



Rucio/BigData Express/SENSE (ROBIN) a Next Generation High Performance Data Service Platform

Dr. Wenji Wu (wenji@fnal.gov)

Thursday, March 11, 2021

Many people's hard work

FNAL: Wenji Wu, Liang Zhang, Qiming Lu, Amy Jin,
Phil DeMar, Robert Illingworth

iCAIR/StarLight: Joe Mambretti, Se-young Yu, Fei Yeh, Jim-Hao Chen

ESnet: Inder Monga, Xi Yang, Tom Lehman, Chin Guok,
John Macauley

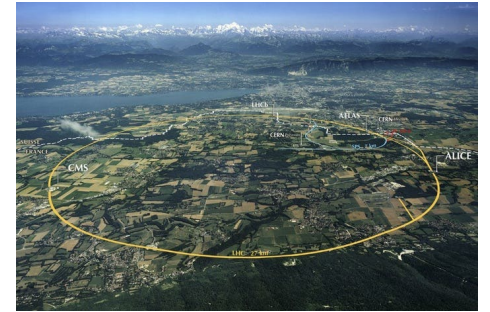


Agenda

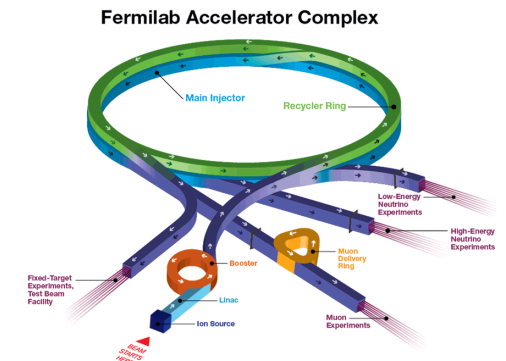
- **Motivation**
- **ROBIN: a next generation high performance data service platform**
 - **Architecture**
 - **Key Mechanisms**
- **Initial Evaluation**
 - **An international testbed**
 - **Experiments**

Motivation (I)

- **Big data has emerged as a driving force for scientific discoveries.**
- **Large scientific instruments generates large amount of data.**
- **Science data must be collected, indexed, archived, shared, and analyzed, typically in a widely distributed, highly collaborative manner.**



The LHC Accelerator Complex



Fermilab Accelerator Complex

Motivation (II)

Managing and moving extremely large volumes of science data worldwide is a special multidimensional challenges!

Need a comprehensive service solution that incorporates:

- **Services designed for scientists**
- **Scientific workflows**
- **Data management**
- **Science DMZs**
- **Large volume data transfer services**
- **Orchestration**
- **High-performance networking**

Our Solution

ROBIN (Rucio/BigData Express/SENSE):

A Next Generation High-performance Data Service Platform



ROBIN (Rucio/BigData Express/SENSE)

A Next Generation High-performance Data Service Platform

Scientific Applications
(e.g., LSST, CMS)

Rucio
Data Management Service

- Smart Namespace
- Storage Support
- Consistency
- Easy Integration
- AA Support
- Proven Track Record

BigData Express
High-performance Data Transfer Service

- Peer-to-peer, Scalable, Extensible
- End-to-end Performance Optimization
- High Performance Data Transfer Engine
- CILogon-base Security

