CHINA
OpenInfra Days

CHINA
OpenInfra Days
IT大咖说
知识共享平台

Trove - OpenStack DBaaS 组件架构分析

议程/Agenda

- 关于/About
- Trove 发展历程/Trove history
- Trove 架构分析/Trove architecture
- Trove 当前开发情况/Trove current status

关于海云捷迅/About AWCLOUD

公司愿景:致力于简化企业上云的进程

企业级云服务提供商

8年专注私有云方向

国内最早的OpenStack云服务提供商

自主知识产权

基于开源架构
OpenStack
Kubernetes
CEPH

完善的云生态

融合Intel创新技术
腾讯云混合云
20家云生态合作伙伴

广泛客户覆盖

销售，支持和服务超过100个客户
5000台服务器
30000万台虚拟机

研发能力

OpenStack社区贡献
AI创新平台
技术创新中心

关于海云捷迅/About AWCLOUD

私有云/混合云产品

为企业用户提供稳定可靠的IaaS云管理平台。使企业能够以最小的初始成本快速实现IT基础设施的“云化”，并实现“积木堆叠式”的弹性扩容，按需升级。

超融合一体机

超融合一体机是一套基于融合架构的IT基础设施平台。使用超融合一体机，企业新业务上线周期可从数月缩减到数小时，使得运营效率大幅提升、远超预期。

容器云和AI云产品

为企业客户提供基于K8S的容器云管理平台，帮助企业实现业务的敏捷交付，持续集成。同时提供AI云产品，可以帮助客户管理GPU集群，利用容器快速构建深度学习所需环境。

技术服务

为企业客户提供“一站式”的技术服务，包括企业上云咨询，定制解决方案，实施，运维，技术培训等。

Trove 发展历程/Trove history

项目基本情况（What is Trove）：

- **基于 OpenStack 的数据库服务**
  OpenStack Database service
- **可扩展性、可靠性，支持关系型和非关系型数据 库引擎**
  To provide scalable and reliable Cloud Database as a Service functionality for both relational and non-relational database engines, and to continue to improve its fully-featured and extensible open source framework.
- **主要特性（main features）**：
  - 数据库引擎管理（datastore & datastore version）
  - 基础数据库实例生命周期管理（basic database instance lifecycle）
  - 配置管理（configuration）
  - 备份管理（backup & restore）
  - 数据复制（replications）
  - 集群（clustering)

Trove 发展历程/Trove history

- **起源于 Rackspace 内部项目（2011），HP 数据库团队也参与早期开发**
  Started by Rackspace(2011), the HP database team also helped a lot.
- **成为 OpenStack 孵化项目，Grizzly & Havana**
  Incubated during Grizzly and Havana
- **成为 OpenStack 正式项目，Icehouse**
  Integrated in Icehouse
- **早期主要开发人员来自 Rackspace 、GlobalLogic 、HP 、Tesora 、eBay ，Icehouse ~ Kilo**
  Initial contributions mainly from Rackspace, GlobalLogic, HP, Tesora (from Icehouse to Kilo).

Trove 发展历程/Trove history

- **从 Juno 开发，由 Tesora 主导，其他贡献者主要来自：HP 、HPE 、eBay 、Red Hat 、IBM 、Mirantis 、AT&T 、EasyStack 、SUSE 等**

  Begin with Juno, most contributions come from Tesora, other contributors are from HP, HPE, eBay, Red Hat, IBM, Mirantis, AT&T, EasyStack, SUSE, etc.

- **在 Pike 周期，Tesora 淡出社区参与，IBM 投入开发人员，Queens 结束，主要贡献者转移到中国，分别来自海云捷迅、中国移 动、中国电信、易捷思达**

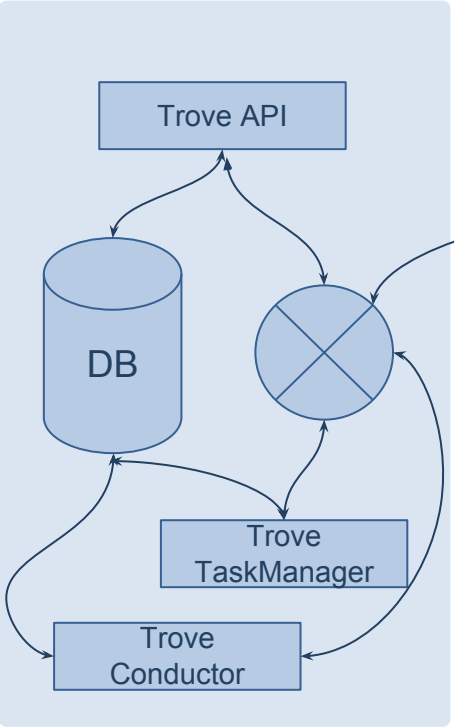
  Tesora moved on during Pike, leadership transferred to IBM, and after Queens, mostly contributors are from China(Awcloud, China Mobile, China Telecom, EasyStack).

