

PROJETO DE PESQUISA

Programa de Iniciação Científica e

Nome do Orientador e Coordenação (Pesquisador/Tecnologista/Pós-doc):
ERALDO SILVA JUNIOR

Nome do pesquisador ou tecnologista e Instituição de Pesquisa Externa: (Coorientador ou Colaborador externo, se houver):

NÃO APLICÁVEL

Nome do Supervisor e Coordenação: (Pesquisador/Tecnologista):
ERALDO SILVA JUNIOR - COHEP

Título do Projeto: Ambientes de Infraestrutura Computacional para apoio a pesquisa em HEP

Palavra-chave:
Computação Científica, Apoio Computacional à Pesquisa, Infraestrutura de TI

Área de conhecimento:
Ciência da Computação
Subárea: Sistemas de Computação

Pré-requisitos desejáveis (se houver):

- Lógica de programação (qualquer linguagem)
 - Inglês técnico para leitura de documentação

Pré-requisitos obrigatórios (se houver):

- Noções básicas de Linux e linha de comando;
 - Fundamentos de redes de computadores;
 - Curiosidade e disposição para aprender em ambiente colaborativo e multicultural.

Possibilidade de orientação remota:

Rio de Janeiro, 15 de Agosto de 2025

Introdução

O avanço da Física de Altas Energias, especialmente no estudo de decaimentos raros de hadrons contendo o quark b, exige não apenas detectores sofisticados, mas também uma infraestrutura computacional capaz de lidar com volumes massivos de dados e algoritmos cada vez mais complexos. Nesse contexto, a Coordenação de Física de Altas Energias

(COHEP) do Centro Brasileiro de Pesquisas Físicas (CBPF) tem consolidado uma arquitetura computacional estratégica, articulando tecnologias de computação em GRID, Cloud Científica do INCT e o desenvolvimento de um piloto para Farm HLT2, voltada à análise em tempo real de eventos do experimento LHCb dentro do DataGRID Javier Magnin.

Resumo

Este projeto multidisciplinar oferece ao aluno a oportunidade de atuar em uma das três frentes de pesquisa e desenvolvimento tecnológico de Computação Científica na COHEP: **Grid Computing, Cloud Computing ou Farm HLT2**. Cada trilha aborda um conjunto distinto de tecnologias e objetivos, permitindo ao aluno aprofundar-se em uma área específica de infraestrutura computacional aplicada à física experimental.

Objetivos

A proposta é acolher estudantes que desejam construir seu domínio técnico por meio de atividades orientadas, contato com especialistas e participação em ambientes computacionais avançados. Ainda que não possuam domínio técnico consolidado, o aluno terá contato com a trilha de ambientação e após isso poderá escolher uma das trilhas disponíveis e desenvolver habilidades específicas com suporte contínuo, orientação prática e acesso a uma infraestrutura real, promovendo uma formação progressiva e contextualizada. Em ambientes de alta complexidade, contribuindo para a evolução da colaboração internacional com o CERN.

Metodologia

Independentemente da trilha escolhida, o aluno terá contato com fundamentos essenciais da área de infraestrutura e operações, incluindo: administração de sistemas operacionais Linux, conceitos de redes de computadores, práticas de DevOps, observabilidade de sistemas, infraestrutura como código (IaC) e gestão de ambientes computacionais. Essas competências serão desenvolvidas de forma transversal ao longo do projeto, garantindo uma formação técnica sólida e aplicável em múltiplos contextos.

Trilha de nivelamento: (Obrigatória)

O aluno será introduzido à operação de ambientes computacionais com uma abordagem sistêmica, explorando Linux, redes e segurança. Terá contato com os serviços de computação científica da COHEP, automação de infraestrutura com IaC e práticas de monitoramento e observabilidade de sistemas.

Trilha 1: Grid Computing

Objetivo: Ter contato com as atividades de administração da infraestrutura do Site de GRID, utilizando ferramentas como HTCondor para submissão de jobs e análise de desempenho. Conhecer o middleware e demais componentes que integram o WorldWide LHC Computing Grid (WLCG).

Trilha 2: Cloud Computing

Objetivo: Explorar ambientes virtualizados e escaláveis, com o uso de técnicas de técnicas de gerenciamento de recursos em nuvem privada usando Openstack e JupyterHub.

