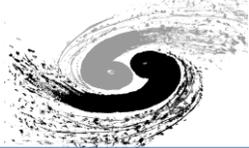


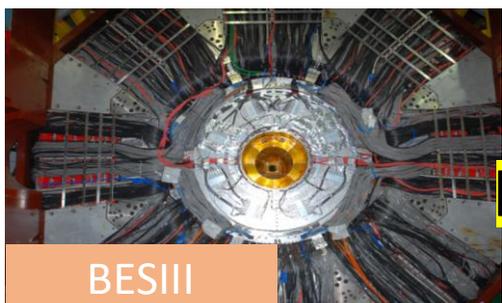
面向高吞吐的异构数据获取与处理架构—— RADAR

张叙
高能物理研究所

大规模实验数据采集



大型高能物理实验产生海量的数据，对**数据读出以及在线处理**提出了更高的要求。



BESIII

MB/s量级



LHAASO

TB/s量级

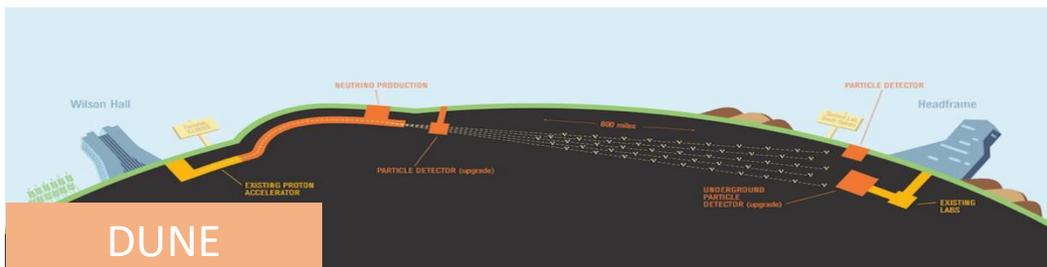
实验名称	读出数据率	存盘数据率
BESIII	~ 50 MB/s	~50 MB/s
LHAASO	~ 5 GB/s	~300 MB/s
JUNO	~ 40 GB/s	~60 MB/s
DUNE	~ 1.8 TB/s	~ 1 GB/s
LHCb	~ 5 TB/s	~ 10 GB/s
CEPC	~ X TB/s	~X GB/s



LHCb



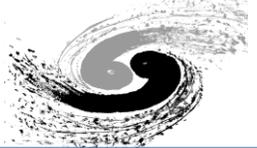
JUNO



DUNE



CEPC



流式读出是当前高能物理实验发展的趋势。

流式读出：无硬件触发，将全部数据交由软件处理。

流式读出的优点

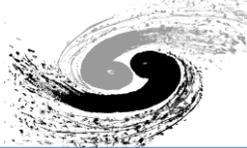
1. 简化硬件的布线。
2. 增强系统的可扩展性，可监测性。
3. 软件处理数据，更方便部署复杂算法。

流式读出带来的挑战

数据不经硬件触发就传入软件，数据量大。



Christian Färber,
HighRR seminar, Heidelberg-10.01.2018



Radar V1.0 (应用于LHAASO实验 (5GB/s数据量))



服务框架升级 (含高可用)

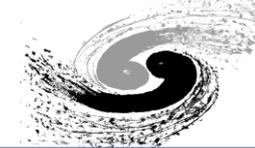
Radar V2.0 (应用于JUNO实验 (40GB/s数据量))



高带宽流式读出服务升级

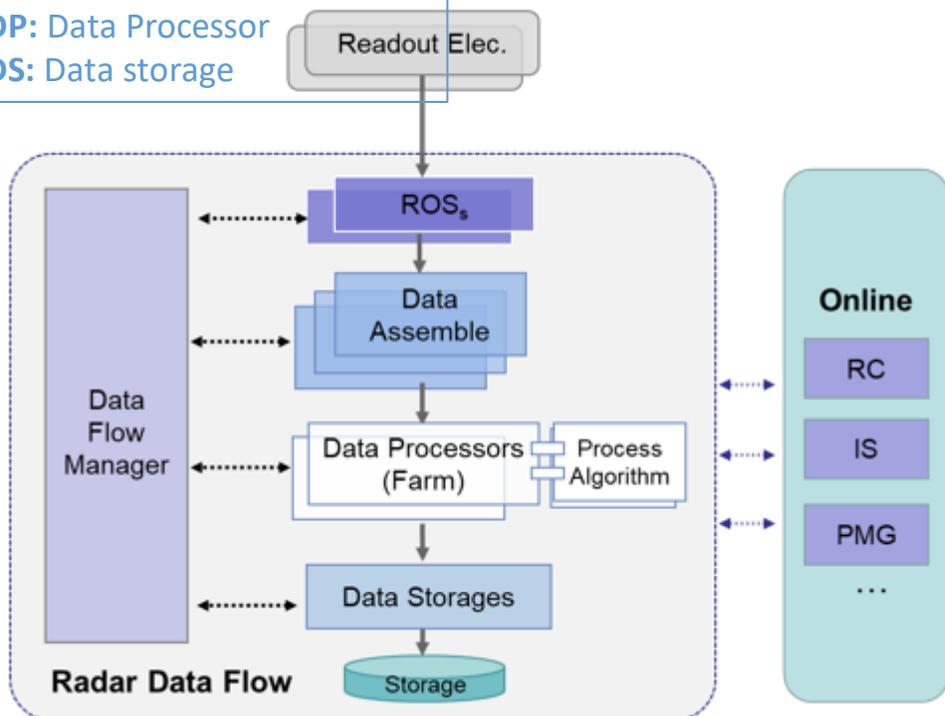
Radar V3.0 (面向数据在TB/s量级实验, 正在开发)

