



Proposta de Serviço Piloto

Grupo de Trabalho □ Segunda Fase

mc^2 – minha *cloud*

científica!

Antônio Tadeu Azevedo Gomes (LNCC/MCTI) e
Francisco Vilar Brasileiro (UFCG)

05 de setembro de 2012

1. Concepção do serviço

1.1. Resumo

O resultado principal deste projeto na Fase 1 foi o desenvolvimento e implantação de uma plataforma de computação na nuvem voltada para aplicações de e-ciência denominada mc². Essa plataforma provê facilidades de: (i) acesso, em diferentes perfis de permissão, a uma variedade de recursos computacionais para a execução de experimentos, (ii) armazenamento, compartilhamento e publicação de resultados de experimentos, (iii) reprodutibilidade de experimentos, e (iv) controle de proveniência dos dados consumidos e gerados pelos experimentos. A plataforma segue um modelo PaaS (*Platform-as-a-Service*), permitindo o fácil desenvolvimento e implantação de ambientes de usuário personalizados, oferecidos em um modelo SaaS (*Software-as-a-Service*). A pilha de serviços também inclui um *broker* para mediar a alocação de recursos em diferentes tipos de provedores de IaaS (*Infrastructure-as-a-Service*), atendendo requisitos de segurança específicos de cada aplicação e demandas de carga de trabalho típicas de aplicações de e-ciência. A proposta de trabalho na Fase 2 contempla atividades de: (i) multiplicação de desenvolvedores aptos a operar a plataforma mc² e divulgação dos ambientes de SaaS desenvolvidos na mesma nas instituições usuárias; (ii) manutenção corretiva, perfectiva e evolutiva dos diversos módulos que compõem a plataforma mc²; (iii) desenvolvimento de soluções para uma maior descentralização da administração da plataforma mc²; e (iv) estudo de soluções de integração com iniciativas da RNP ligadas a computação em nuvem, armazenamento, autenticação federada e provisionamento dinâmico de circuitos.

1.2. Abstract

The main result of the project in its first phase has been the development and deployment of a cloud computing platform tailored to e-science applications. This platform, so called mc², provides e-science facilities such as: (i) access, in different permission profiles, to a variety of computing resources for the execution of experiments, (ii) storage, sharing and publishing of experiment results, (iii) experiment reproducibility, and (iv) provenance tracking of all data consumed and generated by the experiments. The platform follows the Platform-as-a-Service (PaaS) model, allowing the development and deployment of customized e-science user environments in a simple way, following a Software-as-a-Service (SaaS) model. The stack of services also includes a broker that is able to mediate the allocation of computing resources in different Infrastructure-as-a-Service (IaaS) providers, catering to specific security requirements of each application, and the typical resource demands of e-science applications. The proposal of a second phase for this project aims at the following activities: (i) training developers to operate the mc² platform and promoting the SaaS environments customized by such platform at the user institutions; (ii) correcting and evolving the various modules that comprise the mc² platform; (iii) developing solutions to decentralize the administration of the mc² platform; and (iv) studying solutions for the integration of the mc² platform with RNP initiatives

related with cloud computing, storage, federated authentication, and dynamic circuit provisioning.

1.3. Descrição do serviço proposto

Esta proposta é para a Fase 2 do Grupo de Trabalho mc². Nesse sentido, o *background* e motivação gerais para esta proposta são os mesmos apresentados na proposta para a Fase 1 (período 2011-2012). Por sua vez, o detalhamento do serviço piloto proposto explora e propõe expandir os resultados obtidos na Fase 1, de modo que esse serviço possa ser ofertado no âmbito de um escopo limitado de instituições usuárias – englobando inicialmente o próprio LNCC, a UFCG, a UFBA, o CENAPADCE/UFC e o CESUP/UFRGS – em conjunto com a RNP.

Background e motivação

É reconhecido que muitos dos avanços recentes em pesquisas científicas somente foram possíveis devido à habilidade dos cientistas em usar eficientemente computadores para gerar e processar grandes quantidades de dados. O paradigma de computação na nuvem (do inglês *cloud computing*) abre novas possibilidades para esse uso. Computação na nuvem é tipicamente oferecida sobre três modelos de serviço distintos, normalmente formando uma pilha de serviços, com cada camada da pilha provendo um nível de abstração diferente.

No nível mais baixo dessa pilha está o que se chama de IaaS (*infrastructure-as-a-service*). Nesse nível, recursos de processamento, armazenamento e rede são ofertados de forma customizada, atendendo aos requisitos de cada cliente. Os serviços de IaaS podem atender clientes que desejam hospedar suas aplicações diretamente na infraestrutura, ou servir de base para a oferta de serviços de mais alto nível como PaaS (*platform-as-a-service*) e SaaS (*software-as-a-service*).

