CHINA OpenInfra Days

CHINA OpenInfra Days | IT大咖说 知识共享平台

裸金属集群上云之路

目录
CONTENTS

为何出发

为何出发

来自客户的问题

引入云平台后，对于已经存在的物理机集群，并且在其上已经运行了业务，如何处理？

一种选择

P2V迁移，将已有物理机上业务离线或者在线迁移到虚拟机或者容器中

另一种选择？

存在的风险

迁移后的性能下降
业务中断风险
云化改造的工作量，时间成本

为何出发

一个真实的案例

客户的硬件

存量物理服务器众多
硬件型号不一，新老混合

02

03

客户的顾虑

不能接受业务在虚拟机或者容器内的性能损失

01

04

客户的业务

大数据分析业务
已在多个地市部署有现网局点

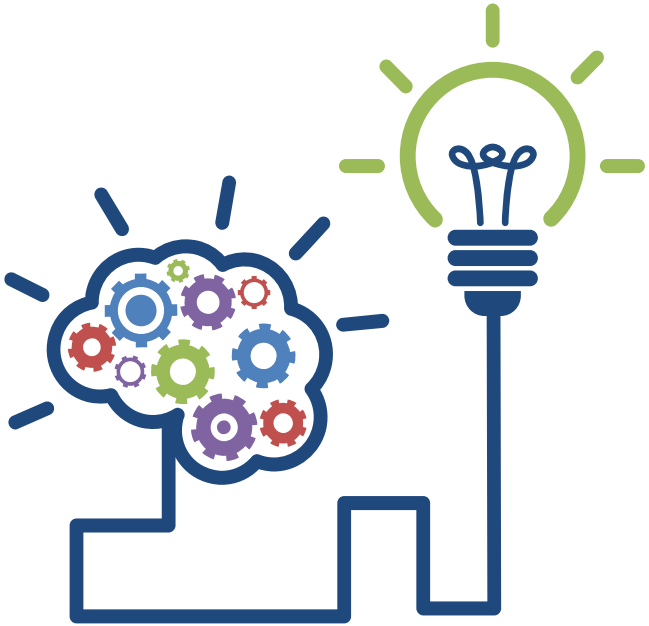
客户的时间

要求支持上云的时间非常短，必须快速交付

踏上旅途

踏上旅途

解决思路和挑战



