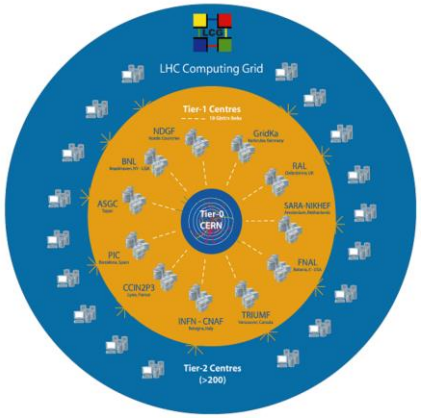
高能物理容器镜像服务

郑伟
计算中心, 中国科学院高能物理研究所

背景

随着容器应用范围越来越广泛普遍，虚拟化容器技术成为高能物理实验不可缺少的一部分，随之对镜像的管理要求越来越高，安全更加严格

HTCondor Software Suite



slurm workload manager

DIRAC THE INTERWARE diracgrid.org

HEPS | jupyterhub

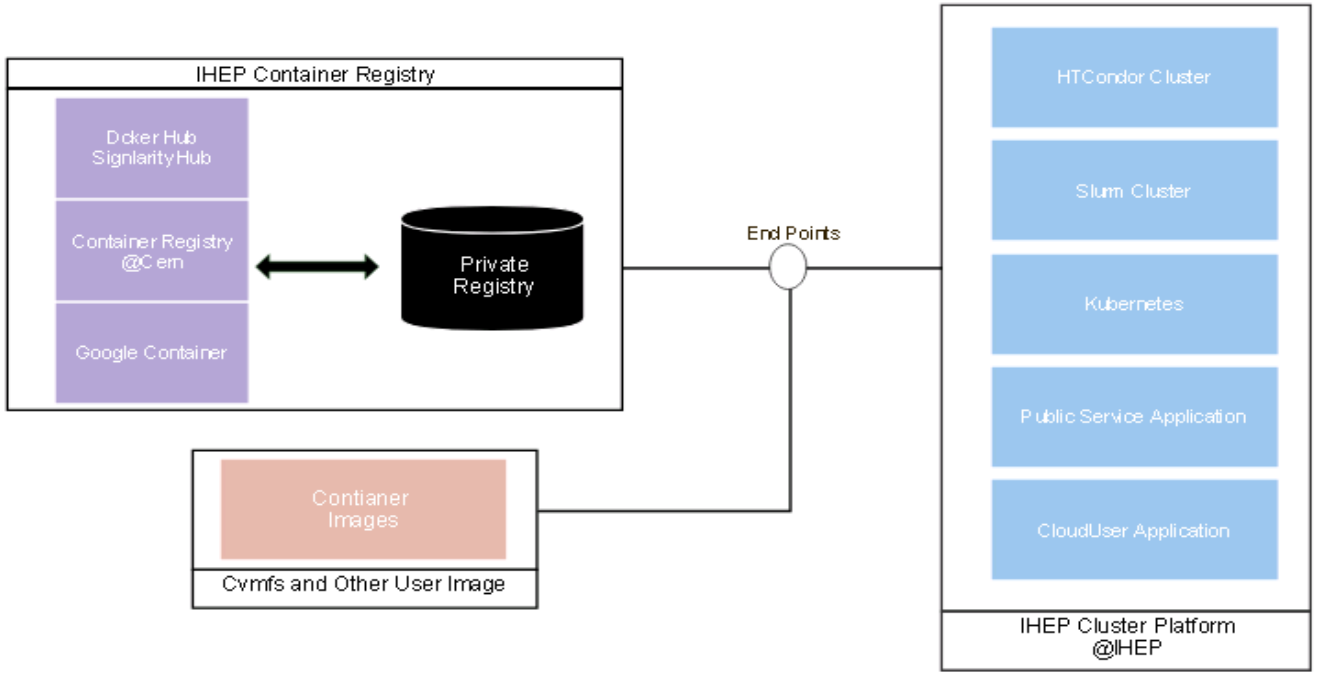
镜像服务需求

高能物理需求特点

- 面向大容量、多版本更新管理，稳定有效分发
- 支持多站点大规模镜像发布（传输性能）
- 支持分布式跨站点镜像分发（多站点协同）
- 支持高能物理特殊的实验计算环境、软件环境、网络环境
- 支持多规格、架构的镜像格式
- 镜像多维度安全保障

