

COMPUTAÇÃO NA NUVEM LEIRT / LEIC / LEIM

SEMESTRE DE VERÃO 2021-2022

José Simão jsimao@cc.isel.ipl.pt; jose.simao@isel.pt

Luís Assunção <u>lass@isel.ipl.pt</u>; <u>luis.assuncao@isel.pt</u>

Objetivos

- Compreender as características distintivas das arquiteturas e do desenvolvimento de software de sistemas distribuídos, as suas potencialidades e desafios;
- Saber utilizar diferentes ofertas de computação em nuvem, nomeadamente serviços de armazenamento e de processamento, tendo em conta aspetos não funcionais relacionados com disponibilidade, replicação e escalabilidade;
- Utilizar diferentes padrões de comunicação e interação, compreendendo as suas características de fiabilidade e desempenho;
- Compreender a necessidade de coordenar e sincronizar a tomada de decisões em ambientes distribuídos



Resultados de aprendizagem

- Compreender as características distintivas das arquiteturas de sistemas distribuídos e saber utilizar diferentes padrões de comunicação e interação entre as componentes dos sistemas distribuídos.
- Conhecer as potencialidades e desafios da computação na nuvem (Cloud Computing);
- Saber desenvolver aplicações utilizando as diferentes ofertas de computação em nuvem, nomeadamente serviços de: armazenamento de dados; de processamento e virtualização de recursos; comunicação por mensagens e eventos, nomeadamente usando o modelo *publish/subscribe*;
- A capacidade de compreender a necessidade de coordenar e sincronizar a tomada de decisões em ambientes distribuídos.



Avaliação dos resultados de aprendizagem

- Avaliação Prática (Trabalhos em grupo com avaliação individual):
 - Aulas práticas laboratoriais com atividades de consolidação de conhecimentos. <u>Para além da entrega de eventual código desenvolvido, os</u> <u>laboratórios serão avaliados com questionários, realizados no Moodle, com</u> <u>resposta individual por cada aluno;</u>
 - Realização, em grupo, de um trabalho de projeto final, com relatório PDF e apresentação final de demonstração da operacionalidade do trabalho;
- Avaliação Teórica (Avaliação individual em Exame Final):
- Classificação Final
 - (15%) Avaliação individual dos Laboratórios;
 - (35%) Trabalho de projeto (nota >= 10);
 - (50%) Exame final em época de exames (nota >=10)



Questionário individual sobre as atividades de laboratório

- O questionário pretende avaliar a consolidação de conhecimentos obtidos com a realização das atividades de laboratório:
 - 1. Realizado online no Moodle com duração de 10 minutos, numa das aulas seguintes à realização do laboratório;
 - 2. Através de questões de resposta múltipla, ou questões de resposta aberta, pretende-se avaliar:
 - Detalhes sobre as tecnologias e as ferramentas (tools) utilizadas;
 - Detalhes sobre as soluções encontradas para resolver os problemas do enunciado do laboratório;
 - Indicação se a solução final é executável e demonstrável
 - Que conclusões e lições foram aprendidas
 - Nas questões abertas poderá ser necessário submeter ficheiros com os artefactos de software desenvolvidos



Sobre o exame escrito

- O exame escrito tem duração de 2 horas com a seguinte estrutura:
 - 45 minutos [Sem Consulta] Questões sobre conceitos fundamentais e aspetos das tecnologias utilizadas;
 - 75 minutos [Com Consulta] Questões práticas de aplicação de conhecimento de aspetos relacionados com as tecnologias utilizadas.
 - Nota: Entende-se por [Com Consulta] a possibilidade de consultar o material em papel (livros, cópias dos slides; exemplos, etc.) que, individualmente, cada aluno leve para o exame, excluindo equipamentos eletrónicos.



