

Puddlestore

Anna Nakai and Amy Pu

CSCI1380 (Distributed Systems) with Theo Benson

Puddlestore is a distributed file system based on OceanStore, a project developed at UC Berkeley in the 2000s. It is designed to provide continuous access to persistent information, and is built off of Tapestry, a Distributed Object Location and Retrieval system, and Zookeeper, a service offering a hierarchical key-value store.

The contents of files and directories are stored as data blocks in Tapestry, but are accessed by the Puddlestore Client through Zookeeper, which stores each file's inode. Each inode stores a map of data blocks, which are identified using globally unique identifiers (GUIDs). Accessing the data blocks for a specific file requires querying for the file's inode in Zookeeper to get the map of GUIDs, and then retrieving the data blocks from Tapestry using the GUIDs. Each inode is identified by its path in Zookeeper. To use Tapestry through Zookeeper, a "/tapestry" directory is created where inodes containing information required to connect to tapestry nodes are stored. From these inodes, we can create Tapestry clients and interact with the Tapestry system.

In order to ensure consistency in file changes across multiple Puddlestore clients, we implemented a distributed file locking system, utilizing Zookeeper's fully distributed and globally synchronous distributed locks. We create a specific "/locknode" directory within Zookeeper, where each Puddlestore client waiting for a lock has its specific lock file. Puddlestore clients are not able to concurrently access the same file and interfere with each other, because we ensure that only one Puddlestore client has the lock node at any point in time, and only the client with the lock node is able to access the specified file.