Sponsored by Red Hat

Pulp & Operate-First

Gerrod Ubben
Pulp Dev

pulpproject.org

pulp

Agenda

- Overview of Pulp
- Pulp's Introduction to Operate-First & Pulp Operator
- Feedback/Discussion on Experience with Operate-First

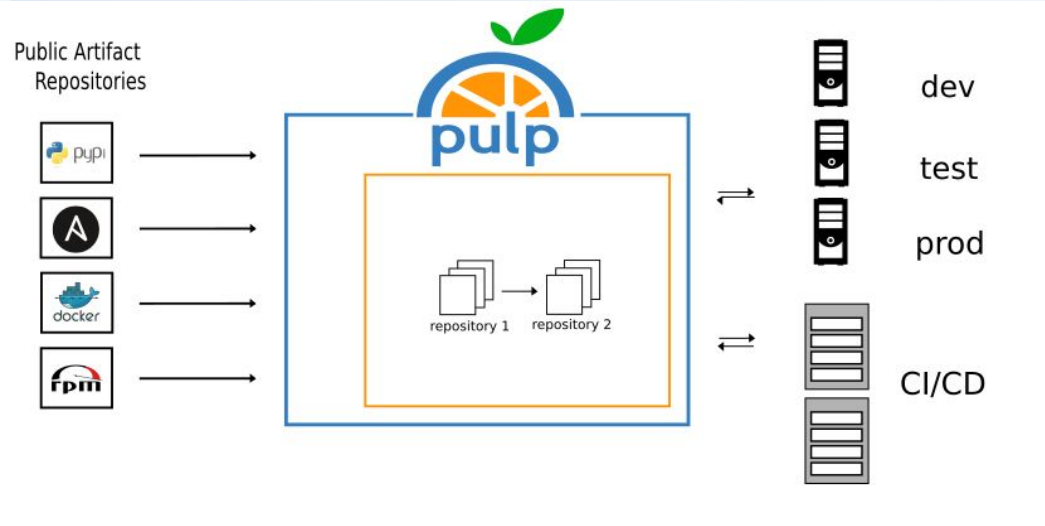
Sponsored by Red Hat

Intro to Pulp

pulpproject.org

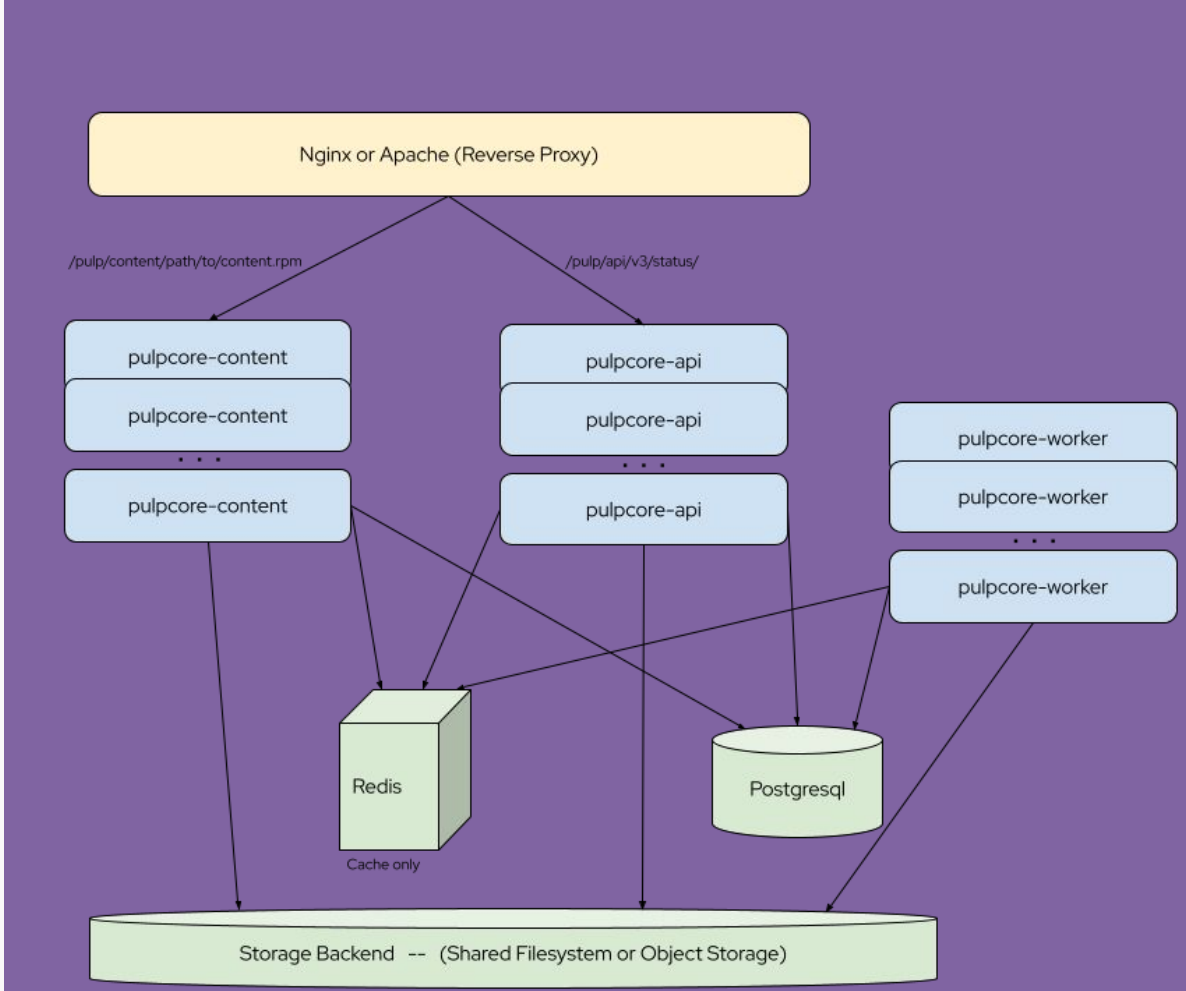
What is Pulp?

