BESIII离线软件系统

邓子艳

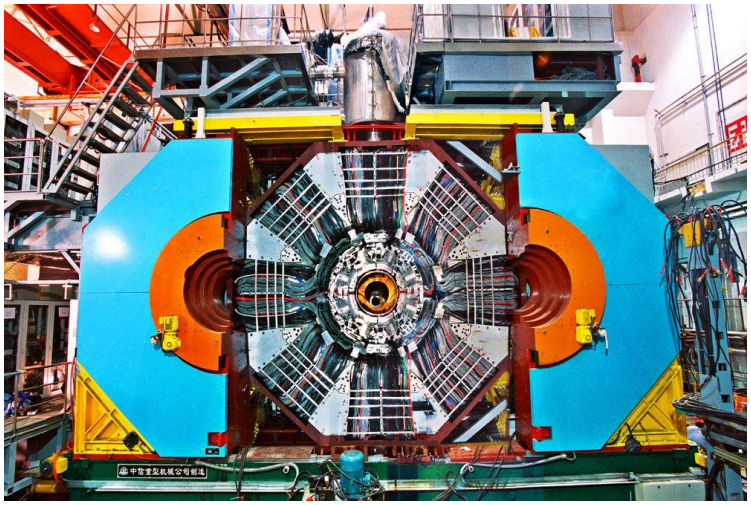
中国科学院高能物理研究所　实验物理中心

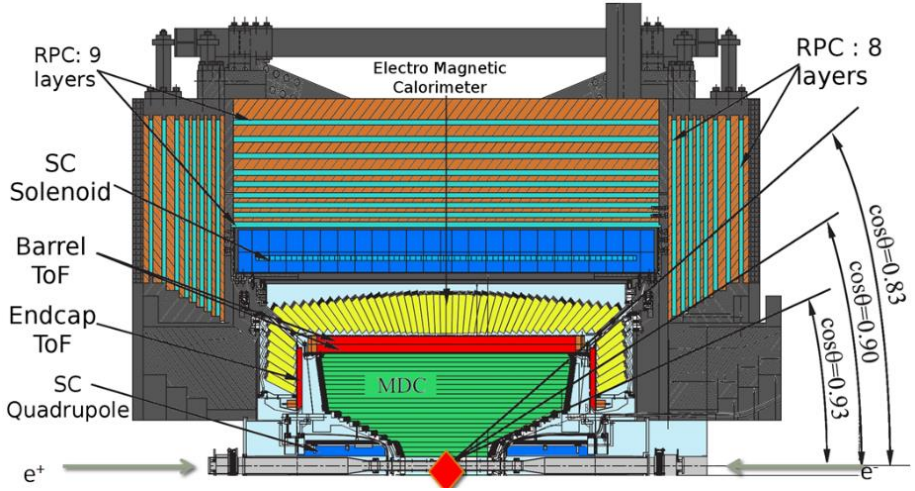
第一届高能物理计算用户研讨会　2024年5月　成都

内容提要

- BESIII实验
- BESIII数据处理流程
- BESIII离线软件系统
- BESIII计算资源使用情况

BESIII实验





北京谱仪III(BESIII)是北京正负电子对撞机重大改造工程BEPCII中的大型粒子探测器，是目前国际上唯一运行在τ-粲能区的正负电子对撞实验

由主漂移室(MDC)、飞行时间计数器(TOF)、电磁量能器(EMC)、缪子计数器(MUC)、超导磁铁和相应的电子学读出、触发、数据获取等系统组成

MDC： 丝分辨 115um，dE/dx分辨<5%(Bhabha)

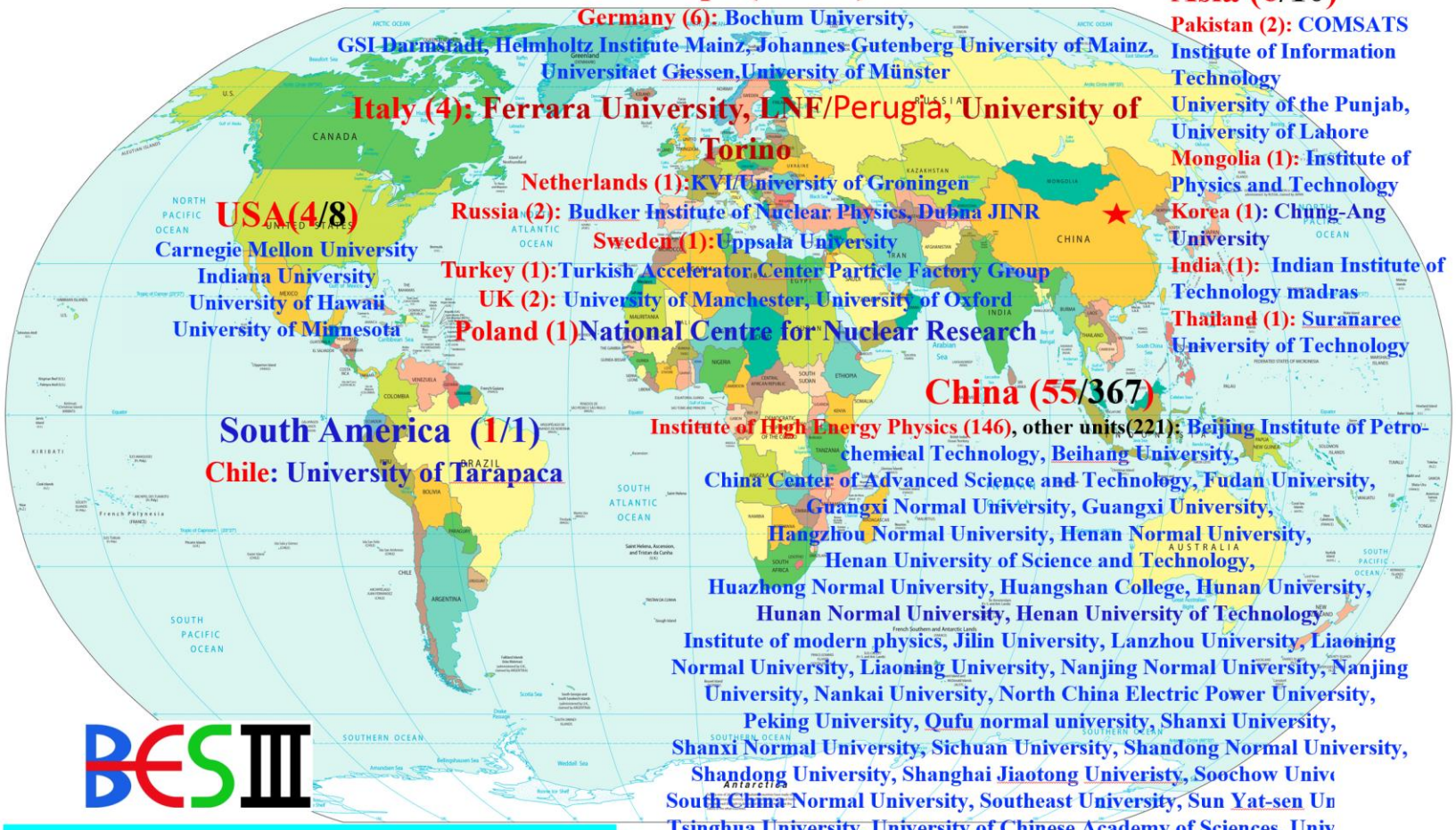
TOF： 桶部时间分辨68ps，端盖闪烁体98ps，端盖MRPC 60ps

EMC： 能量分辨2.3%

MUC： 效率~96%，噪声<0.04 Hz/cm^2(桶部)， <0.1 Hz/cm^2(端盖)

对撞机中的正负电子束团在谱仪中心对撞产生的末态粒子信息由谱仪记录，经过离线数据处理后进行陶粲能区的物理研究

BESIII合作组



Europe (17/115)

Germany (6): Bochum University, GSI Darmstadt, Helmholtz Institute Mainz, Johannes Gutenberg University of Mainz, Universitaet Giessen, University of Münster

Italy (4): Ferrara University, LNF/Perugia, University of Torino

Netherlands (1): KVI/University of Groningen

Russia (2): Budker Institute of Nuclear Physics, Dubna JINR

Sweden (1): Uppsala University

Turkey (1): Turkish Accelerator Center Particle Factory Group

UK (2): University of Manchester, University of Oxford

Poland (1) National Centre for Nuclear Research

Asia (6/10)

Pakistan (2): COMSATS Institute of Information Technology
University of the Punjab, University of Lahore

Mongolia (1): Institute of Physics and Technology

Korea (1): Chung-Ang University

India (1): Indian Institute of Technology madras

Thailand (1): Suranaree University of Technology

USA(4/8)

Carnegie Mellon University
Indiana University
University of Hawaii
University of Minnesota

China (55/367)

Institute of High Energy Physics (146), other units(221): Beijing Institute of Petro-chemical Technology, Beihang University, China Center of Advanced Science and Technology, Fudan University, Guangxi Normal University, Guangxi University, Hangzhou Normal University, Henan Normal University, Henan University of Science and Technology, Huazhong Normal University, Huangshan College, Hunan University, Hunan Normal University, Henan University of Technology, Institute of modern physics, Jilin University, Lanzhou University, Liaoning Normal University, Liaoning University, Nanjing Normal University, Nanjing University, Nankai University, North China Electric Power University, Peking University, Qufu normal university, Shanxi University, Shanxi Normal University, Sichuan University, Shandong Normal University, Shandong University, Shanghai Jiaotong Univeristy, Soochow Univ, South China Normal University, Southeast University, Sun Yat-sen Un, Tsinghua University, University of Chinese Academy of Sciences, Univ Jinan, University of Science and Technology of China, University of Science and Technology Liaoning, University of South China, Wuhan University, Xinyang Normal University, Zhejiang University, Zhengzhou University, YunNan University , China University of Geosciences , Yantai University,

South America (1/1)

Chile: University of Tarapaca

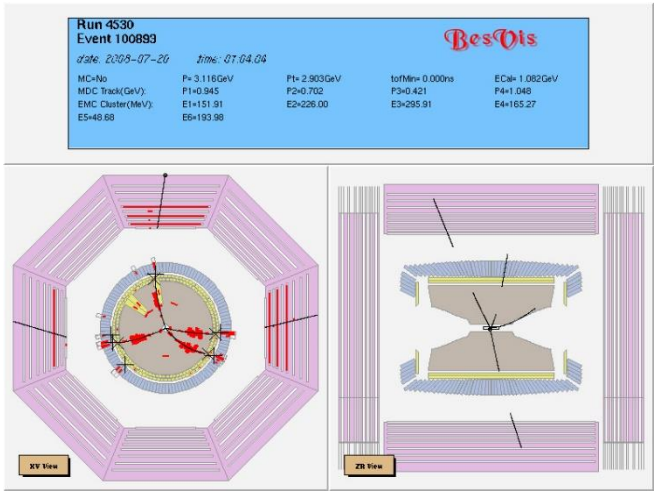
BESIII

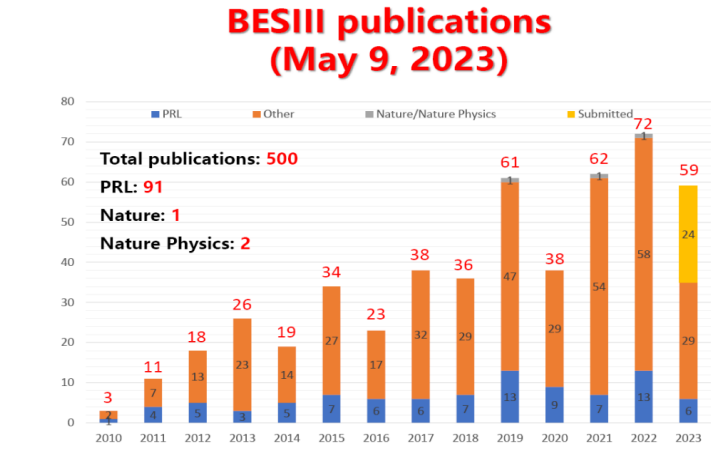
~600 members
From 83 institutions in 16 countries