Radar基本框架——典型DAQ架构



heterogeneous Architecture of Data Acquisition and processing (异构数据获取与处理架构)

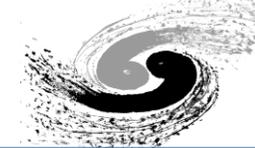
- ROS: ReadOut System
- DA: Data Assemble
- DP: Data Processor
- DS: Data storage



Radar主要分四个层级

- **读出模块 (ROS):** 采集通道级电子学数据。
- **组装模块 (DA):** 将通道级数据组装成全通道数据。
- **处理模块 (DP):** 通过搭载的物理算法处理数据。
- **存储模块 (DS):** 负责数据的存储。

Radar V2.0 应用场景



实验介绍:

江门中微子实验

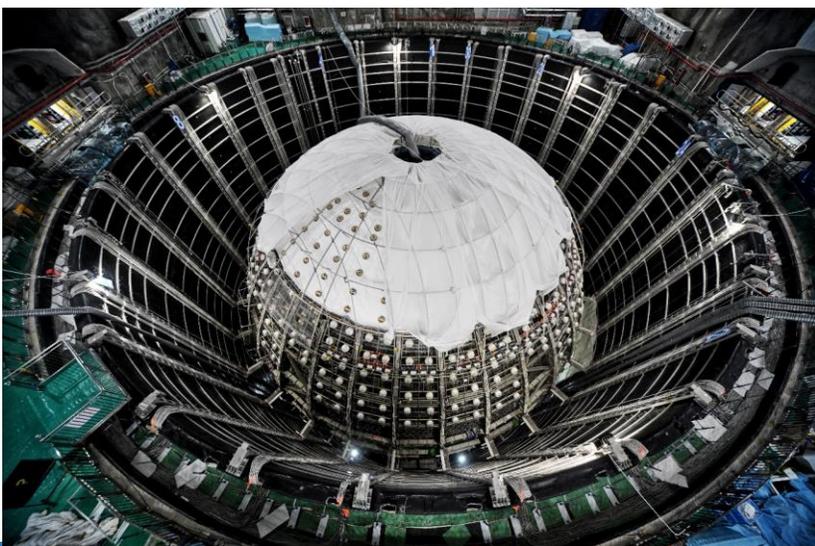
Jiangmen Underground Neutrino Observatory

科学目标:

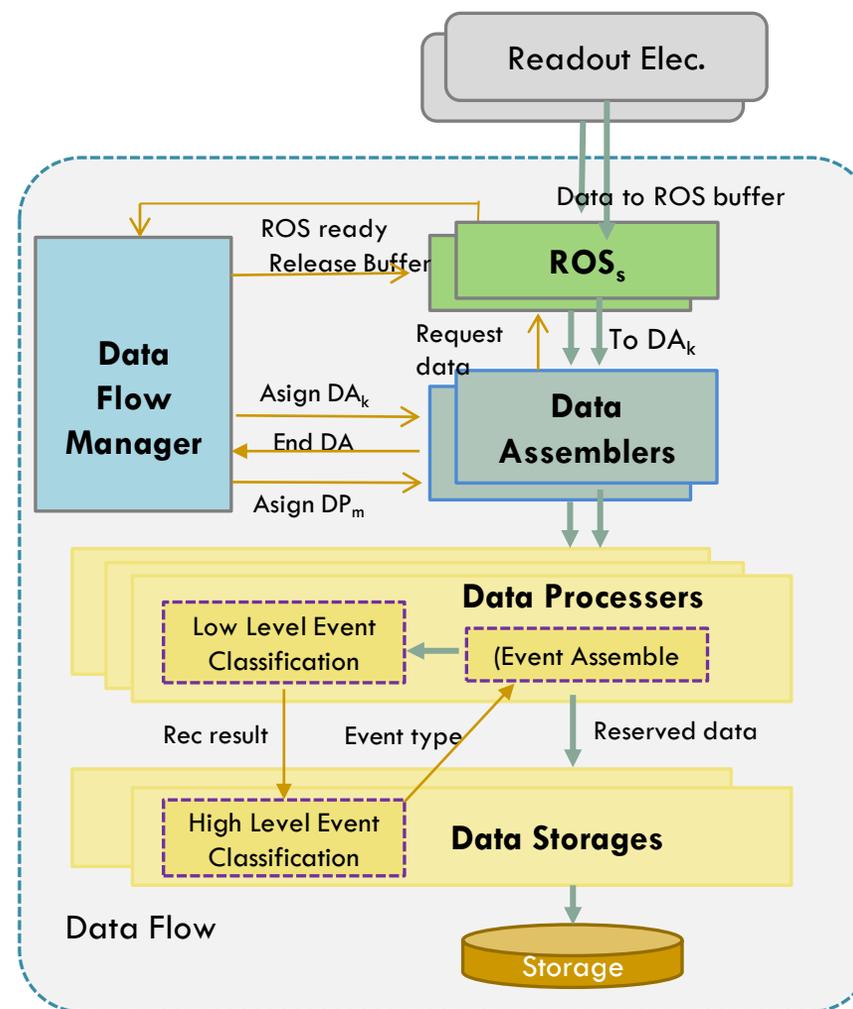
- 测定中微子质量顺序, 精确测量中微子混合参数, 观测超新星等。

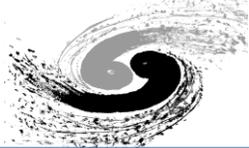
实验装置:

- CD LPMT: 17612 channels
- CD SPMT: 25600 channels
- WP LPMT: 2400 channels



JUNO数据获取系统处理更高吞吐的数据 面向混合触发(软件触发+硬件触发)