- Pulp is a Software Repository Manager allowing you to curate and distribute hundreds to thousands of packages from multiple different content types and sources.



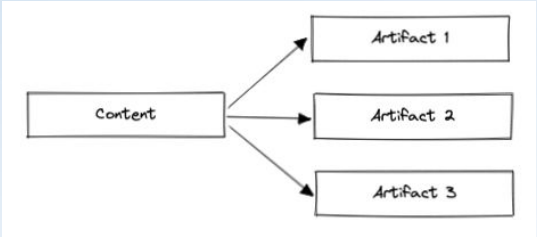
Architecture

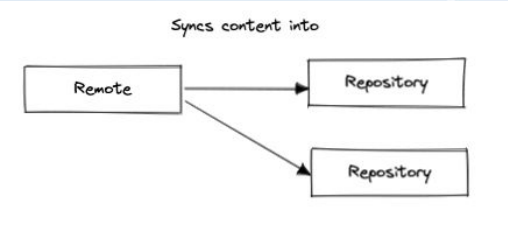
- Pulp is a Django project with a REST API and an aiohttp web server
- Artifacts are stored on a filesystem or in a storage bucket
- Task workers handle long running operations
- Postgres DB is the single source of truth for all of Pulp

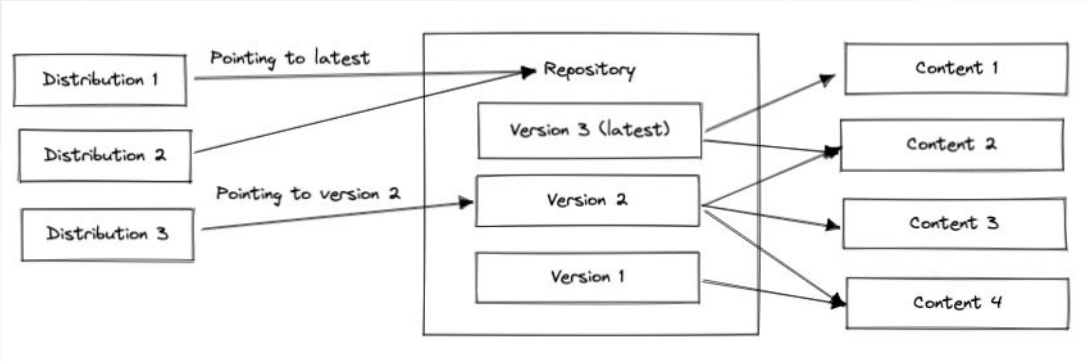


Pulp Objects

- Content - representation of content stored in Artifacts
- Repositories - collection of content, versioned by add/removal diffs
- Remotes - config for external repositories to sync content from
- Distributions - exposes the content in a repository to the public







Sponsored by Red Hat

Pulp Development

- Plugin based architecture with each content type getting its own plugin with its own GitHub repository and package versioning.
- Upstream project used in many Downstream products by Red Hat.
  - Satellite
  - Foreman Katello
  - Ansible Automation Hub
- Have mainly been an on-premise Ansible Playbook deployment project, but we are now switching over to be container first (only) and cloud first.

