University of Engineering and Technology
School of Computer Science
Syllabus of Course – Academic Period 2017-I

1. **Code and Name:** CS3P2. Cloud Computing
2. **Credits:** 3
3. **Hours of theory and Lab:** 1 HT; 4 HP;
4. **Professor(s)**
   Meetings after coordination with the professor
5. **Bibliography**

[Bal+08] Shumeet Baluja et al. "Video Suggestion and Discovery for Youtube: Taking Random Walks Through the View Graph". In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08. Beijing, China: ACM, 2008, pp. 895–904. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367618. URL: http://doi.acm.org/10.1145/1367497.1367618.

[BVS13] Rajkumar Buyya, Christian Vecchiola, and S. Thamarai Selvi. *Mastering Cloud Computing: Foundations and Applications Programming*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2013. ISBN: 9780124095397, 9780124114548.

[Cou+11] George Coulouris et al. *Distributed Systems: Concepts and Design*. 5th. USA: Addison-Wesley Publishing Company, 2011. ISBN: 0132143011, 9780132143011.

[HDF11] Kai Hwang, Jack Dongarra, and Geoffrey C. Fox. *Distributed and Cloud Computing: From Parallel Processing to the Internet of Things*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN: 0123858801, 9780123858801.

[Low+12] Yucheng Low et al. "Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud". In: *Proc. VLDB Endow.* 5.8 (Apr. 2012), pp. 716–727. ISSN: 2150-8097. DOI: 10.14778/2212351.2212354. URL: http://dx.doi.org/10.14778/2212351.2212354.

[Mal+10] Grzegorz Malewicz et al. "Pregel: A System for Large-scale Graph Processing". In: *Proc. ACM SIGMOD*. SIGMOD '10 (2010), pp. 135–146. DOI: 10.1145/1807167.1807184. URL: http://doi.acm.org/10.1145/1807167.1807184.

6. **Information about the course**

   (a) **Brief description about the course** In order to understand the advanced computational techniques, the students must have a strong knowledge of the various discrete structures, structures that will be implemented and used in the laboratory in the programming language.

   (b) **Prerrequisites:** CS370. Big Data. (9^{no} Sem)

   (c) **Type of Course:** Mandatory

7. **Competences**

   - That the student is able to model computer science problems using graphs and trees related to data structures.
   - That the student apply efficient travel strategies to be able to search data in an optimal way.

8. **Contribution to Outcomes**

a) An ability to apply knowledge of mathematics, science. (**Usage**)

b) An ability to design and conduct experiments, as well as to analyze and interpret data. (**Usage**)

i) An ability to use the techniques, skills, and modern computing tools necessary for computing practice. (**Usage**)

j) Apply the mathematical basis, principles of algorithms and the theory of Computer Science in the modeling and design of computational systems in such a way as to demonstrate understanding of the equilibrium points involved in the chosen option. (**Usage**)

9. Competences (IEEE)

C2. Ability to have a critical and creative perspective in identifying and solving problems using computational thinking. ⇒ **Outcome a**

C4. An understanding of computer hardware from a software perspective, for example, use of the processor, memory, disk drives, display, etc.⇒ **Outcome b**

C16. Ability to identify advanced computing topics and understanding the frontiers of the discipline.⇒ **Outcome i**

CS2. Identify and analyze criteria and specifications appropriate to specific problems, and plan strategies for their solution.⇒ **Outcome i**

CS3. Analyze the extent to which a computer-based system meets the criteria defined for its current use and future development.⇒ **Outcome j**

CS6. Evaluate systems in terms of general quality attributes and possible tradeoffs presented within the given problem.⇒ **Outcome j**

10. List of topics

1. Distributed Systems
2. Cloud Computing
3. Centros de Procesamiento de Datos
4. Cloud Computing
5. Cloud Computing
6. Modelos de Programación

11. Methodology and Evaluation
Methodology:

Theory Sessions:
The development of the theoretical sessions is focused on the student, through his active participation, solving problems related to the course with the individual contributions and discussing real cases of the industry. The students will develop throughout the course a project of application of the tools received in a company.

Lab Sessions:
Practical sessions are held in the laboratory. Laboratory practices are performed in teams to strengthen their communication. At the beginning of each laboratory the development of the practice is explained and at the end the main conclusions of the activity in group form are highlighted.

Oral Presentations :
Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

Reading:
Throughout the course different readings are provided, which are evaluated. The average of the notes in the readings is considered as the mark of a qualified practice. The use of the UTEC Online virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.
Evaluation System:

.

