

UNIVERSIDADE FEDERAL DO PIAUÍ (UFPI-CSHNB)

DISCIPLINA DE SISTEMAS DISTRIBUÍDOS

PLANO DE PROJETO FINAL

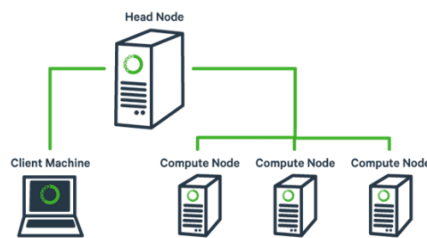
Prof. Francisco Airton Silva

(24 de outubro de 2018)

O projeto final da disciplina será a implementação de um **cluster de computadores**. O seu grupo deve implementar um sistema do zero (não podendo utilizar bibliotecas específicas para clusters disponíveis na internet).

Na computação de cluster, o hardware subjacente consiste em um conjunto de estações de trabalho ou PCs semelhantes, conectados por meio de uma rede local de alta velocidade. Além disso, cada nó executa o mesmo sistema operacional [Andrew Tanenbaum].

A imagem abaixo ilustra a arquitetura do sistema proposto envolvendo cinco máquinas. Basicamente o cliente irá passar uma requisição para o *head node (master)* que irá distribuir aos *computer nodes (slaves)* e ao final retornar o resultado ao cliente.



1. **Escalabilidade** – Deve ser fácil e prático a adição e remoção de máquinas do cluster
2. **Oferecer um Serviço de Conta Palavras** – A entrada do sistema é um conjunto x de palavras e um texto, o retorno deve ser quantas vezes cada palavra aparece no texto.
3. **Segurança** – Os dados trafegados no cluster devem passar por algum processo de criptografia
4. **Interface Gráfica** - O computador cliente e o computador master deve apresentar uma interface gráfica (GUI). No caso da GUI do máster deve ser possível visualizar graficamente os computadores que estão ativos e os computadores que não estão ativos.
5. **Aplicação do Método MapReduce** – O projeto deve aplicar o modelo MapReduce que é um modelo de programação desenhado para processar grandes volumes de dados em paralelo, dividindo o trabalho em um conjunto de tarefas independentes. Possui duas etapas, o map e o reduce. Novamente, você não pode utilizar plataformas prontas de mapReduce (como o Hadoop) deve apenas implementar a ideia com seu próprio código. Pense bem onde irá alocar cada função (quais máquinas).
6. **Análise de Desempenho** – O sistema deve mostrar de alguma forma se há ganho de desempenho quando se adiciona mais máquinas no cluster. A sugestão é medir e exibir o tempo total de processamento.

Datas Importantes

Apresentação Parcial: 06/11

Apresentação Final: 27/11