Exame de época especial

- Na época especial, pode prestar provas de exame final todo o aluno que, estando inscrito na disciplina, reúna as condições previstas em diploma legal, e nos casos dos pontos 3.5.1 e 3.5.2 das Normas de avaliação de conhecimentos do ISEL
- O exame de época especial tem também uma componente teórica e uma componente prática



Conteúdos programáticos

- Conceitos fundamentais de sistemas distribuídos: modelos de comunicação e interação, latência, throughput, speedup, middleware e teorema CAP;
- Modelos de computação na nuvem. Serviços de Infraestrutura, Plataforma e Software. Flexibilidade na alocação a pedido de recursos e aspetos económicos sobre a utilização de serviços na nuvem;
- Acesso e organização das plataformas públicas de computação na nuvem, usando uma delas como caso de estudo;
- Introdução ao armazenamento distribuído de grandes volumes de dados.
 Repositórios de objetos binários, relacionais, e não relacionais.
- Virtualização de recursos computacionais. Diferentes tipos de isolamento e execução: máquinas virtuais de processo e de sistema, contentores;
- Modelos de comunicação publish/subscribe e padrões de utilização;
- Conceito de serverless computing: Cloud Functions



Concretização dos conteúdos programáticos

- Enquadramento histórico dos sistemas distribuídos e as plataformas de computação na nuvem
 - Potencialidades e Desafios: Recursos (computação, armazenamento, virtualização, modelo *pay-per-use*, centros de dados geo-distribuídos, energia, *vendor lock-in*, segurança ...
- Desenvolvimento de software distribuído e APIs de acesso programático às plataformas de computação na nuvem
 - Sockets, contratos, chamada remota de métodos, RPCs e Cloud APIs
- Armazenamento distribuído de Binary Large Objects blobs (Google Storage)
 - Organização em *buckets*
 - Distribuição e acesso aos objetos; partição por hash
- Modelos de dados n\u00e3o relacionais (Google Firestore)
 - Chave-valor; Documental; Colunas; Grafos
 - Partição e Replicação de dados, Consistência, Consensos
- Comunicação distribuída por mensagens/eventos
 - Modelo Publish/Subscribe (Google Pub/sub)
- Máquinas virtuais (Compute engine)
 - Máquinas virtuais de sistema, de processo e contentores;
 - Alocação dinâmica de recursos para escalabilidade horizontal (Google Instance-groups)



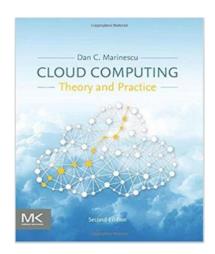
Tecnologias e *Tools*

- Plataforma Java (<u>JDK 11</u>)
- IntelliJ IDEA Community Edition 2020.3
- Build System (suportado pelo Intellij)
 - Maven (<u>https://maven.apache.org/</u>),
 - Maven in 5 minutes: https://maven.apache.org/guides/getting-started/maven-in-five-minutes.html
- Google Cloud Platform APIs e Tools
- Outros



Bibliografia / Recursos

- Capítulos selecionados de:
 - [1] Cloud Computing Theory and Practice, 2nd Edition, Dan Marinescu,
 Morgan Kaufmann, 2017
 - [2] Google Cloud Platform in Action, JJ Geewax, Mannung, 2018





- Artigos e Documentação online
- Slides e outros documentos de acompanhamento das aulas (Moodle)
- Créditos para acesso à Google Cloud Platform (GCP)



Computação Distribuída: Buzzwords!

Sockets TCP/IP

Client/Server applications

Distributed Objects

Java Remote Method Invocation

Remote Procedure Calls - gRPC

Web Services

Web applications

JSON Protobuf

XML

SOAP

ReST (Representational State Transfer)

Model View Control (MVC)