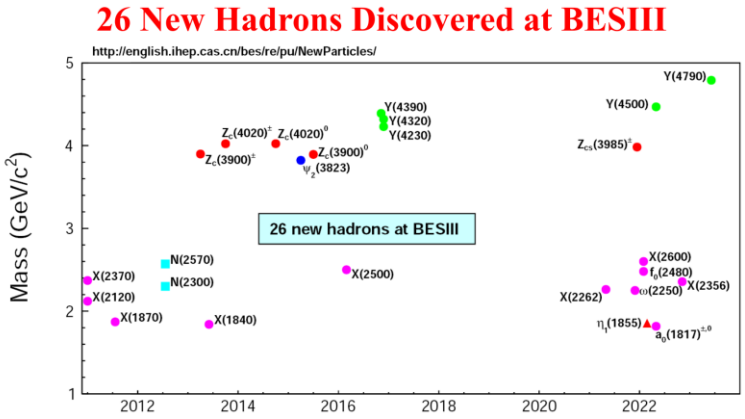
BESIII物理

- BESIII物理
  - Charm physics
  - Charmonium decays
  - Light hadron
  - Tau & R QCD
  - New Physics

First Event in BESIII: 2008-7-19

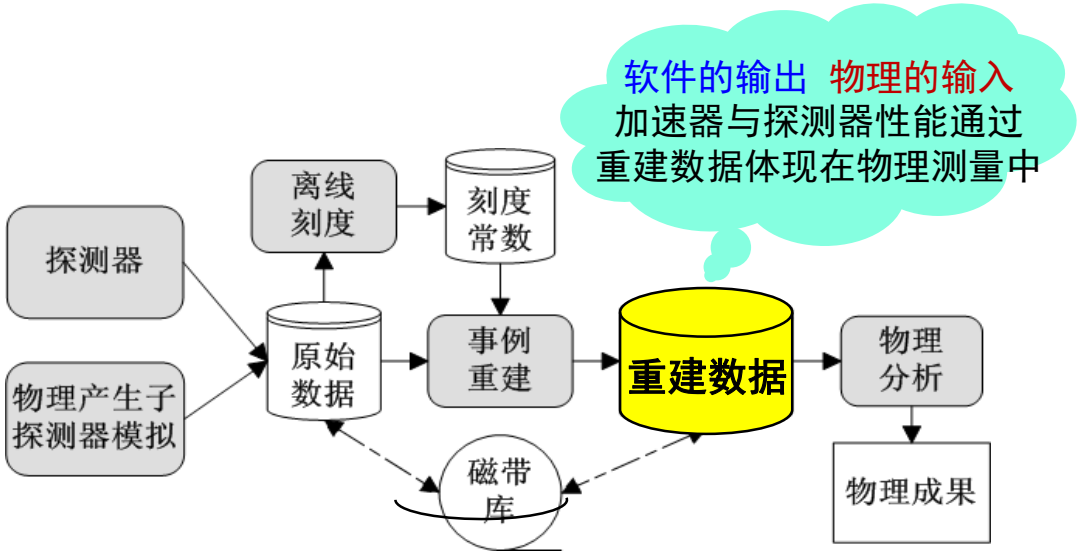






海量的优质数据和前沿物理研究的开展形成了重大物理发现的基础，取得了一系列具有重要物理意义和广泛国际影响的成果

BESIII数据处理流程

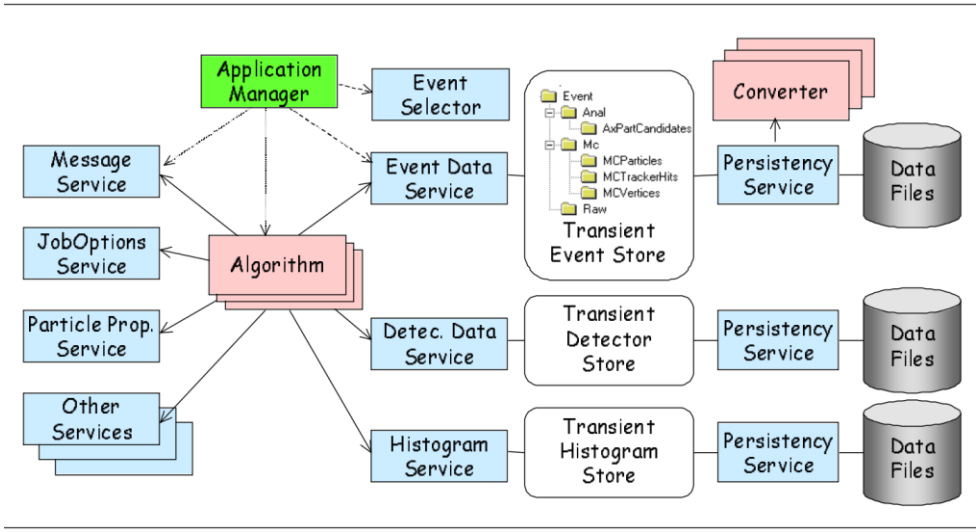


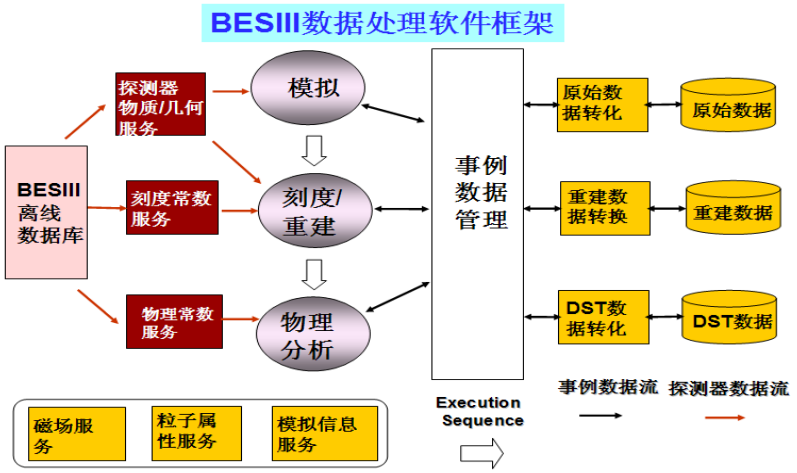
- 软件框架提供有效的数据管理工具，不同软件模块的组合和动态库的链接机制
- 通过刻度与重建，压低各种条件的影响和排除噪声本底，最大限度挖掘加速器和探测器性能
- 实现精确的探测器模拟，为事例选择效率计算、选择条件优化和本底估计等提供可靠依据
- 利用优秀数理方法联合各种实验信息开发物理分析工具软件，进一步提高实验精度

离线软件主要研究内容

1. 物理产生子
2. 探测器模拟
3. 事例重建
4. 探测器刻度
5. 物理分析工具软件
6. 软件框架
7. 数据管理
8. 数据库服务
9. 可视化

BESIII离线软件系统（BOSS）

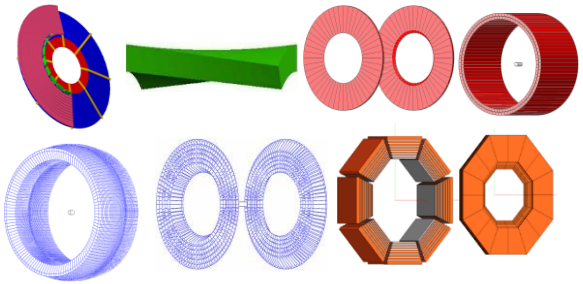


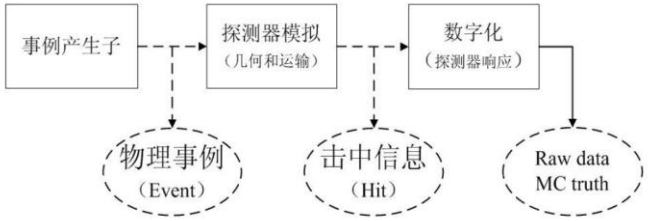


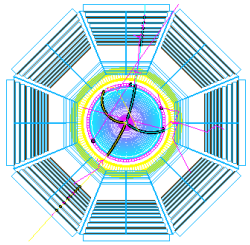
- 基于通用底层框架GAUDI框架开发BESIII离线软件系统，粒子物理实验国际通用软件库和工具软件的支持：Geant4，ROOT，MySQL，CERNLIB，CLHEP，……
- 算法(Algorithm)与服务(Service)分开，算法与数据分开，提供简单易用、安全可靠、服务齐全的数据处理环境
- 模拟、刻度、重建和物理分析算法是数据处理和物理分析的核心
- 成功处理BESIII实验运行以来获取的所有实验数据，支持了基于软件和计算平台的物理分析研究

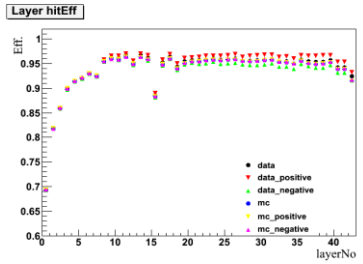
探测器模拟

- 在物理课题研究中，产生数倍于真实数据的模拟样本用于：事例选择效率计算、选择条件的优化、本底污染水平估计和物理结果的系统误差分析
- 模拟与数据的不一致性是系统误差的最主要来源，物理测量结果中有公共的系统误差，如带电径迹重建效率、粒子鉴别效率和光子重建效率等。
- BESIII探测器模拟软件（BOOST）
  - 探测器的几何与物质描述
  - 粒子在探测器中的传输和相互作用
  - 探测器响应机制
  - 探测器运行的真实化模拟
- 目前带电径迹重建效率和粒子鉴别效率的系统误差 1~2%的水平，与国际其他粒子物理实验修正前误差水平相当
- 探测器的响应机制和实验条件及环境的变化过于复杂。在当前精度下，进一步精细调试探测器模拟的研究极具挑战









探测器刻度

- 探测器固有性能和刻度软件水平是决定物理测量结果精度/信号显著性的最主要贡献之一
- 通过刻度算法把探测器的优良的空间分辨，时间分辨等转化为物理分析中的动量/能量分辨，粒子鉴别能力等性能指标
- 各个子探测器性能达到或超过设计指标；达到同类型探测器国际先进水平，部分指标处于国际领先
- 研究目标已经实现
- 探索利用先进数理方法，如深度学习等进一步优化探测器性能

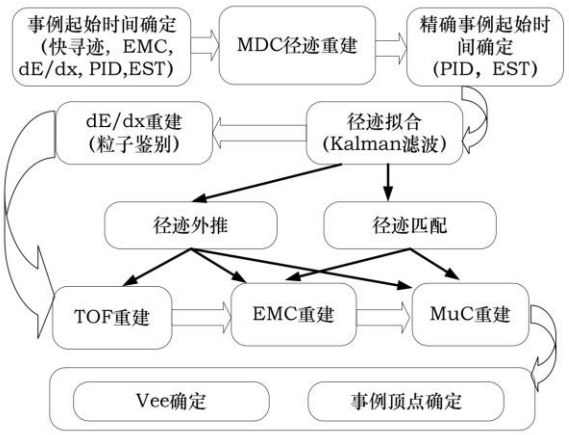
Exps.	MDC Spatial resolution	MDC dE/dx resolution	EMC Energy resolution
CLEOc	110 μm	5%	2.2-2.4 %
Babar	125 μm	7%	2.67 %
Belle	130 μm	5.6%	2.2 %
BESIII	115 μm	<5% (Bhabha)	2.4%

