PROCEEDINGS OF SCIENCE

China-Europe data transfer challenge with upgrading LHCONE Network at IHEP

Yifan Li,[a,b,*] **Xuantong Zhang**[a] **and Xiaomei Zhang**[a]

[a] *Institute of High Energy Physics, Chinese Academy of Science,*
19B Yuquan Road, Shijingshan District, Beijing, China.

[b] *University of Chinese Academy of Sciences,*
1 Yanqihu East Road, Huairou District, Beijing, China.

E-mail: liyifan@ihep.ac.cn

The data center of the Institute of High Energy Physics, Chinese Academy of Sciences(IHEP, CAS), stores a large number of experimental data from major scientific instruments such as the Jiangmen Neutrino Observatory, which are transferred and backed up among the collaborating data centers. IHEP Data Center upgraded the transfer link between IHEP and GEANT from the original 10Gbps to 100Gbps, LHC Open Network Environment as part of link, its network has also become better. In order to prevent large-scale transfer failures and ensure transfer quality, it is necessary to conduct transfer tests to understand the transfer performance. In the previous transfer tests, only multi- parallel transfers using "gfal-copy" were conducted, without high-bandwidth testing or utilizing the CERN File Transfer Services for transfers in a real transfer environment. This leads to a lack of understanding of the actual application bandwidth situation after the network upgrade. To solve this problem, we have done data transfer challenges and tests. This study introduces the previous data transfer challenge, large bandwidth transfer test, and data transfer challenge in a real transfer environment. We developed a data injection tool for the data transfer challenge in a real environment. Using this tool, we can adjust the data transfer volume and the intervals between data transfer, among other parameters. The results of the data transfer challenge were viewed through transfer monitoring deployed at IHEP.

International Symposium on Grids and Clouds (ISGC2024)
24 -29 March, 2024
Academia Sinica Computing Centre (ASGC), Institute of Physics, Academia Sinica Taipei, Taiwan

*Speaker

PoS(ISGC2024)036

1. Introduction

IHEP serves as a tier-1 and tier-2 site of the WLCG[1]. Many experiments at IHEP, such as JUNO [2], Belle II [3], etc., require large-scale data transfer between China and Europe, which demands high-quality network connection and ample bandwidth support.

Therefore in September 2023, IHEP upgraded the network services between IHEP and GEANT, increasing the bandwidth from 10G to 100G. It is necessary to understand the transfer performance of the upgraded network. The JUNO experiment is nearing the data taking stage. An extensive distributed computing infrastructure (DCI) [4] has already established. Some sufficient test results before network upgrade have been accumulated. Therefore, we decide to use JUNO experiment DCI to complete this data transfer challenge. In JUNO experiment, data transfer primarily occurs from China to Europe, with the involved data centers mainly including CNAF, IN2P3, and JINR. The data transfer challenge needs to assess the performance of the transfer between IHEP and the other three data centers. The JUNO experiment is expected to generate approximately 3PB of data annually, equivalent to around 102MiB/s [5]. Figure 1 shows the current state of the network connection for the JUNO experiment among all JUNO data centers after the upgrade. The upgraded 100Gbps network bandwidth only supports the transfer between IHEP and CNAF/IN2P3. From IHEP to JINR, the network bandwidth is still 10 Gbps. To understand the performance after the network upgrade in actual application, we conducted a data transfer challenge between China and Europe. In this study, we will introduce how to conduct the data transfer challenge between China and Europe, the developed data injection tool, and the results of the data challenge.