Um provedor de PaaS oferece uma plataforma que permite ao cliente criar e desenvolver aplicações elásticas capazes de atender um grande número de requisições de maneira facilitada e sem ter que se preocupar com os detalhes da plataforma de execução. Um cliente de PaaS tem acesso a um ambiente de desenvolvimento e de gerência de aplicação que permite o desenvolvimento, teste e implantação de aplicações customizadas, oferecidas em um modelo de SaaS, ou seja, através de uma interface Web. As ferramentas de desenvolvimento e as aplicações desenvolvidas são acessadas através de um navegador Web, o que leva a uma necessidade reduzida de instalação de software tanto para o desenvolvedor das aplicações quanto para os usuários dessas aplicações. Essa característica facilita questões de gerência de software e dos dados associados às aplicações.

A comunidade científica não está indiferente a este novo paradigma e várias iniciativas em todo o mundo já investigam a aplicabilidade de ambientes de computação na nuvem para execução de cargas de trabalho científicas (e-ciência) [Fox 2011] [Evangelinos

2008] [Juve 2009] [Keahey 2010] [Oliveira 2011][Osterman 2008] [Walker 2008] [Watson 2011].

Os resultados preliminares indicam que embora sejam ambientes muito flexíveis e simples de se configurar, disponibilizar uma plataforma de computação na nuvem para dar suporte a atividades de e-ciência não é tão automático como desejado, considerando-se as implementações disponíveis. Em particular, a sua efetiva utilização e controle por parte do pesquisador ainda apresenta alguns desafios [Lee 2010]. As principais dificuldades residem principalmente no fato de que as propriedades mais atrativas do paradigma de computação na nuvem para o contexto acadêmico – conveniência e facilidade de uso, ampla elasticidade e níveis adequados de abstração – ainda não são plenamente endereçados pelo estado-da-prática do segmento. Isso ocorre por que os atuais serviços públicos de computação na nuvem foram concebidos tendo em vista o atendimento de requisitos bem específicos como, por exemplo, o suprimento de infraestrutura para pequenas e médias empresas e o suporte computacional para aplicações do estilo Web 2.0, fazendo com que alguns contextos, inclusive o de e-ciência, não sejam ainda apropriadamente atendidos.

Por outro lado, plataformas de suporte a atividades de e-ciência mais tradicionais, como aquelas baseadas em grades computacionais, também têm limitações. Em particular, iniciativas como o GT-VCG da RNP no Brasil e o TeraGrid da NSF nos EUA foram primordialmente orientadas à tecnologia de grades adotada para essas plataformas, o que dificulta a customização dessas plataformas para atender as necessidades específicas de seus usuários. Uma indicação nesse sentido foi o anúncio pela NSF, em julho de 2011, do encerramento do programa TeraGrid e do lançamento de um novo programa chamado XSEDE (EXtreme Science and Engineering Discovery Environment – <http://www.xsede.org>), focado no oferecimento de um ambiente em que cientistas possam usar e compartilhar tanto recursos computacionais quanto aplicações, ferramentas, e serviços.

A plataforma mc²

A plataforma mc², elemento central do serviço piloto proposto, é uma solução de computação em nuvem voltada para atender necessidades de e-ciência. A Figura 1 ilustra os níveis de serviço e os principais atores da plataforma mc², descritos adiante.

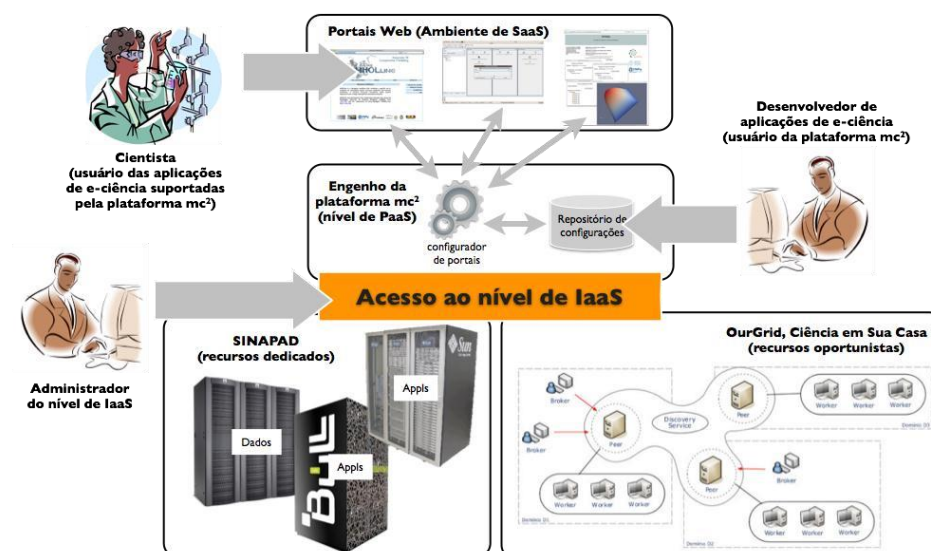


Figura 1: Níveis de serviço e atores da plataforma mc².