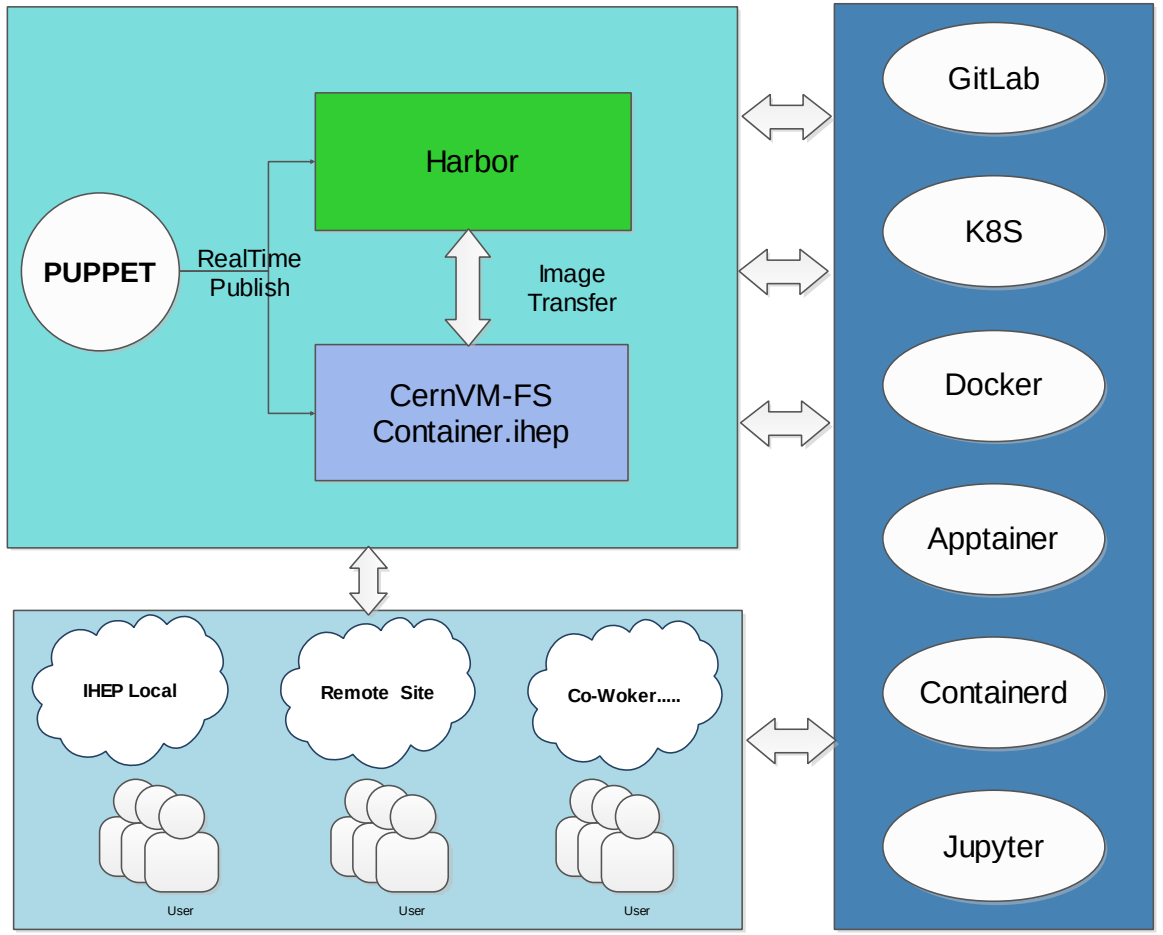
初期容器镜像管理架构

- Docker/Singularity Registry 支持本地访问local，无认证，多个Registry管理分散不统一
- 访问速度限制、空间限制等
- Cern Registry 权限问题等