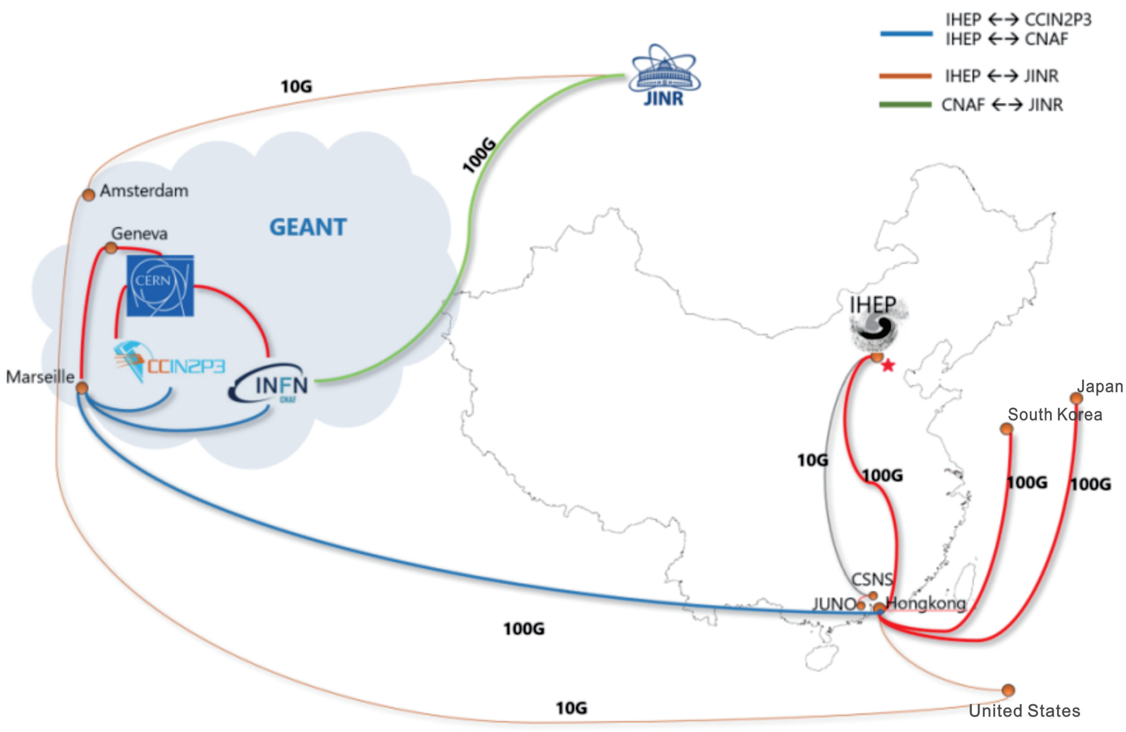

Figure 1: International Network Connection of JUNO

2. Previous data transfer challenge

In the previous data transfer challenge, the data was transferred by "gfal-copy". Transfer between IHEP and JINR/IN2P3 sites was tested by 50 parallel quantities at the same time. Transfer between IHEP and CNAF was tested by 10, 40, and 100 parallel quantities.

In the study of [6], the data transfer challenge used the "gfal-copy" transfer tool, and the results showed a lot of transfer failures between IHEP and CNAF. In the direction from CNAF to IHEP, there were 80% transfer failures with 10 parallel transfers, reaching a maximum bandwidth of 0.28 Gbps. There were 75% transfer failures with 40 parallel transfers, reaching a maximum bandwidth of 1.7 Gbps. And with 100 parallel transfers, there were 75% failures, reaching a maximum bandwidth of 1.92 Gbps. In the direction from IHEP to CNAF, there were 20% transfer failures with 10 parallel transfers, reaching a maximum bandwidth of 1.97 Gbps. There were 20%transfer failures with 40 parallel transfers and the maximum bandwidth was 6.14 Gbps. Furthermore, the failure rate with 100 parallel transfers was 20% and the maximum bandwidth was 7.79 Gbps. These transfer failures were attributed to unstable network connections. Before the network upgrade, we used the same methods and tools as the study of [6] to conduct data transfer tests between IHEP and JINR/IN2P3 sites, with upgraded 100Gbps network to see if the results could be improved.

In order to add the results between IHEP and JINR/IN2P3, which were lacking in the study of [6], we used the same methods and tools as those in the study of tests between IHEP and JINR/IN2P3 sites before the network upgrade. The results indicated significant fluctuations in data transfer in all directions (Table 1, 2). The transfer rate from IN2P3 to IHEP fluctuated between 1.6 Gbps and 8 Gbps, and the single file transfer rate from IHEP to IN2P3 fluctuated between 3.4 Gbps and 6.32 Gbps. The transfer rate from JINR to IHEP fluctuated between 0.59 Gbps and 5.18 Gbps. From JINR to IHEP, there were only 41 out of 50 parallel transfers successfully transferred in the test. The 9 failed transfers were due to network instability causing connection timeouts. The single file transfer rate from IHEP to JINR fluctuated between 1.36 Gbps and 2.4 Gbps. Our tests and the tests in the study of [6] were not satisfactory, indicating that the network connections were unstable before the upgrade and could not meet the transfer requirements for the JUNO experiment, which is 3PB per year or 102MiB per second [6].

Throughput (Gbit/s)		**To IHEP**
From	**CNAF**	10 parallel transfers:0.28 (80% transfer failures) 40 parallel transfers:1.7 (75% transfer failures) 100 parallel transfers:1.92 (75% transfer failures)
	IN2P3	50 parallel transfers:1.60-8.00
	JINR	41 parallel transfers:0.59-5.18

Table 1: Test results of other data centers to IHEP before LHCONE [7] upgrade. In our test, the 50 parallel transfers from IN2P3 to IHEP were successful. From JINR to IHEP, there were 9 transfer failures, failure information was DESTINATION CHECKSUM Connection timed out. These failures are attributed to unstable network connection. The data from CNAF to IHEP are from the previous study, which has already reported in [6].

Throughput (Gbit/s)		**From IHEP**
		10 parallel transfers:1.97 (20% transfer failures)
	CNAF	40 parallel transfers:6.14 (20% transfer failures)
To		100 parallel transfers:7.79 (20% transfer failures)
	IN2P3	50 parallel transfers:3.40-6.32
	JINR	50 parallel transfers:1.36-2.40

Table 2: Test results of IHEP to other data centers before LHCONE upgrade. The data from IHEP to CNAF are from the previous study, which has already reported in [6]. In our test, the 50 parallel transfers from IHEP to IN2P3 and JINR were all successful.

