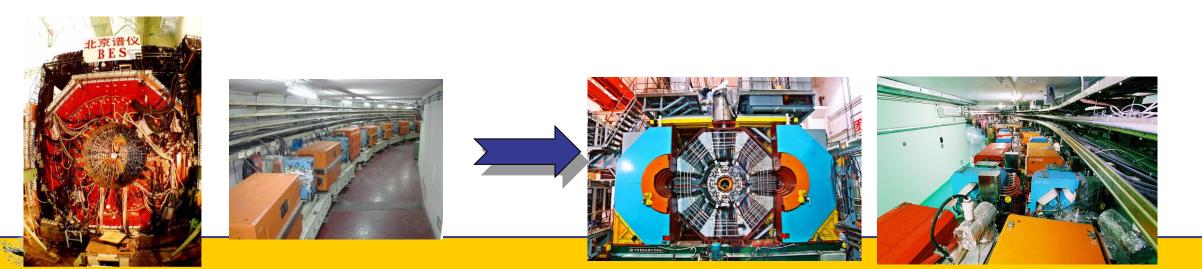
# **BESIII** Data Preservation

### Gang Chen, Beijiang Liu IHEP



# **BESIII/BECPII**

- BEPC: Beijing Electron Positron Collider
  - Started in 1989, and upgraded to <u>BEPCII</u> in 2004
  - Dual-Ring, 2~5GeV/C
  - Luminosity (3~10) × 10<sup>32</sup> cm<sup>-2</sup>s<sup>-1</sup>
- BES: Beijing Spectrometer
  - Upgraded to <u>BESIII</u> with BEPCII
  - Data-taking from May 2004
  - <u>Decommission in 2030</u>



## **BESIII** Collaboration

Europe (17/115)	Asia (6/10)
Germany (6): Bochum University,	Pakistan (2): COMSATS
GSI-Darmsfadt, Helmholtz Institute Mainz, Johannes Gutenberg University of Mainz,	Institute of Information
Universitaet Giessen, University of Münster	Technology
Italy (4): Ferrara University, LNF/Perugia, University of	University of the Punjab,
CANADA CANADA	University of Lahore
	Mongolia (1): Institute of
Netherlands (1):KVI/University of Groningen	Physics and Technology
PACIFIC USA(4/8) Russia (2): Budker Institute of Nuclear Physics, Dubna JINR	Korea (1): Chung-Ang
Ceanegie Mellon University - Sweden (1): Uppsala University - CHINA	University OCEAN
Indiana University Turkey (1): Turkish Accelerator, Center Particle Factory Group	India (1): Indian Institute of
University of Hawaii UK (2): University of Manchester, University of Oxford	Technology madras
University of Minnesota Poland (1)National Centre for Nuclear Research	Thailand (1): Suranaree
	University of Technology
China (55/367)	
	Exactly Revealed and
South America (1/1) Institute of High Energy Physics (146), other units (22)	Being Institute of Petro-

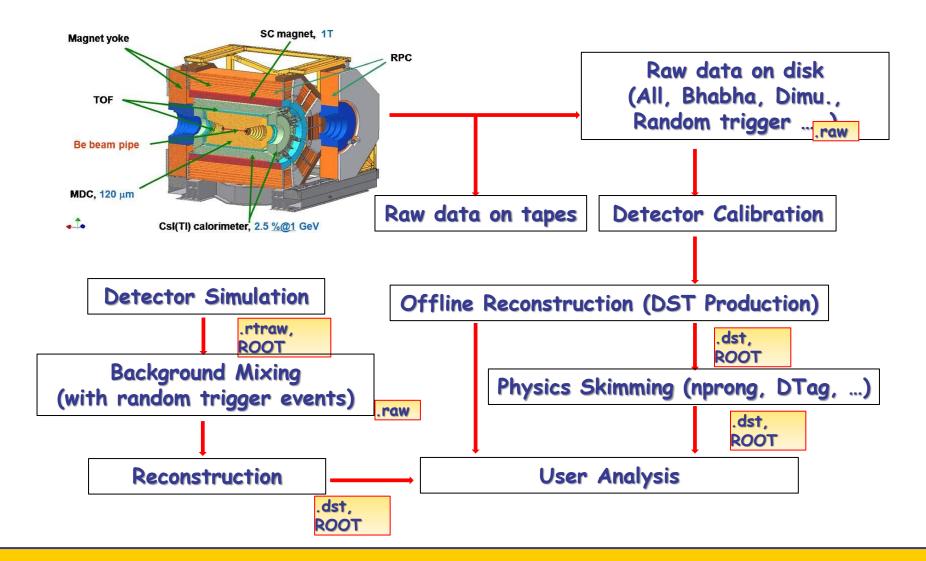
South America (1/1) Chile: University of Tarapaca

