

面向海量数据的天文科研云平台

李长华

中国科学院国家天文台
国家天文科学数据中心

中国科学院国家天文台简介



中国科学院国家天文台（简称“国家天文台”）成立于2001年4月，国家天文台本部设在北京，直属单位包括中国科学院云南天文台、南京天文光学技术研究所、新疆天文台和长春人造卫星观测站。

国家天文台主要研究方向包括天文学基础前沿研究、天文技术方法、观测装置建造运行和空间探测科学应用。

报告提纲

- 天文领域的大科学装置
- 中国虚拟天文台平台现状
- 天文科研云平台构想与展望

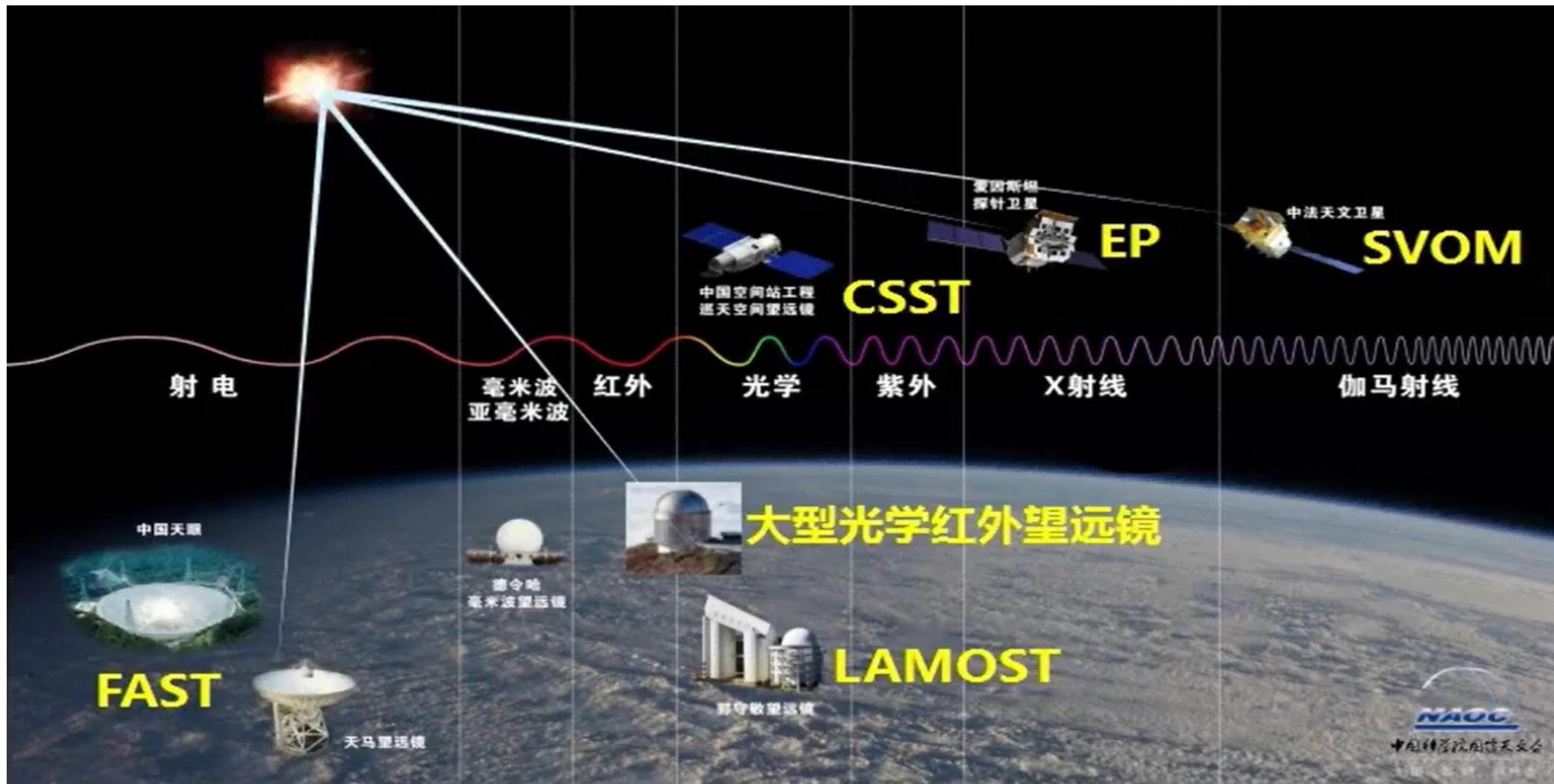


天文学：数据驱动的科学



China-VO

天文学：数据驱动的科学



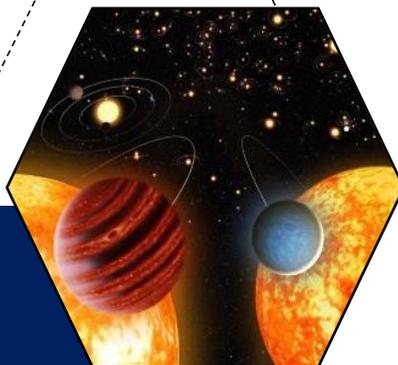
郭守敬望远镜 (LAMOST)

大天区面积多目标光纤光谱天文望远镜

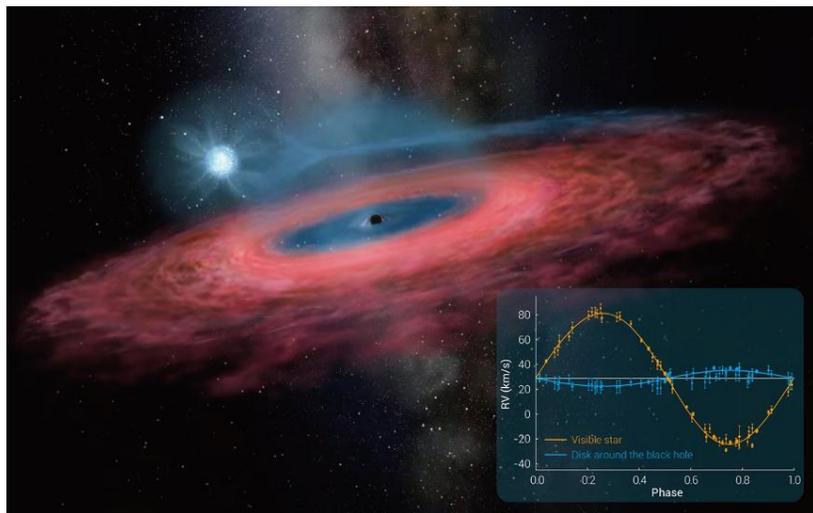
- “九五”国家重大科学工程，2008年落成
- 国际口径最大的大视场光谱巡天望远镜
- 世界上光谱获取率最高的望远镜
- 攻克系列技术壁垒，突破“大视场与大口径难以兼得”难题
 - 首次在一块大镜面上同时应用薄变形镜面和拼接镜面主动光学技术
 - 首次实现六角形的主动可变形镜
 - 首次在一个光学系统中同时采用两块大口径拼接镜面
 - 首次应用4000根光纤定位技术

已完成天体光谱2000+万条，是世界上
其它巡天项目发布光谱数总和的1.8倍

发表SCI论文700多篇，在搜寻奇异天体、给银河系重新画像等
研究领域取得了一批重大成果



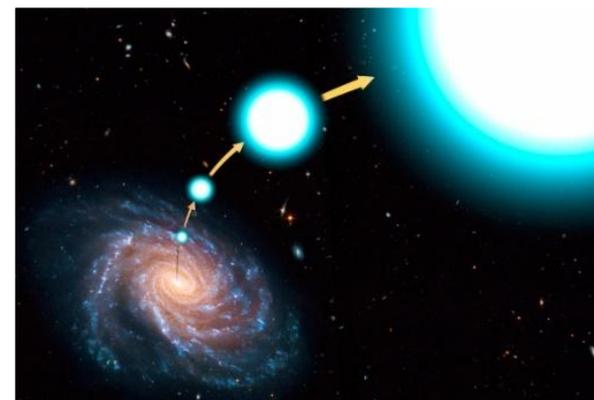
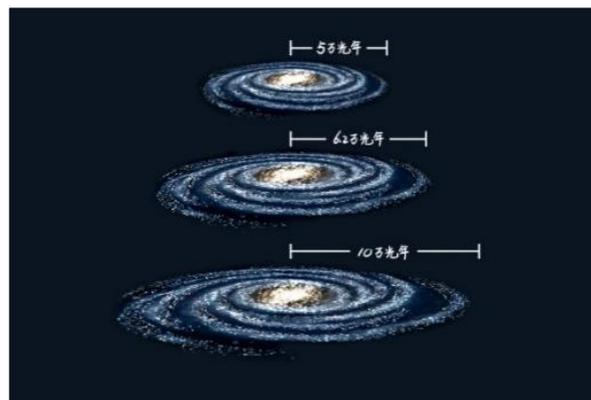
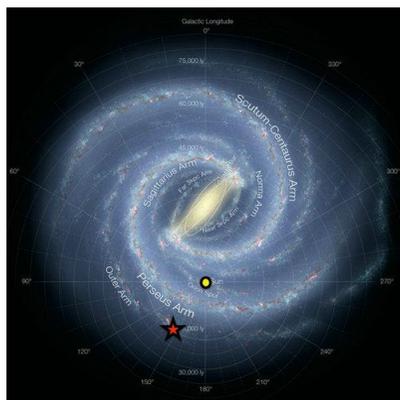
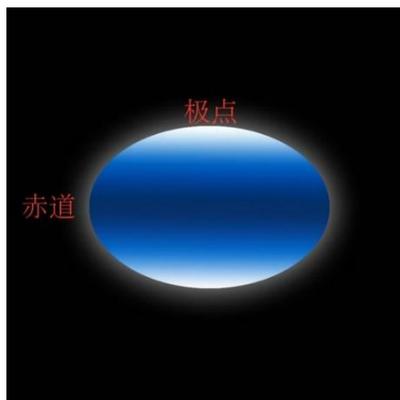
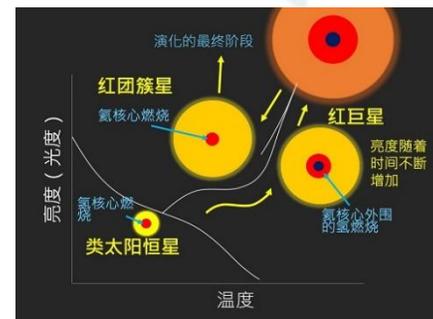
基于LAMOST数据的典型科学发现



←
发现迄今最大的恒星级黑洞



揭示银河系外盘翘曲起源与演化 ↑



↑ 发现迄今银河系中自转最快恒星

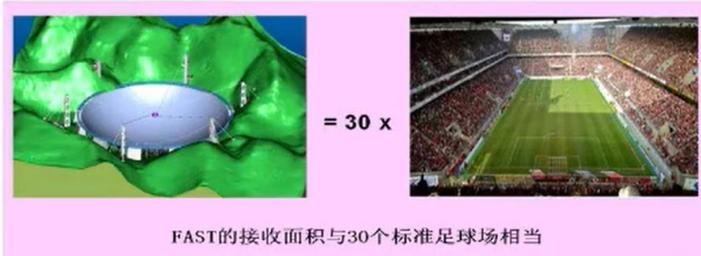
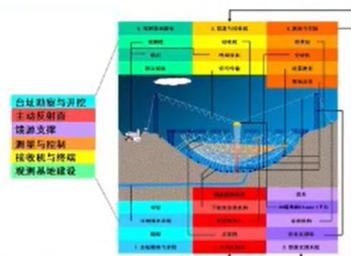
↑ 重新认识银河系的大小

↑ 发现超高速星

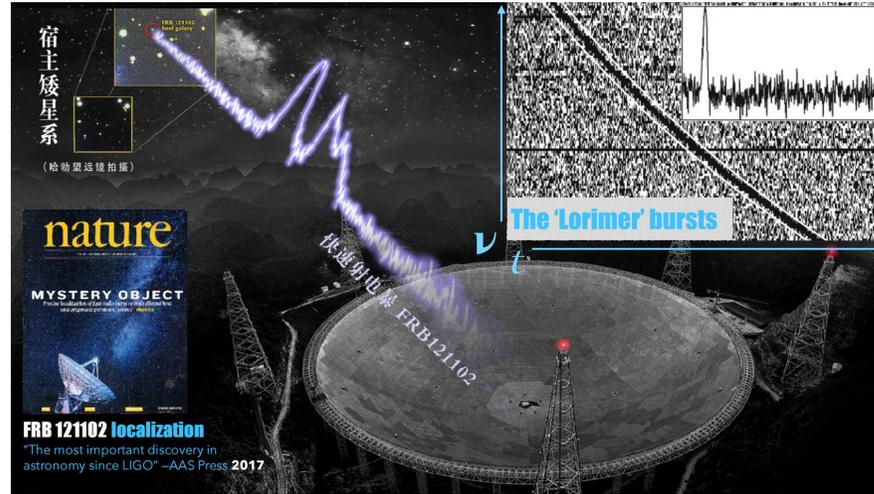
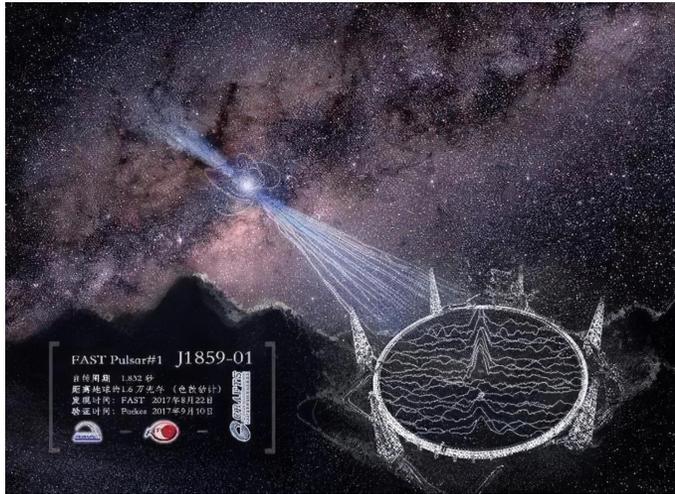
FAST-中国天眼

500 米口径球面射电望远镜

- “十一五”国家重大科技基础设施，2016年落成
- 世界最大单口径射电望远镜，灵敏度最高的射电望远镜，使我国中低频射电天文实现从追赶到领先的跨越发展
- 三项自主创新
 - 利用贵州天然喀斯特巨型洼地独特台址
 - 自主发明主动变形反射面
 - 自主提出轻型索拖动馈源支撑系统和并联机器人



基于FAST数据的典型科学成果



FAST 捕获世界最大快速射电暴样本

Li et al. 2021 Nature
Largest sample of FRB bursts, first complete energy distribution.

nature FIELD GUIDE
Magnetic crucible for the interstellar medium and how stars form

Ching et al. 2022 Nature
First HINSA Zeeman detection, challenges canonical star formation picture.

Science

Feng et al. 2022 Science
A unified picture of polarization evolution of repeater FRBs based a joint FAST-GBT survey

CRABTS Discovers Younger and Weirder Brother of the first repeater Fast Radio Burst

Niu et al. 2022 Nature
The first persistently active FRB, hinting at an evolutionary picture of FRB origin.

- ### China's Top 10 scientific breakthroughs in 2021
1. Tianwen 1 landed on Mars
 2. China's space station core Tianhe in orbit
 3. Synthesizing starch from carbon dioxide
 4. Chang'e-5 returned with lunar rocks
 5. Cryo-EM structure of an extended SARS-CoV-2
 6. FAST caught largest set of fast radio bursts
 7. High-performing woven lithium-ion fiber batteries
 8. Programmable superconducting quantum processor
 9. Soft robot 10,000 meters under the ocean's surface
 10. Spatio-temporal dynamics of bird migration routes

FAST 2021 捕获世界最大快速射电暴样本

2021年度中国科学十大进展

FAST 2021 捕获世界最大快速射电暴样本

FAST 2021 捕获世界最大快速射电暴样本

FRB Detection through direct AI

AI detection framework

Taking detection as image classification problem

Classification: True/False

Network location

Using techniques to improve both the efficiency and accuracy of AI algorithm

Knowledge Distillation Semi-supervised

AI与快速射电暴

Superior AI

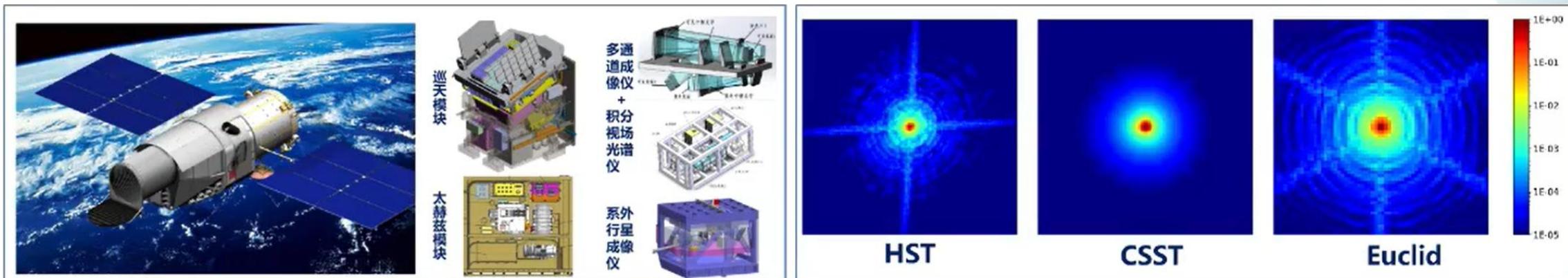
Algorithm	Detected FRB	Processing Time
Traditional	5595	16.1
AI	6439	

Signal detected by traditional (left) and AI (right)

noisy signals newly detected by AI

中国空间站工程巡天空间望远镜 (CSST)

- 可在轨维护的 2 米空间望远镜
- 分辨率与哈勃空间望远镜相当，**视场超过其300倍**
- 兼具大视场巡天和精细观测能力
 - ✓ 大面积、高分辨率的多色成像与无缝光谱巡天观测（普查）
 - ✓ 对遴选天体/天区的精细研究（详查）
- 计划 2024 年发射

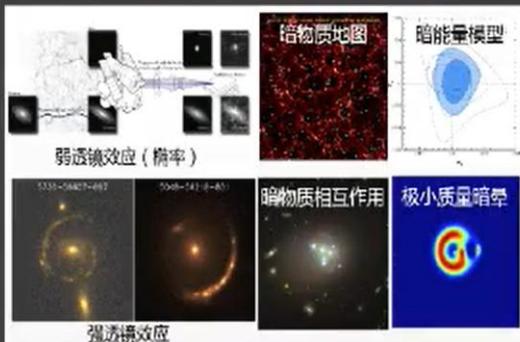


中国空间站工程巡天空间望远镜 (CSST)

- 可
- 分
- 兼
- ✓ 计
- ✓ 划
- 计

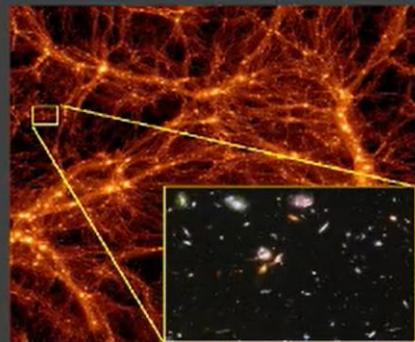


宇宙学：引力透镜效应



测绘宇宙暗物质三维分布；验证暗能量是否演化；精确测量哈勃常数，检验广义相对论

宇宙学：大尺度结构



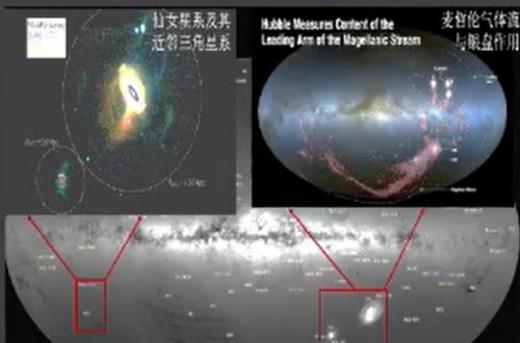
建立最大规模星系红移样本，验证大尺度结构演化理论，探究暗物质和暗能量属性

星系与活动星系核



构建完备的早期宇宙的星系和类星体样本，揭示其形成与演化的规律

近邻宇宙与恒星科学



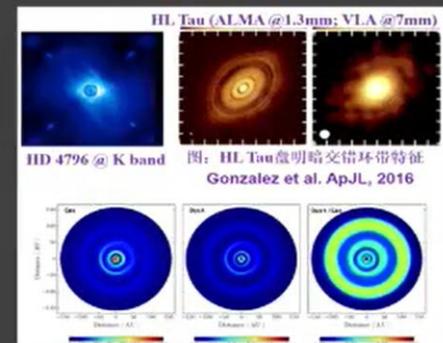
精细刻画银河系与近邻星系的结构、近邻星系的分布与相互作用、恒星的形成与演化

天测、太阳系与暂现源

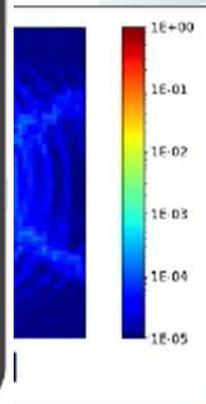


大幅提高暗弱天体的天球坐标和距离精度，发现大量太阳系小天体和暂现源

系外行星



搜寻围绕类太阳恒星的成熟系外行星，精确测量行星和行星盘物理属性



China

天文大数据时代

➤ TB 时代 (~2000)

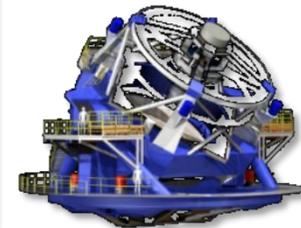
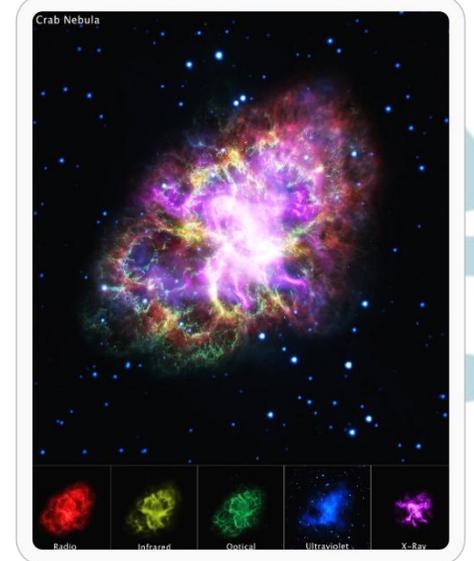
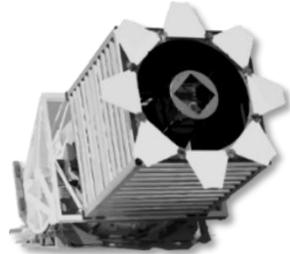
- **SDSS**
- **LAMOST**
- **Gaia**

➤ PB时代 (~2010)

- **FAST**
- **SKA**
- **LSST**
- **Euclid**
- **CSST**

➤ EB时代 (~2025)

- **FASTA**
- ...



报告提纲

- 天文领域的大科学装置
- **中国虚拟天文台平台**
- 天文科研云平台构想与展望



虚拟天文台与天文信息学

□ 虚拟天文台与国际虚拟天文台联盟

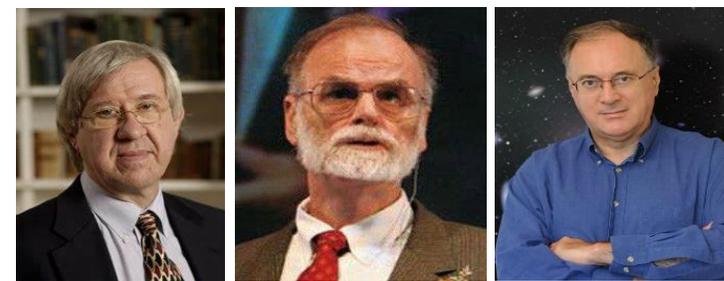
- ✓ 数据密集型网络化天文学研究与应用平台
- ✓ 数据发现-高效访问-互操作→数字宇宙

□ 中国虚拟天文台

□ 天文信息学

- ✓ 把先进的计算和分析方法应用到天文学领域，从海量数据中获得新发现

并行计算、分布式计算、网格计算、云计算、机器学习与人工智能技术 ...



Prof. Szalay

Dr. Jim Gray

Prof. Djorgovski



中国虚拟天文台平台

Development achievements in last decade

VO-DAS

VOFilter

OpenOffice.org

FitHAS

SkyMouse

FITS Manager

Astronomy Astrophysics

Candidate Milky Way satellites in the Galactic halo

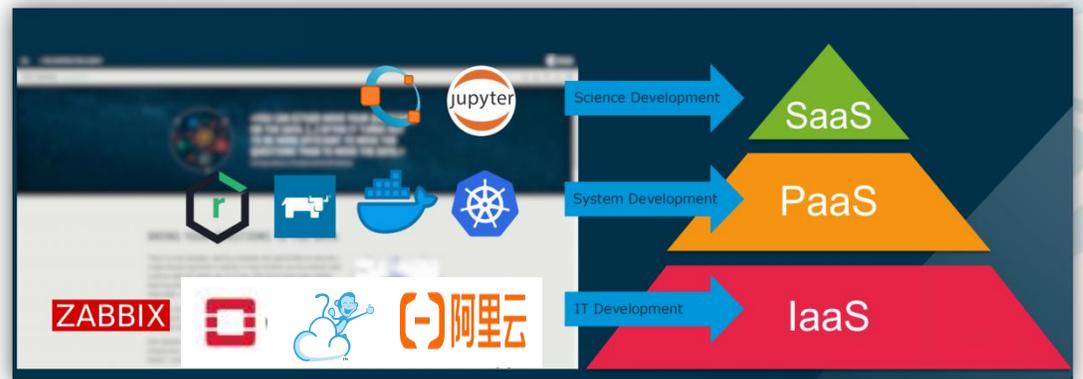
The Multiphase Halo of NGC 311: WYN H-alpha and BVI imaging

China 中国虚拟天文台

申请观测 科学数据 软件工具 **云资源** 公众频道

公众起新星搜寻项目

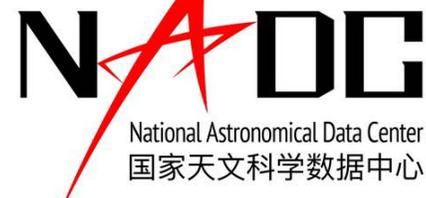
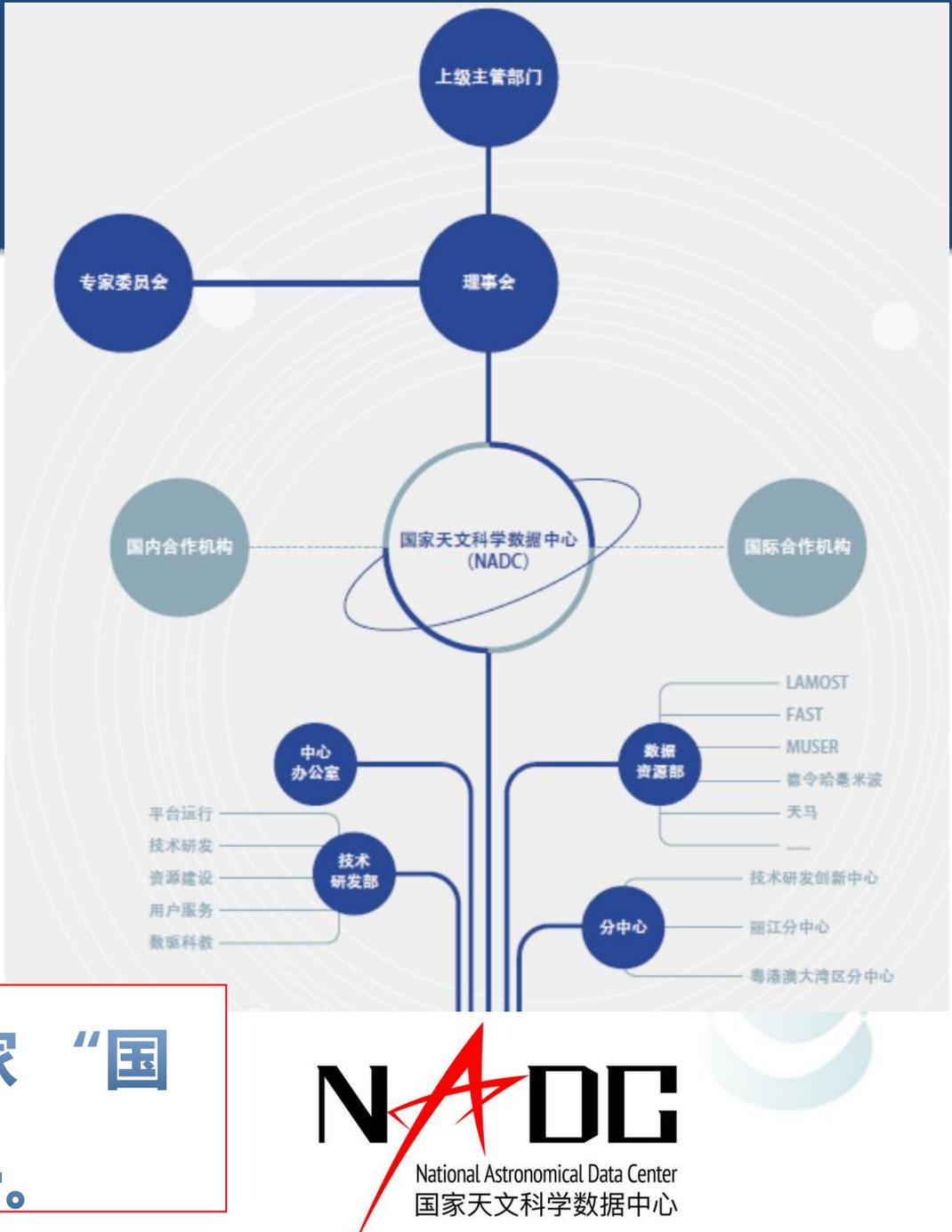
From *bring the data to the user*
 To *bring the user to the data*



The screenshot shows the Chinese version of the NADC website. It features a top navigation bar with 'Home', 'About', 'Data', 'Services', and 'Science Platform'. The main content area is divided into several sections: 'Selected Collections' with a list of data releases, 'Latest Collections', and 'Latest paper data'. Below these are 'Latest News' articles with images and titles, a 'Platform Statistics' table, and a 'Membership and Certification' section. The footer includes contact information and a QR code.

The screenshot shows the English version of the NADC website. It features a top navigation bar with 'Home', 'About', 'Data', 'Services', and 'Science Platform'. The main content area is divided into several sections: 'Selected Collections' with a list of data releases, 'Latest Collections', and 'Latest paper data'. Below these are 'Latest News' articles with images and titles, a 'Platform Statistics' table, and a 'Membership and Certification' section. The footer includes contact information and a QR code.

科技部认定的首批20家“国家科学数据中心”之一。



天文数据服务

NADC National Astronomical Data Center
国家天文科学数据中心

Observation Data Featured Service Cloud Public

input coordinate or target name 30 arcsec Find Dataset

Sample: 15.51967109 -0.28502335, m31

Domestic Dataset

AST3 Image	🔍	AST3 Light Curve	📄	🔍
AST3 Survey	📄	BASS DR1 gradd	★	🔍
BASS DR1 Images	📄	BASS DR2 gradd	★	🔍
BASS DR2 Images	★	BASS DR2 Stack	★	🔍
BASS DR3 coadd	★	BASS DR3 Images	★	🔍
BASS DR3 Rawdata		BASS DR3 Stack	★	📄
CSTAR	🔍	GMG2-4m		🔍
GPS1 Plus	🔍	LAMOST Quasar Catalog	★	🔍
Legacy Plate	🔍	SCUSS Catalog		🔍
SCUSS Image	🔍	SCUSS Proper Motion		🔍
SEDRS	🔍	Wide Binary		🔍

Solar Dataset

LoFRE	🔍	MUSER		🔍
SBR5	🔍			

Other

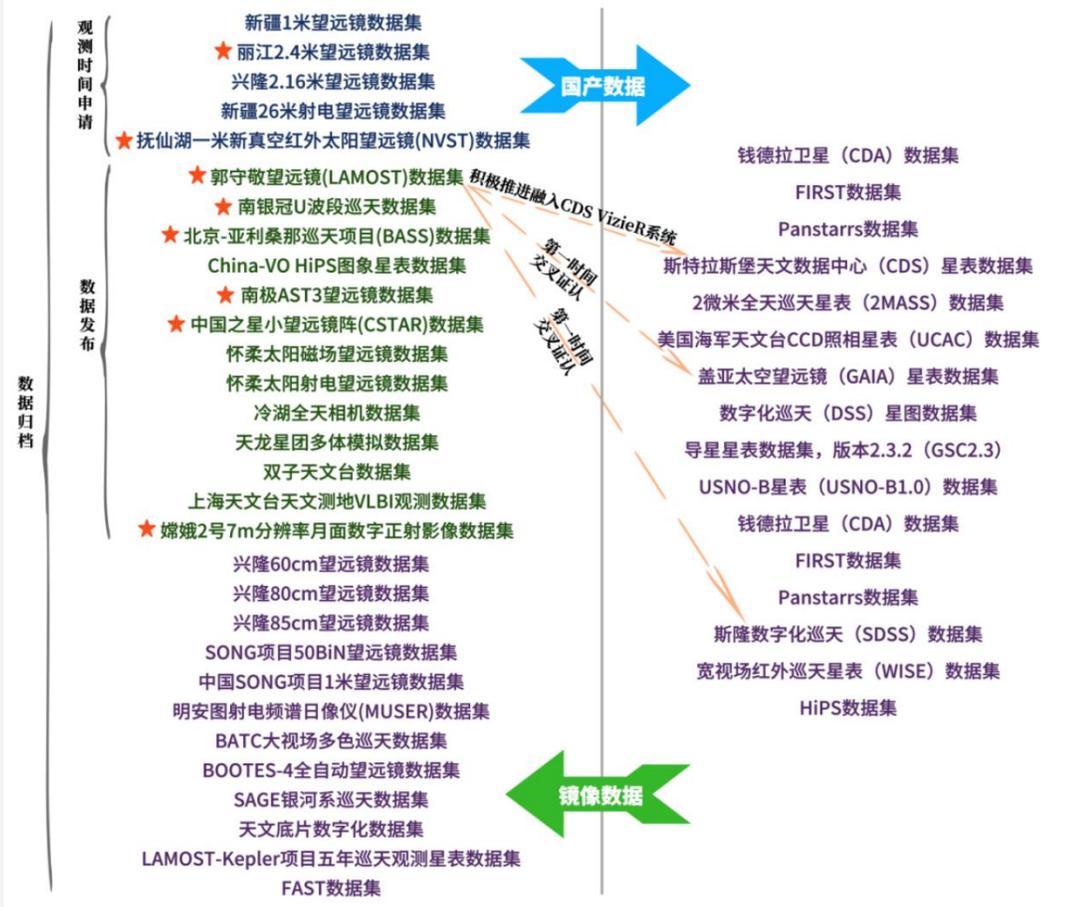
GAIA DR2 Source	📄	GAIA EDR3 Source	📄	🔍
LegacySurvey DR9 Tractor	🔍	Pan-STARRS DR1 OTMO		🔍
Pan-STARRS DR1 Stack Object Thin	🔍	SkyMapper DR2		🔍
TwoMass	📄	UCAC	📄	🔍
WISE	📄	Dragon on Silkroad	📄	🔍
GAIA DR1	📄	Cajobs		🔍
SDSS DR12	📄	SDSS DR14	📄	🔍
VizieR	📄			🔍