Trove 发展历程/Trove history

实际部署情况（Trove adoptions）：

- **Rackspace Cloud Databases**
  https://developer.rackspace.com/docs/cloud-databases/v1/
- **HP Cloud**
  https://www.slideshare.net/tesoracorp/5-hp-presentation-final
  https://www.tesora.com/deploying-trove-public-cloud-hp/
- **eBay**
  https://www.slideshare.net/tesoracorp/4-open-stack-trove-day-ebay-final
  https://www.stratoscale.com/blog/dbaas/deploying-openstack-trove-dbaas-ebay/

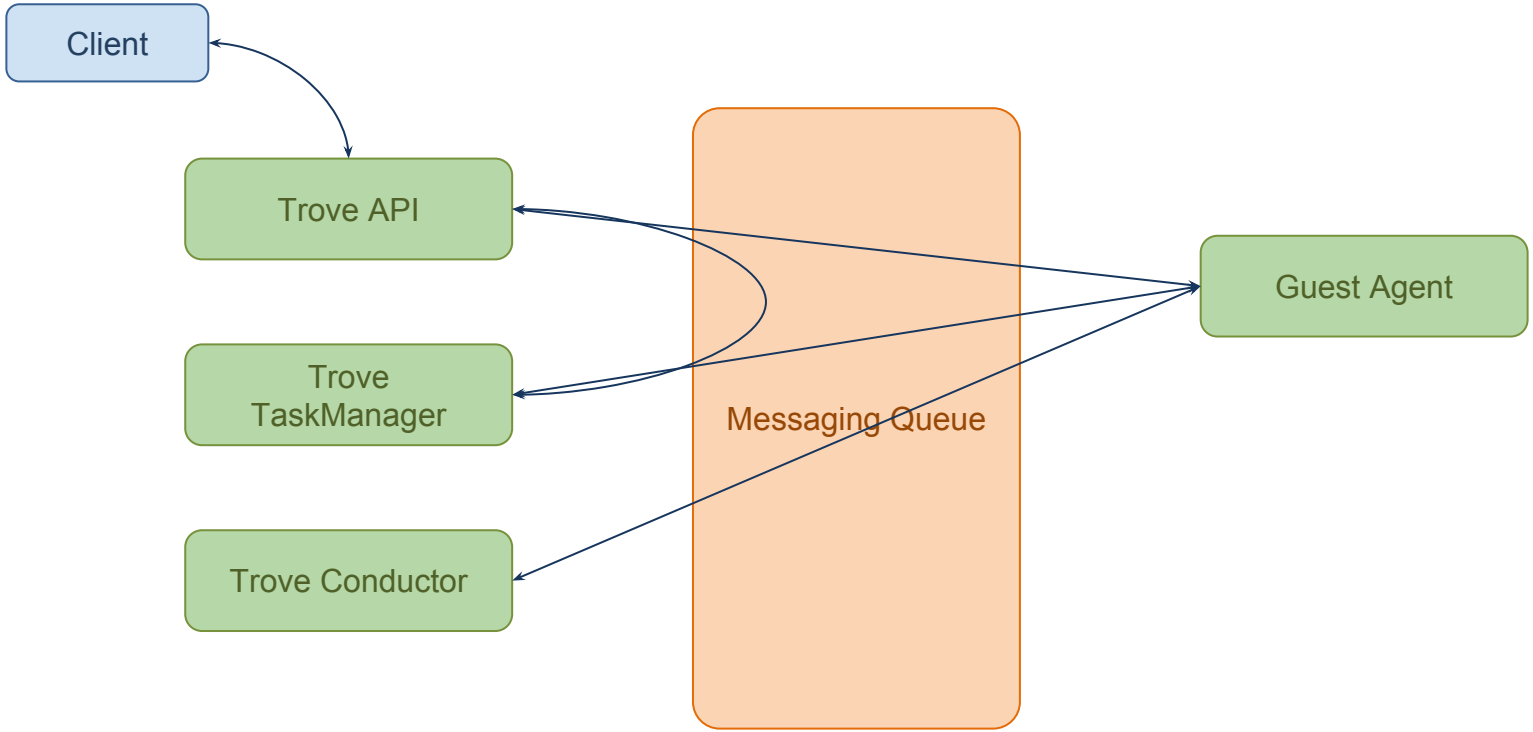
Trove 架构分析/Trove architecture



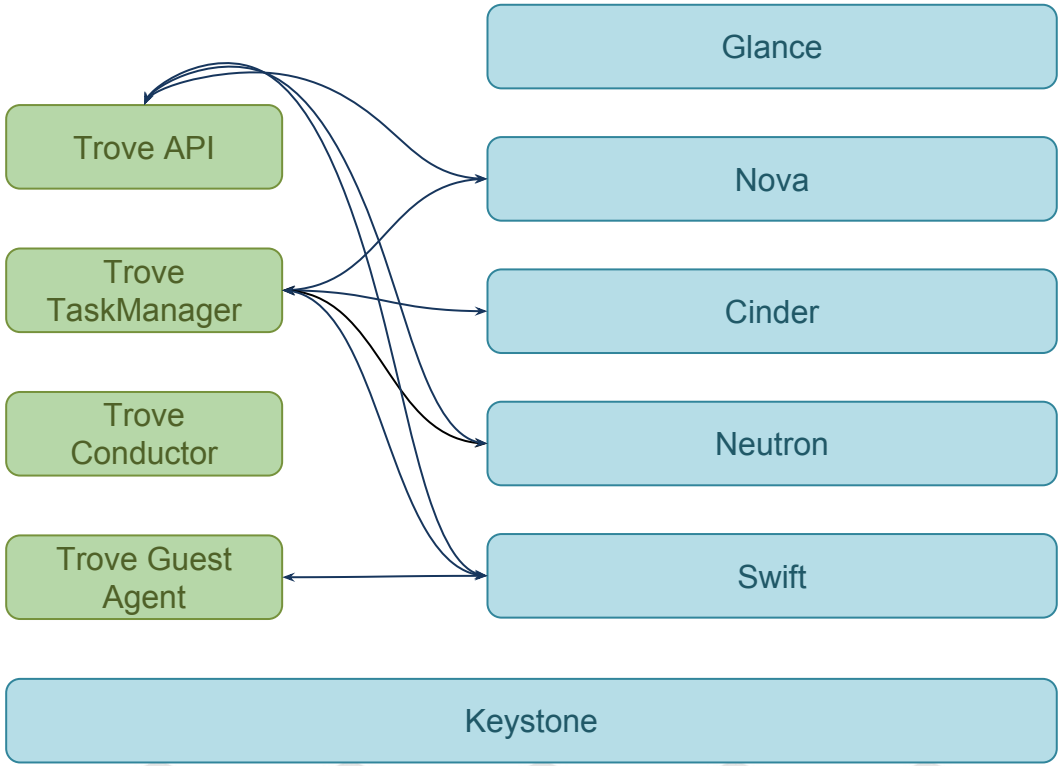
DataStore Image	Glance Image
Guest Agent	Nova Compute
Data Volume	Cinder Volume
Network Connection	Neutron Network
Backups/Logs	Swift Objects