ACCOUNT, ALL

~600 members From 83 institutions in 16 countries

Petrochemical Technology, Beihang University, China Center of Advanced Science and Technology, Fudan University, Guangxi Normal University, Guangxi University, OCEAN Hangzhou Normal University, Henan Normal University, Saint Helena, Ascension and Tristan da Cunhe 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, Oufu normal university, Shanxi University, Shanxi Normal University, Sichuan University, Shandong Normal University, Shandong University, Shanghai Jiaotong University, Soochow Unive 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,

# Data Analysis Model





# **BESIII** Data and software

- 1.8 PB raw data collected up to 2023.
- 3.6 PB reconstructed data and MC data correspond to three different versions of BOSS (BESIII Offline Software System).
- BEPCII is being upgraded and the luminosity will be increased.
   750 TB/year of raw data will be collected from 2024.



# Data preservation

- BESIII adopts DPHEP Level-4 model
  - The full potential of data
    - RAW, DST (data, incl.MC),
    - metadata (calibration databases, .....),
    - software, documents
  - Adhere to the FAIR principles

DPHEP Collaboration: T. Basaglia, M. Bellis J. Blomer et al.: Data Preservation in High Energy Physics Eur.Phys.J.C 83 (2023) 9, 795

Level	Model	Use Case
1	Provide additional information	Publication-related information search
2	Preserve the data in simplified form	Outreach, simple training analysis
3	Preserve the analysis-level software and data format	Full scientific analysis based on existing reconstruction
4	Preserve the reconstruction and simulation software	Full potential of the experimental data
	and raw data	



# Data preservation (cont.)

Т	asks	Needs	To-do
Data	<ul> <li>Raw: data</li> <li>DST: data/ incl. MC</li> </ul>	Storage	<ul> <li>Define the requirements</li> <li>Develop a prototype</li> </ul>
Metadata	<ul><li>Databases</li><li>Bookkeeping</li></ul>	<ul> <li>Storage</li> <li>With checksum</li> <li>Long time</li> </ul>	<ul> <li>Storage services</li> <li>Containers with</li> </ul>
Documents	HN, docdb, webpages, indico, codes,	Contoinona	orchestration <ul> <li>Verification tools</li> <li>Get the resources</li> </ul>
Software	<ul> <li>System</li> <li>Analysis framework</li> <li>External libs</li> </ul>	<ul> <li>Containers</li> <li>Virtualization</li> <li>No longer maintenance</li> </ul>	

However, permanently preserved does not mean practically useable

- Large scale resources for now are not forever (20000 CPU cores, ...)
- Lack of expertise with BESIII specific detailed information

# Bit preservation

- Media: Tape LTO9, IBM-3584 tape library
- Raw data/Random trigger data:
  - Flows from on-line farm to tape directly,
  - A copy on disk after reconstruction.
- Rec/DST/skimmed DST/tag data :
  - Replicated from disk to tape when a certain software version is stable.
- Condition/document databases:
  - Snapshots are copied to tape routinely
- Integrity check
  - A MD5 integrity check is done when data is copied from disk to tape
- Annual examination of tape library and LTO9 tapes



