

## Plano de Ensino – Atividades Remotas Emergenciais – 2021/1

Código	Tópicos de 60 horas
Disciplina	Processamento de dados massivos
Turma	A definir
Professor	Dorgival Olavo Guedes Neto
Horário	Segundas e quartas, 14:55-16:35
Público	BCC, BSI e PPGCC (isoladas serão consideradas)

### Proposta:

Abordar conceitos básicos relacionados ao desenvolvimento de aplicações distribuídas para processamento de volumes massivos de dados. (Basicamente, o que costuma se chamar Big-Data.)

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.

### Ementa:

Modelos de armazenamento e processamento distribuído para grandes volumes de dados. Programação em ambientes para processamento de dados massivos.

### Programa

Aula	Data		Conteúdo previsto	Modalidade	Interação
1	18-out.	Seg	Apresentação, leitura de artigos, perfil prof.	Síncrona Assíncrona	Reunião virtual Atendimento remoto (moodle)
2	20-out.	Qua	Intro MD, AM	assíncrona	Atendimento remoto (moodle)
3	25-out.	Seg	Intro MD, AM	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
4	27-out.	Qua	Lemonade	assíncrona	Atendimento remoto (moodle)
5	1-nov.	Seg	Desenvolvimento do exercício prático 1	assíncrona	Atendimento remoto (moodle)
6	3-nov.	Qua	Algoritmos de MD	assíncrona	Atendimento remoto (moodle)
7	8-nov.	Seg	GFS/HDFS, MapReduce/Hadoop	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
8	10-nov.	Qua	GFS/HDFS, MapReduce/Hadoop	assíncrona	Atendimento remoto (moodle)
9	15-nov.	Seg	Desenvolvimento do exercício prático 1	assíncrona	Atendimento remoto (moodle)

<b>Aula</b>	<b>Data</b>	<b>Conteúdo previsto</b>	<b>Modalidade</b>	<b>Interação</b>
10	17-nov.	Qua Soluções de armazenamento	assíncrona	Atendimento remoto (moodle)
11	22-nov.	Seg Introdução ao Spark	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
12	24-nov.	Qua Pares chave-valor e junções	assíncrona	Atendimento remoto (moodle)
13	29-nov.	Seg Pares chave-valor e junções	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
14	1-dez.	Qua Dataframes e datasets, bibliotecas	assíncrona	Atendimento remoto (moodle)
15	6-dez.	Seg Dataframes e datasets, bibliotecas	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
16	8-dez.	Qua Desenvolvimento do exercício prático 2	assíncrona	Atendimento remoto (moodle)
17	13-dez.	Seg Partições e shuffling; questões de desempenho	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
18	15-dez.	Qua Partições e shuffling; questões de desempenho	assíncrona	Atendimento remoto (moodle)
<b>Recesso de Natal e de Ano Novo.</b>				
19	3-jan.	Seg Ecossistemas de processamento de dados massivos	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
20	5-jan.	Qua Ecossistemas de processamento de dados massivos	assíncrona	Atendimento remoto (moodle)
21	10-jan.	Seg Processamento de consultas	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
22	12-jan.	Qua Processamento de consultas	assíncrona	Atendimento remoto (moodle)
23	17-jan.	Seg Processamento de grafos	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
24	19-jan.	Qua Processamento de grafos (incluindo Fractal)	assíncrona	Atendimento remoto (moodle)
25	24-jan.	Seg Processamento de streams	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
26	26-jan.	Qua Processamento de streams (incluir storm, kafka?)	assíncrona	Atendimento remoto (moodle)
27	31-jan.	Seg Sistemas de armazenamento (armazéns, datalakes)	Síncrona assíncrona	Reunião virtual Atendimento remoto (moodle)
28	2-fev.	Qua Sistemas de armazenamento (armazéns, datalakes)	assíncrona	Atendimento remoto (moodle)
29	7-fev	Seg Observações finais	Síncrona	Reunião virtual Atendimento remoto (moodle)
30	9-fev	Qua Apresentações do trabalho final	Síncrona	Reunião virtual Atendimento remoto (moodle)

O conteúdo será organizado por semana, com aulas presenciais apenas nas segundas-feiras que não forem feriados. A maior parte do conteúdo será fornecido em vídeos.

**Bibliografia**

- [Mining of Massive Datasets](#), por Jure Leskovec, Anand Rajaraman e Jeff Ullman - Cambridge University Press

- [Learning Spark](#), por Matei Zaharia, Patrick Wendell, Andy Konwinski e Holden Karau – O'Reilly

Os dois livros estão oficialmente disponíveis gratuitamente na internet

Artigos da literatura da área

**Material de apoio**

Moodle (slides e outros materiais)

Videos

Listas de discussão no moodle

**Avaliações**

	Descrição	Valor	Data
1	Resenhas/questões sobre artigos	40 pontos	A cada aula
2	Exercícios de programação: Lemonade	10 pontos	22/11/2021
3	Exercícios de programação	15 pontos	03/01/2022
4	Trabalho final	35 pontos	23/02/2022

Todas as avaliações serão remotas, com entregas pelo moodle. Pequenos ajustes na distribuição dos valores podem ocorrer durante o curso.

**Modo de interação:**

Toda comunicação escrita com os alunos será realizada via Moodle, inclusive a divulgação deste plano e a subsequente especificação de quais ferramentas serão utilizadas para as aulas remotas.

As aulas remotas serão transmitidas via Microsoft Teams e gravadas.