Overlayfs

Current status and upcoming future

Author/ Email: Yi Zhang / yi.zhang@huawei.com
Version: 2018.6

HUAWEI TECHNOLOGIES CO., LTD.

Overlayfs

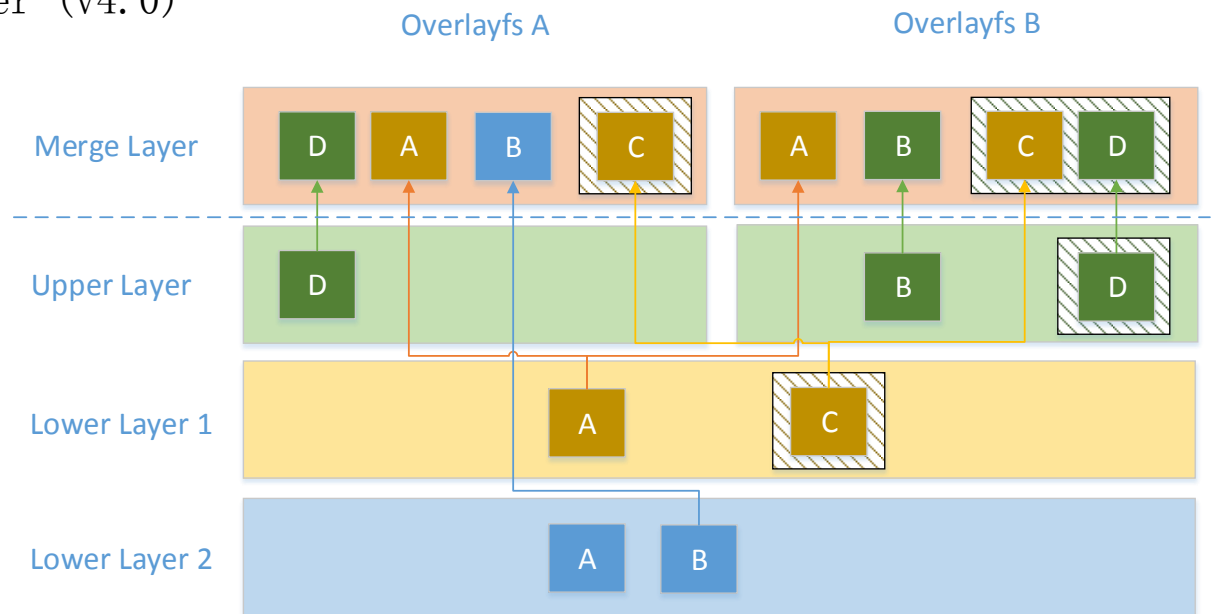- Introduction

  - What is overlayfs ?