Através da plataforma mc², um desenvolvedor poderá rapidamente customizar **ambientes de SaaS** para atender necessidades específicas de um cientista ou um grupo de cientistas com demandas similares. Dessa forma, cada usuário de aplicação de e-ciência – o cientista – terá uma visão única do serviço de computação na nuvem que dá suporte à sua aplicação, dando-lhe a impressão de que ele opera sobre a sua própria nuvem particular (“a minha *cloud* científica!”).

Esses ambientes de SaaS são ofertados na plataforma mc² sob a forma de portais Web, também chamados de *science gateways* no contexto de aplicações de e-ciência, permitindo uma fácil interação do cientista com os serviços computacionais e de armazenamento da infraestrutura subjacente (nível de IaaS).¹ Ao diminuir sensivelmente a curva de aprendizado para que o cientista comece a realizar trabalho útil com a ferramenta, o ambiente de SaaS aumenta substancialmente a produtividade do cientista, que pode se concentrar majoritariamente no objeto de sua investigação, e não nos detalhes do ferramental tecnológico que lhe é disponibilizado.

O desenvolvimento de ambientes de SaaS sobre a plataforma mc² consiste em acessar interfaces Web de fácil utilização e configurar as facilidades necessárias para dar suporte aos requisitos específicos de uma aplicação particular. Essas facilidades são implementadas no nível de PaaS e apenas customizadas pelos usuários desse nível de serviço – os desenvolvedores de aplicações de e-ciência.

Para o sucesso do serviço piloto, duas atividades fundamentais a serem desenvolvidas ao longo da Fase 2 do projeto, ligadas à customização de ambientes de SaaS mencionada acima, serão: (i) a multiplicação de desenvolvedores de aplicações de e-ciência aptos a operar a plataforma mc² nas instituições usuárias do serviço piloto proposto, e (ii) a divulgação dos ambientes de SaaS configurados com esses sistemas

¹ Uma técnica computacional usualmente empregada no nível de IaaS é a virtualização. No contexto de e-ciência a virtualização pode ou não ser interessante dependendo da característica da aplicação científica que se quer dar suporte. Para a presente proposta, o fato da virtualização ser ou não explorada é deixado a cargo do gerenciamento das diferentes infraestruturas subjacentes que compõem o nível de IaaS na plataforma mc².

junto aos usuários dessas instituições. Um usuário em particular já contactado para a Fase 2 do projeto é o GT-AAAS, coordenado pela UFPB, que apresenta claras demandas em termos de capacidade de processamento e armazenamento. O suporte a esse usuário será realizado pela parte da equipe da Fase 2 deste projeto baseada na UFCG.

1.4. Identificação do público alvo

O público alvo do serviço piloto proposto engloba toda a comunidade acadêmica nacional que usa computação de alto desempenho ou de alta vazão para gerar, processar, armazenar e/ou analisar grandes quantidades de dados. Mais especificamente, serão potenciais beneficiários não somente os cientistas – usuários das aplicações de e-ciência – como também os próprios desenvolvedores dessas aplicações.

2. Definição do serviço piloto

2.1. Arquitetura do serviço piloto

A arquitetura proposta para a plataforma mc² é apresentada na Figura 2. Os principais módulos funcionais que compõem essa arquitetura são listados abaixo. Os módulos em verde na figura representam produtos desenvolvidos na Fase 1 deste projeto, os módulos em azul produtos pré-existentes que foram adaptados ao longo da Fase 1 deste projeto, e os módulos em amarelo produtos pré-existentes que foram integrados sem modificações à plataforma mc² na Fase 1 deste projeto.

