

Computação na Cloud (11481)

Mestrado em Engenharia Informática

Ano Letivo de 2019/2020, 2º Semestre

Projeto Laboratorial 2 – Desempenho da plataforma Apache Flink

Organização dos Grupos de Estudantes: Trabalho a realizar por um grupo de, no máximo, 2 estudantes.

Identificação dos Estudantes: Indique, na capa do relatório, o título do trabalho (acima indicado) e os nomes e os números dos estudantes que integram o grupo que realizou o trabalho.

Cotação do Trabalho: Este trabalho contribui com 4,0 valores (20%) para a classificação final. A discussão deste trabalho, a agendar com o docente de cada turno, contribui com outros 4,0 valores para a classificação final do estudante expressa na escala de 0,0 a 20,0 valores.

Formato do Relatório: O relatório deve obedecer ao formato IEEE, conforme especificado na página de apoio a esta unidade curricular: <http://www.di.ubi.pt/~mario/tcc.htm>.

Aulas Práticas Para a Realização do Trabalho: O trabalho deve ser realizado nas aulas práticas até ao dia 5 de junho (inclusive) de 2020.

Data e Forma de Entrega do Trabalho: Um estudante de cada grupo deve enviar o relatório, num ficheiro em formato pdf, até 15 de junho de 2020. Este relatório deverá ser entregue na plataforma [Moodle](#), ou na eventualidade de não ter acesso, enviar por email para os endereços mario@di.ubi.pt (para os alunos do turno 1) e tiago.simoes@ubi.pt (para os alunos do turno 2) com o *subject*: CC 2019/2020 – Trabalho 2. Este trabalho é objecto de discussão oral online, devendo um estudante de cada grupo enviar a apresentação em formato powerpoint até 24 horas antes da discussão.

Objetivos

O objetivo central deste trabalho consiste em instalar e configurar o Apache Flink e comparar respetivo desempenho usando dois benchmarks diferentes.

Descrição

Neste trabalho pretende-se instalar e configurar o Apache Flink [1]. A arquitetura do Apache Flink inclui duas APIs, a DataSet API para processamento em batch e a DataStream API para processamento de fluxos (streams). Neste trabalho pretende-se

comparar o desempenho do Apache Flink para processamento de fluxos, usando dois benchmarks [2] diferentes, por exemplo, o Yahoo Streaming Benchmark [3] e o StreamBench [4], de modo a avaliar e comparar a capacidade de processamento (throughput) e a latência (latency).

Trabalhos análogos foram realizados em [5], [6], podendo ser usados para suporte à realização deste trabalho.

Organização do Relatório

O relatório deve ter, no máximo, 8 páginas no formato do IEEE e deve ser organizado de acordo com a seguinte estrutura: Abstract, I. Introdução, II. Arquitetura do Apache Flink, III. Instalação e Configuração do Flink, IV. Benchmarks Usados para Avaliação do Desempenho, V. Experiências e Resultados, VI. Conclusões, Referências. O relatório pode ser escrito em português ou inglês. O template, formato de artigo de conferências IEEE, está disponível na página de apoio a esta unidade curricular: <http://www.di.ubi.pt/~mario/tvd.htm>.

Referências

- [1] Apache Flink® — Stateful Computations over Data Streams, <https://flink.apache.org/>.
- [2] T. Ivanov et al., Big Data Benchmark Compendium. In: Nambiar R., Poess M. (eds) Performance Evaluation and Benchmarking: Traditional to Big Data to Internet of Things. TPCTC 2015. Lecture Notes in Computer Science, vol 9508. Springer, Cham, 2016, https://link.springer.com/chapter/10.1007/978-3-319-31409-9_9, <http://people.ac.upc.edu/npoggi/publications/SPEC-RGBD-TPCTC-2015-Big%20Data%20Benchmark%20Compendium.pdf>
- [3] Yahoo, Yahoo Streaming Benchmark, <https://github.com/yahoo/streaming-benchmarks>
- [4] Y. Wang, Stream Processing Systems Benchmark: StreamBench, Aalto University, Espoo, Finland, 2016, <https://aaltodoc.aalto.fi/handle/123456789/20991>
- [5] Elkhan Shahverdi, Comparative Evaluation for the Performance of Big Stream Processing Systems, Master's Thesis, University of Tartu, Tartu, Estonia, 2018, <https://pdfs.semanticscholar.org/4d28/87513b538812809d5c7336978a5a189009d9.pdf>
- [6] Isabel Soqui Bongo, Avaliação de Desempenho das Plataformas Apache Hadoop, Apache Spark e Apache Flink Usando o Benchmark Hibench-master, Dissertação de mestrado, Universidade da Beira Interior, Covilhã, Portugal, 2019, <https://ubibliorum.ubi.pt/handle/10400.6/10049>