



JUNO raw data management

Xiaomei Zhang

On behalf of JUNO collaboration

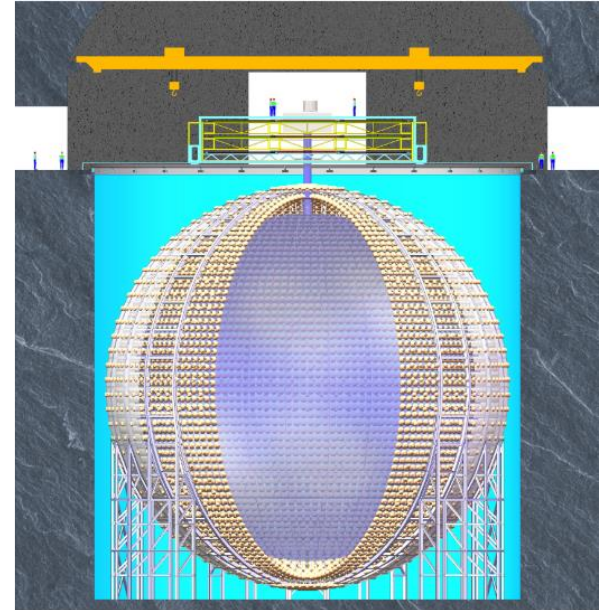
Institute of High Energy Physics

ACAT 2024

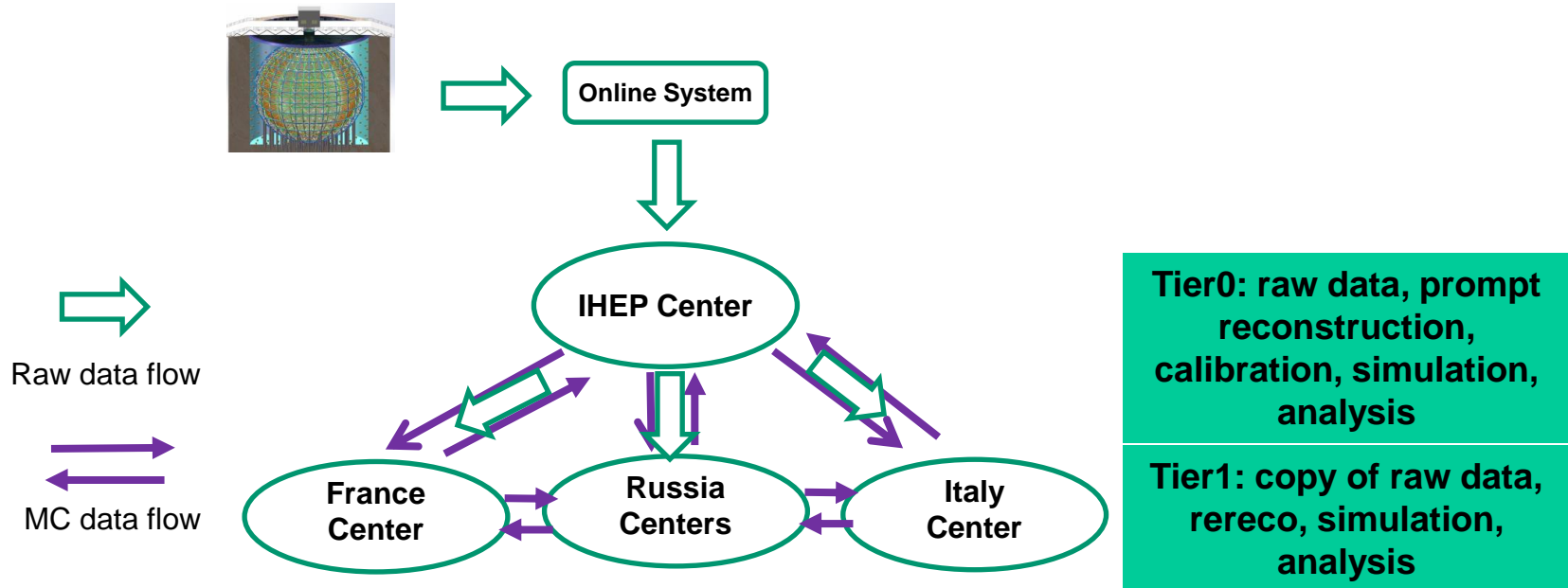
2024-3-12

JUNO experiment

- ❖ A multi-purpose neutrino experiment
 - Measure neutrinos (solar neutrinos, supernova neutrinos, atmospheric, geo) mass hierarchy and mixing parameters
 - Located at Jiangmen, South of China
 - Expect to take data in 2024
- ❖ JUNO-TAO is a satellite detector
 - Precisely measure reactor energy spectrum, improve sensitivity of JUNO on mass hierarchy study



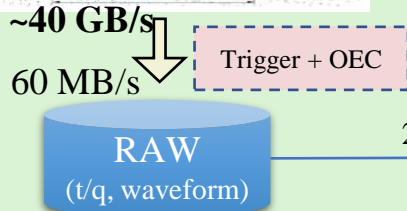
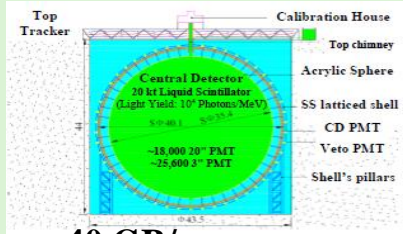
JUNO distributed computing (DCI)



