

# 多波段、多信使天文数据高效 融合关键技术研究与应用

答辩人：许允飞

导师：崔辰州 研究员

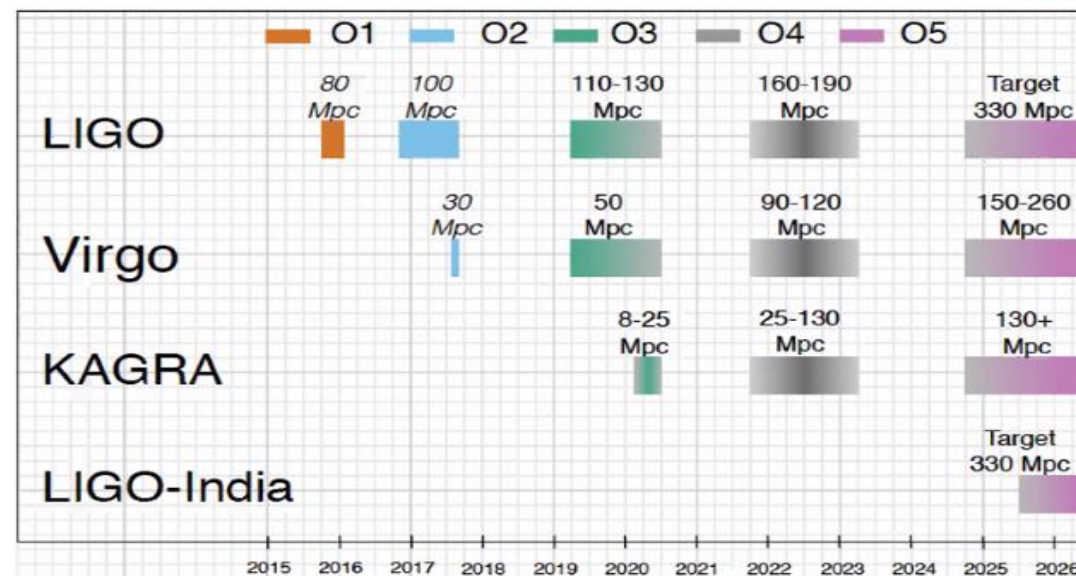
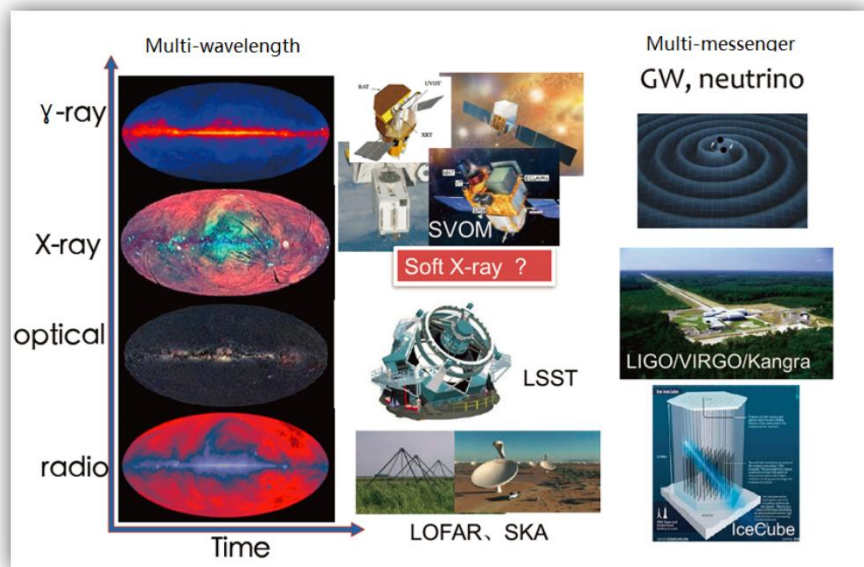
樊东卫 副研究员

国家天文科学数据中心  
中国科学院国家天文台  
中国科学院大学

# 提纲

1. 选题意义
2. 研究现状
3. 论文工作
4. 研究总结
5. 未来展望

# 1. 选题意义



项目名称	波段	日新增数据量	年新增数据量
FAST	射电	~15TB	~5PB
EP	X-Ray	~17GB	~5TB
LAMOST	光学	~22GB	~8TB
空间站望远镜	光学	~1TB	~360TB
GWAC/SVOM	光学/X-Ray、γ-Ray	~4.9TB	~1.74PB
天籁计划	射电	~4TB	~1.5PB

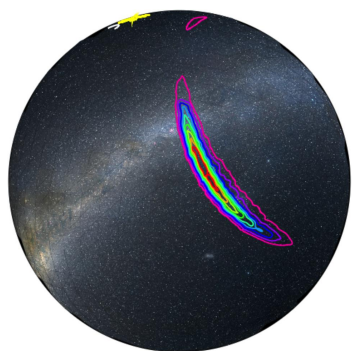
# 1.选题意义



爱因斯坦探针卫星 (EP) X射线暂现源证认

时域天文学和高能天体物理领域的小型天文卫星，系统性发现和探索宇宙暂现和爆发天体、监测天体光变。预计于2022年发射。

1. EP暂现源的多波段交叉证认
2. 多信使事件的EP观测证认
3. EP暂现源多信使观测证认



引力波电磁对应体后随观测

## 2018 – 2019 run (O3):

aLIGO 120 – 170 Mpc,

AdV 65 – 85 Mpc,

90%置信度定位天区: 120 – 180平方度

## 2020+ runs:

aLIGO 190 Mpc,

AdV 65 – 125 Mpc

90%置信度定位天区: 110 – 180平方度

1. 筛选引力波定位天区中宿主星系候选体
2. 宿主星系候选体的观测优先级排序
3. 从观测数据中高效证认引力波电磁对应体



# 1.选题意义



面向时域天文观测和多信使天文观测的实际应用

多信使事件电磁对应体高效搜寻

爱因斯坦探针卫星暂现源证认

核心技术需求

暂现源的实时多波段交叉证认

多信使事件的观测证认

暂现源的多信使观测证认

多波段、多信使天文数据高效融合

星表数据：基本物理属性、光变、光谱、能谱

图像数据：波段、格式、投影、坐标体系异构

海量星表高效检索方法

星表交叉证认及置信度估计

异构多波段图像的高效组织、检索及可视化

应用驱动

需求凝练

核心目标

关键技术

国家自然科学基金天文联合基金重点项目支持

# 2.研究现状



## 多波段数据

- 星表数据量: 数十亿~数亿      数千万~数百万      数万~数千
- 星表存储方式: 关系型数据库、NoSQL数据库、FITS文件、HDF5文件
- 图像组织方式: FITS、HDF5、CASA等, 文件组织、关系型数据库等
- 数据获取方式: IVOA SIA、SCS、SSA、TAP、HiPS等标准

多信使数据: 引力波信号、引力波源估计距离、引力波定位天区

## 2.研究现状



星表检索方法

星表交叉证认

图像组织及可视化

现有数据融合服务

# 2.研究现状

星表检索方法

星表交叉证认

图像组织及可视化

现有数据融合服务



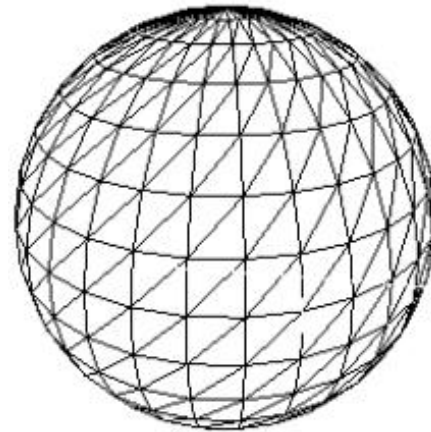
HEALPix  
天球划分

开源实现  
应用广泛



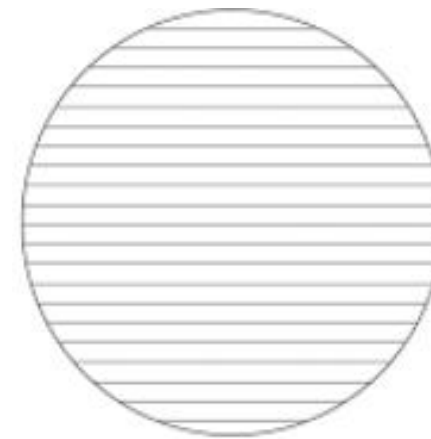
Q3C  
天球划分

开源实现  
应用广泛



HTM  
天球划分

基于商业数  
据库



Zones  
条带索引

基于商业数  
据库

# 2.研究现状

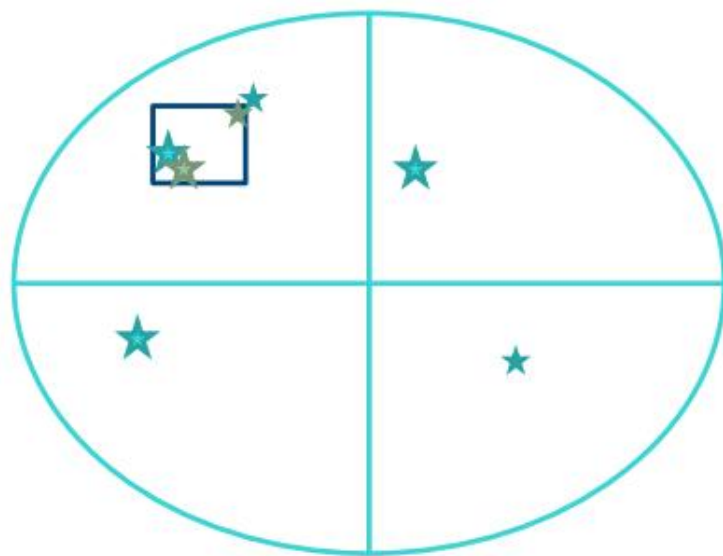
星表检索方法

星表交叉证认

图像组织及可视化

现有数据融合服务

基于位置的交叉证认



基于概率的置信度估计

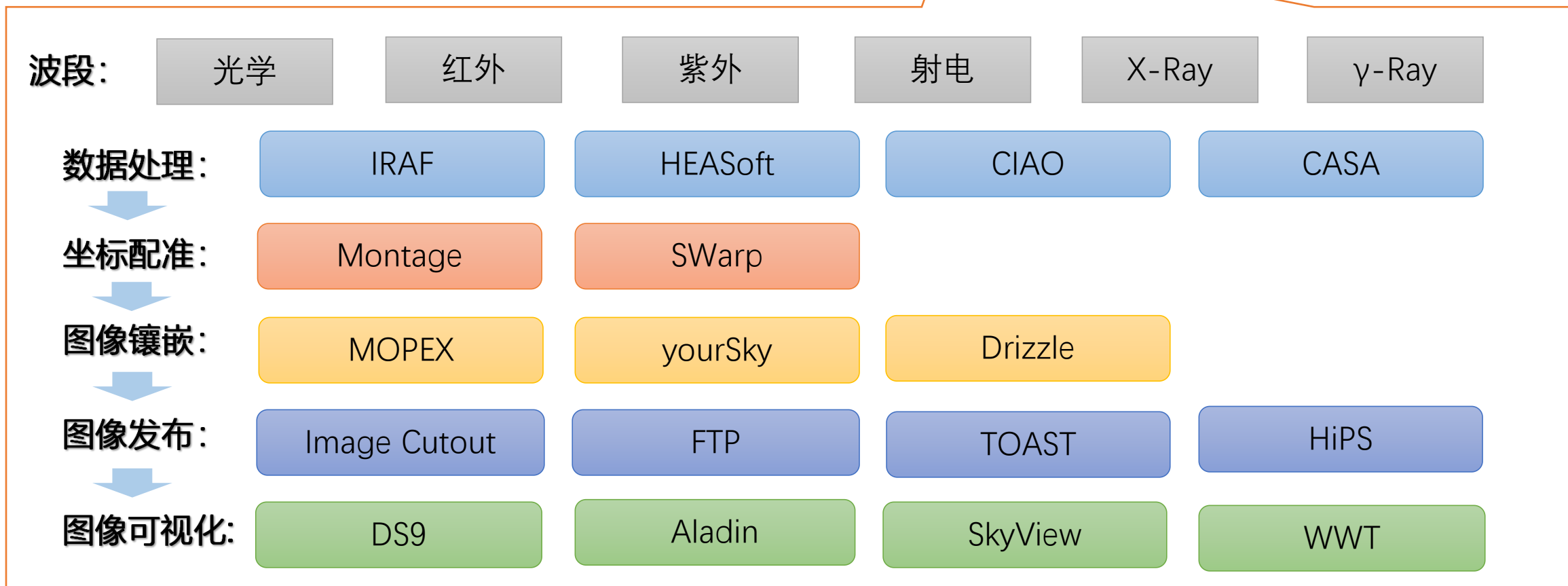
多波段星表位置误差：  
• 亚角秒级~数十角秒

基于概率的方法给出天体匹配置信度：  
• Likelihood Ratio方法  
• 贝叶斯方法

时效性



# 2.研究现状



# 2.研究现状



星表检索方法

星表交叉证认

图像组织及可视化

现有数据融合服务

多波段星表交叉证认:

VizieR

ARCHES

Starlink  
Tables

CDS-Xmatch

SIMBAD

NWay

主要面向科研目的小规模数据集的获取和交叉证认，尚无法满足多波段多信使观测的大尺度和高时效需求

多波段图像全天组织与可视化:

TOAST

HiPS

WWT

Aladin

ESASky

HiPS是当前应用最为广泛的图像数据组织标准，但数据标准化处理效率低，且部分平台未支持

## 3.论文工作

- ① 海量星表高效检索方法
- ② 星表交叉证认及置信度计算
- ③ 异构多波段图像的高效组织、检索及可视化
- ④ EP 暂现源判别及多波段证认
- ⑤ 引力波电磁对应体高效搜寻

# 3.1 海量星表高效检索方法



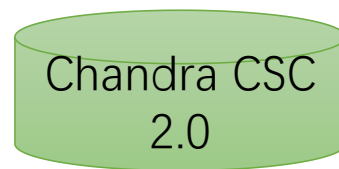
SDSS DR14



2MASS PSC



FIRST 14



Chandra CSC  
2.0



XMM DR6

记录数: 1,231,051,050

470,992,971

946,432

317,167

468,440

存储方式: 关系型数据库

关系型数据库

FITS文件

CSV文件

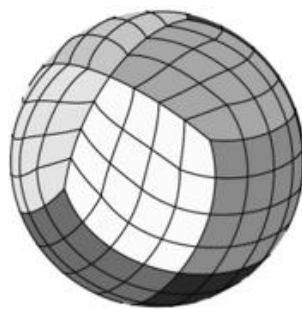
CSV文件

检索方式: 基于RA、DEC空间检索, 字段检索 (B+树)

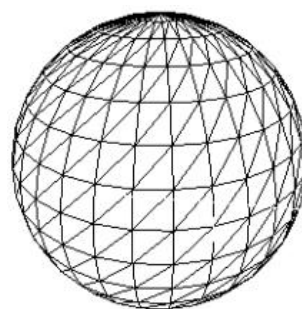
文件搜索



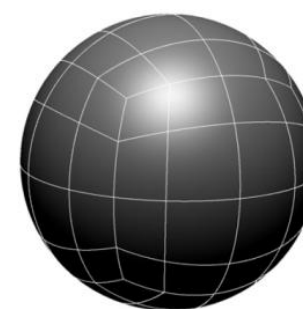
Zones  
条带索引



HEALPix  
天球划分



HTM  
天球划分



Q3C  
天球划分

# 3.1 海量星表高效检索方法



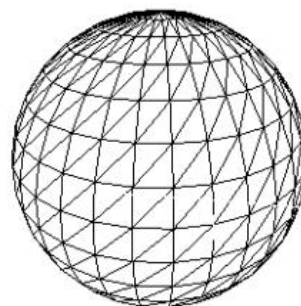
Zones  
条带索引



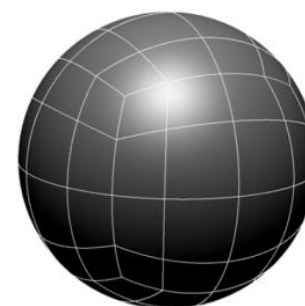
HEALPix  
天球划分

开源实现、应用广泛

SIMBAD、  
VizieR、  
H3C、HiPS、  
mcatCS



HTM  
天球划分



Q3C  
天球划分

Hierarchical Equal Area isoLatitude Pixelization



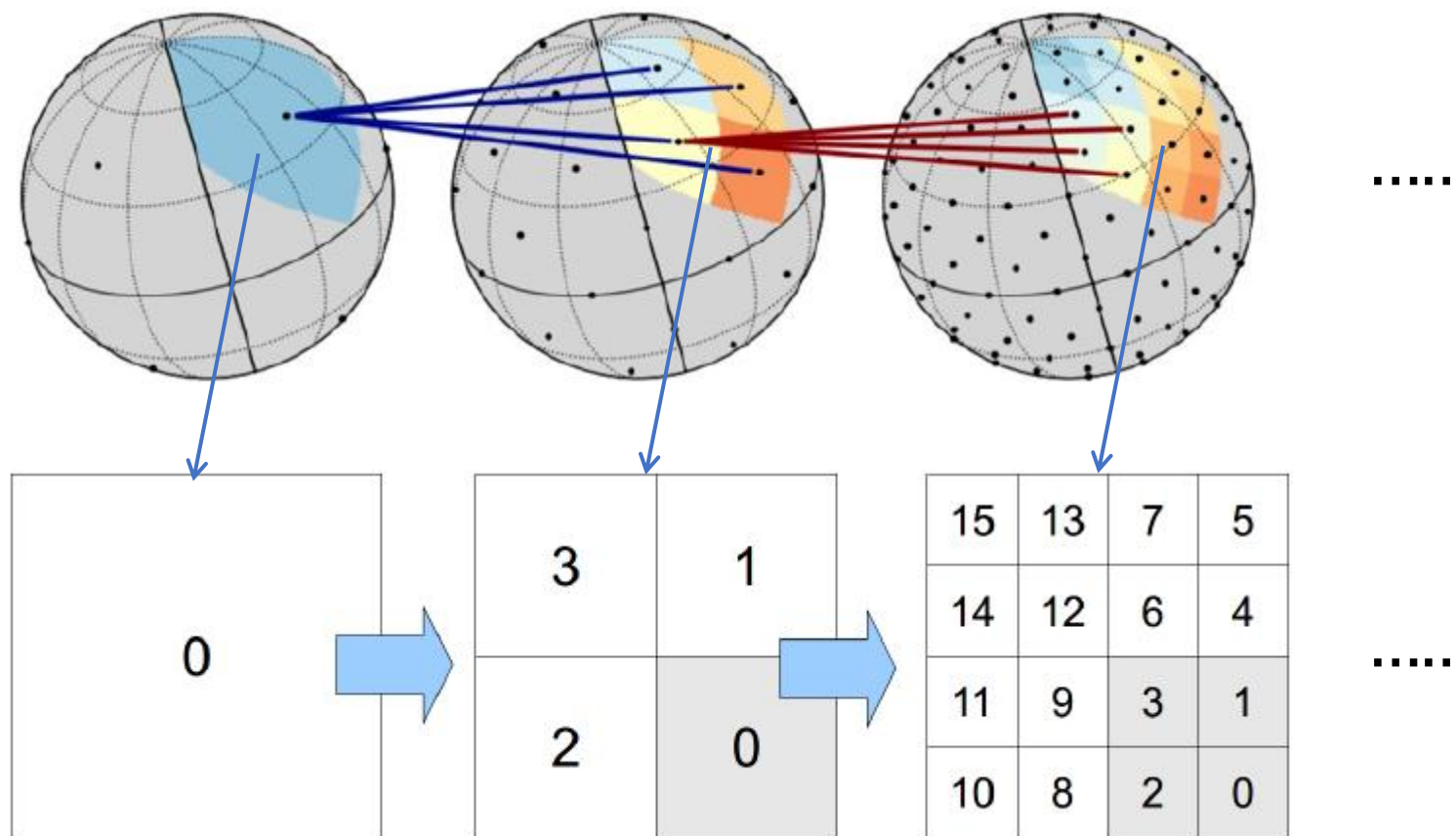
# 3.1 海量星表高效检索方法



**HEALPix  
天球划分**

开源实现、应用广泛

SIMBAD、  
VizieR、  
H3C、HiPS、  
mcatCS



# 3.1 海量星表高效检索方法



针对星表构建基于HEALPix的空间索引:

星表A	
Id 1	Ra Dec
Id 2	Ra Dec
Id 3	Ra Dec
Id 4	Ra Dec
Id 5	Ra Dec
Id 6	Ra Dec
.....	

$$\theta = \frac{(90 - Dec) \times \pi}{180}$$

$$z = \cos\theta$$

$$\phi = \frac{RA \times \pi}{180}$$

$$i = I \left( \left( 2 - \frac{3z}{2} \right) 2^k \right)$$

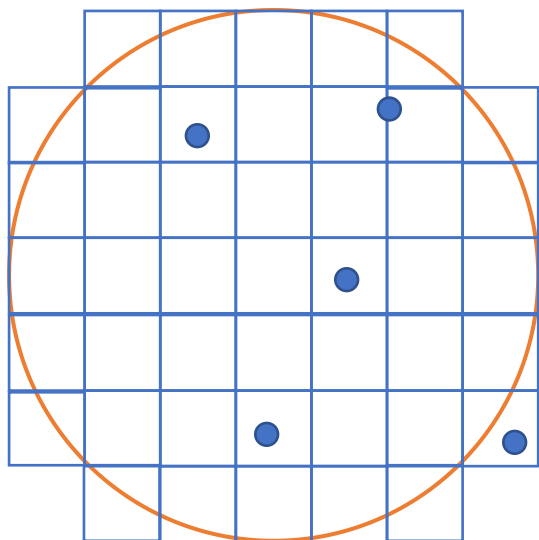
$$j = I \left( \frac{2^{k+1}\phi}{\pi} + \frac{(i - 2^k + 1) \bmod 2}{2} \right)$$

$$hpx = 2^k (i - 2^k - 1) + j + 2^{k+1} (2^k + 1)$$

星表A		
Id 1	Ra Dec	hpx
Id 2	Ra Dec	hpx
Id 3	Ra Dec	hpx
Id 4	Ra Dec	hpx
Id 5	Ra Dec	hpx
Id 6	Ra Dec	hpx
.....		

hpx与划分层级k相关，以hpx作为星表空间索引首先要确定划分层级k

# 3.1 海量星表高效检索方法



## 基于HEALPix索引的空间检索方法:

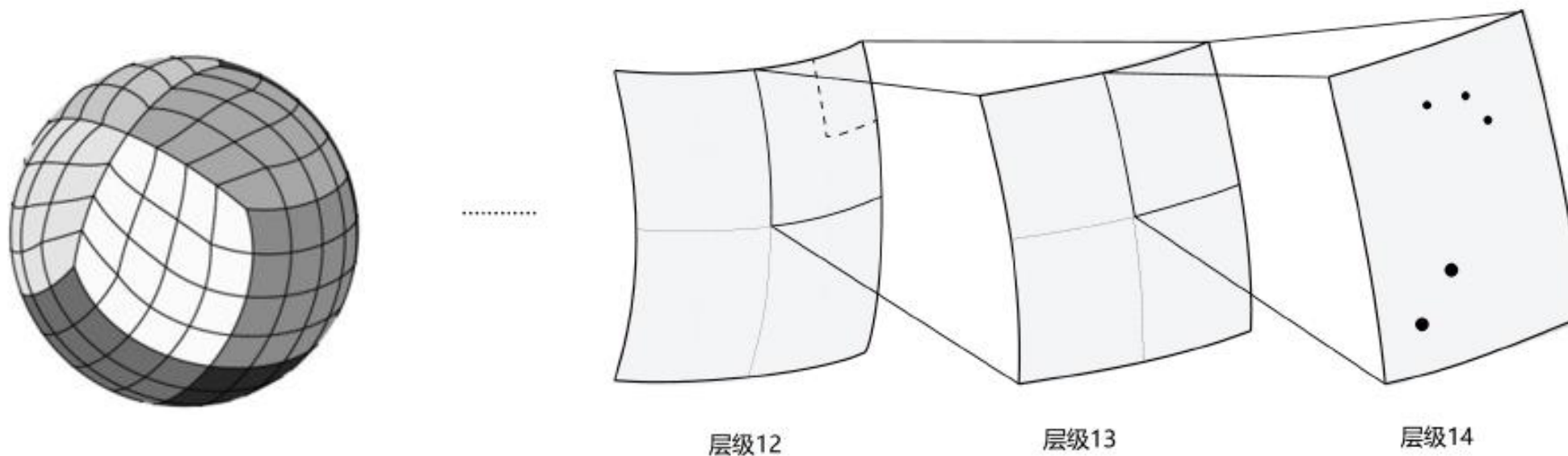
1. 以HEALPix网格拟合检索区域, 如圆盘、多边形
2. 将拟合区域的hpx id与星表中hpx id 进行对比, 返回一致的条目
3. 进一步计算各条目是否实际落在检索区域中, 如左图右下方的天体

## 主要缺陷:

- 需要多次扫描索引, 扫描次数与拟合检索区域的网格数相当
- 需要额外计算条目与检索区域的覆盖关系
- 对于大范围天区检索 (如引力波定位天区), 将大幅降低检索效率

# 3.1 海量星表高效检索方法

基于多层次覆盖天区的空间索引 MOC-Tree (Multi-order Coverage Tree) :



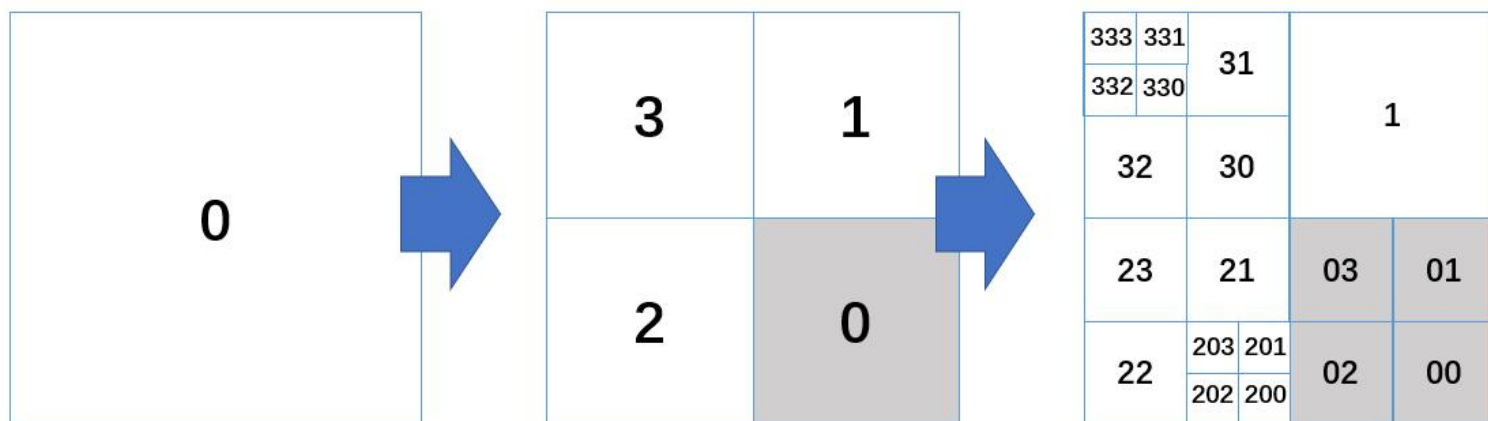
- 整个天球是由12颗MOC-Tree构成的森林，对应HEALPix的12个基础天区
- 每棵MOC-Tree记录了对应基础天区内，所有天体从层级0-层级14的相对位置
- 层级14共包含 $3.22 \times 10^9$ 个网格，分辨率为12.9"，适用于十亿级体量的星表

# 3.1 海量星表高效检索方法



MOC-Tree的索引编码方式:

编码位数与层级数一致, 每一位编码与该网格在其父网格的相对位置相关

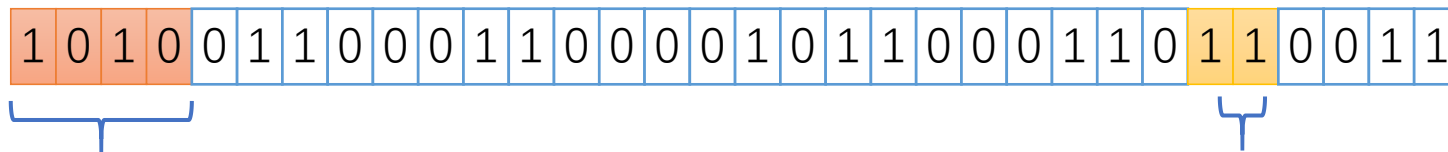


$$hpx_0 = hpx - (2^{2k} \times F)$$

$$N_k = I \left( \frac{hpx_0}{2^k} \right)$$

MOC-Tree索引位数对应层级, 其中每一位对应层级的相对位置

索引形如:  $KN_1N_2 \dots N_k$  共计32位



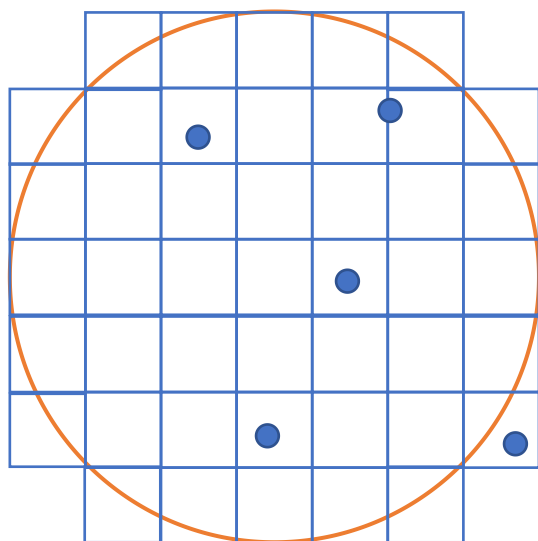
表示所在的基础天区0-11

表示层级12中的相对位置3

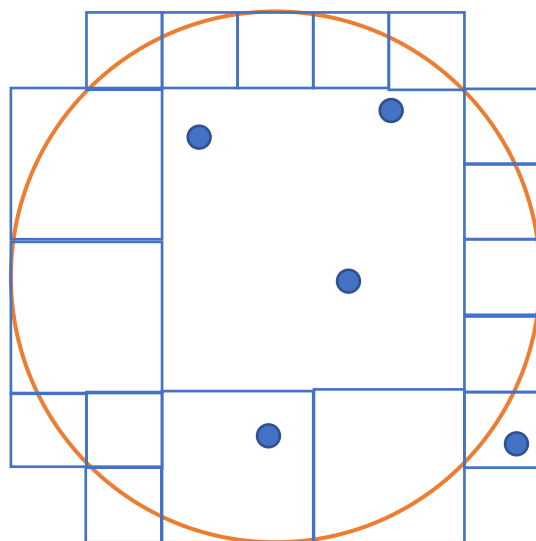


# 3.1 海量星表高效检索方法

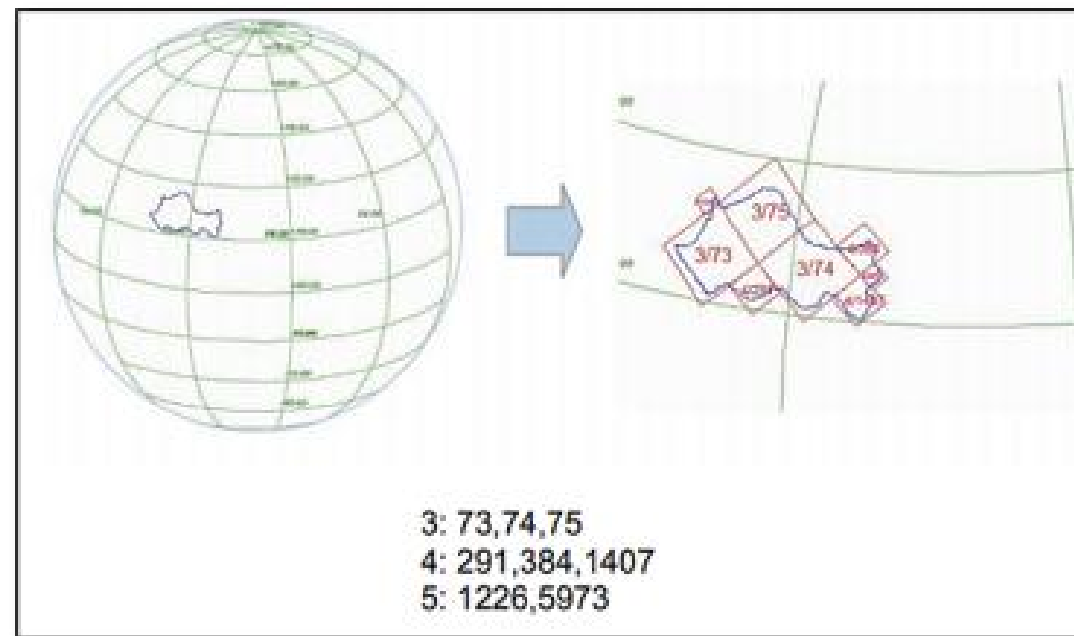
基于多层次覆盖天区的空间检索：



HEALPix索引下的检索区域



MOC-Tree索引下的检索区域



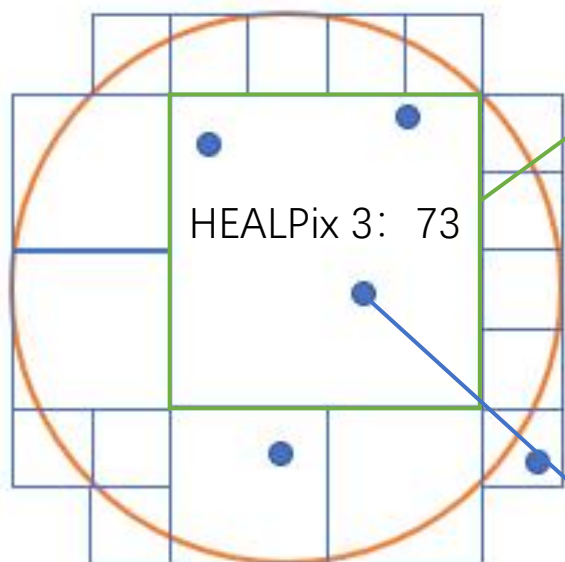
# 3.1 海量星表高效检索方法



## 基于多层次覆盖天区的空间检索:

通过位运算实现索引比较:

不会出现1100、1101、1110、1111，且填充位数为偶数，可得到填充位数



MOC-Tree: 1:101

1 0 0 0 1 0 1 0 0 0 1

填充位

基础天区1

网格号101

0 0

与检索网格编号执行异或运算，若落在该网格中，则结果全为0

1 0 0 0 1 0 1 0 0 0 1

右移，并填充1

0 0 0 1 0 1 0 0 0 1 0 1 0 1 0 0 1 0 1 1 0 0 0 1 1 0 1 1 0 0 1 1

XOR

# 3.1 海量星表高效检索方法

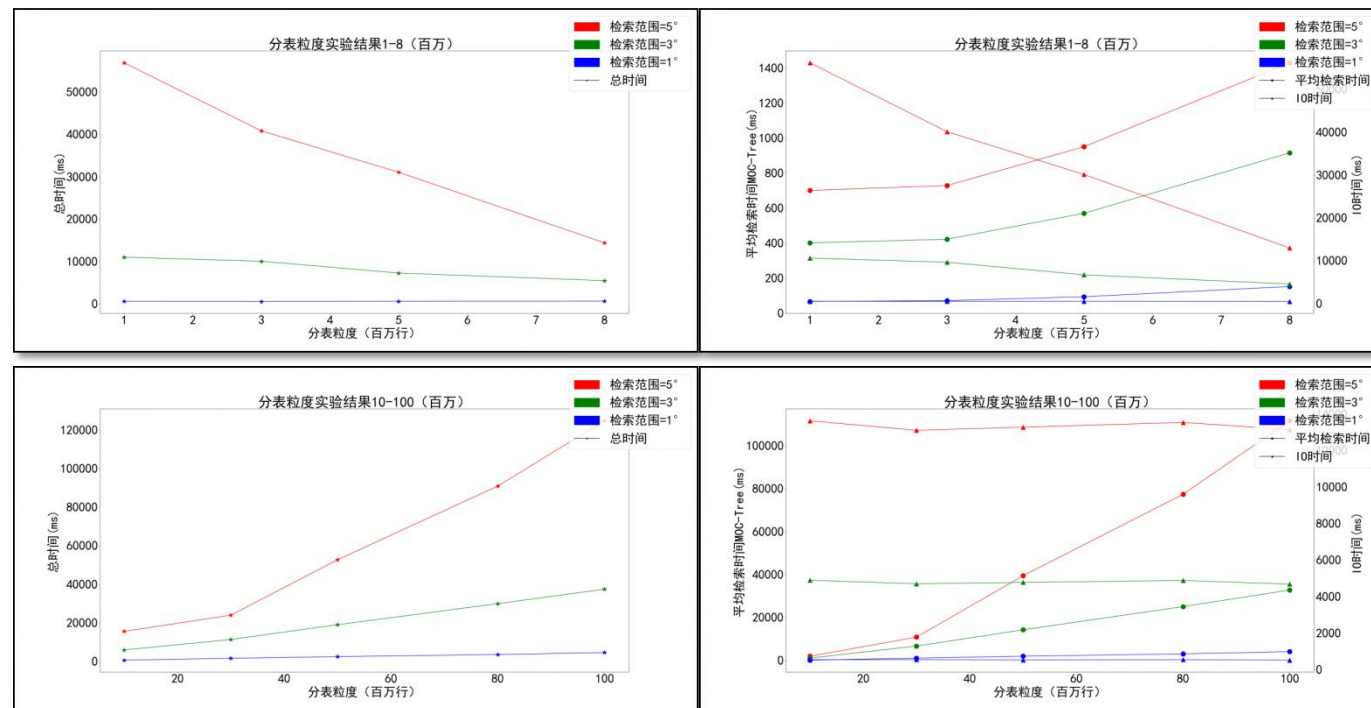


## 海量星表的分治策略：

- 十亿级体量的星表，单个关系型数据表的查询检索开销巨大
- 单台服务器的CPU、内存、网络也会成为瓶颈
- 采用**分表分区策略**可有效提高海量星表的检索效率
- 将海量星表横向划分，提高数据库并发检索能力