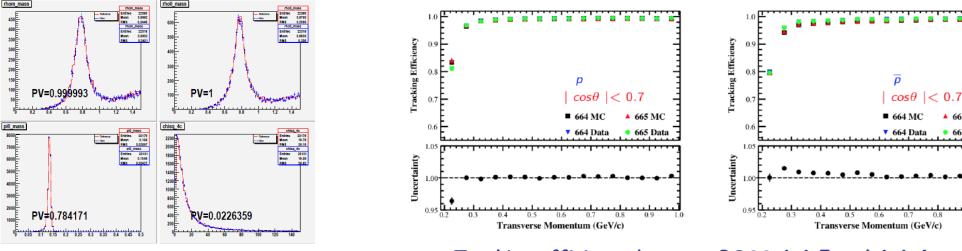
# Software preservation

- BOSS is an integrated software package that includes all the blocks required in BESIII data processing.
  - Simulation, reconstruction, calibration, analysis...
  - functions could be preserved all together in a package
- The stable versions of BOSS are preserved following items:
  - A complete package of software,
  - A runnable virtual machine image
  - The puppet template and RPM repository from which a runnable OS is created,
  - Release documents, bookkeeping parameters...
- Functional validation is done according to the standard software release process.



# Validation Process

- Several MC data samples are generated to do MC validation
  - For example, Jpsi to Rhopi, KsKpi, ee, mumu, ppbar
- Related data samples are reconstructed to check the consistency between MC and real data