- ❖ JUNO DCI was built based on **DIRAC**, use **both WMS and DMS**
- ❖ Five data centers joined: IHEP, CC-IN2P3, INFN-CNAF, JINR, MSU
- ❖ **Raw data** flows from Online to IHEP which then distributes to other centers
- ❖ **1st Reconstruction and Calibration** will run in IHEP
- ❖ **MC Simulation, 2nd Reconstruction and Analysis** will run in all data centers

RAW data processing

JUNO onsite



DAQ Event Rate:
1kHz

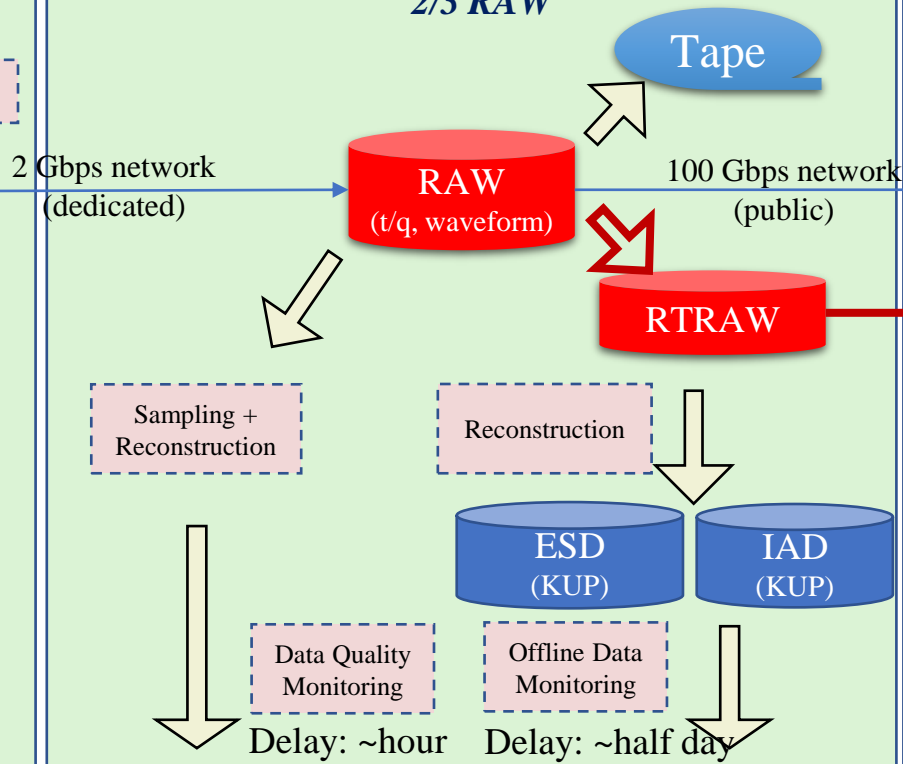
Data Volume:
60 MB/s

Each file size:
5 GB/file

N (Events) per file:
83,333 (=5GB /
60MB/s * 1kHz)

IHEP

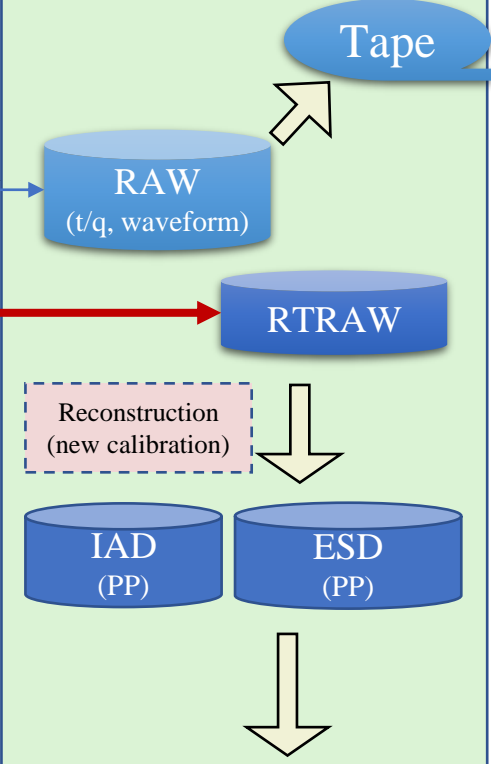
RAW data Volume: *5.2 TB/day, 2 PB/year*
Number of files: *~1k/day, 400k/year*
RTRAW data Volume: ROOT format of RAW,
2/3 RAW



