



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

PROGRAMA DE DISCIPLINA

ANO: 2020

DISCIPLINA: Processamento de Dados Massivos em Nuvem

CÓDIGO: **DCC831**

CLASSIFICAÇÃO:

CRÉDITOS: **04**

CARGA HORÁRIA: TEÓRICA: **060** horas

PRÁTICA: **000** horas

TOTAL: **060** horas

PRÉ-REQUISITO: Pela característica do curso, conhecimento anterior de processamento distribuído, como apresentado em AEDs 3 e/ou Sistemas Operacionais ou Fundamentos de Sistemas Paralelos e Distribuídos é pré-requisito. Isso normalmente significa ter cursado as disciplinas, mas em casos excepcionais, conforme avaliação do professor, isso pode ser liberado.

PERÍODO: 2º.

EMENTA: Os tópicos abordados pretendem cobrir os principais elementos para o desenvolvimento de novas aplicações no contexto de *Big-data*, com aplicação no ambiente em nuvem. Para esse fim serão discutidos princípios gerais de organização de sistemas distribuídos, organização das novas aplicações que estão surgindo na Internet e soluções para simplificar a tarefa dos desenvolvedores, bem como aspectos de sistemas operacionais avançados, como virtualização.

A - OBJETIVOS

O objetivo da disciplina é abordar tópicos avançados relacionados ao desenvolvimento de aplicações distribuídas para processamento de volumes massivos de dados. (Basicamente, o que costuma de chamar de *Big-Data*.)

Com o surgimento de diversas fontes que oferecem enormes volumes de dados, bem como outras que oferecem dados em tempo real, as demandas de processamento têm crescido de forma significativa. Ambientes de processamento como Hadoop se tornaram comuns e outros ambientes mais modernos ou mais específicos também surgem a cada dia.

Nesse contexto, esta disciplina pretende discutir os conceitos básicos relacionados com a área de processamento de dados massivos. O foco será no desenvolvimento de aplicações usando ambientes de processamento como Hadoop e Spark, bem como no desenvolvimento de soluções em termos de novos ambientes de processamento ou melhorias nos ambientes existentes.

B - PROGRAMA

- . Conceitos básicos de sistema sdistribuídos;
- . Princípios de processamento e armazenamento de volumes de dados massivos;
- . Definição de modelos de consistência;
- . Hadoop;
- . Spark;
- . Outros ambientes de processamento;
- . Aspectos gerais de computação virtualização
- . Computação em nuvem aplicada a dados massivos

C - AVALIAÇÃO

A primeira parte do curso será baseada em aulas expositivas apresentadas pelo professor. Ao final dessa fase (aproximadamente na metade do semestre) será realizada uma prova (com consulta, para fazer em casa) sobre os temas discutidos. Durante a segunda parte do curso os alunos deverão apresentar seminários preparados sobre temas correntes de pesquisa e artigos recentes. Todos os alunos deverão ler os artigos assinalados para cada aula e participar das discussões. Em paralelo, os alunos deverão desenvolver um trabalho com três entregas ao longo do semestre sobre um tema relacionado à disciplina. Uma lista de possíveis tópicos para trabalhos será apresentada posteriormente.

Distribuição dos pontos

Esta distribuição ainda está sujeita a alterações mas, a princípio, os pontos serão distribuídos da seguinte forma:

- **Prova:** uma, 15% da nota final
- **Leitura dos artigos:** pelo menos um por aula, normalmente dois, 25% da nota final
- **Seminários:** cada aluno deverá apresentar um seminário, 15% da nota final
- **Trabalho:** desenvolvido ao longo do semestre, 20% da nota
- **Participação em sala e no moodle:** aproximadamente 5% da nota final.

Projeto final

Esse projeto (a ser desenvolvido em grupo) deverá abordar temas relacionados aos assuntos discutidos em sala durante o curso. O formato preciso dependerá muito do tema escolhido. Simulações, implementações, desenvolvimento de modelos formais e avaliações de desempenho são todos possíveis. Em alguns casos, levantamentos bibliográficos também podem ser aceitos.

O trabalho constará de três entregas:

1. **Especificação:** proposta do que será implementado, objetivos, possíveis problemas, referências relevantes, resultados esperados, cronograma, critérios para aprovação do trabalho. 5% da nota.

2. **Andamento:** relatório de andamento do projeto, atualização do cronograma, discussão de problemas e resultados positivos. 10% da nota
3. **Relatório final:** relatório final, no formato de artigos da SBC, com aproximadamente 15 páginas, descrevendo os resultados alcançados em comparação com os critérios de aprovação propostos. 25% da nota

D - BIBLIOGRAFIA

Não há livro texto para a disciplina. Os artigos a serem lidos para cada aula, bem como outros artigos de apoio para temas específicos, serão publicados através do espaço do curso no minha.ufmg.