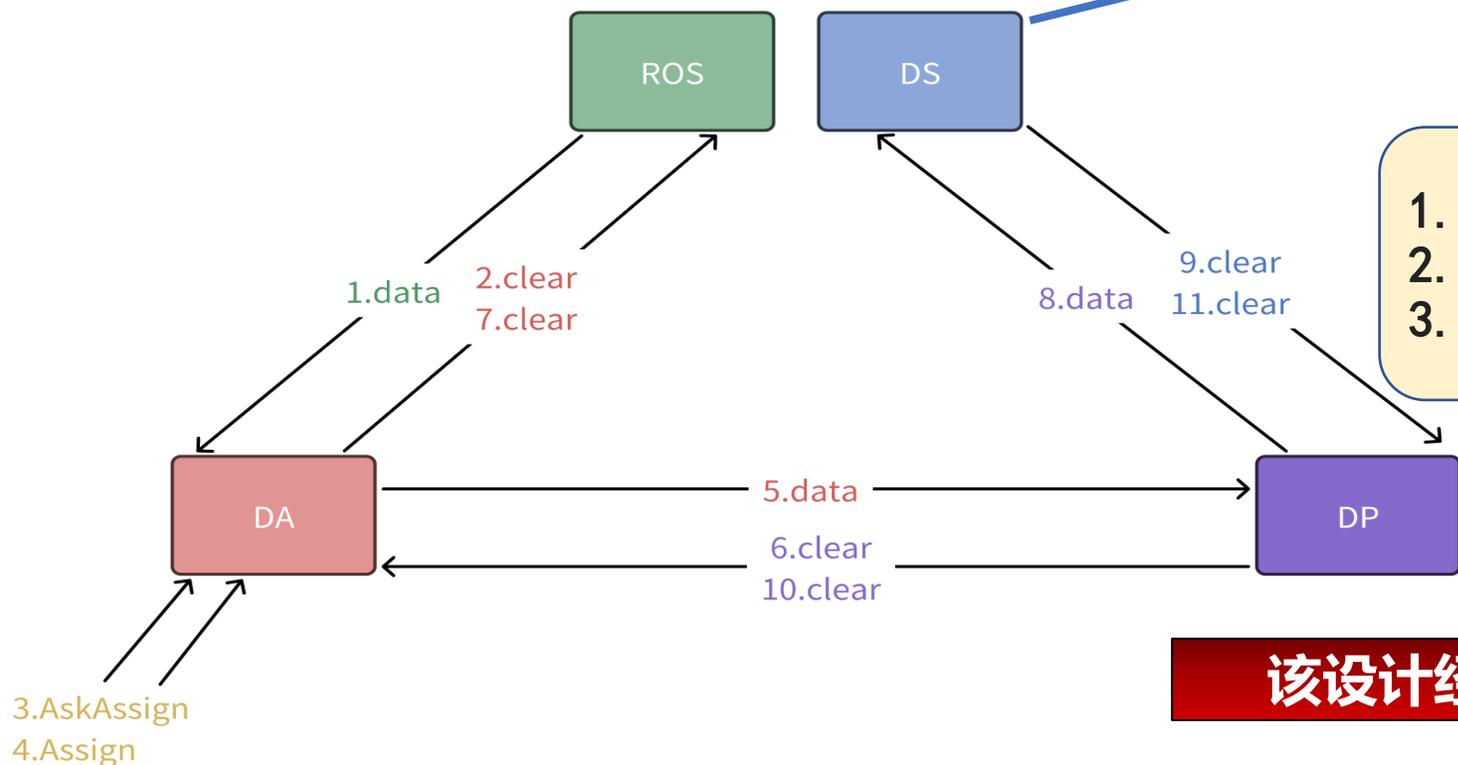


Radar V2.0 服务框架升级

Radar V2.0 基本服务框架 (容器化部署) 升级已完成, **全高可用数据流的升级正在进行中。**

数据处理单点故障恢复设计

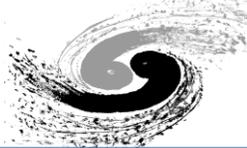
全数据流容器化运行, 便于进程的自我修复



- 1. 数据沿着ROS DA DP DS顺序传输。
- 2. 上一个节点备份当前节点的数据。
- 3. 当上一个节点收到两个clear时, 删除备份。

双备份结构

该设计经过测试验证, 未出现丢数情况



面向**更高吞吐更高带宽的高能物理实验**，例如**环形正负电子对撞机 (CEPC)**。

实验介绍：

环形正负电子对撞机

Circular Electron-Positron Collider

科学目标：

- 主要目标是产生大量的希格斯粒子，研究其性质。

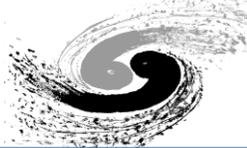
该实验当前在预研阶段



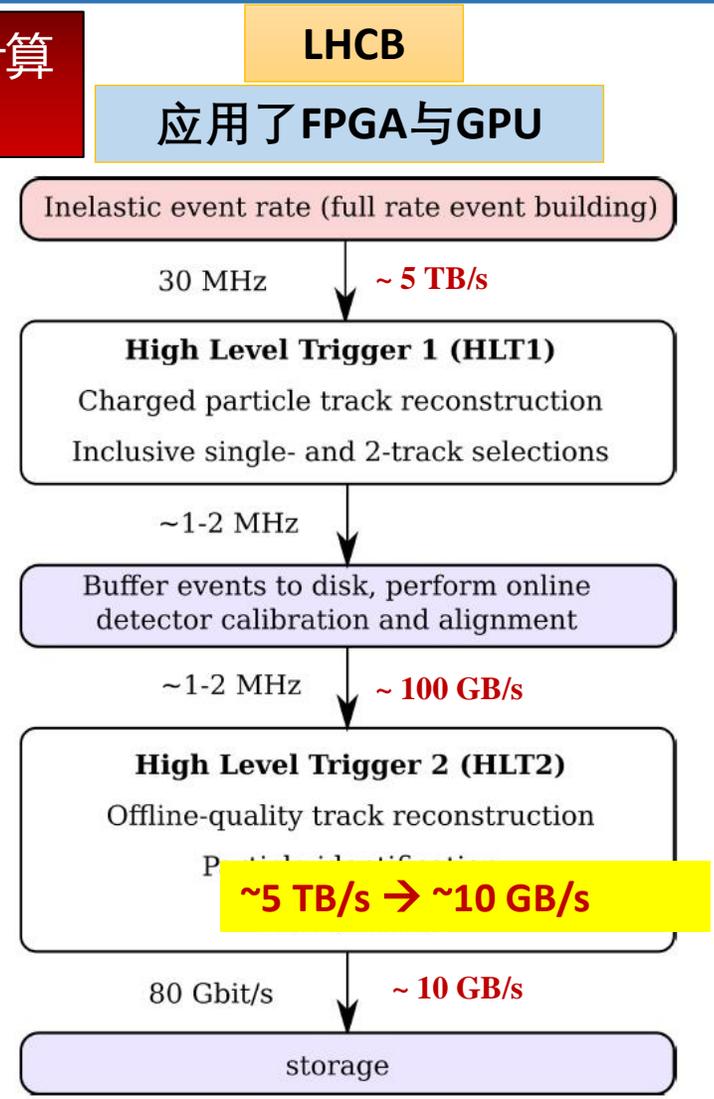