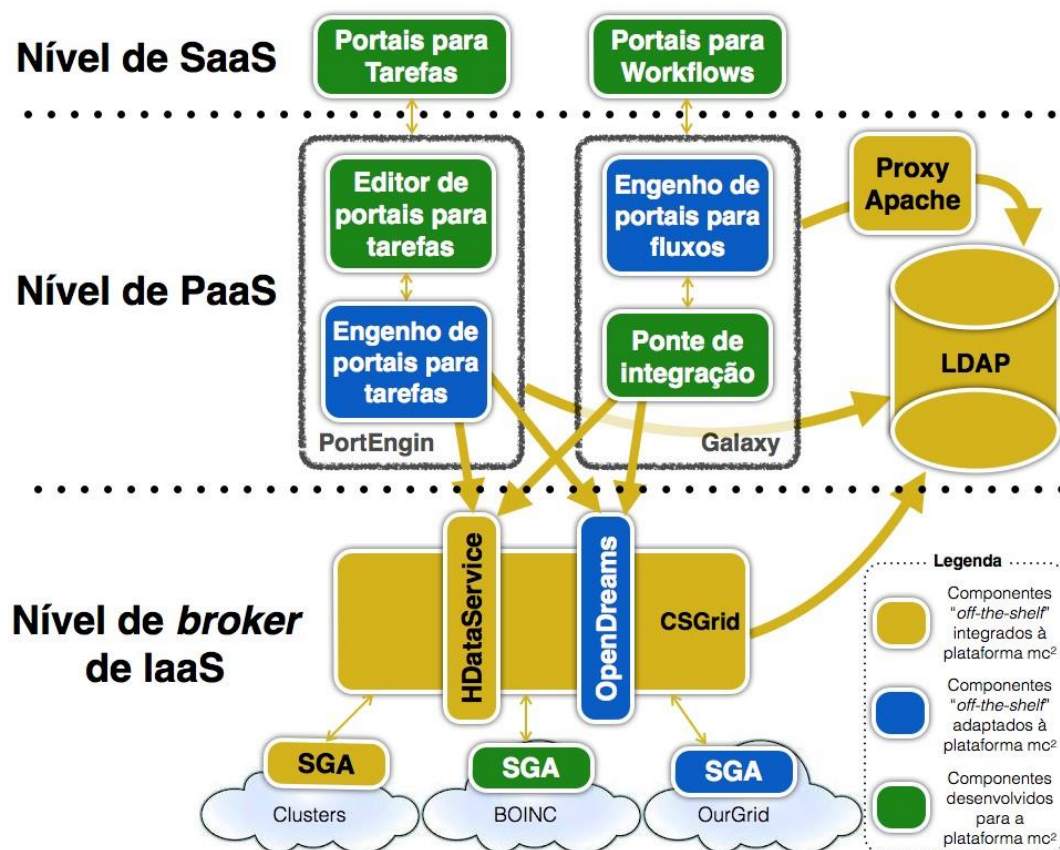


Figura 2: Arquitetura da plataforma mc².

- **Middleware CSGrid** (<http://jira.tecgraf.puc-rio.br/confluence/display/CSGrid>): desenvolvido pela PUC-Rio e estendido na Fase 1 deste projeto pelo LNCC com recursos próprios bem como com recursos do Programa de GTs da RNP. Esse middleware é responsável pela integração no nível de IaaS da plataforma mc² de três infraestruturas computacionais: (i) uma grade P2P (a Comunidade OurGrid, operada pela UFCG – <http://www.ourgrid.org>), (ii) um serviço de computação voluntária (o Serviço “Ciência em sua casa”, operado pela UFCG e baseado no middleware BOINC – <http://mamarreis.lsd.ufcg.edu.br/cienciaem-sua-casa/>) e (iii) recursos de alto desempenho providos pelo SINAPAD (operado pelo LNCC – <http://www.lncc.br/sinapad>). As principais extensões feitas no CSGrid durante a Fase 1 deste projeto foram na interface de serviço do CSGrid para submissão de tarefas (OpenDreams) e na inclusão de novos agentes controladores de infraestruturas (SGAs) para o OurGrid e para o BOINC. O middleware CSGrid já se encontra em execução em servidores do LNCC, onde permanecerá hospedado ao longo da Fase 2 deste projeto.
- **Sistema PortEngin** [Bastos e Gomes, 2012]: desenvolvido pelo LNCC e estendido na Fase 1 deste projeto pelo LNCC com recursos do Programa de GTs da RNP. Esse sistema visa o oferecimento de ambientes de SaaS sob a forma de portais para aplicações científicas implementadas como um único executável ou como um script que combina diferentes executáveis em um fluxo não gerenciável de atividades. Essa abordagem visa simplificar ao máximo o

processo de implantação de um ambiente de SaaS para essas aplicações por meio de uma estratégia de *zero programming*. O LNCC desenvolveu dois portais Web durante a Fase 1 do projeto utilizando o editor de portais para tarefas do PortEngin (<http://www.lncc.br/sinapad/PortEditor> – por enquanto acessível somente na rede interna do LNCC), e que estão em operação em servidores do LNCC: <http://www.lncc.br/sinapad/Profrager> e <http://www.lncc.br/sinapad/DANCE>. Ao longo da Fase 2 propõe-se expandir o portfólio de portais atendidos por essa abordagem junto às demais instituições usuárias do serviço piloto proposto, bem como instalar servidores nessas instituições para a hospedagem desses novos portais.