通过将物理主机快速纳管接入云平台，保持物理主机和业务正常运行，由云平台统一管理，统一监控

- 性能损失最小
- 统一监控与运维
- 现网局点影响最小
- 业务云化代价最小

踏上旅途

FitOS裸机纳管路标

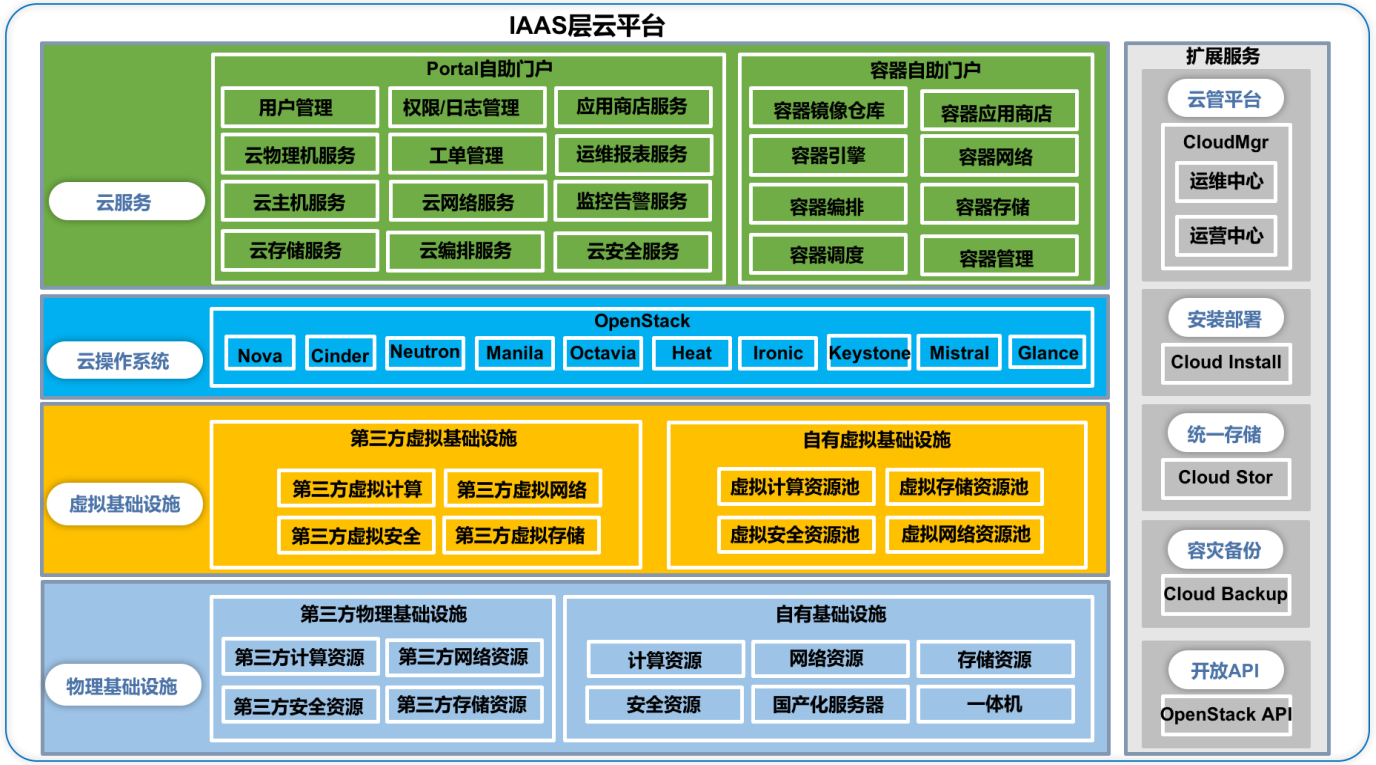
01 尽量减少纳管参数的输入，支持不同的硬件配置服务器

02 纳管后通过云平台可以支持对裸机集群的统一监控

03 整个纳管过程要通过界面展示

踏上旅途

FitOS简介



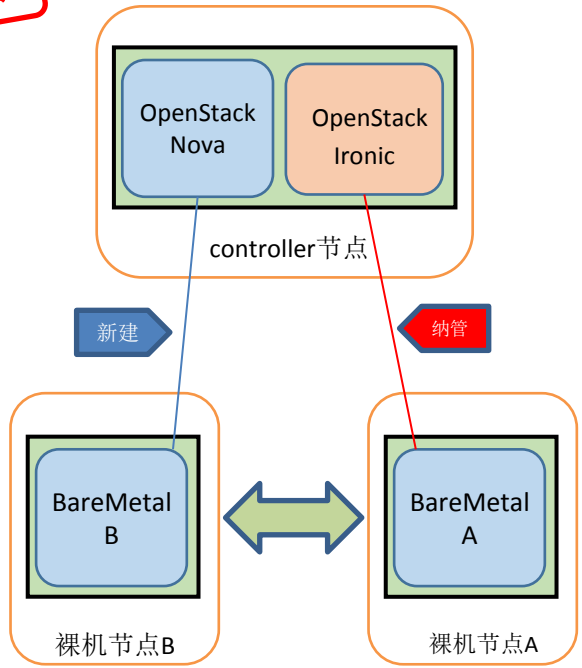
踏上旅途

裸机纳管第一步

已经支持

- 原生Ironic纳管裸机方案：
  - 实现：使用Ironic命令行，建立Ironic node信息，更新node状态。
- 方案约束：
  - 由于Nova目前不支持对裸机的纳管，因此对于纳管后裸机的操作只能通过Ironic下发。与通过Nova新创建的裸机管理上有区别。

