Puddlestore

Jack Kates and Duncan McManus

CSCI1380 (Distributed Systems) with Theo Benson

May 2020

Puddlestore is a distributed file system inspired by OceanStore, developed at UC Berkeley. It provides a highly available and persistent storage mechanism using Tapestry, a Distributed Object Location and Retrieval system, and Zookeeper, a key-value store developed by Apache. Files are divided into blocks and stored on a network of servers which can be distributed around the globe. Participating servers store the blocks that make up files. The servers are connected in a dynamic mesh network using Tapestry. A small number of servers also run Zookeeper, which stores important file metadata like size as well as the unique identifiers of the blocks making up a file. Zookeeper uses distributed consensus to replicate this information and ensure that no file data is lost in the event of a server failure. A user uses a Puddlestore client to access and update their files. When they read or modify a file, their client contacts a Zookeeper node and learns the unique identifiers of all the blocks that make up the file. The client contacts one of the storage nodes over the Internet and requests the blocks. Since the blocks may be distributed across many servers, the requests are routed efficiently across the network until all blocks are found. Puddlestore uses distributed locking to allow multiple clients to access the filesystem while maintaining consistency. Only one Puddlestore client may open a file at a time, ensuring that updates from two clients do not conflict with one another. The internal implementation of Puddlestore is opaque to the client, who experiences a familiar Unix-like API.