

THE COST OF A CLOUD

Introducción

- alto interés → - mano de obra
→ optimización espacio & equipos
- reducir costos?
↓

Descomposición Costos

Servidores

→ 3k USD c/u → 53M USD data center

→ 10% utilización → aplicaciones no aprovechan
→ mal estimación demanda
→ pérdidas virtualización mem

Infraestructura

→ abastecimiento de energía

→ 18,5M USD por año

Potencia

→ PUE: eficiencia entrega energía (~ 1.7)

→ 50k servidores → 9.3M 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

anqui.
conv.
Cloud
/ \
Cisco
2004

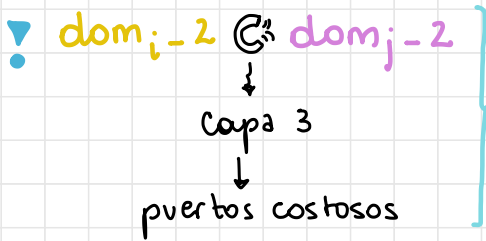
... PROBLEMAS ... ¿?

! TRÁFICO INTERNO → VLANs


políticas { gestión tráfico
seguridad
aislamiento

sobrecarga tráfico
enlaces altos

! balanceo carga → IP dest C dominio_2
↓
aplicación ≠ otro dominio_2



- ancho banda es limitado
- controlar tráfico \neq fácil

! Balanceador de carga OVERflow \rightarrow  nuevo par

Objetivos de Diseño

Req. properties ()

$\rightarrow [0]$: services:

$server.ip() \neq server.realip()$

\neq restricción pertenencia servidor dominio

$\rightarrow [1]$: $S_1, S_2 \in indep. \circ \Rightarrow S \in$ distribuir evitando congestión



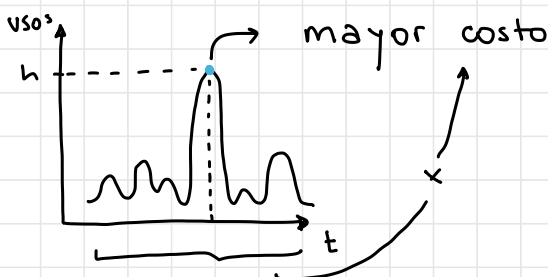
$\neq \text{mml} \sim U \therefore \text{servicios} == \text{rend.}$

$\rightarrow [2]$: $\forall S_i : S_i \in s_i \therefore s \rightarrow$ aislados \therefore ~~afecta rend.~~ \neq disp

Buen Comportamiento

Δ consumo recursos

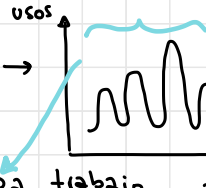
Llenado de Bajos



$\$ \rightarrow V$ (disponibilidad recursos)

\neq demandas Δ ejecución

desplazar carga trabajo t

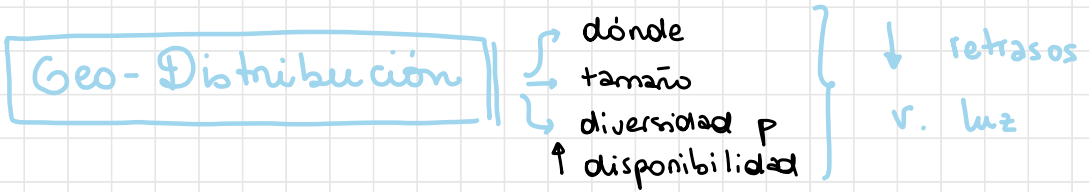


Graph showing usage (usos) on the y-axis and time (t) on the x-axis. A dashed line indicates a peak labeled "desplazar carga trabajo".

• Asignación de Servidores

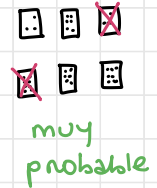
≠ acaparar servidores → devolver libres


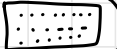
\$+ \rightarrow d(s), d \gg \sigma\$



★ Ubicación & Dimensionamiento Óptimos factores

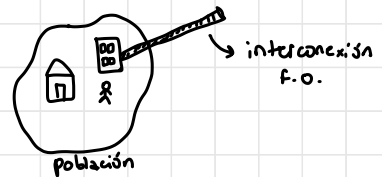
→ áreas separadas \Rightarrow ↓ latencia & ⇄ redundancia



tam () \neq tam ()

↓
+ flexibles
ubicar

equilibrio ubicación  cerca



★ Diversidad Geográfica \leadsto Geo-redundancia

estado crítico aplicaciones $\left\{ \neq \right.$ sitios ...



- compensar costos comunicación con rendimiento al replicar o particionar datos
- dispersar si ≠ 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