

Sistemas de Armazenamento de Dados

Trabalho Prático

Universidade do Minho

2025

Informações gerais

- Cada grupo deve ser constituído por 3 elementos;
- O trabalho deve ser entregue até às 23:59 do dia 1 de Junho;
- Deve ser entregue o código fonte e um artigo de até 6 páginas de conteúdo no formato PDF seguindo o template ACM sample-sigconf <https://www.acm.org/publications/proceedings-template>
- O trabalho deve ser realizado tendo por base o sistema operativo Linux como ambiente de desenvolvimento e de execução;
- O trabalho deve ser submetido num arquivo Zip com nome grupo-xx.zip, em que xx deve ser substituído pelo número do grupo de trabalho (p.ex., grupo-01.zip);
- A apresentação do trabalho ocorrerá em data a anunciar, previsivelmente entre os dias 3 e 5 de Junho;
- Cada grupo pode preparar um conjunto pequeno de slides como material de suporte para a apresentação (detalhes relativamente ao tempo da apresentação a anunciar posteriormente);
- O trabalho representa 40% da classificação final.

Resumo

Pretende-se desenhar e implementar um sistema de ficheiros que dê suporte ao treino de modelos de aprendizagem profunda. A solução proposta deve ter em conta os requisitos de desempenho e resiliência deste tipo de cargas de trabalho. No final é esperado que este sistema de ficheiros seja avaliado experimentalmente, validando o desenho e potenciais funcionalidades propostas.

Requisitos Base (6 valores)

Interface e ambiente de desenvolvimento. O sistema de ficheiros deve ser implementado recorrendo à plataforma FUSE e deve suportar a interface POSIX, garantindo compatibilidade com aplicações como o PyTorch. A linguagem de programação a utilizar é da escolha de cada grupo.

Desenho e implementação. O sistema proposto deve seguir uma arquitetura distribuída composta por um cliente do sistema de ficheiros (i.e., instalado nas máquinas cliente) e por vários servidores que gerem os dados armazenados.

Aplicação e carga de trabalho alvo. O desenho, funcionalidades e implementação do sistema de ficheiros devem centrar-se em fornecer uma solução eficiente e confiável para o treino de modelos de aprendizagem profunda num único nodo (p.ex., com a plataforma PyTorch).

Tolerância a Falhas (4 valores)

A solução proposta deve garantir a disponibilidade dos dados e metadados armazenados aquando a ocorrência de falhas (p.ex, de um dos servidores). Cada grupo deve explorar e implementar as estratégias que pareçam mais apropriadas, tendo consciência dos pontos fortes e limitações das suas escolhas.

Desempenho (4 valores)

Cada grupo deve escolher implementar pelo menos uma otimização de desempenho que permita tornar o sistema de ficheiros mais eficiente no suporte ao treino de modelos de aprendizagem profunda. A otimização a implementar fica ao critério de cada grupo.

Avaliação (3 valores)

A solução desenvolvida deve ser avaliada considerando tanto cargas sintéticas como realistas (p.ex., treino de modelos de IA com o PyTorch). Cada grupo deve escolher quais as cargas de trabalho, métricas e experiências mais relevantes para validar o desempenho e resiliência da sua solução. São valorizadas experiências que demonstrem o impacto de diferentes otimizações/configurações da solução.

Artigo (3 valores)

O artigo final deve seguir o formato e estrutura de um artigo científico. Brevemente, o mesmo deve incluir:

- Um resumo breve do artigo.
- Uma secção de introdução com contexto do problema a ser abordado, com os desafios concretos de investigação a serem explorados, e com as principais contribuições e resultados do trabalho.
- Uma secção incluindo o desenho, arquitetura e algoritmos da solução proposta. Esta secção pode incluir ainda uma breve discussão da implementação do protótipo da solução.
- Uma secção de avaliação experimental com as questões a que a avaliação responde, com a metodologia de testes e com a apresentação e discussão de resultados.
- Uma secção de trabalho relacionado que não precisa de ser extensa mas deve referir possíveis soluções ou artigos que o grupo ache relevantes para o projeto em questão.
- Uma secção de conclusão que pode incluir pontos para trabalho futuro.

O tamanho máximo para o artigo é de 6 páginas. É possível a utilização de anexos para mostrar resultados extra que não tenham sido discutidos no artigo principal. Referências bibliográficas e anexos não contam para o limite de 6 páginas. O artigo deve ser escrito em português ou inglês.