Trilha 3: Farm High Level Trigger 2 (HLT2)

Objetivo: Participar do desenvolvimento e testes da infraestrutura dedicada ao Piloto da Farm HLT2, parte do experimento LHCb, com foco em automação, redes e provisionamento remoto de servidores.

Cronograma Proposto

Meses 1 a 3 – Trilha Comum (Fundamentos Técnicos)

Mês 1: Linux, redes básicas, segurança da informação, estrutura da COHEP.

Mês 2: Containers, automação com shell script, monitoramento básico.

Mês 3: Infraestrutura como código (Ansible/Terraform), Git, escolha da trilha.

Trilha Grid Computing	Trilha Cloud Computing	Trilha Farm HLT2
<p>Mês 4: Introdução ao Grid e submissão de jobs</p> <p>Mês 5: WLCG Middleware e configuração de ambiente</p> <p>Mês 6: Autenticação e segurança (certificados, proxies)</p> <p>Mês 7: Monitoramento de jobs e troubleshooting via HTCondor</p> <p>Mês 8: Accounting (APEL)</p> <p>Mês 9: Integração com sistemas da COHEP</p> <p>Mês 10: Testes avançados e análise de desempenho</p> <p>Mês 11: Documentação técnica e refinamento dos resultados</p> <p>Mês 12: Apresentação final e avaliação técnica</p>	<p>Mês 4: Fundamentos de cloud pública e privada</p> <p>Mês 5: Componentes do Openstack</p> <p>Mês 6: Administração de usuários e grupos</p> <p>Mês 7: Monitoramento, segurança e controle</p> <p>Mês 8: Integração com serviços da COHEP e INCT-CERN</p> <p>Mês 9: Técnicas de escalabilidade e alta disponibilidade</p> <p>Mês 10: Ambientes de análise de dados avançados JupyterHub e Focus</p> <p>Mês 11: Documentação técnica e refinamento dos resultados</p> <p>Mês 12: Apresentação final e avaliação técnica</p>	<p>Mês 4: Ambientação no ambiente HLT: arquitetura, requisitos e dependências</p> <p>Mês 5: Configuração de acesso remoto via VPN e validação de conectividade segura</p> <p>Mês 6: Provisionamento de máquinas físicas via PXE boot e gerenciamento de imagens</p> <p>Mês 7: Instalação automatizada de sistemas operacionais e serviços</p> <p>Mês 8: Integração com sistemas de monitoramento e coleta de métricas</p> <p>Mês 9: Testes de desempenho e validação de ambientes bare metal</p> <p>Mês 10: Suporte à execução de jobs e workflows de benchmark do HLT</p> <p>Mês 11: Troubleshooting, análise de logs e otimização de recursos</p> <p>Mês 12: Documentação final, apresentação técnica e proposta de melhorias para o ambiente</p>

Conclusão

Este projeto vai além de uma formação técnica: é sua porta de entrada para uma iniciativa científica de vanguarda e de estado da arte, com impacto internacional. Durante 12 meses, você terá contato direto com um ambiente real de infraestrutura computacional para física de altas energias, em conexão com o CBPF e o CERN, vivenciando desafios e práticas que movem a pesquisa de ponta.

Essa experiência permitirá participar de redes globais e conhecer, na prática, o funcionamento da computação distribuída e de alto desempenho, contribuindo para o avanço da infraestrutura científica mundial.

Bibliografia

1. LHCb Collaboration. Framework TDR for the LHCb Upgrade II. CERN-LHCC-2023-005, Geneva: CERN, 2023.
2. LHCb Collaboration. The HLT2 Topological Lines. CERN-LHCb-PUB-2011-002, Geneva: CERN, 2011.
3. NEMETH, E.; SNYDER, G.; HEIN, T.; WHALEY, B. UNIX and Linux System Administration Handbook. 5. ed. Boston: Addison-Wesley, 2017.
4. KIM, G.; HUMBLE, J.; DEBOIS, P.; WILLIS, J. The DevOps Handbook: How to Create World-Class Agility, Reliability, & Security in Technology Organizations. 2. ed. Portland: IT Revolution Press, 2021.
5. TANENBAUM, A. S.; WETHERALL, D. J. Redes de Computadores. 6. ed. São Paulo: Pearson, 2021.