Jpsi to Rhopi

Tracking efficiency between BOSS 6.6.5 and 6.6.4

▲ 665 MC

665 Data



# **Bookkeeping for calibration constants**

BeamPipe308CalTask252CalVbLumVer311DataFmtVer311DataFmtVer268DataFmtVsPkg268DataFmtVsPkg268DedxCalConst277DedxCalSft267DedxCalSft267DedxCalSft267DedxCalConst278EsTmeCalConst310EstTofCalConst311LogisticNames333McNextEventID332McAlgmment316McCalSft2675MdcCalSft329McCalSft323McCalSft333McVextEventID332McCalSft303McCalSft303MucCalSft303MucCalSft303MucCalSft303MucCalSft306RanTgData306Replicas276RunParams294RunParams64238SftPkg254TofCalConst307TofCalSft314TofCalConst307TofCalSft314TofCalConst307TofCalSft314TofCalSft314TofCalSft314TofCalSft314TofCalSft315WirePositionCaliDD522seed237test255236245	BossRelease	DataType	RunFrom	RunTo	SftVer
BeamPipe308CalTask252CalVbLumVer311DataFmtVer311DataFmtVer268DataFmtVsPkg268DataFmtVsPkg268DedxCalConst277DedxCalSft267DedxCalSft267DedxCalSft267DedxCalConst278EsTmeCalConst310EstTofCalConst311LogisticNames333McNextEventID332McAlgmment316McCalSft2675MdcCalSft329McCalSft323McCalSft333McVextEventID332McCalSft303McCalSft303MucCalSft303MucCalSft303MucCalSft303MucCalSft306RanTgData306Replicas276RunParams294RunParams64238SftPkg254TofCalConst307TofCalSft314TofCalConst307TofCalSft314TofCalConst307TofCalSft314TofCalSft314TofCalSft314TofCalSft314TofCalSft315WirePositionCaliDD522seed237test255236245	6.6.4	EsTime	9947	10878	6.6.2.b
CalTask333CalTask252CalVbLumVer311DataFmtVer268DataFmtVsPkg268DataSets259DedxCalConst277DedxCalSft267DedxCalConst261EncCalConst260EncCalSft278EsTimeCalConst310EstTorCalConst311LogistCNames333McNextEventID332MdcAlignment264MdcCalConst271MdcCalSft275MdcCalSft303McCalSft303McCalSft303McCalSft303Metadata324MucCalSft303NextIDs240NextIDs_bak313OfflineLum253QualityParams294StipKg254TofCalConst307TofCalConst307TofCalConst307TofCalSft314TofCalConst307TofCalConst307TofCalSft315WirePositionCalibD237seed237test225236245		EsTime	23463	24177	
CalVtxLumVer311DataFmtVer311DataFmtVer268DataFmtVsPkg268DataFmtVsPkg268DataFmtVer281DedxCalConst277DedxCalSft260EncCalSft260EncCalSft278EsTimeCalConst310EstTorCalConst256Files312Jobs274LeakCorr331McAcalgnment264MdcCalSft275MdcCalSft275MdcCalSft303Metadata324MucCalSft303NextIDs240NextIDs_bak313OfflineLum253QualtVParams294RanTrgData306Replicas276RunParams664238SftPkg254TofCalConst307TofCalConst307TofCalSrt314TofCalConst307TofCalSrt313Stafkg254TofCalConst307TofCalSrt315WirePositionCalibD237seed237test226245		EsTime	28649	80000	
DataFmtVer         268           DataFmtVsPkg         259           DedxCalConst         277           DedxCalSft         267           DedxCalSft         267           DedxCalSft         261           EmcCalConst         278           DedxSim         260           EmcCalConst         278           EstTmeCalConst         256           Files         3110           Jobs         274           LeakCorr         333           McNextEventID         3322           MdcAlignment         264           MdcCalConst         275           MdcCalConst         271           MdcCalConst         273           MdcCalConst         275           MdcCalConst         275           MdcCalConst         275           MdcCalConst         275           MdcCalConst         239           MucCalSft         303           NextIDs         240           NextIDs         240           NextIDs         240           NextIDs         240           NextIDs         240           NextIDs         240		EsTime	20448	23454	
DatabrityskigDatabrityskigDataSetsDedxCalConstDedxCalSftDedxCalSftDedxCalSftDedxCalConstEmcCalConstEmcCalConstEsTimeCalConstEsTimeCalConstEventTypesJobsCreationMcCalSftDatasCorrJobsPilesJobsMcCalSftEventTypesJatasCorrJobsMcCalConstMcCalConstMcCalSftMcCalSftMcCalSftMcCalSftMcCalSftMuc		EsTime	24897	28648	
DedxCalConst         277           DedxCalConst         267           DedxCalSft         267           DedxCalConst         261           DedxCalConst         260           EmcCalConst         260           EmcCalConst         278           EsTimeCalConst         310           EventTypes         312           Jobs         274           LeakCorr         331           LogisticNames         333           McNextEventID         332           McCalSft         275           MdcCalConst         271           MdcCalSft         221           MdcCalSft         323           MdcCalSft         333           MucCalSft         303           MucCalSft         303           MucCalSft         303           QualityParams         253           RanTrgData         306           Replicas         276           RunParams664         238           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         522		EsTime	8093	9779	
DedxCalSft         267           DedxCalSft         267           DedxCalSft         261           EmcCalConst         260           EmcCalSft         278           EsTimeCalConst         260           EncCalSft         278           EsTimeCalConst         310           EventTypes         312           Jobs         274           LeakCorr         331           LogisticNames         332           MdcAlignment         266           MdcCalConst         275           MdcAlignment         264           MdcCalSft         521           MdcCalSft         323           MucCalSft         333           MucCalSft         303           MucCalSft         303           MucCalSft         303           MucCalSft         303           MucCalSft         303           OffineLum         253           QualityParams         294           RunParams664         238           SftPkg         254           TofCalConst         307           TofSimSvc         515           236         237		EsTime	11414	14604	
DedxCurvePar281DedxSim261EmcCalConst260EmcCalSft278EsTimeCalConst310EstTofCalConst310EventTypes256Files312Jobs274LeakCorr331LogistdNames333McNextEventID332MdcAlgnmentMdcCalConstMdcAlgnment316MdcAlgnment316MdcAlgnment316MdcAlgnment316MdcAlgnment316MdcCalSft303MuccalSft303MuccalSft303MuctDs_bak313OffineLum253QualityParams294RunParams664238SftPkg254TofCalConst307TofCalSft314TofQELec315VirePositionCalibD522seed237test255236245		EsTof	24897	28648	
DedxSim         261           EmcCalConst         260           EmcCalSft         278           EsTimeCalConst         310           EstTimeCalConst         310           EstTimeCalConst         256           Files         312           Jobs         274           LeakCorr         331           McNextEventID         332           McNextEventID         332           MdcCalConst         271           MdcCalSft         275           MdcDataConst         521           MdcCalConst         232           MucCalSft         303           MextIDs         240           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         294           NuParams664         238           SftPkg         254           TofCalConst         307           TofCalSft         314           TofQeLec         315           TofSimSvc         522           WirePositionCalibD         2237           seed         235           236         245		EsTof	-28648	-27147	
