THE COST OF A CLOUD

Introducción

- alto interés → - mano de obra
  → optimización espacio ∄ equipos
- ¿reducir costos?

Descomposición Costos

Servidores

- 3k USD c/u → 53 M USD data center
- 10 %. utilización → aplicaciones no aprovechan
  - mal estimación demanda
  - pérdidas virtualización mem

Infraestructura

- abastecimiento de energía
- 18,5 M USD por año

Potencia

- PUE: eficiencia entrega energía (~ 1.7)
- 50 k servidores → 9.3 M USD por año

59 %. IT

8 %. lost energy

33%. enfriar

Red

↳ cantidades precisas → metros fibra óptica

→ volumen tráfico

Agilidad

Capacidad de asignar servidores a servicios de manera dinámica manteniendo la seguridad y el aislamiento de rendimiento.

Redes en los Centros de Datos

TRÁFICO fluye → servidores internos

→ sistemas finales externos

... PROBLEMAS ... ¿?

! TRÁFICO INTERNO → VLANs

políticas → gestión tráfico

→ seguridad

→ aislamiento

→ sobrecarga tráfico

→ enlaces altos

! balanceo carga → IP dest C dominio_2

↓

aplicación ≠ otro dominio_2

! dom_{i-2} ⇄ dom_{j-2}

↓ Capa 3

↓ puertos costosos

- ancho banda ∈ limitado
- controlar tráfico ∉ fácil

! Balanceador de carga OVERFLOW →  nuevo par

Objetivos de Diseño

Real.properties()

↳ [0]: services :

server.ip() != server realip()

∉ restricción pertenencia servidor dominio

↳ [1]: s_1  s_2 ∈ indep.  ⇒ s ∈ distribuir evitando congestión

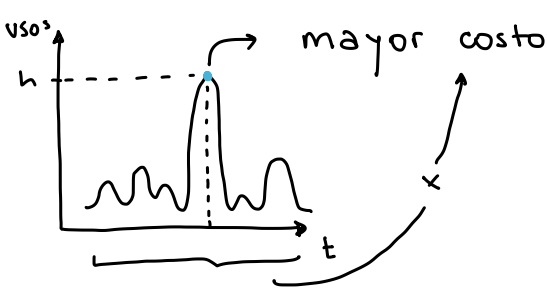
⇓

 ∄ ∼ U ∴ servicios == rend.

↳ [2]: ∀ s_i : s_i ∈ s_i ∴ s → aislados ∴ ~~afecta~~ rend. ∄ disp

Buen Comportamiento ⇝ △ consumo recursos

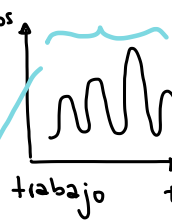
- **Llenado de Bajas**



$ → V (disponibilidad recursos)

↓

≠ demandas
△ ejecución
→



desplazar carga trabajo

- **Asignación de Servidores**

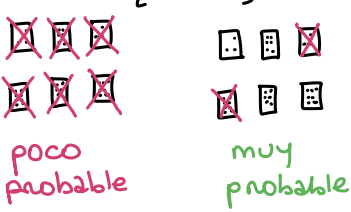
∉ acaparar servidores → devolver libres

$+ → d(s), d >> σ

Geo-Distribución { dónde, tamaño, diversidad p, disponibilidad } ↓ retrasos v. luz

★ Ubicación & Dimensionamiento Óptimos — factores

áreas separadas ⇒ ↓ latencia & + redundancia

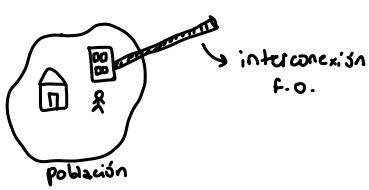


tam (▯) != tam (▭)

↓

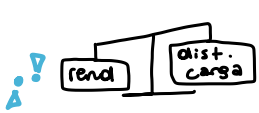
+ flexibles ubicar

equilibrio ubicación ▯ cerca



★ Diversidad Geográfica ⇝ Geo-redundancia

estado crítico aplicaciones { = ≠ } sitios ...



¡! compensar costos comunicación con rendimiento al replicar o particionar datos

¡! dispersar s_i ∄ manejar → ↑ latencia
↳ costo

Conclusiones ↑ costos ; ↓ utilización

+ eficiencia ∄ + aprovechamiento :

⊕ agilidad red interna → − fragmentación
↳ + trabajo − servidores

✓ buen diseño algoritmos Δ consumo recursos
mejoran eficiencia

% diversificación geográfica → mejora rendimiento total
↳ + confiabilidad fallas