ChubaoFS CNCF

Introduction

The Cloud Native Computing Foundation (CNCF) is **a Linux Foundation project that was founded in 2015 to help advance container technology and align the tech industry around its evolution**.

ChubaoFS(Chubao File System) is a distributed file system that is designed to natively support large scale container platforms.

ChubaoFS offers an innovative new option for general-purpose distributed/shared storage infrastructure regardless of file size or file access pattern

Project Summary

| website | https://www.cncf.io/ |
| --- | --- |
| Organization/Foundation Name | CNCF |
| License | Apache License, Version 2.0 |
| Open/Proprietary | Open |
| Source Path(if open Source) | Kubernetes is an open source system used by CNCF. |
| Brief Description | Kubernetes (sometimes shortened to K8s with the 8 standing for the number of letters between the “K” and the “s”) is an **open source system** to deploy, scale, and manage containerized applications anywhere. |

* 
* Open media 1 in modal

**PROJECT DETAILS**

Key Features

*High Performance* - Optimized for both large and small files.

G*eneral Purpose Storage Engine* - Different file access patterns.

*POSIX-Compliant APIs* - Comply with POSIX semantics wherever possible.

Multi-Tenancy Kubernetes volume support.

Highly Scalable- Separate metadata cluster.

Object Storage APIs Operate files by using S3-Compliant APIs.

Architecture

1.metadata sub system

• The subsystem has a set of meta nodes

• Each meta node has a set of meta partitions

• Each meta partition has several replicas (usually 3) in different meta nodes to form a Raft group

• Each meta partition belongs to a specific volume

• Raft logs and snapshot • Multi-raft based (https://www.cockroachlabs.com/blog/sc aling-raft/)

2.data subsystem

Place to persist the file contents

• Data Node and Data Partition

• Sequential or random access

Large files

One file -> a set of new extents

• Extent: a storage unit of data node.

Small files

• One extent -> multiple small files

• Deleting a file and free disk space through punch hole .

Dual replication protocol

• Primary-Backup for sequential write: better performance than Raft

• Multi-Raft for overwrite: ensure strong replication consistency at the cost of performance.

3. Object subsystem

*Operate files like using object storage*

• HTTP based Rest APIs

• Compatible with native Amazon S3 SDKs *Fusion Storage*

• Expose two interface (POSIX and object storage interface).

*Semantic convert*

*• Bucket -> Volume • Key -> Path*

4. Performance –large file

Setup

• 64 processes in random read/write tests

• 16 processes in the sequential read/write tests

• Each process operates a separated 40 GB file.

ChubaoFS outperforms CephFS in the random read/write tests.Each meta node caches all file metadata in memory to avoid expensive disk IOs , Overwrite in ChubaoFS is in-place (no need to update metadata) and therefore ChubaoFS outperforms CephFS in the random read or write.

5.Perfomance -small file

Setup

• File size ranging from 1 KB to 128 KB.

**CURRENT USAGE**

ChubaoFS is a distributed file system that is compatible with most POSIX file system semantics. When ChubaoFS is mounted, it can be as simple as using a local file system. Basically, it can be used in any case where a file system is needed, replacing the local file system, and realizing infinitely expandable storage without physical boundaries. It has been applied in various scenarios, and the following are some of the extracted scenes.

## **Machine Learning**

Disadvantages of using a local disk to store training data sets:

* The local disk space is small, and there are multiple models. The training data set of each model reaches the TB level. If you use a local disk to store the training data set, you need to reduce the size of the training data set.
* Training data sets need to be updated frequently, requiring more disk space.
* Risk of loss of training data set if machine crashes.

The advantages of using ChubaoFS to store training data sets:

* Unlimited disk space, easy to expand capacity. It can automatically expand disk capacity according to the percentage of disk usage, enabling storage system capacity expansion on demand, greatly saving storage costs.
* Multiple replicas of data to ensure high data reliability without worrying about losing data.
* Compatible with POSIX file system interface, no changes required by the application.

## **ElasticSearch**

Using local disks to store data often encounters the following problems:

* Disk usage is uneven and disk IO cannot be fully utilized.
* Local disk space is limited.

The advantages of using ChubaoFS as a backend storage:

* Unlimited disk space, easy to expand capacity. It can automatically expand disk capacity according to the percentage of disk usage, enabling storage system capacity expansion on demand, greatly saving storage costs.
* Disk IO usage is uniform and disk IO is fully utilized.
* Ensure data is highly reliable without worrying about losing data.

## **Nginx Log Storage**

Do you often worry about running out of local disk space? With ChubaoFS, you can store datas in a distributed file system without worrying about running out of disk space. If you use a local disk to store logs, you may often worry about the following issues:

* Docker local disk space is small.
* If the docker container crashes, the log is lost and unrecoverable.
* The mixed deployment of physical machine and docker machine is difficult to manage and has high operation and maintenance cost.

The advantages of using ChubaoFS to store nginx logs are:

* The disk space is unlimited and easy to expand. The disk capacity is automatically expanded according to the percentage of disk usage. The storage system can be expanded on demand, which greatly saves storage costs.
* Ensure data is highly reliable and do not worry about losing data.
* Multiple replicas to solve the problem of unable to write to the log caused by disk error and datanode crashes.
* Compatible with the POSIX file system interface, no changes required by the application.
* The operation and maintenance of ChubaoFS are simple, and one person can easily manage the cluster of tens of thousands of machines.

## **Spark**

In the big data set scenario, are you worried about the amount of data stored in Spark intermediate calculation results in order to carefully calculate each task? You can store the shuffle results to ChubaoFS, and no longer worry about the disk has no free space which causes the task to fail. This enables the separation of storage and computation. The pain points of using local disk to store shuffle intermediate results:

* Insufficient disk space.
* Too many temporary directory files to create new files.

The advantages of using ChubaoFS:

* Unlimited disk space, easy to expand capacity. It can automatically expand disk capacity according to the percentage of disk usage, enabling storage system capacity expansion on demand, greatly saving storage costs.
* Meta node manages file metadata, which can be expanded horizontally and the number of files is unlimited.

## **MySQL Database Backup**

Disadvantages of using OSS(Object Storage Service) to back up MySQL database:

* Need to use the OSS SDK or RESTful API to develop backup programs which increases the difficulty of operation and maintenance.
* If backup file fails, troubleshooting is more difficult.
* After backing up files to OSS, it is not convenient to check whether the files are successfully uploaded.
* Backup files are processed by multiple layers of services, which affects performance.

Advantages of using ChubaoFS to back up MySQL databases:

* Easy to use, compatible with POSIX file interface, and can be used as a local file system.
* Complete and detailed operation logs are stored in the local file system, making it easy to troubleshoot problems.
* Simply execute the ls command to verify that the file was successfully uploaded.
* Supports PageCache and WriteCache, and file read and write performance is significantly improved compared to OSS.

**TECHNICAL DETAILS**

ChubaoFS consists of a **metadata subsystem, a data subsystem, and a resource manager**. Offering both object and file storage, ChubaoFS offers strong replication consistency and is particularly well-optimized for quickly handling small files — another favorable trait for supporting cloud native workloads.

In a POSIX-compliant distributed file system, the behavior of serving multiple processes on multiple client nodes should be the same as the behavior of a local file system serving multiple processes on a single node with direct attached storage. ChubaoFS provides POSIX-compliant APIs.