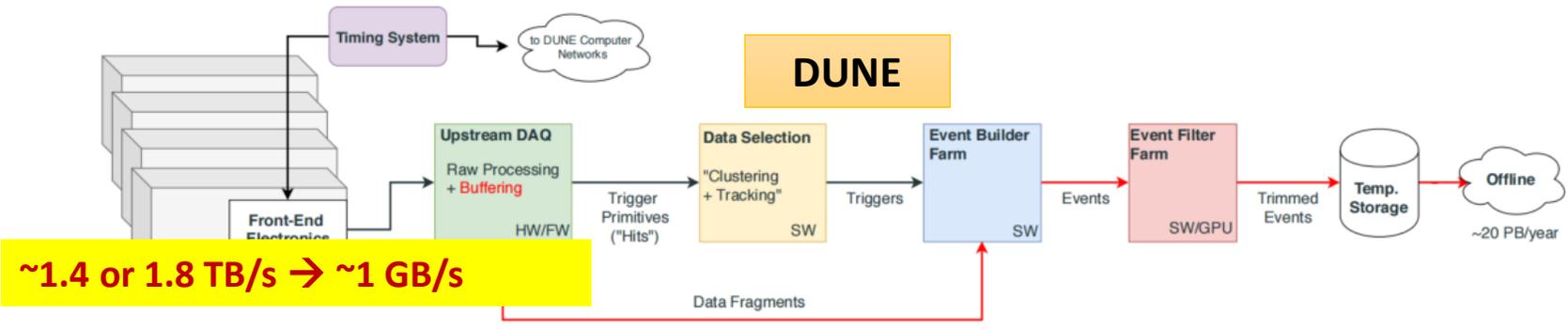
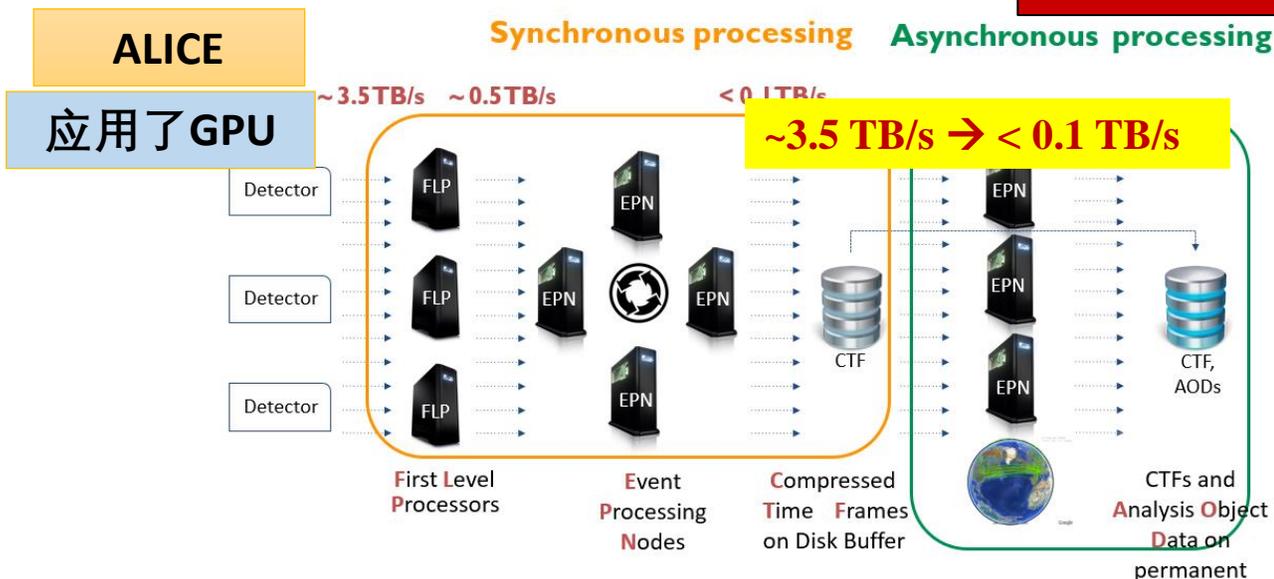
CEPC各探测器预估数据量(TB/s量级)

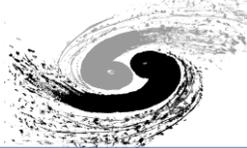
	Vertex	Tracker	TPC	DC	CAL	TOF	LumiCal
Data rate per chip	160Mbps @trigger / 4Gbps @triggerless	1.28Gbps@triggerless	-	-	-	$6.4 \cdot 10^{-3}$ bit/event/chip	-
Data rate per module	3.2Gbps@Trigger / 80Gbps@Triggerless	O(Gbps)@module O(10Gbps)@structure	-	-	-	-	-
Overall data rate	205Gbps@Trigger / 5.12Tbps@Triggerless	~40Tbps	110Gbps	800Gbps@Trigger / 6.4Tbps@triggerless	93Gbps + 72Gbps		33.6Gbps

