

Título do Projeto

Análise de dados em Flink

Orientador: Paula Prata (pprata@di.ubi.pt)

Co-orientador: Maria Eugénia Ferrão (Departamento de Matemática)

Objetivos

Numa sociedade que produz cada vez maiores quantidades de dados, a análise desses dados de forma a deles retirar conhecimento útil é crucial. A análise de enormes quantidades de dados exige grande capacidade de processamento, sendo a programação paralela uma via óbvia para acelerar o processo. Pretende-se analisar os dados do Exame Nacional do Ensino Médio (ENEM) [1] do Brasil, com cerca de 8 milhões de registos, mostrando os resultados dessa análise graficamente. Por exemplo, mostrar o aproveitamento dos estudantes por disciplina, mostrar a distribuição dos estudantes por género ou por estado federal, etc. Para o processamento dos dados deve ser usado o Apache Flink, isto é, uma framework para programação paralela e distribuída de grandes quantidades de dados [2] [3].

Tarefas a Realizar e Cronologia

T1 - Estudo da framework Apache Flink. (1 mês);

T2 - Análise preliminar dos dados e especificação das análises de interesse. (0,5 mês);

T3 - Implementação (1 mês);

T4 - Testes e escrita de relatório (1 mês).

Requisitos Técnicos / Académicos

Gostar de programar;

Ter aprovação às disciplinas de programação.

Elementos de Avaliação a Entregar

- Relatório de projecto;

- Código desenvolvido.

Resultados Esperados

- Uma aplicação funcional;

- Um relatório de projeto.

Referências Bibliográficas

[1] http://inep.gov.br/en_US/web/guest/enem

[2] <https://ci.apache.org/projects/flink/flink-docs-stable/>

[3] <https://data-flair.training/blogs/flink-tutorial/>