Keystone Auth

Trove 架构分析/Trove architecture



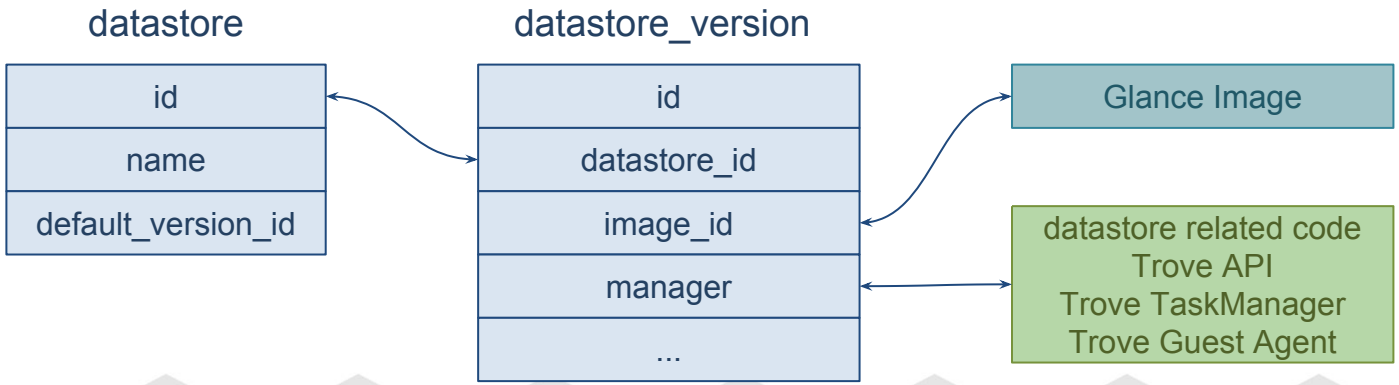
Trove 架构分析/Trove architecture



Trove 架构分析/Trove architecture

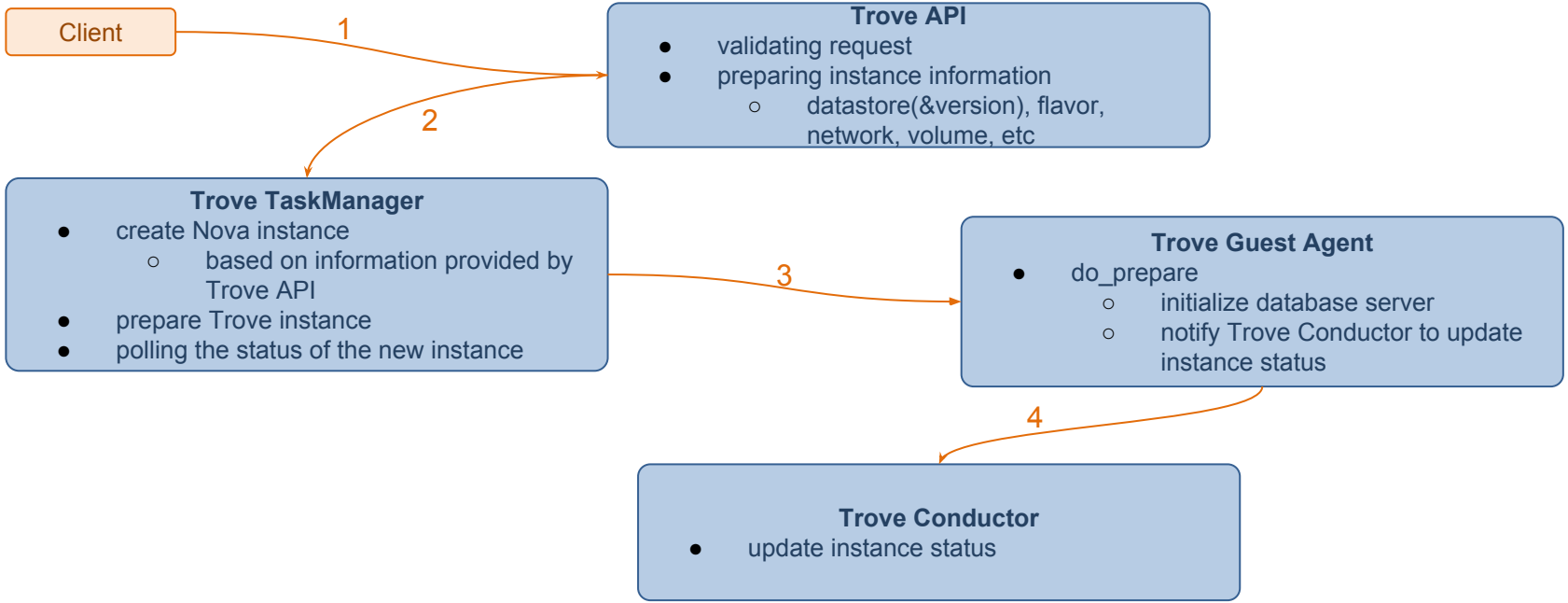
创建数据库实例/creating trove instances

- **datastore**
- **datastore version**
  - 和 Glance 镜像对应（mapping to Glance image）
  - 和 Trove 代码中的数据库引擎对应的配置项以及代码对应（mapping to trove datastore manager code）