- **Sistema Galaxy** (<http://galaxy.psu.edu/>): desenvolvido pela Pittsburgh State University e adaptado na Fase 1 deste projeto pela UFCG com recursos do Programa de GTs da RNP. Esse sistema visa o oferecimento de interfaces de programação visual para fluxos gerenciáveis de tarefas a partir da combinação e execução de uma série de diferentes processos (usualmente simples) que manipulam de forma incremental os dados de interesse do cientista. Um portal Web para o Galaxy foi implantado em caráter experimental em um servidor do LNCC durante a Fase 1 do projeto: <http://galaxy.sinapad.lncc.br> (por enquanto acessível somente na rede interna do LNCC). Ao longo da Fase 2 propõe-se estudar outros sistemas gerenciadores de fluxos de tarefas visando atender diferentes comunidades nas instituições usuárias do serviço piloto proposto (uma vez que o Galaxy é conhecido principalmente pela comunidade de bioinformática), bem como instalar servidores nessas instituições para a hospedagem desses sistemas.

2.2. Instituições participantes

Laboratório Nacional de Computação Científica (LNCC)

Coordenação de Ciência da Computação (CCC)

Sistema Nacional de Processamento de Alto Desempenho (SINAPAD) **Contato:**

Antônio Tadeu Azevedo Gomes (Coordenador Geral do GT mc²).

Antônio Tadeu A. Gomes é pesquisador do LNCC e secretário executivo do SINAPAD (<http://www.lncc.br/sinapad>), que tem por missão prestar serviços de processamento de alto desempenho (PAD) sob demanda para a comunidade acadêmica e científica nacional, bem como difundir a cultura e os conhecimentos associados a essa tecnologia. Esse consórcio é coordenado pelo LNCC e constituído pelos centros de computação operados por CENAPAD-CE/UFC, CESUP/UFRGS, LCC/UFGM, CENAPAD-PE/UFPE, NACAD/COPPE-UFRJ, CENAPAD-SP/Unicamp, CPTEC/INPE e o próprio LNCC. A experiência do LNCC com aplicações de e-ciência e a grande rede de usuários do SINAPAD serão fundamentais para que um grande número de aplicações possam ser portadas para execução no serviço proposto. Antônio Tadeu também foi coordenador adjunto do GT-VCG nas edições de 2007-2008 (Fase 1) e 2008-2009 (Fase 2) do Programa de Grupos de Trabalho da RNP. Por fim, Antônio Tadeu atua como coordenador técnico da célula do LNCC no projeto JiT Clouds. Esse projeto é financiado

pelo CTIC/RNP e tem por objetivo investigar uma arquitetura alternativa para a construção de provedores de computação em nuvem, onde os mesmos apenas incorrem em custos de propriedade quando os recursos usados para prover a sua infraestrutura são demandados pelos seus clientes, permitindo uma ampliação de algumas ordens de magnitude no limite que precisa ser imposto aos clientes de tais serviços.

Universidade Federal de Campina Grande (UFCG)

Departamento de Sistemas e Computação (DSC)

Contato: Francisco Vilar Brasileiro (Coordenador Adjunto do GT mc²)

Francisco V. Brasileiro é o coordenador geral do projeto JiT Clouds, formado por 15 instituições nacionais e internacionais. Ele também lidera o projeto OurGrid (<http://www.ourgrid.org/>), que vem sendo executado desde 2003, e foi responsável pelo desenvolvimento de tecnologia inovadora em grades computacionais, particularmente para a implantação de grades de livre acesso baseadas na arquitetura P2P (*peer-to-peer*). O Laboratório de Sistemas Distribuídos da UFCG, por ele coordenado, desenvolve há algum tempo parceria com o projeto Ibercivis (<http://www.ibercivis.es/> e <http://www.ibercivis.pt/>). Esse projeto é um dos maiores projetos de computação voluntária da Europa. Atualmente congrega mais de 70.000 computadores domésticos na Espanha e em Portugal, já tendo sido utilizado por mais de 10 aplicações de e-ciência diferentes, gerando vários resultados de pesquisa importantes. Essa parceria culminou com a implantação de um serviço de computação voluntária para o Brasil (“Ciência em sua casa”), que opera de forma federada com os recursos do projeto Ibercivis, permitindo a imediata utilização dos recursos dessa infraestrutura por pesquisadores brasileiros.