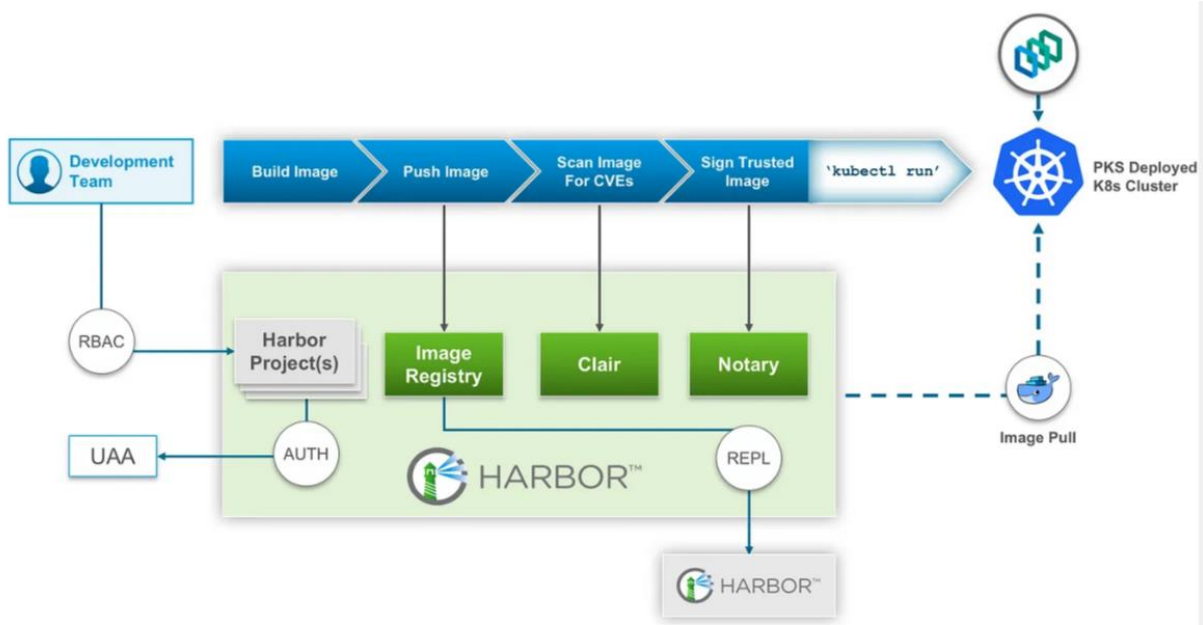
当前容器管理架构

- 底层基于Harbor&Cvmfs-FS管理
- 镜像底层转换
- 实时发布安全审核
- 用户统一认证
- 分布式发布和加速
- 云环境、一平台多中心



Harbor

- Harbor：由VMware公司的中国团队开发的，开源镜像仓库,扩展了开源Docker Distribution功能，提高Image传输效率。
  - 镜像分发管理，支持Docker/Apptainer Helm Chart 等符合OCI规范的制品管理
  - 访问控制，多租户项目管理
  - Webpotal
  - 漏洞扫描、签名
  - 设置Quota
  - ……