  - Overlayfs use cases

  - Layers

- Current status

- Upcoming future

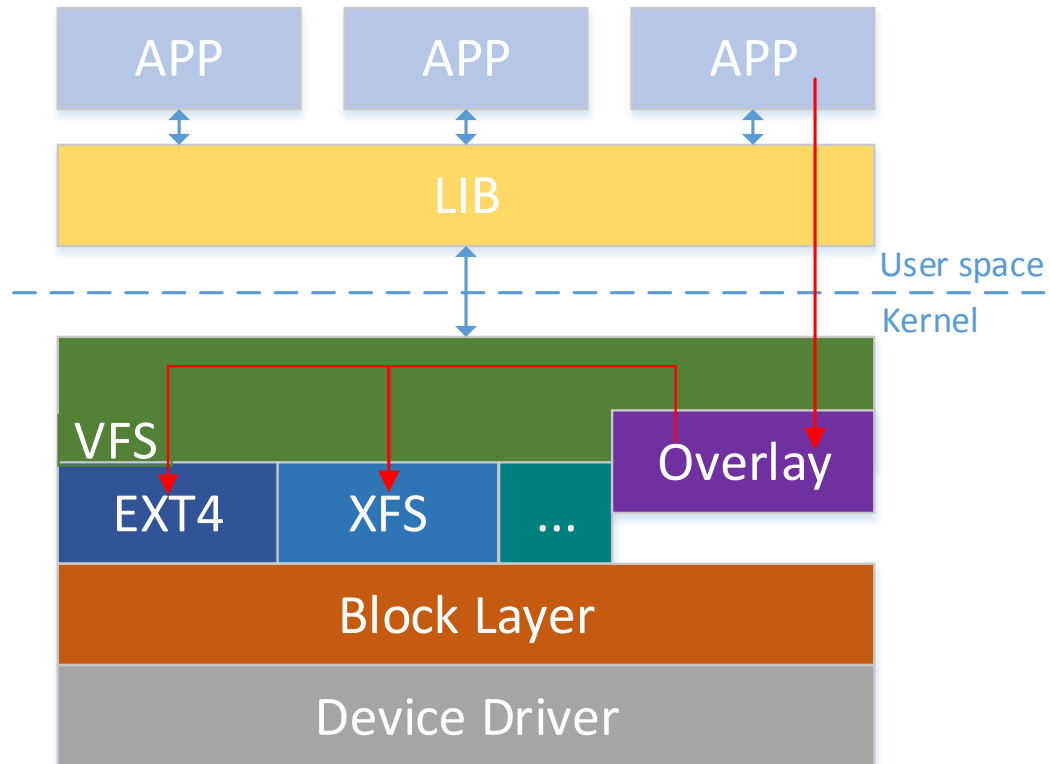HUAWEI

Overlayfs - Introduction

- What is overlayfs ?
    - OverlayFS is a modern union fs implementation for Linux (v3.18)
    - Cover up file in lower dir
    - Merge dir
    - Copy up
    - Mulit lower layer (v4.0)
- Use cases
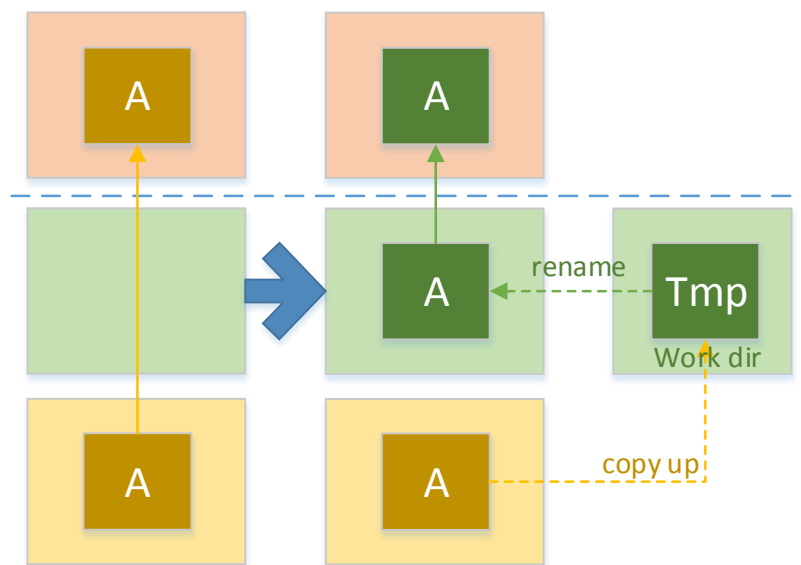   - Docker
    - …

Overlayfs - Layers

Overlayfs

- Introduction
- Current status
  - Basic features: copy up, whiteout, opaque dir.
  - New features: redirect dir, clone up, concurrent copy up, consistent st_ino & st_dev & d_ino, index, nfs export, consistent fd, delayed copy up.
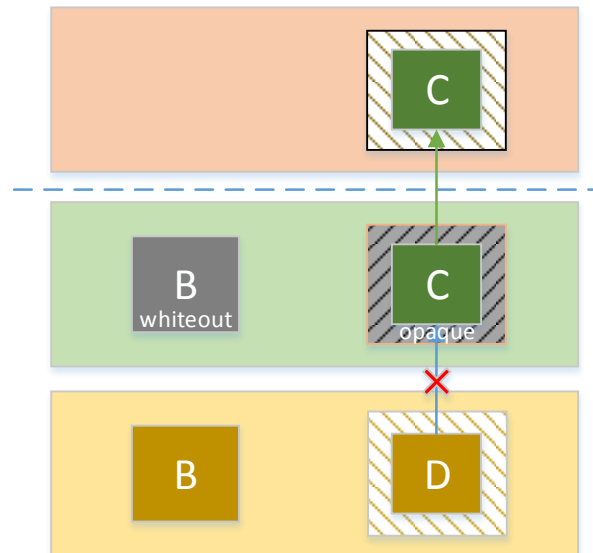- Upcoming future