3. JUNO 2024 data transfer challenge

Data Challenge 2024 (DC24) is a challenge organized by CERN's WLCG to prepare for the High-Luminosity LHC (HL-LHC) experimental operations [8]. To verify whether European sites can still meet the transfer requirements of the JUNO experiment under high load when various LHC experiments are being tested simultaneously, we conducted the JUNO Data Challenge concurrently with DC24.

In the previous data transfer challenge, the test has high fail rate, short test duration and insufficient test content. To better understand the actual performance of data transfer after the upgrade, it is significant to conduct the JUNO 2024 data transfer challenge.

3.1 Large bandwidth data transfer test

In order to understand the large bandwidth performance of the network after the upgrade, a large bandwidth data transfer test was conducted based on the existing DCI. The testing utilized the same "gfal-copy" transfer method as the previous data transfer challenge. We got the test results that the bandwidth of a single file transfer is low, but the transfer bandwidth can be improved by increasing the number of parallel transfers. Because the single-direction transfer stations have I/O performance bottlenecks that prevent full utilization of the network bandwidth, we decided to use a method of parallel transfers from multiple sites to maximize the utilization of the accessible bandwidth. The test changes from 1000 parallel transfers in one direction to 500 parallel transfers in each direction of IHEP to CNAF and IHEP to IN2P3.

In high-bandwidth data transfer test from IHEP to CNAF and from IHEP to IN2P3, there were 3 failures and 36 failures, respectively. The failures were almost SSL handshake failed: connection timed out. There were about 15 routes between IHEP and CNAF/IN2P3, and the instability of the long-distance routing network caused these transfer failures. In the actual transfer environment, FTS [9] tool was used for data transfer. Once the FTS retry function is enabled, the impact of such failures will greatly reduce.

The transfer rate from IHEP to CNAF is 2949.2 MB/s, and the transfer rate from IHEP to IN2P3 reaches 3022.4 MB/s. These two transfer directions share a common link, with a total transfer rate of approximately 47Gb/s(transfer rate = (2949.2 + 3022.4) * 8 / 1000). Figure 2 shows the IHEP LHCONE traffic-out monitoring during our test, the max transfer rate reaches around 47.90Gb/s.

Throughput		From IHEP
To	CNAF	2949.2 MB/s (0.6% transfer failures)
	IN2P3	3022.4 MB/s (7.2% transfer failures)

Table 3: High-Bandwidth Data Transfer Test after LHCONE Upgrade. This table is the transfer rate of 500 parallel transfers from IHEP to CNAF and IN2P3.

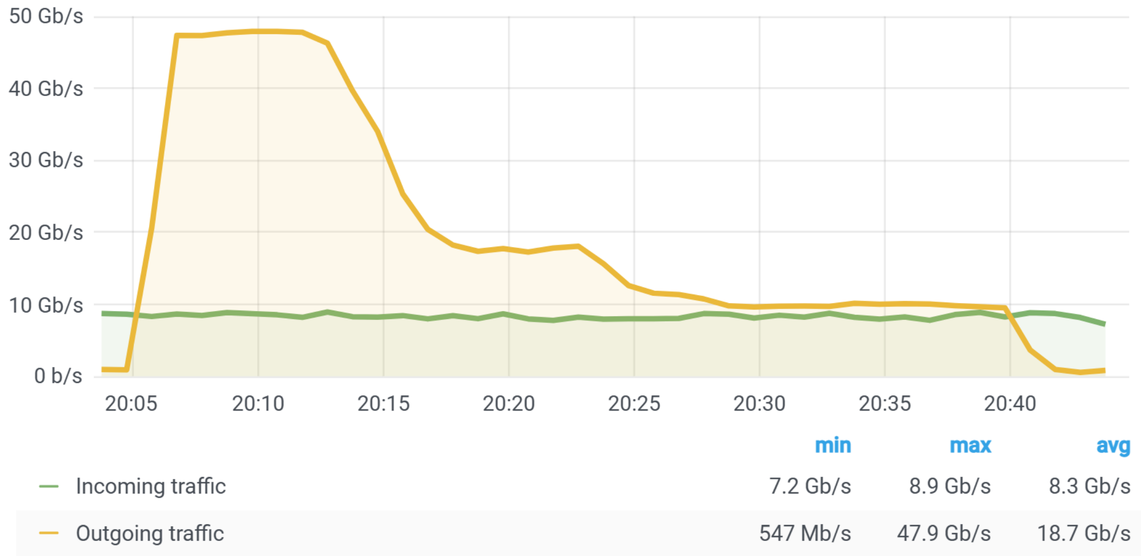

Figure 2: IHEP LHCONE traffic monitoring chart. The x-axis is time, y-axis is network bandwidth (Gb/s). The green and orange curves are IHEP LHCONE incoming and outcoming traffic, respectively. The min, max and avg are the minimum bandwidth, the maximum bandwidth and the average bandwidth within the time range depicted in the graph.