Harbor@IHEP

- IHEP统一认证管理
- 30个项目（25公开，5私有）
- 131镜像仓库（118公开，13个私有）
- BES、LHAASO、HEPS、JUNO…


```
[root@dockerhub harbor]# docker-compose ps -a
NAME                IMAGE                                   COMMAND                  SERVICE             CREATED
chartmuseum         goharbor/chartmuseum-photon:v2.5.1      "./docker-entrypoint…"   chartmuseum         4 weeks ago
harbor-core         goharbor/harbor-core:v2.5.1             "/harbor/entrypoint.…"   core                4 weeks ago
harbor-db           goharbor/harbor-db:v2.5.1               "/docker-entrypoint.…"   postgresql          4 weeks ago
harbor-jobservice   goharbor/harbor-jobservice:v2.5.1       "/harbor/entrypoint.…"   jobservice          4 weeks ago
harbor-log          goharbor/harbor-log:v2.5.1              "/bin/sh -c /usr/loc…"   log                 4 weeks ago
harbor-portal       goharbor/harbor-portal:v2.5.1           "nginx -g 'daemon of…"   portal              4 weeks ago
nginx               goharbor/nginx-photon:v2.5.1            "nginx -g 'daemon of…"   proxy               4 weeks ago
 0.0.0.0:80->8080/tcp, :::80->8080/tcp, 0.0.0.0:443->8443/tcp, :::443->8443/tcp
notary-server       goharbor/notary-server-photon:v2.5.1    "/bin/sh -c 'migrate…"   notary-server       4 weeks ago
notary-signer       goharbor/notary-signer-photon:v2.5.1    "/bin/sh -c 'migrate…"   notary-signer       4 weeks ago
redis               goharbor/redis-photon:v2.5.1            "redis-server /etc/r…"   redis               4 weeks ago
registry            goharbor/registry-photon:v2.5.1         "/home/harbor/entryp…"   registry            4 weeks ago
registryctl         goharbor/harbor-registryctl:v2.5.1      "/home/harbor/start.…"   registryctl         4 weeks ago
trivy-adapter       goharbor/trivy-adapter-photon:v2.5.1    "/home/scanner/entry…"   trivy-adapter       4 weeks ago
[root@dockerhub harbor]#
```

支持OCI制品

- 支持管理以ORAS（OCI Registry As Storage）客户端制作的符合OCI规范的自定义制品，包括镜像、Helm Chart和自定义的OCI制品

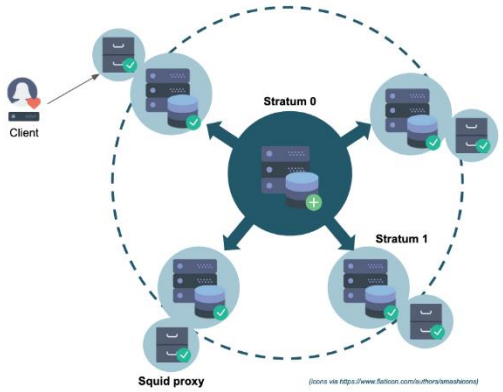
```
oras push dockerhub.ihep.ac.cn/slc7_cluster/yourimagename:tagname  /path/yourimagename.sif
```

```
apptainer push slc75base_org.sif  oras://dockerhub.ihep.ac.cn/slc7_cluster/slc75base22:org
apptainer pull  docker://dockerhub.ihep.ac.cn/slc7_cluster/slc75base:org
```

Cvmfs@IHEP

- The CernVM File System (CernVM-FS) 网络文件系统，只有管理员有权限修改， 主要目的实现在世界范围内轻松分发软件
- 客户端（/cvmfs）是一个虚拟文件系统，文件数据仅在实际访问时才下载，支持容器安装
- /cvmfs/container.ihep.ac.cn

```
[zhengw@ccopt container.ihep.ac.cn]$ df -h |grep cvmfs
cvmfs2                          40G  2.7G   38G   7% /cvmfs/cvmfs-config.cern.ch
cvmfs2                          40G  2.7G   38G   7% /cvmfs/bes3.ihep.ac.cn
cvmfs2                          40G  2.7G   38G   7% /cvmfs/juno.ihep.ac.cn
cvmfs2                          40G  2.7G   38G   7% /cvmfs/atlas.cern.ch
cvmfs2                          40G  2.7G   38G   7% /cvmfs/cms.cern.ch
cvmfs2                          40G  2.7G   38G   7% /cvmfs/common.ihep.ac.cn
cvmfs2                          40G  2.7G   38G   7% /cvmfs/cepc.ihep.ac.cn
cvmfs2                          40G  2.7G   38G   7% /cvmfs/lhaaso.ihep.ac.cn
cvmfs2                          40G  2.7G   38G   7% /cvmfs/sft.cern.ch
cvmfs2                          40G  2.7G   38G   7% /cvmfs/container.ihep.ac.cn
```