Exps.	TOF Time resolution
CDFII	100 ps
Belle	90 ps
BESIII	68 ps (BTOF) 60 ps (ETOF)

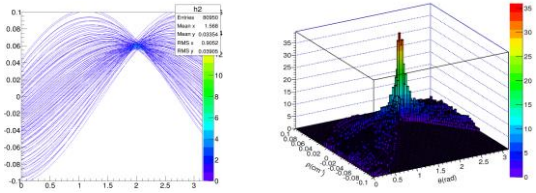
MUC: Efficiency ~ 96%
BG level:
< 0.04 Hz/cm^2(B-MUC),
< 0.1 Hz/cm^2(E-MUC)

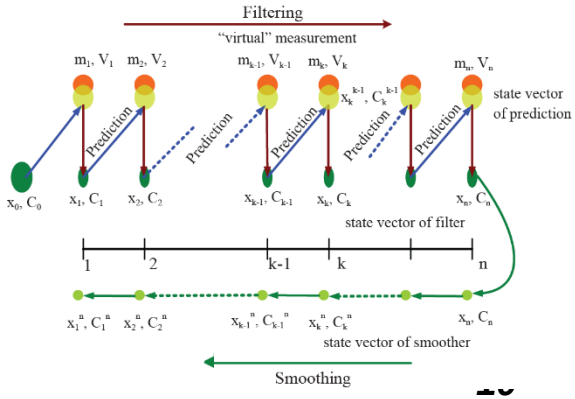
事例重建

- 事例重建软件将原始数据中记录的电子学信号转化为粒子的动量、能量等物理量，生成重建数据，供物理课题研究使用
- BESIII事例重建：
  - 快寻迹和事例起始时间
  - 漂移室带电径迹重建
  - Kalman滤波径迹拟合
  - dE/dx重建
  - EMC、TOF、MUC 重建和径迹外推与匹配
- 开发和应用基于模式匹配、共形变换、霍夫变换的径迹寻找算法，采用卡尔曼（Kalman）滤波的径迹拟合算法，实现精确计算带电径迹参数和误差矩阵的目标
- 加速器和探测器的噪声本底是限制重建效率和精度的最关键因素，探索利用探测器自身特点和信号噪声特征，调试与优化噪声本底排除机制是未来主要研究方向之一

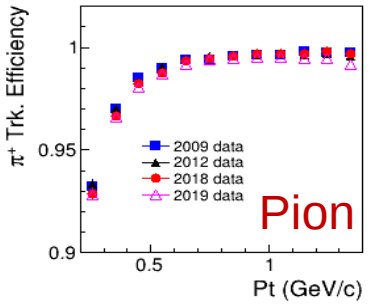


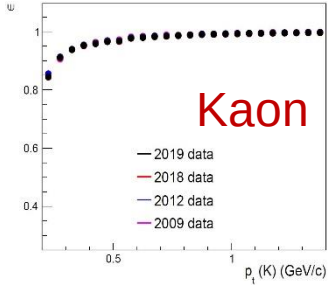
BESIII事例重建流程

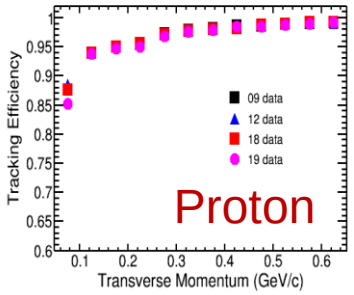




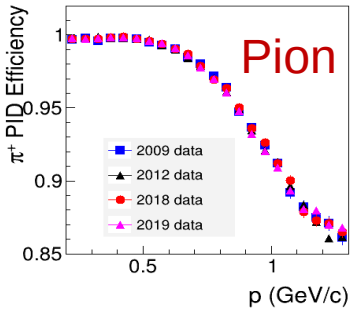
重建效率

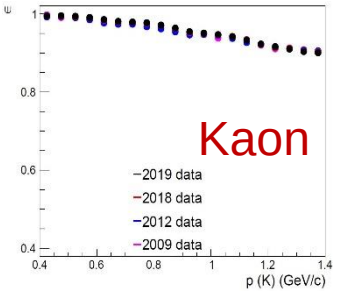


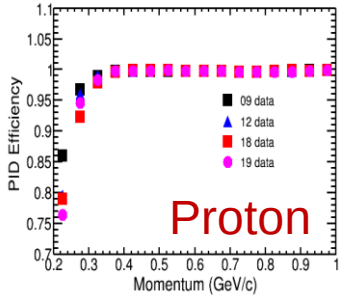




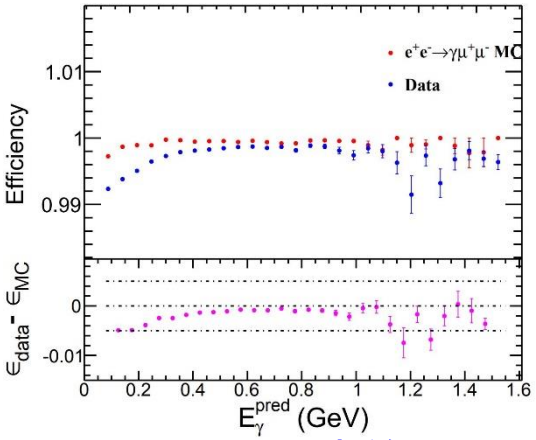
带电径迹重建效率



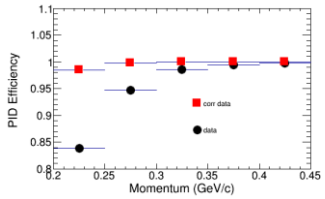




粒子鉴别效率



光子重建效率

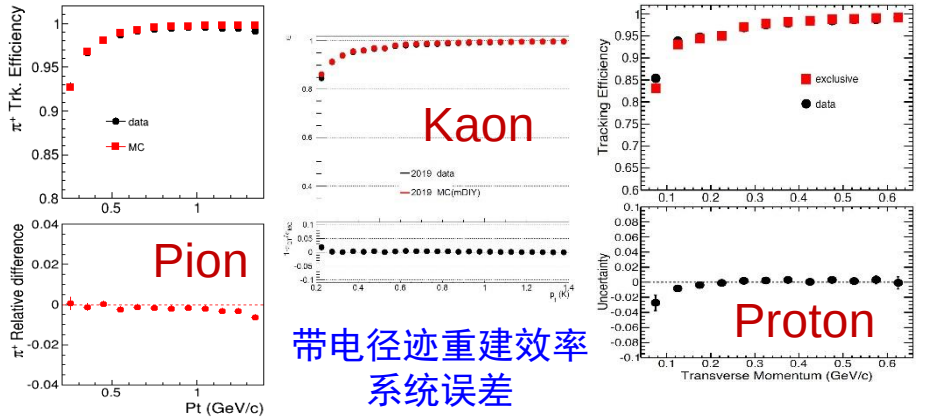


低动量质子修正

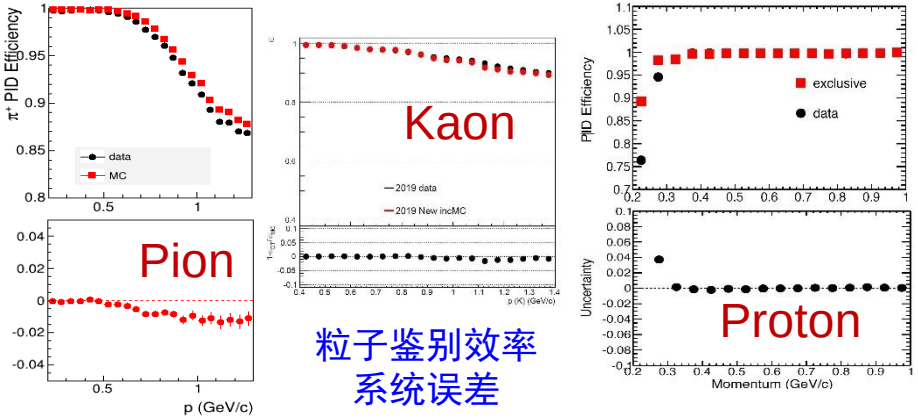
- 高横动量带电径迹重建效率接近100%；提高低横动量（<200MeV）径迹重建效率，压低假径迹比例可提高事例选择效率，对粲重子和粲介子研究十分重要
- 光子重建效率接近100%
- 粒子鉴别效率实现探测器设计要求；通过低动量质子修正获得正确粒子鉴别效率；尝试利用先进的机器学习/深度学习技术提高粒子鉴别效率

系统误差

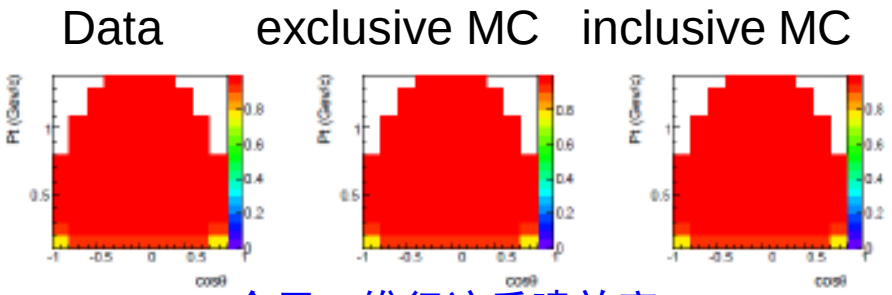
- BESIII已经获取和即将获取的前所未有的高统计量实验数据大幅降低了物理测量结果的统计误差。进一步降低系统误差达到与统计误差相匹配的需求十分迫切
- 发挥BESIII实验高统计量的优势，获得随粒子种类、电荷、横动量和出射角度依赖的效率修正因子，实现系统误差 <0.5%的目标。