3.2 The mimic data transfer challenge based on FTS

Mimic Data Transfer is a data transfer that we built to simulate the transfer environment in reality. It is equipped with the same facilities, tools, and conditions as the real transfer environment.

The large bandwidth data transfer used test "gfal-copy" transfer method, which is not the real transfer environment for JUNO. In real transfer environment, we used FTS for data transfer. FTS has retry functionality and an optimizer to perform retries, while "gfal-copy" does not have these function. Furthermore, IHEP has deployed FTS for transfer monitoring, providing a more intuitive way to view the results of the transfer. Therefore, it is necessary to conduct a mimic data transfer challenge based on FTS.

3.2.1 Data injection tool

To better control the data transfer challenge, we developed a tool that can control the average bandwidth achieved by injecting intermittent pauses and injecting additional data volume within an hour. This tool uses the FTS-REST API [10] to submit transfer jobs in bulk, and then FTS transfers files to the destination data center. When submitting the jobs, the tool sets the maximum number of retries to 3, allowing for automatic retries in case of transfer failures. Additionally, the tool assigns each file a random name to prevent transfer failures due to the same file names. Before injection, the tool deletes previously transferred files to avoid taking up more space. The data injection tool flowchart is shown in Figure 3. The link for this tool is located at [11].

- Delete transferred files: Delete previously transferred files in the destination directory to prevent excessive use of data center space.
- Read config file: Get the source path, destination path, average speed, injection interval, and additional injected data percentage from the configuration file.
- Calculate the injection volume and frequency: Calculate the volume and frequency of data injection based on the injection interval and the volume of additional injected data.
- Generate and submit transfer jobs: Create transfer jobs and submit them to FTS for execution.

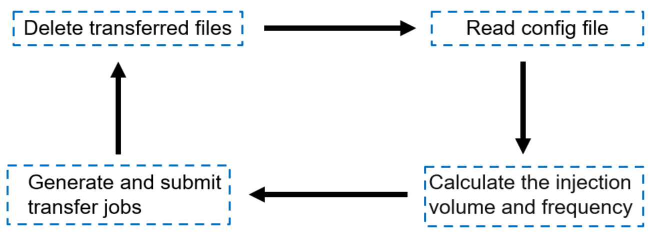

Figure 3: The flowchart of the data injection tool

3.2.2 Data transfer challenge result visualization

The FTS monitoring process shown in Figure 4 involves submitting transfer tasks to FTS by using the data injection tool. Then FTS transfers the files to the destination data center, and upon completion, the generated logs are written into Elasticsearch. The transfer data from Elasticsearch is queried using Grafana to create the monitoring interface. FTS monitoring metrics include transfer efficiency (success number/transfer number), status, throughput (transfer volume per hour), file size, file number, and logs of failed transfers. By utilizing FTS monitoring, the results of the data challenge can be observed(Figure 5).

Figure 4: FTS Monitoring Process Flowchart

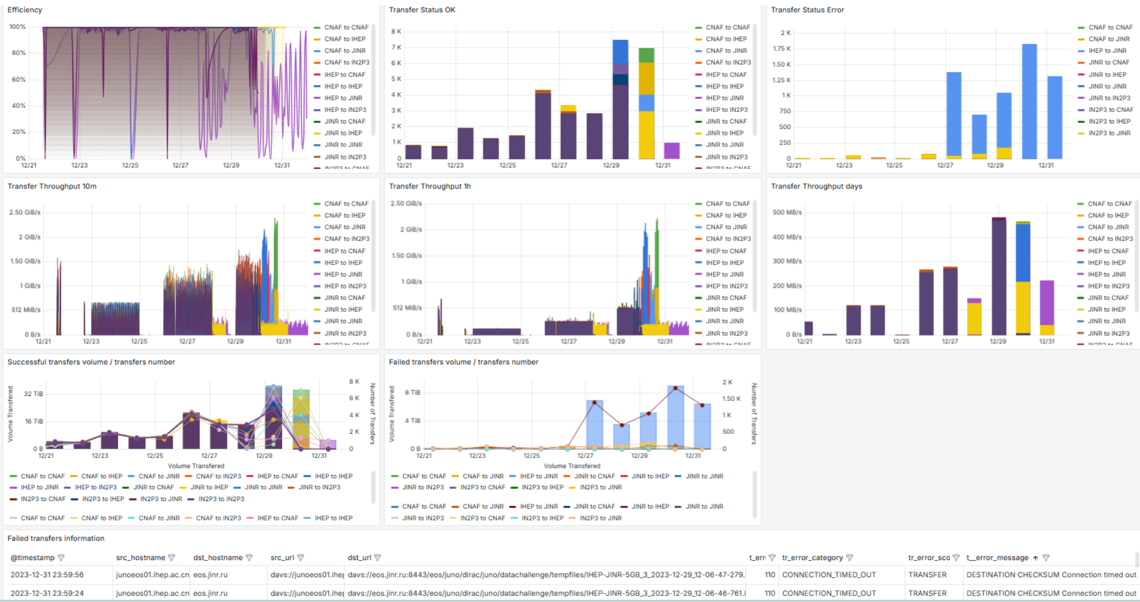

Figure 5: Partial Visualization of FTS Monitoring. The monitoring includes transfer success rate, bandwidth, transfer number of files, transfer volume and transfer failure details.

