3º Workshop Brasil 6G

AVANÇOS DA PESQUISA RUMO AO 6G: CONECTIVIDADE DO FUTURO EM CONSTRUÇÃO

Apoio:

INF UFG Universidade Federal do Ceará UFSC UNICAMP UNISINOS cpqd

Realização:

RNP Inatel Ministério da Ciência, Tecnologia e Inovação Governo Federal Brasil União e Reconstrução

3º Workshop Brasil 6G

Atividade 4.5 - Implantação e disponibilização de uma infraestrutura de HPC para os pesquisadores do projeto

Bruno Ciro do Nascimento

RNP

Atividade 4.5

“Implantação e disponibilização de uma infraestrutura de HPC para os pesquisadores do projeto”

- Implantar uma infraestrutura centralizada de computação de alto desempenho;
- Otimizada por GPUs e dotada de ferramentas de inteligência artificial;
- Projetada para permitir o uso remoto e simultâneo pelos pesquisadores das diferentes instituições envolvidas no projeto.

Pesquisa

“Coleta de informações referente a utilização de recursos computacionais”

https://forms.office.com/r/nFrGGW0iXQ

Implantação e disponibilização de uma infraestrutura de HPC

- **Aplicações e finalidades de uso**
- **Requisitos e especificação técnica da solução**
- **Soluções de mercado e escolha da plataforma**
- **Camada de software**
- **Requisitos de instalação e locais candidatos**
- **Elaboração do Termo de Referência para aquisição**
- **Próximos passos**

Aplicações e finalidades de uso

Será um pilar fundamental que permitirá a realização de simulações complexas, o treinamento de modelos de IA avançados e o processamento de grandes volumes de dados.

- **Aplicação de IA e Machine Learning (ML) nas camadas física e de rede**, com ênfase no uso de GPUs para o treinamento de redes neurais capazes de otimizar o desempenho da rede em tempo real
- **Processamento de sinais em larga escala**, uma vez que a tecnologia 6G exigirá sistemas de antenas Massive MIMO e algoritmos de beamforming extremamente complexos para direcionar o sinal com alta precisão aos usuários
- **Simulação de ambientes complexos e gêmeos digitais** (digital twins)

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- Camada de software
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Requisitos da solução

A solução computacional foi desenhada de maneira a atender tanto as necessidades computacionais atuais quanto as futuras:

- **Executar modelos avançados de IA**, incluindo arquiteturas generativas que podem ser aplicadas na otimização de topologias de rede
- **Interoperar em um ambiente federado e distribuído**, podendo se conectar e colaborar com futuros novos nós, permitindo a execução de projetos conjuntos e o compartilhamento de dados e modelos em larga escala