Links

IVOA China-VO NAOC CDS NASA/IPAC NED NASA/IPAC IRSA MAST Portal NASA/ADS arXiv.org (Astrophysics)

Certified **Member of**

中国虚拟天文台数据资源一览表



注：加星标的数为曾获十大天文科技进展项目所产出的数据集

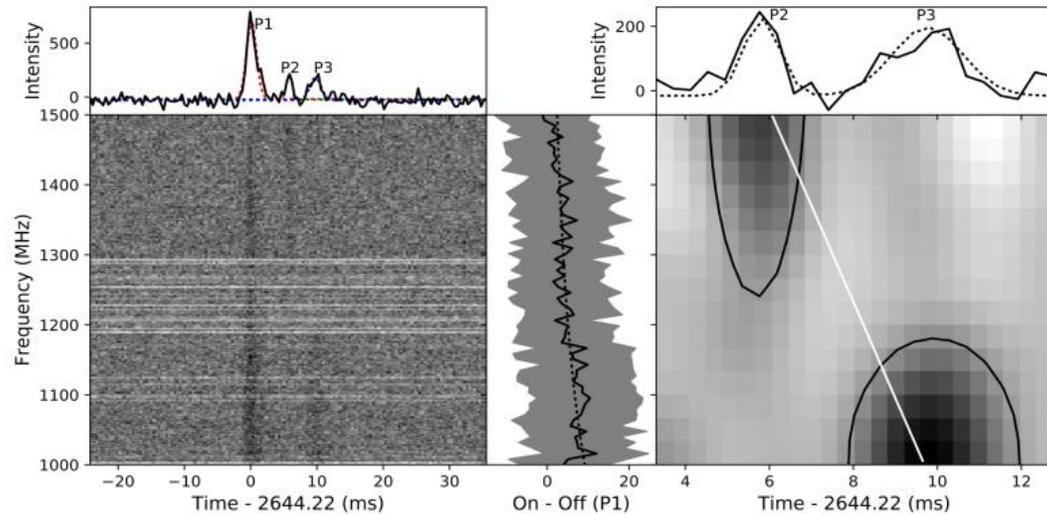
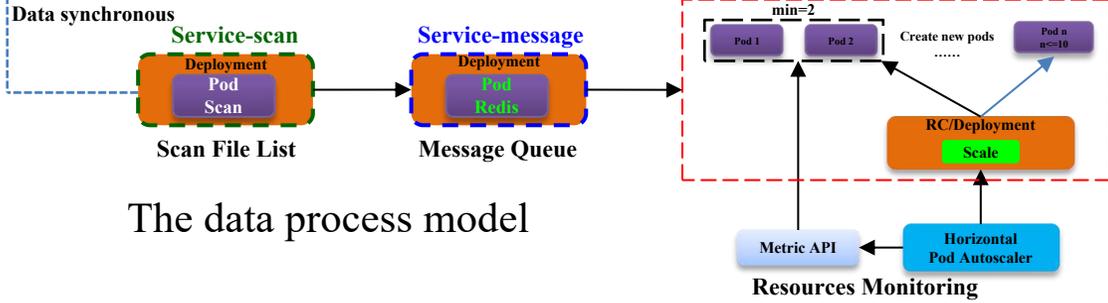
天文云资源服务

The screenshot shows the NADC cloud resource management interface. At the top, there is a navigation bar with '登录' (Login), '退出' (Logout), '注册' (Register), and 'English'. Below this, the NADC logo and name are displayed, along with menu items like '申请观测' (Apply for observation), '科学数据' (Scientific data), '主题服务' (Thematic services), '云资源' (Cloud resources), and '公众频道' (Public channel). The main content area displays the current node as '国台' (Mainland China) and provides a summary of VM status: 5 running VMs (正在运行的 VM), 28 stopped VMs (已停止的 VM), and a total of 33 VMs (总 VM 数). A sidebar on the left contains navigation options such as '控制板' (Dashboard), '虚拟机' (Virtual machines), '我的VO空间' (My VO space), '论文库' (Paper library), '资源节点' (Resource nodes), '模板' (Templates), and '事件' (Events).

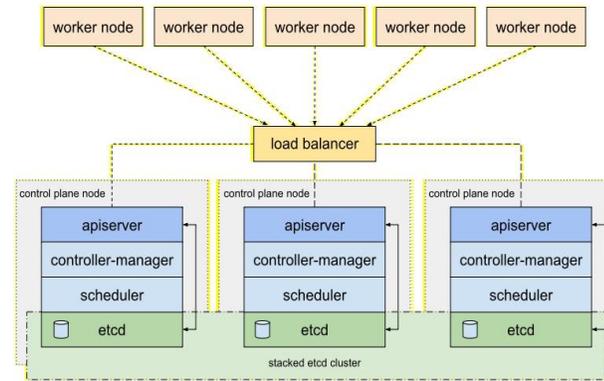


天文云资源服务

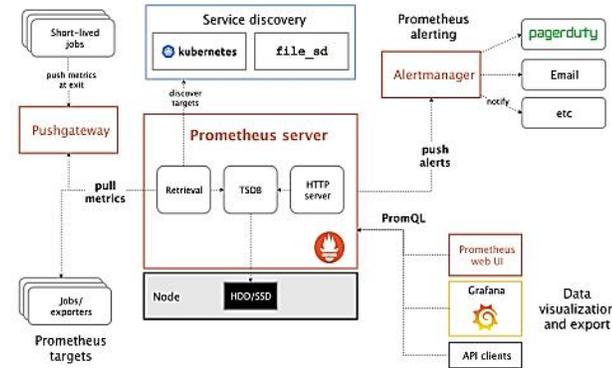
Build Data Processing Pipeline for FAST Key Project and SVOM



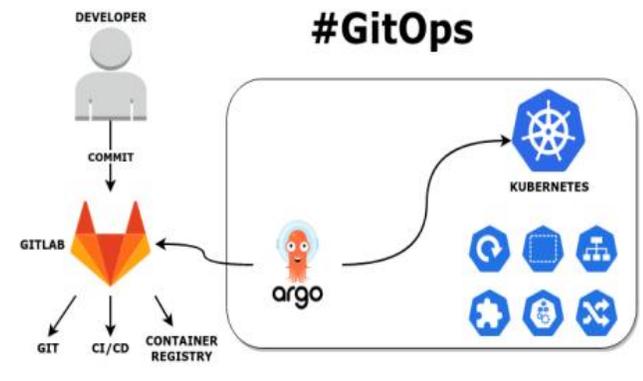
A Fast Radio Burst Discovered by FAST



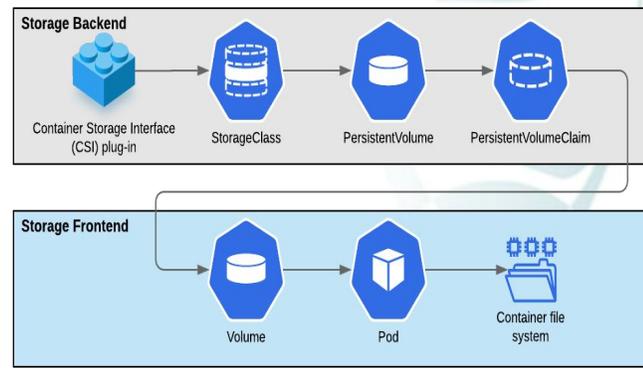
High Availability K8S Cluster



Resources Monitoring



Continuous Integration and Delivery



Distributed Storage System

报告提纲

- 天文领域的大科学装置
- 中国虚拟天文台平台
- **天文科研云平台构想与展望**



天文大数据面临的挑战

- 海量数据的精确检索与有限访问
- 不同种类数据的交叉融合
- 跨平台环境下的数据读写与共享存储
- 天文数据处理与分析软件的部署
- 不同种类计算资源的协同
- 基础设施的可持续建设与运行