3.3 Pre-Data transfer challenge

To better simulate the experimental data transfer environment, the data injection tool was used for data injection, and FTS was used for data transfer. FTS transfer monitoring was used to observe the test results. The JUNO experiment will generate approximately 3PB of data per year, which roughly equals 102MiB/s. In the JUNO experiment, data accumulates over a period of time and is then transferred all at once, requiring a larger transfer bandwidth. To simulate the actual data transfer process, we set the size of single transfer file to 5GB. In order to avoid the impact of short-term network fluctuations, we need to conduct a test for at least one day. The test was conducted by injecting 450GB of data per hour for one day at a bandwidth of 128MiB/s. Additionally, to address the situation of excessive data accumulation in the JUNO experiment, we continuously injected 900GB of data per hour and 1800GB of data per hour for one day, with respective bandwidth of 256MiB/s and 512MiB/s.

In the test at 128MiB/s bandwidth with 450GB of data injected per hour, the transfers from IHEP to CNAF and IHEP to IN2P3 were very smooth (Figure 6). However, the transfer from IHEP to JINR fluctuated significantly, with many transfers accumulating and completing on the following day (Figure 6). This is because the network upgrade did not include the link between IHEP and JINR. The comparison with the good transfer results from IHEP to CNAF and IHEP to IN2P3 illustrates that the network has become more stable after the upgrade.

In the test at 256MiB/s bandwidth, the transfers from IHEP to CNAF and from IHEP to IN2P3 showed very good results, The sudden increase in bandwidth is due to other transfer tasks causing the bandwidth to rise, which also indicates that FTS has the capability to achieve higher bandwidths(Figure 7). However, the transfer from IHEP to JINR was unable to complete the transfer task of 900GB data injected per hour, with approximately 50% of the transfers remaining incomplete (Figure 7). This is also due to the network between IHEP and JINR not being upgraded.

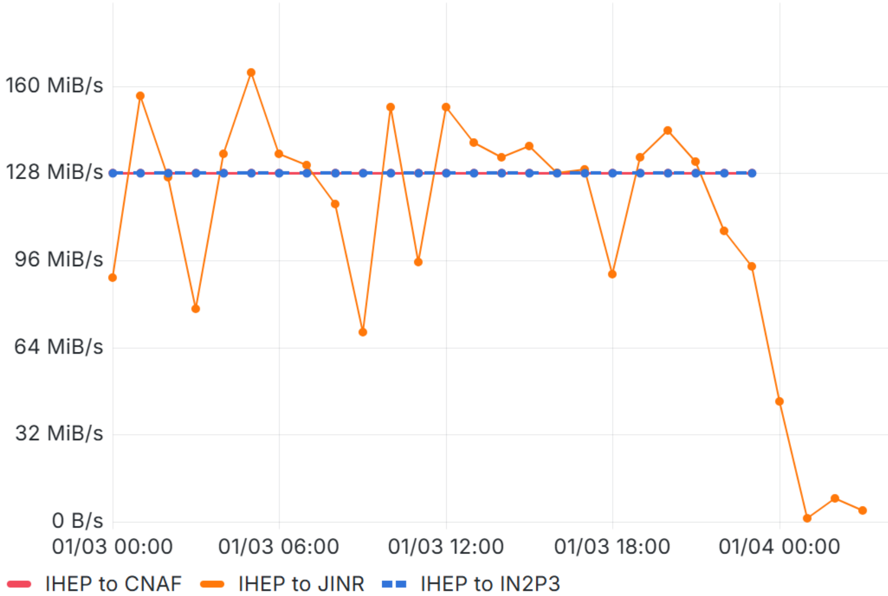

Figure 6: FTS Transfer Monitoring graph for three transfer directions at 128MiB/s bandwidth. The x-axis is time, y-axis is network bandwidth (MiB/s). The legends indicate transfer directions. The transfers from IHEP to CNAF and from IHEP to IN2P3 showed very good results, the bandwidth is consistently maintained at 128 MiB/s. The transfer from IHEP to JINR fluctuated significantly. This is because the network upgrade did not include the link between IHEP and JINR.