HUAWEI

Overlayfs - Basic features

- Copy up
  - Copy when open for write
  - Copy the entire file to upper layer
  - Inconsistent fd, st_dev & st_ino
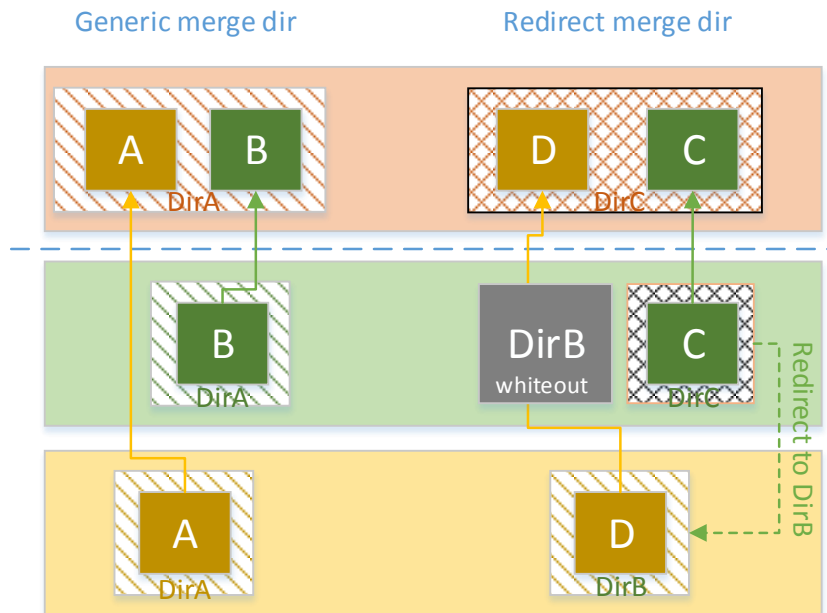 - Hard link breakup

Overlayfs - Basic features

- Whiteout
  - Cover a deleted object in lower layer
  - Requires underlying fs support dtype and exchange rename

- Opaque dir
  - New dir cover deleted object in lower
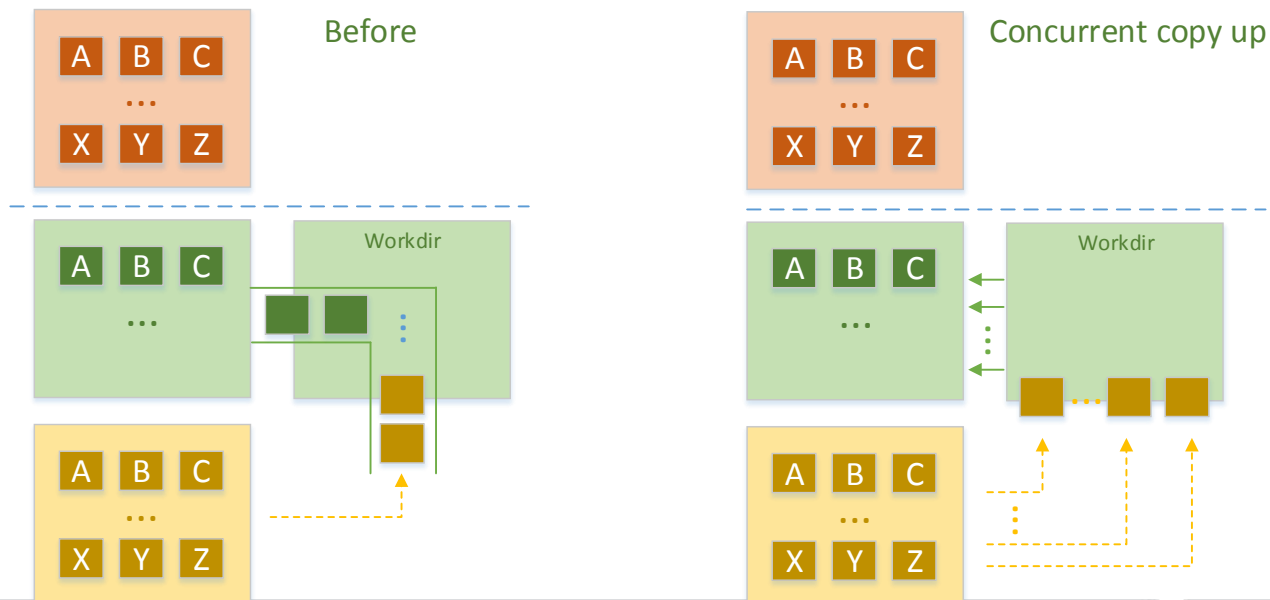 - Save opaque xattrs in upper object

