Giant-Scale Services

Brewer advocates extreme symmetry. Why?

Cluster benefits:
- Key question: what is the lowest-level building block of a cluster?
- Services that can scale to global use
- Simplified service updates
- Lower overall cost
- Groupware support
- Availability via multiple devices
- Access anywhere, anytime

Benefits of Network Services
- Are clouds = clusters?
- Independent components
- Incremental scalability
- Commodity processors or higher-end "super-computer"?
- Do you think this is still true? (What about P2P? Clouds?)
- Update the service in one place, or 100 million?
- Central administrative burden, simplified end devices
- Dedicated resources are typically at least 96% idle
- Multiplex infrastructure over active users
- Calendaring, teleconferencing, messaging, etc.
- Also more true…

Cloud computing?
- This is even more true now than it was in 2001

Clusters as Building Blocks
- Adding one machine typically linearly improves performance

F5 Networks 3-DNS

Load Management Option 1:
- Map hostname to multiple IP addresses, hand out
- Perform mapping of requests to back end servers based on dynamically changing membership information
- Does not hide failure or inactive servers

Exposes structure of underlying service

F5 Networks 3-DNS

Site 1
New York

3-DNS Controller S20

Router

DNS

BIG-IP 2400

Corporate Servers

Corporate Users

Site 2
London

Site 3
San Francisco

User
Seattle

Local DNS
Brewer advocates extreme symmetry. Why?

Cluster benefits:

- Key question: what is the lowest-level building block of a cluster?
- No alternative to clusters for building network
- Lower overall cost
- Availability via multiple devices
- Access anywhere, anytime

Cost and performance:

- Independent components
- Commodity processors or higher-end "super-computer"?

Do you think this is still true? (What about P2P? Clouds?

Update the service in one place, or 100 million?

Central administrative burden, simplified end devices

Dedicated resources are typically at least 96% idle

Multiplex infrastructure over active users

Calendaring, teleconferencing, messaging, etc.

Also more true…

This is even more true now than it was in 2001

Clusters as Building Blocks

F5 Networks 3DNS

"Load Management: Partitioning"

Load Management: Partitioning

Load Management Option 1:

Load Management Option 2:

Vertical partitioning

Uptime: fraction of time service is handling traffic (usually measured in "nines")

Can improve availability by increasing MTBF or by reducing MTTR

Availability = (MTBF – MTTR)/MTBF

Today, "fancy" L4 and L7 switches can inspect TCP session state or HTTP session state

Partitioning Keywords in Search

Multi-word search

Map keywords to a set of documents containing all words

Exposes structure of underlying service

Does not hide failure or inactive servers

Split inverted index across multiple nodes (nodes=data storage devices)

Has to visit every node in system to perform full join (or search)

Essentially like splitting table up into multiple tables (with same number of columns) by putting different (complete) rows in different tables

Essentially like reducing the number of columns in table, and using extra columns to store remaining columns

Each node contains as much of index as possible for a particular keyword

Each node contains portion of inverted index for a particular keyword

Has to visit every node in system to perform full join (or search)

Has to visit every node in system to perform full join (or search)

Starts with "round-robin" DNS in 1995

Performs mapping of requests to back end servers based on dynamically changing membership information

Map hostname to multiple IP addresses, hand out a particular mapping in a round robin fashion to clients

Perform mapping of requests to back end servers based on dynamically changing membership information

Load Management: Replication

Load Management: Replication

Simple replicated store

Round-robin DNS

Single-site server

IP network

Client

Client

Client

Client

Load manager

Partitioned data store

Myrinet backplane

Program

Program

Program

Program

IP network

Single-site server
Measuring Availability

Admission Control

Twitter is over capacity.
Please wait a moment and try again. For more information, check out Twitter Status.