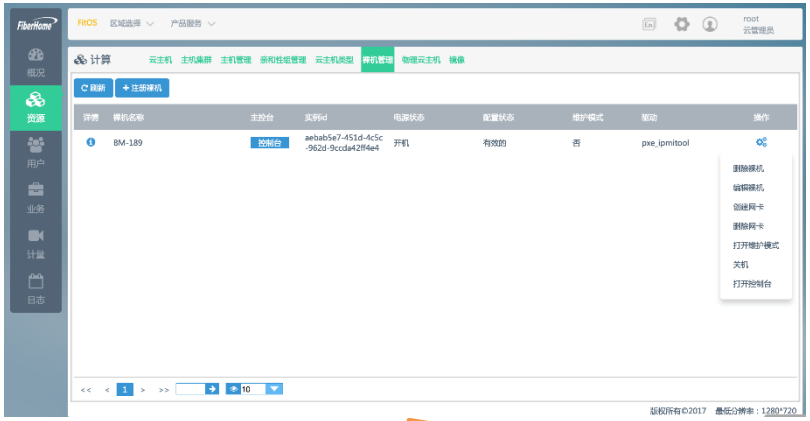
```
# Explicitly set the client API version environment variable to
# 1.17, which introduces the adoption capability.
export OS_BAREMETAL_API_VERSION=1.17

openstack baremetal node create --name testnode \
    --driver agent_ipmitool \
    --driver-info ipmi_address=<ip_address> \
    --driver-info ipmi_username=<username> \
    --driver-info ipmi_password=<password> \
    --driver-info deploy_kernel=<deploy_kernel_id_or_url> \
    --driver-info deploy_ramdisk=<deploy_ramdisk_id_or_url>

openstack baremetal port create <node_mac_address> --node <node_uuid>

openstack baremetal node set testnode \
    --instance-info image_source="http://localhost:8080/blankimage" \
    --instance-info capabilities="{\"boot_option\": \"local\"}"

openstack baremetal node manage testnode --wait

openstack baremetal node adopt testnode --wait
```



踏上旅途

裸机纳管第一步

Ironic界面：裸机管理



Nova界面：物理云主机管理

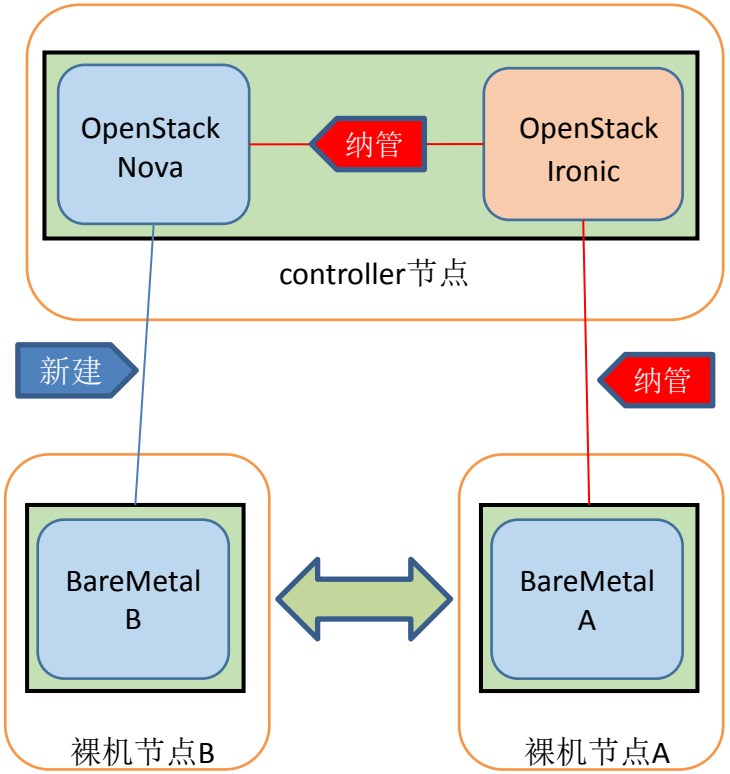
裸机纳管后，被纳管裸机只在裸机管理界面展示，而在物理云主机界面上看不到对应的信息

