

Распределенная транзакционная версионированная web-ориентированная файловая система Djarvir

Даниил Подольский, Git in Sky



Постановка задачи, или откуда вообще взялась такая идея

- десятки терабайт данных, сотни миллионов файлов
- экономные заказчики - commodity hardware и хостер на букву X
- WEB контент - потребность в атомарном обновлении
- Отказоустойчивость - это не когда хранится, а когда отдается
- Проблема резервного копирования

Пара слов о существующих решениях

- Дорого
- Медленно
- Плохо работает
- POSIX не нужен

Подробности реализации: Распределенность

Проблемы:

- Проблема файлового кеша
- Проблема отказоустойчивости
- Проблема репликации
- Проблема ребалансинга
- NoSQL кластер - наше все

Подробности реализации: Распределенность

Решения:

- Почему NoSQL
- Как мы выбирали базу
- Почему Aerospike
 - Да, Aerospike нынче OpenSource и бесплатен
- 3.1.2.3. Почему Cassandra
- 3.1.2.4. Составные индексы
- 3.1.2.5. Транзакции

Подробности реализации: Транзакционность

Проблемы

- Атомарное обновление
- Откат
- Уровень изоляции

Подробности реализации: Транзакционность

Решения

- примерно как на симлинках, только без симлинков
- создание файла
- удаление файла
- commit и rollback как процессы
- самописная система транзакций – это весело



Подробности реализации: Версионирование

Проблемы

- Сисадмины бывают двух сортов, или проблемы бекапа

Подробности реализации: Версионирование

Решения

- Раз есть транзакции - их можно использовать как точки восстановления
- На самом деле, транзакции ни при чем. Просто мы храним несколько версий файла с одним именем

Подробности реализации: WEB-ориентированность

- Прямой доступ по полному пути
- Иерархические файловые системы не нужны

Подробности реализации: Файловость

- FUSE как интерфейс
- Совмещение абстракций “транзакция” и “точка восстановления” с абстракцией “файл”
- На самом деле - файловая система не нужна

Под капотом: Проблема больших файлов

- Chunks
- Bunches

Под капотом: Кое что задаром

- Дедупликация
- Сжатие на лету
- Экономия канала

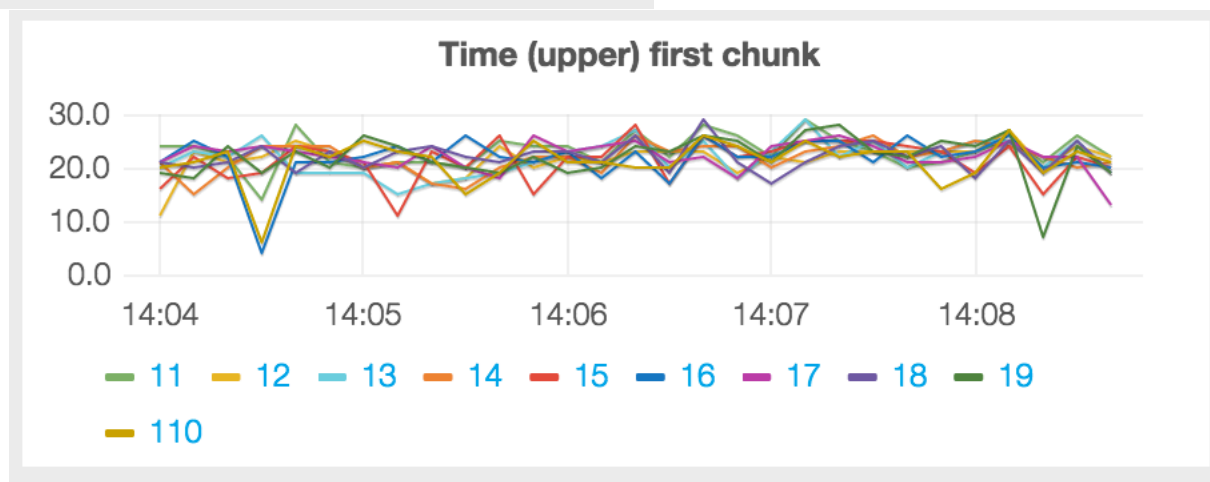
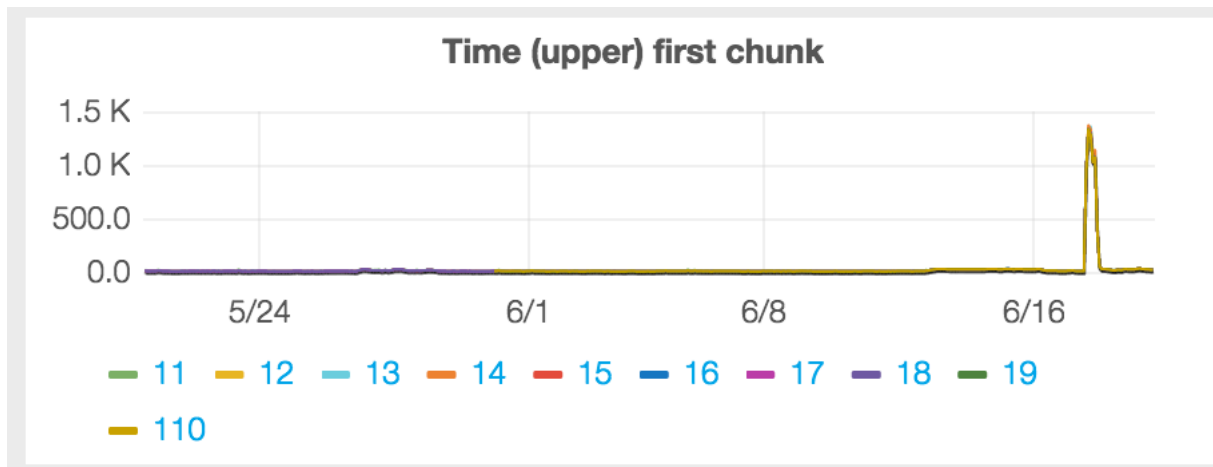
Под капотом

- Двухфазное удаление
 - удаление как таковое не нужно

Под капотом

Transaction keeper как единая точка отказа

Коротко об опыте эксплуатации



Где скачать

- <https://github.com/Djarvur/djarvurfs>

Djarvur / djarvurfs Unwatch 1

Distributed versioning transactional web-oriented file system — Edit

2 commits 1 branch 0 releases 1 contributor

branch: master djarvurfs / +

readme update	
onokonem authored 3 hours ago	latest commit 102e49815b
LICENSE	Initial commit 3 hours ago
README.md	readme update 3 hours ago

README.md

djarvurfs

Distributed versioning transactional web-oriented file system

will be here soon

Вопросы?