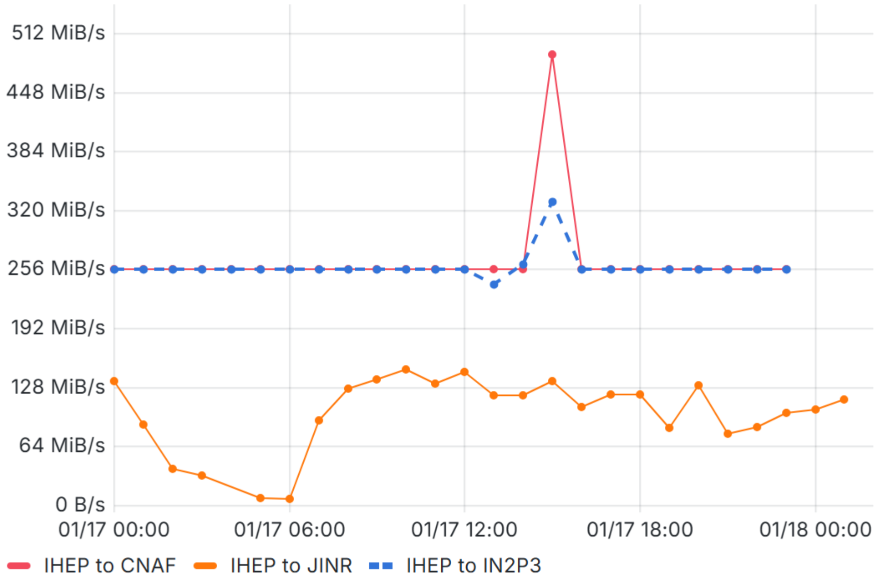

Figure 7: FTS Transfer Monitoring graph for three transfer directions at 256MiB/s bandwidth. The x-axis is time, y-axis is network bandwidth (MiB/s). The legends indicate transfer directions. The transfers from IHEP to CNAF and from IHEP to IN2P3 showed very good results. The transfer from IHEP to JINR was unable to complete the transfer task. The sudden increase in bandwidth is due to other transfer tasks causing the bandwidth to rise, which also indicates that FTS has the capability to achieve higher bandwidths. The transfer from IHEP to JINR was unable to complete the transfer task, with approximately 50% of the transfers remaining incomplete.

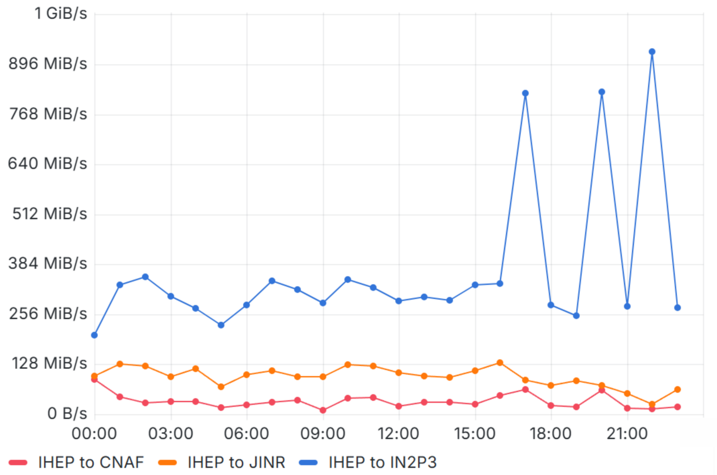

Figure 8: FTS Transfer Monitoring graph for three transfer directions at 512MiB/s bandwidth. The x-axis is time, y-axis is network bandwidth (MiB/s). The legends indicate transfer directions. None of the three types of transfers reached the expected bandwidth of 512MiB/s. This is due to the limited performance of the FTS server, which can activate a maximum of 400 jobs at a time. Another reason is that all transfers share active jobs, the transfer from IHEP to JINR had a very slow transfer rate. This slow transfer occupied the FTS active jobs for an extended period, impacting other transfers and preventing the timely completion of the injected data volume.

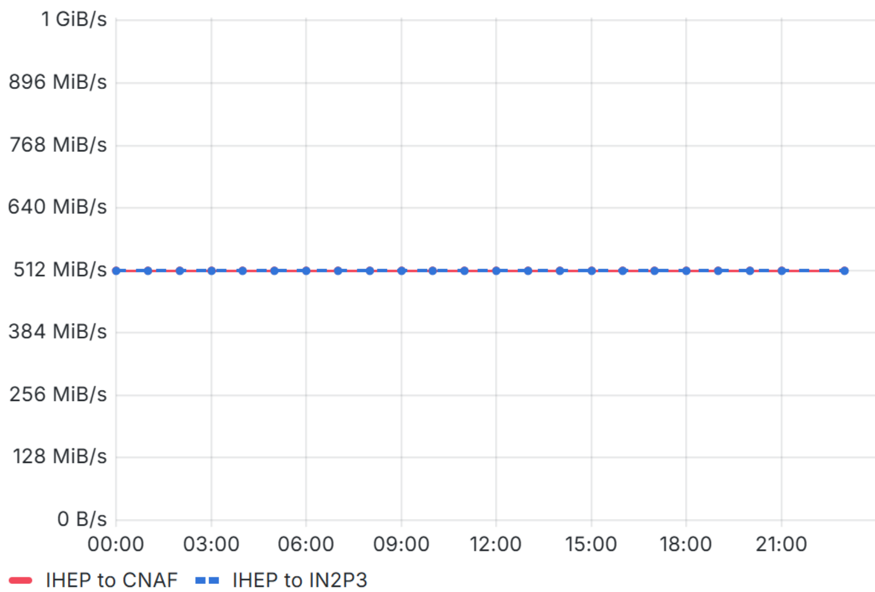

Figure 9: FTS Transfer Monitoring graph for IHEP to CNAF and IHEP to IN2P3 transfer directions at 512MiB/s bandwidth. The x-axis is time, y-axis is network bandwidth (MiB/s). The legends indicate transfer directions. In this test, we removed the direction from IHEP to JINR. The transfers from IHEP to IN2P3 and from IHEP to CNAF were smooth, the bandwidth reached 512MiB/s.