Trove 架构分析/Trove architecture

创建数据库实例/creating trove instances



Trove 架构分析/Trove architecture

创建数据库实例常见问题/common problems for creating trove instances

- **镜像/datastore images**
  - 操作系统（OS）
  - Guest Agent
  - Database engine
- **网络/network issues**
  - Guest Agent 与 Message Queue 之间的连接（connection between Guest Agent and underlying Message Queue broker）
- **其它（Other）**
  - Cinder Volume
  - Quota（Trove, Nova/Cinder/Neutron）

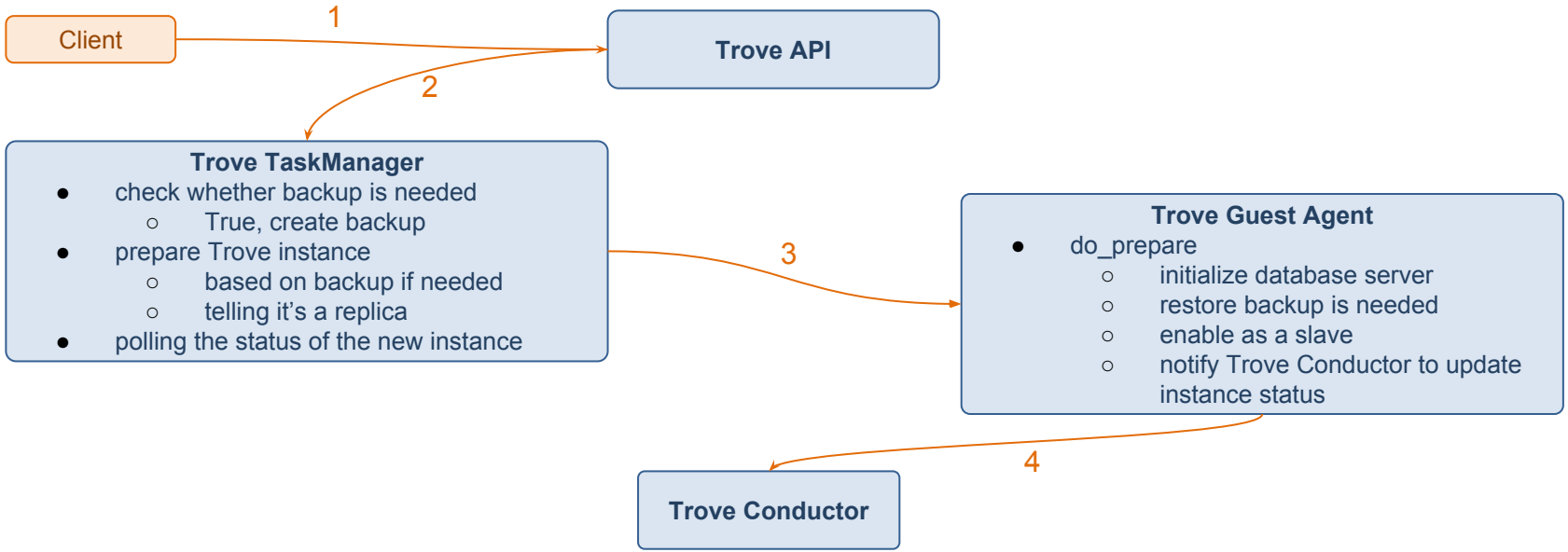
Trove 架构分析/Trove architecture

备份/backups

- **创建备份/Creating backups**
  - Trove API -> Trove TaskManager -> Trove GuestAgent -> Swift
- **基于备份创建新的实例/Backup restoring(creating new instances)**
- **Strategies**
  - backup
    - trove/guestagent/strategies/backup/
  - restore
    - trove/guestagent/strategies/restore/
  - storage
    - trove/common/strategies/storage/
    - 如何保存备份数据（how is backups stored）
    - 目前只支持 Swift（currently only Swift is supported）
- **incremental backup**
  - 目前只有 mysql/postgres strategies 支持（currently only supported by mysql/postgres_impl)

Trove 架构分析/Trove architecture

数据复制/replication



Trove 架构分析/Trove architecture

数据复制/replication

- **Strategies**
  - trove/guestagent/strategies/replication/
  - MySQL(mysql_binlog & mysql_gtid), MariaDB
  - PostgreSQL
  - Redis (backup not needed)
- **Replication v2**
  - https://specs.openstack.org/openstack/trove-specs/specs/kilo/replication-v2.html
  - promote_to_replication_source
  - eject_replica_resource

