizações de segurança. A partir disso os próximos passos foram dados no sentido de criar as VMs seguindo material oficial disponibilizado no site da ferramenta. Depois de criar as máquinas virtuais foi possível testar o recurso de auditoria.

A auditoria serve pra gerar relatórios de consumo, por exemplo, de disco e CPU em uma faixa de tempo determinada. Na Figura 7 observa-se o consumo de CPU pelas VMs, listando a quantidade total consumida e a utilizada em cada dia.

CPU hours

Date UTC ▲	Total	MV 176 PGFA1 - Rafael	MV 196 Windows7-1	MV 197 Windows7	MV 198 Windows	MV 199 Win	MV 200 Win	MV 201 Windows.
2016/1/1	96	96	0	0	0	0	0	0
2016/1/2	96.89	96	0.34	0.14	0.2	0.21	0	0
2016/1/3	96	96	0	0	0	0	0	0
2016/1/4	103.96	96	0	0	0	0	0	5.23
2016/1/5	119.24	78.22	0	0	0	0	0	21.49
2016/1/6	27.41	18.99	0	0	0	0	0	4.74
2016/1/7	143.6	96	0	0	0	0	0	0.29
2016/1/8	141.5	95.64	0	0	0	0	0	0
2016/1/9	144	96	0	0	0	0	0	0
2016/1/10	144	96	0	0	0	0	0	0

Showing 1 to 10 of 33 entries

Previous 1 2 3 4 Next 10

Figura 7: Tabela de consumo de CPU por VM.

Os gráficos podem ser gerados em função dos recursos selecionando usuário, grupo ou VMs. Na Figura 8 destaca-se o consumo de CPU por um determinado usuário, no intervalo do dia 01/01/2016 a 02/02/2016.

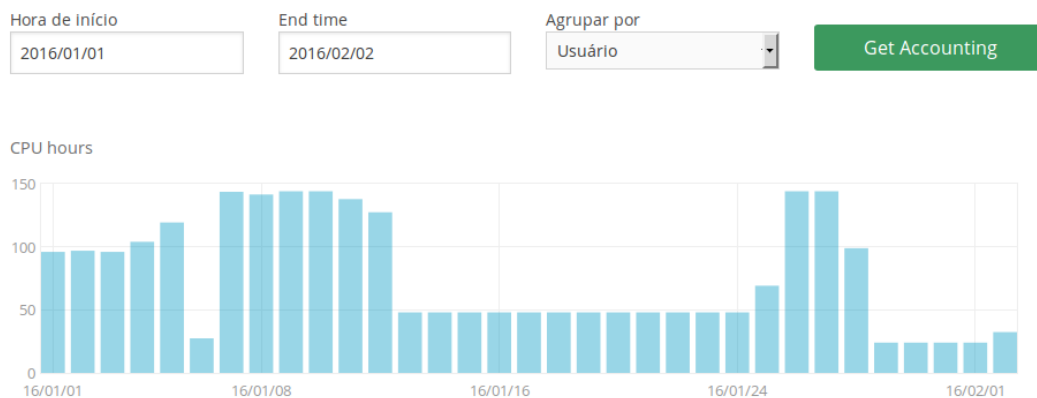


Figura 8: Gráfico de consumo de CPU de um determinado usuário.

A segunda parte do capítulo de resultados trata sobre as máquinas virtuais criadas dentro do ambiente *e-science*, isto é, instâncias, exemplificando seus tipos e possibilidades, também foi tratado sobre alguns *softwares* ligados a área ambiental.

4.2 Instâncias

Após a configuração para o funcionamento da infraestrutura foi necessário criar instâncias, ou seja, máquinas virtuais que ficam dentro do ambiente. Na interface gráfica do OpenNebula existe um espaço, chamado loja, que oferta VMs, também chamadas de imagens, prontas como: Debian, CentOS, e Ubuntu, conforme Figura 9.

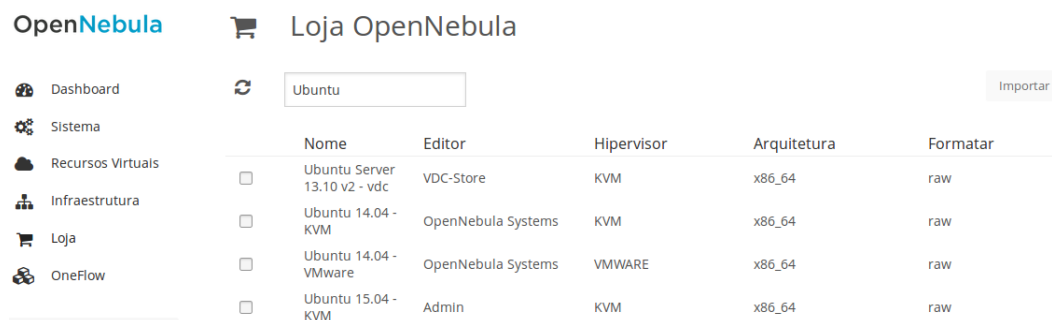


Figura 9: Espaço para download de imagens.

É possível também escolher entre diferentes hipervisores³, por exemplo, KVM (*Kernel-based Virtual Machine*) e VMWARE, arquiteturas de 32 ou 64 bits e, por fim, pode-se optar por formatos: raw, vmdk e qcow2. O passo a seguir foi definir qual o tipo de virtualizado que seria usado nas instâncias.

As imagens ofertadas na loja possuíam tamanho de apenas 10 Giga *Bytes* (GB), espaço insuficiente, visto que é comum apenas a instalação de uma interface gráfica consumir cerca de 2 GB. Então, foi necessário criar um *template* a partir da imagem de tamanho padrão adicionando mais discos e usando LVM (*Logical Volume Management*) para, por fim, criar um bloco único de 30 GB, assim ofertando um espaço maior pra poder manipular dados.

Na sequência foi configurada a comunicação das VMs, que foi feita utilizando uma rede virtual criada na máquina *frontend*, chamada "br0", que faz o papel de criar um *switch* virtual, ou seja, chavear a rede para todas as instâncias.

³Plataforma ou camada de controle da virtualização que permite o uso de diferentes sistemas operacionais concomitantemente no mesmo hardware

As VMs foram configuradas para usar virtualização do tipo KVM, pois de acordo com Che et al. (2010) é uma tecnologia com reconhecida evolução e tem ótimos resultados em comparação com outras tecnologias (YOUNGE et al., 2011). Algumas características devem ser escolhidas no momento da criação de uma nova instância como: nome, quantidade de instâncias, RAM, CPU (*Central Processing Unit*) e *template* a ser usado.

Para que OpenNebula funcione integralmente em todos servidores os serviços RPC (*Remote Procedure Call*) e *bind* devem estar disponíveis já na inicialização do SO, e, para isso, foram colocados *scripts* que forcem esse início. Já outros serviços como: NFS (*Network File System*) e opennebula-sunstone foram colocados na lista de serviços que o Ubuntu deve inicializar por padrão. Sem o NFS carregado os nós não montam a partição que deve armazenar as VMs e por consequência nenhuma instancia consegue ficar *online*.

Depois que o SO estava pronto para ser usado alguns *softwares* relacionados à área ambiental foram instalados:

- R - é uma linguagem para computação estatística e gráficos.
- Rstudio - ambiente de desenvolvimento integrado ao R.
- Phyton - linguagem de computação em alto nível, interpretada, interativa e orientada a objetos.
- QuantumGIS - sistema de informação geográfica com código aberto e licença pública geral GNU. É uma opção ao ArcGIS.
- Matlab - *software* para realização de cálculos numéricos.
- Octave - linguagem de alto nível interpretada, usado principalmente para cálculos numéricos. É uma alternativa ao Matlab.
- ENVI-met - *software* de desenho urbano.

4.3 Serviços

No ambiente *e-science* foram disponibilizados serviços que podem estar relacionados a segurança da informação, como o *backup*, que contribui para disponibilidade dos dados. O próprio ambiente criado é um serviço, uma infraestrutura como serviço, que provê recursos computacionais para pesquisa.