## 基于SDSS DR14测光星表进行分表粒度实验

项目	说明
测试星表	SDSS DR14 测光星表
服务器环境	4核3.0GHz主频、16GB RAM, SSD存储
检索方式	1°、3°、5°锥形检索
分表粒度	100万至800万，100万1级 1000万至1亿，1000万一级

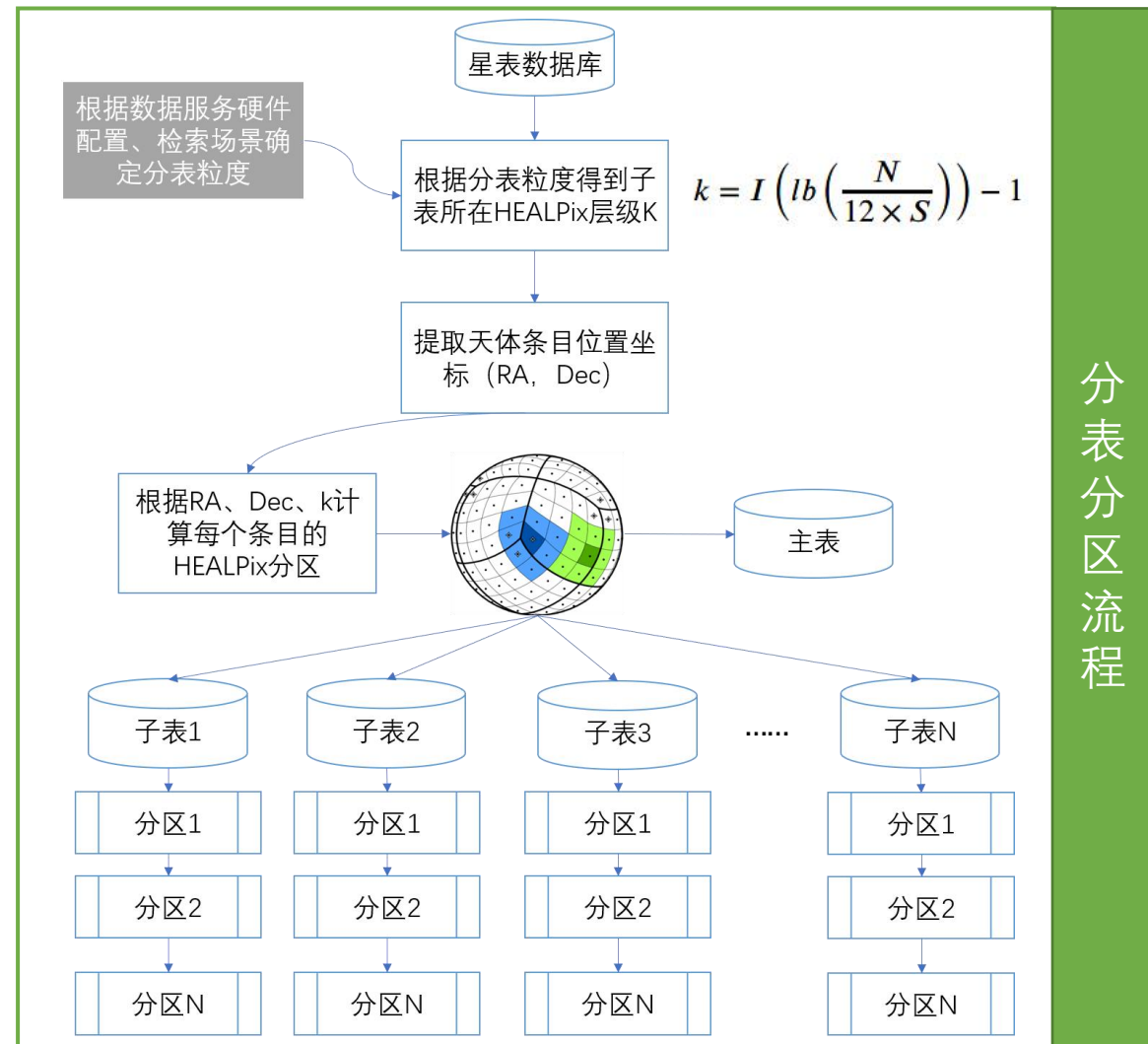
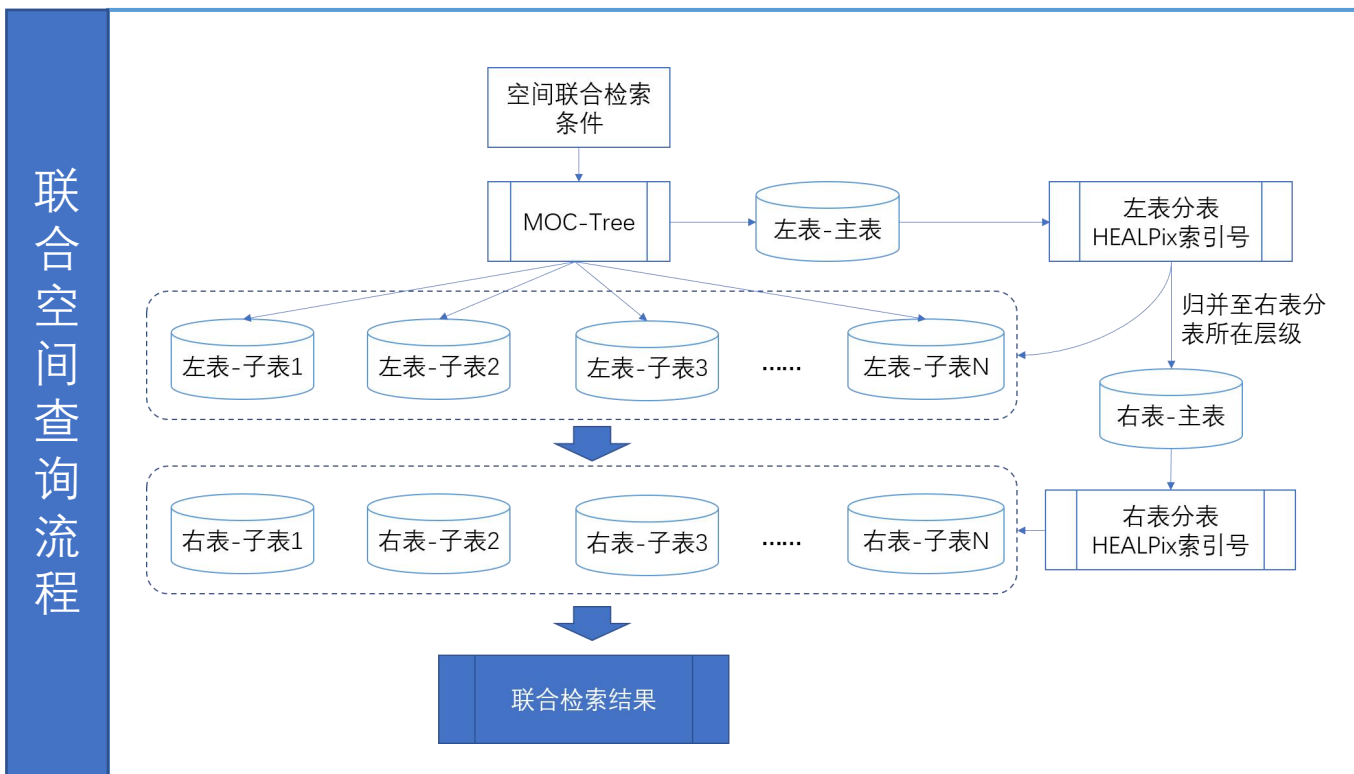


在检索半径1°至5°的锥形检索条件下，对于十亿级的星表分表，子表条目数在1000万为宜。

# 3.1 海量星表高效检索方法



- 如果两个星表的分表粒度不一致，无法实现联合空间检索
- 基于HEALPix 天球划分的分表方法及联合检索策略解决上述问题



# 3.1 海量星表高效检索方法



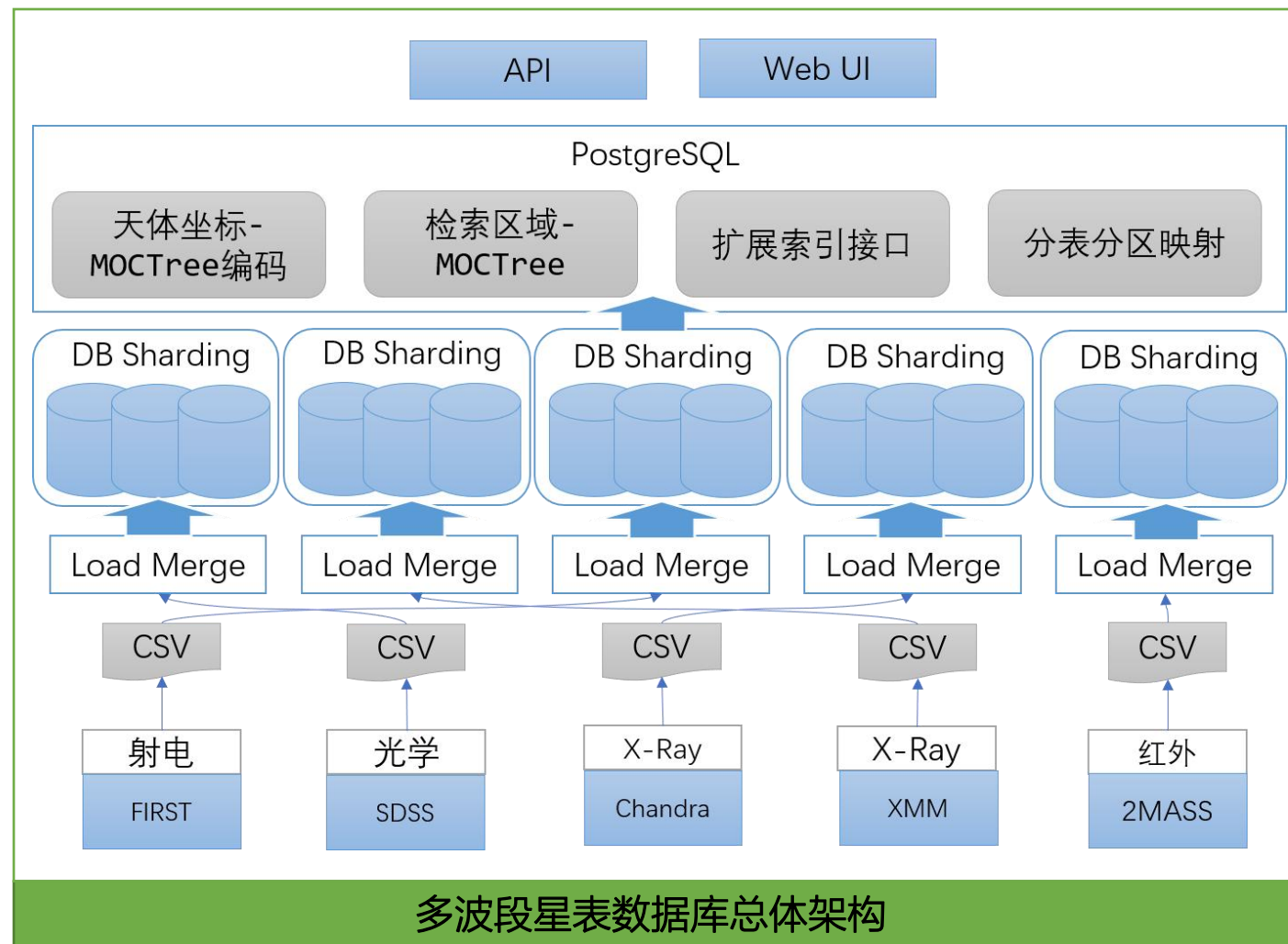
## 多波段星表数据库设计实现及测试

星表信息

星表名称	波段	条目数
SDSS DR14测光星表	光学	1,231,051,050
2MASS PSC	红外	470,992,971
FIRST14	射电	946,432
Chandra CSC 2.0	X射线	317,167
3XMM DR6 源表	X射线	468,440

数据库软件

类型	名称	版本/型号
数据库管理软件	PostgreSQL	10.1
操作系统	ubuntu	14.04
检索插件	Q3C	2.0.0
检索插件	H3C	1.1.9
软件运行环境	python	3.6



多波段星表数据库总体架构



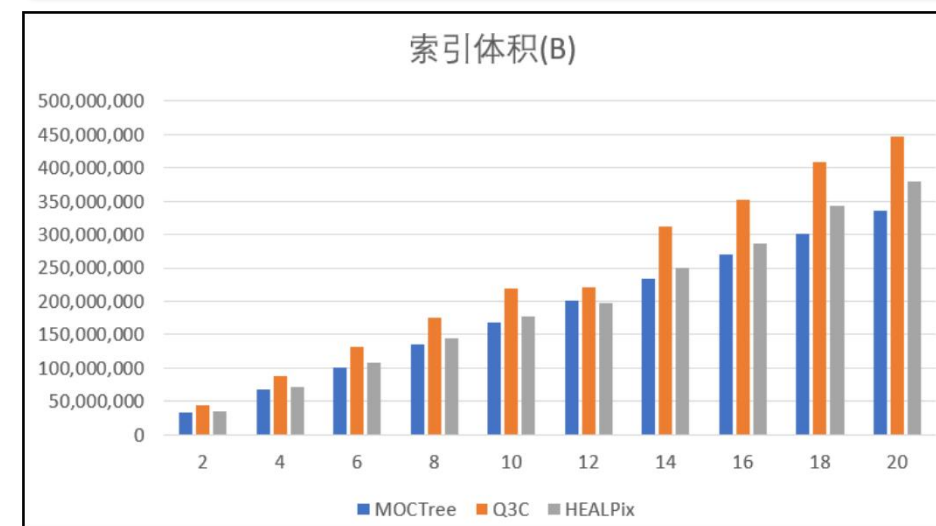
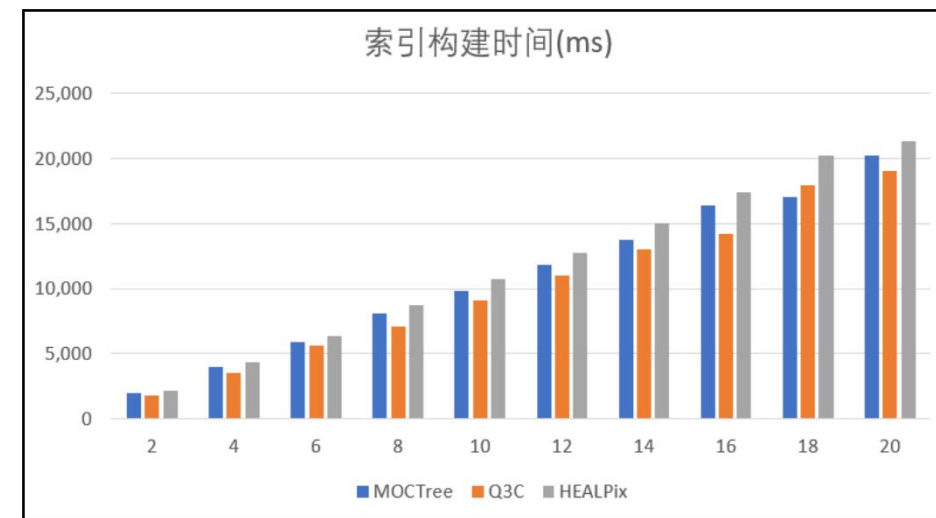
# 3.1 海量星表高效检索方法



## 多波段星表数据库设计实现及测试

对比了MOC-Tree、Q3C、原生HEALPix索引构建时间与索引体积

星表名称	星表条目	索引类型	索引构建时间(ms)	索引体积(B)
SDSS	1,231,051,050	MOCTree	1,983,738	28,847,389,230
		Q3C	1,845,984	35,562,253,312
		HEALPix	2,137,743	27,651,407,872
2MASS	470,992,971	MOCTree	739,280	10,311,903,920
		Q3C	697,812	12,738,298,393
		HEALPix	803,973	9,837,366,273
FIRST	946,432	MOCTree	992	17,377,837
		Q3C	891	21,282,816
		HEALPix	1,048	17,648,593
Chandra	317,167	MOCTree	514	5,538,179
		Q3C	372	7,151,616
		HEALPix	597	5,572,648
XMM	468,440	MOCTree	763	8,173,839
		Q3C	689	10,543,104
		HEALPix	801	8,273,937



