Cloud Constructor Teacher Notes

The Cloud Constructor game provides an interactive game that introduces a number of cloud computing, client/server, and distributed systems concepts. The game features 10 levels labeled Levels 0 - 9 where each level introduces one or more concepts. The game begins by selecting the "PLAY" button. The 10-levels serve as an introduction to concepts as well as the mechanics of the game itself. The premise of the game is that the player serves as the computer "systems administrator" and they must adapt their cloud computing infrastructure in each level to provide resources sufficient to meet user (client) demand. The primary tasks involve: storing messages, movies, and pictures, backing up and replicating data elements across multiple servers, and routing client (user) requests to the appropriate server with the requested data. Once completing all 10 levels, students have the option of playing the game in "ENDLESS" mode where each level is randomly generated and features additional challenges. Students are challenged to see how many levels they complete. In endless modes players have three lives or strikes where each time they fail to meet client demand results in failing the level and earning a strike. After three strikes, the game ends.

The Cloud Constructor Main Menu features PLAY, QUIZ, and ENDLESS buttons.

Cloud Constructor Quiz

After students have completed all 10 levels, and have played the game in endless mode, teachers are STRONGLY encouraged to have students complete the cloud computing "QUIZ" accessed from the main menu using the quiz button. The quiz features 12 multiple choice questions for a total of 15 points. The quiz tests whether students have learned the concepts the game has conveyed. Students may repeat the quiz multiple times.

Cloud Computing Vocabulary Terms:

The Cloud Constructor game introduces the following terms, or vocabulary, in each level as defined below:

Level 0:
Computer Server - These are computers that are in the cloud that process client jobs/requests. They feature central processing units (CPUs) to process data, disk storage to store data, and networking to support communication with clients.

Storage - computer data is stored using disks (hard disk drive-HDD) or (solid state drive-SSD). Hard disk drives consist of physical cylinders that rotate and a head that reads and writes data to/from the physical surface. Solid state drives store data.

Processing - processing on servers is performed by central processing units (CPUs). These are the central "brain" chips of the computer that execute computer programs. Programs work with data provided by the client via the network, or from a disk.

Networking - networking is used to connect computer servers in the cloud to form a cluster of computers that work together to solve problems. Networking is also used to connect clients and servers.

Level 1:
Client - clients submit requests to servers to perform work. This involves requests to process data, play a movie, show a picture, or even to play a game.

Upload/Download data - client requests to save or retrieve data appear as an individual widget depicted with either an upload (arrow pointing up) or download (arrow pointing down).

Client Upload Request   **Client Download Request**
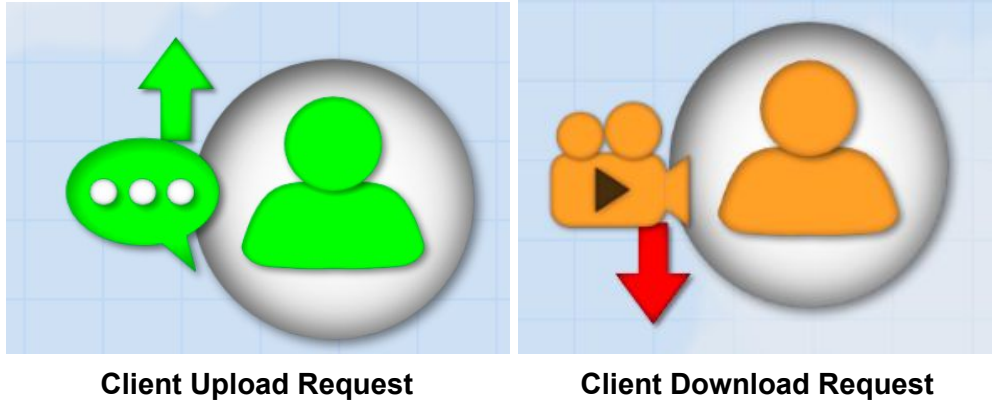
Three types of data include:

Messages   **Pictures**   **Movies**

The player must respond to requests by building enough server capacity with respect to storage, processing, and networking to satisfy demand for client requests.

