

Uso de gRPC y algoritmos de exclusión mutua en sistema de descarga y subida de archivos

Qué se hizo

Para el sistema de descarga y subida de ficheros se utilizaron cuatro máquinas, tres de ellas fueron usadas como DataNode y una fue usada como NameNode, además se contaba con dos tipos de clientes, uno que se encargaba de subir archivos al que llamaremos Cliente Uploader, y otro que se encargaba de descargarlos al que llamaremos Cliente Downloader.

Los DataNodes estaban encargados de recibir chunks de los archivos, revisar los otros DataNodes que estaban disponibles, proponer la distribución de los chunks al NameNode o a los demás DataNodes disponibles, una vez recibida una propuesta aprobada distribuir los chunks y finalmente escribir en el registro del NameNode el lugar en el que guardaron cada chunk.

Por otro lado, el NameNode estaba encargado de recibir mensajes de los DataNodes con la ubicación de los archivos y escribirlos en su registro, además de responder a clientes que solicitan la ubicación de algún archivo en específico con la ubicación de cada chunk que necesita descargas.

En el caso de los clientes, el Uploader se encargaba de separar el archivo en chunks de 250kB antes de enviarlo al DataNode, y en el caso del Downloader se encargaba de solicitar al NameNode la ubicación de cada chunk del archivo que quería, descargarlo de los DataNodes, para finalmente reconstruirlo en el archivo original.

Cómo se hizo

Para la comunicación entre distintos nodos se utilizó gRPC y Protocol Buffers para el lenguaje de programación Go. Para la escritura en el registro del NameNode, los DataNode debían ponerse de acuerdo entre ellos para decidir en que orden lo hacían, esto se lograba a través de un algoritmo de exclusión mutua que podía ser centralizado o distribuido.