

3.7 Programa de Disciplina (Formulário SUPAC/UFBA)

UNIVERSIDADE FEDERAL DA BAHIA

SUPERINTENDÊNCIA ACADÊMICA PROGRAMA DE DISCIPLINAS SECRETARIA GERAL DOS
CURSOS**OBS: A carga horária básica deve ser de 17h ou múltiplo de 17h, conforme Res. 05/03 do CONSEPE.**

DISCIPLINAS

Código		Nome		Crédito	Assinatura do Chefe do Departamento	Ano
T	P	E	Total			
Banco de Dados II Processamento Massivo Paralelo (Hadoop e mapReduce)						
	34	0	34			

Ementa / Objetivos

Ementa:

Visão geral sobre Hadoop, Instalação do Apache Hadoop, Hadoop distributed File System – HDFS, MapReduce,

Objetivos:

Levar aos alunos uma visão geral do uso do Hadoop e seu ecossistemas de aplicações e sua implementação e uso na solução de problemas corporativos.

Metodologia

Todos os dias serão apresentados novos recursos e conceitos e avaliados através de exercícios práticos em todas as aulas.

Conteúdo Programático

1. Introdução
 - 1.1. Visão geral sobre Hadoop
 - 1.2. Características do Hadoop
 - 1.3. Sistema distribuído de arquivos.
 - 1.4. Ecossistema Hadoop
 - 1.5. Quem usa o Hadoop
 - 1.6. Classes do uso Hadoop
 - 1.7. Uso de Hardware comum.
 - 1.8. Distribuições do Hadoop
2. Instalação do Apache Hadoop
 - 2.1. Requisitos de Instalação
 - 2.2. Adquirindo os pacotes de Instalação
 - 2.3. Modo de Instalação (SingleCluster, DistribuidMode)
 - 2.4. Configuração do Ambiente de Rede
 - 2.5. Configurando Yarn.
 - 2.6. Criando diretórios físicos para o FileSystem
 - 2.7. Formatação do FileSystem
 - 2.8. Inicializando Serviços
 - 2.9. Iniciando o cluster com seus nós
 - 2.10. Testando Processos ativos
3. Hadoop distributed File System – HDFS
 - 3.1. Conceitual HDFS
 - 3.2. HDFS - Hadoop FileSystem
 - 3.3. HDFS - Arquitetura
 - 3.4. Comandos de manipulação do FileSystem
 - 3.5. Copiando arquivos para o FileSystem
 - 3.6. Listando arquivos no HDFS
 - 3.7. Criando e Removendo Diretórios
 - 3.8. Interface Web do HDFS
4. MapReduce
 - 4.1. Conceitual Map Reduce
 - 4.2. MapReduce X Hadoop
 - 4.3. MapReduce - Função Map
 - 4.4. MapReduce - Função Reduce
 - 4.5. Fluxo de Trabalho
 - 4.6. Executando um MapReduce
 - 4.7. MapReduce no cluster
 - 4.8. Configurando a IDE Eclipse para MapReduce
 - 4.9. Criando um novo MapReduce
 - 4.10. HDFS - MapReduce Data Flow

Bibliografia

Bibliografia básica:

1. V. Prajapati. Big data analytics with R and Hadoop. Packt publishing: BIRMINGHAM – MUMBAI. 2013.
2. L. Programmer. Big Data, MapReduce, Hadoop, and Spark with Python. 2016