pulp

Sponsored by Red Hat

Pulp's Introduction to Operate-First

pulpproject.org

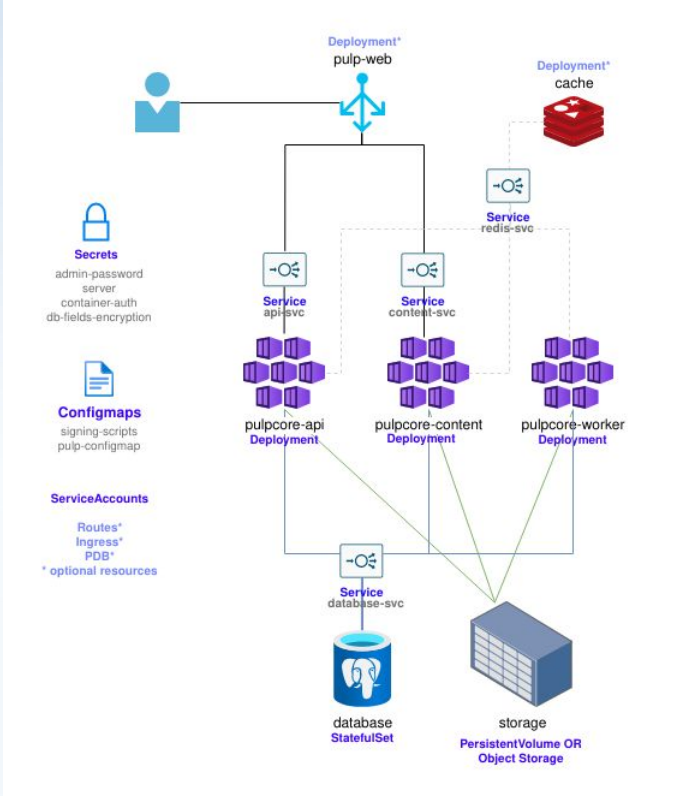
pulp

A Need for an Internal PyPI

- Around March 2021 there was email inquiring for hosting an Internal PyPI by Red Hat. Pulp with its python plugin was suggested to fill this role.
- With the nature of many teams projects being upstream and open source it was decided that the instance should live on Operate-First.
- Goal was to create an index used by the OpenStack and AI/ML communities.
- Thus began our journey to create a Pulp instance on Operate-First infrastructure: https://github.com/operate-first/support/issues/176

Pulp Operator

- When we began this journey we were just getting started with developing our operator.
- The onboarding process provided valuable feedback in helping guide our development and squash bugs within the operator.
- Now we are currently rewriting the operator in Go: https://discourse.pulpproject.org/t/rewriting-pulp-operator-in-go/561



The Deployed Instance

- After 10 months we finally got a Pulp instance up and running in Operate-First:
  - https://pulp.operate-first.cloud/pulp/api/v3/status/
  - Many thanks to all on the Operate-First team!
- We announced the new instance allowing teams to open an issue to create a Python index for them
  - https://developers.redhat.com/articles/2022/01/17/how-self-host-python-package-index-using-pulp#managing_software_repositories_with_pulp

Sponsored by Red Hat

Experience on Managing the Instance

pulp

An Empty Instance

- The Pulp instance is currently unused, not a single object created.
- Hard to have any experience managing the instance when no uses it.
- I've identified 3 areas that are current roadblocks preventing others from using this instance:
  - Lack of Users & Usecases
  - Potential Insufficient Storage
  - No/Little Documentation and Advertisement

No Users and Usecases

- Currently only one user, admin. Every REST endpoint requires permission so nobody can create & curate content. User creation has to be done manually by the admin.
  - Pulp supports external authentication, with tested support for RH-SSO and Keycloak.
  - **What authentication system does other Operate-First services use? What would be best to allow more users to join and use Pulp?**
- Current lack of usecases besides setting up a Python Package index.
  - Pulp can easily be used in many project's CI/CD pipelines as an artifact cache or repository.
  - **Is there any Operate-First usecases that we should look to explore?**

Potential Trouble with Storage

- Currently the instance has around 24 GB of storage available.
  - This severely limits usecases of Pulp with many common Deb/RPM/Container repositories being multiple gigabytes in size.
  - There are endpoints for reclaiming space, but frequent usage defeats the purpose of having a Pulp Instance.
- We support cloud object storages AWS S3, Azure Blob and soon Google Cloud.
  - **Is there an Operate-First cloud object storage instance?**
  - **Are there other services that use lots of storage? If so how do they tackle this problem?**

Help with Documentation and Visibility

- The Pulp Operate-First instance is currently not actively advertised, so it is not surprising that it has no users.
- A link in our “Host a Python Index” article links to a 404:
  - https://www.operate-first.cloud/community-handbook/pulp/usage.md
- **What areas (documentation/community rooms) would be best to work on to help better expose Pulp to the Operate-First Community?**

Questions?

https://discourse.pulpproject.org/

Come chat with us on Matrix #pulp & #pulp-dev