Pontificia Universidad Católica de Chile
School of engineering
Computer Science Department

**IIC2523 Distributed Systems (II/2018) - Parallel 1**

**Teacher: Gabriel Vidal Salazar**

# Test 1

Date: September 24th, 2018

## Section 1 (10 pts): Questions

1. (2 pts) According to Flynn's taxonomy. What is the classification of modern processors? What parallelism does it take advantage of?
2. (2 pts) What is SPMD and MPMD? Mention an example for each of them.
3. (2 pts) Why is neccesary comunicate processes on distributed systems? What role does `MPI` play?
4. (2 pts) What technology are used by `IaaS` system providers (such as AWS, Google, etc.)? What characteristics does this technology have that allows it to serve different customers?
5. (2 pts) What is the main problem of the directory memory coherence mechanism?

## Section 2 (50 pts): Helping a friend

One of your friends is openning a new company, with 5 people, in Chile and wants to centralize the file storage. The files in the company are small and always are edited by one user (at this time). Your friend did a research and is convinced that `GFS` (`Google File System`) is a good idea for this distributed file system. He told you that `GFS` allows him to replicate information, the low cost and the growth opportunity. Also, he told you that `Lustre` is a good idea too, because the low cost, performance and growth opportunities.

For the next questions, the following characteristics are relevant: cost, performance, consistency, concurrence, security, scalability and replication.

1. (4 pts) Do you agree with your friend? Explain your answer for `GFS` and `Lustre`.
2. (4 pts) In this case. What is your recommendation? Why?
3. (4 pts) If, in the future, the company growth (with around 50 people). Does you recomendation change? Why?

4. (4 pts) If we consider a future growth of the company (50 people). Does your recommendation change? Why?

After a while, your friend contacts you because the company has grown a lot, and he is ready to start operating in Europe. Currently the distributed file system is working very well, but your friend wants to be connected with the new office.

4. (3 pts) Is a good idea to use a `VPN` for this case? Base your answer on speed and safety aspects. What type of configuration should be done?

5. (3 pts) And use the `TOR` network? Perform the same previous analysis.

Now, you are already part of your friend's company as a technical advisor and, with engineering team, decided to organize the main application with microservices architecture, like next image.

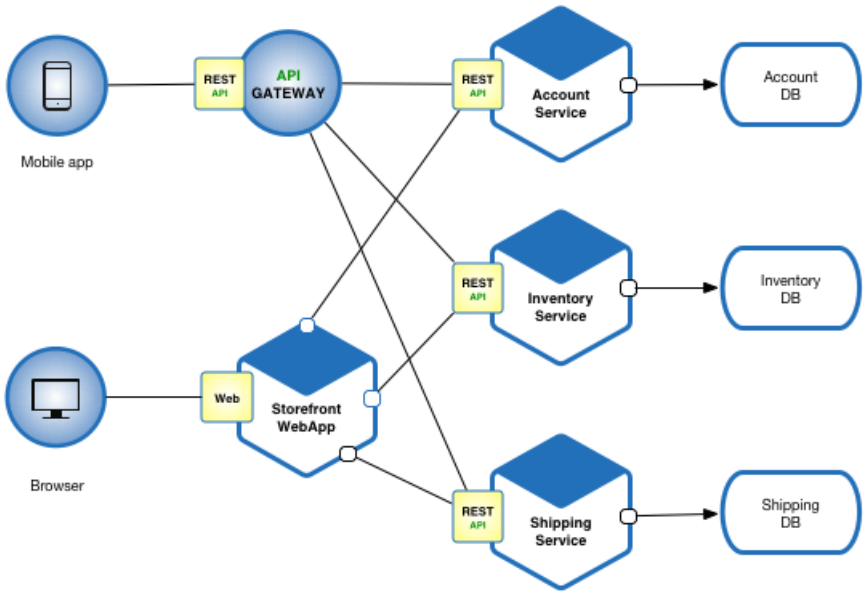


Figure 1: Microservices example (https://microservices.io/patterns/microservices.html)

6. (5 pts) Analyze the fulfillment of the `CAP` theorem for the architecture shown in the image. You can make an assumption in case you need it.

7. (2 pts) If we add replication of the databases of each microservice, How is the `CAP` theorem affected?

Currently the company has two physical servers and the engineering team is divided between virtualizing or installing the system directly on the servers. Finally, the main objective is to have availability of the service.

8. (5 pts) Is it possible to achieve availability with both options (described above)? How?

9. (5 pts) They have decided that virtualizing is the best way for the company. Now the discussion focuses on virtualizing based on containers or hypervisor. What is the biggest difference between the two? Which would you use if another objective is greater performance? Why?

A problem of the company is related to the purchase of supplies. Therefore, they have determined that it is a priority to be able to determine when it is required to buy products for sale. For this, it was decided:

- Study the demand of the last 12 months for each product
- Get the average demand
- If in the last 3 months it is above the average, that product should be purchased.
- The purchase is only made if more than a quarter of the products require that they be purchased.

For this you will have a text file that in the first column will have the product identifier and, in the next columns, the demand for the previous 12 months.

10. (10 pts) Write (in pseudocode) a program using `MPI` to do the explained above.
11. (5 pts) How does your program change if the purchase is made by each product considering the product's demand only?