

FTP Distribuido

Arquitectura o el problema de como diseñar el sistema

Cualquier arquitectura que garantice la comunicación directa entre nodos es válida, la elección va a depender en mayor medida de la facilidad del proceso de descubrimiento de nodos que de la operación misma del sistema distribuido. Se necesita poder consultar que nodos están en línea y consultar si un nodo específico está en línea.

Procesos o el problema de cuantos programas o servicios posee el sistema

Un servidor ftp consiste en solamente un proceso en espera de clientes y multiples sesiones (una sesión por cliente), el funcionamiento del servidor y las sesiones es multiplexado en solamente un hilo, de modo que la creación de nuevas sesiones y el comienzo de la ejecución de los comandos ftp son realizados de forma síncrona y en orden de llegada, casi todos los comandos ftp (excepto los de transferencia de datos) terminan inmediatamente, el resto provoca la creación de un hilo que completa la transferencia de datos y la ejecución del comando ftp sin interferir con el funcionamiento normal del servidor y el resto de sesiones. De esta forma un servidor ftp soporta multiples endpoints, multiples clientes, multiples sesiones por cliente y multiples operaciones concurrentes por cliente (incluso si esto último no está definido en la especificación del protocolo).

Comunicación o el problema de como enviar información mediante la red

La comunicación cliente - servidor está definida por el protocolo. El servidor consiste en un solo proceso (no existe IPC). Cada servidor asume que la comunicación directa con el resto de servidores es posible y se realiza intercambiando mensajes en cualquier formato predefinido.

Coordinación o el problema de poner todos los servicios de acuerdo

Cada servidor ftp es responsable de sus archivos locales únicamente, la toma de decisiones no es distribuida, cada nodo decide si aceptar peticiones de otros servidores de acuerdo a que acciones son válidas en sus archivos locales. El acceso concurrente a un archivo (local o no) solo está permitido para operaciones de lectura.

Nombrado y localización o el problema de dónde se encuentra un recurso y como llegar al mismo

La estructura de archivos y directorios está bien definida y la ubicación de los recursos en la red viene implícita en su ubicación en el árbol conjunto de directorios.

Consistencia y replicación o el problema de solucionar los problemas que surgen a partir de tener varias copias de un mismo dato en el sistema.

Los archivos se almacenan exclusivamente en su raíz local, lo que significa que durante una desconexión el acceso se limita a los archivos locales de los nodos alcanzables. Lamentablemente la única forma de impedir la pérdida de acceso es replicar todos los archivos en todos los nodos porque no podemos predecir que archivo va a necesitar el usuario ni como ocurrirá la desconexión.

Tolerancia a fallas o el problema de, para que pasar tanto trabajo distribuyendo datos y servicios si al fallar una componente del sistema todo se viene abajo.

En caso de desconexión cada nodo siempre va a poder acceder como mínimo a sus archivos locales, por tanto las operaciones con los archivos locales solamente fallan con la desconexión directa de cliente. Además, el protocolo ftp especifica que ante la desconexión del socket de control las operaciones deben abortarse, lo que significa en nuestro caso que incluso si un archivo no local es accesible, la transferencia debe abortarse si el servidor local se desconecta.

Seguridad o el problema de que tan vulnerable es su diseño

El protocolo ftp es inseguro por diseño, la comunicación es en texto plano y la autenticación y autorización son definidos por la plataforma (generalmente contraseña, de igual forma transmitida en texto plano). Posibles mejoras incluyen utilizar TLS (RFC 2228 FTP Security Extensions) y utilizar algún mecanismo propio de autenticación.