Todas as imagens vem apenas com o *prompt*⁴. Para que fosse mais acessível a utilização das instâncias pelos pesquisadores foram testados diferentes ambientes gráficos, e foi avaliado qual consumia menos memória RAM *Random Access Memory*. O teste foi feito usando uma VM com 3 GB de RAM, 30 GB de disco e 1 vCPU (*virtual Central Processing Unit*). As opções gráficas foram instaladas segundo a documentação dos sites oficiais, as seguintes opções foram testadas: gnome, mate, LXDE (*Lightweight X11 Desktop Environment*) e KDE (*K Desktop Environment*) conforme Tabela 3:

Tabela 3: Avaliação do consumo de memoria RAM utilizando quatro ambientes gráficos

Ambiente gráfico	Consumo de memória RAM em Mega Bytes
Gnome	758
KDE	775
LXDE	398
Mate	602

O LXDE⁵ é um ambiente gráfico de código aberto, licenciado sob os termos GPL (*General Public License*), mantido por uma comunidade internacional de desenvolvedores, destinado a plataformas Unix, desenhado para computadores em nuvem.

A seguir foi criada uma rotina automática de *backup* (cópia) dos dados existentes em cada instância. Isso porque caso acontecesse alguma eventualidade o pesquisador estaria resguardado com pelo menos duas cópias das informações, uma dentro da infraestrutura *e-science*, outra fora. Foram adotados os seguintes modelos de *backup*:

- Completo - todos os dados são copiados.
- Diferencial - faz cópia todos os dados modificados desde última rotina completa.
- Incremental - copia arquivos novos ou modificados desde o *backup* anterior.

A Computação em Nuvem trouxe uma nova forma de se fazer cópia dos dados, mudando o cenário dos processos envolvidos, com suas características de baixo custo e escalabilidade (AMADO; MARCONDES, 2015). Na Tabela 4 pode-se visualizar o plano de *backup* aplicado na infraestrutura *e-science*:

⁴Terminal para inserir comandos de texto, sem nenhuma interface gráfica.

⁵www.lxde.org

Tabela 4: Plano de cópia dos dados

Tipo	Periodicidade	Local
Completo	Semanal	/home/usuário
Diferencial	—	—
Incremental	Diário	/home/usuário

O *backup* diferencial foi preterido em relação ao incremental por gerar uma quantidade maior de informações e isso, com o tempo, pode conflitar com as possibilidades de *hardware* disponível comprometendo o tempo de retenção, período que os dados copiados vão ficar guardados. A pasta "/home/usuario" é local padrão que qualquer usuário guarda suas informações no SO escolhido e por esse motivo é a pasta escolhida pra ser copiada.

Na Figura 10 é possível verificar como ficou a disposição das duas opções *backup*, uma dentro da próprio ambiente e outra na mesma LAN:

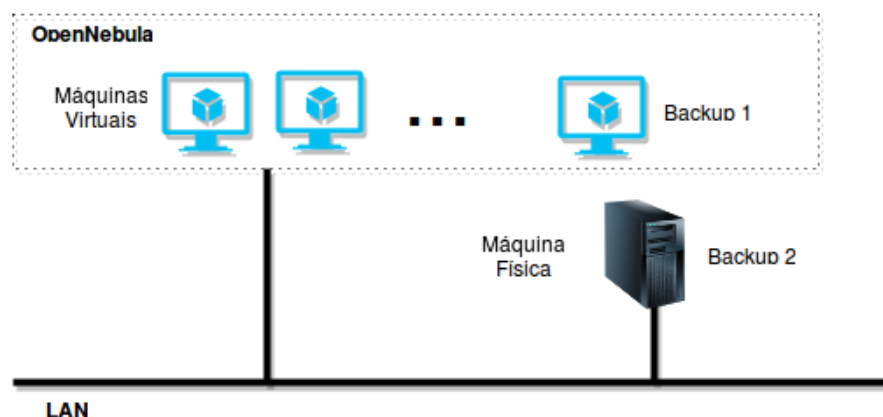


Figura 10: Desenho da estrutura de cópia dos dados na infraestrutura

Depois que os três servidores estavam com suas configurações prontas e com várias instâncias operacionais o passo seguinte foi colocar essa infraestrutura na Internet, assim, proporcionando acesso de qualquer lugar por meio de vários dispositivos, como computadores, *tablets* e celulares, o que proporcionou um ambiente altamente portátil e flexível.

O acesso pela rede mundial de computadores tornou possível atender as necessidades de mudança mais rapidamente, por exemplo, o aumento de disco ou memória de uma determinada máquina, ou até mesmo a criação de várias máquinas utilizando uma única rotina de criação. Isso quer dizer que o objetivo de elasticidade, dentro das capacidades de *hardware*, foi atingido.

Para que fosse possível visualizar as instâncias pela rede mundial de computadores foi necessário optar por dois caminhos: compra de certificado SSL

(*Secure Socket Layer*), protocolo desenvolvido para elevar a segurança de dados transmitidos na Internet, ou uso de IP público, isto é, endereço único na *web*.

Optou-se pelo segundo caso. Assim, o ambiente ficou disponível por nome e IP. Para que o acesso por nome pudesse ser feito o OpenNebula teve que ficar protegido pelo *proxy* reverso conforme Figura 11:

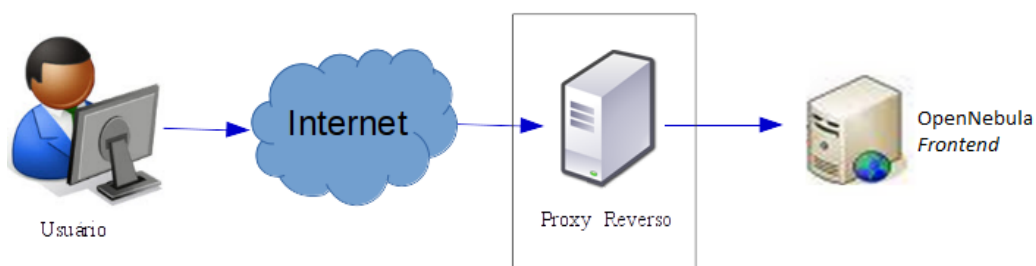


Figura 11: OpenNebula protegido por camada adicional de segurança (*proxy* reverso).

Proxy reverso cria uma camada adicional de segurança criando regras de tráfego mais rigorosas entre usuário e um servidor *web*, por exemplo, permitindo apenas conexões estritamente liberadas.

Outra iniciativa de segurança foi colocar a infraestrutura protegida por *firewall* conforme pode ser observado na Figura 12:

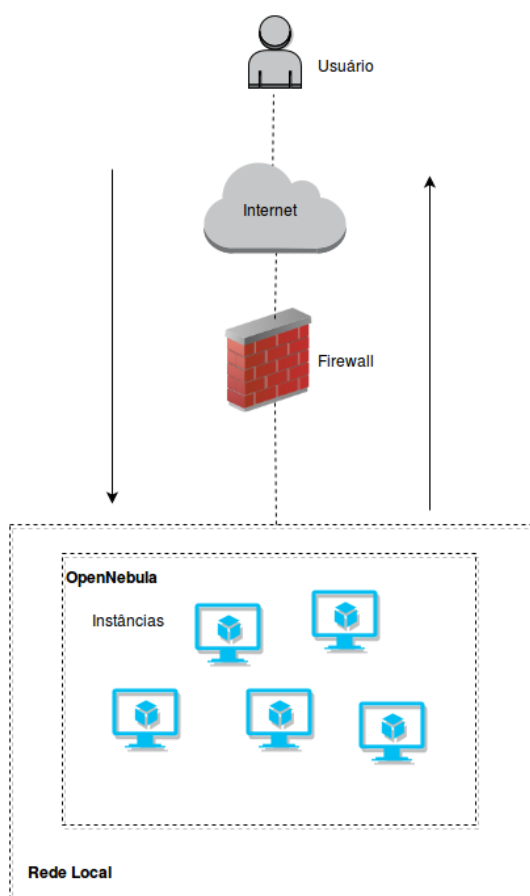


Figura 12: Instâncias protegidas por *firewall*

4.4 Gerenciamento

Gerenciar a infraestrutura computacional envolve questões como controlar criação, edição e deleção de VMs e compartilhamentos. Além disso, possibilitar a elasticidade com o aumento ou diminuição dos recursos de acordo com a demanda.

O gerenciamento do OpenNebula é feito por padrão utilizando um perfil administrador, por exemplo o `oneadmin`, também é possível criar usuário com permissões especiais. O acesso utilizando perfil `oneadmin`, com maior poder administrativo, permite logo após autenticação de usuário e senha visualizar um resumo das informações mais importantes sobre o ambiente, por exemplo, gráficos do consumo de memória e CPU, como pode ser observado na Figura 13:

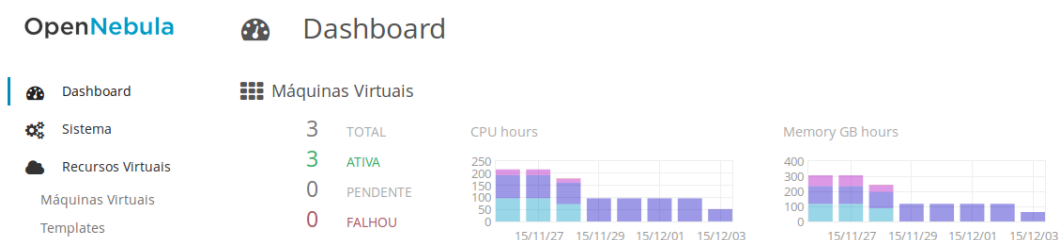


Figura 13: Resumo utilizando gráficos com informações sobre memória e CPU.

As definições de controle do acesso da ferramenta seguem a mesma lógica do SO Linux, permissões são do tipo: leitura, escrita e execução, podem ser atribuídas a: usuário, grupo que ele pertence e todos os outros demais usuários. Na Tabela 5 pode-se observar um exemplo de uma configuração do acesso para usuários e grupos em determinada VM:

Tabela 5: Exemplo de permissão para acesso.

Perfil	Escrita	Leitura	Execução
Dono	X	X	X
Grupo do dono	X	X	—
Outros	X	—	—

O usuário pode acessar sua VM, e, caso configurado pode ter permissões especiais em outras VMs e compartilhamentos. Na figura 14 é possível visualizar área de configuração dos grupos:

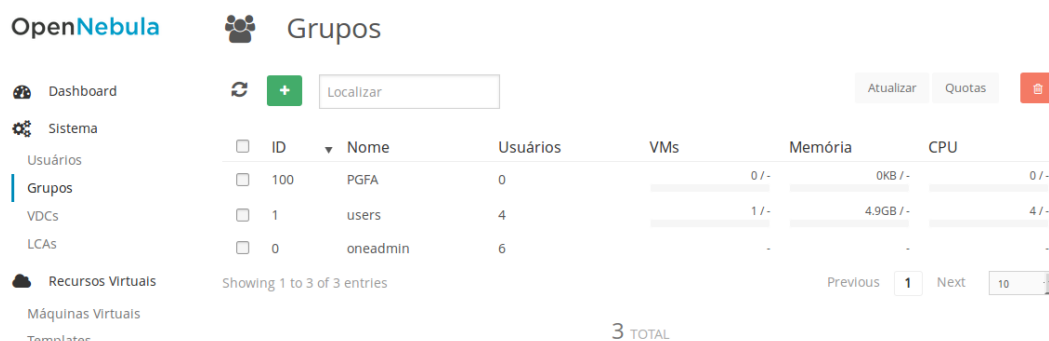


Figura 14: Definição de permissão dos grupos.

A criação de máquinas virtuais dentro da plataforma é feita a partir de *templates*, que são configurados utilizando as imagens copiadas da loja do OpenNebula. Na Figura 15 pode-se observar a tela de criação, exclusão e configuração de VMs, que apresenta entre outras informações o *status*, ou seja, estado, da instância.



Figura 15: Ambiente de edição das instâncias.

Com objetivo de aumentar a disponibilidade dos recursos ofertados aos pesquisadores optou-se por usar o OpenNebula em forma de *cluster*, na Figura 16 vê-se o local de configuração das propriedades desse agrupamento, onde é possível selecionar opções como os *hosts* que devem participar do arranjo.

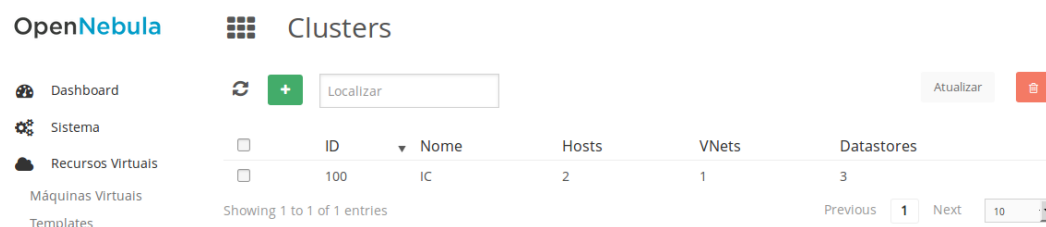


Figura 16: Configuração do cluster.

Na Figura 17 apresenta-se uma máquina virtual que acaba de ser iniciada, na tela de autenticação do sistema operacional, já está com atualizações de segurança, ambiente gráfico e com acesso disponível pela Internet por meio de *software* navegador:

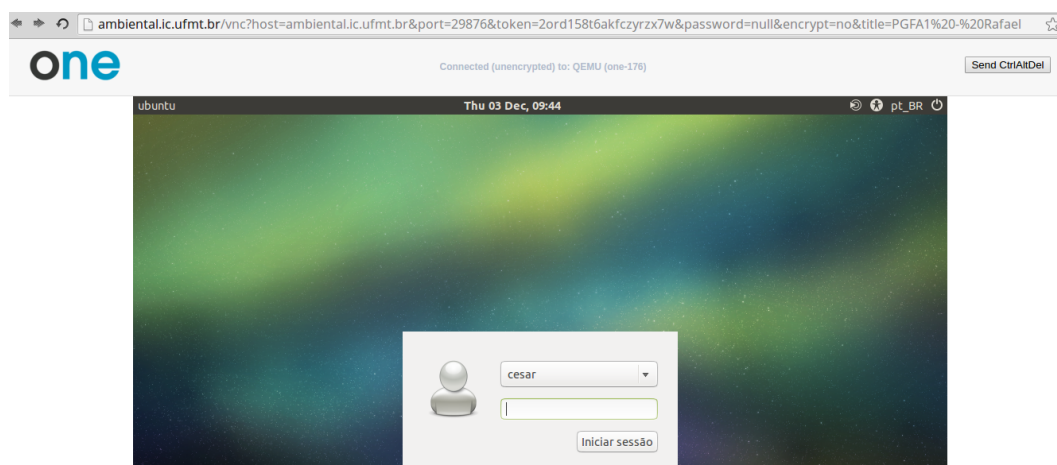


Figura 17: Instância com SO Ubuntu.

Segundo a empresa de levantamentos estatísticos Net Market Share⁶,

⁶<https://netmarketshare.com/>

em pesquisa realizada no ano de 2015, as duas versões, Windows 7 e Windows 8.1, representavam nesse período aproximadamente 70% dos SOs utilizados em computadores domésticos. O Windows também foi testado pelo elevado uso, como evidenciado pela pesquisa. Na loja do OpenNebula não existe nenhuma opção de ISO Windows, e na documentação disponível no site oficial também não existem informações a respeito. Por não existir opção na loja foi necessário pegar uma ISO oficial da Microsoft e criar uma máquina virtual já com *drivers* necessários para o reconhecimento dos recursos de disco e rede, e depois importar no ambiente essa VM.

E por fim a quinta parte do capítulo de resultados busca exemplificar possíveis casos de uso que poderiam utilizar o ambiente *e-science*, são citados *softwares* que possuem utilização na área ambiental.

4.5 Exemplos de uso na área ambiental

O *software* voltado para cálculo numérico Matlab⁷ foi testado no ambiente *e-science* utilizando duas versões de Sistema Operacional da Microsoft. Os testes consistiram em utilizar o recurso plotagem de gráfico e aconteceram nos SOs Windows 7, Figura 18, e Windows 8.1, Figura 19.

