



Infraestructura I

Escalabilidad, tolerancia al fallo y disponibilidad en la nube

Vamos a revisar algunos términos relacionados con las tecnologías de la nube.

¿Qué es la escalabilidad de la nube?

Se refiere a la capacidad de crecer en capacidad en demanda. Por ejemplo, si tenemos 1000 usuarios y de repente crecemos a 2000, la capacidad de escalamiento de la nube se refiere a su habilidad para incrementar su capacidad. Pero no solamente crecer, sino también aumentar su capacidad para ahorrar recursos. A este concepto de crecer y reducir su tamaño se le llama elasticidad, a una infraestructura que permite este comportamiento se la llama elástica. Cuando se utilizan servicios elásticos, la nube debe ser capaz de realizar cobros en base al uso de estos, de forma que no se realicen cargos por recursos que no se encuentran en uso.

¿Qué es la tolerancia al fallo de la nube?

Cuando un servicio es tolerante a los fallos, dicho servicio es capaz de responder a cierto grado de errores sin dejar de ser funcional, es decir que tiene la capacidad de solucionar los problemas que existen o reemplazar elementos que no funcionan de



forma correcta por otros que si lo hacen. Un ejemplo puede ser una red local conectada a un switch de muchos puertos alámbricos. Si uno de estos puertos falla, la red puede seguir operando con el resto de los puertos existentes.

¿Qué es la disponibilidad de la nube?

Cuando, por ejemplo, tenemos múltiples recursos y uno de estos falla, pero hay otros que hagan su trabajo, podemos considerar que existe alta disponibilidad. Si por el contrario, este recurso es único y no puede ser reemplazado, se dice que existe una baja disponibilidad.