国外TB/s量级实验调研



- 1.提高在线计算能力: 异构计算
- 2.减少数据传输: 分级计算





CEPC目前处于预研阶段，我们以JUNO场景为例围绕以下三个方面进行了技术尝试。

1. 分级计算

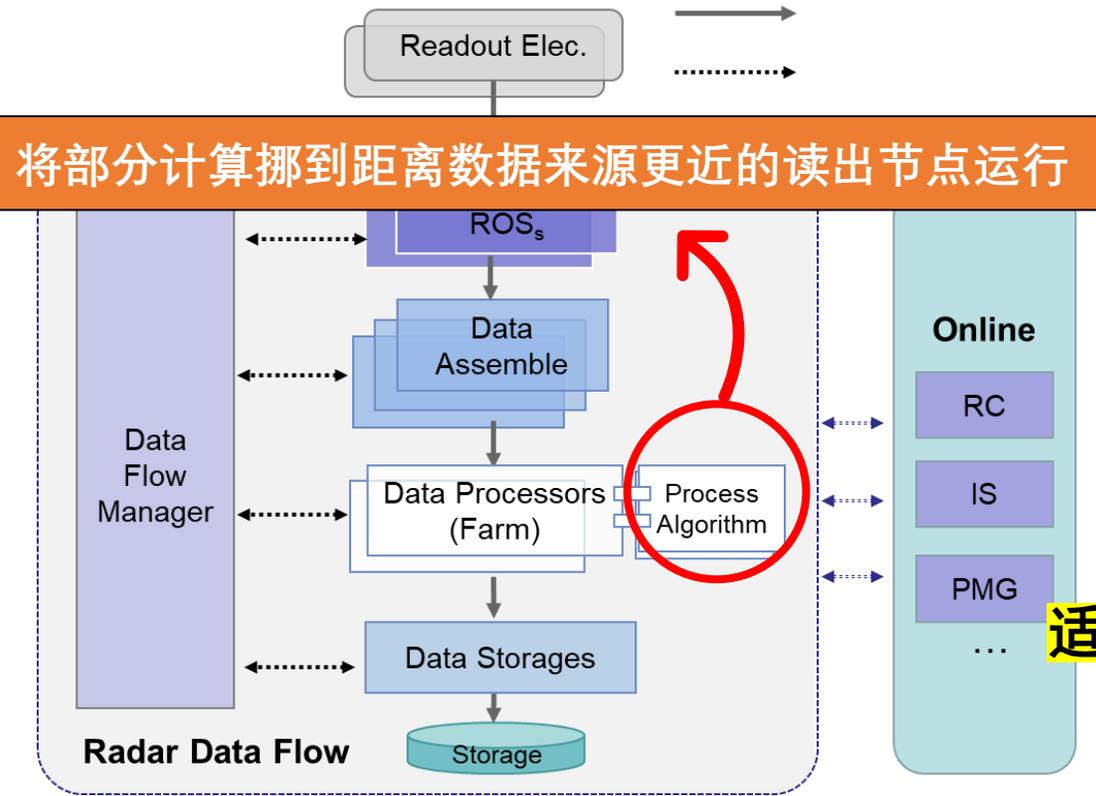
2. 异构计算

3. 灵活性设计

分级计算



分级计算：将计算放到靠近数据来源的一端。

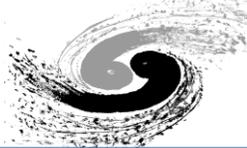


将部分计算挪到距离数据来源更近的读出节点运行

- 1.减少网络的总传输量，降低系统的IO负担。
- 2.充分利用计算资源，提高读出节点的cpu利用率。

适用于通道级计算要求较高的场景，例如波形的处理。

分级计算



基于JUNO，我们对其进行了分级计算方案评估：将Wave Reconstruction计算提前。

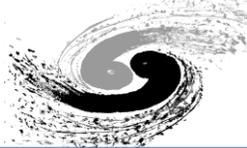