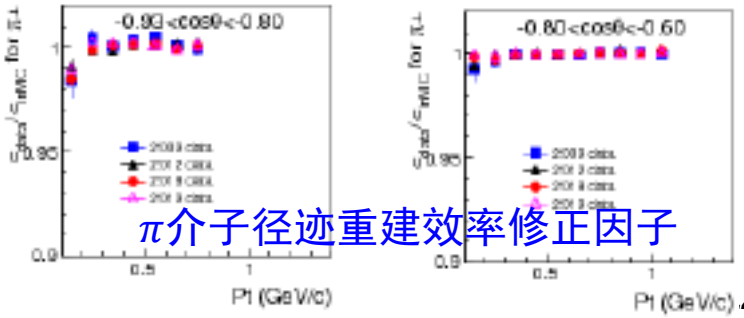
带电径迹重建效率
系统误差



粒子鉴别效率
系统误差



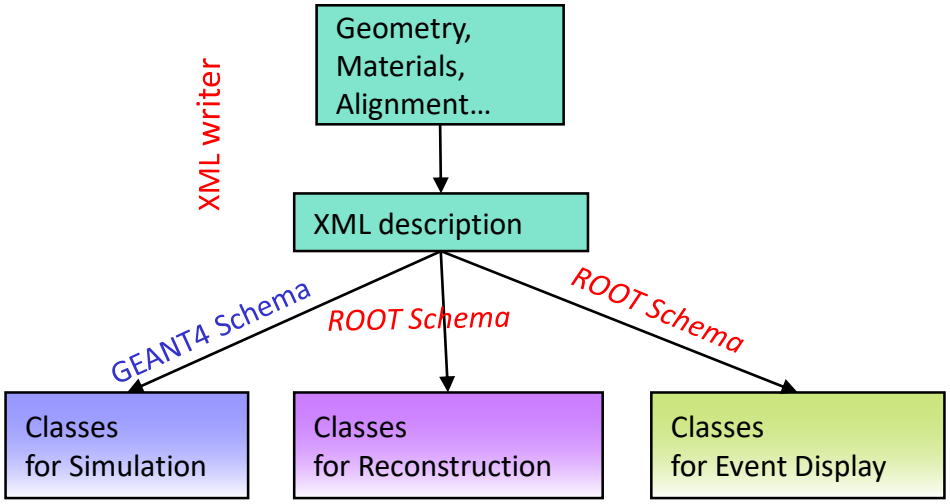
π介子二维径迹重建效率

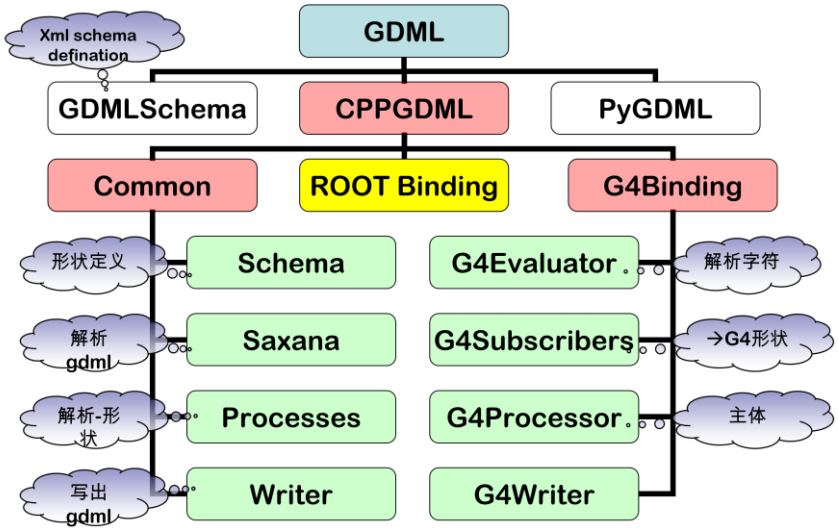


π介子径迹重建效率修正因子

探测器几何描述

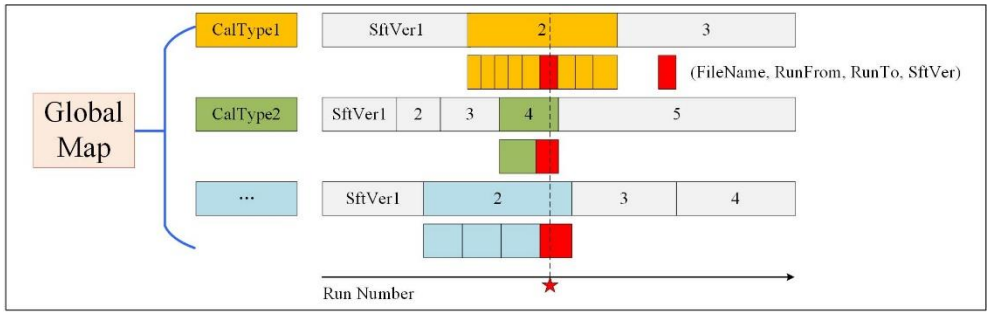
- 基于GDML (Geometry Description Markup Language), 由Geant4组开发
- 针对BESIII探测器拓展了GEANT4 Schema并开发了ROOT Schema
- GDML文件用于探测器模拟、事例重建和事例显示，保证软件系统中几何的一致性

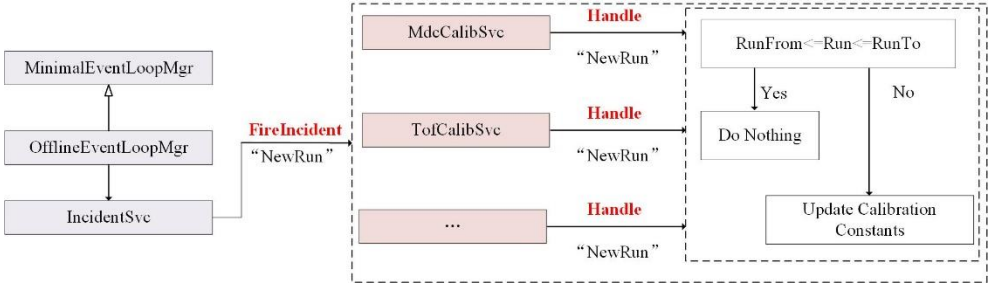




刻度框架

- 利用Gaudi的Incident机制
  - 每个刻度常数对应一段run号范围，超出范围则触发Incident，自动切换刻度常数
- GlobalMap机制
  - 每个软件版本对应一套刻度常数的快照



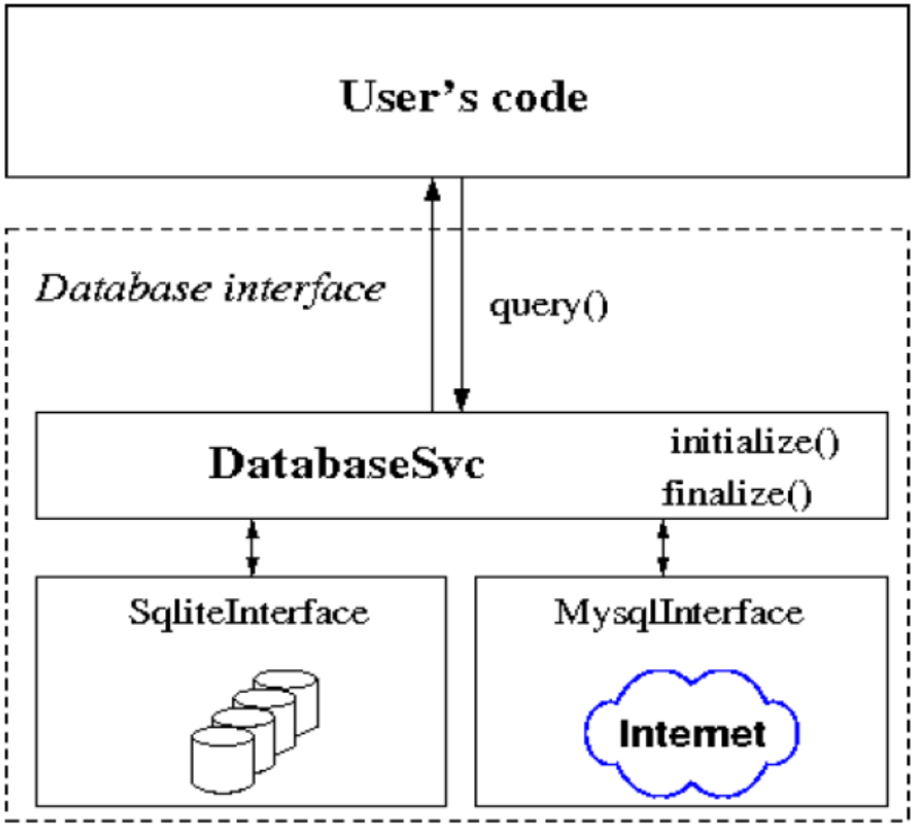


BossRelease	DataType	RunFrom	RunTo	SftVer
6.6.5	Muc	4989	10878	6.6.1
6.6.5	Muc	11414	29676	6.6.2
6.6.5	Muc	29677	31257	6.6.3
6.6.5	Muc	31258	33772	6.6.4
6.6.5	Muc	33999	38140	6.6.4.p01
6.6.5	Muc	39355	80000	6.6.5

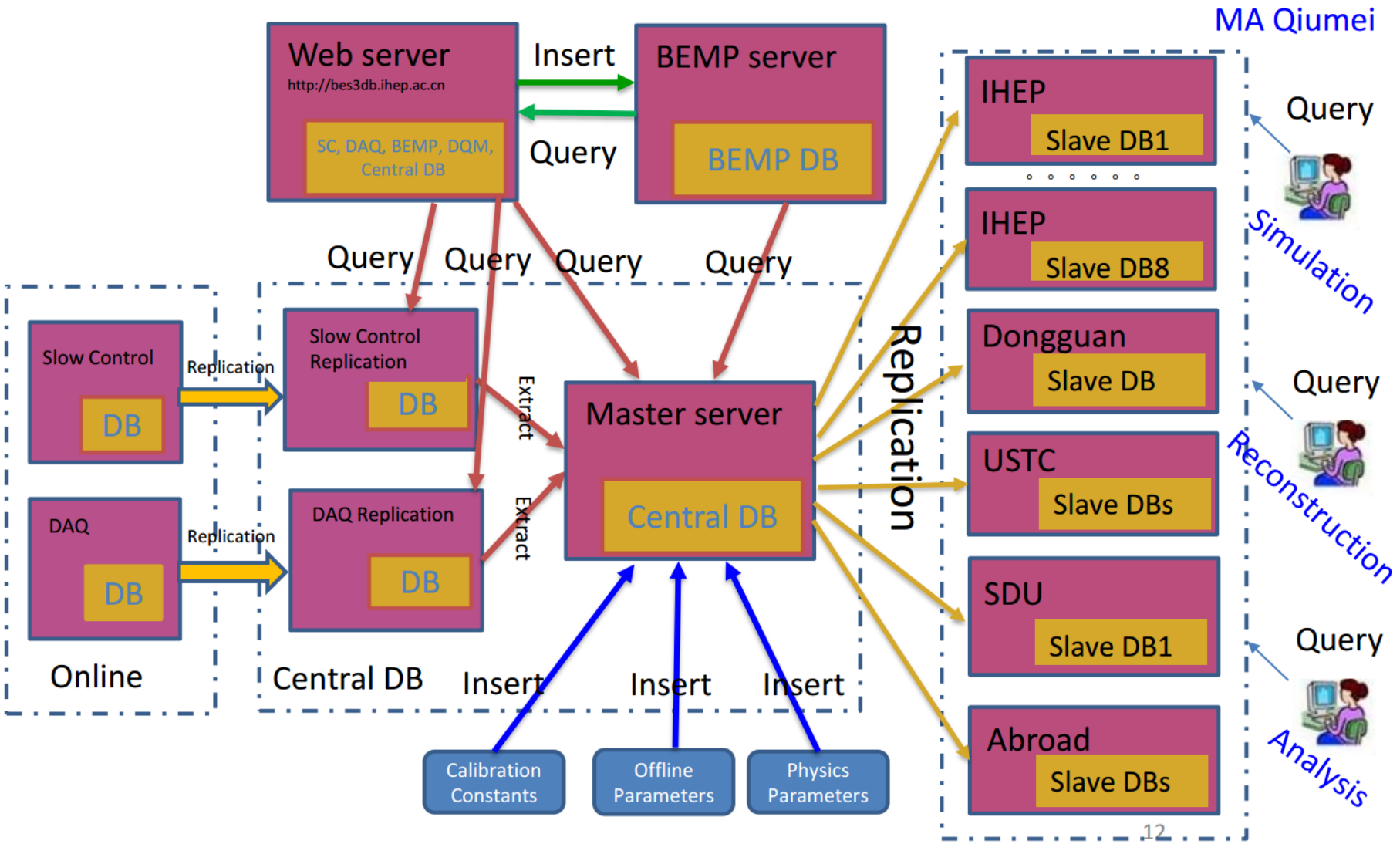
每个软件版本一套刻度常数，数据处理可重复
发布新版本时避免对旧刻度常数不必要的拷贝
降低出错几率，便于版本管理

数据库访问

- 整个离线软件系统内通过统一的接口访问数据库
- DatabaseSvc支持两种后端
  - Sqlite
  - Mysql
- 支持两种访问模式
  - One connection per job
  - One connection per query