EmcCalConst260EmcCalSft278EsTimeCalConst310EsTimeCalConst310EventTypes256Files311Jobs274LeakCorr331LogisticNames333McNextEventID332MdcAilginment264MdcCalSft275MdcDataConst221MdcCalSft303MextIDs240NextIDs241TofCalConst307TofCalSft314TofCalSft315VirePositionCal		EsTof	29628	80000	
EmcCalSft         278           EsTimeCalConst         310           EsTimeCalConst         256           EventTypes         256           Files         312           Jobs         274           LeakCorr         333           McNextEventID         332           McAlignment         264           MdcCalSft         275           MdcCalConst         221           MdcCalConst         223           MucCalSft         303           NextIDs         240           StPkg         253           RunParams664         238           SftPkg         254           TofCalConst         307           TofCalConst         3		EsTof	-10878		6.6.2.b
EstTimeCalConst310EstTiofCalConst256EventTypes312Jobs274LeakCorr331LogisticNames333McNextEventID332McCalIgnment264MdcCalConst275MdcCalSft301Mutzls240McCalSft303NextIDs240NextIDs_bak313OffineLum253QualityParams294RunParams664238SftPkg254TofCalConst307TofCalSft314TofCalConst307TofCalSft314TofSimSvc512WirePositionCalibD522seed237test256236245		EsTof	11414	14604	
Est1orCalConst         256           EventTypes         312           Jobs         274           LeakCorr         331           LogisticNames         333           McVextEventID         332           MdcAlignment         264           MdcCalConst         275           MdcCalSft         521           MucCalSft         308           MucCalSft         303           NextIDs         240           NextIDs         306           RunParams         294           StPkg         254           TofCalConst         307           TofCalSft         314           TofQalEec         315           VirePositionCalibD         522           seed         <		EsTof	23463	24177	
Event Vpes         312           Files         312           Jobs         274           LeakCorr         331           LogistcNames         333           McNextEventID         332           MdcAlignment         264           MdcCalSft         271           MdcCalSft         274           MdcCalSft         275           MdcCalSft         276           MdcCalSft         333           Metadata         324           MucCalSft         333           NextIDs         240           NextIDs         306           Replicas         276           RunParams664         238           SftPkg         254           TofCalConst         307		EsTof	23463		6.6.2.b
Jobs         274           Jobs         274           LeakCorr         331           Logisto/Names         333           McNextEventID         332           MdcAlignment         264           MdcCalf         275           MdcCalfst         275           MdcCalfort         264           MdcAlignment         264           MdcCalfort         275           MdcCalost         279           MucCalSft         303           Metadata         324           MucCalSft         303           MettDs         240           NextIDs         240           RanTrgData         306           Replicas         276           RunParams         294           TofCalSft         314           TofQeLec         315           TofQeLec         315           VirePositionCalibD         252           seed <td< td=""><td></td><td>EsTof</td><td>20448</td><td>23454</td><td></td></td<>		EsTof	20448	23454	
LeakCorr         331           LogisticNames         333           McNextEventID         332           MdcAilgnment         264           MdcCalConst         275           MdcCalSft         275           MdcCalSft         331           McVatEventID         332           MdcCalConst         264           MdcCalSft         375           MdcTaltConst         521           MdcTaltata         324           MucCalSft         303           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         276           RunParams         276           Stft/kg         254           TofCalConst         307           TofCalConst         307           TofCalConst         307           TofCalConst         307           TofQELec         315           TofSimSvc         315           Seed         237           test         255           236         245		EsTof	-80000	-29677	
LogisticNames         333           McNextEventID         332           MdcAilgnment         264           MdcCalConst         275           MdcCalSft         275           MdcCalSft         324           MdcCalConst         323           MdcCalSft         333           MdcCalSft         336           MdcCalConst         229           MucCalSft         303           NextIDs         240           NextIDs_bak         313           OffineLum         253           QualityParams         253           RunParams         294           StPkg         254           TofCalSft         307           TofCalSft         314           TofQELec         315           StPkg         252           seed         237           test         255           236         245		Es lor Es Tof	-80000	-29677	
McNextEventID         332           McNextEventID         332           MdcAlignment         264           MdcCalConst         275           MdcCalSft         275           MdcCalConst         521           MdcCalConst         521           MdcCalConst         521           MdcCalConst         521           MdcCalConst         239           MucCalSft         303           NextIDs         240           RanTrgData         306           RunParams664         238           SftPkg         254           TofCalSft         314           TofCalSft         314           TofSimSvc         252           WirePositionCalibD         237           test		Es lof Es Tof	-2/146	-20333	
MdcAlignment332MdcCalConst264MdcCalSft275MdcCalSft521MdcDataConst316Metadata324MucCalSft303NextIDs240NextIDs240NextIDs240NextIDs240NextIDs240RanTrgData306Replicas276RunParams294SftPkg254TofCalConst307TofCalSft318TofCalSft317TofQELec315VirePositionCalibD522seed237test256236245		Es lor Es Tof	-9809	-8093	
