

CCC-So Processamento Massivo de Dados

PROJETO PRÁTICO

Profa. Sahudy Montenegro González

DESCRIÇÃO DO PROJETO

O projeto é de tema livre.

Os grupos podem escolher entre um projeto de tipo teórico ou prático.

Um projeto teórico pode ser um estudo ou estudo comparativo de algum(ns) tópico(s) relacionado(s) ao escopo da disciplina com pouca ou nenhuma parte de implementação.

Um projeto prático envolve justificativas teóricas para sua motivação e viabilidade, mas a maior parte dele é de implementação. Pode ser um protótipo de uma aplicação, um estudo de *benchmarking*, a reprodução de algum trabalho técnico/científico encontrado, dentre outros.

Requisitos mínimos para projetos práticos:

- Utilizar dois modelos de dados NoSQL diferentes para armazenamento **ou** um modelo + Apache *Spark*.
- Integração direta entre as tecnologias escolhidas: usar os conectores disponíveis.
- Interface não é necessária!

Fase Intermediária I – **Planejamento**: Definição da proposta

O intuito desta fase é formalizar a proposta e o planejamento do projeto prático e, para isto, precisa ser exposto como os requisitos mínimos serão atendidos (em casos de projetos práticos). Espera-se que nesta fase os grupos entreguem: quais são seus objetivos (O que vai ser feito? O que queremos atingir?) – **quanto mais detalhes, melhor!**

Quais as tecnologias escolhidas para a implementação? **Por que** as tecnologias escolhidas são as melhores para atender a sua aplicação? Caprichem nas justificativas!

O caso de projetos práticos, quais as **fontes** de dados que serão usadas?

Para ajudar a construir o minimundo, criem um **fluxograma** de como os dados vão transitar pela sua aplicação, das fontes até o resultado das consultas definidas.

Dica: os dados disponíveis nas fontes devem guiar o objetivo da aplicação escolhida.

Fontes de dados **sugeridas**, mas não obrigatórias:

- <https://www.kaggle.com/>
- <https://www.google.com/publicdata/directory>
- <https://ourworldindata.org/>

Fase Intermediária II - **Primeiros passos**

- Entrega das correções da fase anterior sugeridas pela professora.
- Projetos teóricos tem que ter a estrutura da proposta completa e pelo menos os tópicos introdutórios completos.

CCC-So Processamento Massivo de Dados

PROJETO PRÁTICO

Profa. Sahudy Montenegro González

- Projetos práticos ou teóricos com alguma implementação: precisam preparar o ambiente de implementação da prática. Nesta fase devem ser apresentadas as tecnologias já instaladas e configuradas no seu computador ou seu acesso e configuração em caso de uso de *Cloud*.

Façam os primeiros testes com a intenção de mostrar a conexão da aplicação com os bancos de dados, podendo ser testes de leitura ou escrita.

Não tem entrega, só apresentação informal na sala de aula!

Recomendações:

- Para Apache Spark (se não for usar *Databricks*), recomendação de instalação em qualquer um dos SOs: <https://sundog-education.com/spark-python/>.
Fazer leituras das fontes usando o PySpark
(leiam <https://spark.apache.org/docs/latest/sql-data-sources-load-save-functions.html>
e <https://towardsdatascience.com/pyspark-import-any-data-f2856cda45fd>).

Fase Final

Todos os membros do grupo devem ser capazes de mostrar o projeto funcional como um todo. O grupo deve mostrar que os objetivos foram atingidos, que a proposta planejada foi totalmente desenvolvida e que o fluxograma entregue na fase inicial de planejamento foi respeitado (no caso de projetos práticos).

RELATÓRIO

**O projeto precisa ser disponibilizado no GitHub
com nome (UFSCar-CC-So-PMD2025-<GrupoXX>).**

O relatório tem formato livre, porém deve ser respeitada a estruturação tradicional de um relatório: introdução, objetivos, fundamentação teórica, desenvolvimento e conclusões. Deve ser acrescentada uma seção de planejamento inicial (condizente com a fase intermediária I). Pode ser acrescentado um tópico de dificuldades encontradas (não necessariamente uma seção).

O relatório deverá ser entregue na primeira e última fase do projeto, com os tópicos que o grupo achar relevante de acordo com a fase. O relatório deve ficar contido, a princípio, no arquivo `readme.md` do projeto criado no GitHub e, quando pertinente, pode/deve destacar trechos de códigos a discutir e até *links* a arquivos específicos do projeto.

A palavra-chave do projeto é **POR QUÊ?** Encontrar as respostas que explicam os resultados enriquece a discussão sobre o tema e o nosso aprendizado!

CCC-So Processamento Massivo de Dados

PROJETO PRÁTICO

Profa. Sahudy Montenegro González

ENTREGAS

Cronograma do Projeto Prático

	Nota de execução	Entregas no AVA Moodle		Apresentação
		Quando?	O que entregar?	
Fases Intermediárias				
Formação dos grupos		30-abr.-25	Preencha o formulário no AVA	
Definição do tema e do projeto - I	100%	17-jun.-25	Relatório da fase intermediária (ler as instruções)	18-jun.-25
Primeiros passos e testes - II	0% a - 20% da final			2-jul.-25
Fase Final				
Correções da fase anterior	10%	15-jul.-25	- Relatório final	16-jul.-25
Elaboração do relatório	25%		- Código fonte no Github	
Desenvolvimento do projeto	65%		(todos os códigos necessários para executar o projeto)	
Projeto como um todo				

O cronograma poderá variar de acordo com o andamento das aulas!