12. Content

Unit 1: Distributed Systems (15)

Competences Expected: C2, C4

Learning Outcomes	Topics
• Distinguish network faults from other kinds of failures [Familiarity] • Explain why synchronization constructs such as simple locks are not useful in the presence of distributed faults [Familiarity] • Write a program that performs any required marshalling and conversion into message units, such as packets, to communicate interesting data between two hosts [Usage] • Measure the observed throughput and response latency across hosts in a given network [Usage] • Explain why no distributed system can be simultaneously consistent, available, and partition tolerant [Familiarity] • Implement a simple server – for example, a spell checking service [Usage] • Explain the tradeoffs among overhead, scalability, and fault tolerance when choosing a stateful v stateless design for a given service [Familiarity] • Describe the scalability challenges associated with a service growing to accommodate many clients, as well as those associated with a service only transiently having many clients [Familiarity] • Give examples of problems for which consensus algorithms such as leader election are required [Usage]	• Faults (cross-reference OS/Fault Tolerance) – Network-based (including partitions) and node-based failures – Impact on system-wide guarantees (e.g., availability) • Distributed message sending – Data conversion and transmission – Sockets – Message sequencing – Buffering, retrying, and dropping messages • Distributed system design tradeoffs – Latency versus throughput – Consistency, availability, partition tolerance • Distributed service design – Stateful versus stateless protocols and services – Session (connection-based) designs – Reactive (IO-triggered) and multithreaded designs • Core distributed algorithms – Election, discovery

Readings : [Cou+11]

Unit 2: Cloud Computing (15)

Competences Expected: C2, C4

Learning Outcomes	Topics
• Explain the concept of Cloud Computing. [Familiarity] • List some technologies related to Cloud Computing. [Familiarity] • Explain strategies to synchronize a common view of shared data across a collection of devices [Familiarity] • Discuss the advantages and disadvantages of the Cloud Computing paradigm. [Familiarity] • Express the economic benefits as well as the characteristics and risks of the Cloud paradigm for business and cloud providers. [Familiarity] • Differentiate between service models. [Usage]	• Overview of it Cloud Computing. • History. • Overview of the technologies involved. • Benefits, risks and economic aspects. • Cloud services – Infrastructure as a service ∗ Elasticity of resources ∗ Platform APIs – Software as a service – Security – Cost management • Internet-Scale computing – Task partitioning – Data access – Clusters, grids, and meshes

Readings : [HDF11], [BVS13]

Unit 3: Centros de Procesamiento de Datos (10)

Competences Expected: C16

Learning Outcomes	Topics
• Describe the evolution of Data Centers. [Familiarity] • Sketch the architecture in detail of the data center. [Familiarity] • Indicate design considerations and discuss their impact. [Familiarity]	• Overview of a data processing center. • Design Considerations. • Comparison of large data processing centers.

Readings : [HDF11], [BVS13]

Unit 4: Cloud Computing (20)	
Competences Expected: CS2, CS3	
Learning Outcomes	**Topics**
• Virtualization – Shared resource management – Migration of processes . [Familiarity] • Explain the advantages and disadvantages of using virtualized infrastructure. [Familiarity] • Identify the reasons why virtualization is becoming enormously useful, especially in the cloud. [Familiarity] • Explain different types of isolation such as failure, resources and security provided by virtualization and used by the cloud. [Familiarity] • Explain the complexity that management can have in terms of abstraction levels and well-defined interfaces and their applicability for virtualization in the cloud. [Familiarity] • Define Virtualization and Identify Different Types of Virtual Machines. [Familiarity] • Identify CPU virtualization conditions, recognize the difference between full virtualization and paravirtualization,explain emulation as a major technique for CPU virtualization and examine virtual CPU planning in Xen. [Familiarity] • Sketching the difference between the classic OS virtual memory and memory virtualization.Explain multiple levels of page mapping as opposed to memory virtualization. Define over-commitment memory and illustrate VMware memory ballooning as a claiming technique for virtualized systems with overcommitted memory. [Familiarity]	• Virtualization – Shared resource management – Migration of processes • Security, resources, and failures isolation . • Storage as a Service. • Elasticity. • Xen y WMware. • Amazon EC2.
Readings : [HDF11], [BVS13]	

Unit 5: Cloud Computing (12)	
Competences Expected: CS2, CS3	
Learning Outcomes	**Topics**
• Describe the general organization of data and storage. [Familiarity] • Identify the problems of scalability and administration of the big data. Discuss several abstractions in storage. [Familiarity] • Compare and contrast different types of file system. Compare and contrast the Hadoop Distributed File System (HDFS) and the Virtual Parallel File System (PVFS). [Usage] • Compare and contrast different types of databases. Discuss the advantages and disadvantages of NoSQL databases. [Usage] • Discuss storage concepts in the cloud. [Familiarity]	• Cloud-based data storage – Shared access to weakly consistent data stores – Data synchronization – Data partitioning – Distributed file systems – Replication • Overview of Storage Technologies. • Fundamentals concepts of cloud storage. • Amazon S3 y EBS. • Distributed File System. • Database System NoSQL.
Readings : [HDF11], [BVS13]	

Unit 6: Modelos de Programación (12)	
Competences Expected: CS6	
Learning Outcomes	**Topics**
• Explain the fundamental aspects of parallel and distributed programming models. [Familiarity] • Differences between programming models: MapReduce, Pregel, GraphLab and Giraph.. [Usage] • Explain the main concepts in the MapReduce programming model. [Usage]	• Overview of cloud computing-based programming models. • Programming Model MapReduce. • Programming model for graph-based applications.
Readings : [HDF11], [BVS13], [Low+12], [Mal+10], [Bal+08]	