MdcCalConst         275           MdcCalSft         521           MdcDataConst         521           MdcDataConst         316           Metadata         324           MucCalSft         303           MettDs         240           NextIDs         240           NertIDs         306           RanTrgData         306           RunParams664         238           SftPkg         254           TofCalSft         307           TofCalSft         314           TofQalconst         307           TofQalEec         315           WirePositionCalibD         522           seed         255           236         245			-14604 8093	-11414 9779	
MdcCalSft         521           MdcCalsft         316           MdcTuning         316           MdcTuning         316           Mctadata         324           MucCalConst         239           MucCalSft         303           NextIDs         240           NextIDs         240           NextIDs         240           NextIDs         240           QualityParams         253           QualityParams         306           Replicas         276           RunParams         294           TofCalConst         307           TofCalSft         314           TofQELec         315           VirePositionCalibD         522           seed         237           test         236           245         245		EsTof	8093	97/9	
MotDataConst         316           MdcTuning         316           MdcTuning         324           MucCalconst         239           MucCalSft         303           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         256           Replicas         276           RunParams         294           RunParams         294           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSinSvc         315           WirePositionCalibD         237           test         236           245         236		Mdc Mdc	25338		
Metadata         324           Metadata         324           MucCalConst         239           MucCalSft         303           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         256           Replicas         276           RunParams         294           RunParams         294           TofCalConst         307           TofCalSft         314           TofSimSvc         315           Seed         237           test         252           236         245				27090	
MucCalConst         239           MucCalSft         303           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         253           RanTrgData         306           Replicas         276           RunParams         294           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           seed         237           test         236           245         236		Mdc	20683	23454	
MucCalSft         303           NextIDs         240           NextIDs_bak         313           OfflineLum         253           QualityParams         306           RanTrgData         306           RunParams664         238           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           test         255           236         246		Mdc	20448	20682	
NextIDs         300           NextIDs_bak         240           NextIDs_bak         313           OffineLum         253           QualityParams         306           Replicas         276           RunParams         294           TofCalConst         307           TofCalConst         307           TofCalSft         314           TofSimSvc         512           WirePositionCalibD         522           seed         253           test         236           246         236		Mdc	27102		6.6.2.b
NextIDs_bak     240       OfflineLum     313       QualityParams     306       Replicas     276       RunParams     294       RunParams     294       Sffkg     254       TofCalConst     307       TofCalSft     314       TofSimSvc     315       Seed     225       seed     2255       236     246		Mdc	24897	25337	
OfflineLum         313           QualityParams         253           RanTrgData         306           Replicas         276           RunParams         294           RunParams         294           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           seed         255           236         236           245         236		Mdc	9810		6.6.2.b
QualityParams         253           RanTrgData         306           Replicas         276           RunParams         294           RunParams664         238           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           test         252           238         236           245         236		Mdc	11414	14604	
Ramingueta         276           Replicas         276           RunParams         294           RunParams         294           Stftkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           Seed         237           test         255           236         245		Mdc	28649		6.6.3.p01
RunParams         294           RunParams         294           RunParams         294           Stfrkg         254           TofCalGonst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           seed         235           test         236           244         245		Mdc	23463	24177	
RunParams664         238           SftPkg         254           TofCalConst         307           TofCalSft         314           TofSimSvc         315           WirePositionCalibD         237           test         256           238         236           245         236		MdcAlign	8046	9809	
SftPkg 254 TofCalConst 307 TofCalSft 314 TofQELec 315 WirePositionCalibD 522 seed 237 test 255 236 245		MdcAlign	24897		6.6.2.b
TofCalConst 307 TofCalSft 314 TofQELec 315 WirePositionCalibD 2237 test 255 236 245		MdcAlign	27102		6.6.2.b
TofCalSft         3107           TofQalSft         314           TofQELec         315           TofSimSvc         522           WirePositionCalibD         2237           test         255           236         245		MdcAlign	28649		6.6.3.p01
TofQELec         314           TofSimSvc         315           WirePositionCalibD         522           seed         237           test         255           236         245		MdcAlign	23463	24177	
TofSimSvc         315           WirePositionCalibD         522           seed         237           test         255           236         245		MdcAlign	11414	14604	
WirePositionCalibD         522           seed         237           test         255           236         236		MdcAlign	20448	23454	
test 255 236 245		MdcAlign	25338		6.6.2.b
236 245		MdcAlign	9810		6.6.2.b
245	6.6.4	MdcData	28649	80000	6.6.3
	6.6.4	MdcData	8093	28648	6.6.2
285	6.6.4	Dedx	8093	25337	6.6.2
	6.6.4	Dedx	28649	80000	6.6.3
282	6.6.4	Dedx	27091	28648	6.6.3