JUNO Data Monitoring Web

JUNO Data centers

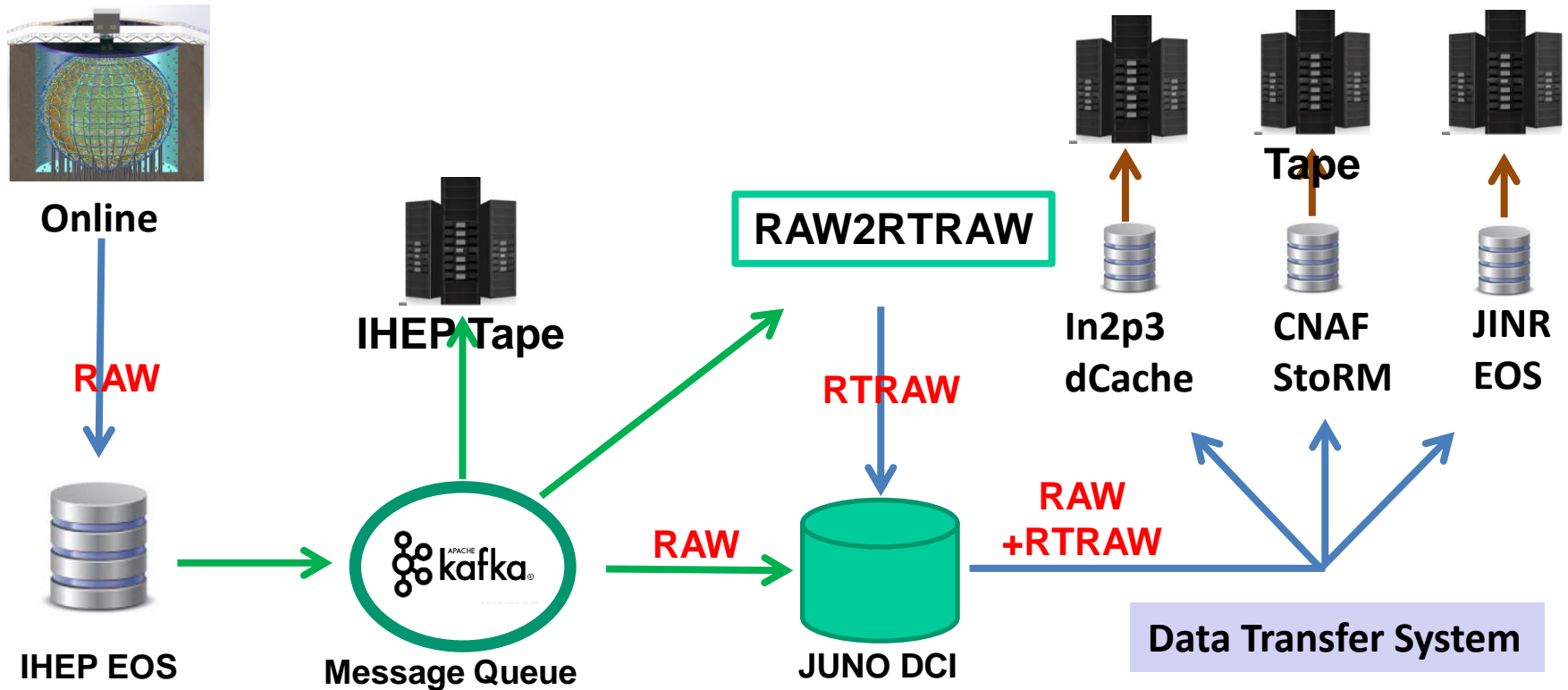
Requirements: one year data
needs to be reconstructed
within 2 months.



Physics Analysis

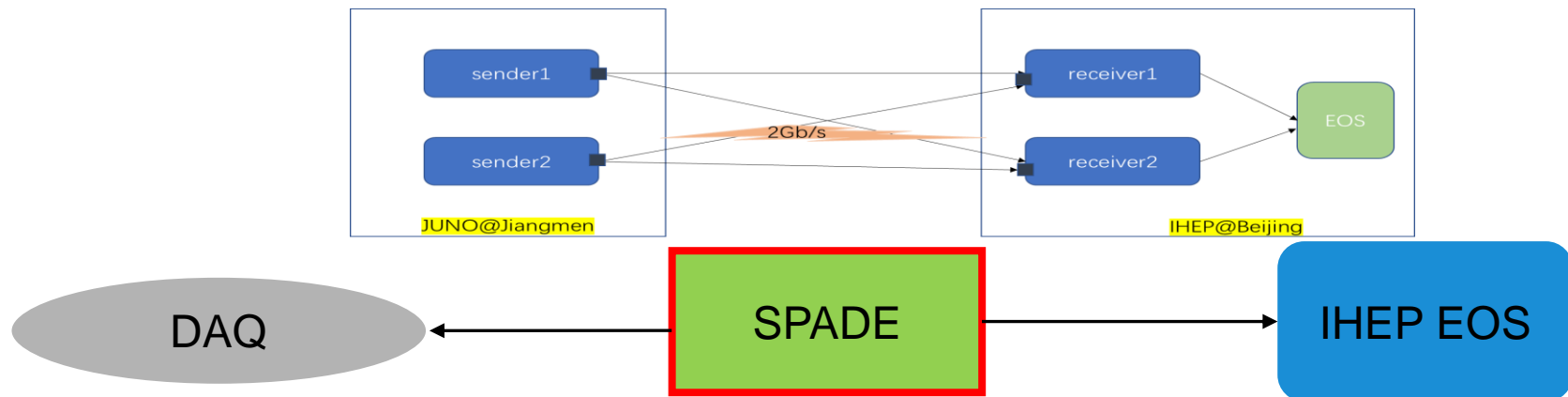
RAW data replication

- ❖ RAW data replicated to IHEP EOS disk system and sent info to MQ
- ❖ With RAW info from MQ, data archived in Tape, converted to RTRAW
- ❖ Both RAW and RTRAW are registered in DCI and replicated to data centers