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Departamento de Informática (DI)

Contato: Renato Cerqueira

Renato Cerqueira é coordenador da área de engenharia de sistemas distribuídos do Laboratório Tecgraf da PUC-Rio, que desenvolve projetos de P&D com diversos parceiros industriais e da academia, tais como PETROBRAS, Microsoft Research, Instituto de Pesquisa da Marinha, Globo, CEPEL, IME/USP e LNCC. Esses projetos têm gerado ferramentas de programação distribuída e sistemas de middleware que são usados por grupos de pesquisa e empresas, tanto no Brasil quanto no exterior. Em particular, Renato é o principal arquiteto do middleware CSGrid (<http://jira.tecgraf.puc-rio.br/confluence/display/CSGrid/>) para integração de aplicações científicas distribuídas, que vem sendo adotada pelo SINAPAD.

Universidade Federal da Bahia (UFBA)

Departamento de Ciência da Computação (DCC)

Contato: Marcos E. Barreto

Marcos E. Barreto é pesquisador do Laboratório de Sistemas Distribuídos (LaSiD), do DCC da UFBA, onde atua como coordenador técnico da célula da UFBA no projeto JiT Clouds (financiado pelo CTIC/RNP). Também atua como pesquisador colaborador num projeto de monitoramento e análise de conteúdo ilícito em redes BitTorrent, coordenado pelo professor Luciano Barreto, em parceria com a ONG SaferNet e a Polícia Federal. Possui experiência com sistemas paralelos e distribuídos, adquirida em aproximadamente 15 anos de estudos e pesquisas junto ao Grupo de Processamento Paralelo e Distribuído do Instituto de Informática da UFRGS, no qual desenvolveu sua tese que versou sobre integração de sistemas paralelos heterogêneos. Também atuou como Assistente 1 no GT-UniT (Monitoramento do Universo Torrent), coordenado pelo professor Marinho Barcelos (UFRGS), desenvolvendo atividades de coordenação técnica e administrativa entre os meses de janeiro a maio de 2010.

Universidade Federal do Rio Grande do Sul (UFRGS)

Centro Nacional de Supercomputação (CESUP)

Contato: Denise G. Ewald

Denise G. Ewald atua no CESUP/UFRGS desde a sua criação, tendo criado o Serviço de Apoio ao Usuário e a Divisão de Divulgação e Treinamento desse centro. Denise atua como diretora do CESUP desde 1996. Denise possui graduação em Tecnólogo em Processamento de Dados pela UFRGS (1975), especialização em Administração pela UFRGS (1985) e especialização em Análise de Sistemas pela Pontifícia Universidade Católica do Rio Grande do Sul (1979).

Universidade Federal do Ceará (UFC)

Centro Nacional de Processamento de Alto Desempenho do Ceará (CENAPAD-CE)

Contato: Rossana M. C. Andrade

Rossana M. C. Andrade é a fundadora do grupo de pesquisa em Redes, Engenharia de Software e Telecomunicações (GREAT) na UFC. Desde 2003, Rossana coordena o CENAPAD-CE. Há mais de 10 (dez) anos desenvolve P&D com empresas do setor de telecomunicações no desenvolvimento de aplicativos móveis. Coordenadora do MDCC da Ciência da Computação da UFC. É também Coordenadora Local da UFC no Projeto Internacional “Learning WhileMoving” (LWM). Ela possui doutorado em Computer Science pela University of Ottawa (2001), mestrado em Ciência da Computação pela Universidade Federal da Paraíba (1992) e graduação em Ciência da Computação pela Universidade Estadual do Ceará (1989). Rossana possui experiência em pesquisa, desenvolvimento e inovação nas áreas de Ciência da Computação e Teleinformática, atuando principalmente nos seguintes temas: engenharia de software (linhas de produto de software, padrões de software, frameworks e middlewares) e redes de computadores (computação móvel, computação ubíqua, grades computacionais, redes de sensores sem fio e segurança).

2.3. Refinamento do protótipo

As facilidades desenvolvidas ao longo da Fase 1 para a plataforma mc² e que serão oferecidas na Fase 2 como um serviço piloto são descritas abaixo. Junto às suas descrições acompanham também propostas de refinamento dessas facilidades no contexto da Fase 2 deste projeto de modo a tornar a plataforma mc² um produto utilizável como serviço piloto.

Suporte à customização de ambientes de SaaS para aplicações científicas implementadas segundo duas abordagens distintas: como tarefas computacionais isoladas ou como fluxos de tarefas.