Trove 架构分析/Trove architecture

集群/Clustering

- **Strategies**
  - trove/common/strategies/cluster/experimental/
  - Cassandra, Galera, MongoDB, Redis, Vertica
  - API/TaskManager/GuestAgent all have different strategies
- **特殊操作/Special Actions**
  - grow/shrink/restart/upgrade/configuration_attach/configuration_detach
- **具体实现示例/MongoDB Cluster as Example**
  - 只支持创建包含 1 个 shard 的集群，replicaset 节点数固定为 3，config server 和 router 的数量由配置项控制

    Only support create a cluster containing 1 shard with 3-node replicaset(node number is hardcoded), the number of config servers and routers are controlled by configuration.
  - 支持 grow/shrink/add_shard

    Extra actions: grow, shrink, add_shard.

Trove 架构分析/Trove architecture

配置管理/Configuration

- **template**
  - trove/templates/<manager>/x.template
  - 支持不同的数据库引擎版本/versioning support
    - trove/templates/<manager>/x.x/
  - 由 Trove TaskManager 进行解析
    Rendered in Trove TaskManager
- **configuration (Trove API)**
  - 参数检查/validation rules
    - trove/templates/<manager>/validation-rules.json
    - 需要在初始化数据库引擎过程中手工导入到数据库
      Need to be loaded manually during datastoe version initialized.
  - 可以在创建实例时加载，也可以在实例运行过程动态加载（部份参数需要重启数据库服务才会生效）
    Could be attached when creating instances, or dynamicaly attached to running ones（some parameters may only be applied after the database reboots).