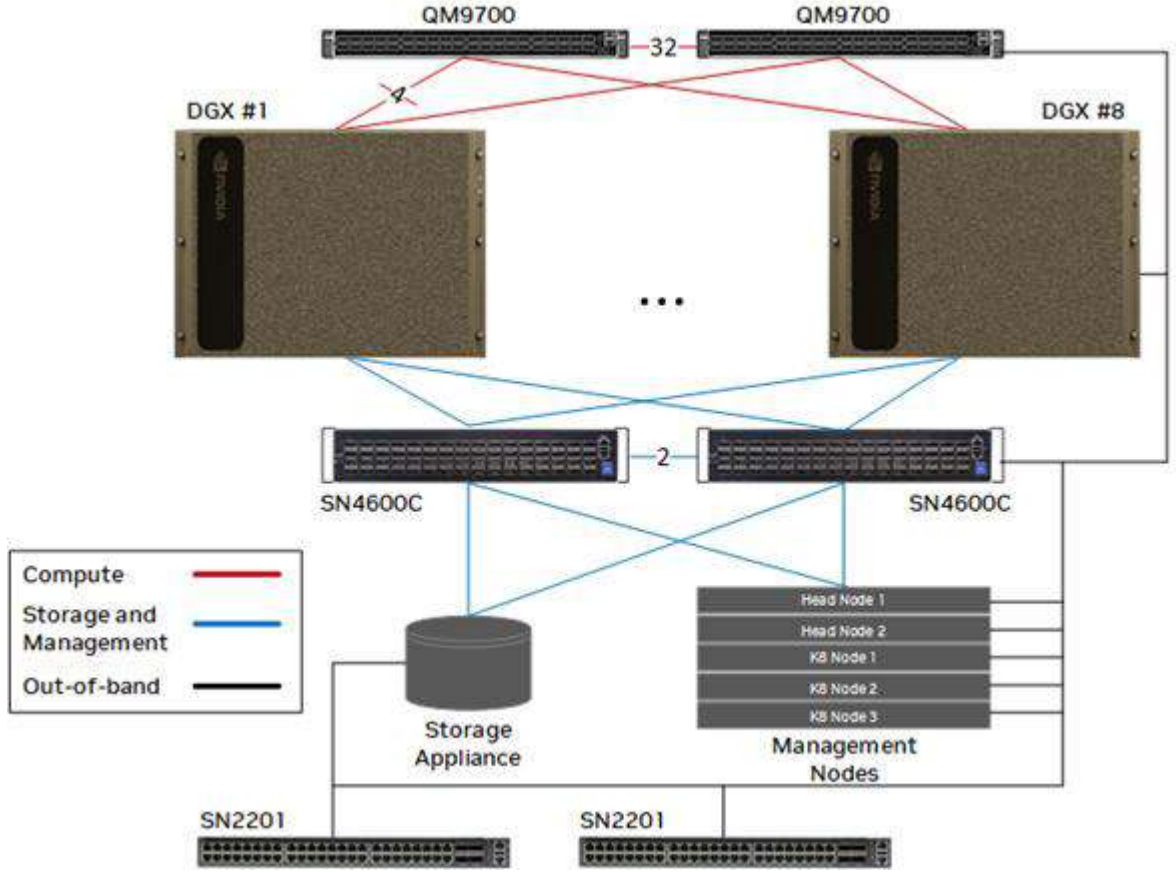
Requisitos da solução

A solução computacional foi desenhada de maneira a atender tanto as necessidades computacionais atuais quanto as futuras:

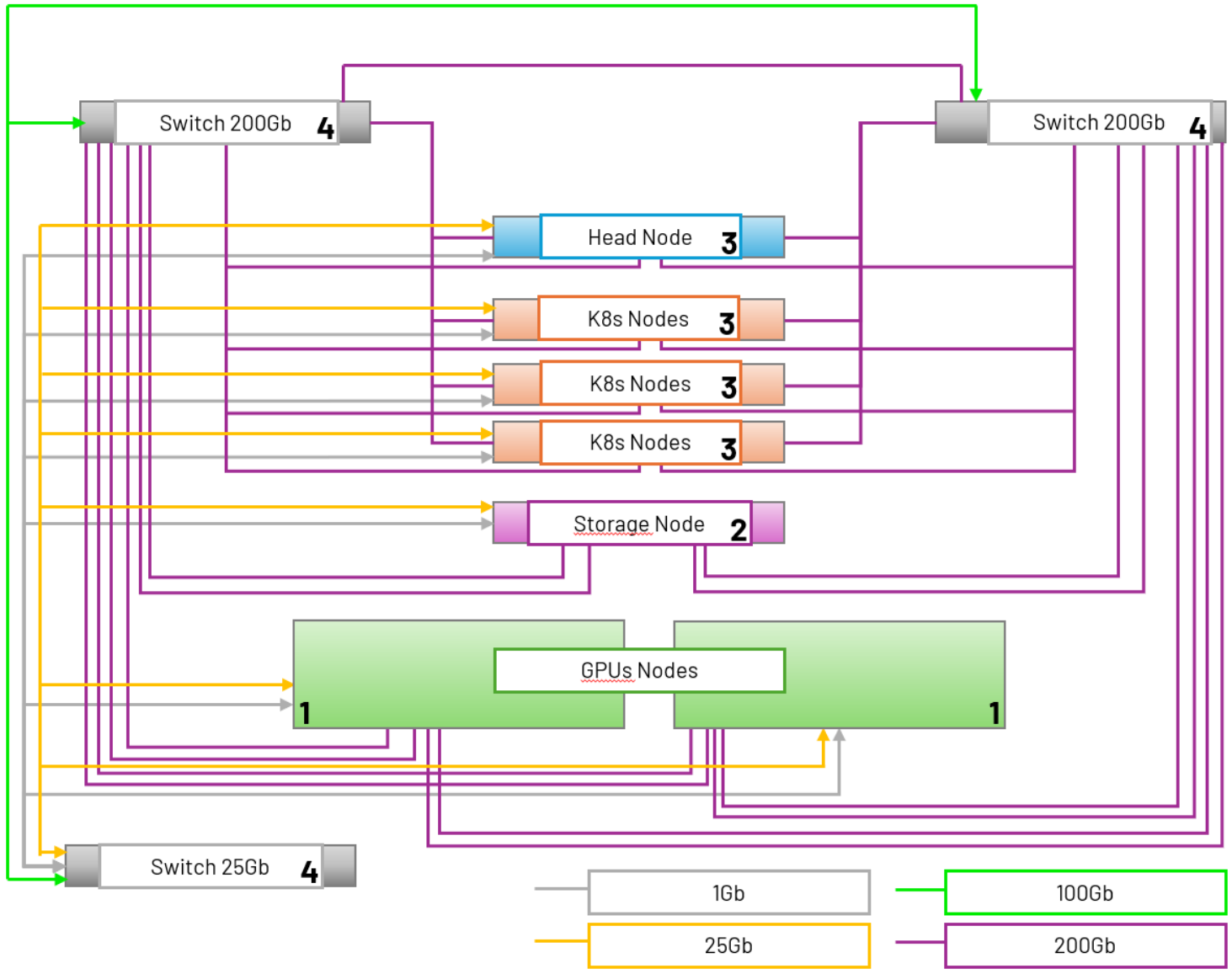
- **Permitir o gerenciamento dinâmico de GPUs virtuais por meio de tecnologias de fatiamento de recursos**. Assegurando que múltiplos pesquisadores possam compartilhar o hardware de forma segura
- Todos os **dados gerados deverão ser armazenados de forma centralizada e segura no ambiente de armazenamento de dados da solução**, garantindo acesso, integridade e persistência

Especificação técnica da solução

O desenho da solução se baseou na arquitetura de referência da NVIDIA, chamada de BasePOD:



Especificação técnica da solução



1. **Dois GPUs Nodes, com 8 GPUs NVIDIA H200 (141 GB) cada**
2. **Sistema de Armazenamento paralelo, 368 TB totalmente SSD, expansível até pelo menos 1 PB**
3. **Um Head Node e três Kubernetes Nodes**
4. **Dois switches de 200 Gbps, para comunicação interna, um switch de 25 Gbps para comunicação externa**

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- Camada de software
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Soluções de mercado

Segundo dados de 2024 da consultoria IoT Analytics*, a distribuição de participação de mercado no segmento de GPUs para datacenters é dada por:

- 92% NVIDIA
- 4% AMD com
- 4% e Intel e outras

Esses números evidenciam a **liderança expressiva da NVIDIA**, conquistada não apenas pelo desempenho de seu hardware, mas também pela **maturidade e consolidação de seu ecossistema** de software.

Escolha da plataforma

Após análise comparativa das soluções de GPU dos três principais players do mercado, optamos pelas soluções da NVIDIA, seguindo critérios técnicos objetivos:

- Maturidade do ecossistema de software
- Compatibilidade consolidada com os principais frameworks de IA
- Ampla disponibilidade de bibliotecas de alto desempenho pré otimizadas para HPC e IA
- Escalabilidade comprovada em arquiteturas multi-GPU e multi-nó, com interconexão GPU-GPU
- Redução do risco tecnológico em função da ampla adoção de mercado e suporte de comunidade

Escolha da plataforma

Outro fator determinante para escolha da NVIDIA foi a disponibilização de um SDK (Software Development Kit) exclusivo para pesquisas 5G/6G.