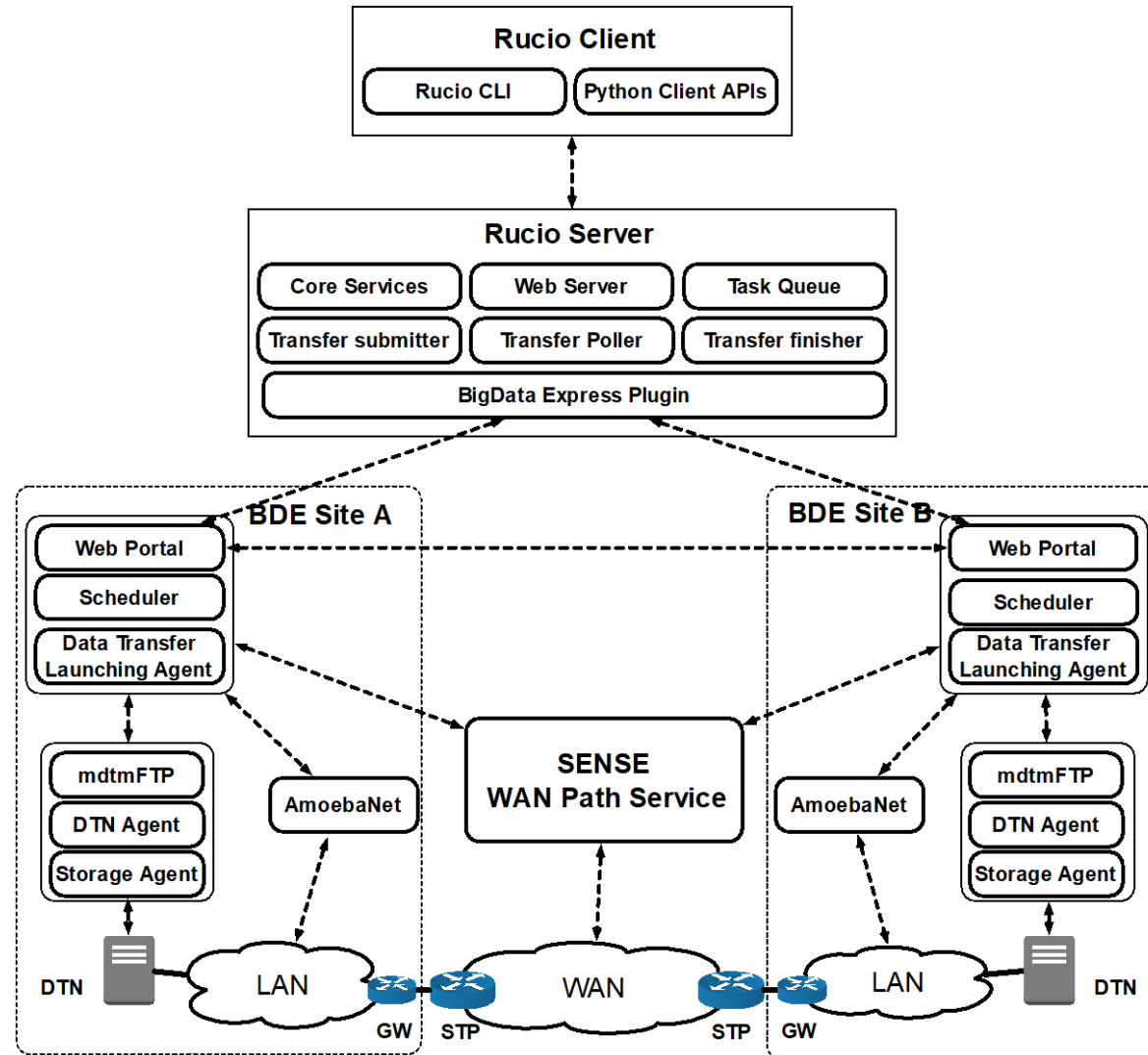
SENSE
Smart Network Service

- Software Defined-Networking (SDN) for End-to-End Virtual Guaranteed Network
- A Highly Intuitive “Intent”-based Interface



ROBIN (Rucio/BigData Express/SENSE)

A Next Generation High-performance Data Service Platform



ROBIN Key Mechanisms

- **Site Registration**
- **Rucio/BigData Express (BDE) job launching mechanism**
- **On-demand provisioning of end-to-end network path with guaranteed Qos**
- **Security**

Site Registration

- **Register a BigData Express site as an RSE with the Rucio server**
 - **The RSE name**
 - **The information necessary to access the new RSE**
 - **Hostname, port, protocol, and local file system path**
 - **The distance metric between the new RSE and other RSEs**

