国家天文科学数据中心

丽江分中心建设介绍

王传军

中国科学院云南天文台

丽江天文观测站

提　纲

一　丽江天文观测站介绍

二　分中心建设背景

三　分中心建设进展

四　下一步计划

一、丽江天文观测站

站址坐标：100°1' 48"(E)
26°41' 42"(N)
海拔高度：3200 m
年均温度：8 °C
年均湿度：71.9%
年均风速：1.36m/s
主要风向：西南风
大气视宁度：1”-1.2"
可观测小时：~2000h

一、丽江天文观测站


1.8米望远镜
中泰望远镜
2.4米望远镜
E100.03°
NAO2.4m
01-13-2022 星期四 10:56:09
Camera 01
生活区
射电
2013
影像拍摄日期：2020/1/19 纬度 26.696650° 经度 100.035085° 海拔 3218 米 视角海拔高度 3.48 公里

二、分中心建设背景及关键需求

背景一：

- 数字经济作为清洁载能产业已经成为丽江发展新动能的突破口。
- 丽江水电资源丰富、平均海拔高、年均气温比较低，适合建设大数据存储和高性能计算的中心。
- 丽江天文观测站是国内重要的观测站点，站内目前已经有15台望远镜在运行，有3台望远镜在建，以丽江天文观测站为依托的南方天文观测集群已经初具规模。
- 为了利用数据密集型天文研究的需求，通过科研带动数字经济产业的起步和发展，云南天文台向省科技厅申请经费进行丽江数字经济产业研究基地的建设，其中的一项重要任务就是建设符合国家标准的数据中心机房。

2019年申请，2020年得到资助。

丽江阿海电站

二、分中心建设背景及关键需求

背景一：

- 中科院多个单位在丽江设有观测台站和科研基地，产生和使用数据。
- 计算能力严重不足。

□云南天文台——丽江天文观测站

□昆明植物所——高山植物园

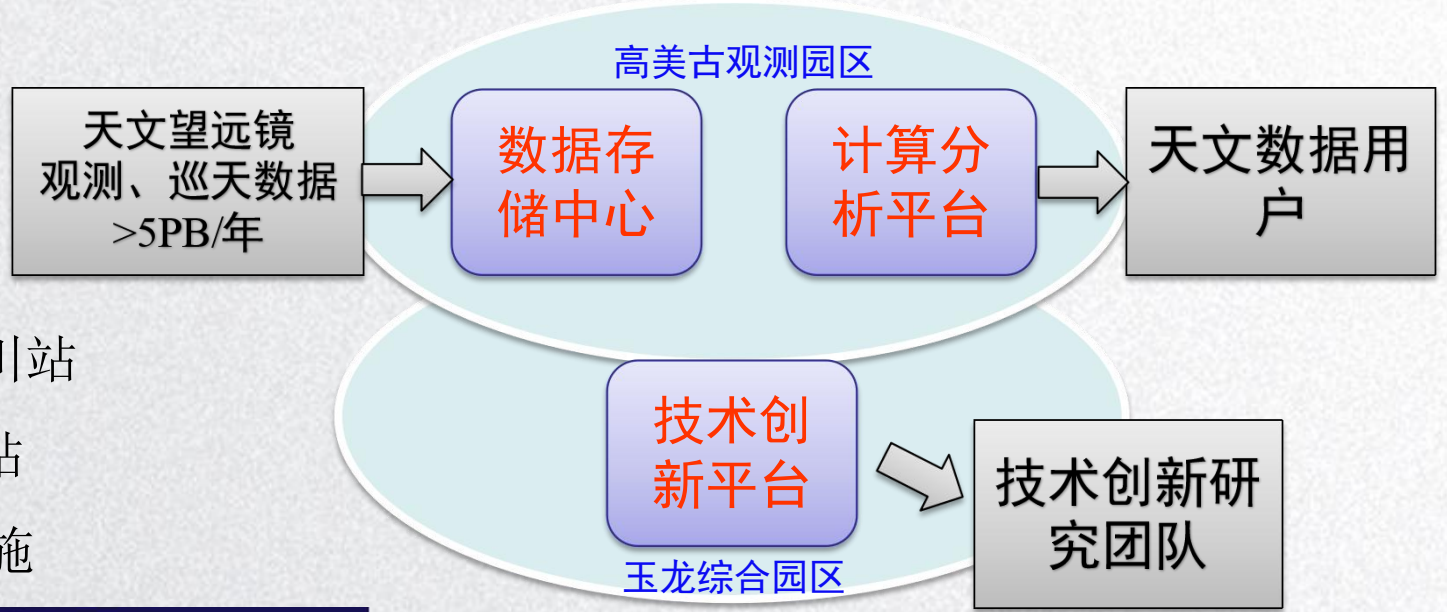
□昆明动物所——繁育基地

□（兰州）寒旱区环境工程所——玉龙冰川站

□空天信息研究院——西南遥感卫星地面站

□以及院、所、高校共建的观测和研究设施

- **建设目标**：丽江数字经济产业研究基地
  - 大数据存储中心
  - 大数据分析计算平台
  - 技术创新平台：信息产品发布、传输、应用




丽江高山植物园


玉龙冰川监测


遥感卫星地面站布局

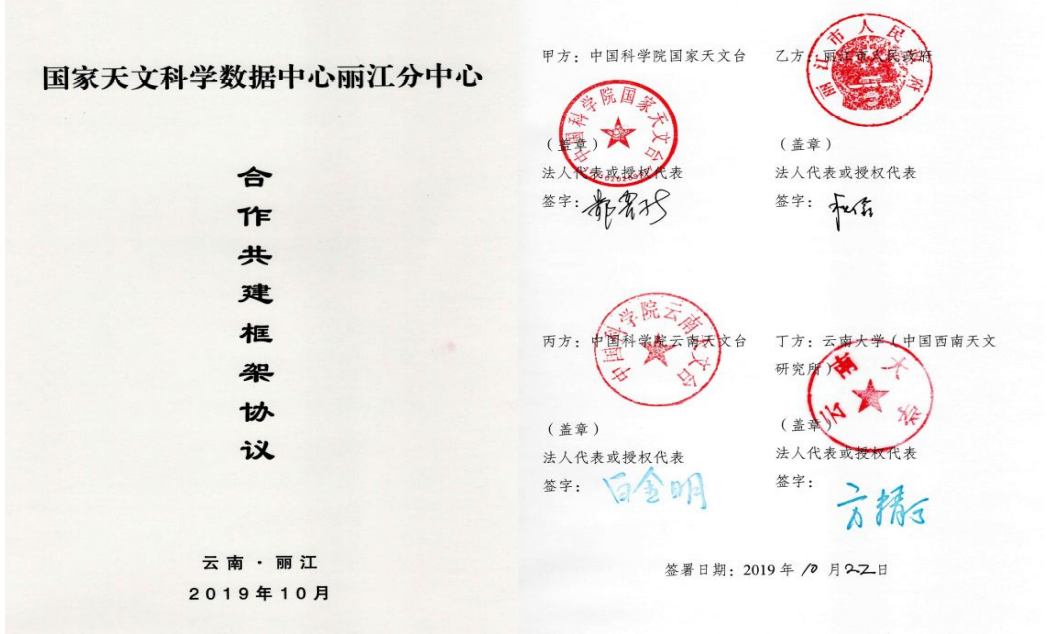
二、分中心建设背景及关键需求

背景一：

国家天文科学数据中心（丽江）分中心建设需求

- 2019年10月，国家天文台、云南大学、丽江市政府和云南天文台签署了合作共建框架协议
- 2020年6月完成了数据中心建设的备案
- 2021年8月，国家天文科学数据中心正式发文

- 天文科学数据中心丽江分中心



国家天文科学数据中心丽江分中心

合作共建框架协议

云南·丽江

2019年10月

甲方：中国科学院国家天文台　　乙方：丽江市人民政府

（盖章）　　（盖章）

法人代表或授权代表　　法人代表或授权代表

签字：　　签字：

丙方：中国科学院云南天文台　　丁方：云南大学（中国西南天文研究所）

（盖章）　　（盖章）

法人代表或授权代表　　法人代表或授权代表

签字：　　签字：

签署日期：2019 年 10 月 22 日

投资项目备案证

项目序号：53879120200600322

项目代码：2020-530721-73-03-044144

项目基本信息			
项目类型	备案		
备案名称	企业投资项目备案		
项目名称	国家天文科学数据中心丽江分中心		
项目（法人）单位	中国科学院云南天文台		
项目法人证照类型	组织机构代码证（事业单位法人）	项目法人证照号码	12100000431204899U
拟开工时间（年）	2020-10-01	拟建成时间（年）	2021-12-31
建设区域	玉龙纳西族自治县		
建设地点	玉龙县太安乡天虹村委会高美古村民小组丽江天文观测站		
跨区域			
所属行业	7310 自然科学研究和试验发展		
建设性质	新建	总投资（万元）	470
建设规模及内容	项目占地面积0.49亩，建筑面积550平方米。主要建设数据中心（456.71平方米，两层建筑）、柴油发电机房（93.44，一层建筑）及相关的室外工程建设（室外硬地：3045.28平方米和道路：647.48平方米）。		
项目符合产业政策申明	2、在线数据与交易处理、IT设施管理和数据中心服务，移动互联网服务，国特网会议电视及图像等电信增值服务		
联系人信息			
姓名	范玉峰	电话	13388884802
身份类型	居民身份	身份号码	130525198101116636
填表人信息			
姓名	[illegible]	手机	18213856325
联系电话		填表时间	2020-06-04 09:54:23

中国科学院 国家天文台文件

国天发字〔2021〕61 号

中国科学院国家天文台关于调整
国家天文科学数据中心组织架构的通知

七、分中心

中心以观测装置和数值模拟设备等数据源和业务发展需要设置分中心。每个分中心设主任 1 位，副主任 1-3 位。

丽江分中心

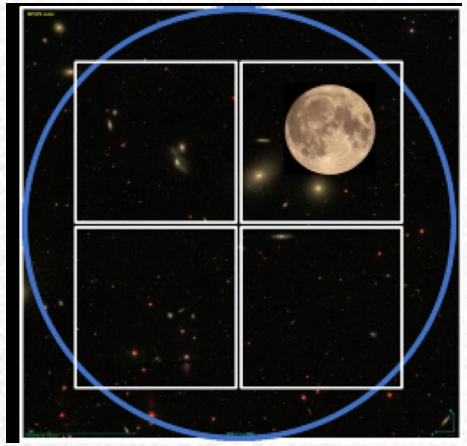
主　任：王传军　云南天文台

副主任：范玉峰　云南天文台

张新华　丽江市科学技术局

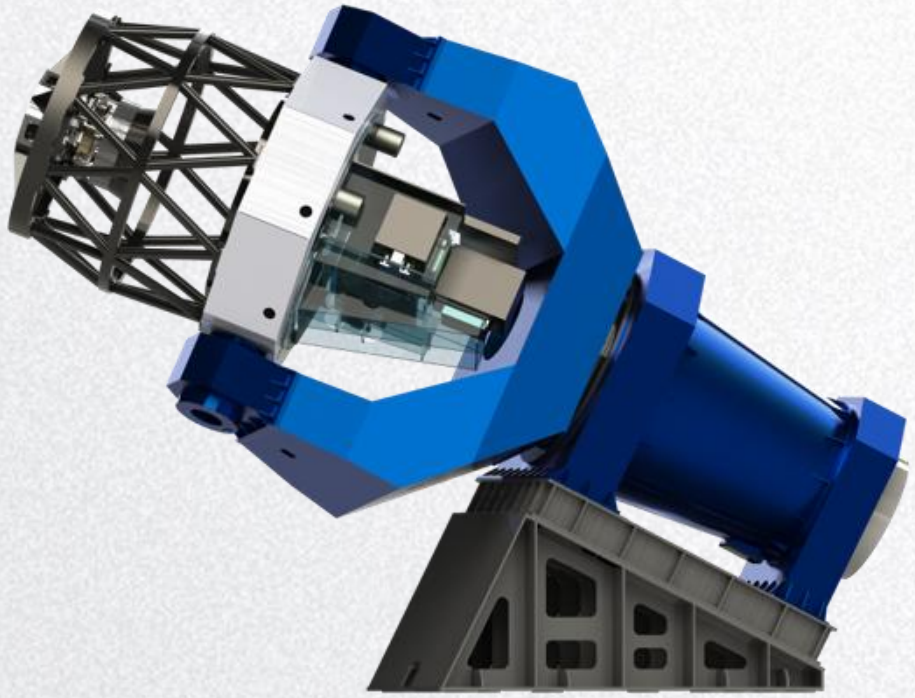
方　圆　云南大学

二、分中心建设背景及关键需求



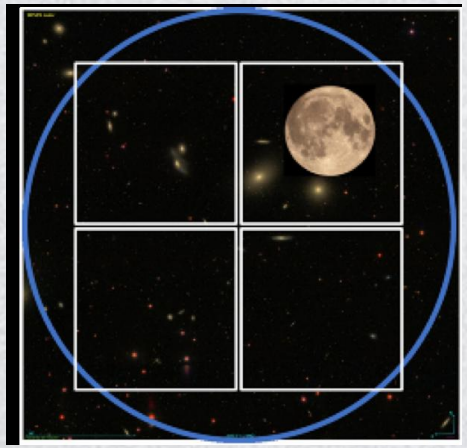
背景三：

- 云南大学1.6米多通道测光巡天望远镜落户丽江天文观测站5号点
- Mephisto 望远镜是主镜口径1.6 米、视场3.14 平方度、配备3 台CCD 相机的大视场成像巡天望远镜。
- 同时在3 个波段（ugi 或vrz）拍摄同一天区的高质量测光图像，获取天体高精度多波段星等及实时颜色信息，录制天体运动、变化的“彩色纪录片”。
- 高效率: 1200平方度/10小时@20秒曝光 x 3个波段，与LSST相当
- 开展 1）北半球可观测天区多波段多历元测光巡天；2）采样频率为天、小时以及分钟量级的变源与暂现源巡天。
- 国际一流水平的科研数据产品



二、分中心建设背景及关键需求

背景三：1.6米望远镜的数据需求



E2v CCD芯片：9216*9232
单台相机：2*2拼接，4片
三台相机，16bit读出
2*20s曝光（1min内）
9216*9232*4*3*2*2 = 4.1GB/min

每晚观测10小时估计：
4.1GB*10hr*60min = 2.46TB（3TB）
每年：300个可观测夜
2.46TB（3TB）*300 = 738TB/yr

Pilot Andor：
芯片：6144*6160，三台相机，16bit读出
6144*6160*3*2*2 = 454MB/min
每晚10小时：454MB/min*10hr*60min = 272GB
每年（300可观测夜）：272GB*300 = 81.8TB

最终的数据中心规模：
数据存储量：**50PB**，
计算节点：**4000个核**，
超算平台内主干网络用IB网络的架构

I期建设规模：
≥2PB存储容量；~600个核计算节点，双精度算力达到60TFlops以上；高性能计算平台内主干网络用IB网络的架构

ZTF算力统计

服务器	主频（GHz）	核/CPU	CPU数	算力/台（TFlops）	数量	总核数	算力（TFLOPS）
Dell E5-2640V3	2.6	16	2	2.6624	32	1024	85.1968
Dell E5-2640V4	2.4	20	2	3.072	34	1360	104.448

二、分中心建设背景及关键需求

一期建设关键需求：

- 建设标准化的机房。（包括机房装修、模块化机房建设）
- 存储容量≥2.0PB的大数据存储设备。
- 存储设备采用多级存储的存储，以满足实时数据处理的要求。
- 一期建设中计算节点只规划CPU计算节点，包含并行计算节点和一台串行大内存计算节点。
- 所有计算节点的双精度计算能力总量达到60-70TFlops。
- 高性能计算平台内部实现全线速主干网络架构，主干网络通信速度达到100Gbps。
- 计算网络核心交换机采用200Gbps的IB交换机，管理网络采用千兆/万兆的交换机。

机房基础装修

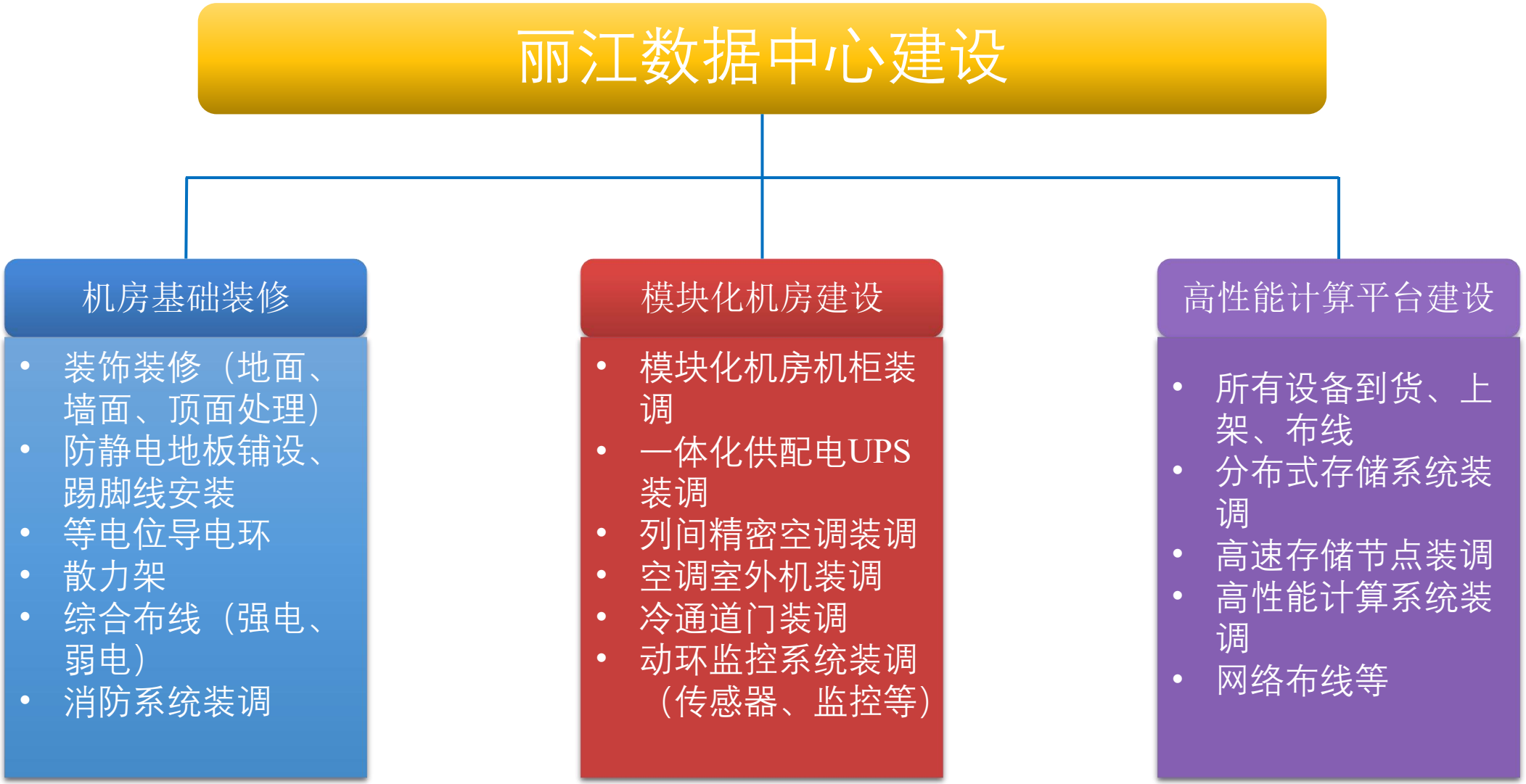
模块化机房建设

高性能计算平台建设

三、分中心建设进展

基础建设（2020年-2021年）

三、分中心建设进展



- **项目从3月30日基础装修进场开始，到6月15日超算系统培训结束，建设任务基本完成，进入试运行**

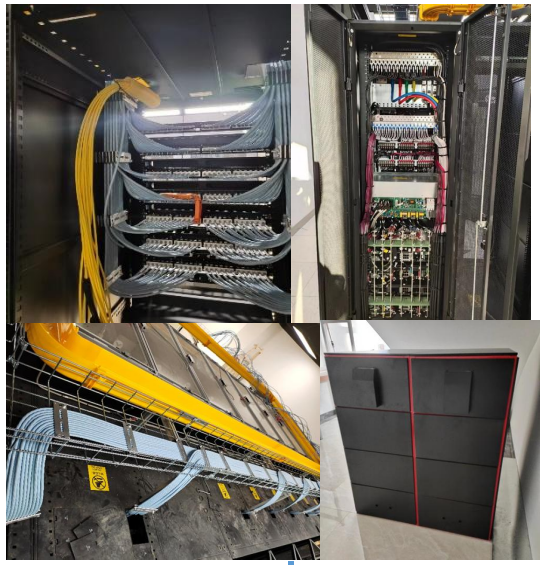
三、分中心建设-模块化机房

模块化机房建设时间线:

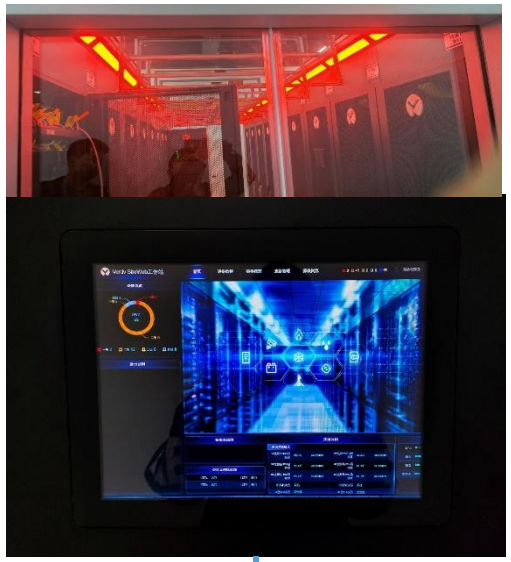
03.30基础装修进场

04.07

04.10



04.25

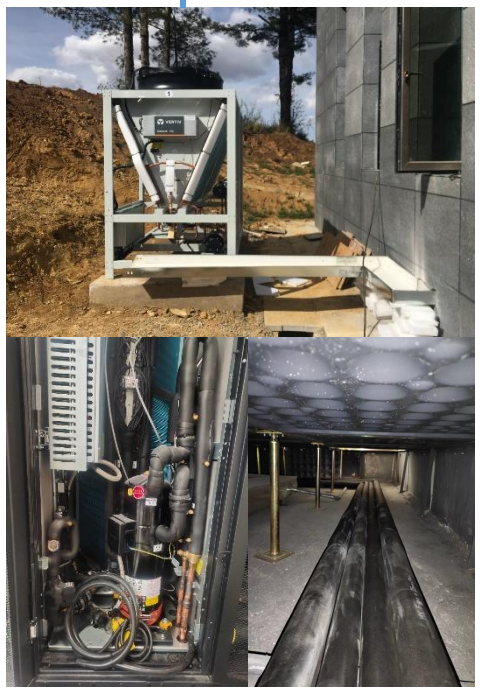


05.31

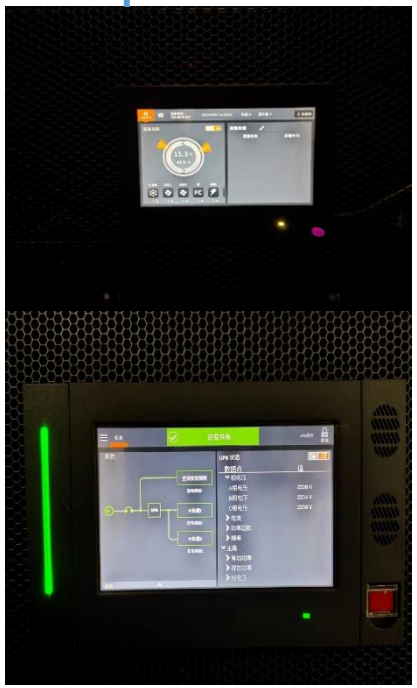
04.09

04.15

04.21

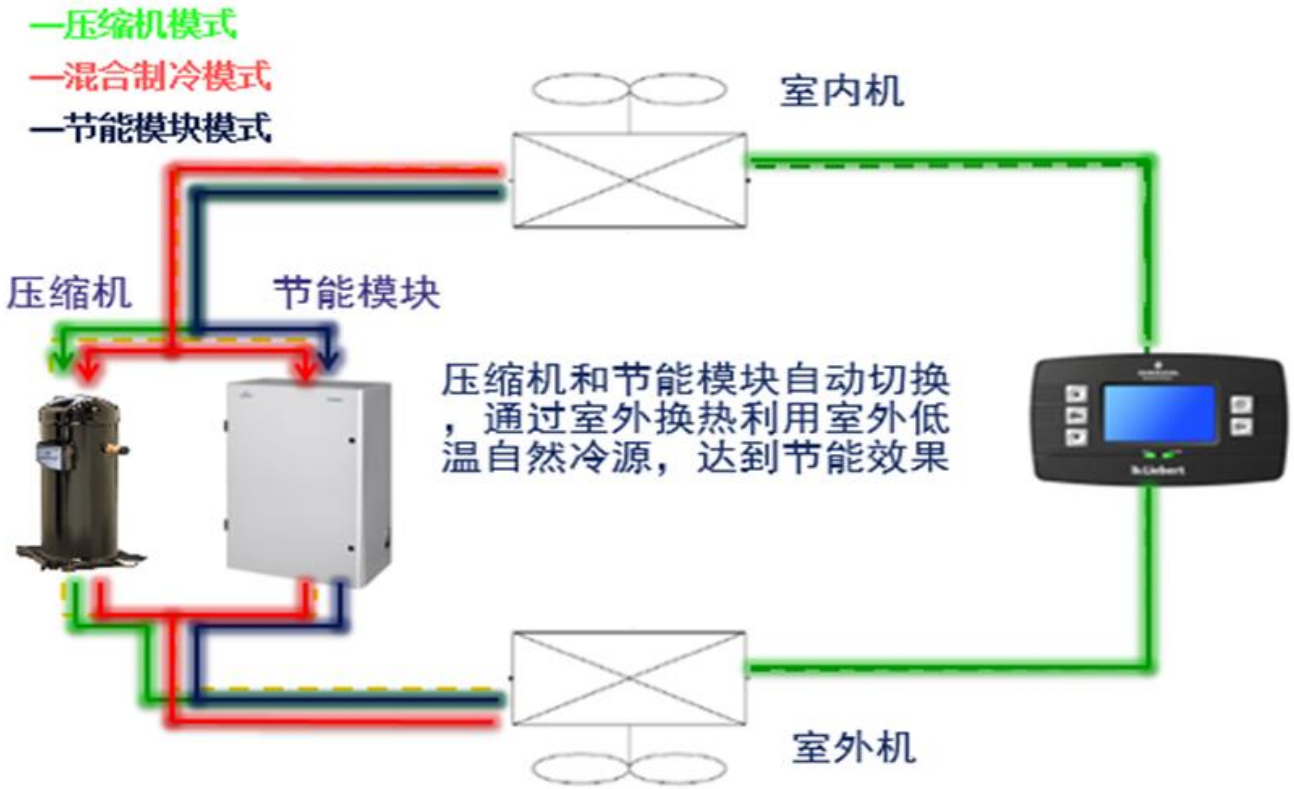


04.25



三、分中心建设-模块化机房

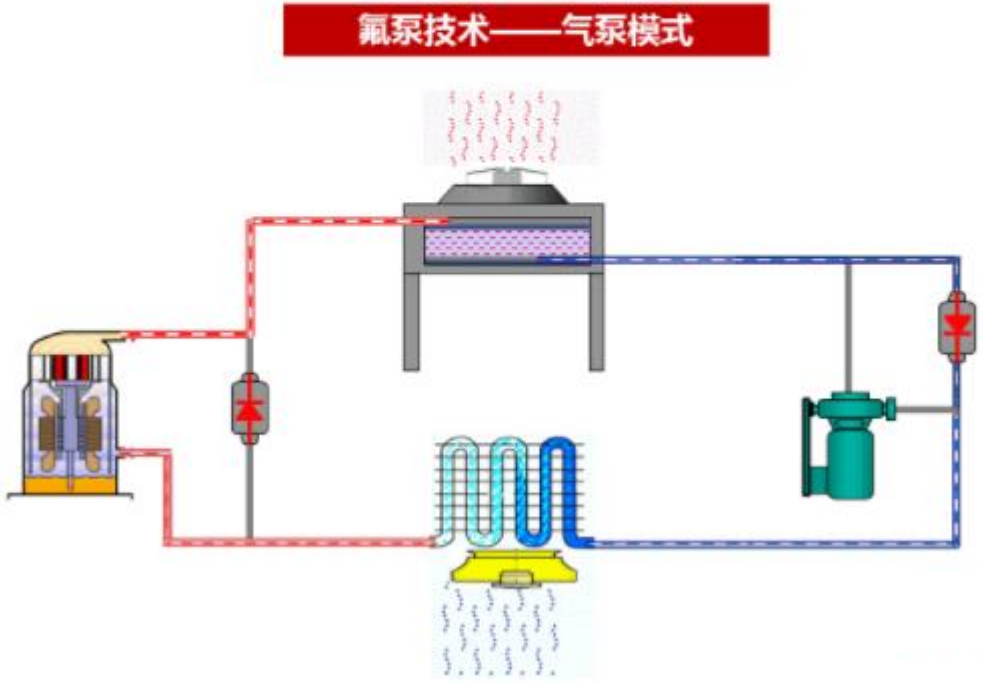
三、分中心建设 - 模块化机房
制冷系统-氟泵



三、分中心建设 - 模块化机房
制冷系统-氟泵

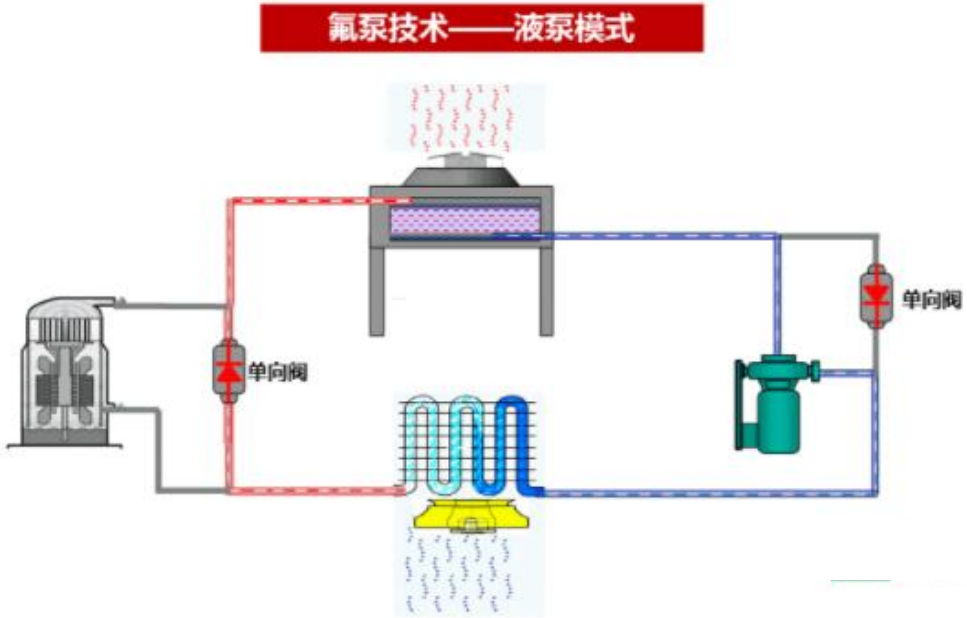
工作模式一：压缩机模式

当室外环境温度较高时，如室外环境温度>20℃时，氟泵开始压缩机制冷模式，这时压缩机正常运行，氟泵停止工作，系统类似于一台普通的机房空调；



工作模式二：节能模式

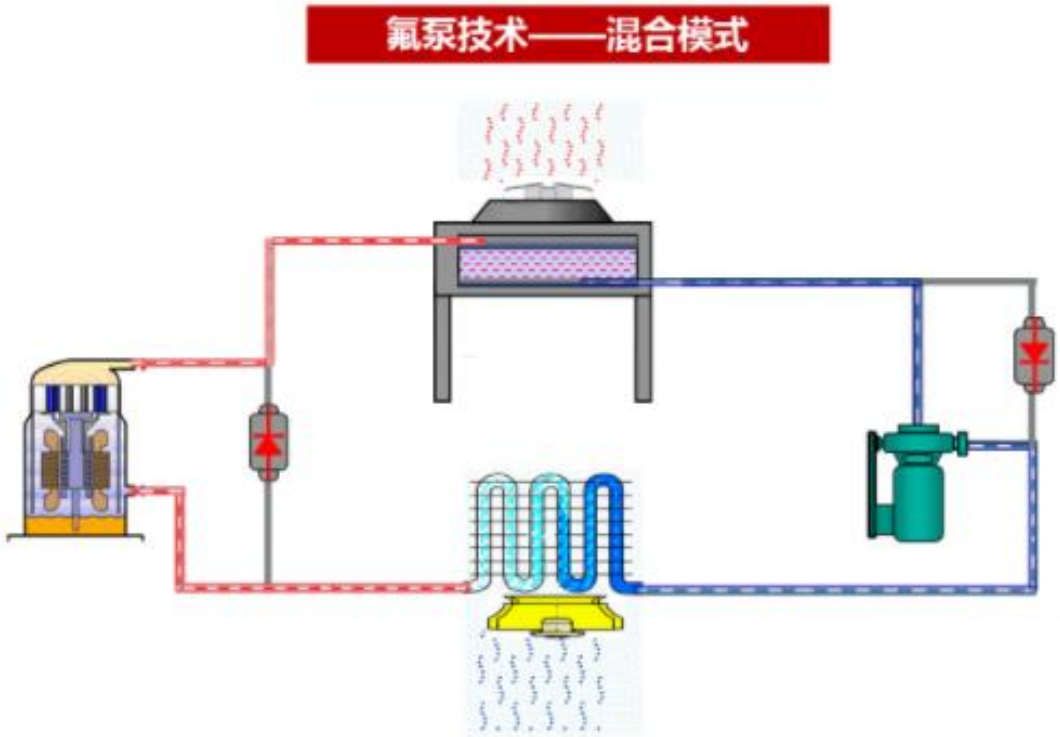
当室外环境温度较低，达到系统控制的设定点时，如室外环境温度<10℃时，,这时压缩机停止工作，氟泵启动。蒸发器中与室内空气换热后的制冷剂，直接进入风冷冷凝器与室外冷源进行换热，冷却成液态后的制冷剂在氟泵的作用下克服管阻回到蒸发器继续换热，达到节能效果。



三、分中心建设 - 模块化机房
制冷系统-氟泵

工作模式三：混合模式

当室外环境温度略低时，如10℃<室外环境温度<20℃时，这时系统处于混合模式，压缩机和氟泵同时工作，不过这时压缩机变频运行，转速较慢，达到节能效果，这方面PEX4.0是其中的佼佼者。

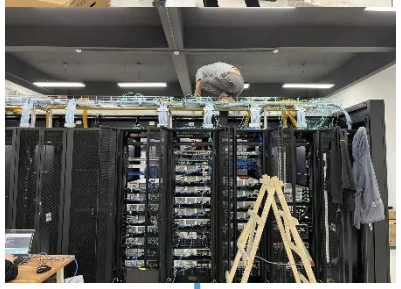


三、分中心建设-超算系统装调

超算系统装调时间线:

05.16

05.23



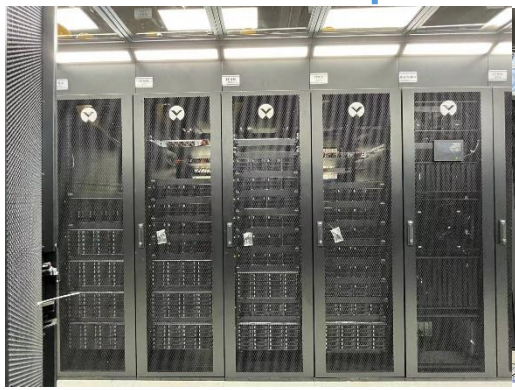
06.10

06.15

05.17

06.08

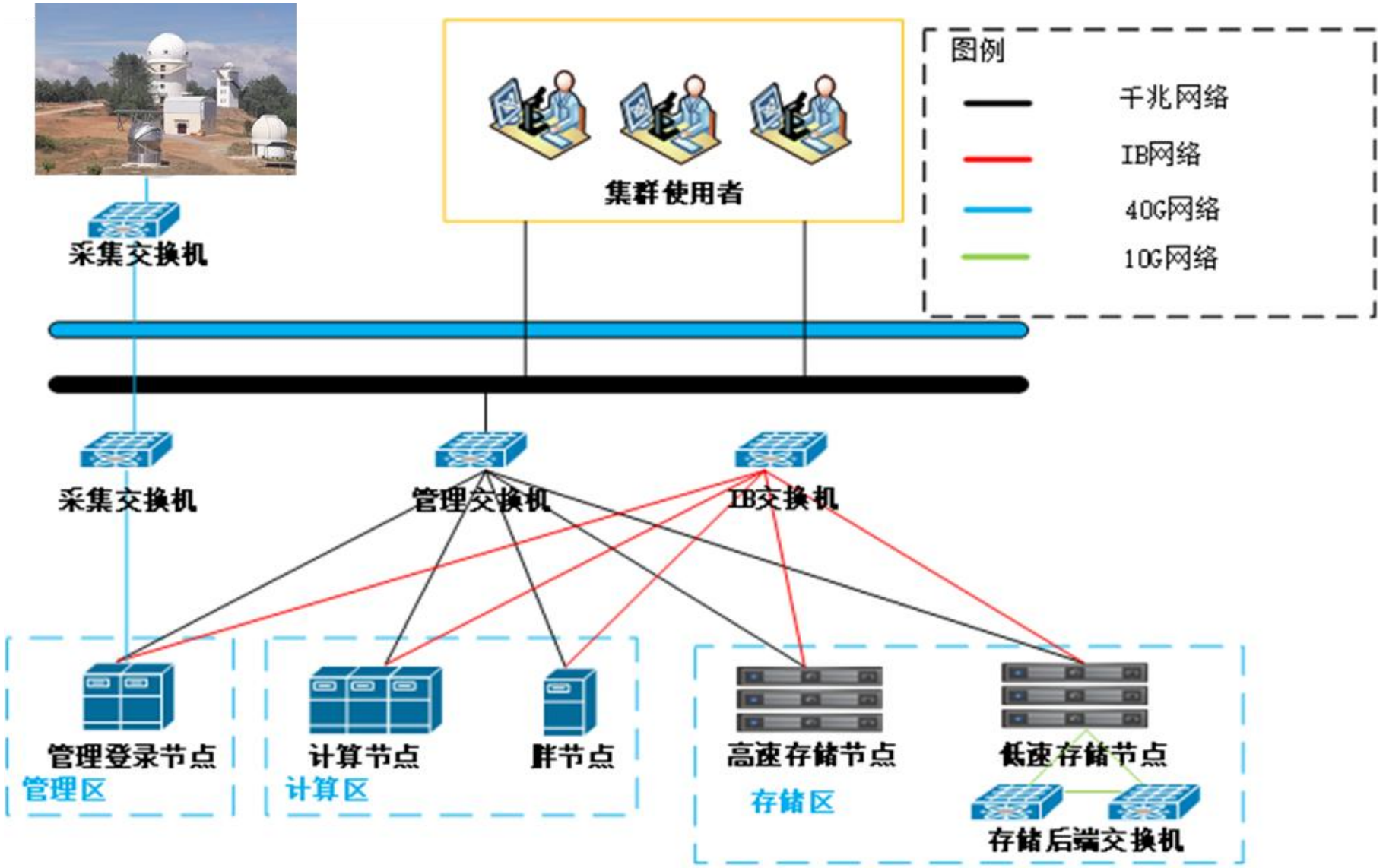
06.14



三、分中心建设-超算系统装调

三、分中心建设 – 超算系统

超算系统架构



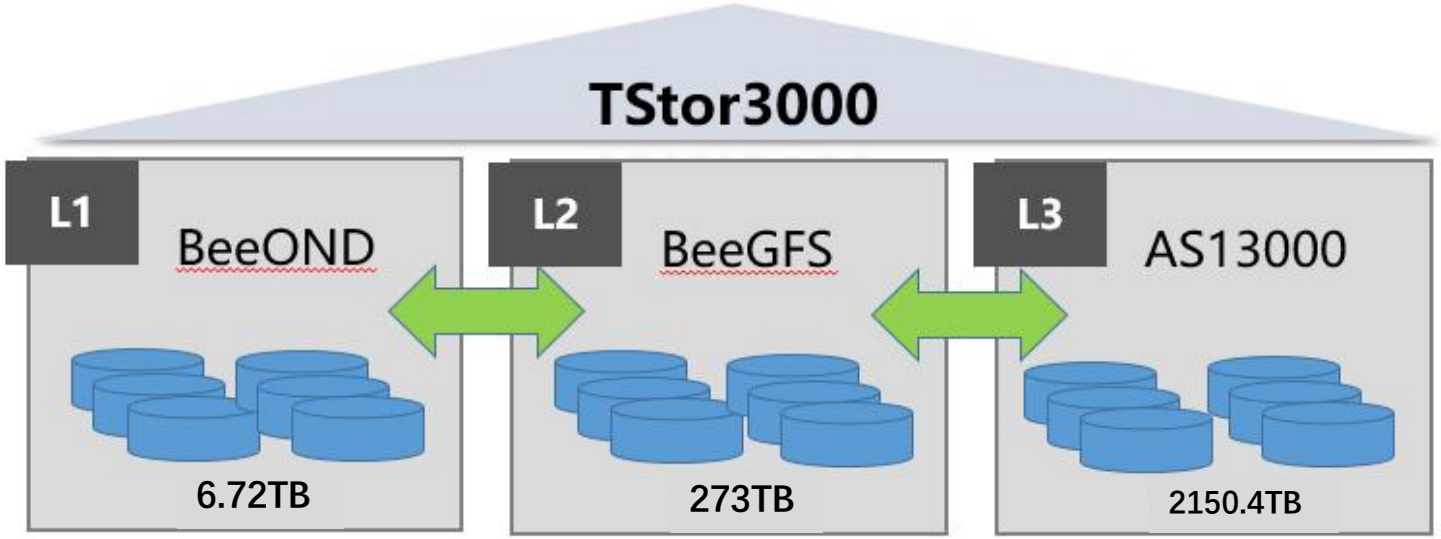
三、分中心建设 - 超算系统

超算系统上架图

机柜	A05	机柜	机柜	A04	机柜	机柜	A03	机柜	机柜	A02	机柜
42		42	42		42	42		42	42		42
41		41	41		41	41		41	41		41
40		40	40		40	40		40	40		40
39		39	39		39	39		39	39		39
38		38	38		38	38		38	38		38
37		37	37	存储/管理万兆交换机1-10GSW1	37	37	存储/管理万兆交换机2-10GSW2	37	37		37
36		36	36		36	36		36	36		36
35	采集交换机2-40Gsw1	35	35	IB交换机 IBSW1	35	35	千兆交换机1-1Gsw1	35	35		35
34		34	34		34	34		34	34		34
33	千兆交换机2-1Gsw2	33	33	管理节点2-ispim	33	33		33	33		33
32		32	32		32	32		32	32		32
31		31	31		31	31		31	31		31
30		30	30	计算节点6-cu06	30	30		30	30		30
29		29	29		29	29	计算节点11-cu11	29	29		29
28	登录管理节点1-ctl01	28	28		28	28		28	28		28
27		27	27	计算节点7-cu07	27	27		27	27	胖节点-fat01	27
26		26	26		26	26	计算节点12-cu12	26	26		26
25	计算节点1-cu01	25	25		25	25		25	25		25
24		24	24	计算节点8-cu08	24	24		24	24		24
23		23	23		23	23	计算节点13-cu13	23	23		23
22	计算节点2-cu02	22	22		22	22		22	22	低速存储节点1-inspur01	22
21		21	21	计算节点9-cu09	21	21		21	21		21
20		20	20		20	20	计算节点14-cu14	20	20		20
19	计算节点3-cu03	19	19		19	19		19	19		19
18		18	18	计算节点10-cu10	18	18		18	18		18
17		17	17		17	17	低速存储节点5-inspur05	17	17	低速存储节点2-inspur02	17
16	计算节点4-cu04	16	16		16	16		16	16		16
15		15	15	低速存储节点8-inspur08	15	15		15	15		15
14		14	14		14	14		14	14		14
13	计算节点5-cu05	13	13		13	13		13	13		13
12		12	12		12	12	低速存储节点6-inspur06	12	12	低速存储节点3-inspur03	12
11		11	11		11	11		11	11		11
10	高速存储数据节点1-mds01	10	10	高速存储数据节点2-mds02	10	10		10	10		10
9		9	9		9	9		9	9		9
8		8	8		8	8		8	8		8
7	高速存储数据节点1-oss01	7	7	高速存储数据节点2-oss02	7	7	低速存储节点7-inspur07	7	7	低速存储节点4-inspur04	7
6		6	6		6	6		6	6		6
5		5	5		5	5		5	5		5
4		4	4		4	4		4	4		4
3		3	3		3	3		3	3		3
2		2	2		2	2		2	2		2
1		1	1		1	1		1	1		1

三、分中心建设 – 超算系统

存储架构

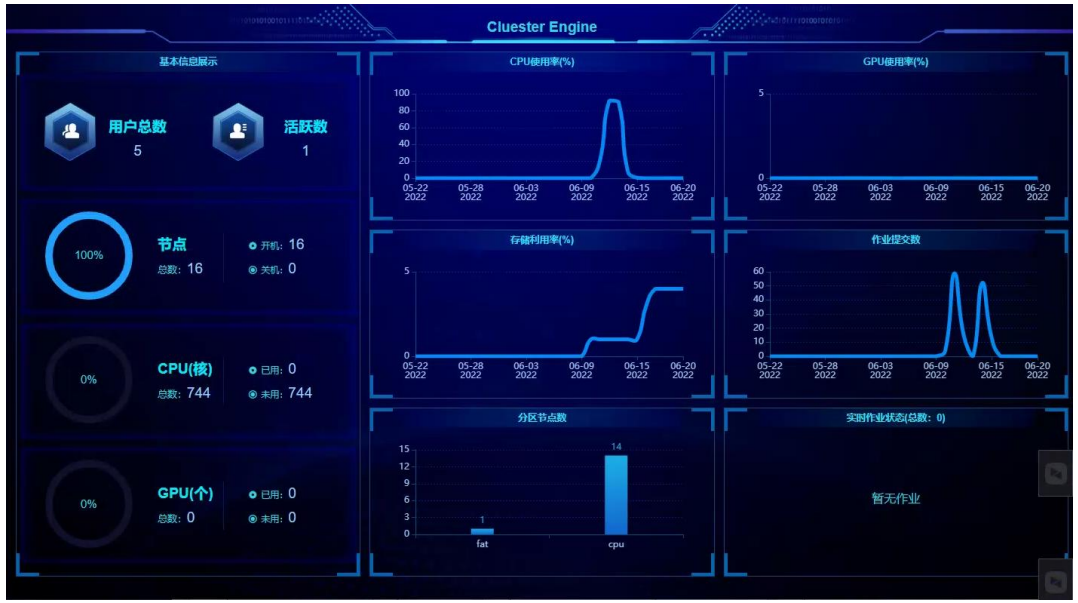


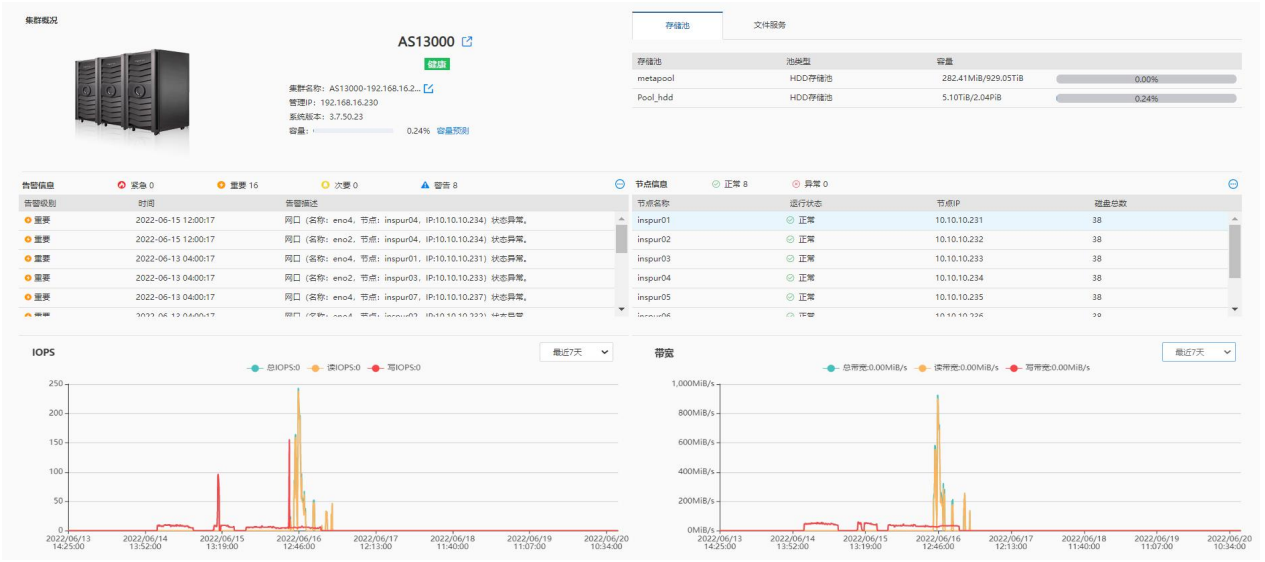
存储架构	可用容量
L1：BeeOND	480GB*14=6720GB
L2：BeeGFS	10TB*36块*1节点*0.76=273TB（双路在线）
L3：AS13000	12TB*32块*8节点*0.8=2150.4TB
共计	2430.12TB（2.43PB）

三、分中心建设 – 超算系统资源

序号	设备名称	型号	数量	备注	
1	并行计算节点	NF5280M6 6342(CPU)	14台	每个节点配置：2颗2.4GHz的CPU，24核/CPU，256GB内存 整个集群：14个节点，共：**672核**，算力达到**60TFlops**	
2	胖节点	NF8480M6	1台	4颗3.1GHz的CPU，18核/CPU，**1TB**内存，算力达到**7TFlops**	
3	管理登录节点	NF5280M6 4314(CPU)	2台	每个节点：2颗2.4GHz的CPU，16核/CPU，128GB内存	
4	高速存储系统	AS13000G6-H12元数据节点	2台	2颗2.4GHz的CPU，16核/CPU，128GB内存	先做RAID，然后元节点和数据节点互为备份，可用存储容量达到**274TB**
		AS13000G6-H36高速储存数据节点	2台	2颗2.3GHz的CPU，20核/CPU，128GB内存	
5	低速存储系统	AS1300G5-M36	8台	每个节点：2颗2.4GHz的CPU，12核/CPU，128GB内存 采用纠删码（6+2:0）模式，可用存储空间达到**2PB**	
6	IB HDR高速网络	Mellanox QM8790	1台	40个200Gb/s全线速的端口	
7	40G数据采集交换机	CN6132Q-V	2台	4个固定SFP+ 10G端口，30个固定40G QSFP+端口	
8	存储/管理万兆交换机	CN93240YC-FX2	2台	48个10G 端口，12个100G端口	
9	千兆交换机	S655OV2-48TQ-AC/D	2台	48个千兆端口	
10	标准机柜	维谛	10台	42U标准列间机柜，柜深1100mm，每个机柜配备2个防雷PDU	

三、分中心建设 – 超算系统性能测试






```
[root@cu01 ~]#
[root@cu01 ~]# ib_write_bw -i 1 -F -d mlx5_0 --report_gbits ibctl01
---------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 PCIe relax order: ON
 TX depth        : 128
 CQ Moderation   : 1
 Mtu             : 4096[B]
 Link type       : IB
 Max inline data : 0[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------
 local address: LID 0x18 QPN 0x002f PSN 0x8560ff RKey 0x002485 VAddr 0x002b82e0110000
 remote address: LID 0x17 QPN 0x0052 PSN 0xf3e952 RKey 0x00f515 VAddr 0x002b519d570000
---------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 65536      5000             98.68              98.68              0.188215
---------------------------------------------------------------------------------
[root@cu01 ~]#
```

```
[root@ctl01 ~]#
[root@ctl01 ~]# ib_write_bw -i 1 -F -d mlx5_0 --report_gbits

************************************
* Waiting for client to connect... *
************************************
---------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 PCIe relax order: ON
 CQ Moderation   : 1
 Mtu             : 4096[B]
 Link type       : IB
 Max inline data : 0[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------
 local address: LID 0x17 QPN 0x0052 PSN 0xf3e952 RKey 0x00f515 VAddr 0x002b519d570000
 remote address: LID 0x18 QPN 0x002f PSN 0x8560ff RKey 0x002485 VAddr 0x002b82e0110000
---------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 65536      5000             98.68              98.68              0.188215
---------------------------------------------------------------------------------
[root@ctl01 ~]#
```

四、下一步工作计划

- **试运行**:

- 对高性能计算集群的性能进行继续的试运行测试
- 对整个模块化机房的性能进行测试
- 对硬件设备在3200米海拔的地方的运行状态进行监控

- **提供服务**:

- 为即将建设的1.6米多通道测光巡天望远镜提供存储和计算服务
- 为其他需要大容量存储和高性能计算设备的应用提供服务
- 总结I期建设的经验、存在的问题
- 根据不断增加的需求，逐步开展II期的建设

四、下一步工作计划

超算平台测试案例征集

有意使用平台的代表请与我联系：

手机：13398884033（微信同号）

邮箱：wcj@ynao.ac.cn


请各位专家和老师批评指正！