踏上旅途

裸机纳管第二步

需要开发

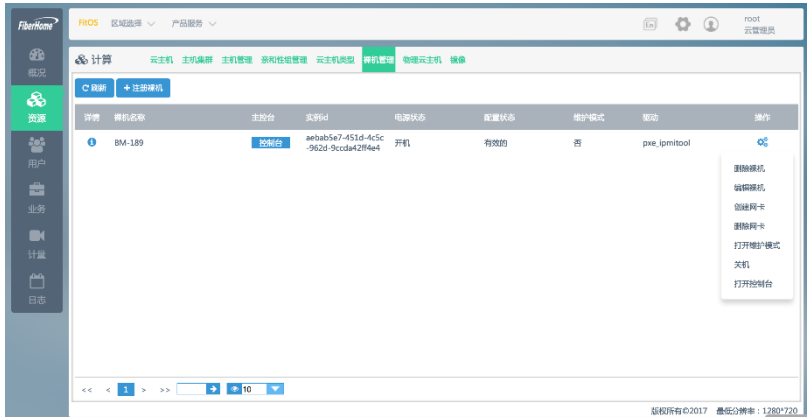
- FitOS创新：
  - 为了解决第一步Ironic纳管后无法通过Nova对裸机计算资源进行统一管理的问题
  - FitOS增加物理云主机纳管能力
- 方案特点：
  - Nova纳管裸机为物理云主机后，可与新创建的物理云主机统一管理，形成裸金属计算资源池。增加包括软重启，优雅关机等高级功能，并且可以方便的通过Nova重新安装操作系统，提高云数据中心中对于裸机的运维管理能力。



踏上旅途

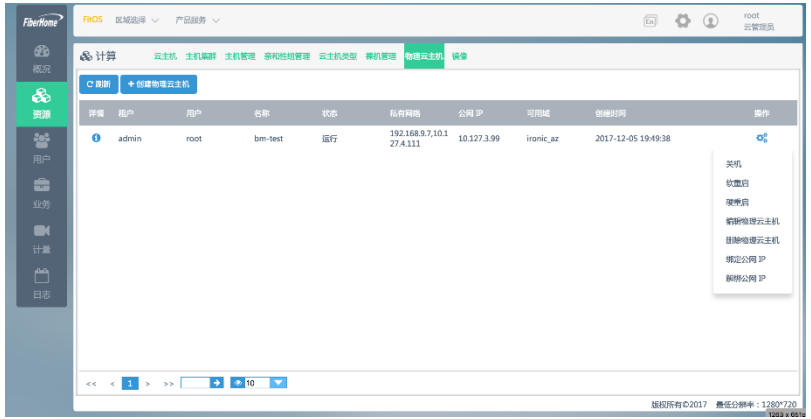
裸机纳管第二步

Ironic界面：裸机管理



裸机纳管后，被纳管裸机只在裸机管理界面展示，而在物理云主机界面上看不到对应的信息

Nova界面：物理云主机管理



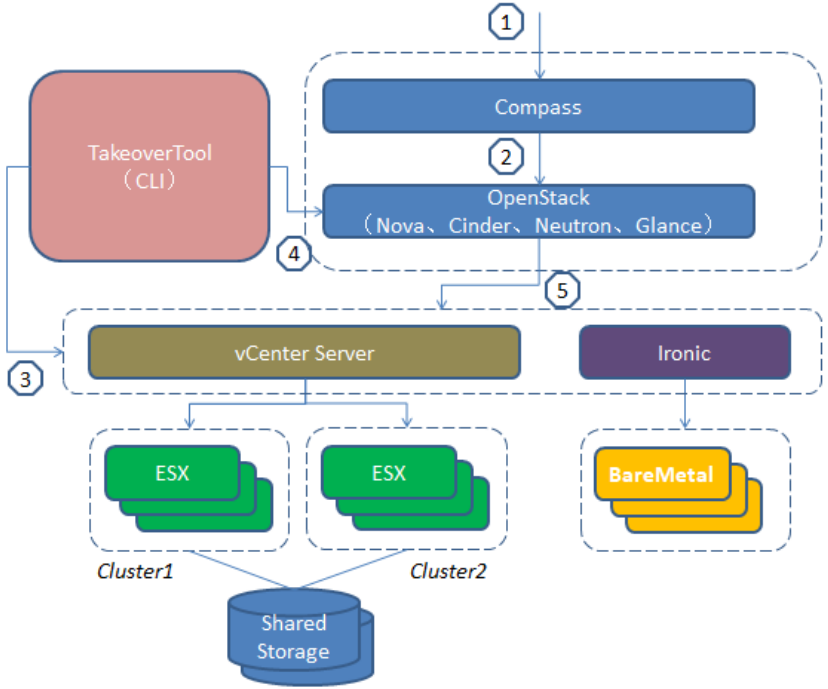
物理云主机纳管后，被纳管裸机在物理云主机管理界面展示，可以支持软重启等高级特性

踏上旅途

关键技术点

通过自研 TakeoverTool工具，来完成对裸机，异构虚拟化等平台/设备的纳管接入：

- Compass工具完成组件Driver切换；
- TakeoverTool 识别底层对接平台，调用对应接口获取需要纳管的资源信息；
- TakeoverTool调用OpenStack的接口将从底层获取的资源纳管到OpenStack平台；
- 修改OpenStack对应配置，建立资源与底层基础设施资源的具体关联。