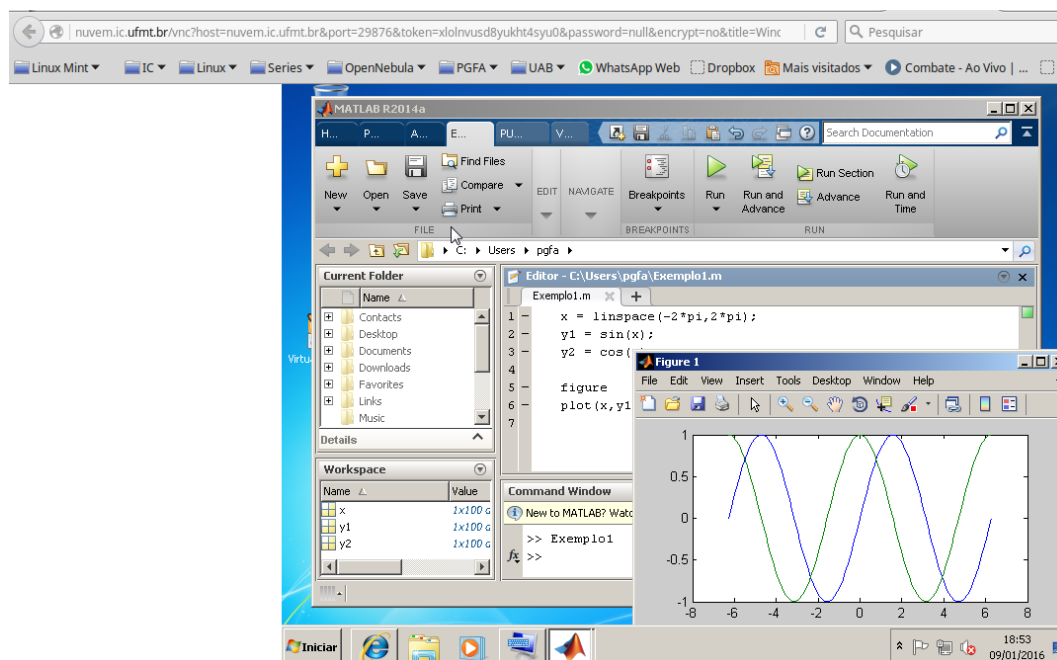


Figura 18: Teste do *software* Matlab no SO Windows 7.

⁷<http://www.mathworks.com/products/matlab/>

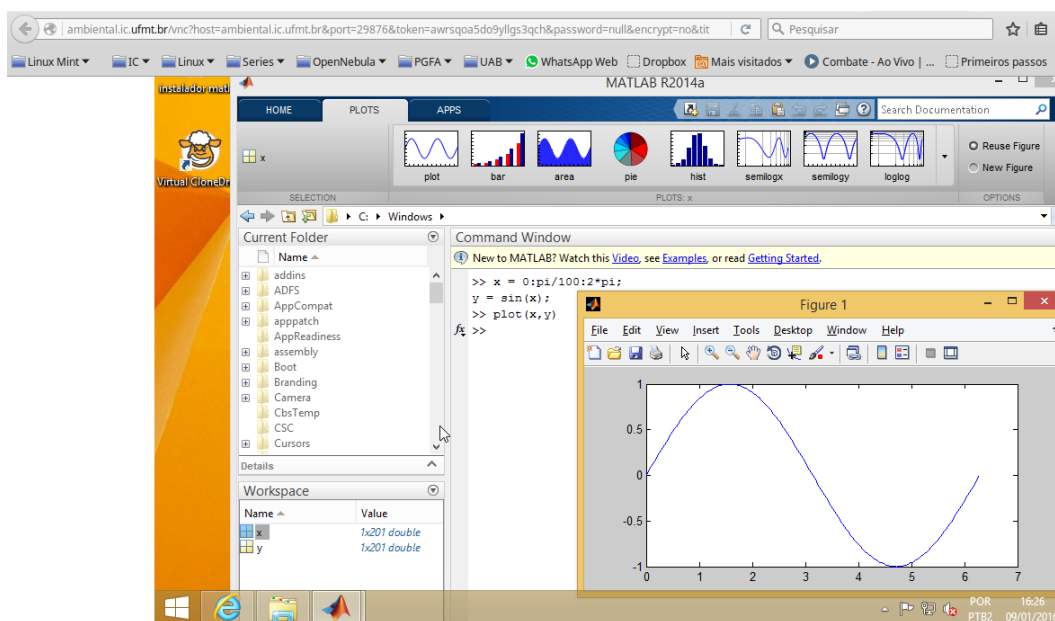


Figura 19: Teste do Matlab no Windows 8.1.

Na Figura 20 visualiza-se uma simulação matemática dentro de uma instância com *software* relacionado à área ambiental, o Octave. No exemplo são usados dois vetores (x e y) para criar um gráfico bidimensional por meio da utilização do comando plot.

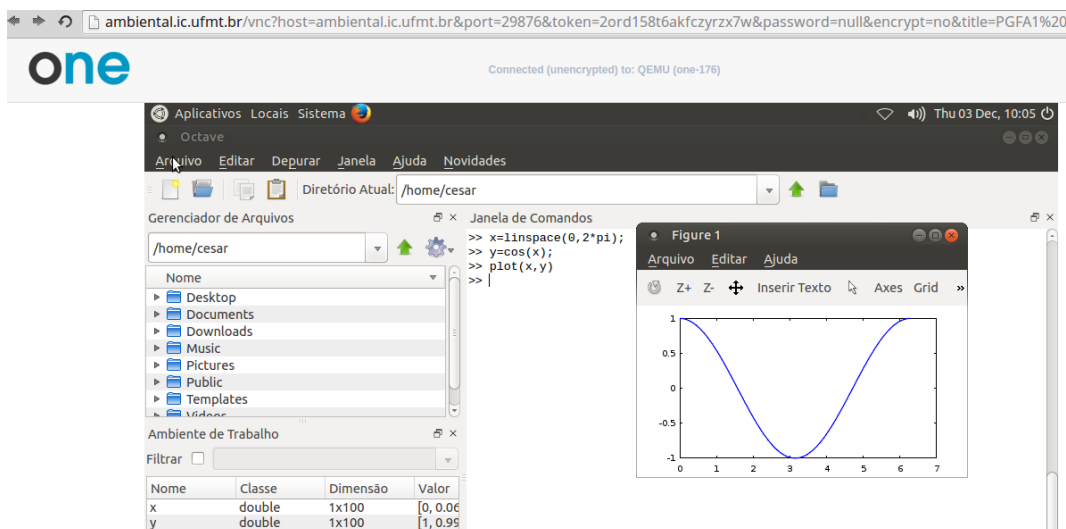


Figura 20: Simulação com Octave.

Também foram testados o R e Rstudio, com exemplo da criação de gráfico, conforme Figura 21:

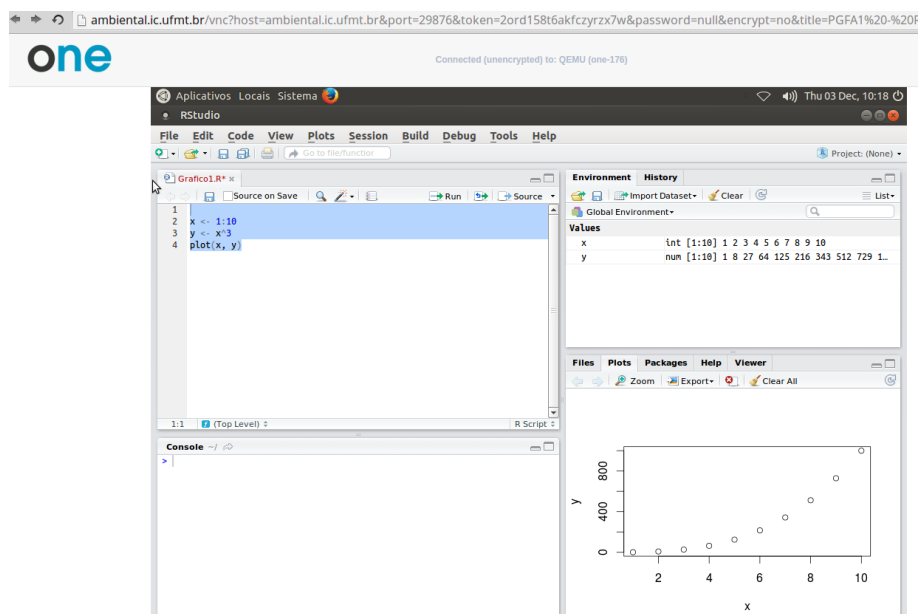


Figura 21: Criação de gráfico com R e Rstudio.