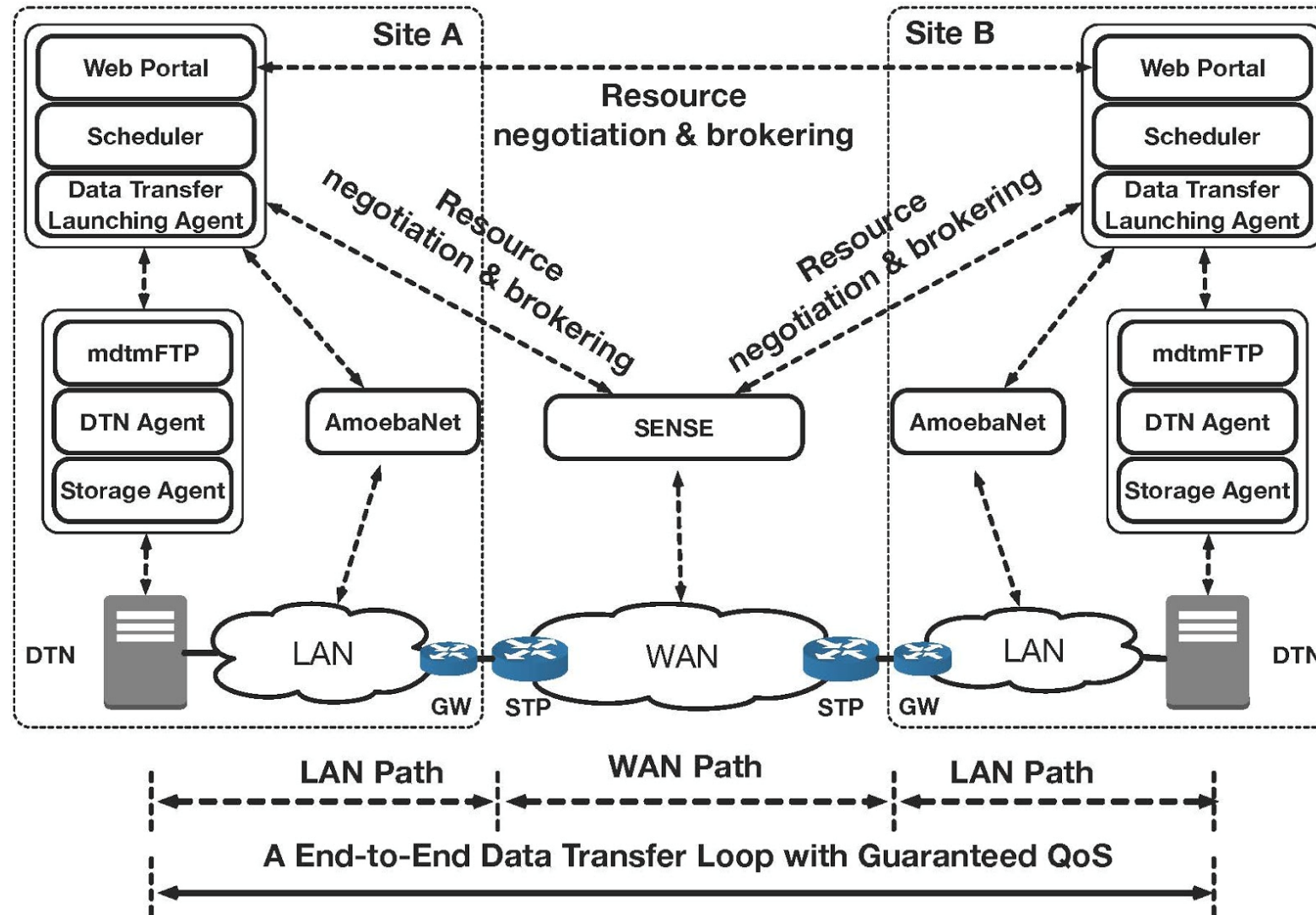
A new protocol “bde” is defined to support BDE-based data transfer.