- Calibration constants and tuning parameters of each sub-detector are got from database according to run number
- A complete set of calibration constants and tuning parameters are set in a table in database for each BOSS Release, to make sure the production of simulated data and reconstruction are reproducible



# Data ecosystem: to use the data

- Core capabilities for data-driven science: storage, access, transfer, process, analysis
  - Build with existing technologies while supporting new tech., e.g. AI
- Long time preservation
- Seamless Data and Compute Infrastructure
  - Flexibility with resources and support of time-critical tasks

Technical Concept		
System services <ul> <li>Containers</li> <li>Orchestration</li> <li>User APIs</li> </ul>	Data services <ul> <li>Data catalog</li> <li>Data management</li> <li>Storage interface</li> </ul>	Interface to National Data Center • APIs • Network
Storage <ul> <li>Composable</li> <li>High-performance/hot data</li> <li>Stable/cold data</li> </ul>	<ul> <li>CPU standard units</li> <li>GPU standard units</li> </ul>	• Infrastructure

### AI-empowered data ecosystem for BESIII

Fully explore the potential of BESIII data sets

Fulfil the user requirement in short/mid/long term

<ul> <li>Permanent storage</li> <li>Retrieval</li> <li>Movement</li> <li>Compression</li> </ul>	<ul> <li>Meta data</li> <li>Document</li> <li>Results</li> <li></li> </ul>	• AI model as a more compact and efficient preservation (distributions, detector effects,)	• Low-coding			
Data management	Knowledge repository	Reinterpretatio n/recasting	Automation analysis			
AI+						