O *software* QuantumGIS, sistema de informação geográfica que usa código aberto, também foi testado utilizando *shapefile*⁸ de Estado da região norte (Amazonas) como pode ser visto na Figura 22:

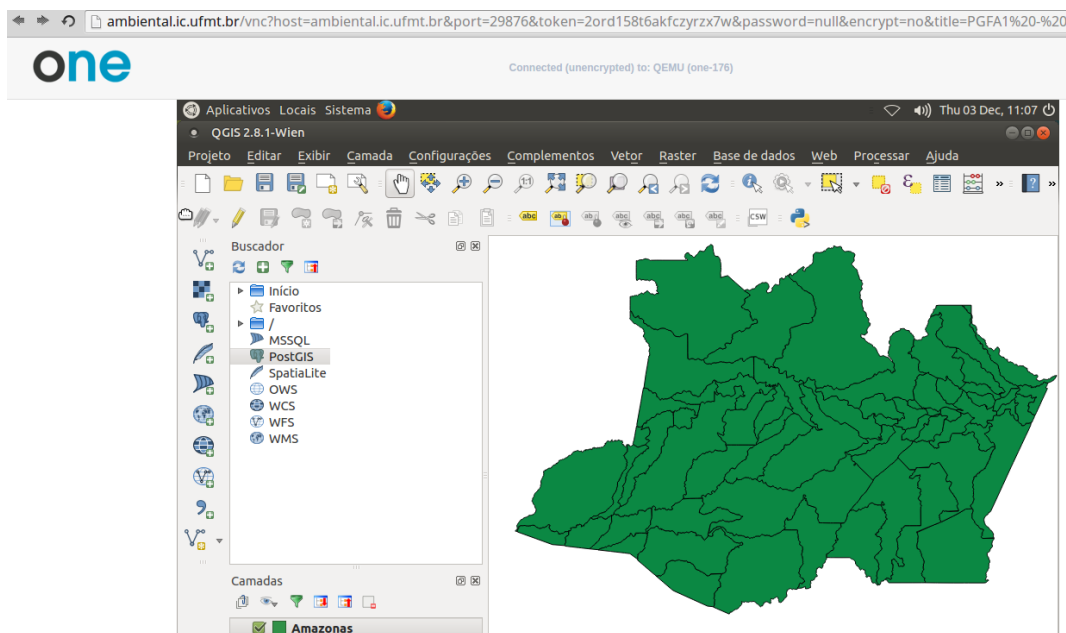


Figura 22: Exemplo de uso do QuantumGIS.

Na Figura 23 um exemplo do uso de código python em uma instância:

⁸Formato de arquivo com informações geoespaciais em forma de vetor.

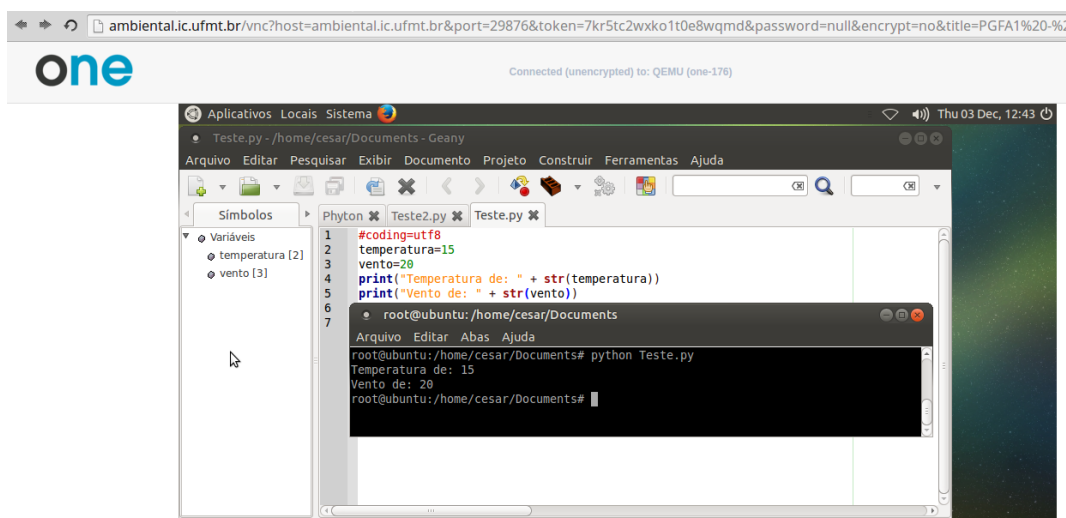


Figura 23: Código python executado via terminal.

O ENVI-met⁹ é um *software*, segundo o site oficial, de desenho urbano e arquitetura sustentável, foi instalado com êxito em uma instância com Sistema Operacional Windows 7, como pode ser observado na Figura 24:

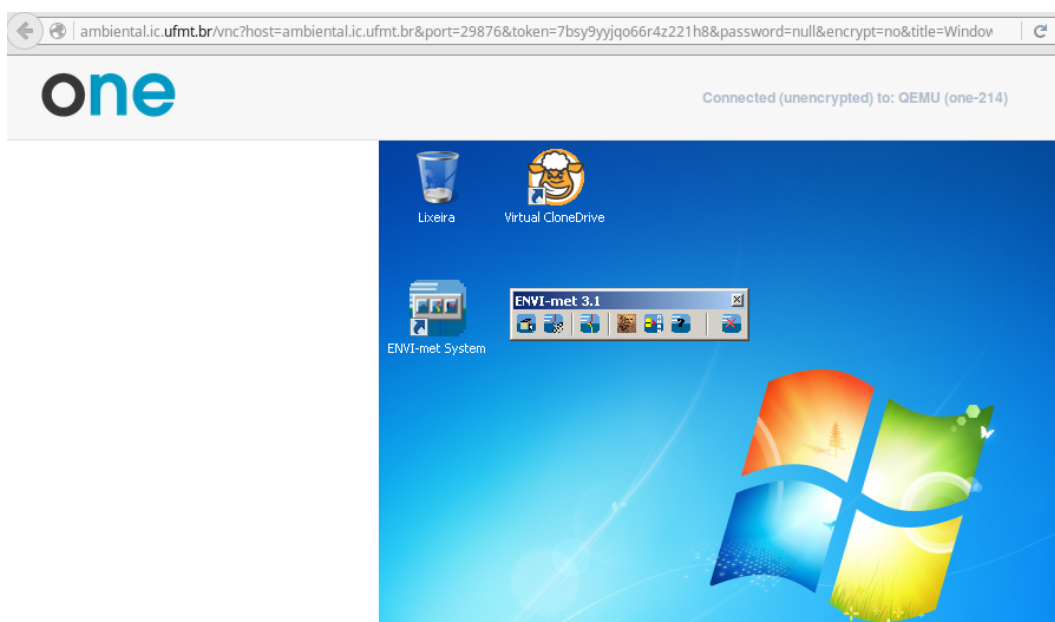


Figura 24: *Software* ENVI-met em execução em VM Windows.

Em 2014, pesquisadores vinculados ao PPGFA, estavam cumprindo uma etapa do doutorado na cidade de Varna, Bulgária. Para os trabalhos publicados em Oliveira et al. (2015) e Ventura et al. (2015) em determinado momento necessitou-se de recursos computacionais como 100 GB de disco e 8 GB de me-

⁹<http://www.envi-met.com>

mória RAM para realização de experimentos utilizando o *software* HTK¹⁰. Como o ambiente não era elástico o prazo para efetuar mudanças e disponibilizar recursos computacionais conforme a necessidade da pesquisa era maior. Além disso, era necessário prover a visualização da máquina virtual, para isso foi preciso efetuar uma série de configurações, por exemplo, do *software* VNC (*Virtual Network Computing*), que possibilita compartilhamento de interface gráfica. Dessa forma, o ambiente *e-science* proposto neste trabalho atenderia essa demanda sem necessitar de novos *softwares* ou configurações adicionais.

