Cloud Computing: Task Parallel Computing

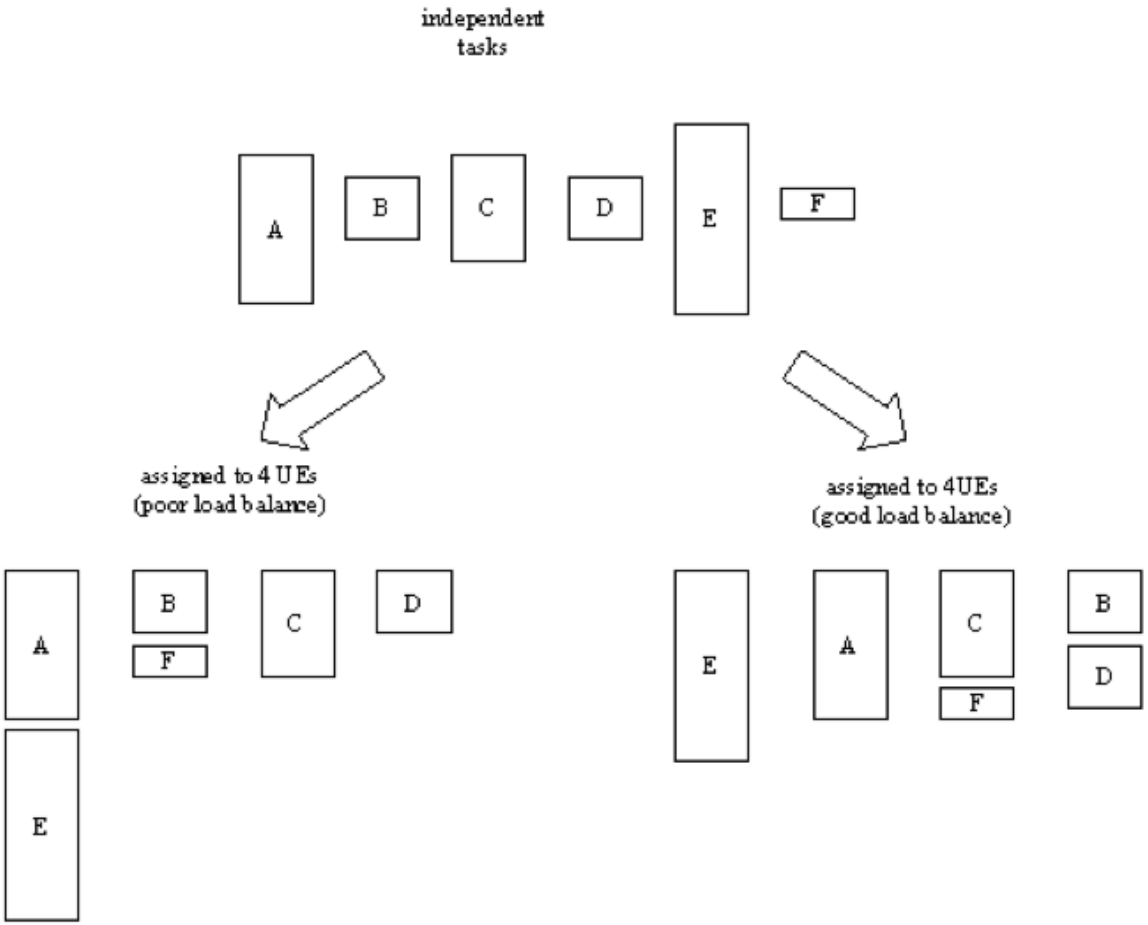
Vijay Dialani, PhD
Boise State University
vijaydialani@boisestate.edu

Some things to catch up …

- Example of session management in Web Servers:
  - http://www.cs.gmu.edu/~offutt/classes/642/slides/642Lec10b-JSP-stateHandling.pdf

- Example of tomcat container when hosting a web service

What is task parallelism?



https://blackboard.boisestate.edu/webapps/portal/frameset.jsp?tab_tab_group_id=_2_1&url=%2Fwebapps%2Fblackboard%2Fexecute%2Flauncher%3Ftype%3DCourse%26id%3D_62145_1%26url%3D

Properties of Task Parallelism

- Each Task is independent of others
- Each Task can be randomly allocated to any of the processing units
- System is elastic
  - Resilient to failure of nodes
  - Resilient to addition of new nodes
- If idempotent - each task can be simultaneously scheduled on multiple nodes
  - To provide resilience
  - To provide guarantees on completion

Which are the allocation modes for the tasks?

- Round Robin
  - All nodes are arranged in a circular array
  - Each incoming request increments the pointer and allocates the request to next node

- Least Loaded
  - All nodes describe their current load in terms of either:
    - Number of pending task
    - Estimated time for completion
    - Least loaded node is selected for processing the request