# 3.1 海量星表高效检索方法



## 多波段星表数据库设计实现及测试

对比了不同星表在MOC-Tree、Q3C、原生HEALPix上的空间检索效率

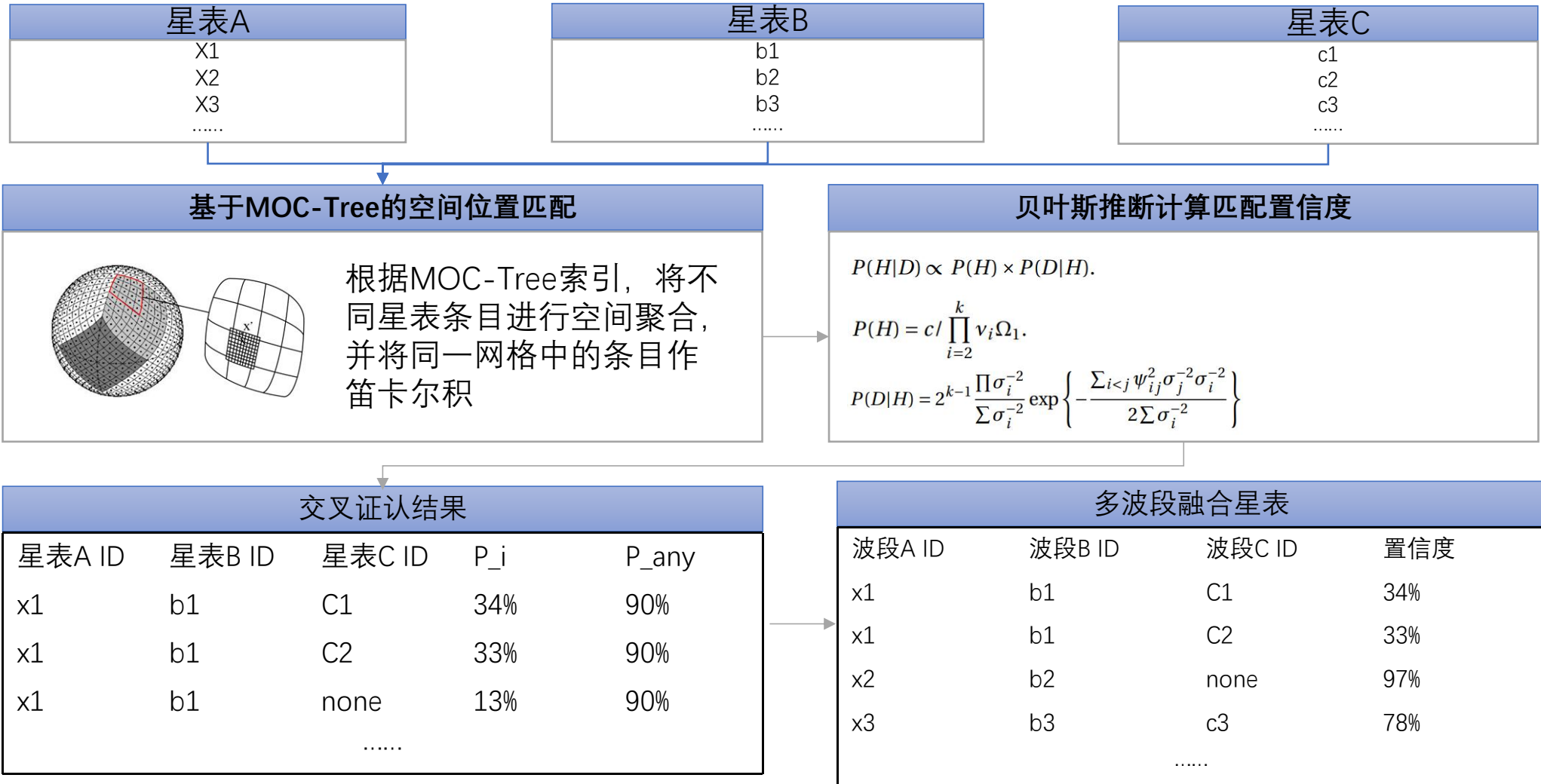
十亿级星表优势明显

星表名称	星表条目	索引类型	1°查询时间 (ms)	1°IO时间 (ms)	3°查询时间 (ms)	3°IO时间 (ms)	5°查询时间 (ms)	5°IO时间 (ms)
SDSS	1,231,051,050	MOCTree-分表分区 (129)	201.80	513.20	1321.20	5312.20	2012.40	13642.20
		Q3C	36780.30	511.60	330020.10	5039.60	911256.30	11649.10
		HEALPix	41592.80	512.90	351473.20	5066.20	1001211.60	11623.00
2MASS	470,992,971	MOCTree-分表分区 (48)	192.30	513.10	1254.30	5260.60	1969.30	12784.20
		Q3C	20158.60	511.20	185695.30	5114.20	506489.20	10231.90
		HEALPix	23329.30	513.00	213898.30	5113.20	583114.20	10232.20
FIRST	946,432	MOCTree	69.50	221.30	391.60	635.60	691.60	1763.80
		Q3C	77.10	220.10	593.10	634.70	1659.20	1762.30
		HEALPix	89.90	221.30	688.30	635.40	1871.20	1762.90
Chandra	317,167	MOCTree	65.30	124.30	139.20	514.30	309.20	783.20
		Q3C	77.20	123.60	187.60	515.60	744.10	782.90
		HEALPix	89.10	122.90	219.10	515.90	831.20	783.60
XMM	468,440	MOCTree	69.20	153.40	183.80	583.40	400.50	892.10
		Q3C	77.60	152.20	201.60	582.20	698.20	891.60
		HEALPix	83.40	152.80	239.80	582.60	765.20	892.50

MOC-Tree 的多层级覆盖天区检索方法降低了进行大范围空间查询时的查询次数。在3° 检索范围上，MOC-Tree 查询时间比HEALPix 低37.7%，比Q3C 低27.3%。在5° 检索范围上，MOC-Tree 查询时间比HEALPix低59.6%，比Q3C 低54.8%

大天区范围检索有较大优势

# 3.2星表交叉证认及置信度估计



## 3.2 星表交叉证认及置信度估计



### 星表间天体位置匹配

- 多星表间的位置匹配可转化为针对空间索引字段的Left Join联合检索问题。即不同星表中，相同空间索引的条目在空间位置上是匹配的；
- 不同星表的定位误差不同，其基于天球划分的空间索引所在层级也不同，因此在联合检索时要进行层级归并；
- MOC-Tree索引包含了不同层级的空间位置信息，免去了归并过程。

星表A (HEALPix层级5)

ld 1	Hpx 2005
ld 2	Hpx 2019
ld 3	Hpx 2037
ld 4	Hpx 2048
ld 5	Hpx 2069
ld 6	Hpx 2101
ld 7	Hpx 2199
ld 8	Hpx 2204
ld 9	Hpx 2227
ld 10	Hpx 2248
.....	



星表A (HEALPix层级3)

ld 1	Hpx 19
ld 2	Hpx 57
ld 3	Hpx 83
ld 4	Hpx 99
ld 5	Hpx 109
ld 6	Hpx 125
ld 7	Hpx 137
ld 8	Hpx 145
ld 9	Hpx 189
ld 10	Hpx 201
.....	

② Left Join



星表B (HEALPix层级3)

ld 1	Hpx 125
ld 2	Hpx 126
ld 3	Hpx 127
ld 4	Hpx 128
ld 5	Hpx 129
ld 6	Hpx 131
ld 7	Hpx 137
ld 8	Hpx 137
ld 9	Hpx 139
ld 10	Hpx 140
.....	

③ 输出结果



位置匹配结果  
(HEALPix层级3)

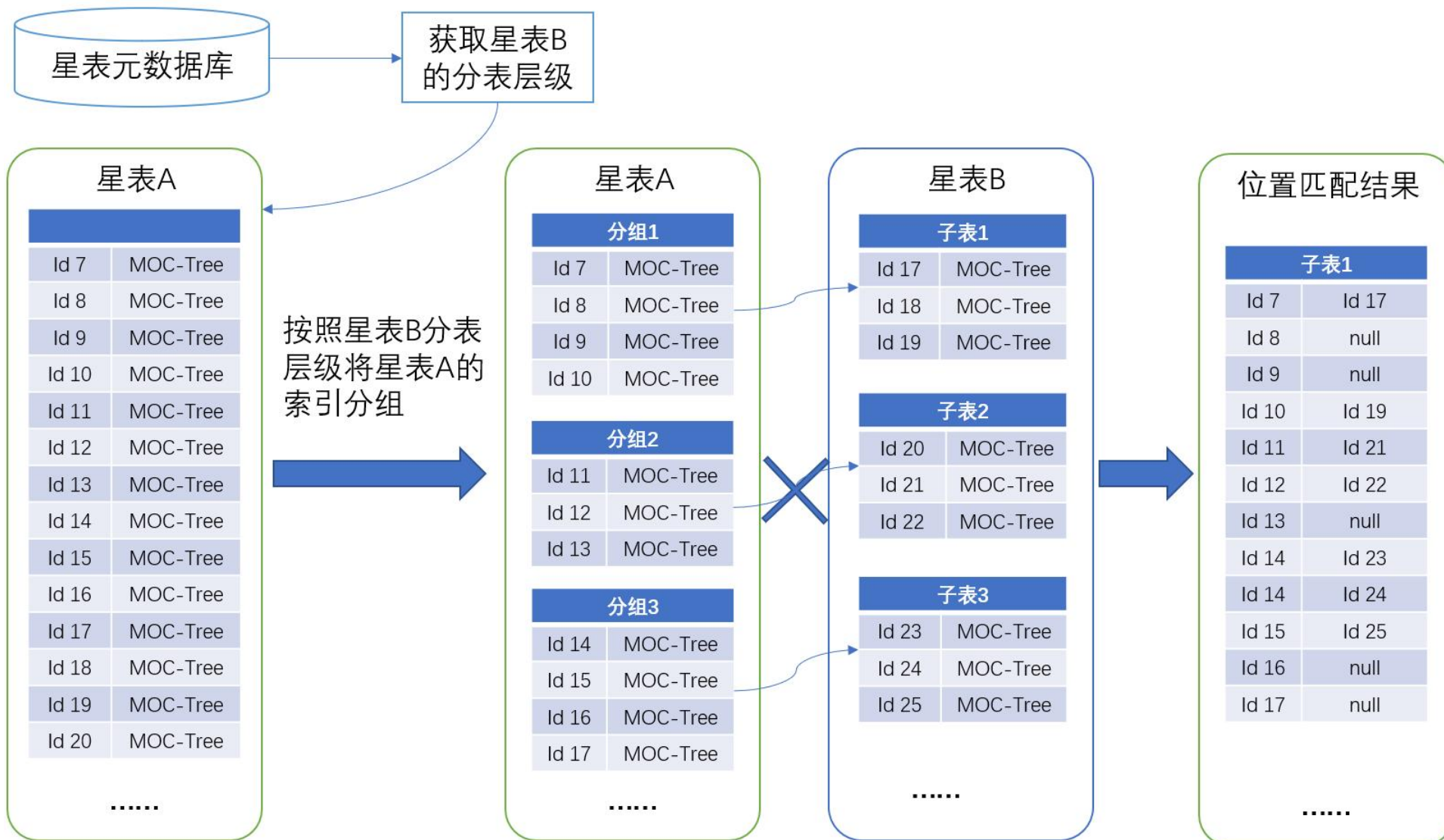
A-ld 6	B-id 1
A-ld 7	B-id 7
A-ld 7	B-id 8
.....	

# 3.2 星表交叉证认及置信度估计



## 基于MOC-Tree索引的多星表位置匹配

- 以定位误差较大、规模较小的星表为主星表进行位置匹配
- 若星表存在分表，需要对参与空间匹配的星表进行拆分归并，再进行位置匹配





# 3.2 星表交叉证认及置信度估计



基于贝叶斯推断对位置匹配的结果进行置信度估计:

贝叶斯公式:  $P(H|D) \propto P(H) \times P(D|H)$

$$P(H) = \frac{N_1}{\prod_{i=1}^k N_i} = \frac{1}{\prod_{i=2}^k N_i} = \frac{1}{\prod_{i=2}^k v_i \Omega_i}$$

先验概率

$$P(D|H) = 2^{k-1} \frac{\prod \sigma_i^{-2}}{\sum \sigma_i^{-2}} \exp \left\{ -\frac{\sum_{i<j} \phi_{ij} \sigma_j^{-2} \sigma_i^{-2}}{2 \sum \sigma_i^{-2}} \right\}$$

似然函数

星表A ID	星表 B ID	星表C ID	P_i	P_any
x1	b1	C1	34%	91%
x1	b1	C2	33%	91%
x1	b1	none	13%	91%
		.....		

$$P_i = \frac{P(H_i|D)}{\sum_{i>0} P(H_i|D)}$$

第i组匹配的置信度估计

$$P_{any} = 1 - \frac{P(H_0|D)}{\sum_i P(H_i|D)}$$

该天体在其它星表中存在对应体的概率

参考文献:  
 Salvato M., Buchner J., Budavári T. et al  
 2018 MNRAS 473  
 4937

# 3.2星表交叉证认及置信度估计



基于XMM-COSMOS多波段星表进行了方法验证

XMM-COSMOS多波段星表涵盖X射线波段、光学波段、红外波段的星表；

以XMM-COSMOS的XMM星表为主星表，分别与多波段星表数据库中的SDSS、2MASS、FIRST、Chandra星表进行交叉证认，并对比了采用Q3C和HEALPix索引方式的时间花费。

- XMM-COSMOS中基于XMM观测数据的X射线星表包含1848个点源，基于Chandra数据对其中1281源进行了对应体确认、128源的对应体进行了修正；
- 基于上述1409个源在光学波段（SDSS）和红外波段（2MASS）来验证交叉证认的准确性；
- 共计1312个正确匹配，其中1259个源的对应体与XMM-COSMOS原始星表中的对应体一致，53个源的对应体与基于Chandra修正后的对应体一致，准确率为93.1%，高于LR方法的90.9%的准确率。

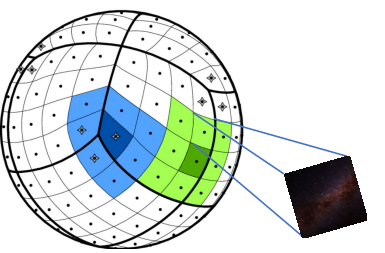
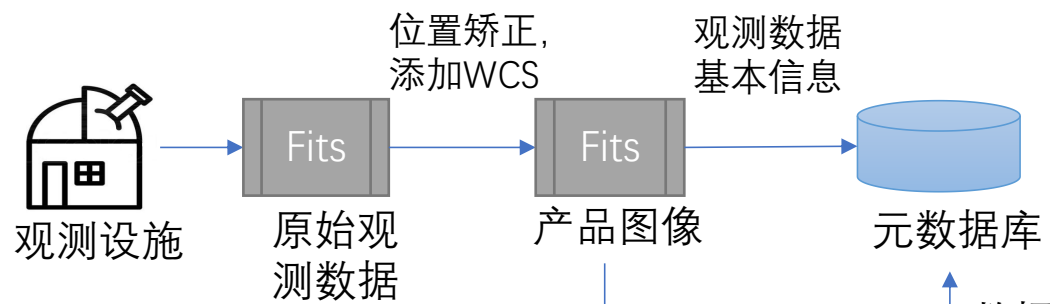
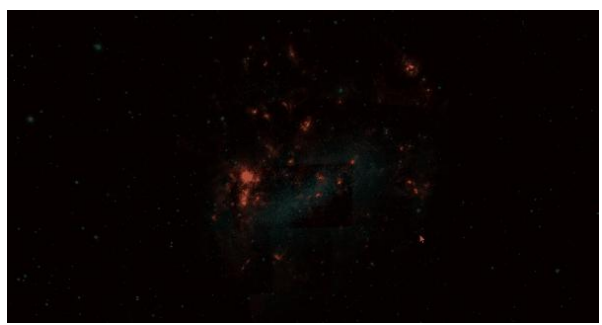
星表名称	星表条目	索引类型	位置匹配时间 (ms)
		MOCTree-分表分区 (129)	419.40
SDSS	1,231,051,050	Q3C	59432.30
		HEALPix	73223.80
		MOCTree-分表分区 (48)	394.20
2MASS	470,992,971	Q3C	39226.60
		HEALPix	46784.30
		MOCTree	173.20
FIRST	946,432	Q3C	198.60
		HEALPix	224.10
		MOCTree	131.40
Chandra	317,167	Q3C	158.20
		HEALPix	195.40



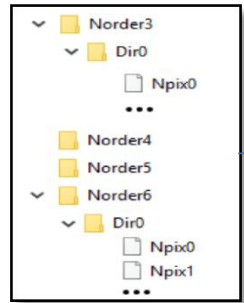
# 3.3 异构多波段图像的高效组织、检索及可视化



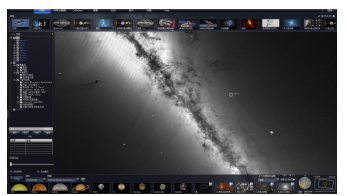
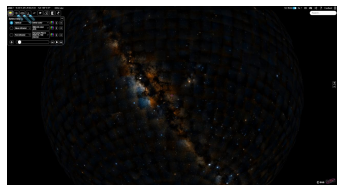
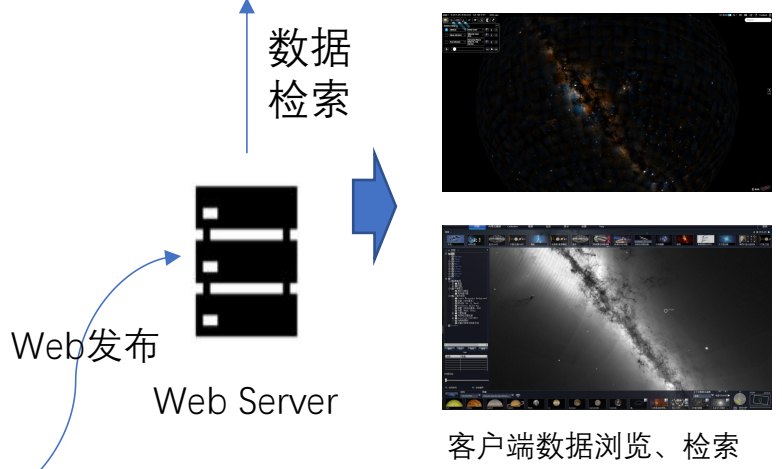
IVOA HiPS标准, 为实现异构多波段图像数据的高效组织、检索及可视化提供了思路