Pelos testes realizados foi possível demonstrar que ambiente *e-science* proveu elasticidade no uso de recursos computacionais como disco, memória e processador podendo dar suporte as pesquisas do PPGFA.

Durante a realização desse estudo diversas dificuldades foram encontradas. A primeira e maior dificuldade foi implementar as soluções de IaaS estudadas, pois a maioria exigia requisitos mínimos de *hardware* superior aos disponíveis.

Outra questão que ofereceu dificuldade foi a implementação da visualização de instâncias criadas dentro do ambiente, isso porque era necessário comprar certificado SSL ou utilizar IP público na VM de *frontend*.

O uso de instancias Windows gerou dificuldades como:

- Licenciamento - Adquirir licença Windows para uso acadêmico.
- Utilização do mouse - Ficava duplicado na tela.
- Usar disco e rede - Dificuldades para encontrar *drivers* que reconhecessem esses recursos.

Por fim pode-se citar o papel de convencimento dos pesquisadores convidados para testar a infraestrutura, visto que estão acostumados a trabalhar isoladamente em computadores pessoais e com apenas um sistema operacional, por mais que outras opções se apresentem como intuitivas.

¹⁰<http://htk.eng.cam.ac.uk>

Capítulo 5

Considerações Finais

A partir dos resultados pode-se concluir que o uso de ambiente *e-science* utilizando Computação em Nuvem, do tipo Infraestrutura como Serviço, permitiu o funcionamento com êxito de *softwares* relacionados a área ambiental, inclusive em diferentes Sistemas Operacionais. Além disso, permitiu a implantação de serviços como *backup* e autenticação de usuários. Então, é possível utilizar essas ferramentas e serviços pela Internet tirando proveito de características como elasticidade e maior facilidade no compartilhamento dos dados.

A implementação da infraestrutura *e-science* do PPGFA permite que grupos de pesquisadores possam participar dos estudos mesmo estando em lugares diferentes. O ambiente favorece a troca de experiências e a interação entre cientistas, visto que o PPGFA já possui parcerias com outras instituições tanto em nível nacional quanto internacional.

Dessa forma, o ambiente *e-science* foi adequado para atender diferentes necessidades de pesquisa da área ambiental facilitando o acesso e o compartilhamento de dados de pesquisa em um cenário científico com crescentes demandas de soluções para facilitar o tratamento de grandes volumes de dados.

O objetivo específico de implantar diferentes opções de nuvem computacional não foi possível de ser alcançado, pois este estudo não possuía os recursos mínimos para a implantação.

5.1 Contribuições

Este trabalho contribuiu para disseminação de conhecimentos na criação de *e-science* aplicado a área ambiental utilizando Computação em Nuvem e disponibilizou recursos computacionais dentro dessa infraestrutura via Internet.

As principais contribuições deste estudo são:

- Criação e disponibilização de ambiente *e-science*: disponível na Internet permitindo acesso de pesquisadores por meio de diferentes dispositivos e de qualquer lugar.
- Nuvem com capacidade multiplataforma: utilização de instâncias com variados Sistemas Operacionais, de código aberto e fechado, dessa forma, permitindo a escolha do SO.
- Ambiente colaborativo: facilidade de trabalho cooperativo independente da localização geográfica permitindo o compartilhamento de dados e informações de pesquisa.
- Ponto único de acesso: utilização de *softwares* para modelagem, processamento e visualização de dados ambientais de forma centralizada e via Internet por meio de um programa navegador, sem a necessidade de instalar nenhum *plugin* adicional.
- Uso otimizado da infraestrutura: por meio de virtualização aloca-se recursos como memória e disco de acordo com as necessidades de cada pesquisa, dessa forma, utilizando recursos que poderiam estar ociosos para auxiliar a criar novos serviços.
- Auditoria de uso: possibilidade de auditar o consumo dos recursos computacionais, possibilitando o levantamento de informações de consumo seja por usuário, ou VM, em determinado período de tempo.
- Baixo impacto na mudança da plataforma de trabalho: visto que os pesquisadores podem continuar usando no ambiente Cirrus o mesmo Sistema Operacional e os mesmos *softwares* que já utilizam no modo de pesquisa tradicional.

5.2 Trabalhos Futuros

Como trabalhos futuros pretende-se:

- Estudar as características do gerenciamento de energia para verificar se é possível economia energética em VMs ociosas, visto que a virtualização é uma tecnologia integrante da Computação em Nuvem e um dos seus atrativos é a de redução de custos.

- Ampliar a opção de *softwares* relacionados à área ambiental dentro do ambiente *e-science*, uma vez que as pesquisas na área ambiental possuem diversas necessidades que necessitam de diferentes *softwares*. Portanto, com a disponibilização de mais opções pode-se ampliar a gama de pesquisas que podem ser atendidas pelo ambiente *e-science*.
- Medir a performance do ambiente *e-science* ampliando o número de nós que compõe um *cluster* na criação de várias instâncias. Com isso é possível testar o comportamento do ambiente, por exemplo, na criação de várias VMs ao mesmo tempo com a utilização de mais recursos computacionais.
- Difusão de *e-science* com outros grupos de pesquisa, dessa forma, compartilhando práticas e experiências da implantação do ambiente no PPGFA.
- A partir deste trabalho é possível ter elementos iniciais no sentido de dimensionar a criação de um *datacenter* científico para atender demandas *e-science* em níveis maiores que do PPGFA, ou seja, ampliar o atendimento para a UFMT e outras instituições de Mato Grosso.

REFERÊNCIAS

AHRENS, C. D. *Meteorology today: an introduction to weather, climate, and the environment*. [S.l.]: Cengage Learning, 2012. Citado na página 4.

AMADO, W. R.; MARCONDES, C. A. C. Análise do software bsn para a realização de backups na nuvem. *Revista TIS*, v. 3, n. 3, 2015. Citado na página 26.

APPELBE, B.; BANNON, D. eresearch-paradigm shift or propaganda? *Journal of Research and Practice in Information Technology*, v. 39, n. 2, p. 83–90, 2007. Citado na página 11.

ARAUJO, B. D. L. de; CAVALCANTE, V. M. R. M. Computação em nuvens: Contribuição para a sustentabilidade econômica e ecológica. *Revista da Escola Regional de Informática*, v. 1, n. 1, 2015. Citado na página 14.

ARDESTANI, S. B.; HAKANSSON, C. J.; LAURE, E.; LIVENSON, I.; STRANAK, P.; DIMA, E.; BLOMMESTEIJN, D.; SANDEN, M. V. D. B2share: An open escience data sharing platform. In: IEEE. *e-Science (e-Science), 2015 IEEE 11th International Conference on*. [S.l.], 2015. p. 448–453. Citado na página 11.

ATKINSON, B. W. *Meso-scale atmospheric circulations*. [S.l.]: Academic press, 1989. Citado na página 4.

ATKINSON, M.; CARPENÉ, M.; CASAROTTI, E.; CLAUS, S.; FILGUEIRA, R.; FRANK, A.; GALEA, M.; GARTH, T.; GEMUND, A.; IGEL, H. et al. Verce delivers a productive e-science environment for seismology research. In: IEEE. *e-Science (e-Science), 2015 IEEE 11th International Conference on*. [S.l.], 2015. p. 224–236. Citado na página 11.

BACIS, E.; MUTTI, S.; CAPELLI, S.; PARABOSCHI, S. Dockerpolicymodules: mandatory access control for docker containers. *IEEE CNS*, 2015. Citado na página 13.

BESERRA, D.; FERNANDES, F.; FALCÃO, F.; LIMA, W.; KARMAN, R.; ENDO, P. T.; BORBA, A.; ARAUJO, J.; SOUSA, E. Como a computação em nuvem pode ajudar a prover infraestrutura para aplicações de escience? *Revista de Exatas e TECNológicas*, v. 1, n. 6, 2015. Citado 2 vezes nas páginas 9 e 21.