Replication from Online to IHEP

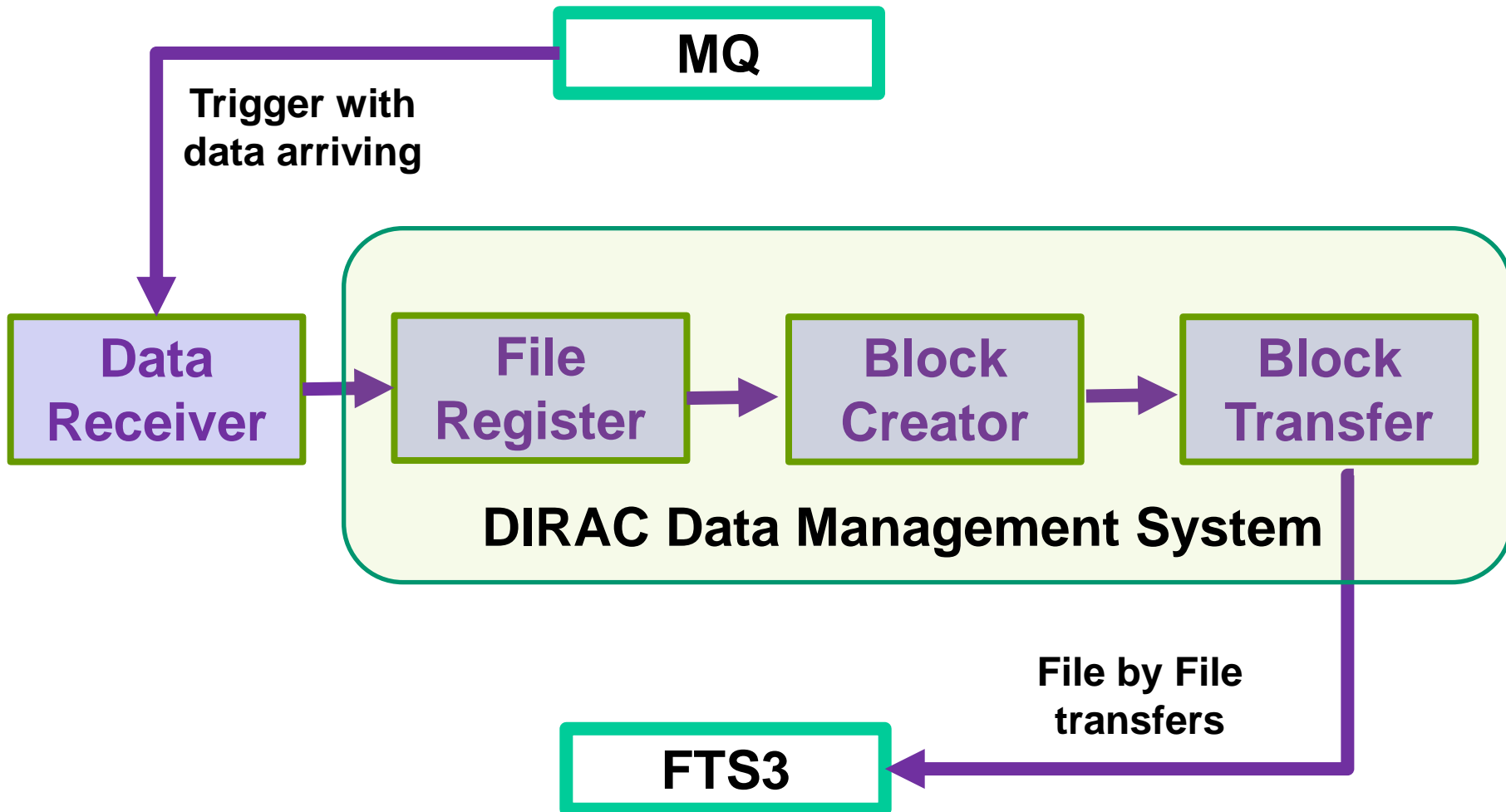
- ❖ **SPADE** -- non-grid transfer system is used for online->IHEP replication
 - Deployed with double-server mode both in Jiangmen and IHEP
 - 100TB onsite local cache is planned to support 15 days' cache
 - This system is also used for other experiments in IHEP
- ❖ **2Gb/s dedicated link** is planned from onsite to IHEP
- ❖ Raw data name space is planned
 - Directory: /<EOSROOT>/juno/raw/YYYY/mmdd/filename
 - Filename: Daq.RunNumber.StreamType.StreamName.FileTag.FileTime.FileSequence.dat



Replication from IHEP to data centers

- ❖ **Grid data transfer system based on JUNO DCI** is developed to implement raw data flow from IHEP to data centers
- ❖ Design goal
 - Implement in a fully automated way
 - MQ -> registration in DCI -> replication in DCI -> archive in Tape
 - Automated checking and validation
 - Easy troubleshooting and error warning
 - Provide monitoring dashboard for shifters

Architecture for data transfer in DCI



RAW and RTRAW data in DFC

- ❖ DFC (Dirac File Catalogue) is used for both metadata catalogue and replica catalogue
- ❖ All the RAW data in data centers will be registered in DFC
- ❖ When files arriving in data centers, file replicas and blocks are visible and validated through DFC
- ❖ Files in disk and tape system can be operated directly through DFC in a longer term with tools developed

```
DFC:/juno/Commissioning_Dryrun/DTS/202212/7750>replicas 7323_0_12-09-2244_1008_7750_-  
fn: /juno/Commissioning_Dryrun/DTS/202212/7750/7323_0_12-09-2244_1008_7750_-  
CNAF-STORM /juno/Commissioning_Dryrun/DTS/202212/7750/7323_0_12-09-2244_1008_7750_-  
IN2P3-DCACHE /juno/Commissioning_Dryrun/DTS/202212/7750/7323_0_12-09-2244_1008_7750_-  
JINR-EOS /juno/Commissioning_Dryrun/DTS/202212/7750/7323_0_12-09-2244_1008_7750_-  
IHEP-JUNOEOS /juno/Commissioning_Dryrun/DTS/202212/7750/7323_0_12-09-2244_1008_7750_-
```

Data Receiving and Triggering



- ❖ DataReceiver is implemented as a independent Daemon
 - Listen to and accept messages from MQ
 - Decode messages from MQ to get file info
 - “FileName”, “FilePath”, “Checksum”
 - Trigger DCI registration and transfer process
 - Basic data check and return status to MQ
 - File size, checksum.....
 - Automatic Email warning on critical problems
- ❖ Separated daemons for RAW and RTRAW