基于天球网格划分的图像切片



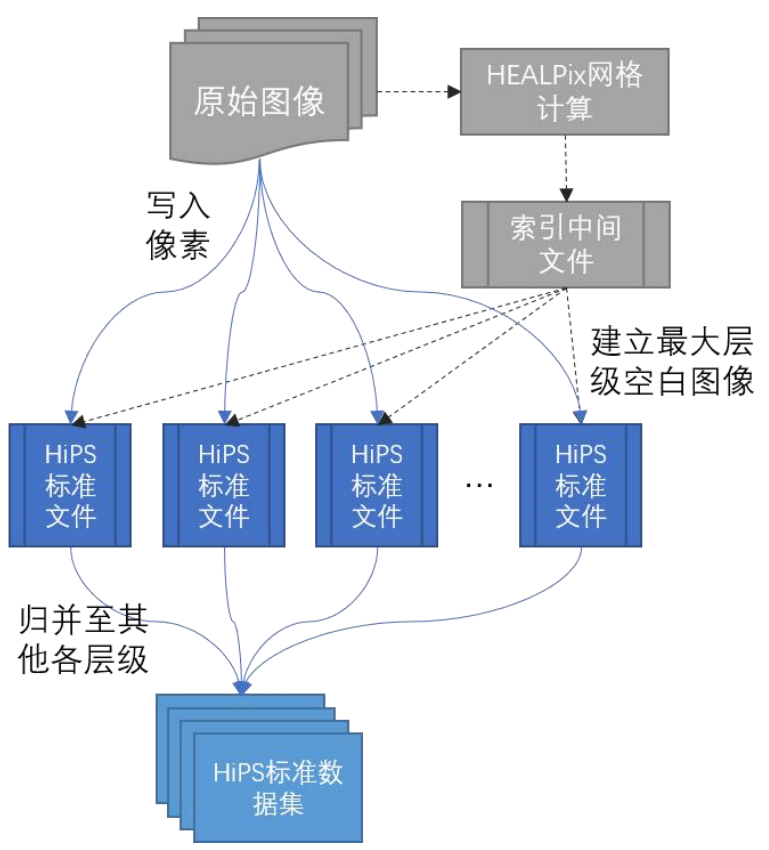
分层文件存储



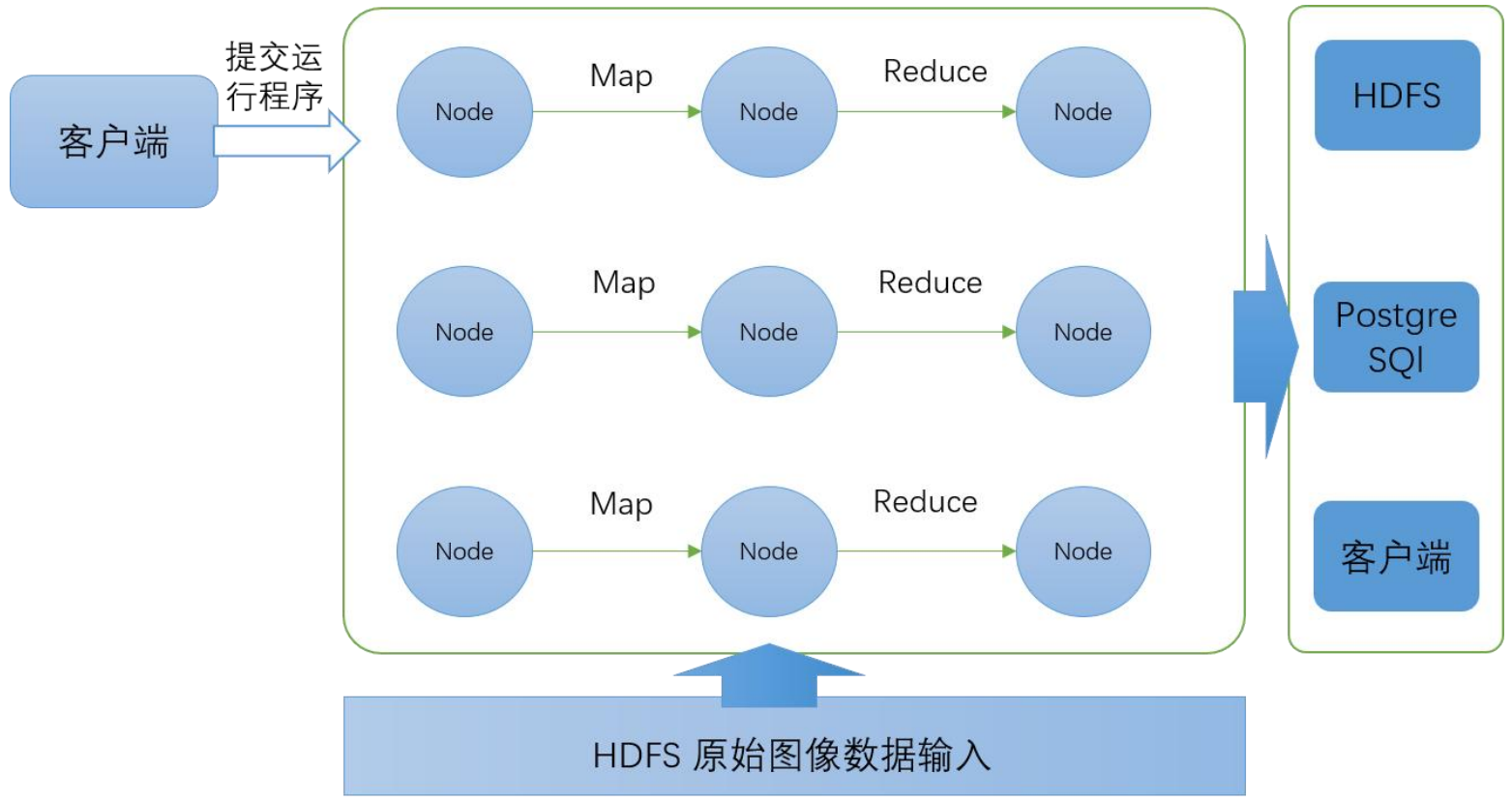
客户端数据浏览、检索

# 3.3 异构多波段图像的高效组织、检索及可视化 China-VO

针对原始图像数据转换至 HiPS标准数据集效率较低的问题开展研究：



基于像素并行写入的转换方法流程



基于Spark内存运算实现像素并行写入方法

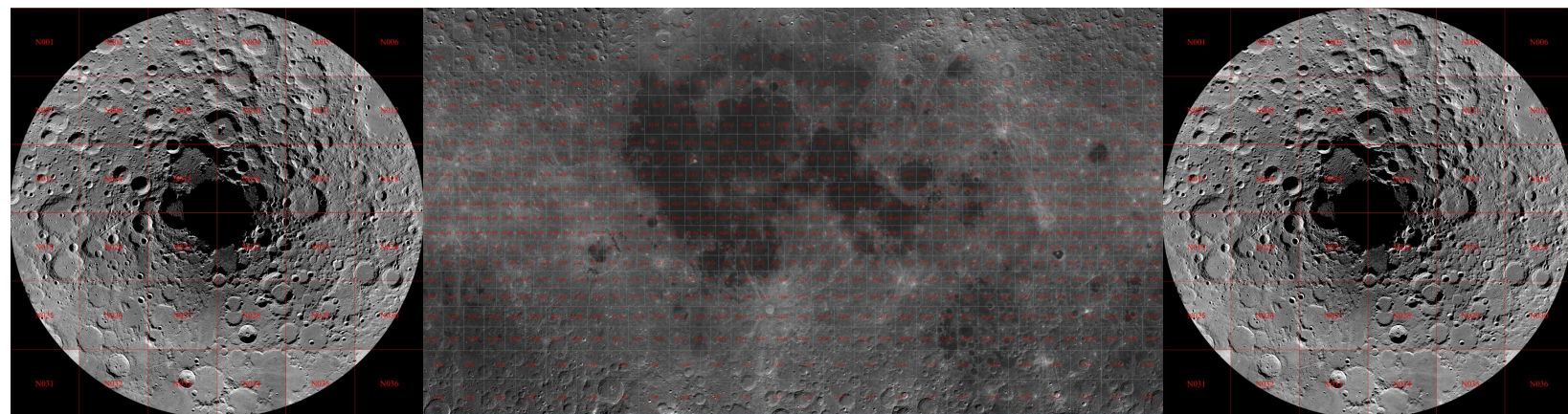
# 3.3 异构多波段图像的高效组织、检索及可视化



## 效率测试:

### 测试数据:

嫦娥2号7m分辨率正射影像  
总数据量747.9GB, 844个分幅  
每幅数据在1GB左右  
数据格式: Tiff、Prj



### HiPSGen 运行服务器配置

类型	详情
CPU	3.0GHz
逻辑核数	4核
内存	64GB

➡ 152h

~4.5h ⬅

### Spark集群配置

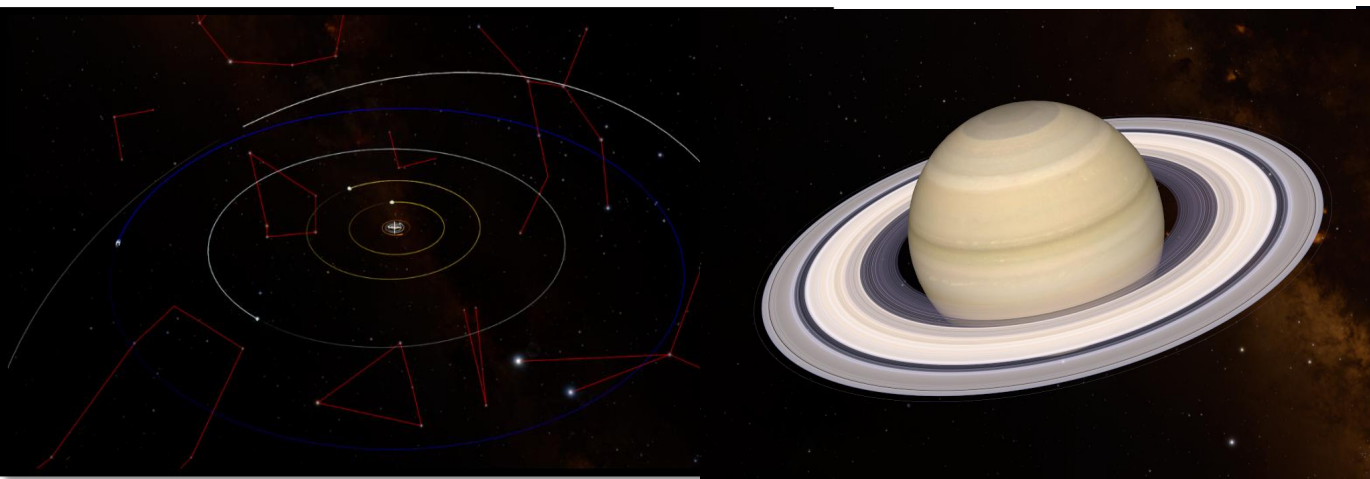
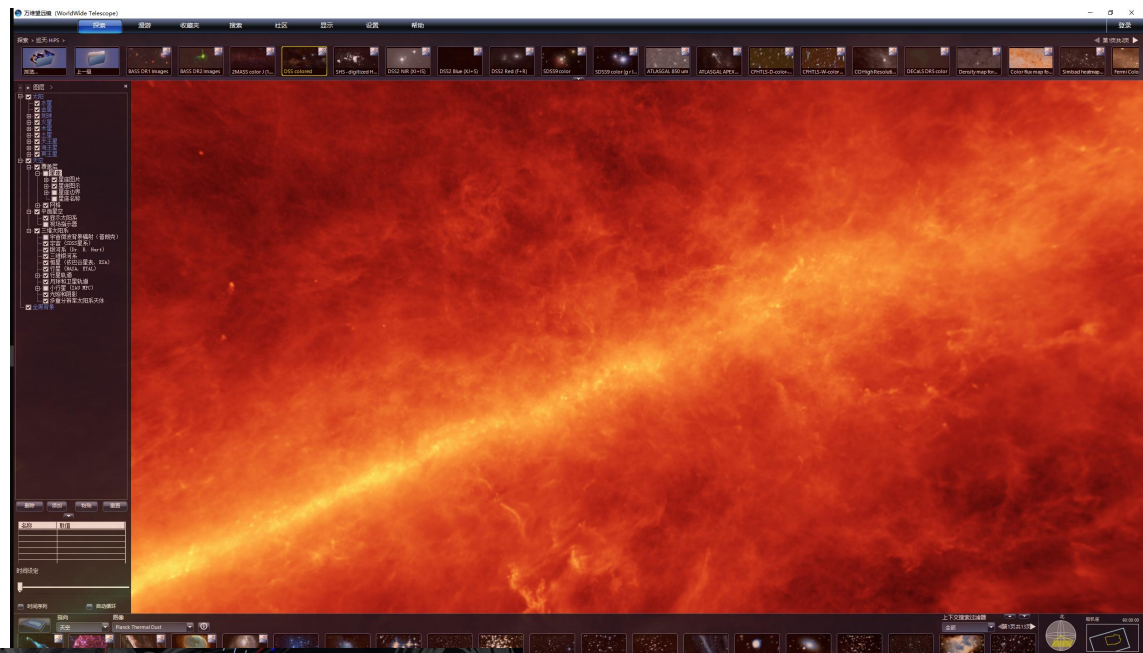
类型	详情
节点数	9个
节点CPU	1 x 3.0GHz
节点逻辑核数	4核
节点内存	8GB
Spark版本	2.2.0



# 3.3 异构多波段图像的高效组织、检索及可视化

- 万维望远镜中实现HiPS标准数据集的加载与可视化

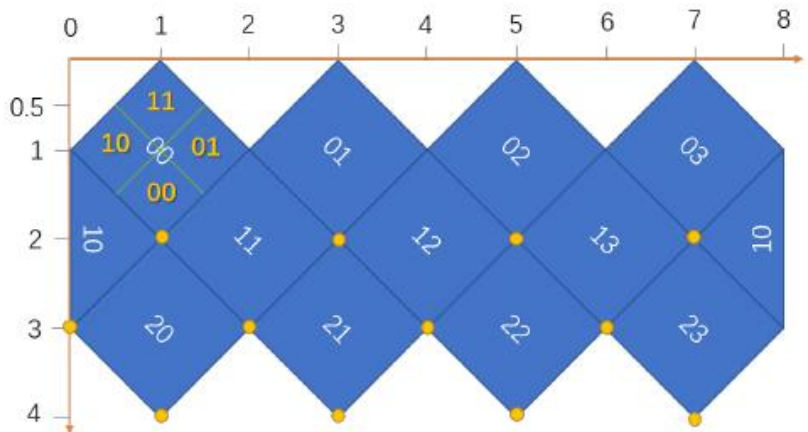
- 天文科学数据可视化软件平台
- 微软研究院于2008年开发，并于2015年开源贡献给天文社区，由AAS负责维护和推广
- 广泛用于天文科学数据可视化、天文科普教育
- 中文版本由China-VO团队维护
- 数据组织及可视化标准为TOAST，不支持HiPS标准





# 3.3 异构多波段图像的高效组织、检索及可视化 China-VO

- 三维虚拟场景中HEALPix多层次网格的绘制



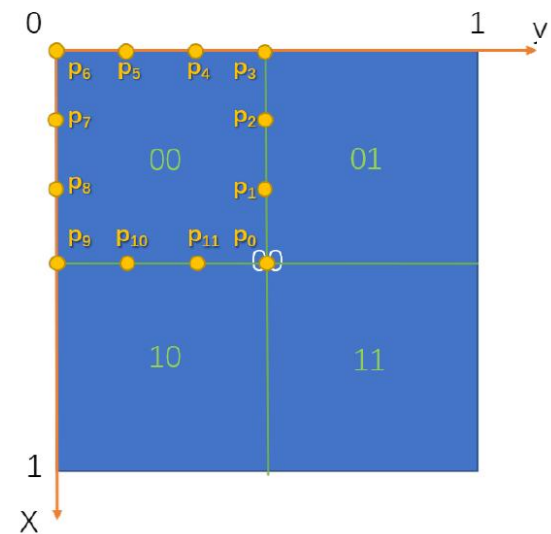
$$z = \frac{(p_x + p_y)^2}{3} - 1$$

$$\phi = \frac{\pi}{2} \left( 2 \text{face} - 15 + \frac{p_x - p_y}{p_x + p_y} \right)$$

$$x = \sqrt{(1 - z)(1 + z)} \cos \phi$$

$$y = \sqrt{(1 - z)(1 + z)} \sin \phi$$

$$z = z$$

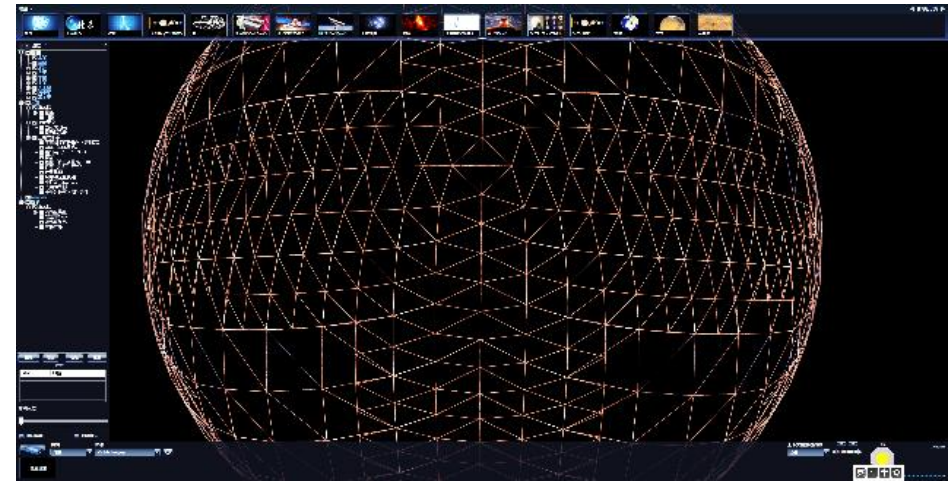
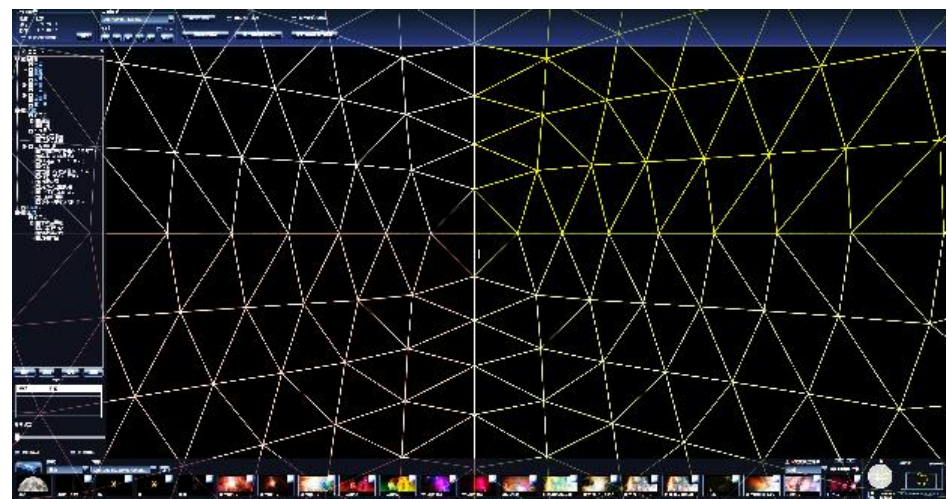


$$(P_{i_x}, P_{i_y}) = \left( \frac{1 + x}{N_{side}} - \frac{i}{step \times N_{side}}, \frac{1 + y}{N_{side}} \right)$$

$$(P_{(i+step)_x}, P_{(i+step)_y}) = \left( \frac{x}{N_{side}}, \frac{1 + y}{N_{side}} - \frac{i}{step \times N_{side}} \right)$$

$$(P_{(i+2step)_x}, P_{(i+2step)_y}) = \left( \frac{x}{N_{side}} + \frac{i}{step \times N_{side}}, \frac{1 + y}{N_{side}} \right)$$

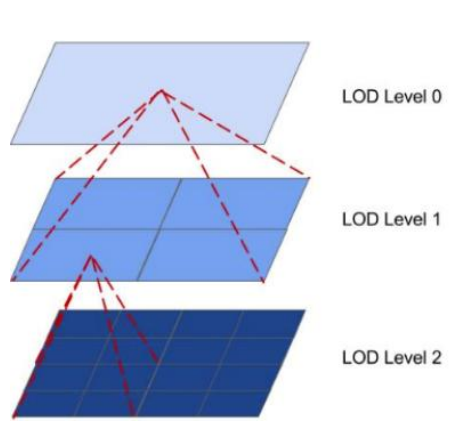
$$(P_{(i+3step)_x}, P_{(i+3step)_y}) = \left( \frac{1 + x}{N_{side}}, \frac{y}{N_{side}} + \frac{i}{step \times N_{side}} \right)$$