RESTful Service

Web Sockets

Message Oriented Middleware

Monolithic

Web API

Publish/Subscribe

Microservices

Relational DB

Message Broker

Containers

Virtual machines

NoSQL

Kafka

Docker

Stateless/Stateful

Choreography

Cloud services

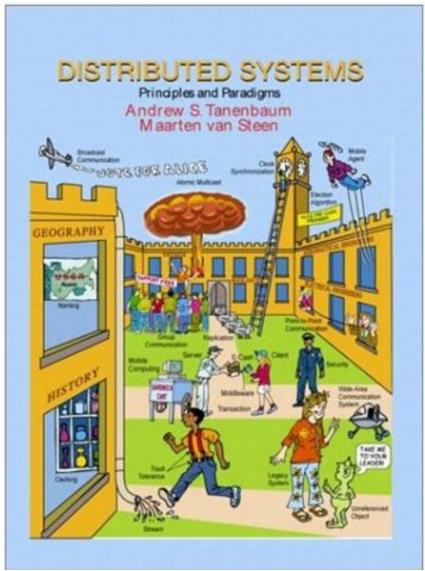
Orchestration

Kubernetes

Serverless



Sistemas distribuídos



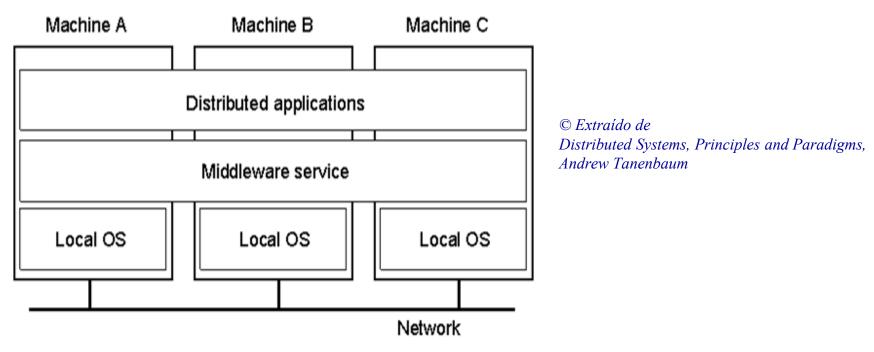
- Cloud Computing
- Grid Computing
- Mobile Computing



- Clock synchronization
- Election Algorithm
- Broadcast/Multicast communication
- Group Communication
- Caching
- Replication
- Peer-to-Peer Communication
- Fault Tolerance
- Transactions
- Security



Aplicações Distribuídas: Uma imagem dos 80/90



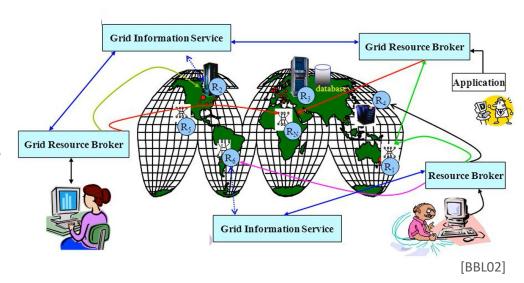
Middleware

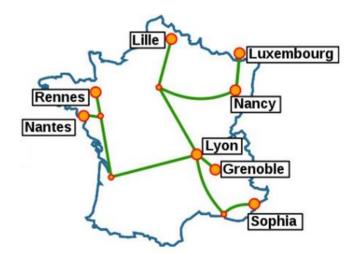
- Camada de software que proporciona um modelo computacional de programação, mascarando a heterogeneidade do hardware, sistemas operativos e linguagens de programação;
- Normalmente disponibiliza um conjunto de blocos (building blocks) ou APIs que facilitam o desenvolvimento de sistemas distribuídos.



O virar do século e a ideia de Grid Computing

«A computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to highend computational capabilities» [FK99]





Grid'5000, a large-scale testbed for distributed computing: 8 sites, 30 clusters, 840 nodes, 8490 cores, dedicated 10-Gbps backbone network

[Nussbaum18]

[FK99] I. Foster and C. Kesselman, editors. The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers, July 1998 [BBL02] Mark Baker, Rajkumar Buyya, and Domenico Laforenza, Grids and Grid technologies for wide-area distributed computing, Software – Practice and Experience, 2002

[Nussbaum18] Lucas Nussbaum, Grid 5000, https://www.silecs.net/wp-content/uploads/2018/04/nussbaum-grid5000.pdf