Block Creating

- ❖ Block level transfer, a block as a transfer unit
 - Easy to control process of transfer and check status of transfer
- ❖ Three level management: File -> block -> dataset
 - A group of files as a block, depending on transfer efficiency
 - A group of blocks as a dataset, depending on physics need
- ❖ Define Block/Dataset with DFC metadata
 - Block use two metadata: Block and Type
 - Block metadata example: {'Block': '20230302', 'Type': 'Raw'}

Block Transferring

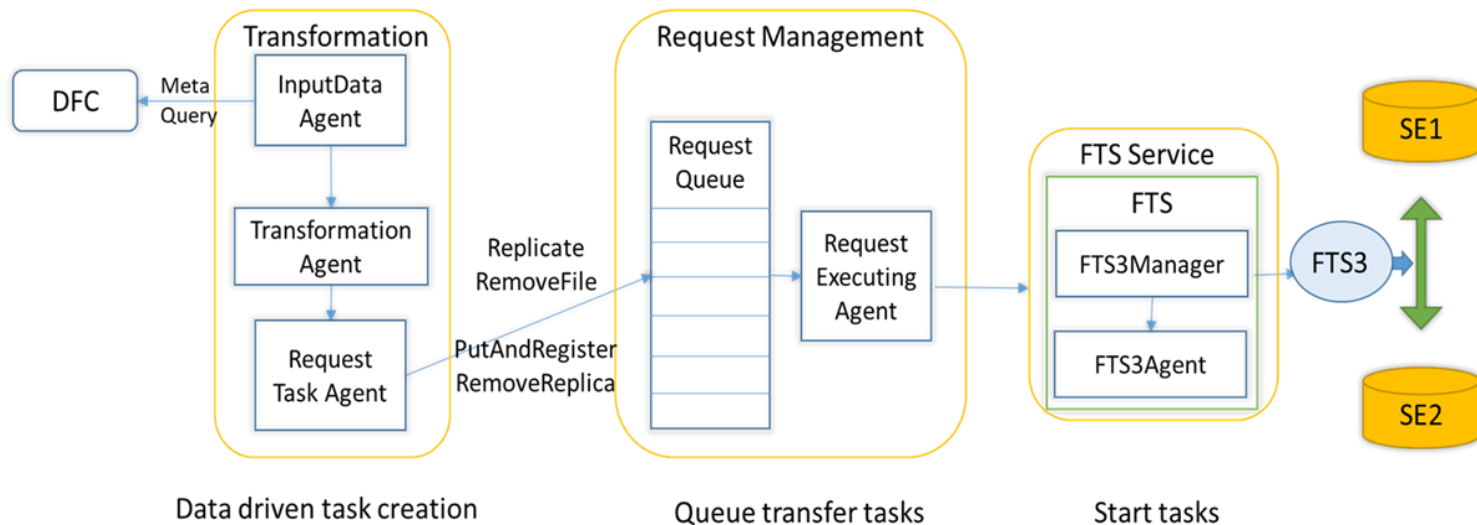
- ❖ Transfer tasks are created automatically block by block
 - **Separated transfers for each data center** (CNAF, JINR, IN2P3), transfer in parallel, avoid single failure and affecting each other
 - No source SEs are defined, allowing multi-channel transfers in case of problems with certain channels

<input type="checkbox"/>	ID	Status	AgentT...	Ty...	Name	Files	Processed (%)
☐ Request: 0							
<input type="checkbox"/>	3972	■ Active	Automatic	Tr...	Commissioning-20230404-JINR-EOS	15	100.0
<input type="checkbox"/>	3971	■ Active	Automatic	Tr...	Commissioning-20230404-CNAF-STORM	15	100.0
<input type="checkbox"/>	3970	■ Active	Automatic	Tr...	Commissioning-20230404-IN2P3-DCACHE	15	100.0

- ❖ Retransfer files in block level for each task in case of problems
- ❖ Provide check and validation block by block

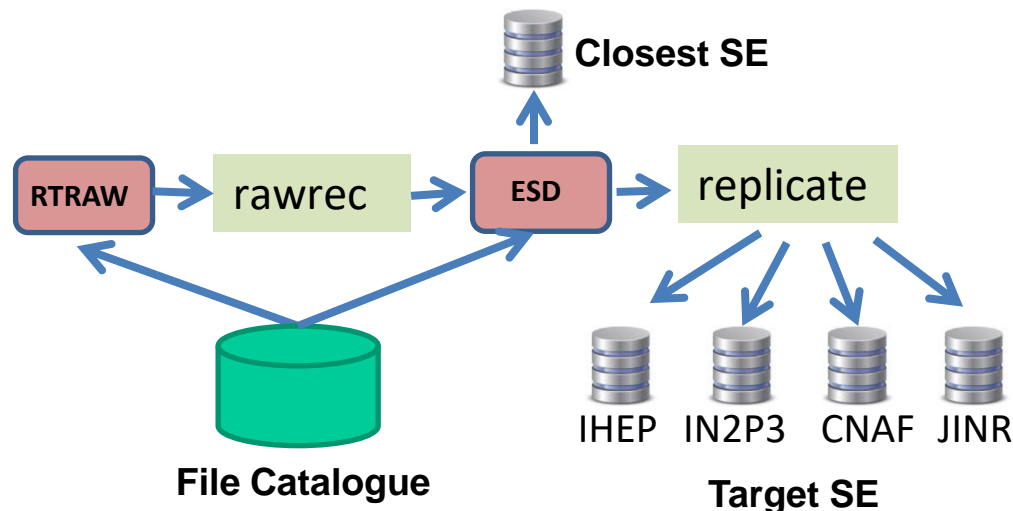
Data-driven Transfer with DIRAC DMS

- ❖ Block level transfer is implemented based on DIRAC DMS
 - Use data-driven transformation infrastructure (TS)
- ❖ Transfer tasks are started in a data-driven way when data is available
 - Don't need to wait files to be fully registered in blocks
 - File register and block transfer will run in parallel