Overlayfs – Redirect dir feature

- Redirect dir (Since v4.10)
  - Support dir rename which contain dirs from lower layer
  - Save origin path in redirect xattr on upper dir
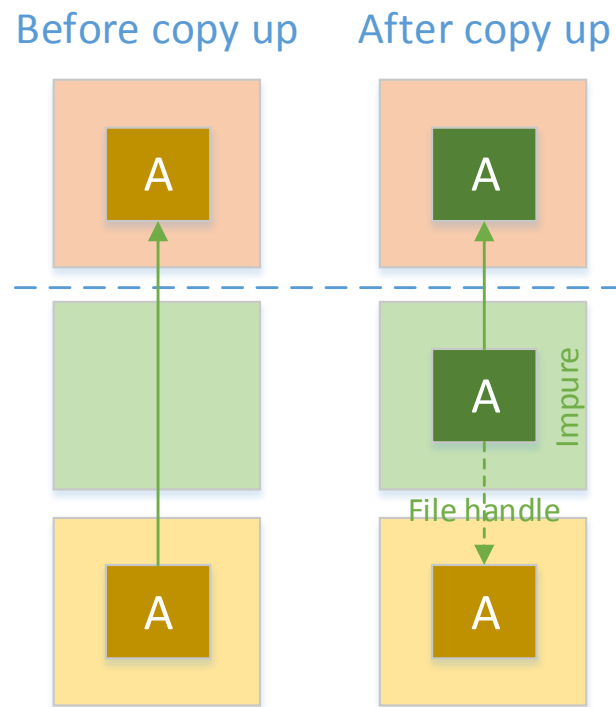 - Lookup by redirect path

Overlayfs – Clone up & Concurrent copy up

- Clone up (v4.10)
  - Requires underlying fs support reflink (xfs, btrfs)
  - Lower layers and upper on the same file system
- Concurrent copy up (v4.11)
  - Base on tmpfile in underlying fs
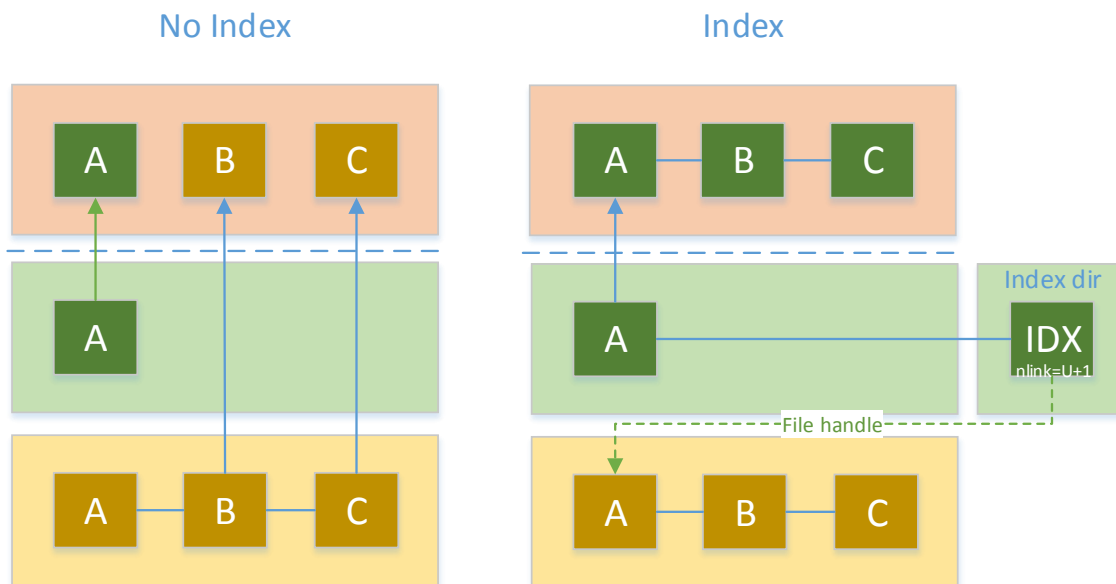 - Relax global copy up lock

Overlayfs – Consistent st_dev & st_ino & d_ino

- Origin Feature (Since v4.12)
  - Requires underlying fs support file handle
  - Use overlay st_dev and origin st_ino
- Impure dir Feature (Since v4.12)
  - Dir may contain copied up object
  - Save impure xattr on upper parent dir
- Xino Feature (Since v4.17)
  - Require underlying fs has enough unused bits in inode number (ext4, xfs)
  - Combine lower fsid and origin st_ino

Before copy up   After copy up

A   A
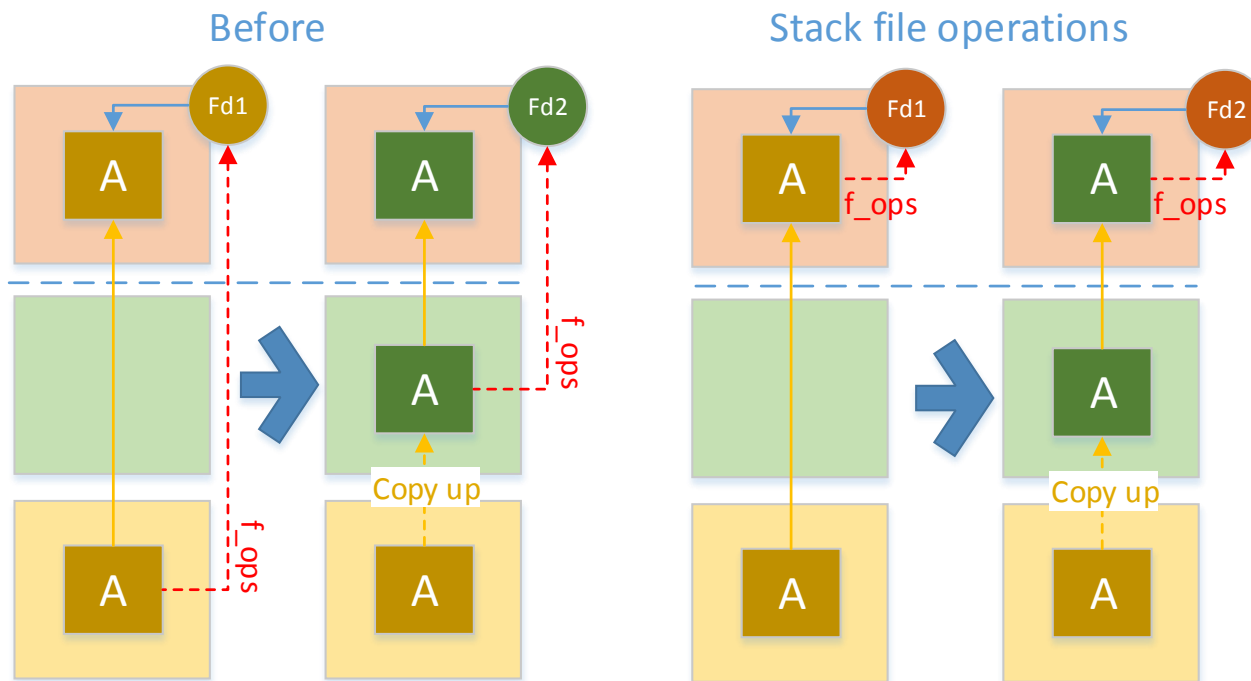
A

Impure

File handle

A   A

Overlayfs – Index (avoid hard link breakup)

- Index feature (Since v4.13)
  - Requires underlying fs support file handle
  - Copy up as index file
  - Save file handle and nlink xattr in index
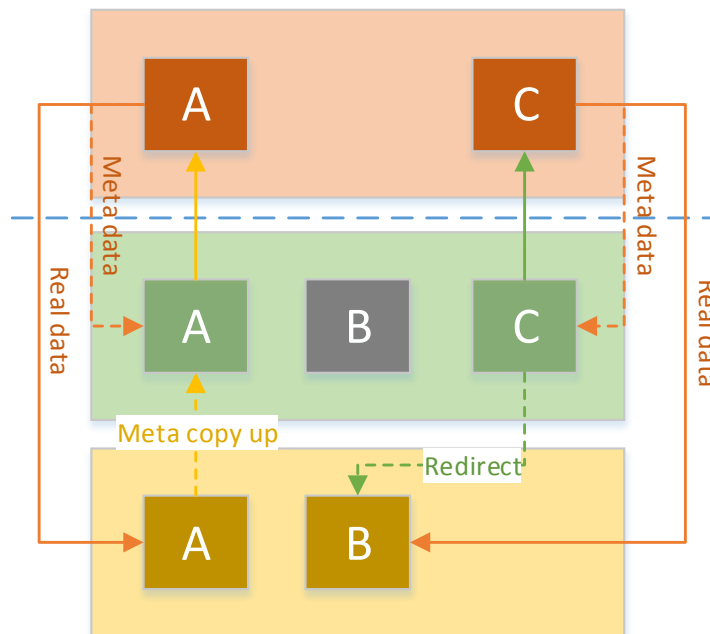 - Link index to upper target

No Index

Index

| A | B | C |

| A | B | C |

Index dir

| A |

| A | IDX nlink=U+1 |

File handle

| A | B | C |

| A | B | C |

HUAWEI

Overlayfs – Constant file descriptor

- Stack file ops (v4.18 ?)
  - Make overlay file struct to unify to overlay fd
  - Find real file and switch to operate the real underlying file

Before

Stack file operations

Overlayfs – Delayed copy up

- Delayed copy up (metadata copy up, v4.18 ?)

  - Copy up metadata only if no data write (chattr, chmod, rename...)

  - Save meta copy xattr on upper file

  - Look up real data by name or redirect xattr
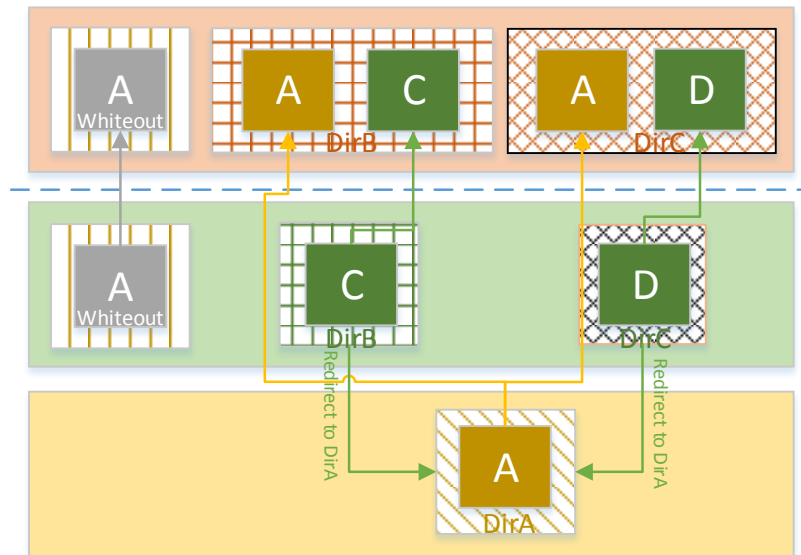
  - Multi metadata copied up files

Overlayfs

- Introduction

- Current status

- **Upcoming future**

  - Overlay Feature set

  - Offline layer check tool

  - Merge file (Partial data copy up)

Overlayfs – Upcoming future

- Overlay feature set (work in progess)
  - Enable features by mount options
  - Cannot disable some features completely once it was enabled now
  - Mark compatible, ro-compatible and incompatible feature set
    - Compatible: origin, impure dir
    - Ro-compatible: index, nfs_export
    - Incompatible: redirect dir, meta data copy up
  - Refuse to mount if unsupported feature detected

HUAWEI

Overlayfs – Upcoming future

- Offline layer check tool (RFC)
  - Overlayfs-progs: fsck.overlay
  - Auto check and fix underlying dir inconsistency
    - Invalid whiteouts: whiteout exposure, fail to remove dir...
    - Invalid/duplicate redirect xattr: merge with the wrong dir...
    - Missing impure xattr: inconsistent d_ino...
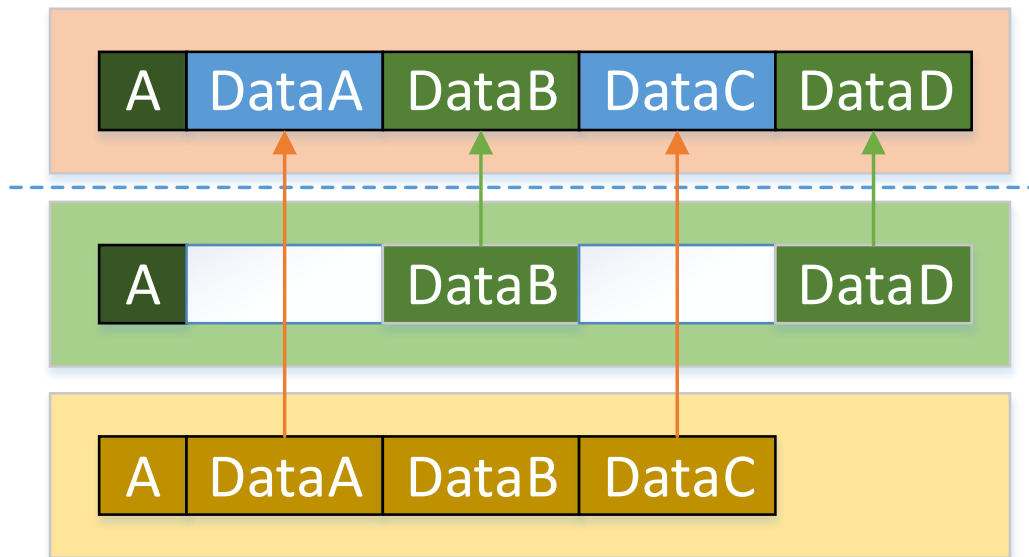   - Test cases: xfstests

Overlayfs – Upcoming future

- Offline layer check tool: Todo
    - Auto fix the missing feature set
    - Detect the overlayfs is already mounted
    - Check and fix origin xattr, index object, metadata copy up
    - Set or clear the indicated overlay features
    - Release the first version!

Overlayfs – Merge file (Partial data copy up)

- Merge file (work in progress)
  - Copy up across fs (or on the fs not support reflink) is time-consuming and waste of space now
  - Copy up metadata and partial data blocks instead of whole file
  - Create data map and merge data blocks between each layer

Thank you

www.huawei.com

Huawei OS Kernel Lab

Huawei Operating System R&D Department

- OS Kernel Lab

➢ Linux Kernel (ARM/x86/ heterogeneous platforms) R&D and Innovation

➢ R&D on a Next-generation OS kernel with Low Latency, High Security, Strong Reliability, Intelligence, etc.

Job Vacancy

Next-generation Operating System Researcher and Senior Engineer

Formal Verification Researcher and Senior Engineer

Linux Kernel Architect and Senior Engineer

Locations

Hangzhou, Beijing, Shanghai

Contact us

Tel: Mr. Wang/18658102676

Email：hr.kernel@huawei.com

HUAWEI