- A abordagem de **tarefas computacionais isoladas** emprega o sistema PortEngin, detalhado na Seção 2.1. Ao longo da Fase 2 propõe-se atividades de manutenção corretiva, perfectiva e evolutiva desse sistema, em particular seu editor de portais (produto da Fase 1 deste projeto), visando torná-lo incrementalmente mais seguro, robusto, confiável e amigável.
- A abordagem de **fluxos de tarefas** emprega o sistema Galaxy, também detalhado na Seção 2.1. Ao longo da Fase 2 propõe-se atividades de manutenção corretiva e perfectiva desse sistema, em particular sua ponte de integração (produto da Fase 1 deste projeto) com o middleware CSGrid, visando torná-lo incrementalmente mais robusto, confiável e com bom desempenho.

Acesso a recursos computacionais variados no nível de IaaS. Inclui-se nessa facilidade acesso a infraestruturas de processamento de alto desempenho como as dos centros que compõem o SINAPAD (sendo que nesta Fase 2 contempla-se o uso somente dos centros das instituições participantes do projeto), e a infraestruturas de processamento de alta vazão como a da comunidade OurGrid e a do serviço Ciência em Sua Casa. Essa facilidade poderá sofrer manutenções evolutivas ao longo da Fase 2 em decorrência da necessidade de melhorias no algoritmo de escalonamento implementado no CSGrid em função da heterogeneidade dos recursos computacionais envolvidos, bem como das oportunidades de inclusão de outras infraestruturas ou do atendimento de demandas de usuários específicos da plataforma mc². Em particular, objetivando uma maior integração com outras iniciativas da RNP, pretende-se investigar:

1. O interfaceamento com projetos da RNP ligados a computação em nuvem, como CDC e JiT Clouds.
2. O interfaceamento com o serviço experimental CIPÓ de provisionamento dinâmico de circuitos.
3. A ampliação do suporte ferramental de projetos de experimentação por simulação/emulação, como os de Internet do Futuro (FIBRE), o que transforma a plataforma mc² em um serviço muito útil para a própria Divisão de Pesquisa e Desenvolvimento da RNP.

Armazenamento de dados de usuários no nível de IaaS, utilizando a infraestrutura de armazenamento massivo de dados do SINAPAD. Essa facilidade também poderá sofrer manutenções evolutivas ao longo da Fase 2 em decorrência das oportunidades de inclusão de outras infraestruturas, como a do projeto CDC.

Compartilhamento e publicação de resultados de experimentos. Esses mecanismos são oferecidos nativamente tanto no PortEngin como no Galaxy. No caso do compartilhamento, usuários registrados de um mesmo ambiente de SaaS podem compartilhar dados entre si de forma segura. No caso da publicação, os dados ficam disponíveis para qualquer pessoa com acesso à Internet, por meio de uma URL aberta (útil para divulgação de resultados em publicações científicas). Ao longo da Fase 2 propõe-se atividades de manutenção corretiva e perfectiva desses mecanismos visando torná-los continuamente mais seguros.

Controle da proveniência e reprodutibilidade de experimentos. Esses mecanismos são oferecidos nativamente tanto no PortEngin como no Galaxy. Eles permitem o registro dos dados de experimentos (p.ex. controle de versão dos arquivos de entrada e saída) bem como de meta-informação associada aos experimentos e seus dados (p.ex. parâmetros de execução, usuário fornecedor/gerador dos dados, data/hora da última atualização desses dados). Ao longo da Fase 2 propõe-se atividades de manutenção corretiva e perfectiva desses mecanismos visando torná-los continuamente mais robustos, confiáveis e amigáveis.

Em resumo, vislumbra-se que a transformação da plataforma mc² em um serviço piloto por meio das atividades de manutenção corretiva, perfectiva e evolutiva apontadas acima servirá para ampliar a inserção da RNP em iniciativas de e-ciência nacionais, aumentando substancialmente os benefícios que a RNP oferece a seus usuários, sobretudo aqueles envolvidos com atividades de pesquisa científica.

2.4. Ferramentas de suporte à operação

As principais atividades de operação, administração e gerenciamento do serviço proposto e as ferramentas que as oferecem na plataforma mc² são listadas abaixo:

- **Cadastramento de aplicações de e-ciência.** Essa atividade é conduzida no âmbito da interface de administração do CSGrid e não demanda alterações para sua implantação no serviço piloto proposto. As aplicações registradas podem ser usadas tanto para a confecção de portais de tarefas isoladas usando o editor de portais do sistema PortEngin, como para a combinação em fluxos de tarefas modelados com o sistema Galaxy. Isso confere grande capacidade de reuso à plataforma mc².
- **Cadastramento de portais desenvolvidos com o editor de portais do sistema PortEngin.** Essa atividade também é conduzida no âmbito da interface

de administração do CSGrid e também não demanda alterações para sua implantação no serviço piloto proposto.