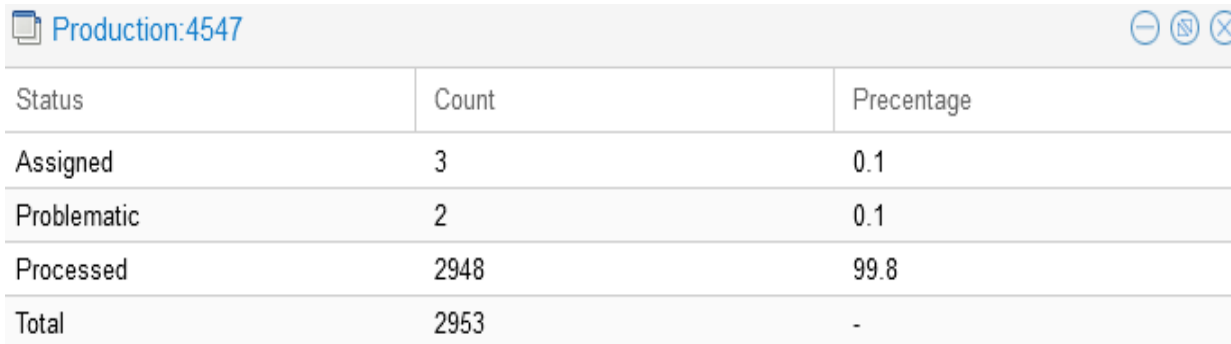
Data-driven Re-Reco system

- ❖ Re-RecoSys was developed based on DIRAC transformation system
- ❖ An automatic chain created to manage the workflow and dataflow
 - “rawrec” workflow Input: RTRAW, output:ESD
 - “replicate” dataflow Input:ESD, transfer data to IHEP, CNAF, JINR
- ❖ The chain is triggered by availability of RTRAW data in DFC
- ❖ The two step are connected by ESD data in DFC



Monitoring and Accounting

- ❖ Mainly use information from DIRAC and FTS
- ❖ Block-level monitoring is using DIRAC TS monitoring
- ❖ File-level monitoring
 - Real-time monitoring using FTS
 - History of FTS monitoring is managed through ActiveMQ+Logstash+ElasticSearch+Grafana



Status	Count	Percentage
Assigned	3	0.1
Problematic	2	0.1
Processed	2948	99.8
Total	2953	-