BOLZAN, M. J. A. Análise da transformada em ondeletas aplicada em sinal geofísico. *Revista Brasileira de Ensino de Física*, SciELO Brasil, v. 26, n. 1, p. 37–41, 2004. Citado na página 4.

BUSKE, D. Solução gillt bidimensional em geometria cartesiana: simulação da dispersão de poluentes na atmosfera. 2008. Citado na página 4.

BUYYA, R.; VECCHIOLA, C.; SELVI, S. T. *Mastering cloud computing: foundations and applications programming*. [S.l.]: Newnes, 2013. Citado na página 13.

CARISSIMI, A. Virtualização: da teoria a soluções. *Minicursos do Simpósio Brasileiro de Redes de Computadores-SBRC*, v. 2008, p. 173–207, 2008. Citado na página 12.

CARR, N. G. *Sera Que Ti E Tudo?* [S.l.]: Editora Gente Liv e Edit Ltd, 2009. Citado na página 13.

CHANDRASEKARAN, K. *Essentials of cloud computing*. [S.l.]: CRC Press, 2014. Citado na página 13.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. Crisp-dm 1.0. *CRISP-DM Consortium*, 2000. Citado 2 vezes nas páginas I e 7.

CHE, J.; YU, Y.; SHI, C.; LIN, W. A synthetical performance evaluation of openvz, xen and kvm. In: IEEE. *Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific*. [S.l.], 2010. p. 587–594. Citado na página 25.

CHEN, J.; WANG, Y.; CHEUNG, K.-H. *Semantic E-science*. [S.l.]: Springer Science & Business Media, 2010. Citado na página 11.

CHEN, X.-Y.; JIN, Z.-G. Research on key technology and applications for internet of things. *Physics Procedia*, Elsevier, v. 33, p. 561–566, 2012. Citado na página 15.

CHU, X.; NADIMINTI, K.; JIN, C.; VENUGOPAL, S.; BUYYA, R. Aneka: Next-generation enterprise grid platform for e-science and e-business applications. In: IEEE. *e-Science and Grid Computing, IEEE International Conference on*. [S.l.], 2007. p. 151–159. Citado na página 15.

CIURANA, E. *Developing with google app engine*. [S.l.]: Apress, 2009. Citado na página 14.

CORDEIRO, D.; BRAGHETTO, K. R.; GOLDMAN, A.; KON, F. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. *Revista USP*, n. 97, p. 71–81, 2013. Citado 5 vezes nas páginas 8, 9, 10, 12 e 14.

CORTES, S. d. C.; PORCARO, R. M.; LIFSCHITZ, S. Mineração de dados—funcionalidades, técnicas e abordagens. *PUC-RioInf, Rio de Janeiro. IN THE BRAZILIAN GOVERNMENT USING CREDIT SCORING* Leonardo sales, 2002. Citado na página 6.

CURADO, L. F. A.; NOGUEIRA, J. de S.; SANCHES, L.; RODRIGUES, T. R.; LOBO, F. de A.; BÍUDES, M. S. Inter seasonality of the energy fluxes in brazilian savana—mato grosso—brazil. *Atmospheric and Climate Sciences*, Scientific Research Publishing, v. 2014, 2014. Citado na página 4.

DEITEL, H. M.; DEITEL, P. J.; CHOFFNES, D. R. *Sistemas operacionais*. [S.l.]: Pearson Prentice Hall, 2005. Citado na página 21.

DOURADO, C. da S.; OLIVEIRA, S. R. de M.; AVILA, A. M. H. de. Análise de zonas homogêneas em séries temporais de precipitação no estado da bahia. *Bragantia*, SciELO Brasil, v. 72, n. 2, p. 192–198, 2013. Citado na página 6.

ESTÉBANEZ, K. M. B. *IDENTIFICAÇÃO DE PADRÕES NA OCORRÊNCIA DE EMERGÊNCIAS E DESASTRES ASSOCIADOS A NÍVEIS DE PRECIPITAÇÃO*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2012. Citado na página 4.

FERNANDES, F. J.; SCHULZE, B.; MURY, A. R. Neblina-espacos virtuais de trabalho para uso em aplicaçoes cientificas. *XIX SBRC*, p. 965–972, 2011. Citado na página 12.

FOKEN, T.; NAPPO, C. J. *Micrometeorology*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 4.

GONÇALVES, A. L. *Utilização de técnicas de mineração de dados em bases de C & T: uma análise dos grupos de pesquisa no Brasil*. Tese (Doutorado) — Universidade Federal de Santa Catarina, Centro Tecnológico., 2000. Citado na página 6.

GUARIENTI, G. S. S. *DESENVOLVIMENTO DE UMA TÉCNICA COMPUTACIONAL DE PROCESSAMENTO ESPAÇO-TEMPORAL APLICADA EM SÉRIES DE PRECIPITAÇÃO*. Dissertação — Universidade Federal de Mato Grosso, 2015. Citado na página 8.

HASHIZUME, K.; ROSADO, D. G.; FERNÁNDEZ-MEDINA, E.; FERNÁNDEZ, E. B. An analysis of security issues for cloud computing. *Journal of Internet Services and Applications*, Springer, v. 4, n. 1, p. 1–13, 2013. Citado na página 16.

KEAHEY, K.; FOSTER, I.; FREEMAN, T.; ZHANG, X. Virtual workspaces: Achieving quality of service and quality of life in the grid. *Scientific programming*, Citeseer, v. 13, n. 4, p. 265–275, 2005. Citado na página 12.

MACHADO, P. A. L. *Direito ambiental brasileiro*. [S.l.]: Malheiros Editores, 1995. Citado na página 5.

MARINO, C. C.; BRENDEN, J.; AMOR, P.; KOTHARI, P. *User-configured on-demand virtual layer-2 network for infrastructure-as-a-service (IaaS) on a hybrid cloud network*. 2015. US Patent 9,154,327. Citado na página 15.

MARTINEZ-ORTIZ, C.; AROYO, L.; INEL, O.; CHAMPILOMATIS, S.; DUMITRACHE, A.; TIMMERMAN, B. Provenance-driven representation of crowdsourcing data for efficient data analysis. In: IEEE. *e-Science (e-Science)*, 2015 IEEE 11th International Conference on. [S.l.], 2015. p. 300–303. Citado na página 11.

MEDEIROS, J. da S.; CAREGNATO, S. E. Compartilhamento de dados e e-science: explorando um novo conceito para a comunicação científica data sharing and e-science: exploring a new concept for scientific communication. *Liinc em Revista*, v. 8, n. 2, 2012. Citado na página 11.

MEIRA, C. A. A.; RODRIGUES, L. H. A. Mineração de dados no desenvolvimento de sistemas de alerta contra doenças de culturas agrícolas. In: *CONGRESSO BRASILEIRO DE AGROINFORMÁTICA*. [S.l.: s.n.], 2005. v. 5. Nenhuma citação no texto.

MELL, P.; GRANCE, T. The nist definition of cloud computing. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, 2011. Citado na página 13.

MORESI, E. A. D.; ALCANTARA, A.; PRADO, H. Cenários prospectivos, monitoração ambiental e metadados. In: *Congresso Anual de Tecnologia da Informação (CATI)*. [S.l.: s.n.], 2005. Citado na página 5.

NAROCK, T.; FOX, P. From science to e-science to semantic e-science: A heliophysics case study. *Computers & Geosciences*, Elsevier, v. 46, p. 248–254, 2012. Citado na página 11.

NETO, R. M.; FAXINA, J. M. Ti verde e sustentabilidade. *Revista de Ciências Exatas e Tecnologia*, v. 7, n. 7, p. 159–174, 2015. Citado na página 14.

NICHOLSON, S. E.; SOME, B.; MCCOLLUM, J.; NELKIN, E.; KLOTTER, D.; BERTE, Y.; DIALLO, B.; GAYE, I.; KPABEBA, G.; NDIAYE, O. et al. Validation of trmm and other rainfall estimates with a high-density gauge dataset for west africa. part ii: Validation of trmm rainfall products. *Journal of Applied Meteorology*, v. 42, n. 10, p. 1355–1368, 2003. Citado na página 4.

OLIVEIRA, A. G. de. *MiMi: Plataforma computacional para mineração de dados micrometeorológicos*. Tese (Doutorado) — Universidade Federal de Mato Grosso, Cuiabá, MT, Brasil, Maio 2015. Citado 3 vezes nas páginas I, 7 e 8.