Rucio/BDE Job Launching Mechanism

1. A Rucio client uses Rucio CLI to request replication on a destination RSE.
2. The client sends the replication request to the Rucio server.
3. The Rucio server creates a replication rule for the request and generates the data transfer tasks. The tasks are temporarily kept in a *task queue*.
4. The Rucio server regularly pulls tasks from the queue. It ranks the sources for each task, selects the protocol “*bde*” for src/dst RSEs, submits the tasks in groups to BDE.
5. BDE schedules and assigns resources (DTNs, network) to execute the data transfer tasks.
 - BDE calls SENSE to provision WAN paths with guaranteed QoS between sites.
6. After the DTNs and the paths have been successfully reserved, BDE launches the data transfer tasks, monitors the progress of the tasks, retries in case of errors, and notifies the Rucio server upon completion.
7. The Rucio server closely monitors the status of the transfers. A failed data transfer will be resubmitted in the *task queue* for retries until the maximum retry limit is reached.
8. The Rucio server updates the internal states and notifies the client upon completion

On-demand Provisioning of End-to-end Path with Guaranteed Qos

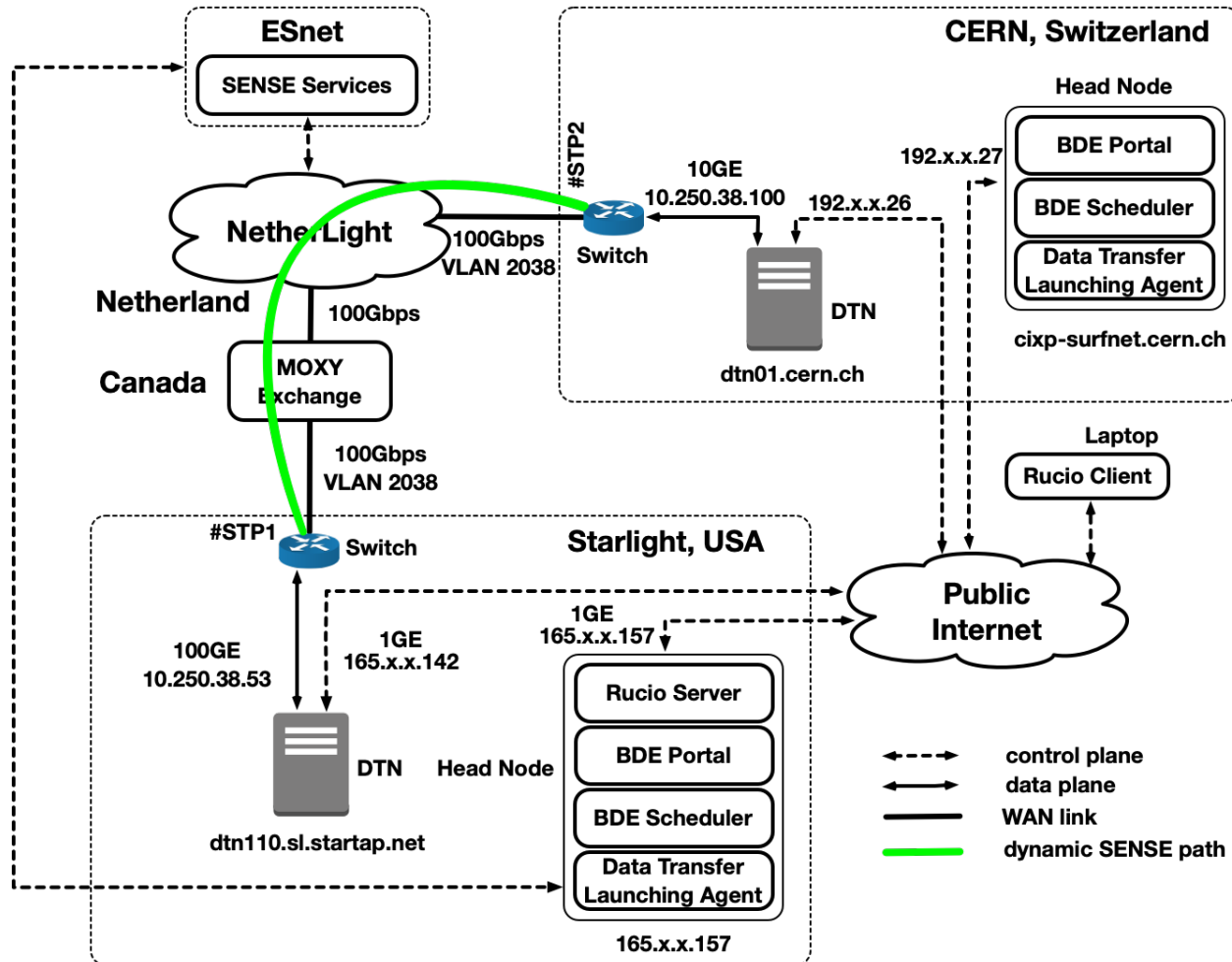


Security

- **Keep each system's security intact**
- **Execute a logic mapping between them to enforce security at all levels**
 - **Direct mapping between Rucio and BDE accounts with X509 certificate delegation**
 - **Each BDE site, acting as a SENSE client, with pre-configured client credential.**