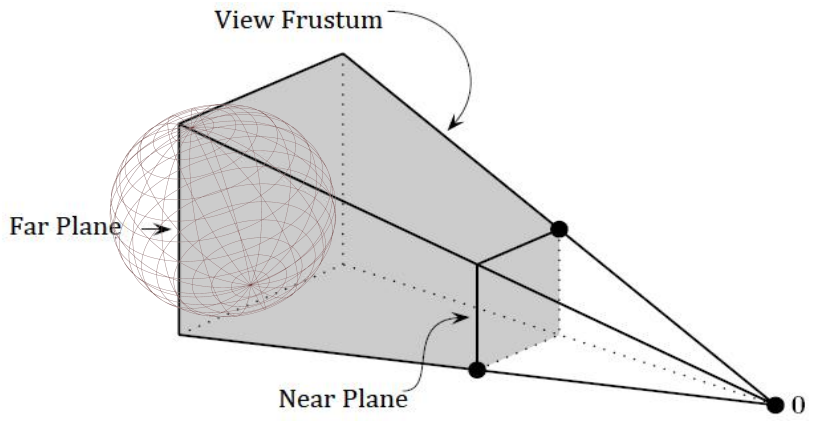


# 3.3 异构多波段图像的高效组织、检索及可视化 China-VO

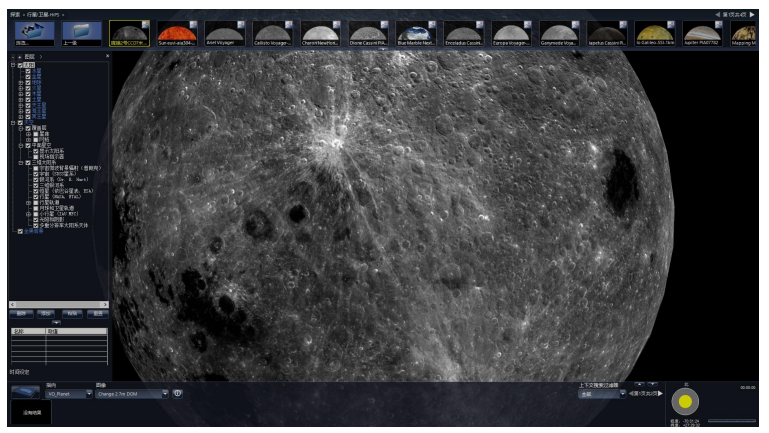
- 层次细节模型与视场剔除



层次细节模型

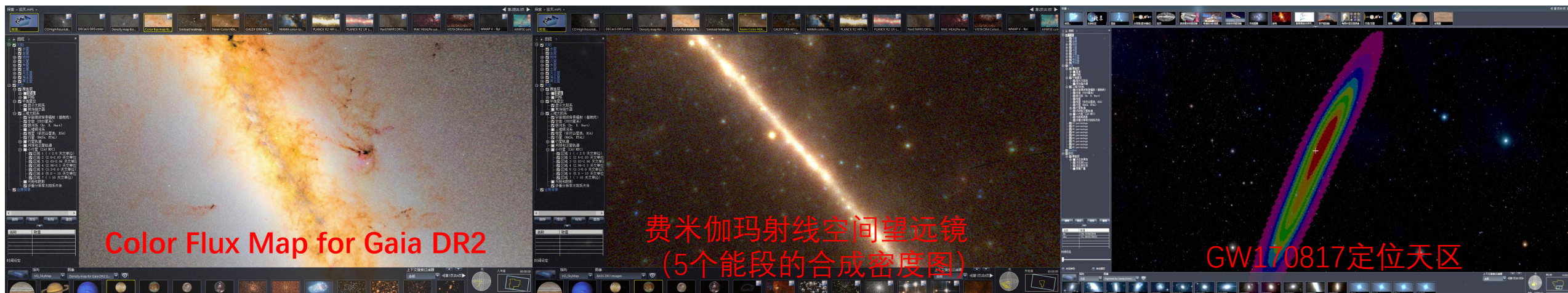


视场剔除



随视场的数据加载及可视化

- 可视化效果





# 3.3 异构多波段图像的高效组织、检索及可视化



- HiPS标准数据集与万维望远镜原生数据渲染效率对比

6组数据，分别包含HiPS标准数据集和WWT原生数据集（TOAST）。结果证明，在相同图像切分分辨率下，HiPS标准数据集在渲染时的内存消耗及加载速度上均有优势。

Data name	Data type	Ti	Tm	Max order	Mi	Mm	Tile size
Digitized Sky Survey	TOAST-Skymap	2.28 s	7.73 s	12	851 920 kB	825 584 kB	256 × 256
	HiPS-Skymap	2.25 s	8.60 s	9	1 327 612 kB	1 488 840 kB	512 × 512
WISE All Sky	TOAST-Skymap	1.90 s	9.58 s	7	821 484 kB	821 832 kB	256 × 256
	HiPS-Skymap	2.10 s	3.86 s	8	1 551 080 kB	1 221 216 kB	512 × 512
2Mass: Imagery	TOAST-Skymap	2.23 s	11.86 s	8	864 448 kB	821 052 kB	256 × 256
	HiPS-Skymap	2.30 s	4.83 s	9	1 335 512 kB	1 306 756 kB	512 × 512
Moon: Lunar Reconnaissance Orbiter WAC Global Morphologic Map	TOAST-Planet	2.12 s	16.66 s	10	817 988 kB	840 520 kB	256 × 256
	HiPS-Planet	1.65 s	1.53 s	6	1 413 816 kB	23 987 602 kB	512 × 512
Jupiter: PIA07782, Cassini's Best Maps of Jupiter	TOAST-Planet	1.43 s	4.46 s	4	846 060 kB	831 084 kB	256 × 256
	HiPS-Planet	2.30 s	0.98 s	3	1 756 340 kB	1 083 836 kB	512 × 512
PLANCK	TOAST-Skymap	0.71 s	0.54 s	6	990 892 kB	1 000 896 kB	256 × 256
	HiPS-Skymap	0.63 s	0.59 s	3	757 152 kB	712 348 kB	256 × 256



# 3.4 EP 暂现源判别及多波段证认



WXT宽视场X射线望远镜采用了龙虾眼结构，视场达3600平方度，大视场监测，角分级定位

~10到几百平方度



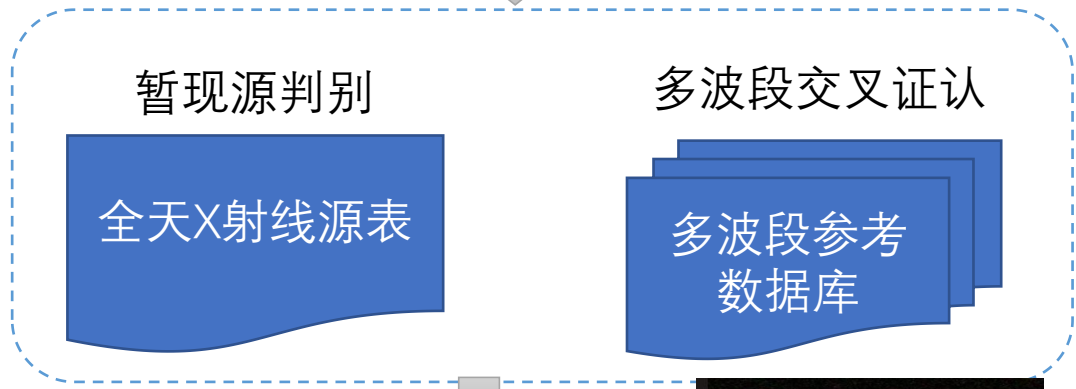
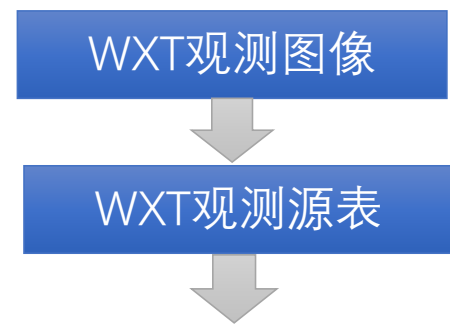
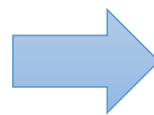
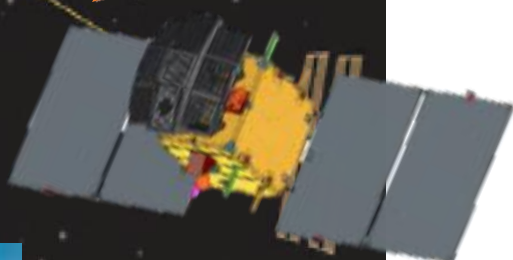
视场完全覆盖  
精度~10角秒-1角分

X射线

引力波

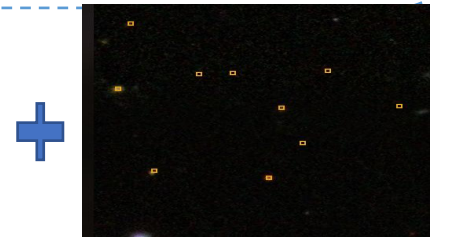
A-LIGO

A-VIRGO



波段A ID	波段B ID	波段C ID	置信度
x1	b1	c1	87%
x1	b1	c2	3%

多波段交叉证认结果



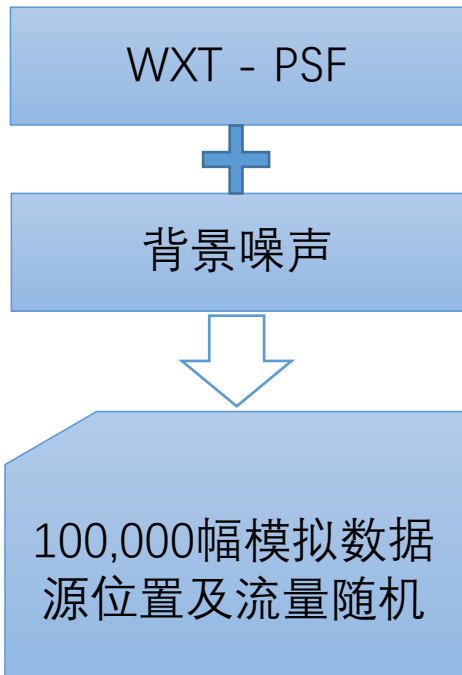
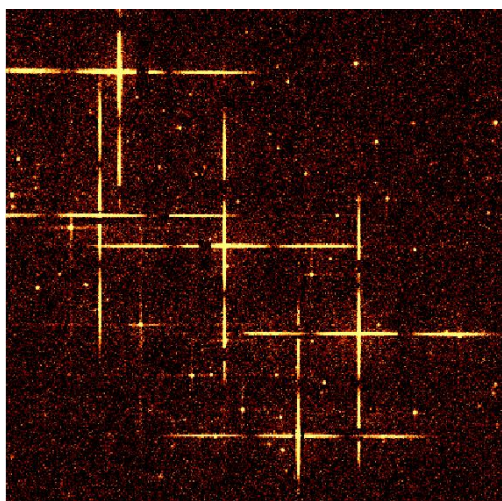
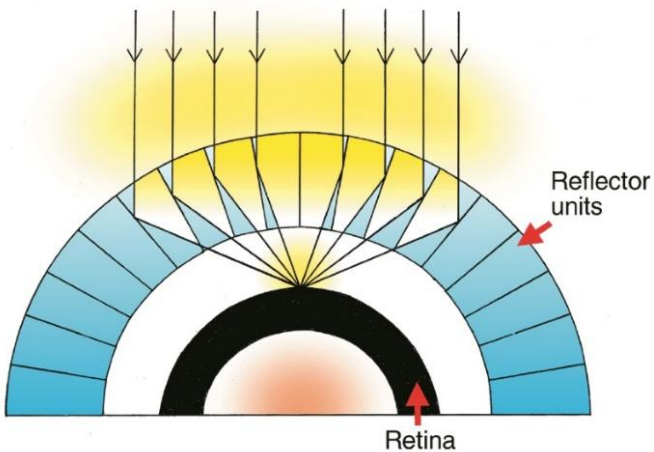
多波段图像辅助证认

# 3.4 EP 暂现源判别及多波段证认

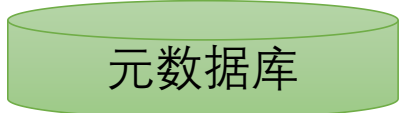
ResNET50



## • EP WXT 观测源识别



源位置  
流量  
对应图像

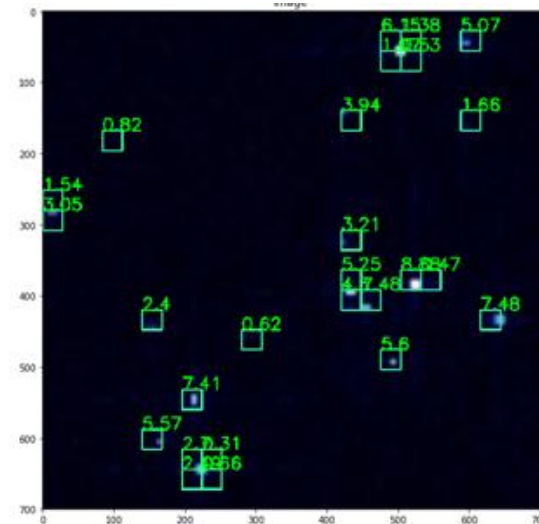
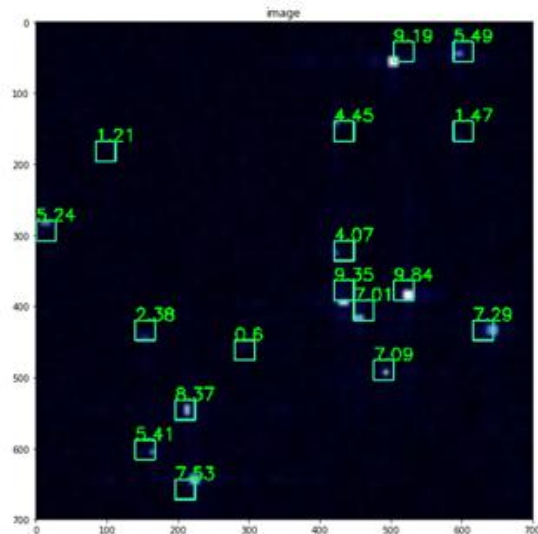


EPNet100

用于源探测和定位

EPNet25

用于源流量测量



- 源探测结果：查准率Precision=0.9259, 查全率Recall=0.9626, F1-Score =0.9439
- 流量预测平均相对误差5.33%
- 该模型将做为WXT源搜索的备选算法之一, 有机会应用到WXT数据处理pipeline之中

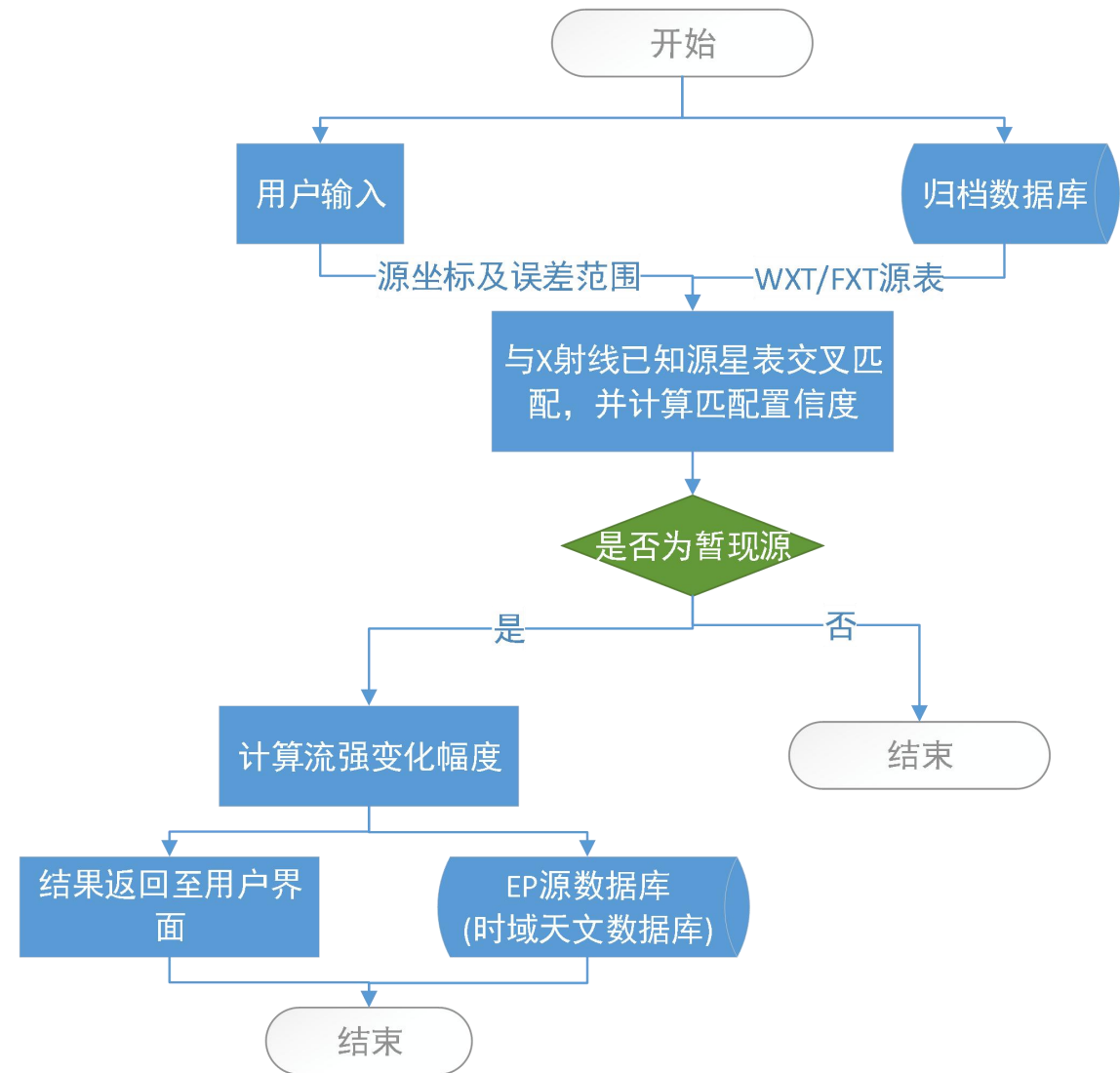
# 3.4 EP 暂现源判别及多波段证认



## • EP WXT 暂现源判别

X射线星表名称	天体数目
3XMM DR6 Source Catalogue	468,440
Chandra Source Catalogue 2.0	317,167
The Second ROSAT All-Sky Survey Point Source Catalog	135,000
ROSAT PSPC WGA Sources	84,000
Swift-BAT 105-Month Hard X-ray Survey	1,632
Swift-XRT point source	151,524
MAXI Master Catalog	338

- 由于已知X射线源表的数据条目并不多，不需要对其进行分表存储，仅需构建空间索引即可实现高效检索
- 将观测源的位置和WXT 观测误差范围（约1角分）作为输入检索已知源表，若检索结果为空，或 $P_{ny}$  值小于30%，则认为该观测源为暂现源