高能物理镜像

- 节点镜像
  - 登陆节点 SL75/65/69/55
  - 计算节点 SL65/69/55/58
- 物理实验镜像
  - BES,HEPS,LHAASO,JUNO ….
- 服务镜像
  - EOS、Mysql，Apache, MonitorAgent, Grafana……
- 软件镜像
  - BOSS、Nagio……
- 系统架构
  - x86-64、x86、ARM、ARM64

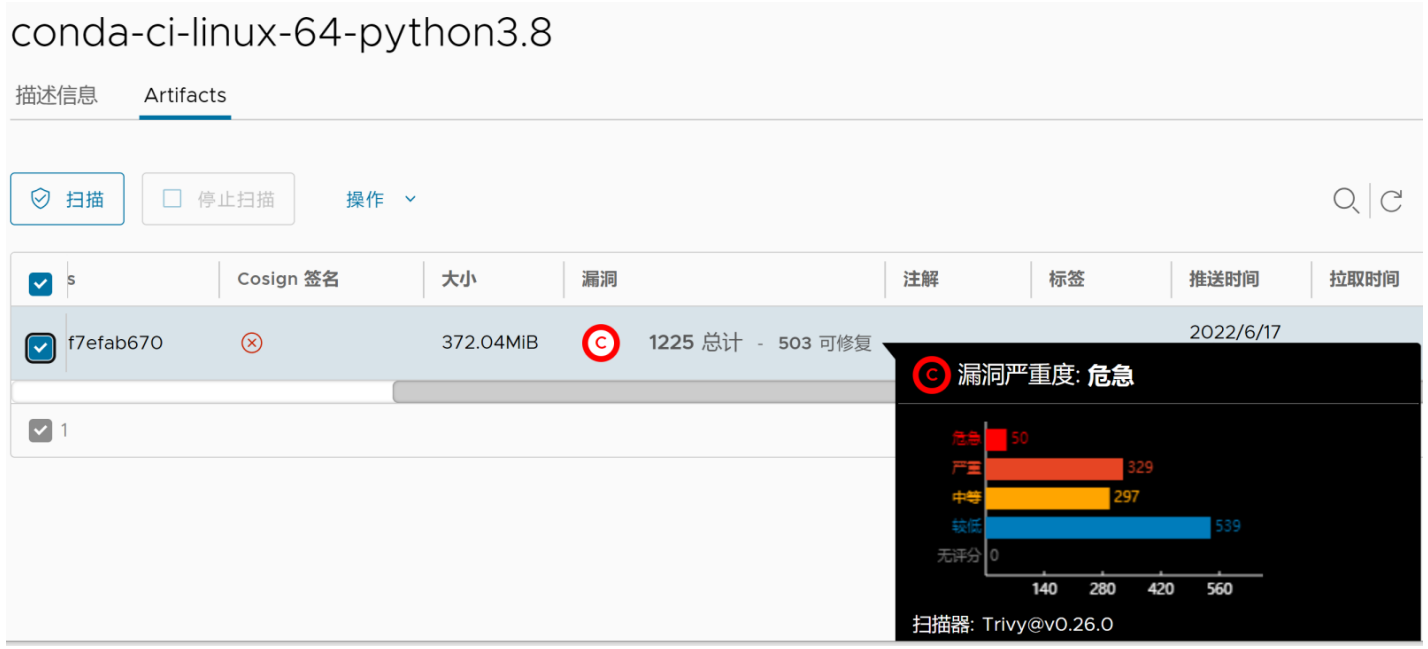
镜像分发

- 通过复制、缓存、代理、P2P等方式，实现本地、异地以及大规模镜像访问下载的稳定高效镜像分发策略
- 异地分布式拉取
- 大规模并发访问

镜像分发策略					
	镜像访问	**镜像缓存**	**异地发布**	**大规模分发**	**其他镜像源**
Harbor	统一认证	Cache	镜像同步	负载均衡 /P2P组网	Proxy代理
Cvmfs	证书认证	按需加载	Client挂载	Squid	分卷挂载