Systems	Authentication/Authorization methods
Rucio	Username/password, X509 certificates, Kerberos tickets, SSH-RSA public key
BigData Express	Username/password, X509 certificates
SENSE	Username/password, OIDC

ROBIN Cross-Atlantic Testbed



StarLight site:

- DTN: `dtn110.sl.startap.net`, with several Intel NVMe drives for data storage, a 100GE Mellanox NIC for data transfer, and a 1G NIC for control.
- Head node: `165.x.x.157`, with a 1G NIC for control.

CERN site:

- DTN: `dtn01.cern.ch`, with a rotational disk for data storage, a 10GE Mellanox NIC for data transfer, and 1G NIC for control.
- Head node: `cixp-urfnet.cern.ch`, with a 1G NIC for control.

Experiments

1. Register each BDE site in the testbed as an RSE with the Rucio server
2. Create an experiment file named *25g-1.bin* and register the file with the Rucio server
3. Use the Rucio client to submit a request to the Rucio server to replicate the registered file from *StarLight* to *CERN*

Results (I) – Rucio Rules

```
$rucio-cmd rucio list-file-replicas test:25g-1.bin
```

SCOPE	NAME	FILESIZE	ADLER32	RSE: REPLICA
test	25g-1.bin	26.844 GB	49576448	STARLIGHT-SITE: bde://165.124.33.157:5000/165.124.33.142/disk0//25g-1.bin

```
$rucio-cmd rucio list-rules test:25g-1.bin
```

ID	ACCOUNT	SCOPE:NAME	STATE[OK/REPL/STUCK]	RSE_EXPRESSION	COPIES	EXPIRES (UTC)	CREATED (UTC)
554acd9b7ddb4a319a37308a1285753f	root	test:25g-1.bin	OK[1/0/0]	STARLIGHT-SITE	1		2020-08-31 04:04:32