数据库管理


MA Qiumei
Web server
http://bes3db.ihep.ac.cn
SC, DAQ, BEMP, DQM, Central DB
Insert
Query
BEMP server
BEMP DB
Query
Query
Query
Query
Slow Control
DB
Replication
Slow Control Replication
DB
DAQ
DB
Replication
DAQ Replication
DB
Extract
Extract
Master server
Central DB
Online
Central DB
Insert
Insert
Insert
Calibration Constants
Offline Parameters
Physics Parameters
Replication
IHEP
Slave DB1
IHEP
Slave DB8
Dongguan
Slave DB
USTC
Slave DBs
SDU
Slave DB1
Abroad
Slave DBs
12
Query
Simulation
Query
Reconstruction
Query
Analysis

Bookkeeping数据管理系统

创建数据集

Name Description

Return

Search

Data Type

Raw Rec Dst

Date From Date To

Run From Run To

Run In

BOSS Version

Event Type

Full Data Bhabha Dimuon Diphoton Random trigger

Search

编号	名称	备注
3637	230606	78231-78248
3636	230605	78201-78230
3635	230604	78192-78200
3634	230603	78171-78191
3633	230602	78154-78170
3632	230601	78128-78152
3631	230530	78094-78112
3630	230531	78113-78127
3628	230529	78075-78093
3627	230528	78054-78074
3626	230527	78034-78053
3625	230526	78016-78033

浏览数据集

数据集名称

230606

备注

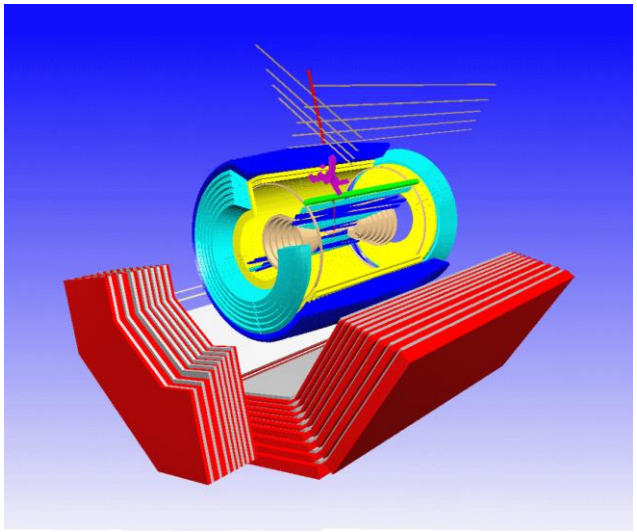
78231-78248

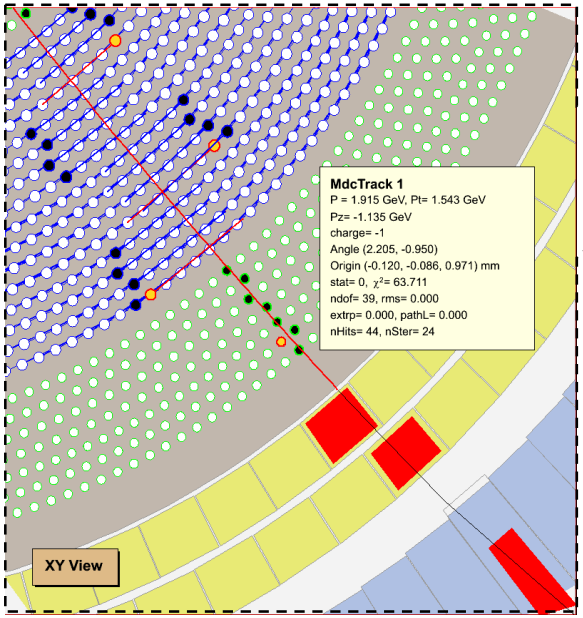
返回

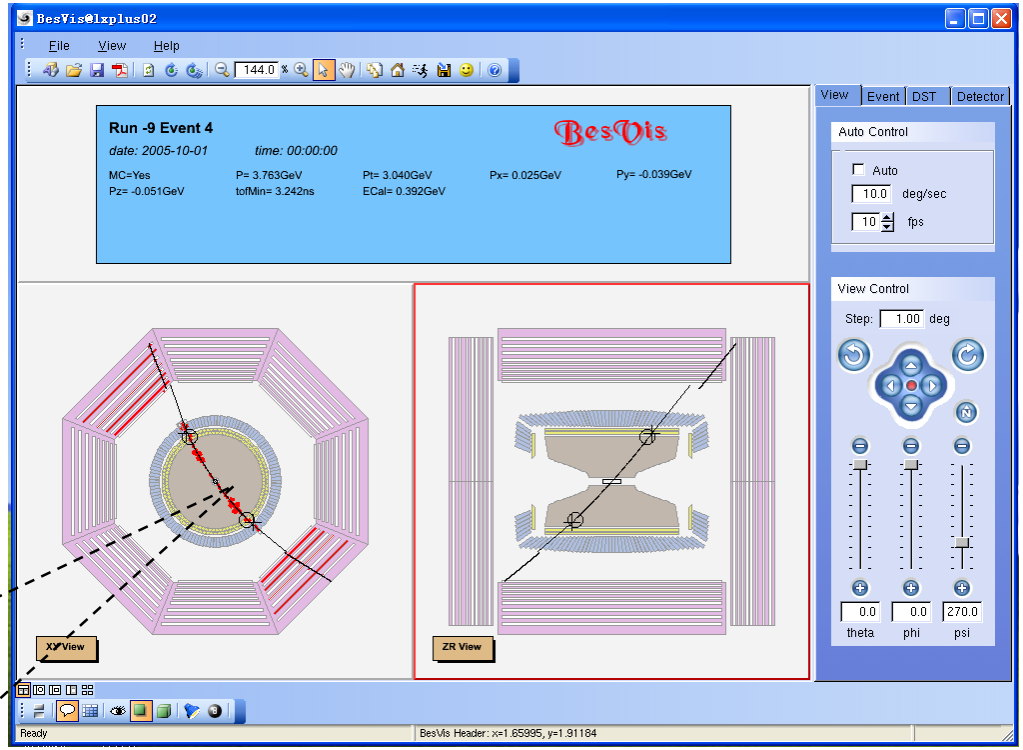
共找到1,103条。 [首页/上一页] 1, 2, 3, 4, 5, 6, 7, 8[下一页/尾页]

编号	运行号	文件名称	文件类型	创建时间	文件格式	事例数	状态	事例类型
5629417	78248	run_0078248_RandomTrg_file002_SFO-1.raw	Raw	2023-06-06	raw	121124	closed	random
5629416	78248	run_0078248_RandomTrg_file001_SFO-1.raw	Raw	2023-06-06	raw	130551	closed	random
5629415	78248	run_0078248_GEBhabha_file005_SFO-1.raw	Raw	2023-06-06	raw	84582	closed	bhabha
5629414	78248	run_0078248_GEBhabha_file004_SFO-1.raw	Raw	2023-06-06	raw	127957	closed	bhabha
5629413	78248	run_0078248_GEBhabha_file003_SFO-1.raw	Raw	2023-06-06	raw	124945	closed	bhabha
5629412	78248	run_0078248_GEBhabha_file002_SFO-1.raw	Raw	2023-06-06	raw	122587	closed	bhabha
5629411	78248	run_0078248_GEBhabha_file001_SFO-1.raw	Raw	2023-06-06	raw	119851	closed	bhabha
5629410	78248	run_0078248_GDiphoton_file001_SFO-1.raw	Raw	2023-06-06	raw	29927	closed	diphoton
5629409	78248	run_0078248_GDimuon_file001_SFO-1.raw	Raw	2023-06-06	raw	10493	closed	dimu
5629408	78248	run_0078248_GBBhabha_file003_SFO-1.raw	Raw	2023-06-06	raw	79670	closed	bhabha
5629407	78248	run_0078248_GBBhabha_file002_SFO-1.raw	Raw	2023-06-06	raw	123999	closed	bhabha
5629406	78248	run_0078248_GBBhabha_file001_SFO-1.raw	Raw	2023-06-06	raw	119711	closed	bhabha
5629405	78248	run_0078248_BBhabha_file001_SFO-1.raw	Raw	2023-06-06	raw	17187	closed	bhabha
5629404	78248	run_0078248_All_file049_SFO-1.raw	Raw	2023-06-06	raw	24680	closed	full
5629403	78248	run_0078248_All_file048_SFO-1.raw	Raw	2023-06-06	raw	130587	closed	full
5629402	78248	run_0078248_All_file047_SFO-1.raw	Raw	2023-06-06	raw	130837	closed	full
5629401	78248	run_0078248_All_file046_SFO-1.raw	Raw	2023-06-06	raw	130082	closed	full
5629400	78248	run_0078248_All_file045_SFO-1.raw	Raw	2023-06-06	raw	128703	closed	full
5629399	78248	run_0078248_All_file044_SFO-1.raw	Raw	2023-06-06	raw	128752	closed	full
5629398	78248	run_0078248_All_file043_SFO-1.raw	Raw	2023-06-06	raw	128229	closed	full

事例显示





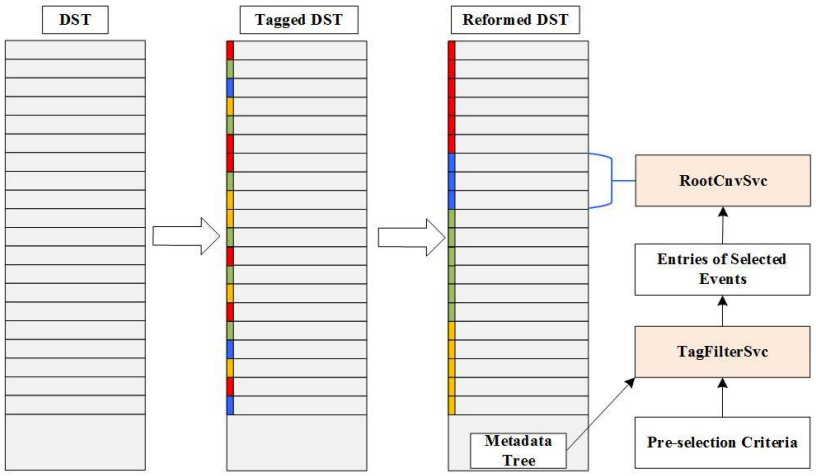


- 基于ROOT，OpenGL， XML
- 支持2D和3D显示
- 通过menu和toolbar进行控制操作

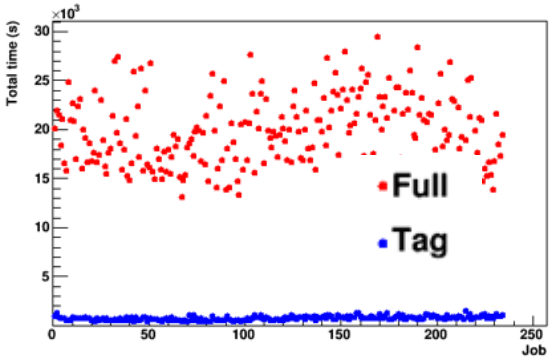
基于事例标记的数据分析框架

- 随着数据量增大和用户增多，计算资源面临严峻挑战
- 对特定衰变过程的事例选择，重建数据文件中的绝大部分事例通过简单选择条件就可以排除，如带电径迹数目，粲介子或粲重子的标记等
- 新分析框架，对事例做标记(TAG)，支持基于TAG的事例预筛选，可在全合作组内广泛使用