镜像安全策略

- 安全扫描容器镜像
  - 安装Trivy安全漏洞扫描器，镜像漏洞，文件系统漏洞，主机漏洞
- 容器镜像签名
  - Notary 对镜像进行加密签名用来保证镜像件来源和镜像内容防篡改
- 开启镜像版本控制
  - 版本回滚
  - 文件镜像md5,权限控制



应用:Hep_container统一管理

- 一平台多中心容器管理工具(Hep_containter)：建立高能物理用户统一、满足多样化需求、安全、对用户透明、跨站点的容器引擎工具
- 特点：
  - 一平台多中心：CSNS、稻城、科大、北大、兰大…
  - 多镜像：Centos7,SL7/6/5….
  - 环境变量自动传递
  - 根据实验组，自动挂载实验数据盘
  - 支持GPU驱动..
  - 顺滑升级策略：升级镜像、引擎等用户使用无感
  - 通过fakeroot 实现用户build 镜像功能
  - 支持用户自定义镜像，如编译 RASER 所需软件环境
  - 适配作业容器站点（兰大全容器站点）


Hep_container 用户使用简要手册

1 简述

Singularity 是目前在高性能计算平台上被大量应用的轻量虚拟化容器技术，能够提供操作系统级的虚拟化。 Hep_container 是基于 singularity 容器管理命令开发的适用于高能所计算集群的容器客户端工具，满足用户使用多种操作系统版本及环境的需求。

说明：本文涉及的命令均需要在 lxslc7 登陆节点上运行，所用命令在以下目录，建议将该目录加入用户个人环境变量 PATH 中。

Lxslc7： /afs/ihep.ac.cn/soft/common/sysgroup/container/bin/

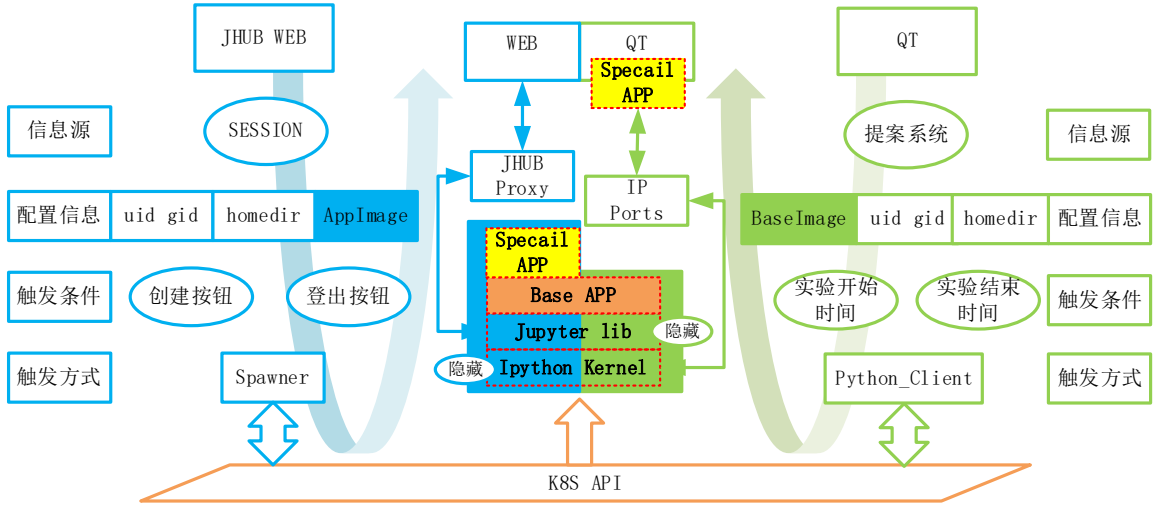
2 容器命令使用说明

Hep_container 的容器命令主要有三种操作 images、shell、exec。可以在命令行中通过 help 参数查看各个命令的使用说明和样例

```
$ hep_container help
Usage : hep_container   <command> [command options...]
CONTAINER USAGE COMMANDS:
    shell       Run a Bourne shell within container
    exec        Execute a command within container
    images      List Support container images
    help        Show command help
EXAMPLES:
    hep_container images
    hep_container shell SL5
    hep_container exec   SL5 cat /etc/redhat-release
    hep_container exec   SL5 python ./yourprograme.py
```