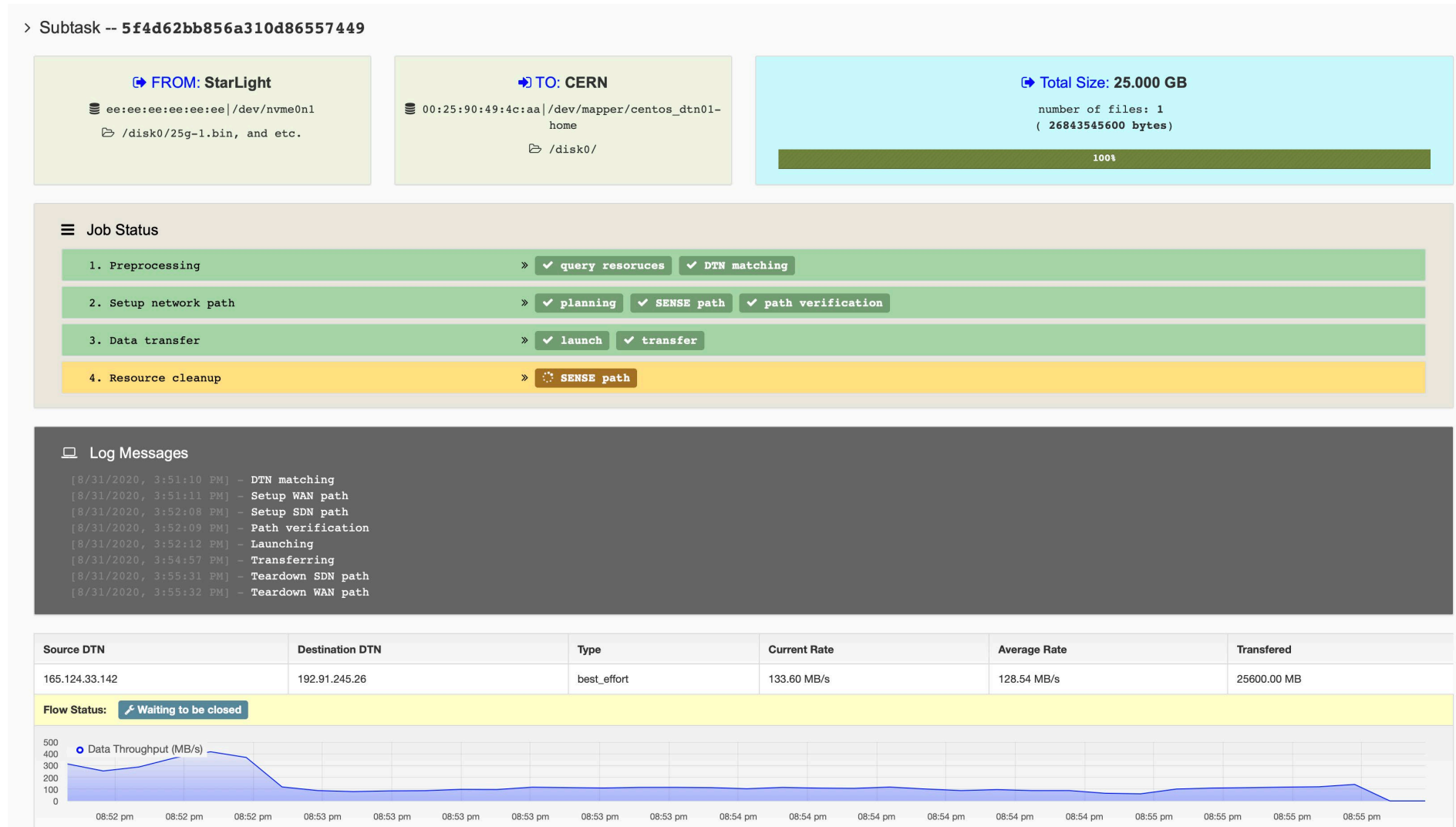
The replica and the replication rule for *25g-1.bin*

```
$rucio-cmd rucio list-rules test:25g-1.bin
```

ID	ACCOUNT	SCOPE:NAME	STATE[OK/REPL/STUCK]	RSE_EXPRESSION	COPIES	EXPIRES (UTC)	CREATED (UTC)
c510e4cd53b44b77b6cdb2c367036766	root	test:25g-1.bin	REPLICATING[0/1/0]	CERN-SITE	1		2020-08-31 04:09:34
554acd9b7ddb4a319a37308a1285753f	root	test:25g-1.bin	OK[1/0/0]	STARLIGHT-SITE	1		2020-08-31 04:04:32

The replication rule created for the Rucio data replication request

Results (II) – BigData Data Transfer Process



Results (III) – BDE/SENSE Interactions

BDE Control Event	Transaction Time	SENSE Control Event	Transaction Time
Service Negotiation	5s	Compute Service Intent (initial)	3s
		Re-Compute Service (negotiate)	2s
Service Reservation	9s	Reserve with RMs	7s
Service Allocation	94s	Commit with RMs	34s
		Verify Service Model	51s
Service Deallocation	60s	Release with RMs	4s
		Commit with RMs	33s
		Verify Service Model	12s

Future Plans (Work in Progress)

- **Continue to evaluate/test ROBIN**
 - 100Gbps international WAN paths
 - High-end DTNs
 - Multiple site deployment
 - Increased automation
 - Enhanced parameter analytics
- **Compare ROBIN with Rucio/FTS**

Conclusion

- **ROBIN (Rucio/BigData Express/SENSE)
A Next Generation High-performance Data Service Platform**
- **A unique comprehensive set of integrated services designed specifically for managing and moving extremely large amounts of data over long distances**

Questions?

Additional Information

[1] Rucio: <https://rucio.cern.ch/>

[2] BigData Express: <http://bigdataexpress.fnal.gov>

[3] SENSE: <http://sense.es.net>