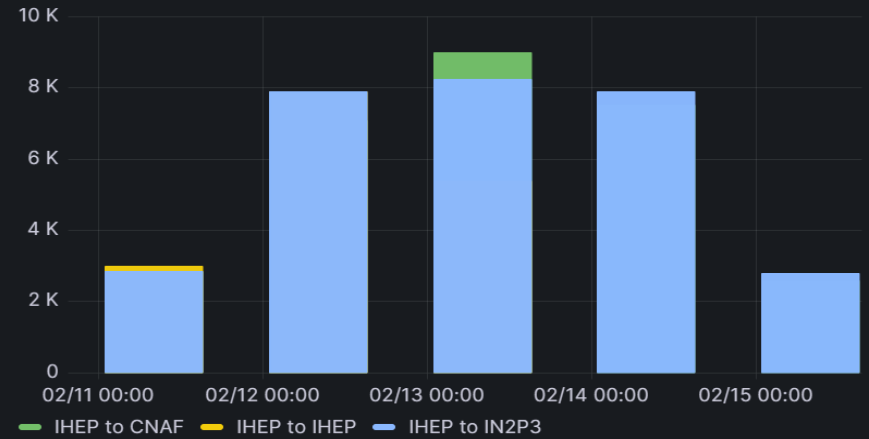
Grafana dashboard for DCI transfers

~ total

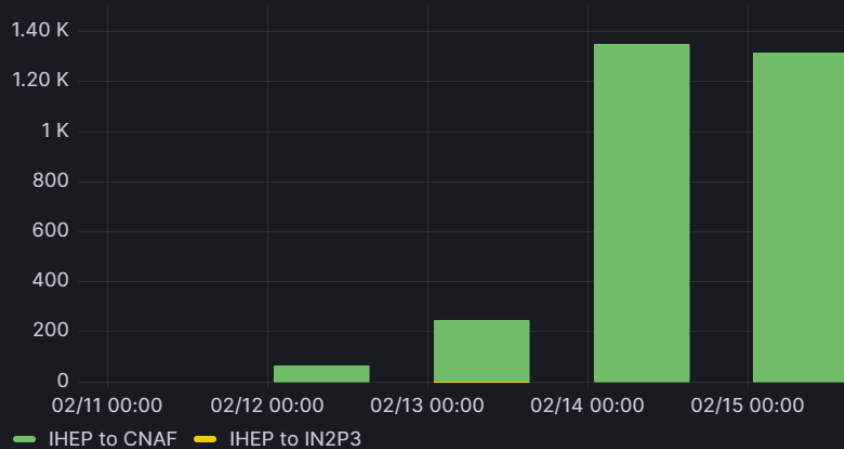
Efficiency



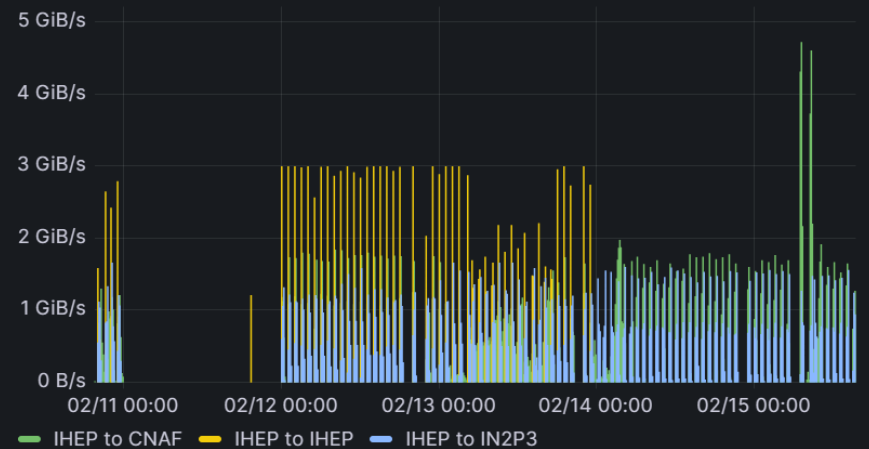
Transfer Status OK



Transfer Status Error

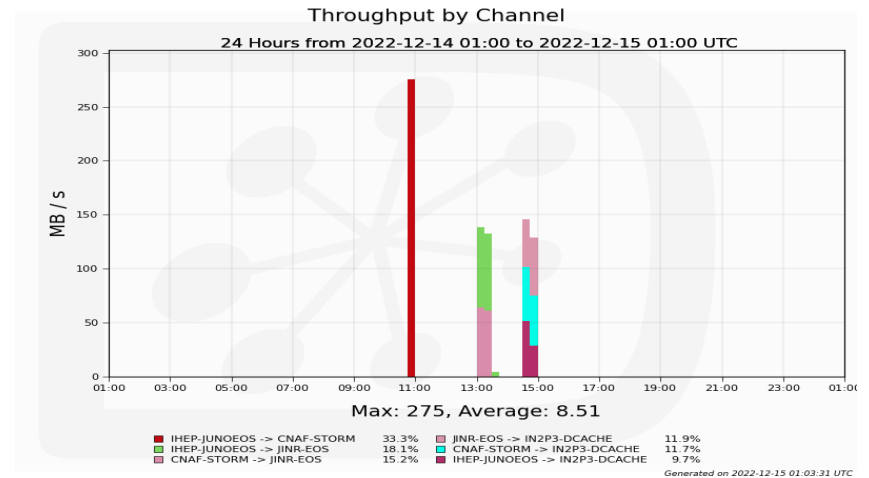


Transfer Throughput 10m



RAW data flow tests

- ❖ Started to be tested with commissioning data
 - First test was started in Dec 14, 2022
 - Full functions was tested since 2023
- ❖ ~80000 files, ~25TB data are transferred
 - last about one year, no human intervention needed
 - Some issues found and fixed in 2024 transfers
 - Bad files, 0 file size, too large files (>10GB).....



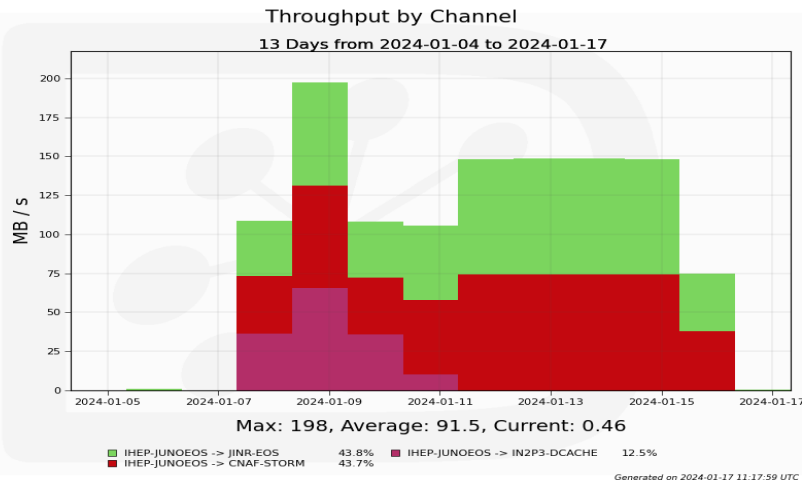
```
FC:/juno/Commissioning_Dryrun>size -l  
directory: /juno/Commissioning_Dryrun  
Logical Size: 25,681,379,886,114 Files: 84129 Directories: 202
```

	StorageElement Size	Replicas	
1	CNAF-STORM	25,680,875,114,937	83501
2	IN2P3-DCACHE	25,249,297,231,841	83499
3	JINR-EOS	24,653,860,054,209	83493
4	IHEP-JUNOEOS	25,681,379,886,114	84129
	Total	101,265,412,287,101	334622

RTRAW flow tests

- ❖ RTRAW transfer was tested during the first offline data challenge
- ❖ 100K files were **registered** in DFC
- ❖ 7 blocks, 103TB data were **transferred** to data centers in time **with 100% success rate**
 - One week's data, each block has 14400 files
 - CNAF and JINR 100%, IN2P3 about 1/3