- **Autenticação e autorização em múltiplos níveis.** No que se refere à autenticação, o LNCC participa do Programa de Gestão de Identidades 2012 da RNP com vistas à integração do sistema PortEngin e do middleware CSGrid ao serviço CAFé. Ao longo da Fase 2 pretende-se investigar a possibilidade de integração do restante da plataforma mc² (em particular o sistema Galaxy) ao serviço CAFé. No que se refere à autorização, foi desenvolvido na Fase 1 deste projeto um conjunto abrangente de perfis de permissão de acesso ao serviço, incluindo usuário convidado (acesso restrito aos serviços de SaaS sem login/senha, disponível somente no sistema PortEngin), usuário registrado (acesso pleno aos serviços de SaaS com login/senha), desenvolvedor (capacidade de customizar no nível de PaaS os serviços de um ambiente de SaaS específico), e administrador. No caso do perfil administrador, será estudada ao longo da Fase 2 a possibilidade de se definir diferentes níveis de administração no CSGrid que permitam uma operação menos centralizada da plataforma mc² tanto pela RNP como pelas instituições usuárias do serviço piloto proposto (p.ex. para gerenciamento de usuários, de ambientes de SaaS, e de infraestruturas de execução no nível de IaaS).

3. Cronograma

O cronograma de atividades da Fase 2 deste projeto é apresentado abaixo:

1. A atividade de multiplicação de desenvolvedores de aplicações de e-ciência aptos a operar a plataforma mc² nas instituições usuárias ocorrerá nos primeiros 2 meses da Fase 2 deste projeto, e envolverá o treinamento das equipes da UFBA, do CENAPAD-CE/UFC e do CESUP/UFRGS no uso dos sistemas PortEngin e Galaxy.
2. A divulgação da plataforma mc² bem como dos ambientes de SaaS configurados com essa plataforma junto aos usuários das instituições participantes ocorrerá de forma continuada ao longo de toda a Fase 2 deste projeto, e será feita pelas equipes locais dessas instituições.
3. As atividades de manutenção do CSGrid, PortEngin e Galaxy também ocorrerão de forma continuada ao longo de toda a Fase 2 deste projeto, sendo concentradas nas equipes do LNCC, da UFCG e da UFBA.
4. O estudo de outros sistemas gerenciadores de fluxos de tarefas (como o Taverna – <http://www.taverna.org.uk/>) será feito nos primeiros 2 meses da Fase 2 do projeto pelo equipe do LNCC. Caso esse estudo mostre a viabilidade de integração de um ou mais desses sistemas à plataforma mc² na Fase 2, um cronograma mais concreto de atividades relacionadas a essa integração será apresentado ao final desses 2 meses.

5. A integração com iniciativas da RNP ligadas a computação em nuvem, armazenamento e provisionamento dinâmico de circuitos demanda um estudo de viabilidade prévio, que será feito nos primeiros 4 meses da Fase 2 do projeto pelas equipes do LNCC, da UFCG e da UFBA. Caso essa integração mostre-se viável para a Fase 2, um cronograma mais concreto de atividades relacionadas a essa integração será apresentado ao final desses 4 meses.
6. A integração com a iniciativa da RNP ligada a autenticação federada (CAFe) consistirá em quatro etapas a serem executadas pela equipe do LNCC: estudo, desenvolvimento, implantação e testes, prevendo-se 1 mês de duração para a primeira etapa e 3 meses para as demais etapas.
7. O desenvolvimento de soluções para uma maior descentralização da administração da plataforma mc² consistirá em quatro etapas a serem executadas pela equipe do LNCC: estudo, configuração e testes, prevendo-se 1 mês de duração para a primeira etapa, 2 meses para a segunda etapa e 3 meses para a terceira etapa.

Atividade	01	02	03	04	05	06	07	08	09	10	11	12
1.	X	X										
2.	X	X	X	X	X	X	X	X	X	X	X	X
3.	X	X	X	X	X	X	X	X	X	X	X	X
4. (Estudo)	X	X										
5. (Estudo)	X	X	X	X								
6. (Estudo)			X									
6. (Desenvolvimento)				X	X	X						
6. (Implantação)							X	X	X			
6. (Testes)										X	X	X
7. (Estudo)				X								
7. (Configuração)					X	X						
7. (Testes)							X	X	X			