将数据处理模块中的计算前移到读出节点

该思路可以沿用到CEPC，减少数据的总传输量。

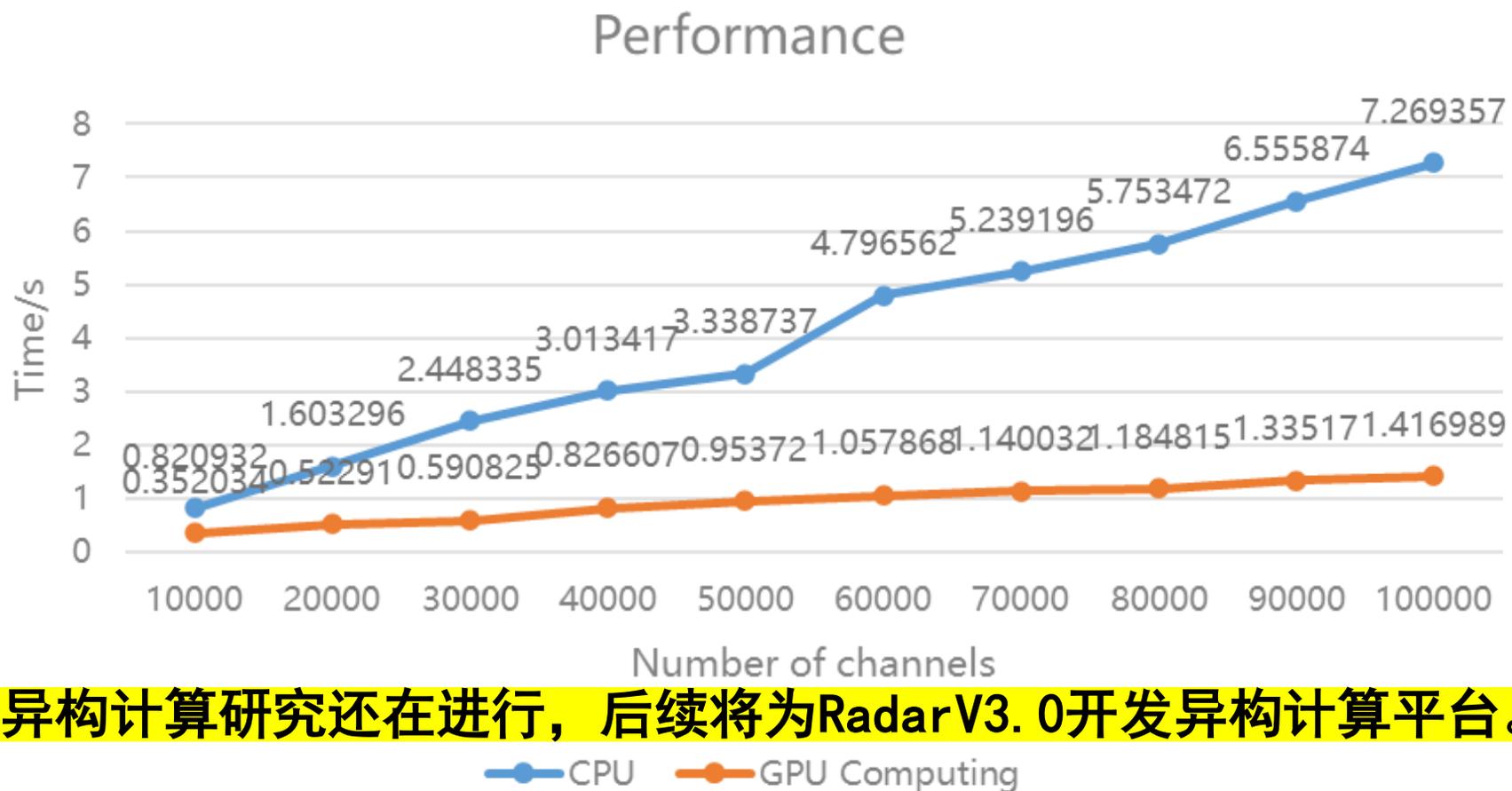
节点数	节点总核数	读出节点所需的基本核数	波形重建所需实际核数	所需总核数
25	1600	~1200	~1400	~2600
30	1920	~1200	~1200	~2400
35	2240	~1200	~1200	~2400
40	2560	~1200	~1200	~2400
45	2880	~1200	~1200	~2400
50	3200	~1200	~1200	~2400

理论上：预计读出节点用纯CPU做波形重建（基于avx-512加速）至少需要40个节点。

在读出节点做波形重建，提高了读出节点的资源利用率，大部分波形数据无需传入下一级。



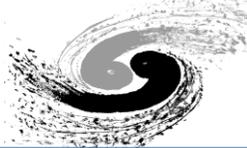
基于JUNO场景，开展了异构计算研究：将Wave Reconstruction用GPU加速。