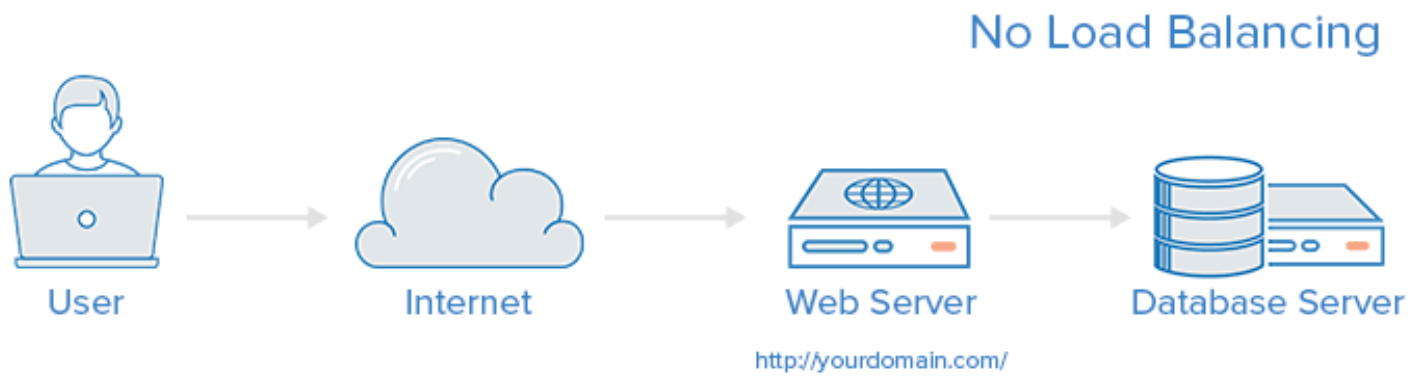
- Consistent hash allocation
  - A field from the incoming request is chosen to identify the hash code
  - A DHT(Distributed Hash Table) is used to allocate the hash code to the node serving the range.

Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications
http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord_sigcomm.pdf

JAX-RS and Load Balancing

No Load Balancing

A simple web application environment with no load balancing might look like the following:



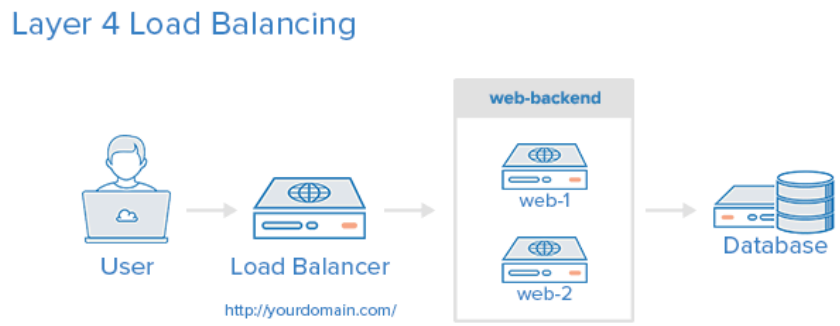
In this example, the user connects directly to your web server, at *yourdomain.com* and there is no load balancing. If your single web server goes down, the user will no longer be able to access your web server. Additionally, if many users are trying to access your server simultaneously and it is unable to handle the load, they may have a slow experience or they may not be able to connect at all.

JAX-RS and Load Balancing

Layer 4 Load Balancing

The simplest way to load balance network traffic to multiple servers is to use layer 4 (transport layer) load balancing. Load balancing this way will forward user traffic based on IP range and port (i.e. if a request comes in for *http://yourdomain.com/anything*, the traffic will be forwarded to the backend that handles all the requests for *yourdomain.com* on *port 80*). For more details on layer 4, check out the *TCP* subsection of our Introduction to Networking.

Here is a diagram of a simple example of layer 4 load balancing:



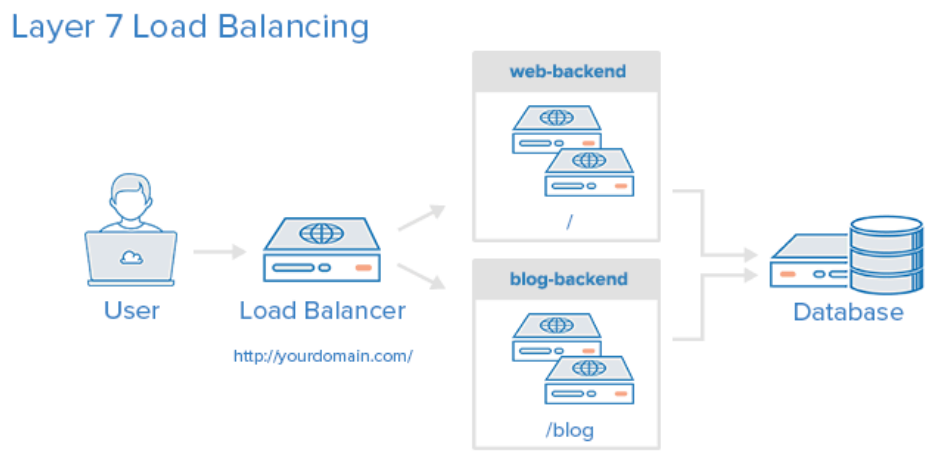
The user accesses the load balancer, which forwards the user's request to the *web-backend* group of backend servers. Whichever backend server is selected will respond directly to the user's request. Generally, all of the servers in the *web-backend* should be serving identical content--otherwise the user might receive inconsistent content. Note that both web servers connect to the same database server.

JAX-RS and Load Balancing

Layer 7 Load Balancing

Another, more complex way to load balance network traffic is to use layer 7 (application layer) load balancing. Using layer 7 allows the load balancer to forward requests to different backend servers based on the content of the user's request. This mode of load balancing allows you to run multiple web application servers under the same domain and port. For more details on layer 7, check out the *HTTP* subsection of our Introduction to Networking.

Here is a diagram of a simple example of layer 7 load balancing:



In this example, if a user requests *yourdomain.com/blog*, they are forwarded to the *blog* backend, which is a set of servers that run a blog application. Other requests are forwarded to *web-backend*, which might be running another application. Both backends use the same database server, in this example.

Example HAProxy based code in git chapter4

Special Thanks to:
https://www.digitalocean.com/community/tutorials/an-introduction-to-haproxy-and-load-balancing-concepts

http://www.haproxy.org