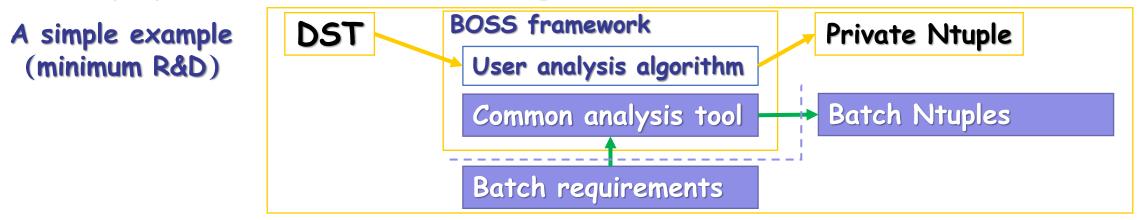
Big data: storage, computing

13

# Data ecosystem: to use the data easily

Further development: Common analysis tool

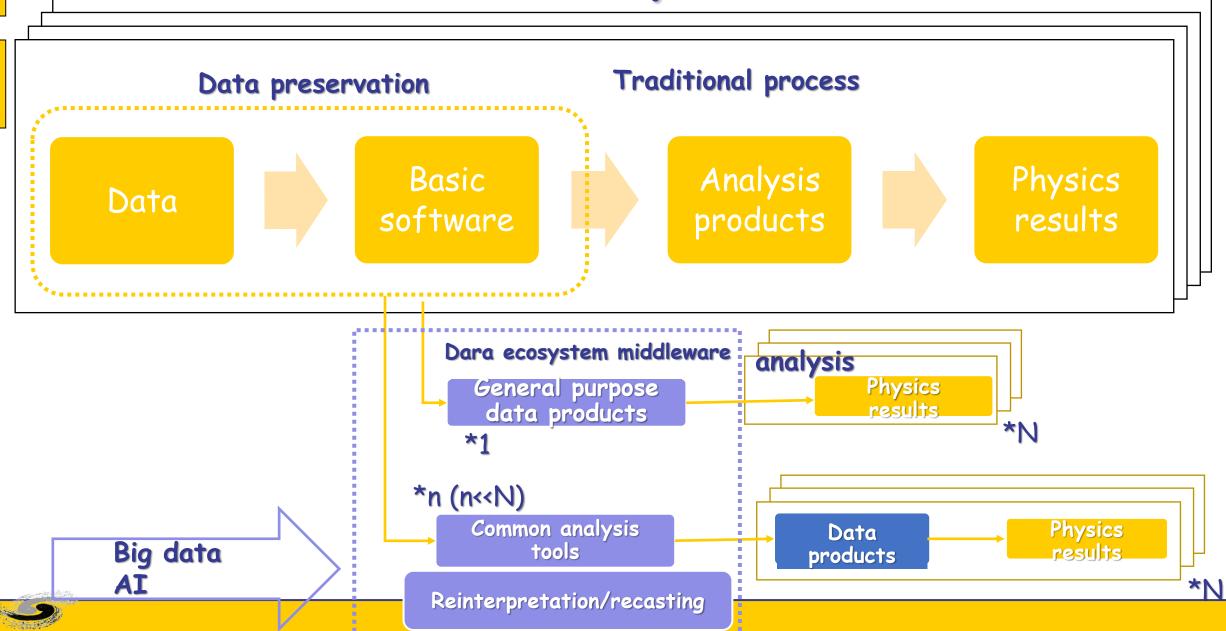
- User-friendly: as a middleware
  - Produce physics information without knowing DST and BOSS
  - Many options. FSFilter can be a good start



- Higher performance: faster I/O, reduced throughput, ...
- Possibility for workflow automation

#### This can be applied to the current analyses

Data ecosystem



**\***N

# Goals and progress

### Goals

- 1-3 years:
  - develop baseline design for data preservation
  - R&D of the data ecosystem
  - establish costs and get more funds for R&D
- 3-5 years:
  - develop baseline design for the full data ecosystem
  - Build it, test it and get ready
- Progress
  - Initial discussion with BESIII Publication Committee
  - Formal discussions with the National HEP Data Center (run by IHEP)
  - Started a task force
  - Applied for new funding from IHEP





## Committee for BESIII data ecosystem

- A commitment of data preservation to the collaboration and the community
- Collective intelligence
  - Define the strategy to fulfil the need for the collaboration
  - Expertise to solve key issues
  - Develop the guidelines for data usage after BESIII shutdown
    - Members/global fits/theorists/...
  - Outreach
- Coordinate the resource collaboration-wide (mainly manpower)



# Timeline

### 2024-2030

#### BESIII data taking

- Data analysis by BESIII collaboration groups
- Current computing
  - resources

DPHEP level 4
Research for future data structure

### 2030-2035

#### BESIII shutdown

- BESIII groups / Global fit
- Limited resources (local+cloud)

### Fast data search and I/O AI-based data compression

### 2035-

#### Virtual BESIII

- BESIII groups / Global fit
- Analysis on demand

- AI-based automatic data analysis
- Integrate to large data ecosystem

IHEP will establish a dedicated project to support the BESIII data preservation. We are also seeking more supports from funding agencies.



### Thanks