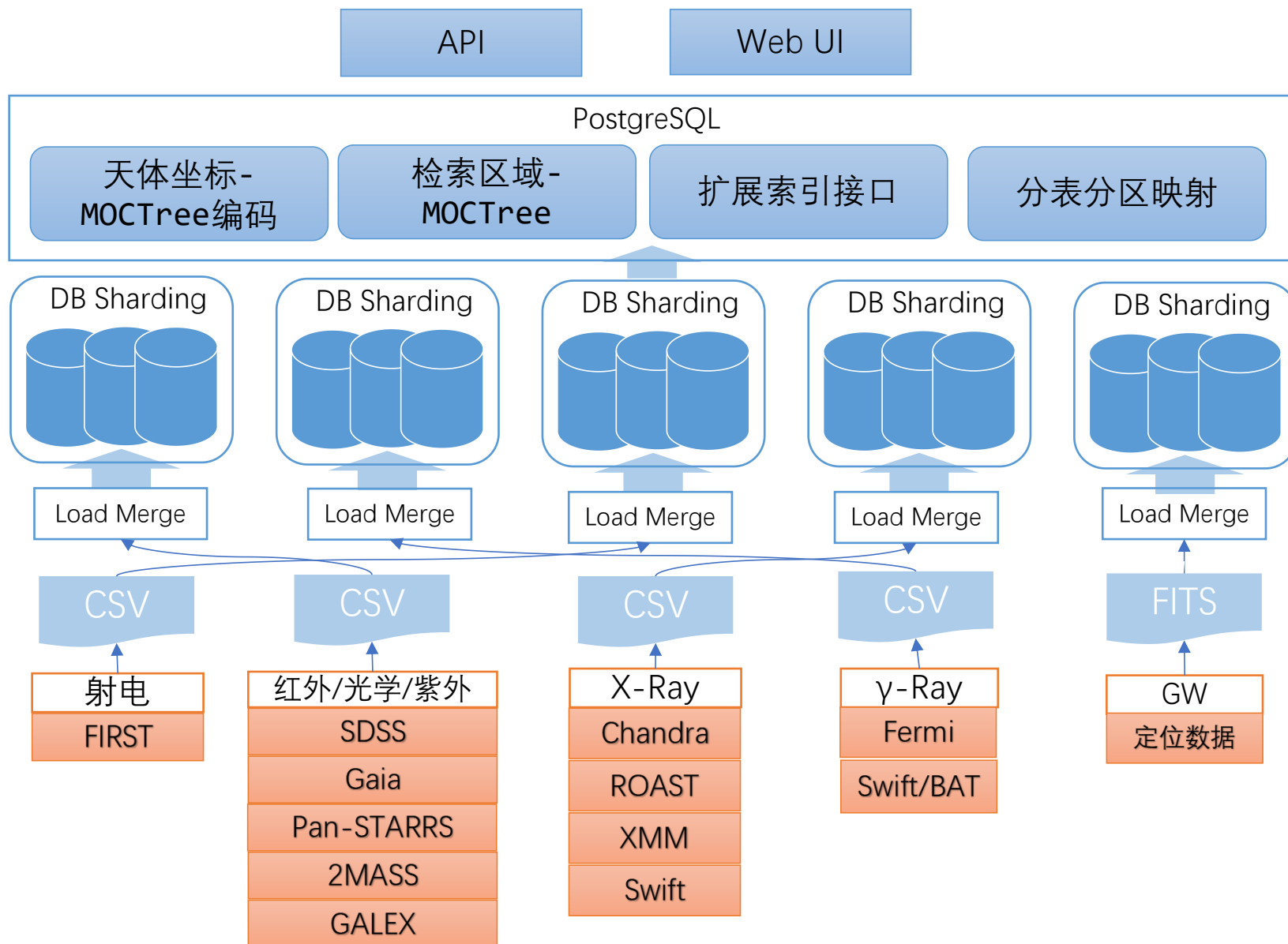


# 3.4 EP 暂现源判别及多波段证认



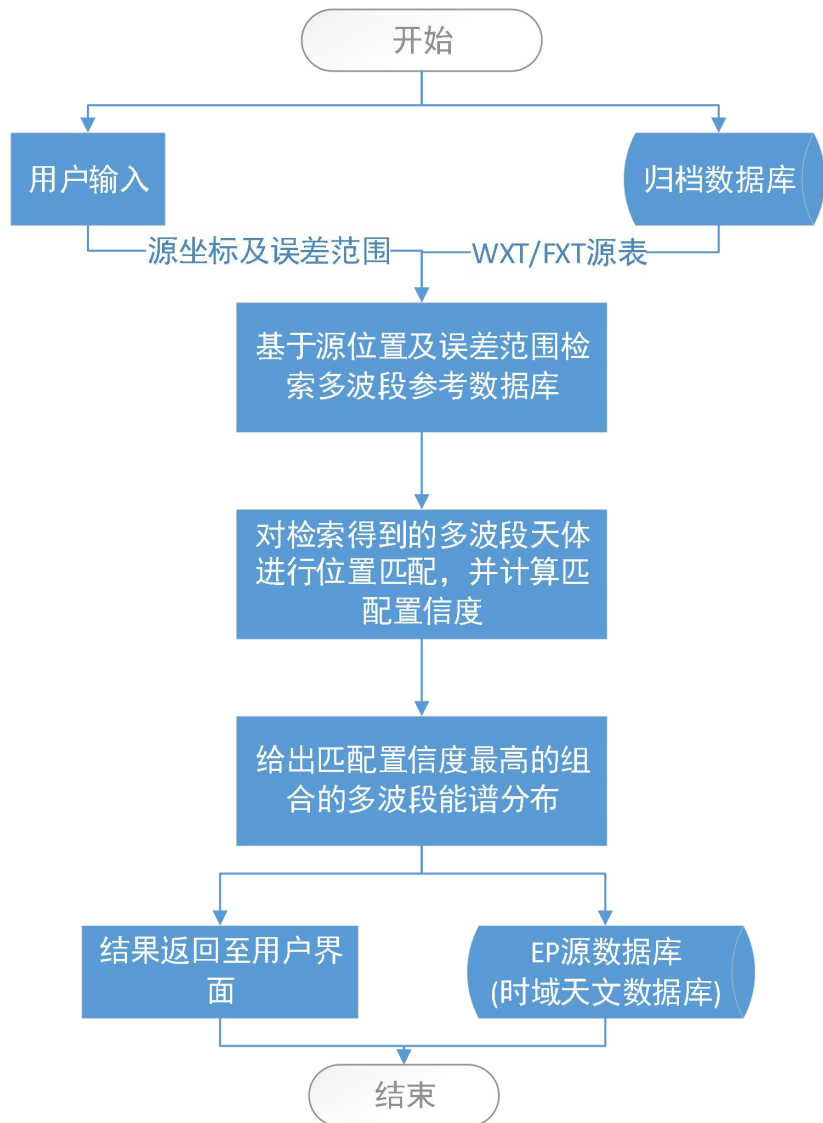
## 暂现源多波段参考数据库

- 集成本研究的主要关键技术
- 提供在线WebUI和API，可供EP数据处理pipeline调用，也可供EP科学用户在线使用
- 基于EP WXT模拟数据实测，完成一组50个暂现源在各个波段对应体搜索时间小于500ms，完成交叉证认置信度估计的时间小于30s
- 良好的可拓展性，可兼容未来发布的多波段星表



# 3.4 EP 暂现源判别及多波段证认

## 暂现源多波段交叉证认及可视化



以暂现源为核心的EP数据展示

- 快视数据
- 能谱数据
- 光变数据
- 图像数据
- 定标数据
- 多波段融合数据
- 质量数据

The interface displays a star field visualization with various data overlays. Below the visualization, there are two tables of transient data.

**HL\_LGRBs List**

ID	Source Name	RA	Dec	Flux	Absflux	Cmosnum	SNR	OBS Number
3	MJ000240-320028	0.67	-32.01	1.50e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	2.40e-11	24	6.10	104
16	MJ000914+384705	2.31	38.78	5.94e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	5.12e-11	33	17.09	8
18	MJ001059-363610	2.75	-36.60	5.00e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	3.95e-10	36	12.52	63
65	MJ003051-022022	7.72	-2.34	3.55e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	2.37e-10	4	8.75	74
66	MJ003110+535736	7.79	53.96	2.17e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	1.14e-10	42	7.22	1

**Lists of Transient: SGRB (two pointing observations per orbit) All 16 Rows**

ID	Source Name	RA	Dec	Flux	Absflux	Cmosnum	SNR	OBS Number	Show Details
1	MJ000330-250202	0.88	-25.03	3.17e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	8.92e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	24	43.26	1	Hide Details
2	MJ010604-055655	16.52	-5.95	2.16e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	3.62e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	1	21.17	1	Show Details
4	MJ012018-400151	20.08	-40.03	1.32e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	1.29e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	8	14.95	1	Show Details
8	MJ021602+714922	34.01	71.82	7.09e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	6.14e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	1	28.70	6	Show Details
10	MJ032433-032642	51.14	-3.45	6.61e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	3.84e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	35	21.92	4	Show Details
11	MJ043247+245909	68.20	24.99	5.24e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	4.07e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	37	16.13	1	Show Details
12	MJ044440+495403	71.17	49.90	1.57e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	2.80e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	41	2.70	59	Show Details
14	MJ050139-692519	75.41	-69.42	1.23e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	3.71e-11 erg cm <sup>-2</sup> s <sup>-1</sup>	1	5.09	1	Show Details
16	MJ054819+492100	87.08	49.35	1.45e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	1.33e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	8	8.81	1	Show Details
21	MJ074316+142638	115.82	14.44	7.27e-9 erg cm <sup>-2</sup> s <sup>-1</sup>	1.14e-10 erg cm <sup>-2</sup> s <sup>-1</sup>	37	11.09	20	Show Details

## 3.5 引力波电磁对应体高效搜寻



### 目标:

对引力波定位天区中的星系进行筛选，并给出观测优先级，支持地面观测网络的快速后随观测

### 主要问题:

1. 如何高效获取引力波定位天区（100-1000平方度）中包含的星系？
2. 如何对筛选出的星系制定观测优先级？
3. 如何从观测数据中快速证认引力波电磁对应体候选体？

引力波事件后随观测规划及暂现源证认系统

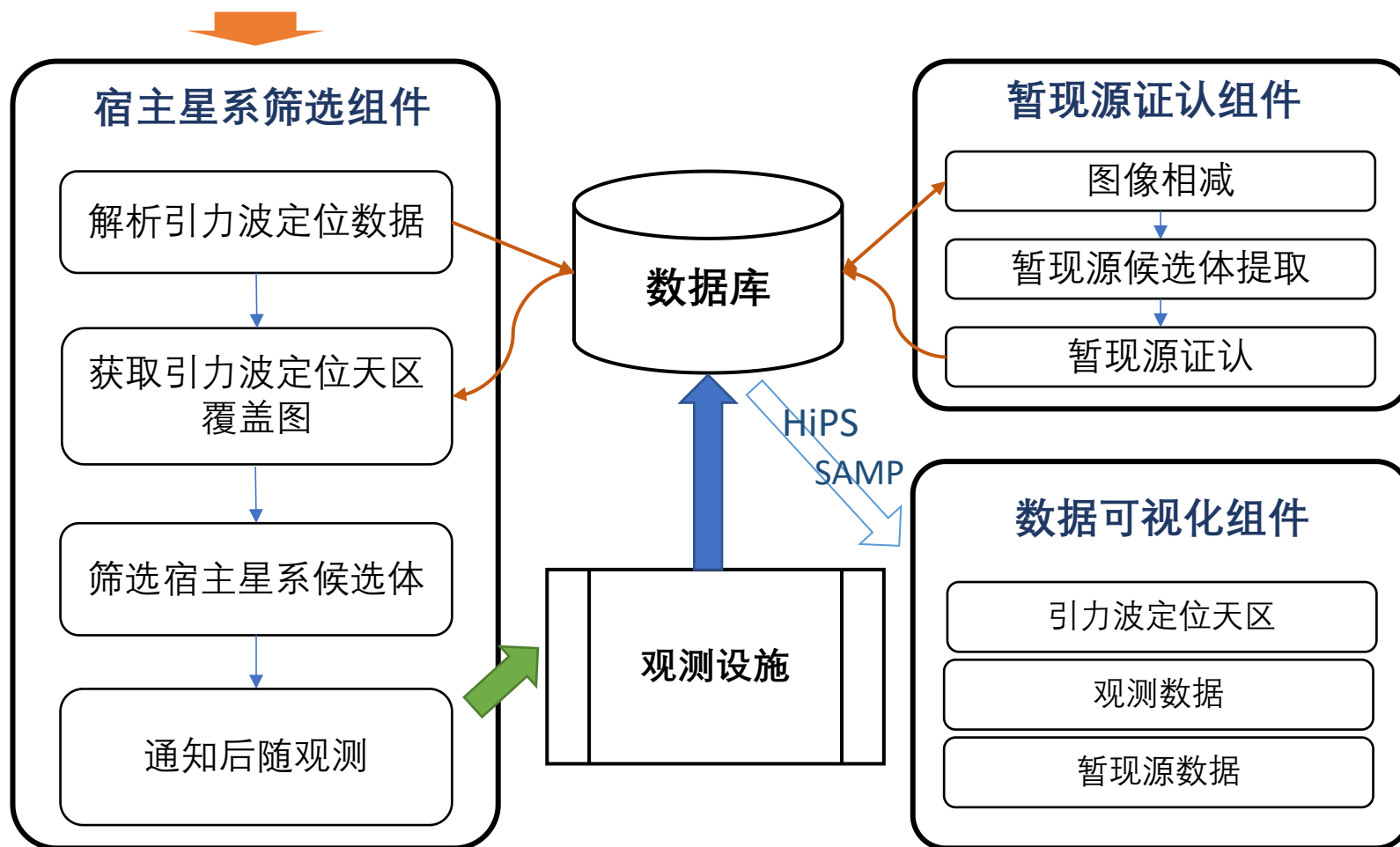
GWOPS

引力波事件后随观测数据库

# 3.5 引力波电磁对应体高效搜寻



引力波事件警报



GWOPS组成

- ① 宿主星系筛选组件
- ② 暂现源证认组件
- ③ 数据可视化组件



# 3.5 引力波电磁对应体高效搜寻



ot_gw [table]	
id	serial[10]
gw_name	varchar[20]
gw_date	timestamp[29,6]
comment	varchar[50]
< 0	191 rows 0 >

ot_classification [table]	
id	serial[10]
classify	varchar[20]
ml	int4[10]
comment	varchar[50]
< 0	0 rows 0 >

ot_object_hpx [table]	
id	serial[10]
ot_object_id	int4[10]
hpx	int4[10]
< 0	364,367 rows 0 >

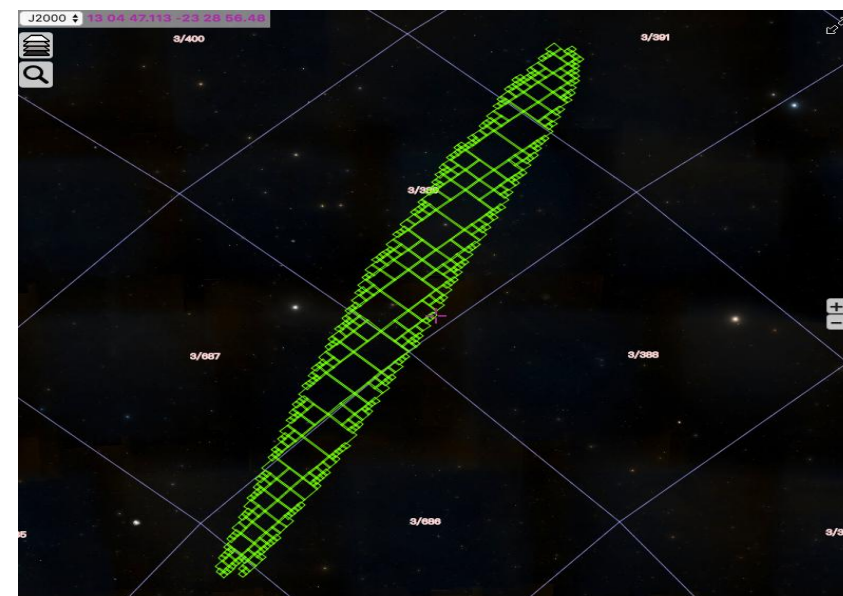
glade [table]	
pgc	int4[10]
gwgc_name	varchar[80]
hyperleda_name	varchar[80]
2mass_name	varchar[80]
sdssdr12_name	varchar[80]
flag1	bpchar[1]
ra	float4[8,8]
dec	float4[8,8]
dist	float4[8,8]
dist_err	float4[8,8]
z	float4[8,8]
b	float4[8,8]
b_err	float4[8,8]
b_abs	float4[8,8]
j	float4[8,8]
j_err	float4[8,8]
h	float4[8,8]
h_err	float4[8,8]
k	float4[8,8]
k_err	float4[8,8]
flag2	bpchar[1]
flag3	bpchar[1]
hpx	int8[19]
< 0	3,262,881 rows 0 >

ot_tmpt [table]	
id	serial[10]
tmpt_name	varchar[100]
tmpt_path	varchar[100]
tmpt_date	timestamp[29,6]
tmpt_id	numeric[14,7]
tmpt_ra_pointing	varchar[20]
tmpt_dec_pointing	varchar[20]
tmpt_ra_center	numeric[10,5]
tmpt_dec_center	numeric[10,5]
tmpt_ra_left_top	numeric[10,5]
tmpt_dec_left_top	numeric[10,5]
tmpt_ra_left_bottom	numeric[10,5]
tmpt_dec_left_bottom	numeric[10,5]
tmpt_ra_right_bottom	numeric[10,5]
tmpt_dec_right_bottom	numeric[10,5]
tmpt_ra_right_top	numeric[10,5]
tmpt_dec_right_top	numeric[10,5]
tmpt_object	varchar[50]
tmpt_airmass	numeric[10,5]
tmpt_imagetype	varchar[20]
tmpt_cod_temperature	numeric[10,5]
tmpt_telescope	varchar[20]
tmpt_instrument	varchar[20]
tmpt_color_band	varchar[20]
tmpt_exptime	numeric[10,5]
tmpt_limit_mag_3sigma	numeric[10,5]
tmpt_limit_mag_5sigma	numeric[10,5]
tmpt_fits_zero	numeric[10,5]
comment	varchar[50]
obj_stars_sum	int4[10]
< 0	15,676 rows 0 >

ot_ot [table]	
id	serial[10]
gw_name	varchar[20]
gw_id	int4[10]
diff_id	int4[10]
ot_number	int4[10]
ot_name	varchar[20]
ot_date	timestamp[29,6]
ot_jpg_name	varchar[100]
ot_image_x	numeric[10,5]
ot_image_y	numeric[10,5]
ot_wcs_ra	numeric[10,5]
ot_wcs_dec	numeric[10,5]
ot_learning_marks	numeric[10,5]
ot_ranking_marks	numeric[10,5]
ot_check	bool[1]
ot_checker	varchar[20]
ot_magnitude	numeric[10,5]
ot_classification	varchar[20]
comment	varchar[100]
ot_jpg_path	varchar[100]
< 0	368,943 rows 0 >



- GWGC
- 2MPZ
- 2MASS XSC
- HyperLEDA
- SDSS DR12Q



将引力波定位天区转化为多层次覆盖天区，检索构建了MOC-Tree索引的GLADE星表，得到该天区内的星列表

# 3.5 引力波电磁对应体高效搜寻



- 基于贝叶斯推断的宿主星系候选体排序

- 先验概率：GLADE星表中星系的位置信息、B波段亮度信息、引力波事件的估算距离
- 后验概率：基于贝叶斯推断计算星系候选体是当前引力波事件宿主星系的后验概率，根据概率大小排序

后验概率计算公式：

$$p(\gamma|D, S, M) = \frac{p(\gamma|S, M)p(D|S, \gamma, M)}{p(D|S, M)} \quad (1)$$

$$\Rightarrow p(\gamma|D, S, M) = \frac{p(\gamma|D, M)p(\gamma|S, M)}{p(D|S, M)} \quad (2)$$

先验概率计算公式：

$$p(\gamma|S, M) \propto \left(\frac{D_{gc}}{D_{GW}}\right)^3 \frac{1}{N} \sum_{j=1}^N \delta(\alpha - \alpha_j, \beta - \beta_j, d - d_j) L_{B_j} + \frac{3L_{Bmean}}{4\pi D_{GW}^3} H(d - D_{gc}) d^2$$