基于TAG的事例预筛选
大大降低读入的事例数



两种模式作业运行时间对比



sel: selected events
ign: ignored events

$N_{total} = N_{sel} + N_{ign}$

$N_{ign} = R_{Nevent} \times N_{sel}$

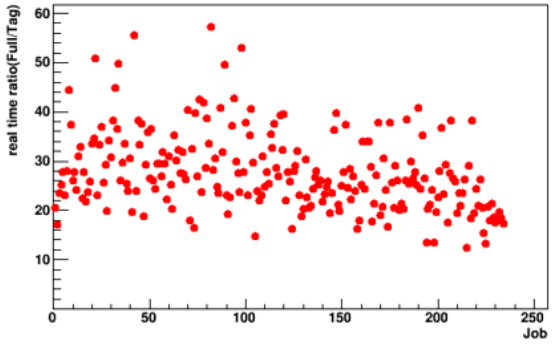
$t_{ign} = R_{time} \times t_{sel}$

$T_{total} = N_{ign} \times t_{ign} + N_{sel} \times t_{sel}$

$T_{sel} = N_{sel} \times t_{sel}$

$$R_{SpeedUp} = \frac{T_{total}}{T_{sel}} = 1 + R_{Nevent} \times R_{time}$$

分析速度提高倍数



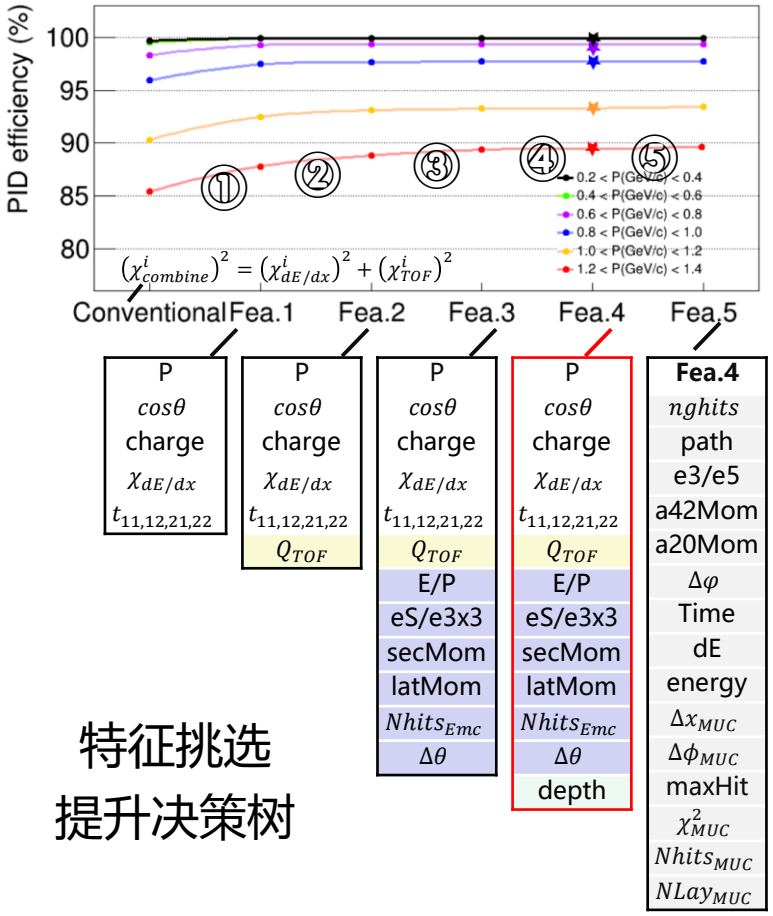
$R_{Nevent} = 194$

$R_{time} = 0.125$

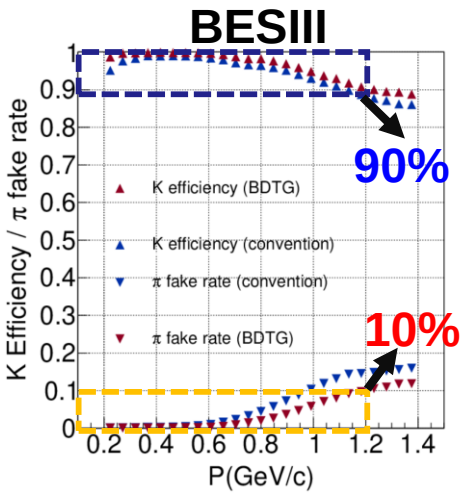
$R_{SpeedUp} = 25$

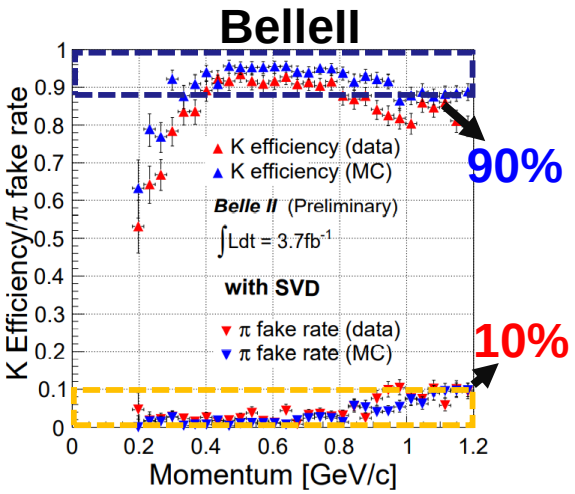
提高分析速度，减轻磁盘压力

机器学习方法提高粒子鉴别能力



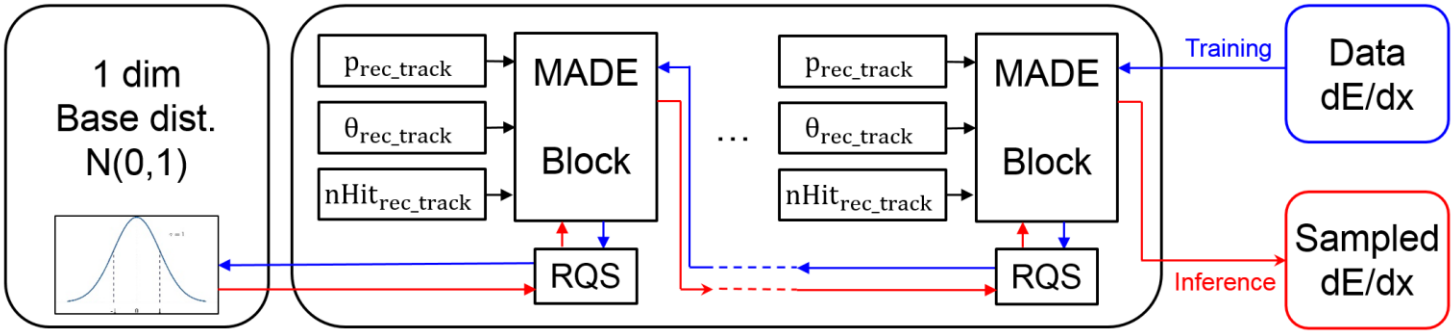
特征挑选
提升决策树



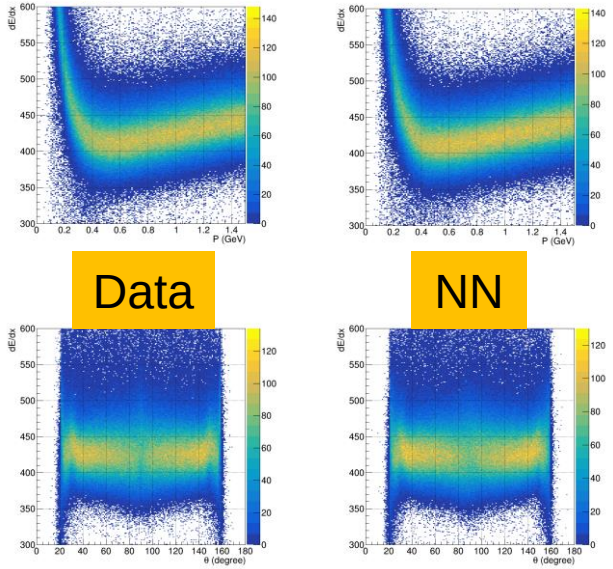


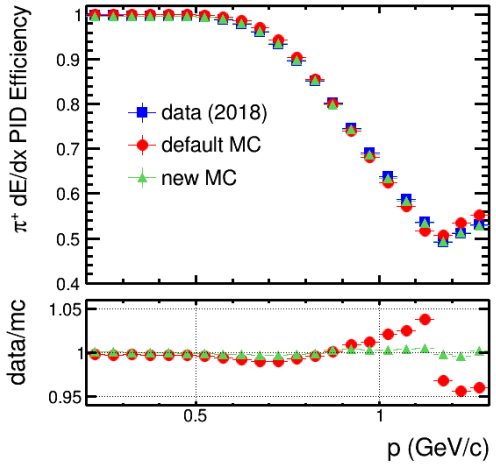
- 与传统方法相比
  - π介子提升了~8%@1.4GeV/c
  - K介子提升了~3%@1.4GeV/c
- 验证了机器学习方法对强子样本没有依赖
- 新的方法对π介子和K介子鉴别的系统误差~1%水平
- 完成相关软件部署

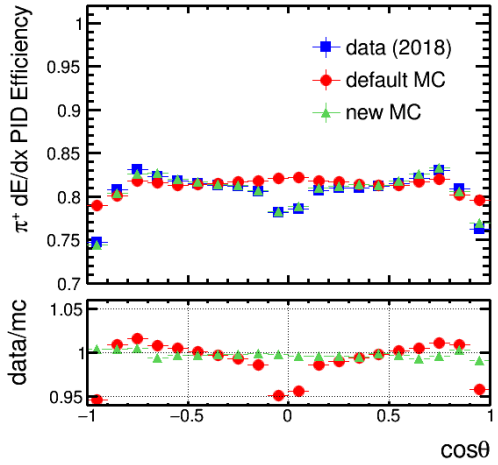
机器学习方法提高模拟精度



- 利用 Normalizing flow方法实现dE/dx的精确模拟
- 粒子鉴别效率在较大动量区间内达到~1%的水平

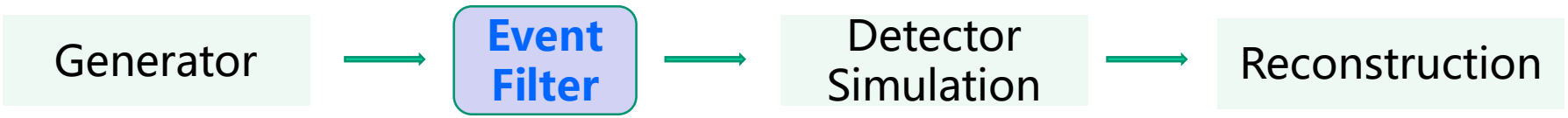






用户定制数据处理流程

(1)产生子级别的Filter，满足条件的HepMC事例才进行探测器模拟



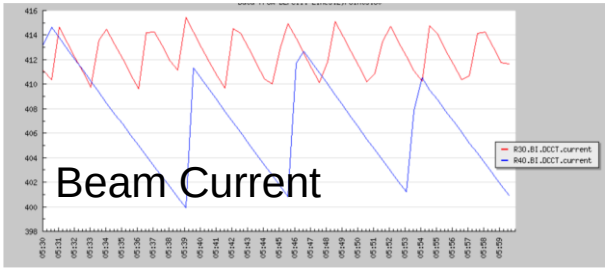
(2)Geant4级别的Filter，Geant4模拟过程中满足某些条件的事例才继续探测器模拟

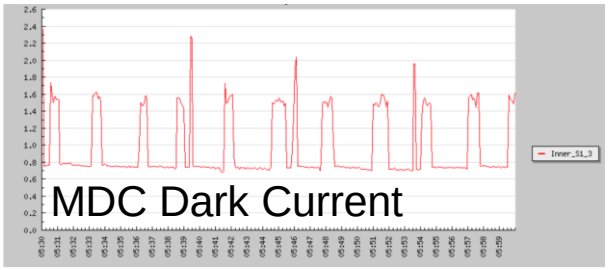
例如：模拟 J/psi->p+ pi- anti-n0 时，只保存 “anti-n0与beam pipe发生非弹性散射”，
50万事例中仅模拟31个事例，节省大量磁盘空间和CPU时间