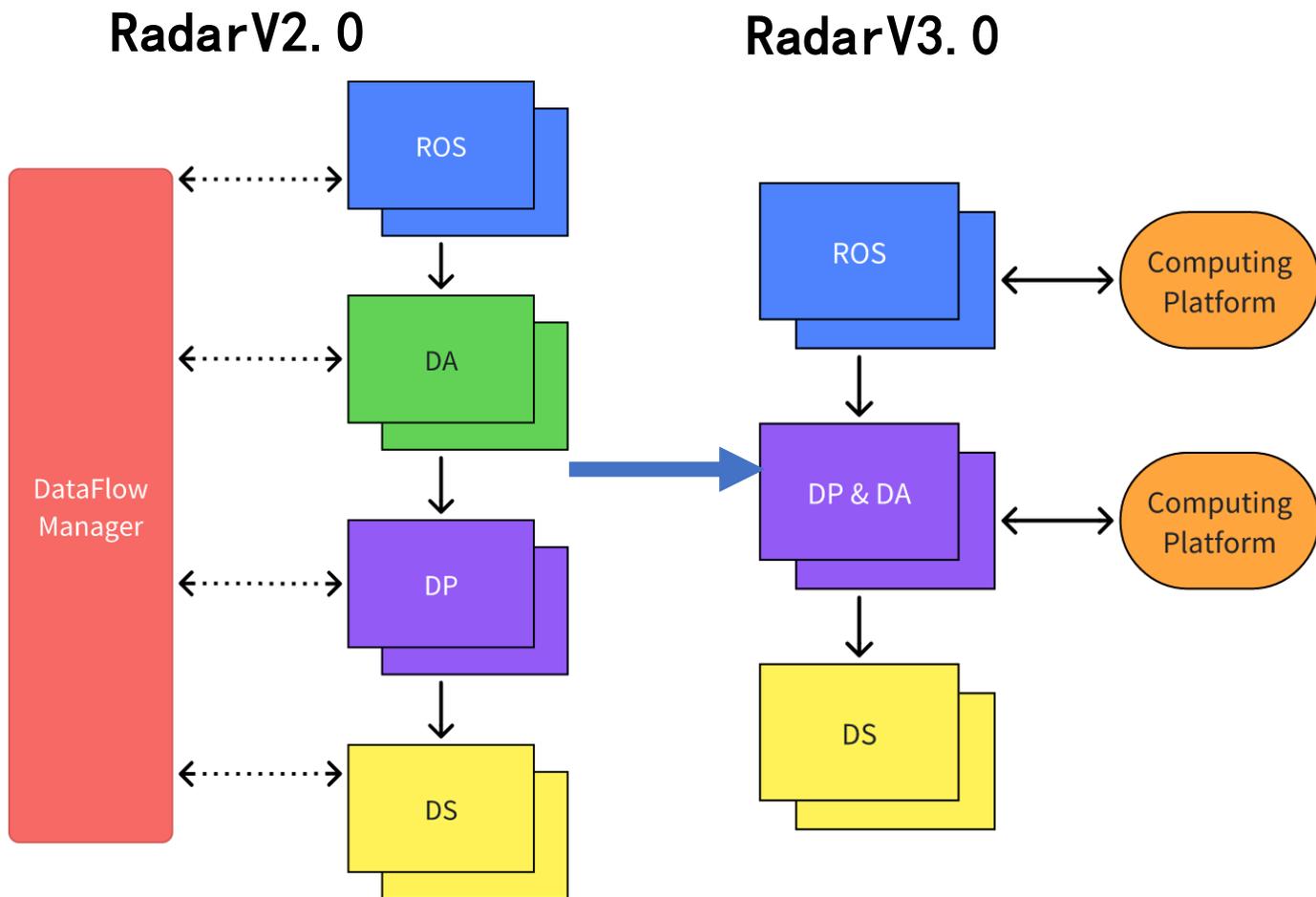
异构计算研究还在进行，后续将为Radar V3.0开发异构计算平台。

详见:24th IEEE Real Time Conference Poster <https://indico.cern.ch/event/940112/timetable/?view=standard#day-2024-04-25>

RadarV3.0 灵活性设计



RadarV3.0 可以为不同的物理实验提供服务。



1. 计算的灵活性。

数据可以在读出节点或者处理节点进行计算。

2. 框架的灵活性。

a. 不同模块之间可以相互结合。

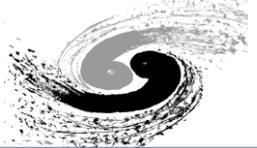
例如将DA与DP结合在一起。

b. 针对不同的需求选择不同的模块。

例如只使用ROS, DP, DS模块。

3. 调度的灵活性

非集中式, 分层调度, 降低单点故障影响。



RadarV1.0:

- 基于RadarV1.0开发的系统应用在高海拔宇宙线观测站实验 (LHAASO) 上, 平稳运行至今。

RadarV2.0 (服务框架升级):

- 基于RadarV2.0开发的系统应用在江门中微子实验 (JUNO) 上, 已实现了基本功能。

RadarV3.0 (高带宽流式读出服务升级):

- 围绕**异构计算**, **分级计算**, **灵活性设计**三大方面。
 - 为下一代**更高带宽更高吞吐**的高能物理实验提供数据获取与在线处理的解决思路。
 - 搭建**稳定, 高速, 实时, 灵活**的高能物理实验数据获取与在线处理框架。



THANKS