Figure 8 shows that none of the three types of transfers reached the expected bandwidth of

512MiB/s. This is due to the limited performance of the FTS server, which can activate a maximum of 400 jobs at a time. Another reason is that all transfers share active jobs, when a slow transfer occupies an active job, it affects transfers in other directions. The transfer from IHEP to JINR had a very slow transfer rate. This slow transfer occupied the FTS active jobs for an extended period, impacting other transfers and preventing the timely completion of the injected data volume. However, when the direction from IHEP to JINR was removed, transfers from IHEP to IN2P3 and from IHEP to CNAF were both able to reach the expected bandwidth of 512MiB/s (Figure 9).

3.4 JUNO Data Challenge 2024

To better ensure the data transfer after JUNO's retrieval, we participated in the data transfer challenge on JUNO Data Challenge 2024. The pre-challenge showed that in the case of single file sizes of 5GB, the transfer from IHEP to JINR was unstable and prone to causing accumulation, which affects transfers in other directions. Conversely, the transfers from IHEP to CNAF and from IHEP to IN2P3 performed very well. Therefore, for the DC24, we set two directions: from IHEP to CNAF and from IHEP to IN2P3. The transfer bandwidth was 512MiB/s. The challenge took place from February 12th to February 23rd. We found the transfer results were good, the bandwidth was increased to 1024MiB/s on the last day.

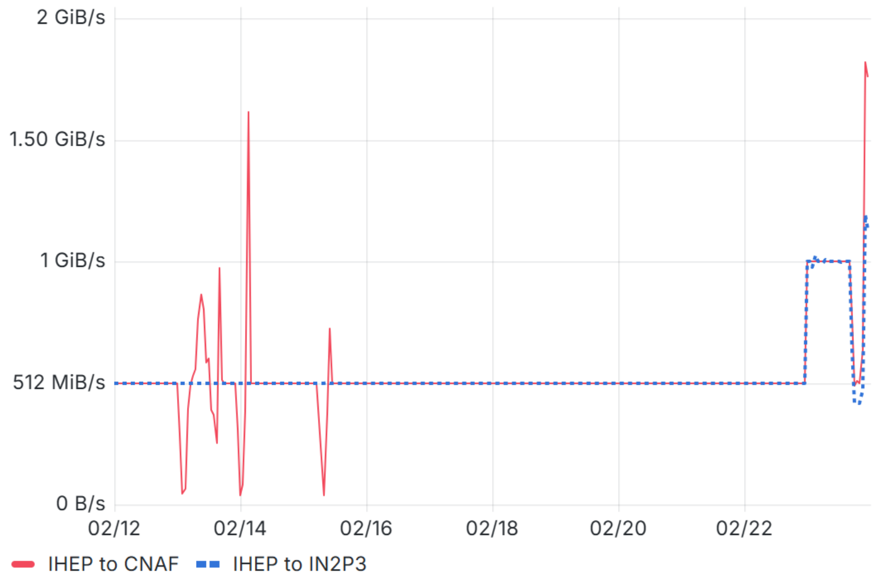

Figure 10: FTS Transfer Monitoring graph for data transfer challenge in DC24. The x-axis is time, y-axis is network bandwidth (MiB/s). The legends indicate transfer directions. The fluctuations observed in the direction of CNAF before February 16th were due to all experiments sharing a single StoRM WebDAV endpoint at CNAF. The endpoint became overloaded during DC24. This overload prevented the establishment of transfer connections, ultimately leading to transfer failures. This issue was also encountered in other experiments during DC24 and has been reported at the ISGC 2024[12].

As shown in Figure 10, the transfer from IHEP to IN2P3 performed extremely well, with no fluctuations from February 12th to February 22nd. The transfer from IHEP to CNAF experienced some fluctuations along the way, but overall performed well. The fluctuations observed in the direction of CNAF before February 16th were due to all experiments sharing a single StoRM WebDAV endpoint at CNAF. The endpoint became overloaded during DC24. This overload prevented the establishment of transfer connections, ultimately leading to transfer failures. This issue was also encountered in other experiments during DC24 and has been reported at the ISGC2024 [12].

Due to the stable transfers between February 16th and February 22nd, we increased the bandwidth to 1024 on the last day to explore the performance. When the bandwidth was increased to 1024MiB/s, there were some fluctuations in the final few hours. We found that when fluctuations occurred, the transfer bandwidth for each file decreased. The reason for this issue has not yet been identified. We suspect it may be due to network fluctuations. This issue does not impede our progress towards meeting the goals of the DC24 test.