踏上旅途

裸机纳管准备

纳管前需要创建网络，只能创建一个network，如果有多个网段，则在此network下创建多个子网，子网的IP池，应该包含所有待纳管主机的管理IP地址。

需先创建好要纳管的租户和用户信息，并将纳管租户配额（包括CPU、Memory、磁盘等）修改为-1

获取待纳管环境所有主机信息，以如下格式保存到hosts.conf中:

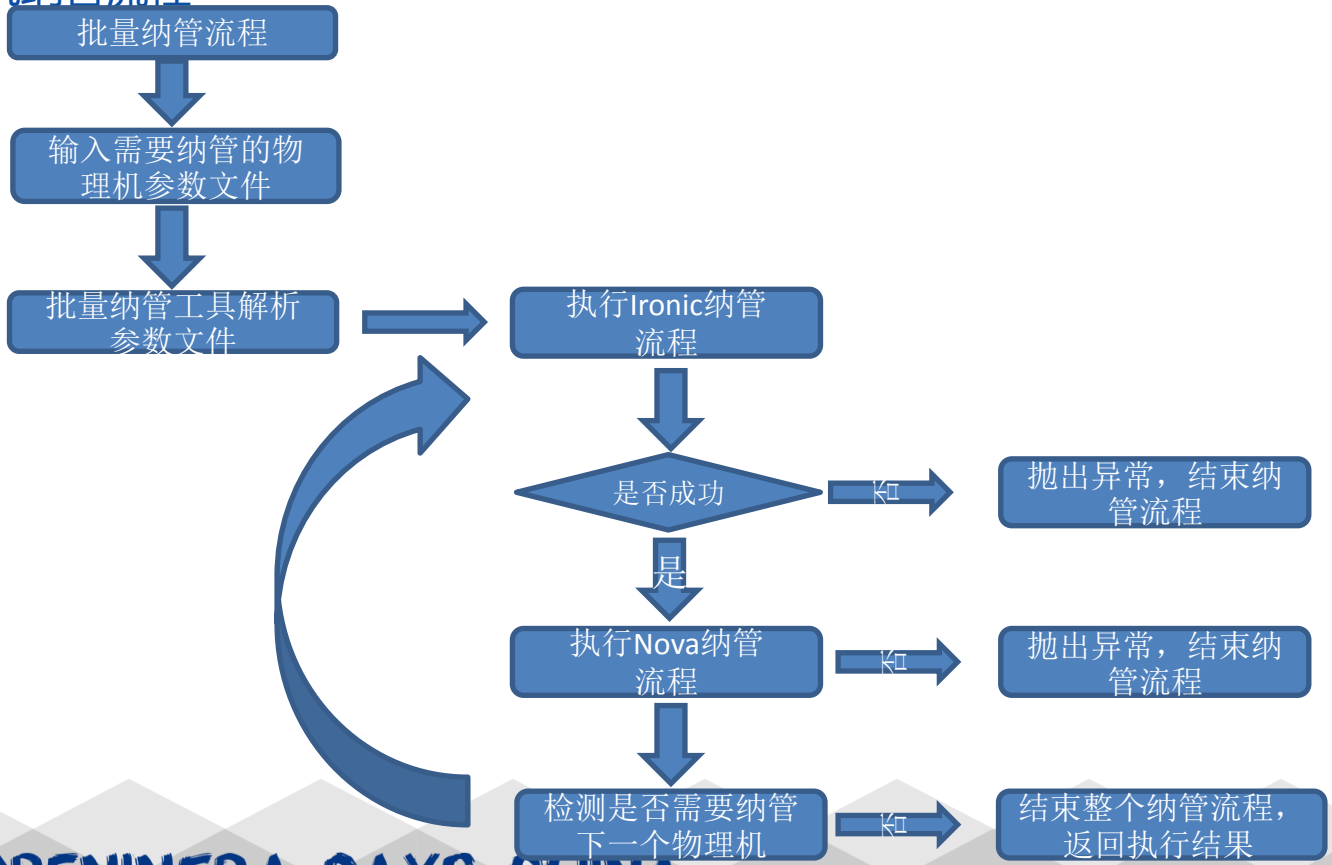
ipmi_ip;ipmi_user;ipmi_password;mgmt_ip;username;passowd