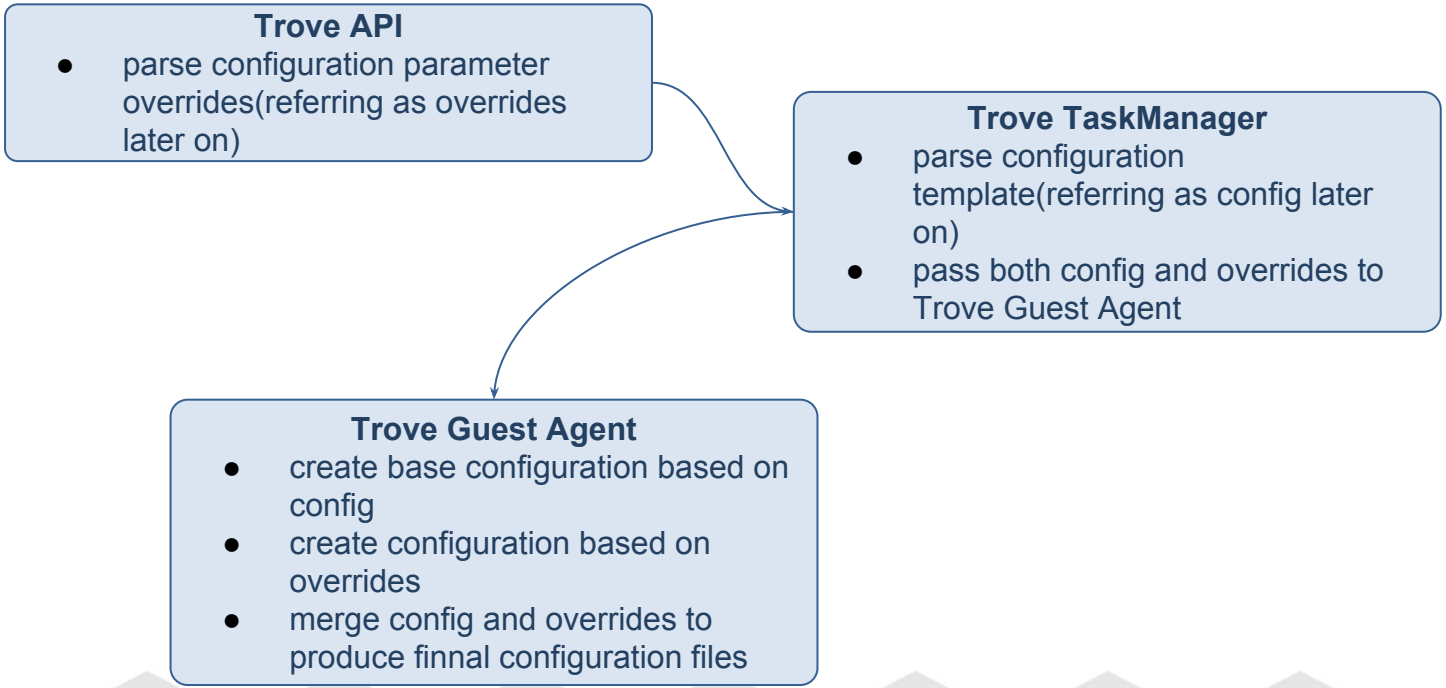
Trove 架构分析/Trove architecture

配置管理/Configuration

- **configuration (Trove Guest Agent)**
  - 基本的版本管理/basic version control
    trove/guestagent/common/configuration.py
  - 基础配置文件根据 Trove TaskManager 解析之后的 template 生成
    Basic configuration comes from rendered templated by Trove TaskManager.
  - overrides 根据加载的 configuration 生成
    The “overrides” comes from attached configuration.
  - 最终配置文件是合并基础配置文件和 overrides 的结果
    Final configuration used by the database server is the merging result of basic configuration and overrides.

Trove 架构分析/Trove architecture

配置管理/Configuration



Trove 架构分析/Trove architecture

面临的问题/Problems

- **创建 Trove 专用镜像/Guest image building**
  - 没有提供官方镜像（资源，优化，加固，自定义）
    No official images(resources, optimization, hardening, customization)
  - DIB/trovestack
    更适用于 CI 或者开发过程
    More suitable for CI or development.
  - 其它工具的链接/links for other tools
    - https://github.com/open-power-ref-design-toolkit/os-services/tree/master/osa/dbaas/dbimage-builder
    - https://github.com/denismakogon/trove-guest-image-elements
  - trovestack 中 Fedora 镜像没有继续维护，CentOS 镜像有计划支持，但仍然没有开始
    Fedora support in trovestack is not maintained for some time, CentOS support is planned but not start yet.

Trove 架构分析/Trove architecture

面临的问题/Problems

- **GuestAgent 和其它组件的通信**
  Communication between Guest Agent and the controll panel
  - 安全性/security
    Ocata 之后支持加密 RPC 通信，但没有解决根本问题
  - agentless guest
    没有最终方案（not worked out）
  - Octavia way communication bwteen
    - https://docs.openstack.org/octavia/latest/index.html
    - https://docs.openstack.org/octavia/latest/contributor/guides/dev-quick-start.html#load-balancer-network-configuration
    - still in review

Trove 架构分析/Trove architecture

面临的问题/Problems

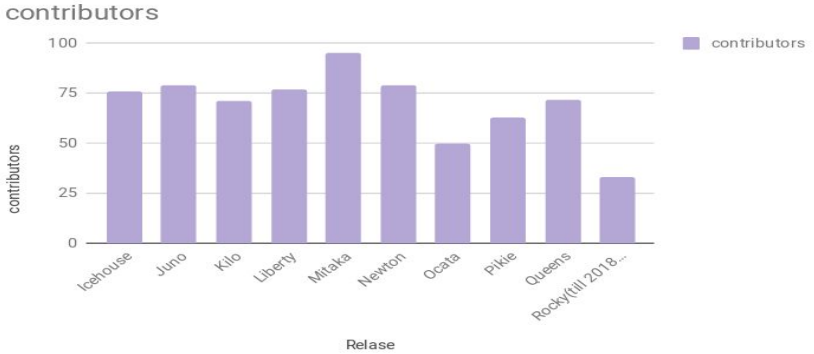
- **集群支持/Clustering support**
  - 用户需求/increasing user demand
    - 仅仅是基本实例或者数据复制不能满足用户对数据库即服务的需求
      Basic database instances and replications are not enough for the user
  - 目前只有初步的讨论/Currently only some initial discussions
    - https://review.openstack.org/#/c/564103/
    - clustering API/TaskManager/GuestAgent refactoring?
- **需要完善对 Datastore 版本的支持**
  Datastore version support need to be improved
  - volume_support/replication_strategy 等通过datastore version 来控制更加理想
    Configurations like volume_support, replication_strategy controlled by datastore version would be much better.
  - conifguration parameters