Level 2:
Horizontal Scaling - horizontal scaling involves adding more computers to handle more requests. This allows requests to be distributed across a larger set of computer servers. Horizontal scaling supports load balancing of client requests; a term introduced in level 7.

Vertical Scaling - vertical scaling involves expanding the resources of a single computer to handle more requests or to process requests more quickly. Compared to horizontal scaling, vertical scaling typically involves a more finite limit where a single server can not be expanded further. For processing, today few computer processors exceed a 4 GHz processing speed. For storage, an individual computer server may be limited in storage capacity to several thousand gigabytes (GB) or terabytes (TB). For networking, an individual computer server may be limited to networking throughput of 1 Gigabits per second. The difference is in the terms gigabyte and gigabit. There are 8 bits in a byte so, in reality, 1 Gbps translates into about 125 mega**bytes** per second (MBps). Remember though, that is just the theoretical limit of a Gigabit network.

Level 3:
Service Isolation - service isolation involves the use of dedicated computer servers to process individual client requests. With service isolation the computer is dedicated to processing a single, individual user request, and performs no other activity. The idea is with service isolation the client's request can complete without interference or slow down from processing other client

3

requests simultaneously. Service isolation can be expensive, however, because it requires dedicated computer resources to process the client's request.

Level 4:
Availability - availability refers to the percentage of time a computer service is available to end users. Availability is usually described as a percentage consisting of a number of 9s. For example 99% availability implies 3.65 days of downtime annually. 99.9% availability implies 8.76 hours of downtime annually, where 99.99% availability implies only about 53 minutes of downtime annually.

Data Replication - when data is copied to multiple servers for redundancy and backup purposes this is called replication. Data replication is also used to improve performance. For example, if multiple users access the same movie at the same time, then with redundant copies of the movie distributed across servers, then client requests can be load balanced to improve movie download performance as a dedicated service (service isolation) could provide data for just one client. If a single server must provide data to multiple clients at the same time, then the available network bandwidth will be shared reducing the download performance.

High Availability - high availability refers to a degree of availability (see above) that is considered to be high. Some consider high availability as having 5 nines, or 99.999% availability in a year. Other interpretations may consider having 3 nines or 99.9% availability as "high" availability for a given year. The degree of service reliability for a service to be considered as offering high availability may depend on the type of computer service offered.

Level 5:
Fault Tolerance - a computer system that is considered to be "fault tolerant" is a computer system that can suffer hardware or software failure without affecting the system's availability. Fault tolerance requires detecting the failure and transparently addressing the failure without the user knowing. Fault tolerance in a distributed computing system is often provided by employing a combination of replication and duplication of data and computing resources.

Level 6:
Parallelism - involves processing multiple (client) requests in parallel (at the same time). A client or user that sends multiple requests to a computing service to process will perceive the entire system as offering faster performance when requests are processed in parallel. Parallel processing usually requires the use of horizontal or vertical scaling to load balance and process multiple requests in parallel.

Level 7:
Load Balancing - load balancing is the concept of distributing work evenly across a set of available computing servers. There are various approaches as to what is considered as the best way to distribute requests across servers.

Level 8:

Databases - databases are special computer servers that are customized for the task fo storing and retrieving data. In computing databases usually fall into two categories: relational databases, where data is stored in a structured form of using a series of tables called data normalization to avoid data duplication, and unstructured key-value databases where the emphasis is on simplifying the data structure to more easily allow data replication across multiple servers.

Level 9:

Gateways - gateways are a key node in a computer network that route client requests to the proper servers to handle requests. Gateway offer an intermediary between clients and servers where functions such as security and data caching can be implemented. Gateways route traffic from a client to the network where the servers hosting requested data are.

Questions or Feedback:

Questions or feedback on the cloud constructor game are encouraged.
Please contact Wes Lloyd at the University of Washington Tacoma: wlloyd@uw.edu

November 2, 2019
Revised to add game description and terms.

September 29, 2019
Initial version.