**RELATÓRIO DE DESEMPENHO TÉCNICO: PROJETO DE COMPUTAÇÃO DE ALTO DESEMPENHO (HPC) PARA PROCESSAMENTO PARALELIZADO DE IMAGENS MÉDICAS DICOM**

**Este documento detalha a concepção, o desenvolvimento e a avaliação abrangente de um pipeline de processamento paralelo projetado especificamente para imagens médicas no formato DICOM. A implementação foi conduzida na infraestrutura de Computação de Alto Desempenho (HPC) do supercomputador Santos Dumont. Esta iniciativa enfrenta o desafio significativo representado pelo processamento em escala massiva de exames médicos, utilizando técnicas avançadas de paralelismo, como MPI e OpenMP. Os resultados destacam melhorias substanciais de desempenho, ao mesmo tempo que aderem rigorosamente aos protocolos de segurança e à anonimização essencial de dados sensíveis dos pacientes.**

**1. Contexto e Justificativa**

**1.1. O Desafio Central**

**A proliferação de imagens médicas digitais cresceu exponencialmente na última década, com instituições de saúde gerando terabytes de dados no padrão DICOM (Digital Imaging and Communications in Medicine) mensalmente. Consequentemente, os métodos de processamento sequencial tradicionais tornaram-se inadequados para lidar com operações de grande escala, que incluem tarefas críticas como:**

* **Anonimização de Dados: Essencial para a conformidade com a Lei Geral de Proteção de Dados (LGPD).**
* **Compressão de Imagens: Fundamental para a otimização dos recursos de armazenamento.**
* **Cálculo de Estatísticas: Necessário para análises de garantia de qualidade.**
* **Conversão de Formatos: Crucial para garantir a interoperabilidade entre sistemas.**

**1.2. Relevância Estratégica**

**Para uma organização como a TechMed Solutions, que lida com o processamento mensal de mais de 100.000 exames, a necessidade de soluções escaláveis e robustas é primordial. Os principais requisitos para tal sistema são:**

* **Reduzir drasticamente os prazos de processamento, de dias para meras horas.**
* **Garantir a anonimização consistente e confiável de informações sensíveis.**
* **Otimizar a alocação e o uso de recursos computacionais.**
* **Facilitar a análise de dados quase em tempo real, permitindo insights mais rápidos.**

**1.3. Objetivos do Projeto**

**Os objetivos principais estabelecidos para este projeto foram:**

* **Implementar um pipeline paralelo totalmente funcional para o processamento de imagens DICOM.**
* **Avaliar a escalabilidade deste pipeline no ambiente HPC do Santos Dumont.**
* **Realizar uma análise comparativa das abordagens de paralelização MPI versus OpenMP.**
* **Validar a reprodutibilidade dos resultados e a eficiência geral do sistema.**

**2. Metodologia e Arquitetura de Sistema**

**O design do sistema baseia-se em uma estratégia híbrida que combina sinergicamente dois modelos distintos de paralelismo:**

* **MPI (Message Passing Interface): Utilizado para a tarefa de alto nível de distribuir arquivos de imagem entre os vários nós de computação do cluster.**
* **OpenMP (Open Multi-Processing): Empregado para o paralelismo de granularidade fina, dentro de cada nó, visando especificamente operações de filtro e processamento em nível de pixel.**

**O fluxo de trabalho de processamento de dados implementado consiste nas seguintes etapas sequenciais, mas internamente paralelizadas:**

1. **Leitura Distribuída de Arquivos: Os arquivos são lidos de forma distribuída usando MPI.**
2. **Anonimização de Metadados: Metadados sensíveis são sistematicamente removidos ou ofuscados.**
3. **Compressão de Imagem: As imagens são comprimidas usando o algoritmo JPEG com perdas (*lossy*).**
4. **Cálculo Estatístico: Métricas-chave (mínimo, máximo, média, desvio padrão) são computadas.**
5. **Agregação de Resultados: Os resultados finais de todos os processos são agregados usando operações MPI Reduce.**

**3. Análise de Resultados e Desempenho**

**A avaliação de desempenho revelou ganhos significativos com a implementação paralela. A abordagem híbrida**

**MPI+OpenMP alcançou o melhor desempenho, reduzindo o tempo médio de execução de 45.2 segundos (serial) para apenas 5.9 segundos, o que representa uma aceleração (*speedup*) de 7.66 vezes.**

**A taxa de transferência (**

***throughput*) de processamento aumentou de 4.4 MB/s na versão serial para 31.7 MB/s com 8 processos MPI. A eficiência paralela permaneceu alta, começando em 98% para 2 processos e caindo para 90% com 8 processos, uma degradação atribuída principalmente à sobrecarga de comunicação.**

**4. Limitações e Desafios Identificados**

**Apesar dos resultados positivos, várias limitações e desafios foram identificados.**

* **Limitações Técnicas: A escalabilidade do sistema apresentou ganhos marginais acima de 16 processos, e as operações concorrentes de E/S (leitura e escrita) representaram um potencial gargalo para o sistema de arquivos. O uso de dados DICOM sintéticos pode não capturar completamente a complexidade e a variabilidade das imagens clínicas do mundo real.**
* **Desafios de Implementação: A estratégia inicial de distribuição de arquivos *round-robin* mostrou-se subótima para arquivos de tamanhos variados, indicando a necessidade de uma abordagem de balanceamento de carga mais sofisticada. O pipeline atual carece de tolerância a falhas robusta, pois arquivos corrompidos ou falhas de nó podem interromper todo o processo.**
* **Restrições Ambientais: O projeto esteve sujeito às restrições operacionais do supercomputador Santos Dumont, incluindo janelas de execução limitadas e tempo de espera em filas.**

**5. Otimizações e Trabalhos Futuros**

**Para abordar as limitações identificadas e aprimorar as capacidades do sistema, diversas linhas de trabalho futuro são propostas.**

* **Otimizações Imediatas: As prioridades incluem a implementação de agregação de E/S por meio de leitura em lote para reduzir a carga no sistema de arquivos e a exploração do uso de aceleração por GPU via CUDA para tarefas computacionalmente intensivas, como a compressão JPEG, que se estima fornecer uma aceleração de 10 vezes.**
* **Expansões do Projeto: O desenvolvimento futuro poderia introduzir novas funcionalidades, como análise automatizada da qualidade da imagem e indexação de metadados para bancos de dados pesquisáveis37. Além disso, integrações planejadas com sistemas PACS hospitalares existentes e extensões para ambientes de nuvem híbrida ampliariam significativamente a aplicabilidade do projeto.**

**6. Conclusão**

**Em suma, este projeto demonstrou com sucesso a viabilidade e os benefícios substanciais da aplicação de técnicas de Computação de Alto Desempenho no processamento em larga escala de imagens médicas DICOM. A arquitetura híbrida**

**MPI+OpenMP provou ser excepcionalmente eficaz, alcançando uma aceleração notável de até 7.17x com oito processos, mantendo uma eficiência paralela superior a 90% na maioria das configurações.**

**Os resultados desta iniciativa não apenas fornecem uma solução escalável para o processamento DICOM, mas também validam o investimento estratégico em infraestrutura HPC para aplicações médicas avançadas, abrindo caminho para uma implantação completa em produção no supercomputador Santos Dumont.**