<input type="checkbox"/>	4535	Active	Automatic	Transfer-J...	RTRaw-20231007-JINR-EOS	14400	100.0
<input type="checkbox"/>	4534	Active	Automatic	Transfer-J...	RTRaw-20231007-CNAF-STORM	14400	100.0
<input type="checkbox"/>	4533	Active	Automatic	Transfer-J...	RTRaw-20231006-JINR-EOS	14400	100.0
<input type="checkbox"/>	4532	Active	Automatic	Transfer-J...	RTRaw-20231006-CNAF-STORM	14400	100.0
<input type="checkbox"/>	4531	Active	Automatic	Transfer-J...	RTRaw-20231005-JINR-EOS	RTRaw-20231006-CNAF-STORM	0.0
<input type="checkbox"/>	4530	Active	Automatic	Transfer-J...	RTRaw-20231005-CNAF-STORM	14400	100.0
<input type="checkbox"/>	4525	Active	Automatic	Transfer-J...	RTRaw-20231004-JINR-EOS	14400	100.0
<input type="checkbox"/>	4524	Active	Automatic	Transfer-J...	RTRaw-20231004-CNAF-STORM	14400	100.0
<input type="checkbox"/>	4522	Active	Automatic	Transfer-J...	RTRaw-20231003-JINR-EOS	14400	100.0
<input type="checkbox"/>	4521	Active	Automatic	Transfer-J...	RTRaw-20231003-CNAF-STORM	14400	100.0

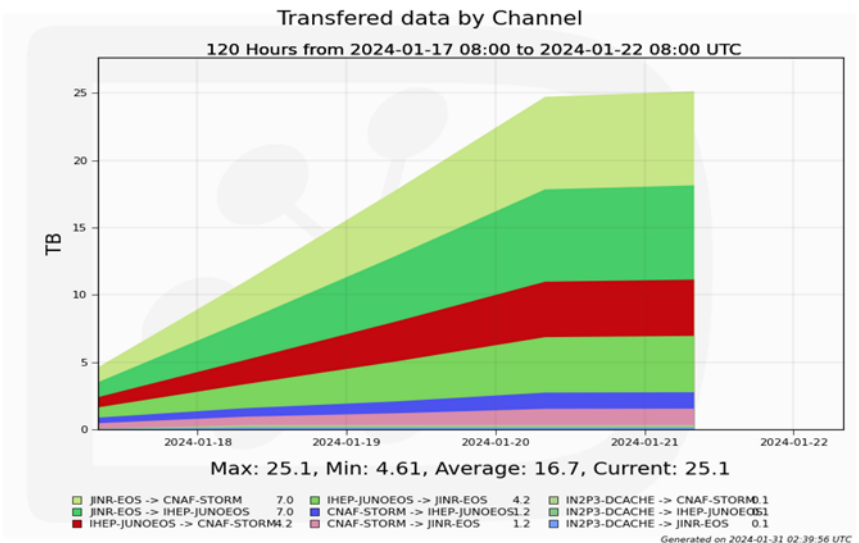
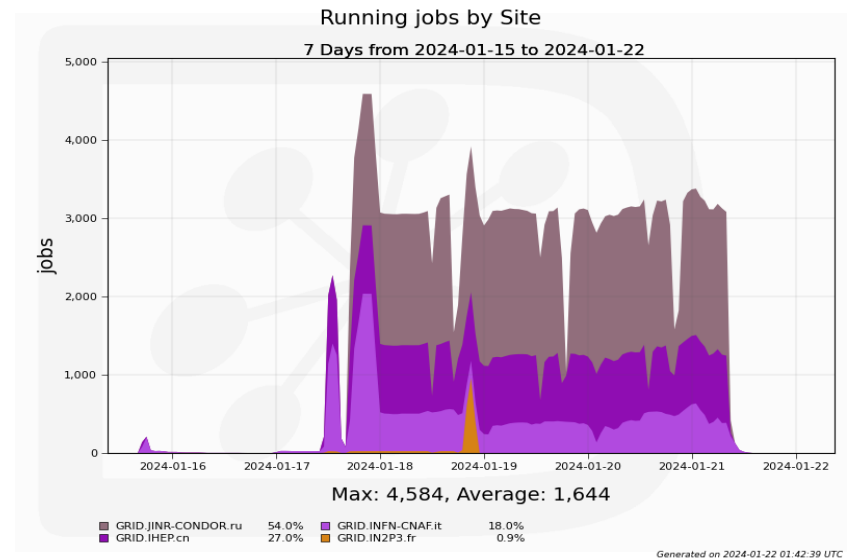


```
FC:/juno/groups/Production/rc5-dc1/2023>size -l
directory: /juno/groups/Production/rc5-dc1/2023
Logical Size: 44,979,377,646,977 Files: 100800 Directories: 7
```

StorageElement	Size	Replicas
1 CNAF-STORM	44,979,377,646,977	100800
2 IN2P3-DCACHE	12,810,923,611,030	28800
3 JINR-EOS	44,979,377,646,977	100800
4 IHEP-JUNEOES	44,979,377,646,977	100800
Total	147,749,056,551,961	331200

RTRAW Re-Reco tests

- ❖ The test started when all RTRAW is ready in DCI and lasted for 4 days
- ❖ Total ~100K Re-Reco jobs were assigned and run in four data centers
 - Re-Reco jobs run in 8-core mode
 - Complete with 99.9% success rate, no error found in grid side
- ❖ 25TB data were transferred to three data centers



Summary and Plan



- ❖ The prototype of JUNO raw data transfer and processing system is ready
- ❖ The RAW-related offline systems is well connected to each other through MQ
- ❖ DIRAC DMS provides an easy solution on automated raw data management system for JUNO
- ❖ More pressure tests simulating data-taking will be carried out in 2024



❖ Thank you for listening!