The network performance demonstrated in this data transfer challenge far exceeds the requirements of the JUNO experiment, and even under high load environment at the site, the network transfer performance can meet the requirements of the JUNO experiment.

4. Conclusion

In the previous data transfer challenge, the test content was single. There are several disadvantages:

- There are no tests after the network upgrade.
- The transfer tools utilized are not those used in the real environment.
- There is no monitoring of the transfers and no data injection tools in the real transfer environment.
- It is unclear whether the existing DCI can meet the experimental needs.

In order to better understand the changes from the network upgrade, we conducted a high-bandwidth test. The test results showed that the transfer bandwidth between China and Europe could reach 47 Gb/s. Compared to the results before the network upgrade, the bandwidth has been significantly increased. To better conduct the data challenge, we used FTS for monitoring and developed a data injection tool for real transfer environment.

To understand the network transfer conditions under the real environment of the JUNO experiment, we conducted a series of tests in the real transfer environment. The tests found that the transfers between IHEP and IN2P3 and CNAF were relatively stable, while the transfers between IHEP and the JINR site experienced significant fluctuations. The tests also confirmed that the existing DCI is capable of meeting the transfer requirements of the JUNO experiment.

Subsequently, we conducted research on JINR and found that a large number of "DESTINATION CHECKSUM Connection timed out" errors occurred in the direction of JINR. The reason for these errors was that the file size was too large, causing the transfer to take too long. Therefore, we changed the single file size from 5GB to 2GB, which increased the success rate by about 20%.

To reduce the fail rate of data transfer, we enabled the retry feature of FTS and set the number of retry to 3. After testing, the fail rate for transfers decreased by approximately 10%.

Finally, we optimized the performance of FTS. We migrated the FTS service to a server with better performance. And to prevent FTS from becoming a bottleneck in transfers, we increased the FTS active jobs limit to 20,000. To avoid affecting transfers at other sites, we also capped the active jobs limit for transfers between IHEP and JINR at 400. After testing, the transfer bandwidth can reach approximately 3.5 GiB/s.

In general, compared to previous data challenge, we have developed data injection tool and FTS transfer monitoring. By using these tools, we have conducted the JUNO Data Challenge 2024. This data challenge has provided a better understanding of the transfer performance after the network upgrade. The research findings indicate that the environmental factors affecting data transfer of JUNO experiment include the following points:

- The storage I/O could possibly be a bottleneck in data transfer.
- The size of single transfer file.
- The number of active jobs in the FTS service and the retry function.

In the future, we will continue to monitor the transfers in the direction of JINR and strive to improve the transfer performance for the JUNO experiment.

References

[1] Jamie Shiers. The worldwide lhc computing grid (worldwide lcg). *Computer Physics Communications*, 177(1-2):219–223, 2007.

[2] T Adam, F An, G An, Q An, N Anfimov, V Antonelli, G Baccolo, M Baldoncini, E Baussan, M Bellato, et al. Juno conceptual design report. *arXiv preprint arXiv:1508.07166*, 2015.

[3] Tetsuo Abe, Ichiro Adachi, Karol Adamczyk, S Ahn, H Aihara, K Akai, M Aloi, L Andricek, K Aoki, Y Arai, et al. Belle ii technical design report. *arXiv preprint arXiv:1011.0352*, 2010.

[4] Giuseppe Andronico. China-eu scientific cooperation on juno distributed computing. In *EPJ Web of Conferences*, volume 245, page 03038. EDP Sciences, 2020.

[5] Giuseppe Andronico, Zhang Xiaomei, and Li Weidong. Jiangmen underground neutrino observatory computing requirements and infrastructure. In *XXIX International Symposium on Lepton Photon Interactions at High Energies*, volume 122, 2019.

[6] Andrea Rendina, Lucia Morganti, Federico Fornari, Lorenzo Chiarelli, Donato De Girolamo, and Stefano Zani. Cnaf experience in support of the juno distributed computing model. In *International Symposium on Grids & Clouds (ISGC)*, volume 19, 2023.

[7] Edoardo Martelli and S Stancu. Lhcopn and lhcone: status and future evolution. In *Journal of Physics: Conference Series*, volume 664, page 052025. IOP Publishing, 2015.

[8] https://indico.cern.ch/event/1225415/contributions/5155042/attachments/2593516/4476291 /data%20challenge%202024.pdf.

[9] ákos Frohner, Jean Philippe Baud, Rosa Maria Garcia Rioja, Gilbert Grosdidier, Rémi Mollon, David Smith, and Paolo Tedesco. Data management in egee. *Journal of Physics Conference*, 219(6):062012, 2010.

[10] Tom Hepworth. Improving observability of fts. Technical report, 2022.

[11] https://code.ihep.ac.cn/liyifan/data-challenge.

[12] https://indico4.twgrid.org/event/33/contributions/1407/attachments/823/1052/isgc24_dc24.pdf.