创建并上传一个裸机纳管使用的假镜像

qemu-img createbaremetal_adopt.raw 1M

踏上旅途

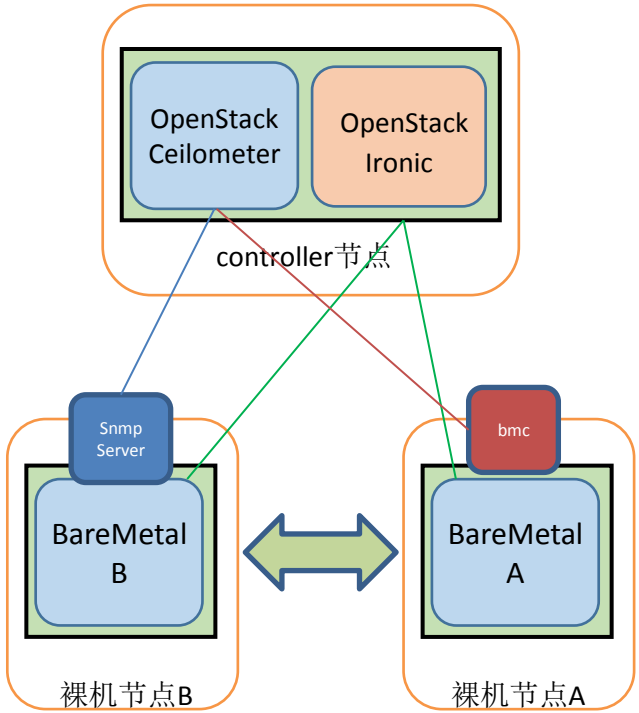
裸机纳管流程


批量纳管流程
输入需要纳管的物理机参数文件
批量纳管工具解析参数文件
执行Ironic纳管流程
是否成功
否
抛出异常，结束纳管流程
是
执行Nova纳管流程
否
抛出异常，结束纳管流程
检测是否需要纳管下一个物理机
否
结束整个纳管流程，返回执行结果

踏上旅途

上云后的统一监控

- SNMP监控:
  - 实现：裸机镜像中安装SNMP Server，裸机部署后，通过Ceilometer agent采集该裸机监控指标
  - 影响：需要在镜像或已有操作系统中部署配置SNMP Server
  - 优势：方案成熟，监控信息对用户可用性较大
- IPMI监控
  - 实现：通过打通管理网络和BMC网络，Ceilometer使用IPMI命令采集物理主机硬件信息，需增加采集指标
  - 优势：无Agent部署，配置简单
  - 影响：IMPI采集信息较少且主要为硬件底层信息，对用户可用性不大。



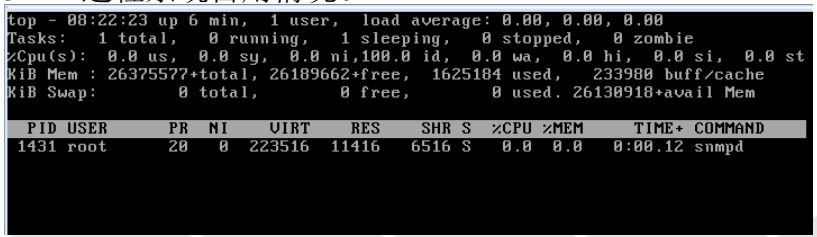
踏上旅途

上云后的统一监控

IPMI

监控指标	描述
hardware.ipmi.fan	风扇转速
hardware.ipmi.temperature	系统温度
hardware.ipmi.current	系统功率
hardware.ipmi.voltage	系统电压

SNMP进程系统占用情况：



SNMP

监控指标	
hardware.cpu.load.1min	一分钟cpu load
hardware.cpu.load.5min	五分钟cpu load
hardware.cpu.load.15min	十五分钟cpu load
hardware.cpu.util	cpu使用率
hardware.memory.total	内存总量
hardware.memory.used	内存已使用
hardware.memory.swap.total	swap总量
hardware.memory.swap.avail	swap可用
hardware.memory.buffer	buffer
hardware.memory.cached	cached
hardware.system_stats.cpu.idle	CPU空闲
hardware.disk.size.total	磁盘总量
hardware.disk.size.used	磁盘已使用
hardware.network.incoming.bytes	网络流入总量
hardware.network.outgoing.bytes	网络流出总量
hardware.network.outgoing.errors	网络发送错误包数