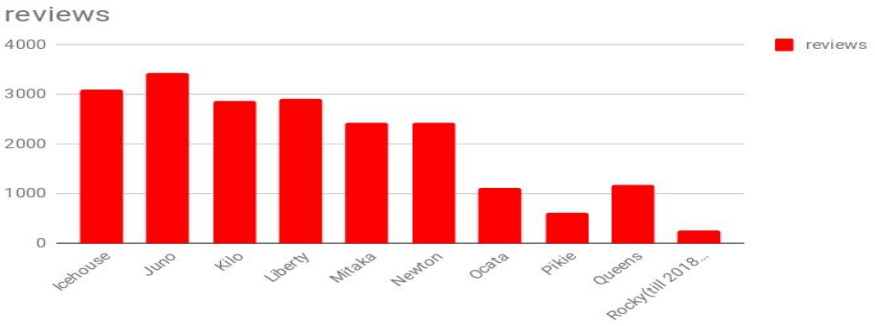
Trove 架构分析/Trove architecture

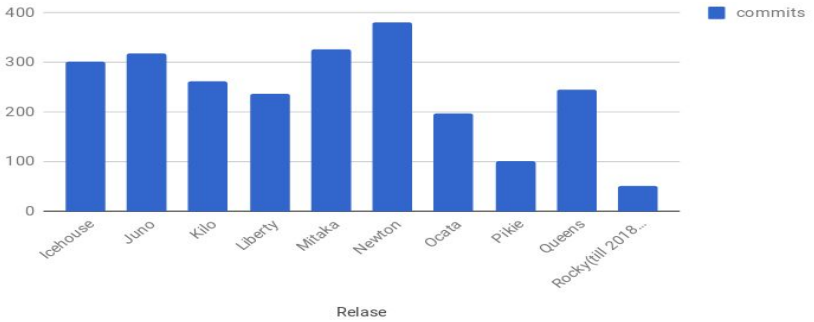
面临的问题/Problems

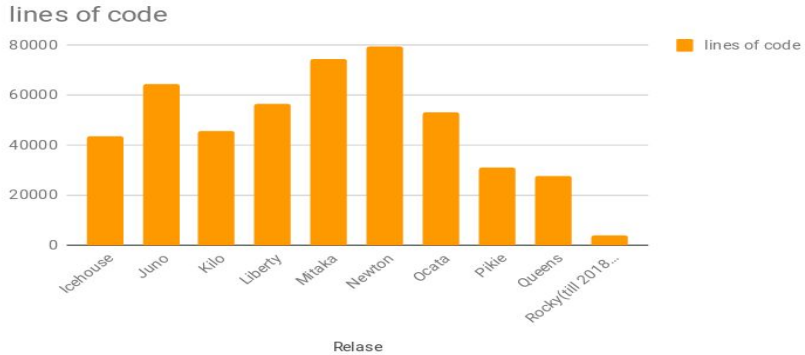
- **Trove Conductor 作用有限**
  Weak Trove Conductor
- **支持 baremetal/容器（？）**
  Add supports for baremetal/container(?)
- **测试代码**
  Testing codes
- **文档和支持**
  Documentaion and supporting
- **和其它 OpenStack 同步演进**
  Integration with the other OpenStack project

Trove 当前开发情况/Current status









Trove 当前开发情况/Current status

- **参与度自 Ocata 周期之后持续下降**
  Participations keep falling after Ocata.
- **当前主要开发人员均来自中国，但是主要精力都并不在 Trove**
  Current active team members are all from China, sadly none of us are focusing on Trove.
- **Queens 版本之后的开发情况**
  Development after Queens
  - Rocky PTG 讨论记录/Rocky PTG notes
    - https://etherpad.openstack.org/p/trove-ptg-rocky
  - Rocky 开发目标跟踪/Rocky goals tracking
    - https://etherpad.openstack.org/p/trove-priorities-and-specs-tracking
  - Weekly meetings
    - https://etherpad.openstack.org/p/trove-rocky-meeting-agendas

Trove 当前开发情况/Current status

参与 Trove/Partipate in Trove

- **代码/Repos**
  - https://git.openstack.org/cgit/openstack/trove/
  - https://git.openstack.org/cgit/openstack/python-troveclient/
  - https://git.openstack.org/cgit/openstack/trove-dashboard/
  - https://git.openstack.org/cgit/openstack/trove-tempest-plugin/
  - https://git.openstack.org/cgit/openstack/trove-specs/
- **开发环境/Development Environment**
  - devstack
  - integration/scripts/trovestack (in trove repo)

CHINA
OpenInfra Days

Thank You