- Chamado de NVIDIA Aerial, esse SDK conta com duas importantes ferramentas:
  - **NVIDIA Aerial Omniverse Digital Twin**: plataforma de simulação de larga escala que permite criar gêmeos digitais de ambientes inteiros, desde uma simples torre até uma cidade
  - **NVIDIA Sionna**: biblioteca de código aberto, construída sobre o TensorFlow, focada especificamente na simulação de sistemas de comunicação, permitindo não apenas prototipagem rápida de algoritmos, mas também otimização de ponta a ponta com IA

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- **Camada de software**
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Camada de software

Para garantir um ‘quick start’ da solução após a entrega e instalação, a aquisição do pacote de software NVIDIA AI Enterprise foi uma premissa, garantindo assim:

- Gerenciamento e monitoramento do ambiente
- Provisionamento de containers pré otimizados
- Escalonamento de tarefas entre GPUs e Nós
- Acesso a suporte especializado

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- Camada de software
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Requisitos de instalação e locais candidatos

Os principais requisitos para instalação da infraestrutura são:

- **Conexão de alta capacidade e baixa latência com a Rede Ipê.** Rede acadêmica da RNP que interliga universidades, institutos de pesquisa, hospitais de ensino e outras instituições em todo o território nacional, além de prover conexão com redes acadêmicas internacionais
- Devido à alta densidade computacional dos servidores com GPUs, a infraestrutura de instalação deverá atender a rigorosos requisitos de ambiente físico e capacidade elétrica, visando garantir a operação contínua e segura do sistema

Requisitos de instalação e locais candidatos

A RNP tem como parte de sua infraestrutura os CNDs (Centros Nacionais de Dados), que estão localizados dentro de datacenters parceiros:

- 2 CNDs já estão em operação: São Paulo e Brasília
- Os equipamentos serão instalados no CND Brasília

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- Camada de software
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Elaboração do Termo de Referência para aquisição

Em conformidade com sua política de governança e transparência, a RNP exige (para aquisições de alto valor) a publicação de um documento chamado Termo de Referência, que estabelece de forma objetiva:

- Requisitos técnicos de hardware e software
- Níveis mínimos de serviço para garantia e suporte
- Requisitos de capacitação técnica
- Condições de fornecimento

Elaboração do Termo de Referência para aquisição

A publicação deste documento fornece assim condições de concorrência igualitárias para todo e qualquer proponente que atenda o conjunto de requisitos definidos no Termo de Referência.

Você pode acessar o Termo de Referência que foi publicado em 17/03/25 através do link:

https://plataforma.rnp.br/documentos/fornecedores/2025/adc/14121/2025-aquisicao-de-hpc-projeto-brasil-6g

Implantação e disponibilização de uma infraestrutura de HPC

- Aplicações e finalidades de uso
- Requisitos e especificação técnica da solução
- Soluções de mercado e escolha da plataforma
- Camada de software
- Requisitos de instalação e locais candidatos
- Elaboração do Termo de Referência para aquisição
- Próximos passos

Próximos passos

Em andamento

- Investigação das melhores práticas em infraestruturas de GPU compartilhadas para definir como o recurso computacional será disponibilizado aos usuários finais
- Criação do Plano de Governança e Sustentabilidade, documento que especificará as políticas de acesso, métodos de escalonamento de tarefas e as regras de compartilhamento dos recursos

Próximos passos

2025

- Solicitação ao fornecedor de um cronograma de entrega detalhado para todos os componentes da solução
- Elaboração do plano de instalação detalhado, em um esforço conjunto entre as diversas áreas da RNP, e validação desse planejamento com o fornecedor

Próximos passos

2026

- Acompanhamento da instalação física e da integração lógica dos equipamentos. Essas atividades serão realizadas pelo fornecedor
- Realização dos testes de aceitação (físicos e virtuais) para validar se a solução entregue está em conformidade com todos os requisitos do TR

Próximos passos

2026

- Execução do treinamento da equipe que atuará diretamente na administração e sustentação da nova infraestrutura
- Promoção dos treinamentos iniciais com os pesquisadores do projeto Brasil 6G e, na sequência, conduzir um período de operação assistida com um grupo piloto para refinar os processos e coletar feedback

3º Workshop Brasil 6G

Obrigado!

Bruno Ciro do Nascimento

bruno.nascimento@rnp.br