踏上旅途

上云后的统一监控

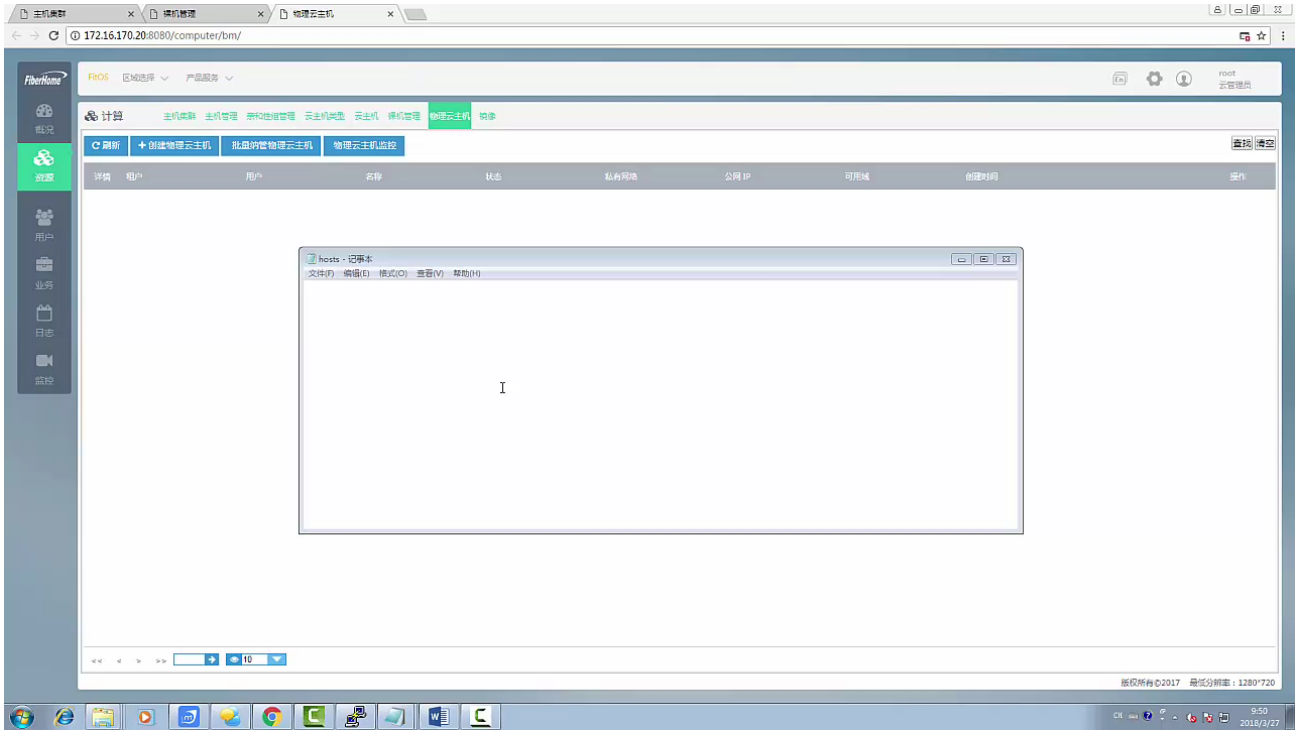
01 特有针对大数据计算云整体监控展示

02 各个物理云主机运行健康状态一目了然

03 负载运行过高自动提示，时刻保持系统稳定

踏上旅途

Demo


172.16.170.20:8080/computer/bm/
FiberHome
计算
刷新
创建物理云主机
批量纳管物理云主机
物理云主机监控
hosts - 记事本

仍在路上

仍在路上

基于一些独立的Project，例如Mogan，可独立使用裸金属计算服务进行新建，纳管，管理等一系列操作，可根据需要选择性安装Nova组件。

TakeoverTool工具的界面优化，支持配置文件的导入导出功能和进度展示等功能

支持裸机网络交换机上的配置更改

CHINA
OpenInfra Days

Thank You