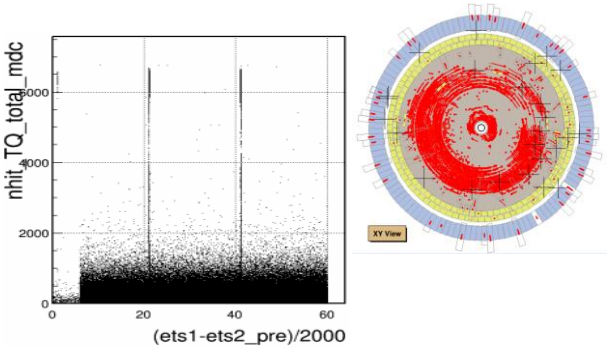
(3)Raw data进行Filter，某些事例不进行重建，例如高噪声事例等等

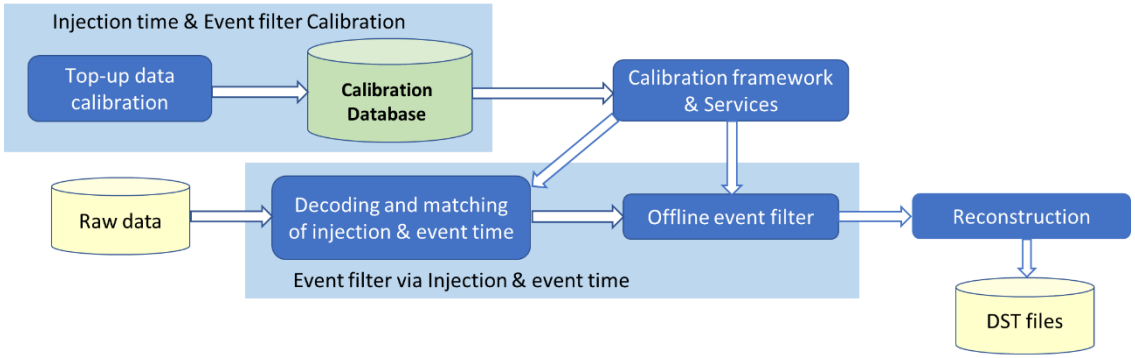
离线事例过滤机制

- BEPCII先进的恒流注入模式（top-up）有效提高30%积分亮度
- 束流注入储存环后的短时间内噪声本底较高，数据质量无法满足数据分析要求
- 在线触发屏蔽排除了大部分高噪声本底事例，缺点是无法保证始终最佳条件下运行
- 离线事例过滤是实验数据质量的最后屏障
- 加速器“50Hz”噪声通过离线事例过滤实现排除，整体取数效率提高了3~5%
- 探测器高压trip和短时束流质量不稳定的事例也可以通过离线事例过滤机制排除





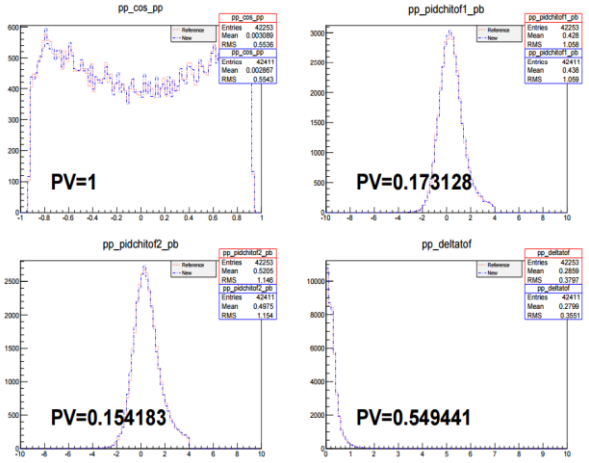




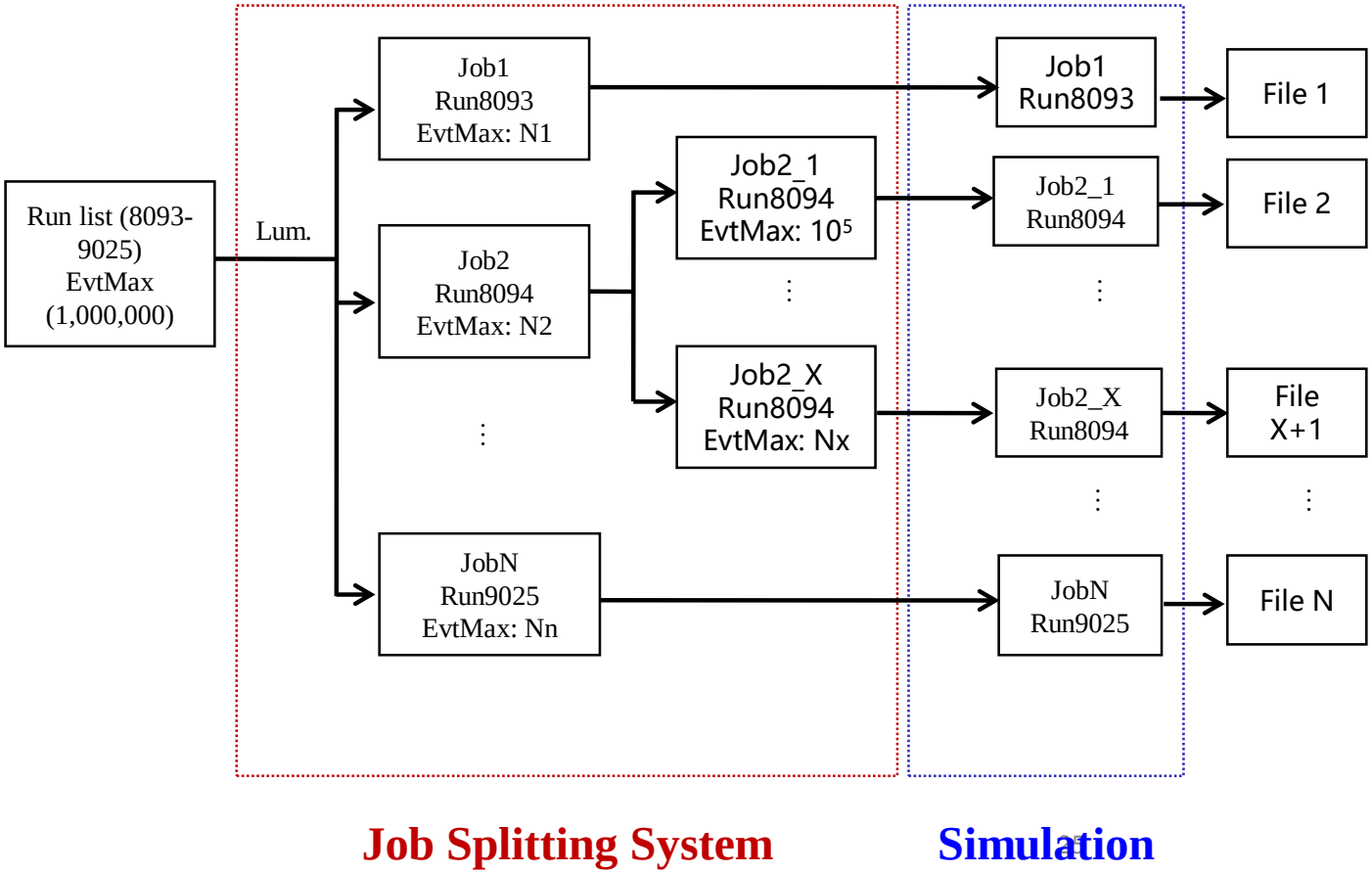
BOSS软件升级

Year	2006-2009	2010	2011	2012	2013	2014	2015	2016	2017	...	2023	2024
BOSS Release	6.0.0 ... 6.5.1	6.5.2 ... 6.5.5	6.6.0 6.6.1	6.6.2	6.6.3 6.6.4 6.6.4.p01	...	6.6.5 7.0.0	7.0.1 7.0.2	7.0.3	...	7.1.0 7.1.1	7.2.0
OS	SL4	SL5	SL5 64-bit				SL6				CentOS7	Alma9
Gaudi	v19r4		v21r6				v23r9				v27r1	v36r14
Geant4	9.0	9.3									10.7.p02	
ROOT	5.14		5.24				5.34				6.06	6.28

- 软件发布流程规范化
  - 7.0.8 -> 7.0.8.a -> 7.0.8.b -> 7.0.8.c -> 7.0.9
  - 补丁版本 7.0.9 ->7.0.9.p01
- 用控制样本进行软件测试，对比新旧版本差别
  - 模拟样本 & 真实数据
  - 检查各个子探测器的多项分布



大规模数据产生

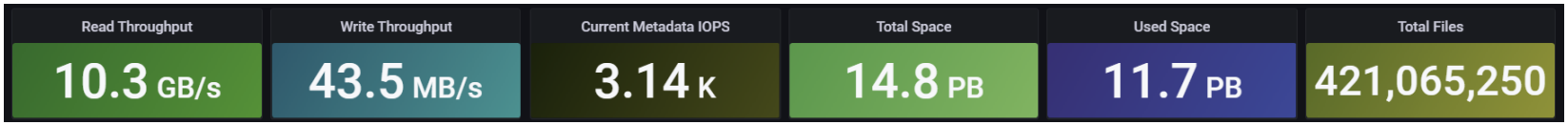


海量模拟数据产生，作业根据run号和亮度自动拆分成一批作业

BESIII计算资源使用情况

Event type	sim cpu time/ 5k events	rec cpu time/5k events
Jpsi->e+e-	3447 s	915 s
Jpsi->mu+mu-	1095 s	919 s
Jpsi->rhopi	4631 s	1167 s
Jpsi->K*K	5242 s	1306 s
Jpsi->Lambda anti-Lambda	7195 s	2251 s
Jpsi->p pbar	6325 s	1289 s
Jpsi->p pbar pi pi	9143 s	3166 s
Psip->pipijpsi, jpsi->e+e-	5840 s	2045 s
Psip->pipijpsi, jpsi->mumu	3694 s	2398 s

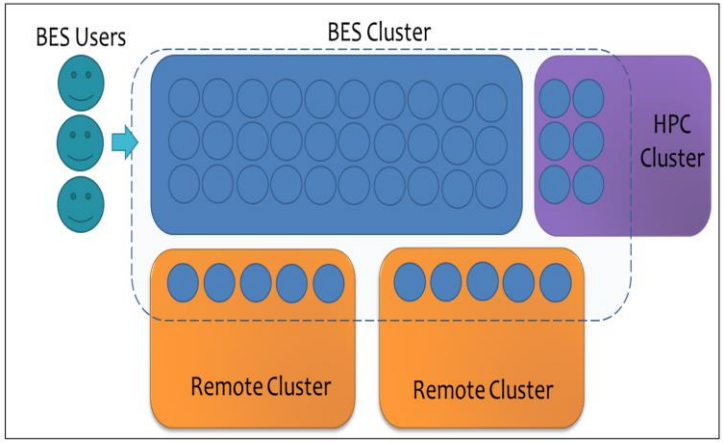
数据量： ~10PB
其中原始数据: ~3PB, (模拟+重建)数据: ~7PB