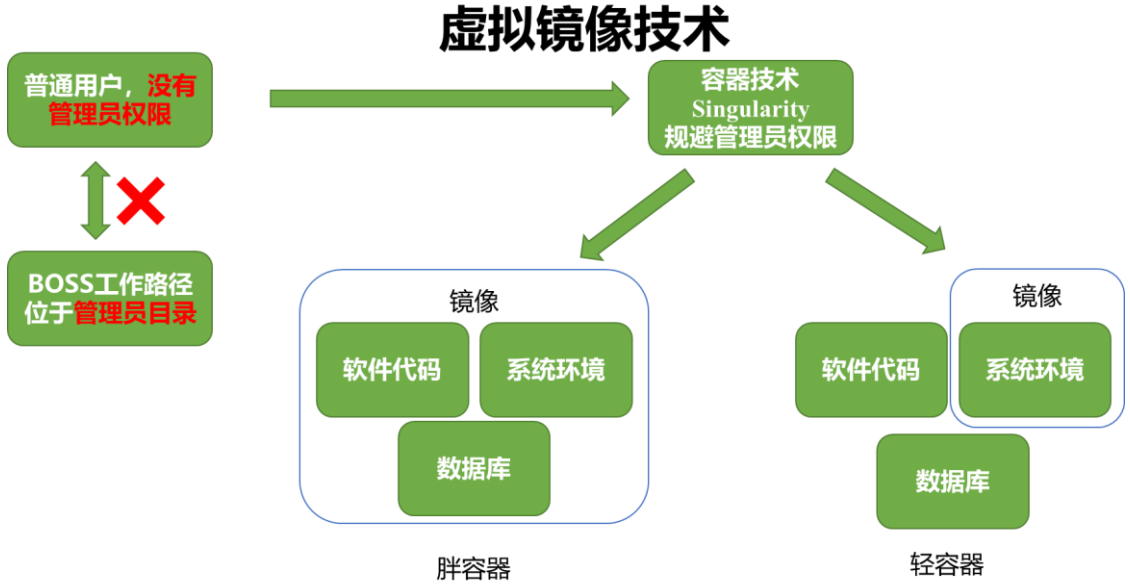
应用:HEPS交互式计算

- HEPS生产环境的交互式计算平台
- 高可用、高可靠k8s集群
- 公网环境容器安全和访问控制策略



应用:BESIII@天河2号

- 目的：实现BESIII软件在天河2号的部署和计算应用，扩展资源
- 难点：编译部署难度大，权限少
- 解决方式：容器化部署，Cvmfs软件+数据库+系统环境镜像
- 方案路线发展：编译部署->全镜像->Cvmfs镜像缓存



长期规划

- 建立一体化镜像软件集、服务集，提供镜像目录Catalog库，通过helm,k8s快速部署
- 扩展认证管理, 对更多合作单位提供服务
- 开发用户需求的CLI提供更灵活的服务
- 提供更详细的统计、访问、日志分析

小结

- 根据高能物理镜像服务需求特点，基于开源Harbor和分布式Cvmfs及相关安全组件、以及镜像转换中间件的总体架构，实现对高能物理容器软件镜像管理服务
- 适用于镜像分布式架构、大规模并发、缓存性能提高
- 自动化分发，底层镜像实时发布和类型转换
- 安全可靠，发布安全认证、扫描、签名
- 已经有了较多的高能物理实验应用案例

谢谢