其中：

$\gamma$ ：GW和EM两组观测的公共参数，指覆盖天区中星系的位置参数， $\alpha, \beta, d$

$D$ ：GW数据集

$S$ ：EM数据集

$M$ ：GW宿主源星系数据

$p(\gamma|S, M)$  为先验概率公式

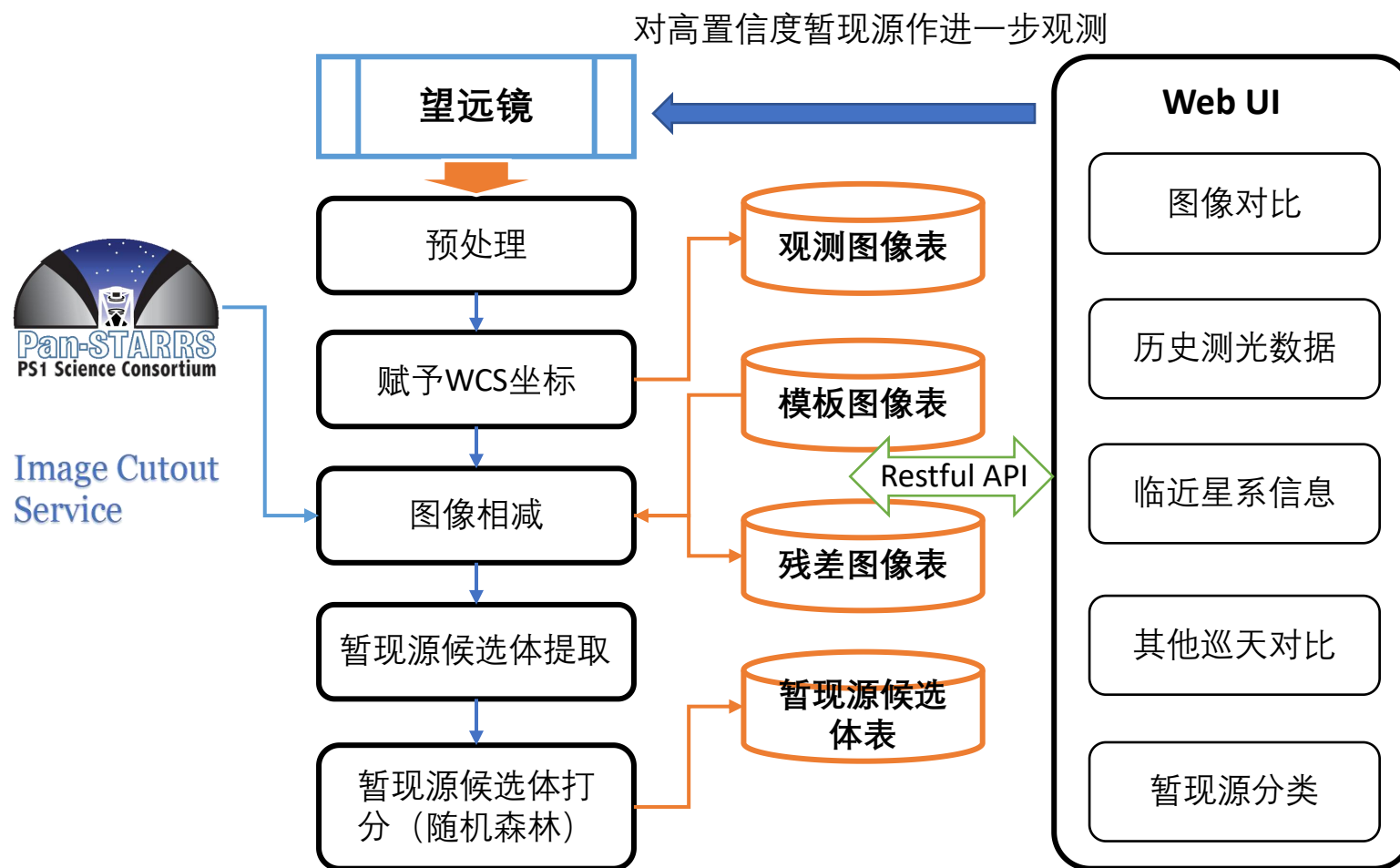
$p(D|S, \gamma, M)$  为给定 $S, \gamma$ 的数据集 $D$ 的似然概率

Fan, X., Messenger, C., & Heng, I. S. (2014). A Bayesian approach to multi-messenger astronomy: Identification of gravitational-wave host galaxies. The Astrophysical Journal, 795(1), 4

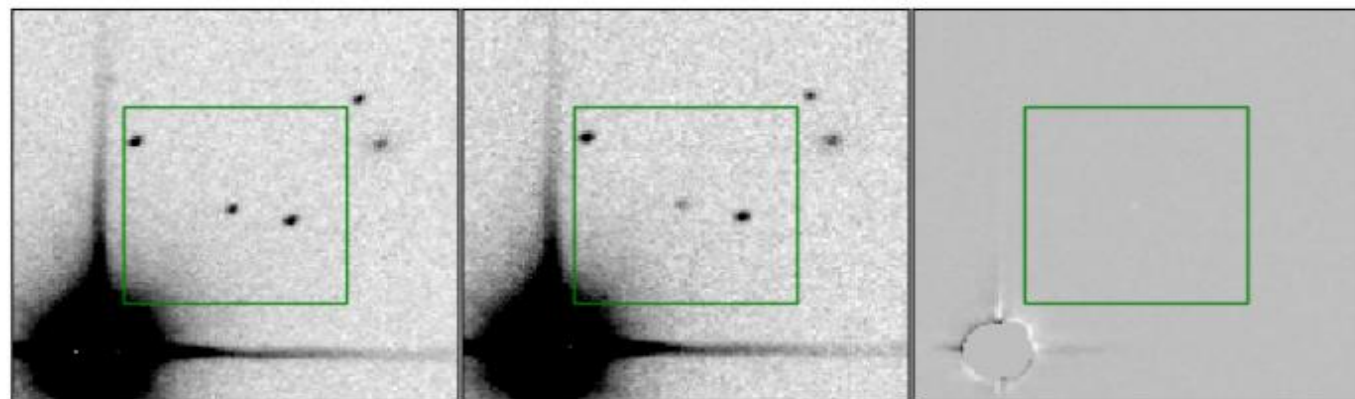
# 3.5 引力波电磁对应体高效搜寻



- 观测数据暂现源提取与证认



# 3.5 引力波电磁对应体高效搜寻

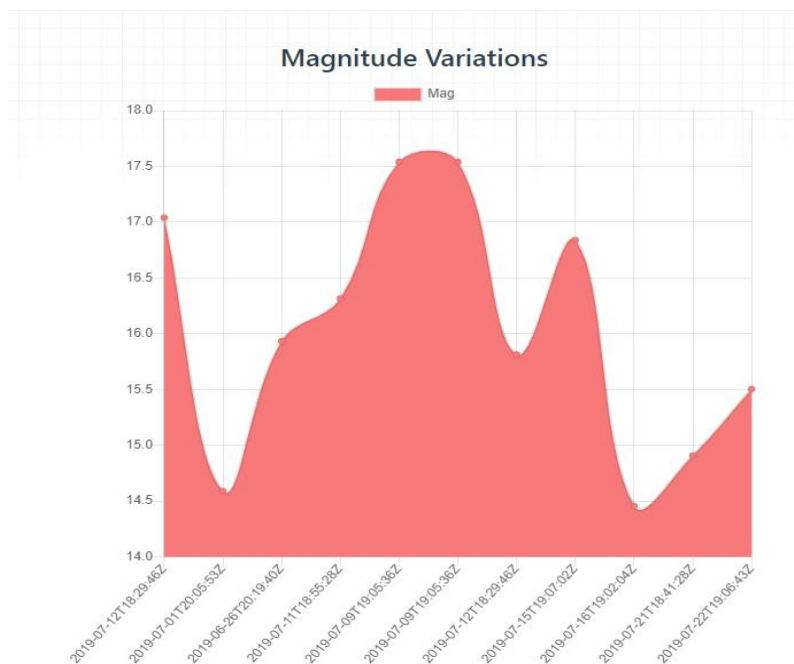
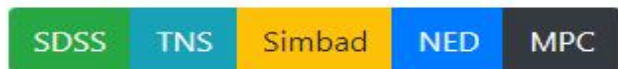


原始观测图像

模板图像

残差图像

对比其它巡天数据



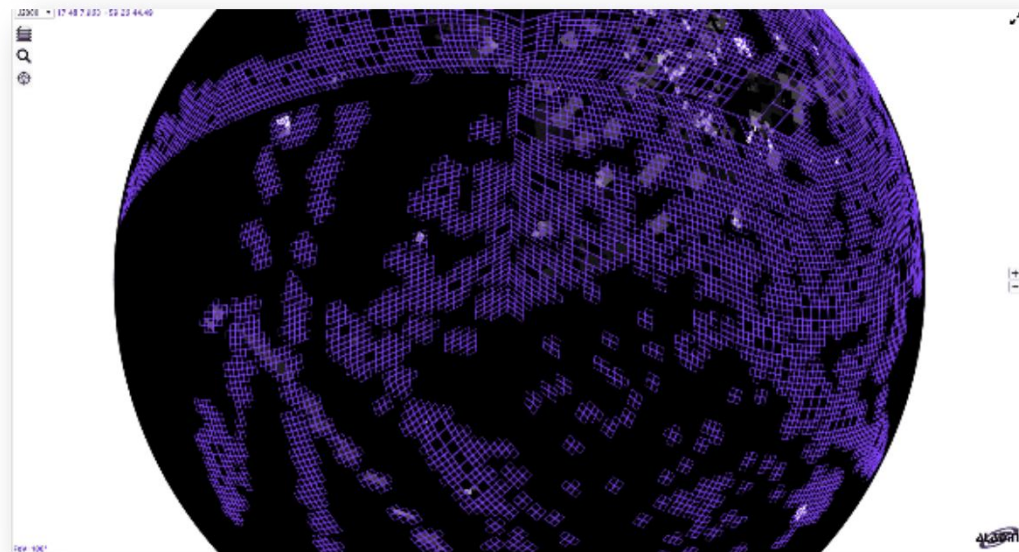
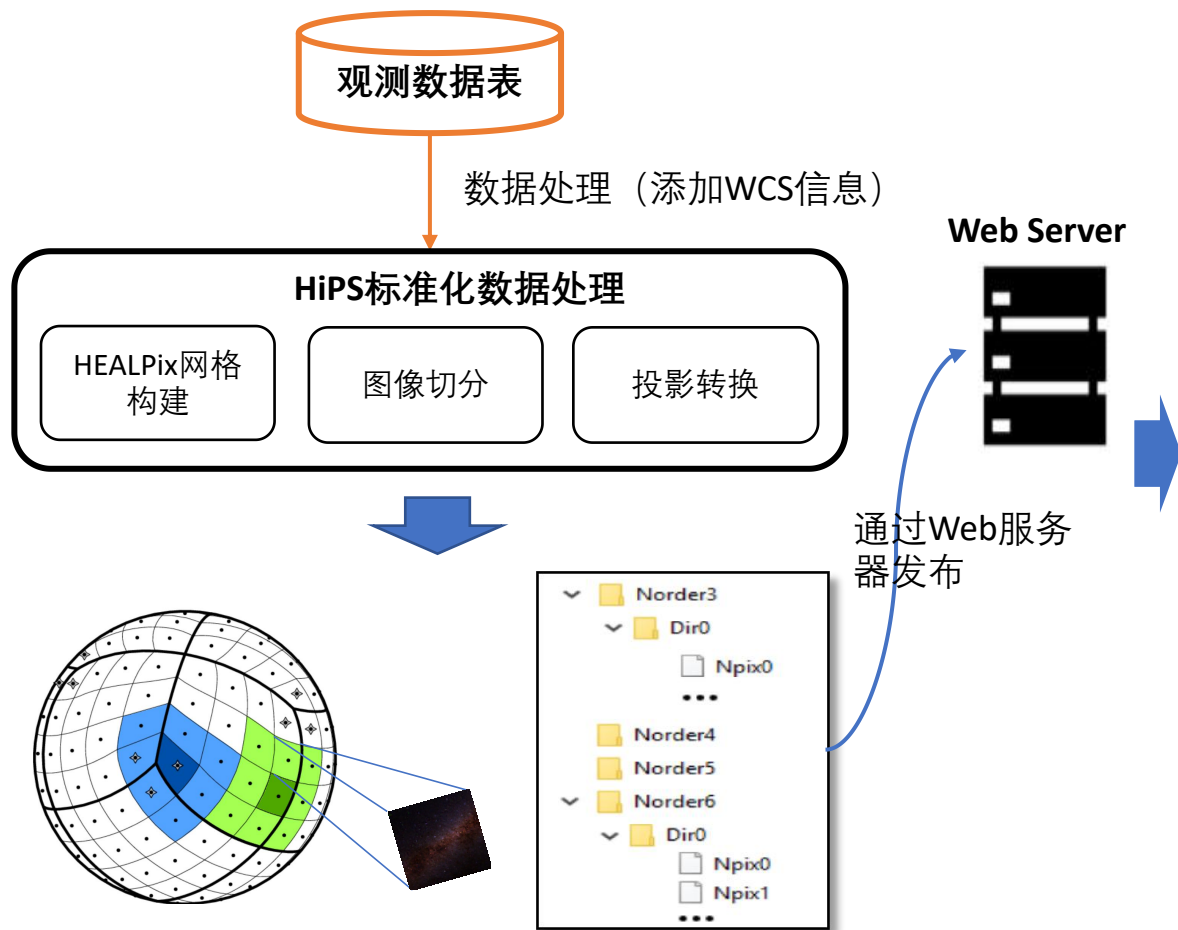
暂现源历史测光数据

## Near Galaxies Info

Pgc	Gwgc Name	Hyperleda Name	Mass Name	Sdssdr12 Name	Ra	Dec	Dist	Dist Err
		SDSSJ003821.27+340131.2			9.58860	34.02530	2594.18000	
				003812.53+340538.1	9.55221	34.09390	19220.10000	



# 3.5 引力波电磁对应体高效搜寻



支持HiPS标准数据集的客户端可浏览观测数据

## 3.5 引力波电磁对应体高效搜寻

**Table 1**  
Telescopes Supported by GWOPS During LVC O3 Run

Telescope	Location	Altitude	FOV	CCD Size	Aperture/m
HMT	88.5772E 45.5224N	1795	60' × 40'	4008 × 2672	0.5
CNEOST	11.46E 32.73N	210.3	3° × 3°	10560 × 10560	1.2/1
Schmidt	117.9640E 40.6508N	950	1.5° × 1.58°	4096 × 4096	0.6/0.9
NOWT	87.1777E 43.4708N	2080	1.3° × 1.3°	4160 × 4136	1
NEXT	87.1777E 43.4708N	2080	20' × 20'	2048 × 2064	0.6
XL216	117.9640E 40.6508N	950	10'' × 10''	2048 × 2048	2.16

- 截止至2020年3月的53个引力波事件中，GWOPS都对事件警报做出了实时的响应，平均在25s之内完成引力波宿主星系候选体的筛选并提交给望远镜进行观测。
- 在此期间，共观测了6133平方度的天区，获取了21,172副观测数据。





# 4.研究总结



## 面向多信使与时域天文观测科学应用

多信使事件电磁对应体高效搜寻

时域天文观测暂现源证认



应用系统构建



关键技术探索

### 海量星表高效检索

- 提出了多层次覆盖天区空间索引方法及基于天球划分的星表分表分区策略
- 方法对大范围天区检索效率有较大提升，综合检索效率也优于同类型方法

### 星表交叉证认及置信度估计

- 将多层次覆盖天区空间索引方法和星表分表分区策略应用至海量多波段星表间的位置匹配
- 针对多波段星表由于空间定位精度不同带来的多对象匹配问题，实现了并行化的贝叶斯推断的星表交叉证认置信度估计方法

### 异构多波段图像组织、检索及可视化

- 基于HiPS标准提出了异构多波段图像数据组织、检索和可视化框架
- 提出了一种并行化的图像数据转换至HiPS 标准数据集的方法，大幅提高数据转换效率
- 在万维望远镜软件平台中实现了HiPS 标准数据集的沉浸式可视化，拓宽了万维望远镜的数据源

# 4.工作总结



## • 发表论文:

1. **Xu, Y.**, Xu, D., Cui, C., et al. (2020). GWOPS: A VO-technology Driven Tool to Search for the Electromagnetic Counterpart of Gravitational Wave Event. Publications of the Astronomical Society of the Pacific, 132(1016), 104501. (SCI)
2. **Xu, Y.**, Cui, C., Fan, D., et al. (2020). IVOA HiPS implementation in the framework of WorldWide Telescope. Astronomy and Computing, 100380. . (SCI)
3. **许允飞**, 樊东卫, 崔辰州,等(2020), 中国虚拟天文台的核心功能需求调查分析. 天文研究与技术, 17(1), pp.111-120. (中文核心)
4. 崔顺, **许允飞**, 苏丽颖, 崔辰州, 樊东卫, 等(2019). 基于卷积神经网络的全天空地基云图分类研究. 天文研究与技术, (2), p.12. (中文核心)
5. 陶一寒, 崔辰州, 张彦霞, **许允飞**, 樊东卫, 等. (2020). 深度学习在天文学中的应用与改进. 天文学进展, 38(2), 168-188. (中文核心)
6. 张磊, 樊东卫, 崔辰州, 何勃亮, **许允飞**, 崔顺, & 王川中. (2019). 海量巡天数据在线可视化技术综述. 天文学进展, (2019年02), 158-177. (中文核心)
7. 樊东卫, 何勃亮, 李长华, 韩军, **许允飞**, 崔辰州. (2019). 球面距离计算方法及精度比较. 天文研究与技术, 16(1):69-76 . (中文核心)
8. Han, J., Wang, C., Fan, D., Cui, C., Li, S., Mi, L., Li, Z., **Xu, Y.**, He, B., Li, C. and Yang, S., (2018). Amateur public observatory I: The observatory and hardware integration system. Astronomy and computing, 25, pp.89-93. (SCI)

## • 申请专利:

1. 发明专利名称: 获取全天相机的自动曝光参数的方法, 状态: 已公开, 发明人: **许允飞**; 崔辰州; 樊东卫; 韩军; 崔顺; 李长华; 李正

# 5 未来展望



- 在应用实践中对本文提出的方法不断进行改进及完善
  1. 海量星表空间索引的物理组织方式可进一步优化，减少检索的对比次数；
  2. 多波段星表交叉认证的置信度估计可结合实际应用引入更多参量，进一步提高交叉认证的准确率；
  3. 异构多波段图像数据转化至HiPS数据集会出现全天图像亮度不一致问题，需要进一步解决。
- 进一步凝练关键技术，结合IVOA针对天文数据可获取性的技术标准，以国家天文科学数据中心为基础平台提供面向天文科研工作者的多波段多信使数据融合在线服务
- 开放数据融合相应的第三方开发接口，吸引开发人员构建更多应用，以实现资源与技术向服务的快速转换
- 随着EP的发射和引力波探测的深入，对论文研究内容在真实科学应用中进行检验和完善

**敬请批评指正!**