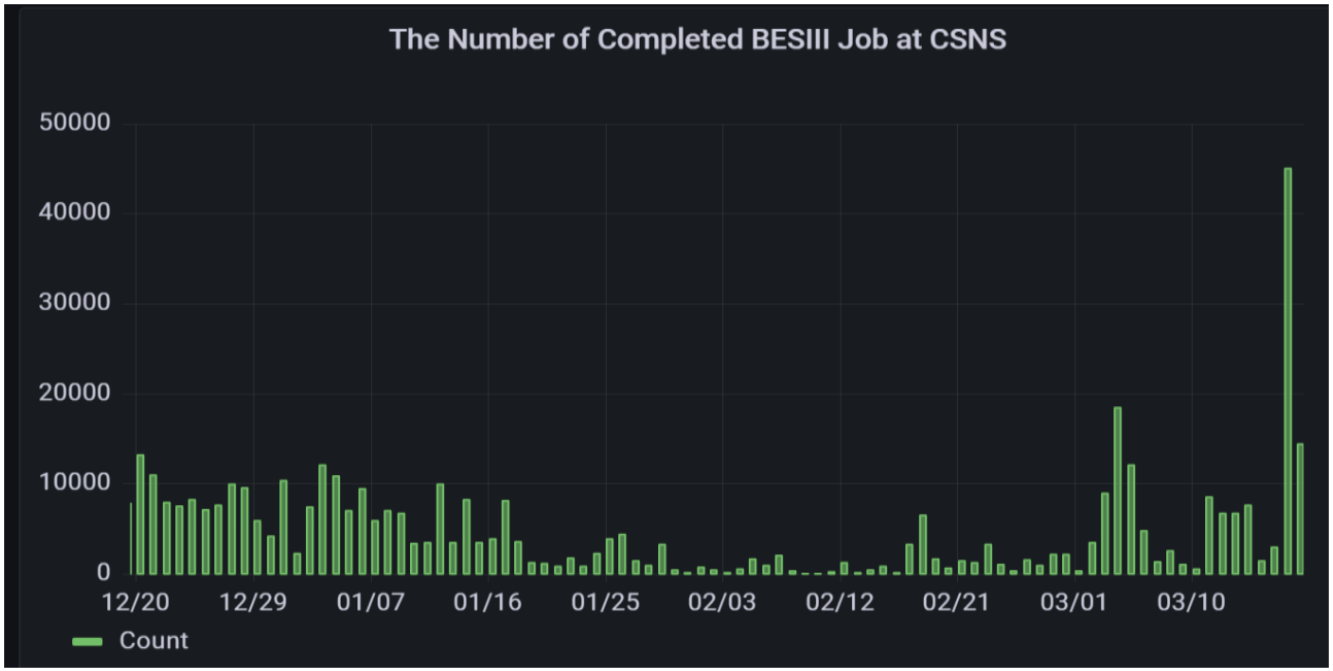


BESIII计算资源使用情况

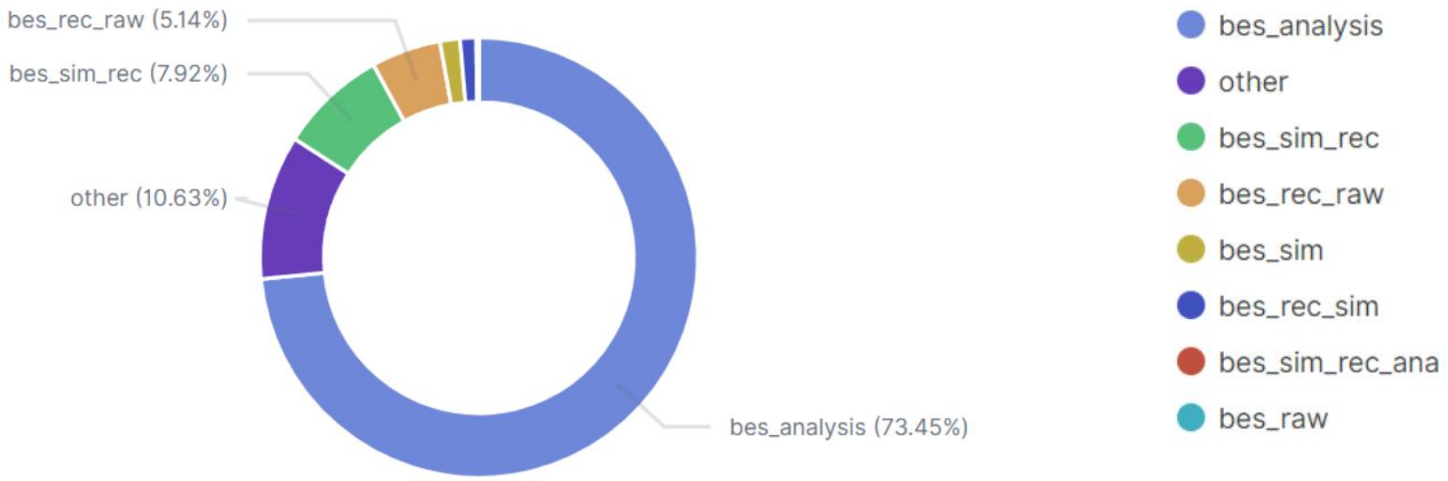
不改变用户提交作业习惯
增加BESIII可用计算资源

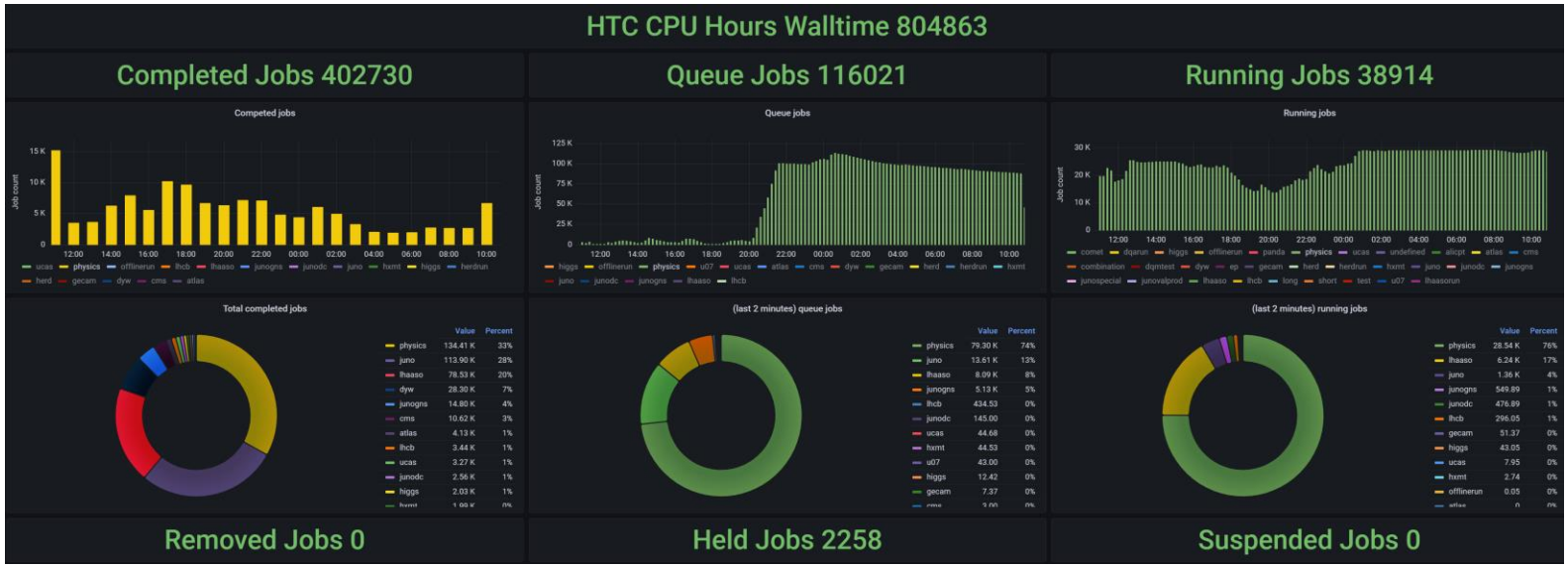
~90%模拟作业已转移到
东莞集群(~6000CPU核)





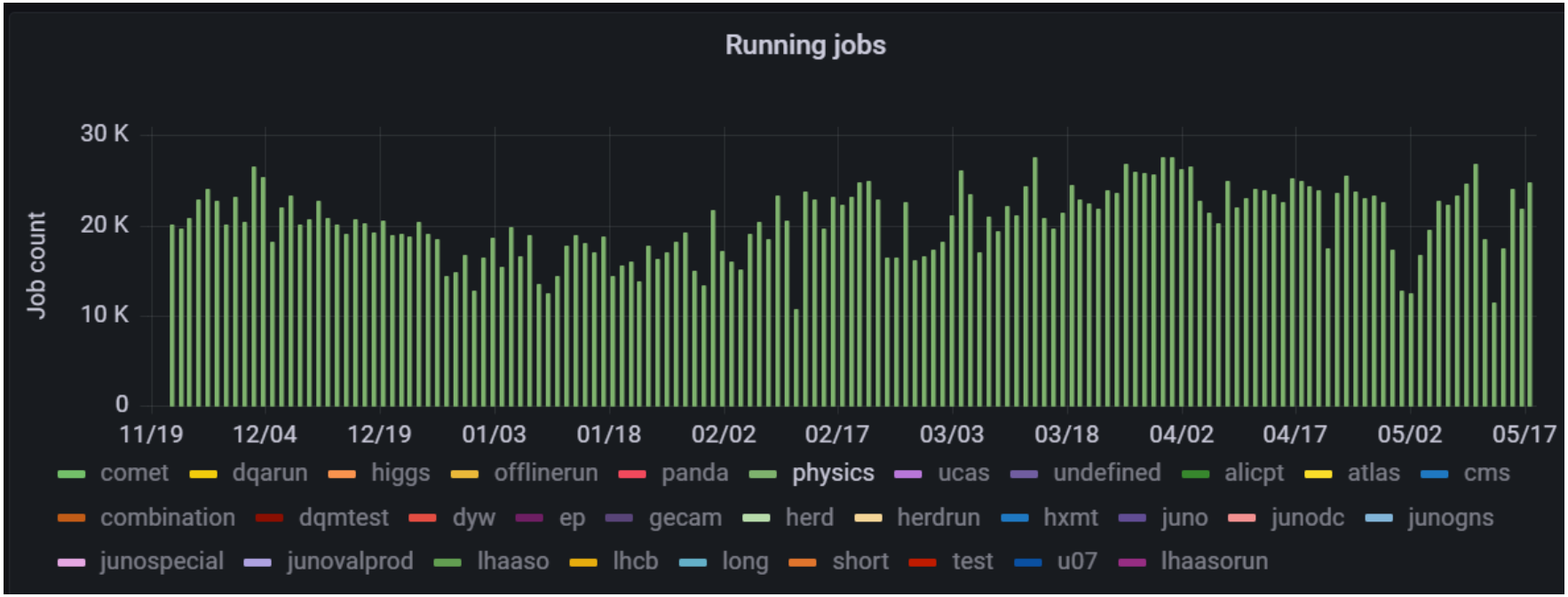
BESIII计算资源使用情况





总结

- BESIII离线软件系统包括软件框架，模拟，重建，刻度和物理分析工具等，是BESIII实验的重要组成部分
- 通过软件的不断改进，深入挖掘实验探测能力，提高数据分析的精度与速度，强有力支持了探测器稳定高效取数和物理成果的获得
- 海量实验数据的积累，显著改善了物理结果的统计误差，进一步改进系统误差是软件研究的重点内容


Running jobs
Job count
30 K
20 K
10 K
0
11/19
12/04
12/19
01/03
01/18
02/02
02/17
03/03
03/18
04/02
04/17
05/02
05/17
comet
dqarun
higgs
offlinerun
panda
physics
ucas
undefined
alicpt
atlas
cms
combination
dqmtest
dyw
ep
gecam
herd
herdrun
hxmt
juno
junodc
junogns
junospecial
junovalprod
lhaaso
lhcb
long
short
test
u07
lhaasorun