天基观测设施



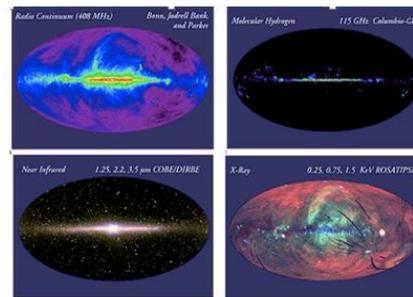
地面观测设施



协同
观测
数据

数据融合

不同观测设备
不同观测波段
不同观测模式
不同观测时间



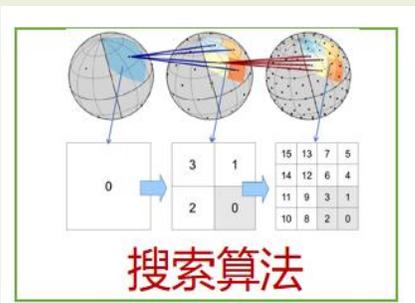
天文科研云平台



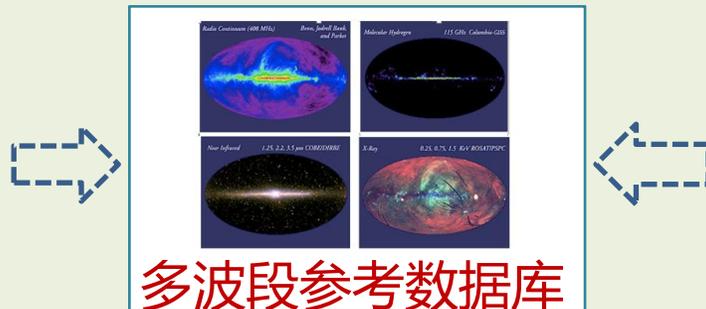
交互工具



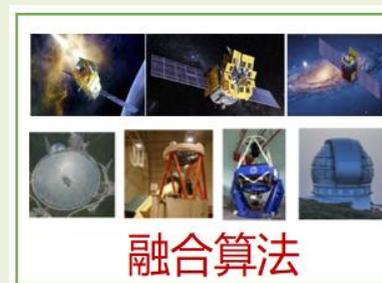
分析工具



搜索算法



多波段参考数据库



融合算法

便于天文学家在线分析使用多波段的天文大数据，更专注于科学发现



算力

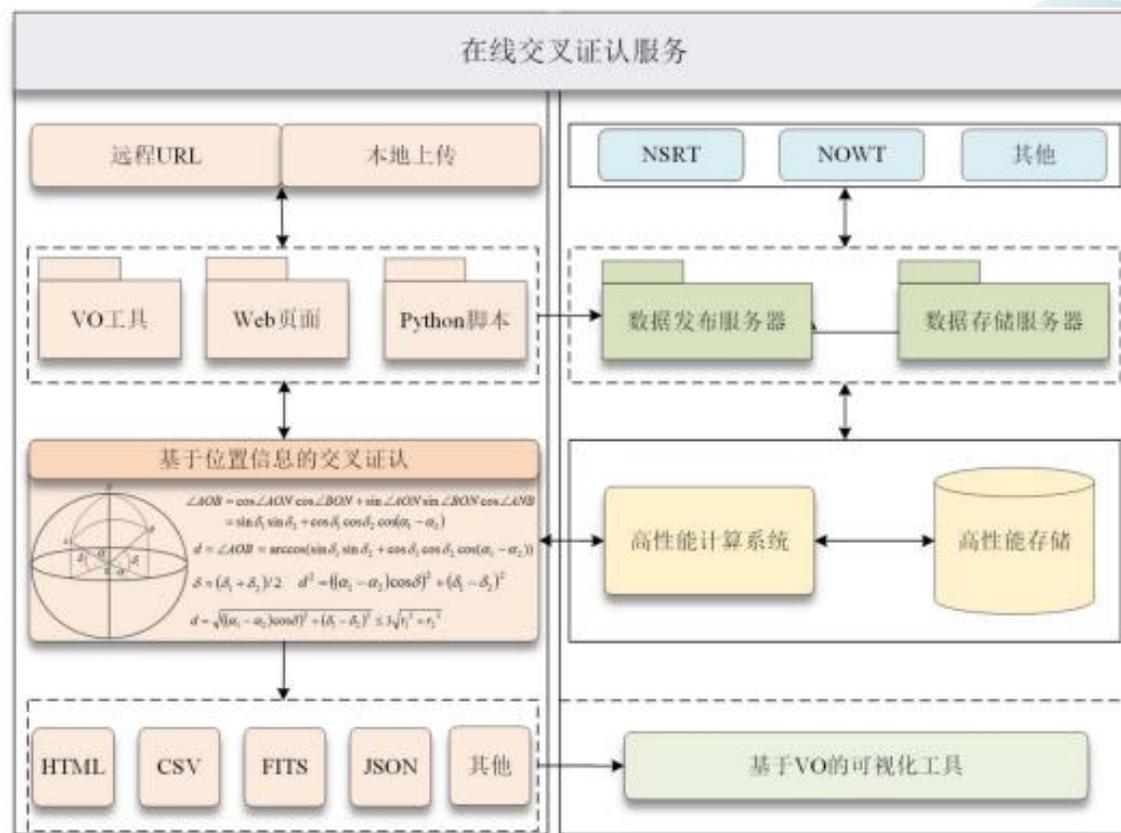
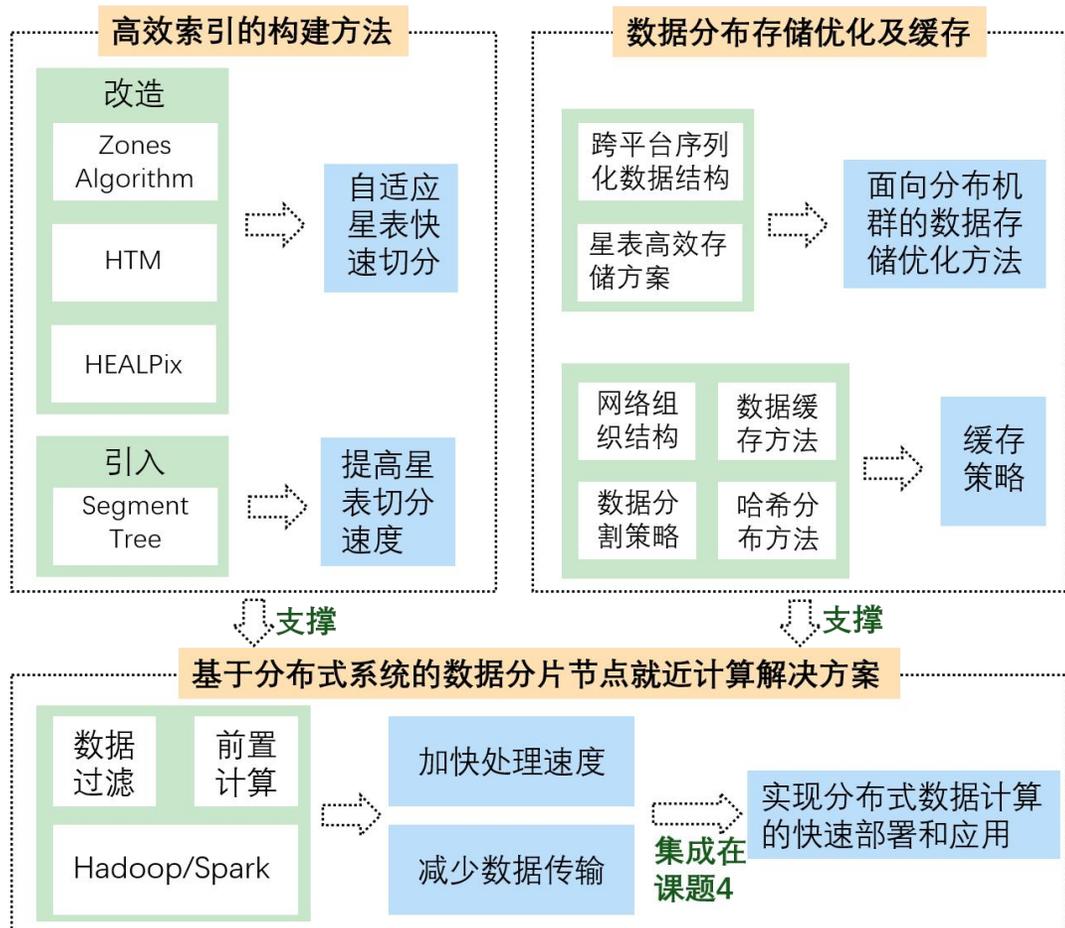


存储

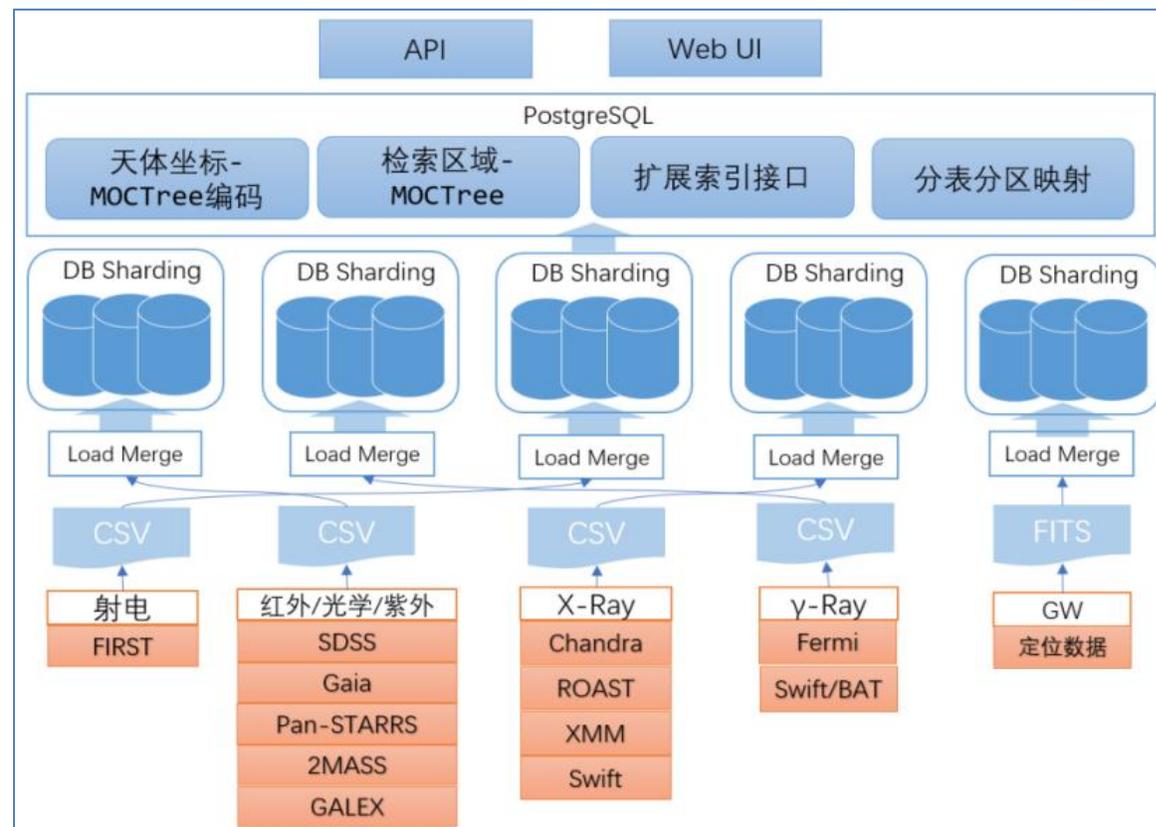
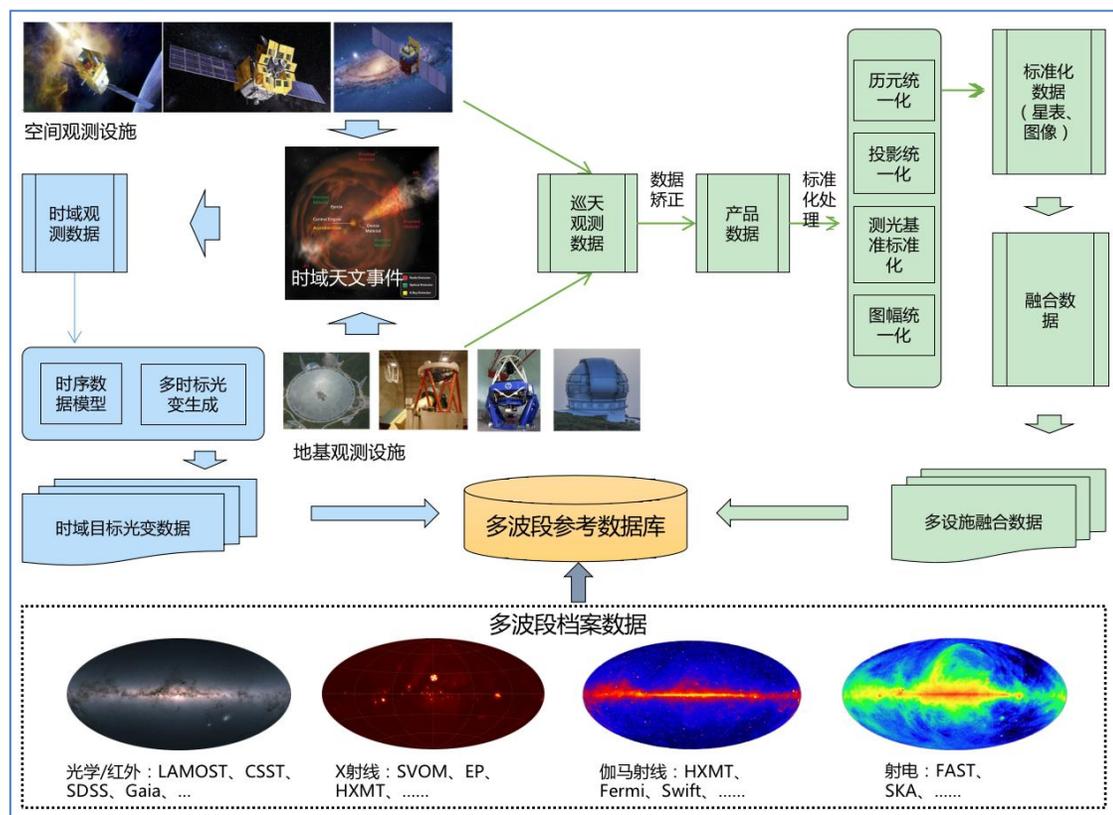
天文大数据科研云平台

国家天文科学数据中心

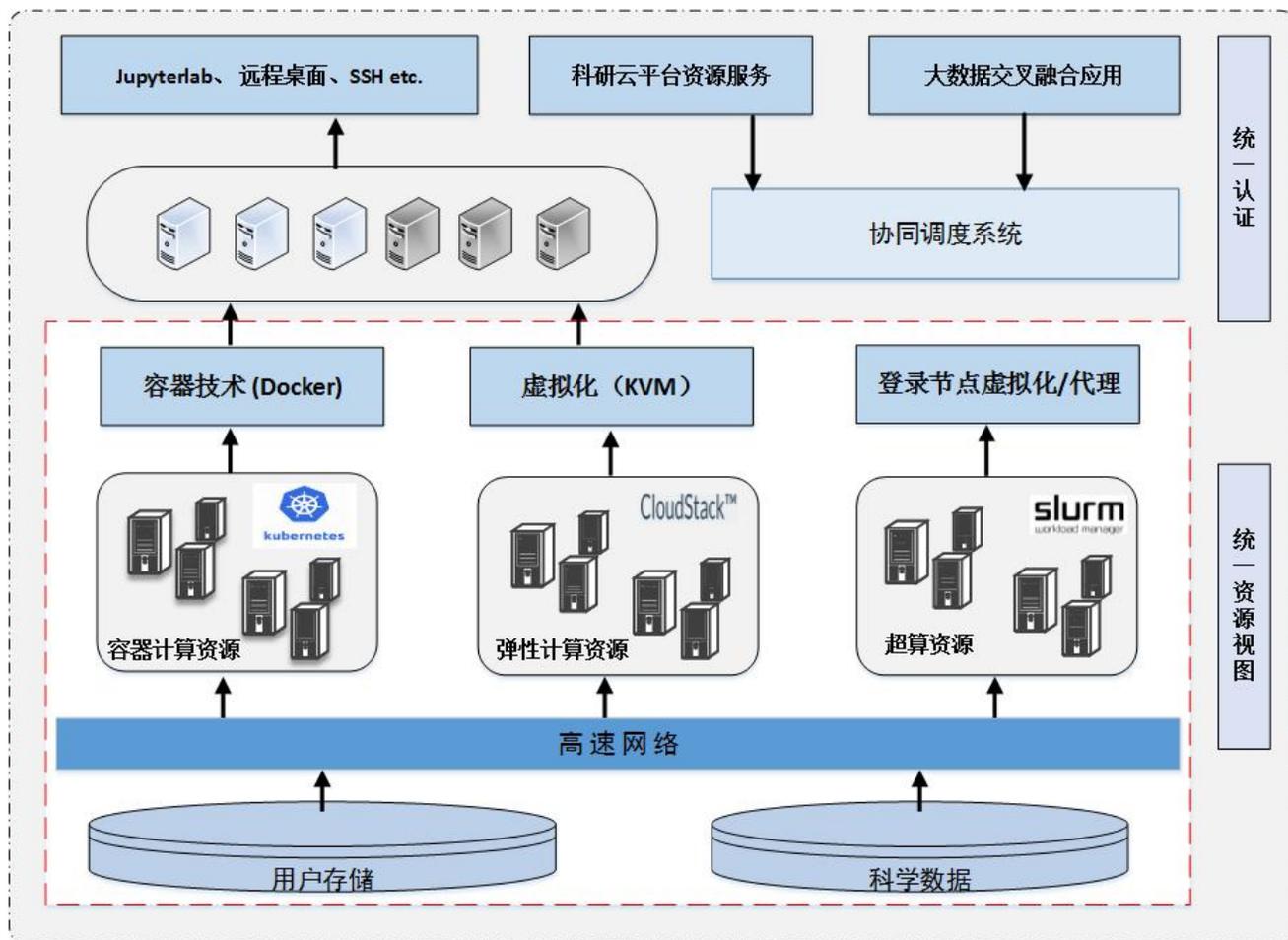
高性能检索与交叉证认



多源异构天文数据高效融合

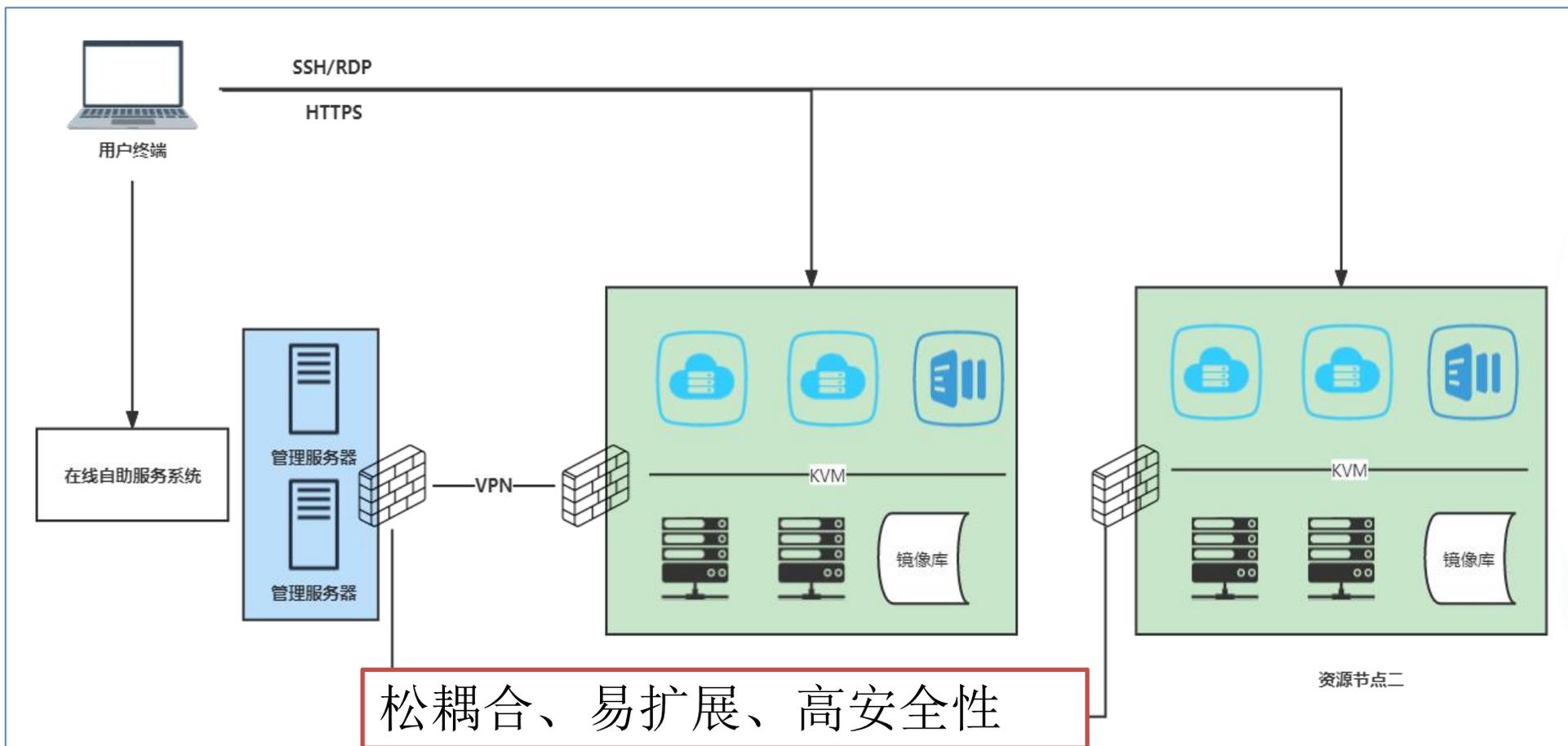


天文科研云平台资源整合框架

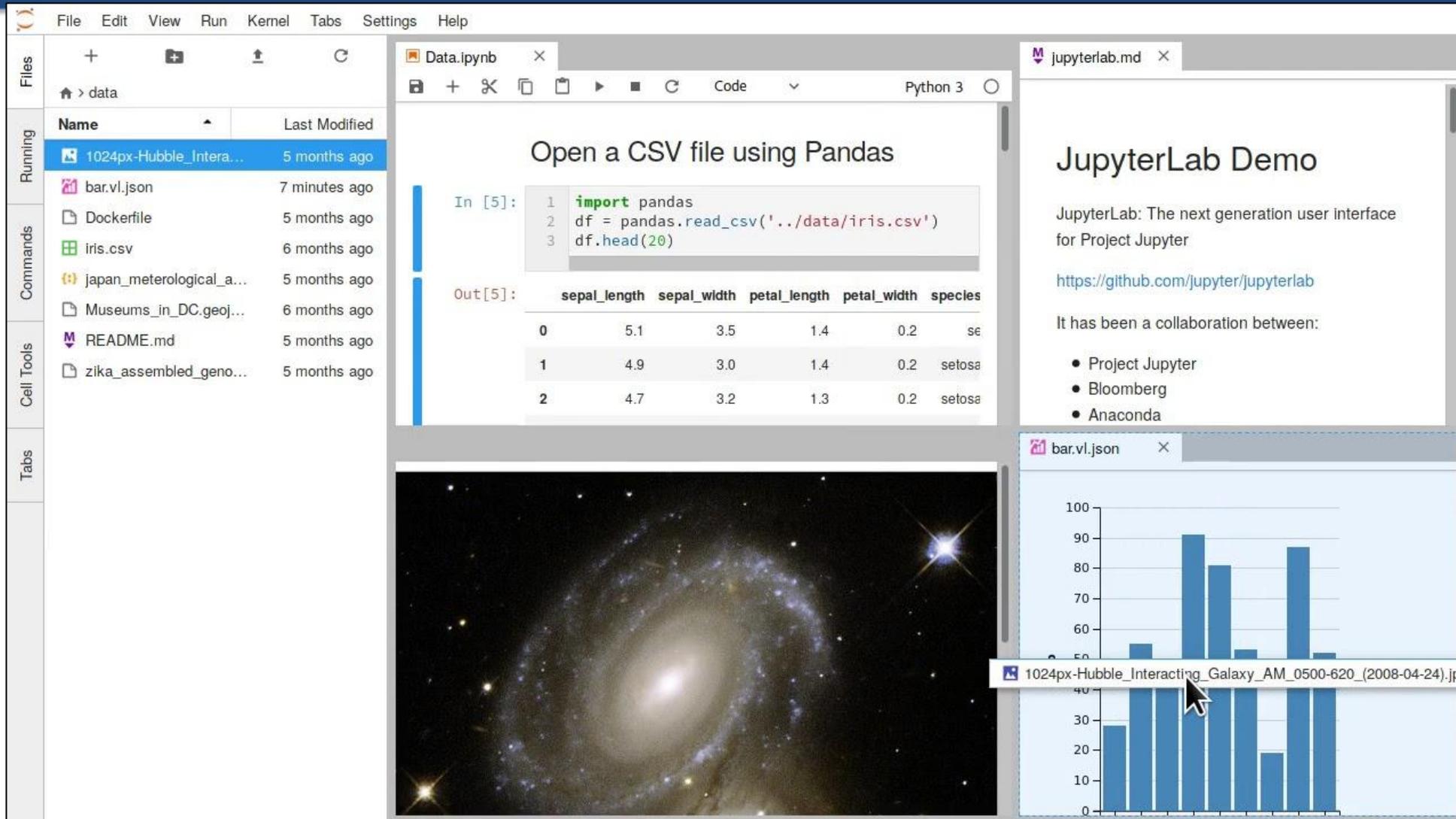


- 统一资源管理系统管理跨地域的异构资源，提供统一资源视图及资源申请自助服务
- 基于WEB的在线科研环境提供以Jupyter为交互手段的科研环境接入
- 计算与数据的协同调度根据用户任务的数据、计算需求，以就近计算为基本原则分配计算资源

统一资源管理与互操作



在线数据分析



The screenshot displays the JupyterLab web interface. On the left is a file browser showing a directory named 'data' containing files like '1024px-Hubble_Interacting_Galaxy_AM_0500-620_(2008-04-24).jpg', 'bar.vl.json', 'iris.csv', and others. The main area is split into a code editor and an output area. The code editor shows a Python snippet using pandas to read a CSV file. The output area displays the resulting DataFrame with columns for sepal_length, sepal_width, petal_length, petal_width, and species. Below the code editor, there are two preview windows: one showing a large image of an interacting galaxy and another showing a bar chart of the data.

Data.ipynb Python 3

Open a CSV file using Pandas

```
In [5]: 1 import pandas
         2 df = pandas.read_csv('../data/iris.csv')
         3 df.head(20)
```

Out[5]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	se
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa

JupyterLab Demo

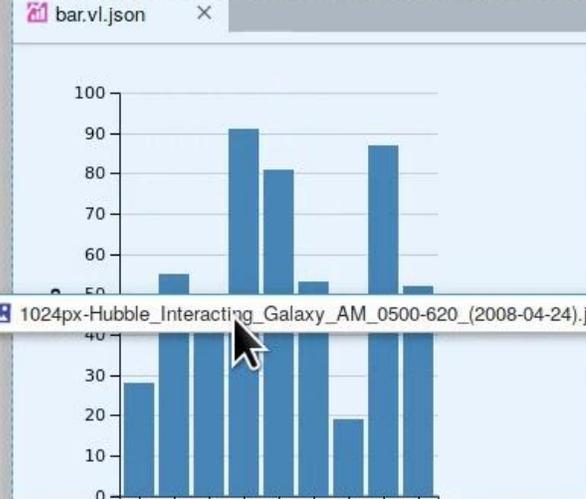
JupyterLab: The next generation user interface for Project Jupyter

<https://github.com/jupyter/jupyterlab>

It has been a collaboration between:

- Project Jupyter
- Bloomberg
- Anaconda

bar.vl.json



Category	Value
0	28
1	55
2	92
3	80
4	52
5	88
6	18
7	50

China-VO

以数据为主线的科学平台



谢谢!

