

Escalamiento Horizontal

Para la realización del escalamiento horizontal sobre la arquitectura de nuestra aplicación será necesario en primer instancia la confección de un cluster de la aplicación con un número N de nodos asociado al escalamiento requerido.

Para administrar eficientemente la carga sobre los distintos nodos acudiremos al agregado de un load balancer que se encargará de la óptima derivación de los usuarios a las diferentes instancias del cluster.

Aplicando esta modalidad, uno de los posibles problemas será la sincronización de los nodos de la aplicación ante una eventual actualización del sistema. A fin de evitar el proceso manual de sincronización, se optará por la configuración de un servicio de deployment continuo y automatizado mediante la utilización de herramientas como Jenkins o similares.

Base de datos (Storage)

Acorde a los lineamientos de la estrategia de escalamiento antes definida, será necesaria la separación de la base de datos alojada inicialmente en cada servidor y destinarla a un “storage” unificado, a fin de que los múltiples nodos puedan compartir el mismo servidor de base de datos manteniendo la consistencia de los datos y ayudando a su vez a eliminar la carga de los respectivos nodos.

Llegado el caso de alta demanda a la DB, también sería posible aplicar de similar manera un escalamiento horizontal y/o vertical exclusivamente sobre el servidor de storage.

Consistencia de sesión (SessionManager)

Resuelta la separación y consistencia de la base de datos, otra problemática resultará como consecuencia de la administración del loadBalancer, que puede caer en la necesidad de derivar a un cliente en diferentes nodos durante la misma sesión de usuario.

Para solucionar el conflicto del manejo de la consistencia de la sesión de usuarios a través de los diferentes nodos será requerido implementar un servidor con la finalidad exclusiva de la administración de las sesiones de usuario (SessionManager) mediante la implementación de tecnologías como Memcache o Redis.

Naturalmente, todos los nodos del cluster deberán tener conexión con este nuevo servidor de manejo de sesiones garantizando la consistencia de la misma.

Consistencia de archivos de usuario (FileServer)

Inicialmente los archivos subidos serán almacenados en cada nodo en el que se encuentre el usuario al momento de realizar la carga, lo cual trae aparejado un conflicto de consistencia de archivos, ya que al intentar referenciar a un archivo desde un nodo puede ocurrir que este no sea encontrado como consecuencia de haber sido cargado desde otro nodo.

Para solucionar este conflicto utilizaremos como opción un servidor dedicado exclusivamente a la carga de archivos, que al ser referenciado desde todos los nodos, se encargará de mantener la consistencia de archivos cargados por los usuarios de manera unificada, independientemente del nodo en el que este haya realizado la carga.

Diagrama de escalamiento horizontal:

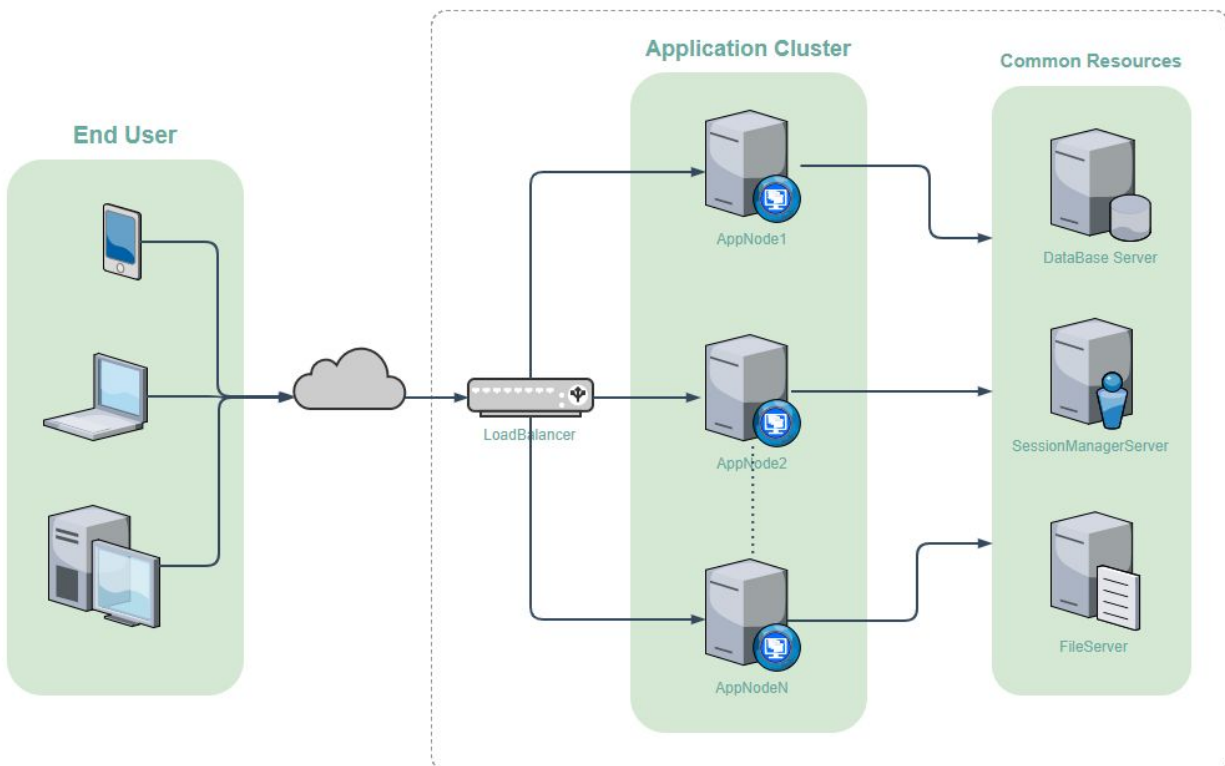


Diagrama de arquitectura actual:

