

# Intro: Current State of containerd

Wei Fu (Alibaba Cloud)



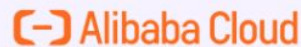
# What is containerd?

- Container Runtime
  - Below platforms (Docker, Kubernetes)
  - Above lower level runtimes (runC, Kata, Firecracker, gVisor)
- Resource Manager
  - Container processes
  - Image artifacts
  - Filesystem snapshots
- Smart Client Model



ecosystem

Platform



Client



kubelet



CRI Runtime



Container Engine



containerd client



BuildKit



containerd client



ctr



containerd client

containerd

API



CRI



containerd client



containerd



Service Handlers



Prometheus

Metrics

Core

Services

Containers Service

Content Service

Diff Service

Images Service

Leases Service

Namespaces Service

Snapshots Service

Tasks Service

Metadata (namespaced)

Containers

Content

Images

Leases

Namespaces

Snapshots



Backend

Content Store

plugin

local

Snapshotter

overlay

btrfs

devmapper

native

windows

plugin

Runtime

v2 shim client

trtpc



containerd-shim



runc



runhcs



kata



Firecracker



gVisor



shim

system



# Who uses containerd?

- Public Clouds
- Kubernetes Infra
- End Users
- DevOps Tools
- Custom Sandboxes



# containerd matures

- 5th project to graduate within the CNCF - February 2019
- Broad support and contribution from across the ecosystem
  - Over 200 individual contributors; represent > 100 companies
  - 13 maintainers represent 9 different companies
- All major cloud providers using containerd
- Supports Linux and Windows platforms
- Added sub-projects to governance



# Release 1.4 - Preview



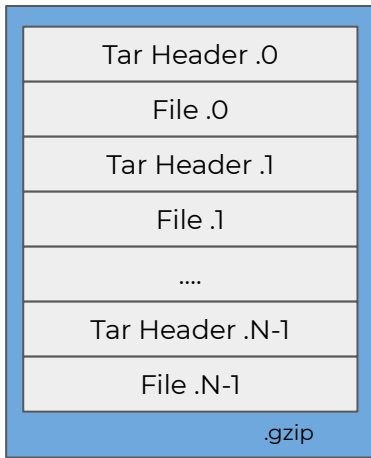
CLOUD NATIVE + OPEN SOURCE

*Virtual Summit China 2020*



# stargz-snapshotter

- Facing issue with current Image Spec
  - Tar ([Tape Archive](#)) is not seekable data
  - Deduplicate on file level
  - Build from distribution image
    - Most files are for operation if debugging



Transitional tar.gz



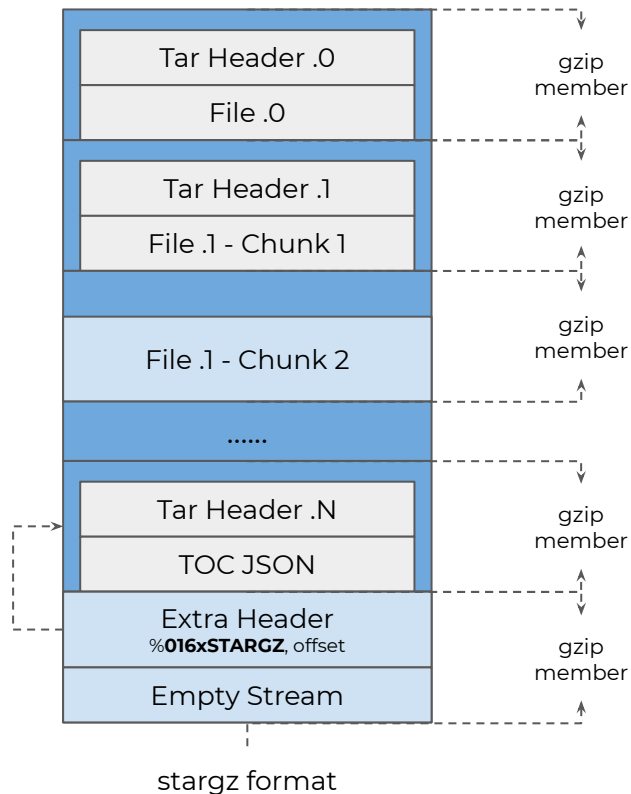
CLOUD NATIVE + OPEN SOURCE

*Virtual Summit China 2020*



# stargz-snapshotter

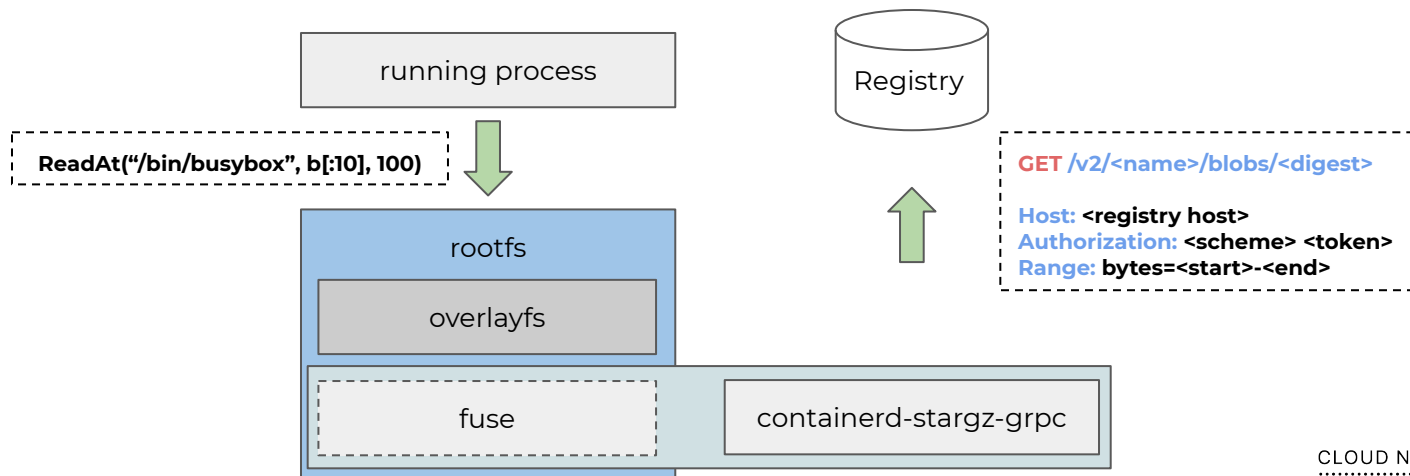
- [stargz](#) - a seekable tar.gz format
  - A tar entry as a gzip member
    - Chunked into several members if file is large
  - TOC(Table of Contents) as index
    - Record offset for each file
  - Last gzip member is empty
    - Extra header stores offset of TOC
    - Only 47 bytes
  - Compatible with tar.gz





# stargz-snapshotter

- stargz-snapshotter is sub-project of containerd
  - Maintained by [Kohei Tokunaga](#)
  - Out-of-tree snapshotter plugin - fuse-overlaysfs
    - Initialize TOC JSON with last two gzip member of each layer
    - Read content by **on-demand** http request



# stargz-snapshotter

- As Remote Snapshotter
- OCI [Image Spec v2 Discussion](#)

# shimV2 as default runtime

- Legacy shim only supports runC command line
  - New runtime, like sandbox runtime, must support runC sub-commands
    - *Introduce new abstract layer between shim and runtime*
  - Can't clean up resource if shim is down unexpectedly



# shimV2 as default runtime

- A first class shim API for runtime authors to integrate with containerd
  - Required start/delete sub-commands
    - Resources created by container, including mount points, will be cleanup by delete sub-command
  - Support pluggable logging via URL
  - Manage container(s) by one shim
    - Per-pod container shim
- runc.v2 as CRI plugin's default runtime

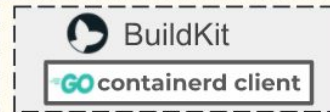


ecosystem

Platform

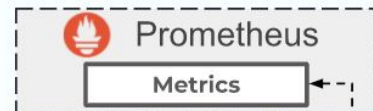
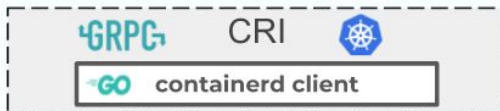


Client

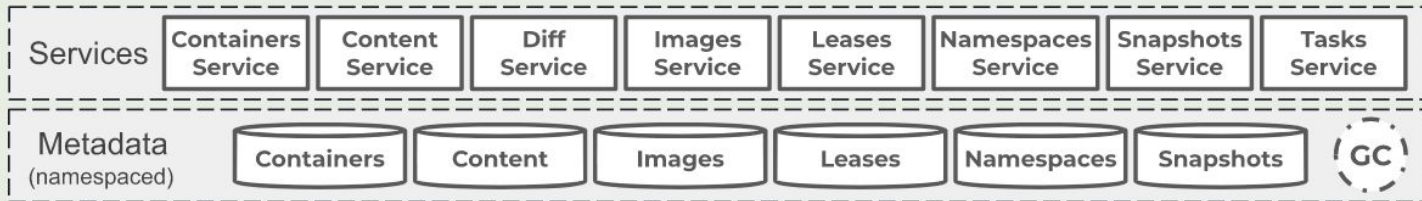


containerd

API



Core



Backend



system



# Support cgroup v2

- Support most controllers
  - pids/ cpu, cpuset/ memory/ freezer
  - OOM event
  - rmda
  - hugetlb
  - io
  - device
- Integrated with CRI plugin (cpu/memory metrics)
- Provide **cgctl** debugging tool



# What is next?



CLOUD NATIVE + OPEN SOURCE

*Virtual Summit China 2020*



# Next

- Introduce sandbox API - [#4131](#)
- Merge CRI plugin into upstream - [containerd/cri#1479](#)
- Mountable resource manager





Thank you



CLOUD NATIVE + OPEN SOURCE

*Virtual Summit China 2020*