OLIVEIRA, A. G. de; VENTURA, T. M.; GANCHEV, T. D.; FIGUEIREDO, J. M. de; JAHN, O.; MARQUES, M. I.; SCHUCHMANN, K.-L. Bird acoustic activity detection based on morphological filtering of the spectrogram. *Applied Acoustics*, v. 98, p. 34 – 42, 2015. ISSN 0003-682X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0003682X15001309>>. Citado na página 35.

RODRIGUES, T. R.; PAULO, S. R. de; NOVAIS, J. W.; CURADO, L. F.; NOGUEIRA, J. S.; OLIVEIRA, R. G. de; LOBO, F. d. A.; VOURLITIS, G. L. Temporal patterns of energy balance for a brazilian tropical savanna under contrasting seasonal conditions. *International Journal of Atmospheric Sciences*, Hindawi Publishing Corporation, v. 2013, 2013. Citado na página 4.

SAAD, A.; CERIN, J. A self-organized volunteer cloud for e-science. *The Journal of Supercomputer*, Springer, 2015. Citado na página 11.

SHEARER, C. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, THE DATA WAREHOUSE INSTITUTE, v. 5, n. 4, p. 13–22, 2000. Citado na página 6.

SILVA, C. E. P. *CAPTURE DE DADOS DE PROVENIÊNCIA DE WORK-FLOWS CIENTÍFICOS EM NUVENS COMPUTACIONAIS*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2011. Citado na página 10.

SOUSA, F. R.; MOREIRA, L. O.; MACÊDO, J. d.; MACHADO, J. C. Gerenciamento de dados em nuvem: Conceitos, sistemas e desafios. *Temas em sistemas colaborativos, interativos, multimídia, web e bancos de dados, Sociedade Brasileira de Computação*, p. 101–130, 2010. Citado na página 15.

TANENBAUM, A. S.; STEEN, M. V. *Princípios e Paradigmas. 2ª Edição. Autores: Andrew S. Tanenbaum e Maarten V. Steen*. [S.l.]: Editora Pearson Prentice Hall, 2007. Citado na página 8.

THAMADA, T. T.; NETO, C. D. G.; MEIRA, C. A. Sistema de alerta da ferrugem do cafeeiro1. 2013. Citado na página 6.

VAQUERO, L. M.; RODERO-MERINO, L.; CACERES, J.; LINDNER, M. A break in the clouds: towards a cloud definition. *ACM SIGCOMM Computer Communication Review*, ACM, v. 39, n. 1, p. 50–55, 2008. Citado na página 15.

VENTURA, T. M. *Criação de um Ambiente Computacional para Detecção de Outliers e Preenchimento de Falhas em Dados Meteorológicos*. Tese (Doutorado) — Universidade Federal de Mato Grosso, Cuiabá, MT, Brasil, Fevereiro 2015. Citado na página 8.

VENTURA, T. M.; MARQUES, H. O.; OLIVEIRA, A. G. de; MARTINS, C. A.; NOGUEIRA, M. C.; TEIXEIRA, W. R. et al. Detecção de outliers em dados micrometeorológicos. 2013. Citado na página 5.

VENTURA, T. M.; OLIVEIRA, A. G. de; GANCHEV, T. D.; FIGUEIREDO, J. M. de; JAHN, O.; MARQUES, M. I.; SCHUCHMANN, K.-L. Audio parameterization with robust frame selection for improved bird identification. *Expert Systems with Applications*, v. 42, n. 22, p. 8463 – 8471, 2015. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417415004625>>. Citado na página 35.

VILANI, M. *Análise de Fourier e Wavelet em Variáveis Micrometeorológicas em Diferentes Tipologias de Ocupação*. Tese (Doutorado) — PhD thesis, Universidade Federal de Mato Grosso, 2011. Citado na página 4.

VOGEL, A. J.; GRIEBLER, D.; ROVEDA, D. Understanding, discussing and analysing the opennebula's and openstack's iaas management layers. *Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação*, v. 1, n. 3, 2015. Citado na página 18.

VUKOJEVIC-HAUPT, K.; GOMEZ-SAEZ, S. A middleware-centric optimization approach for the automated. In: IEEE Proceedings of the 2015 IEEE 7th International Conference on. [S.l.], 2015. p. 174–179. Citado na página 12.

WIRTH, R.; HIPPEL, J. Crisp-dm: Towards a standard process model for data mining. In: CITESEER. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. [S.l.], 2000. p. 29–39. Citado na página 6.

YADAV, S. Comparative study on open source software for cloud computing platform: Eucalyptus, openstack and opennebula. *International Journal Of Engineering And Science*, v. 3, n. 10, p. 51–54, 2013. Citado na página 16.

YOUNGE, A. J.; HENSCHER, R.; BROWN, J. T.; LASZEWSKI, G. V.; QIU, J.; FOX, G. C. Analysis of virtualization technologies for high performance computing environments. In: IEEE. *Cloud Computing (CLOUD), 2011 IEEE International Conference on*. [S.l.], 2011. p. 9–16. Citado na página 25.

YU, H.-E.; HUANG, W. Building a virtual hpc cluster with auto scaling by the docker. *arXiv preprint arXiv:1509.08231*, 2015. Citado na página 13.

Anexo A: Material complementar

DOCUMENTAÇÃO

Versão do OpenNebula: 4.12

SO: Ubuntu 14 Server

Obs: qualquer tarefa administrativa deve usar o usuário oneadmin.

- Infraestrutura utilizada

Frontend: VM 4GB RAM - Rede: IP qualquer da LAN - nome: opennebula

Nó: 1 - IBM 8GB RAM - 400 gigabytes de disco - Rede: IP qualquer da LAN - nome: opennebula1

Nó: 2 - IBM 8GB RAM - 400 gigabytes de disco - Rede: IP qualquer da LAN - nome: opennebula2

- Verificar serviços Checar se rpcbind e nfs estão ON:
/etc/init.d/nfs-kernel-server status

- Logar como usuário oneadmin
su - oneadmin

- Acompanhar os logs
tail -f /var/log/one/oned.log

- Verificar se o front end se conecta sem senha com os nós
ssh oneadmin@10.5.0.27

- Colocar scripts importantes (nfs, opennebula-sunstone) pra iniciar automático
update-rc.d opennebula-sunstone defaults

- Colocar nome dos nós no “/etc/hosts”
Exemplo:
10.1.0.2 opennebula1.ic opennebula1

10.1.0.3 opennebula2.ic opennebula2

- Checar com mount se carregou a montagem

```
mount /var/lib/one/ -vvv
```

- Visualizar a senha do usuário oneadmin

Checar pasta /home/.one/one_auth

- Exemplo de criação de imagem:

```
oneimage create --name "ubuntu12"
```

```
--path "http://cloud-images.ubuntu.com/releases/precise/release/ubuntu-12.04"
```

```
--server-cloudimg-amd64-disk1.img"
```

```
--driver qcow2
```

```
--datastore default
```

- Exemplo de criação de template:

```
onetemplate create --name "ubuntu12--cpu 1 --vcpu 1 --memory 512"
```

```
--arch x86_64 --disk "ubuntu12" --nic "private" --vnc --ssh
```

- Permissão de acesso por ssh. Editar o arquivo alterando da seguinte forma:

```
etc/ssh/sshd_config PasswordAuthentication = yes
```

- Instalação de Ambientes gráficos

LXDE

```
sudo apt-get install lxde
```

XFCE

```
sudo add-apt-repository ppa:xubuntu-dev/xfce-4.10 apt-get install xubuntu-desktop
```

MATE

```
sudo apt-add-repository ppa:ubuntu-mate-dev/ppa sudo apt-add-repository ppa:ubuntu-mate-dev/trusty-mate apt-get update && sudo apt-get upgrade sudo apt-get install --no-install-recommends ubuntu-mate-core ubuntu-mate-desktop ; reboot
```

- Ajuste de ponteiro do mouse em instância Windows

Editar template Windows em “INPUT/OUTPUT” adicionar entrada do tipo “tablet” e bus “USB”, clicar em adicionar.