CERN FTS site report

XRootD and FTS Workshop 2023 at JSI

Steven Murray on behalf of the CERN FTS team

Monday 27th March 2023

Deployment architecture

Central machines shared by all instances

FTS watchdog

- Monitor and alarm each FTS instance
- Close idle database connections

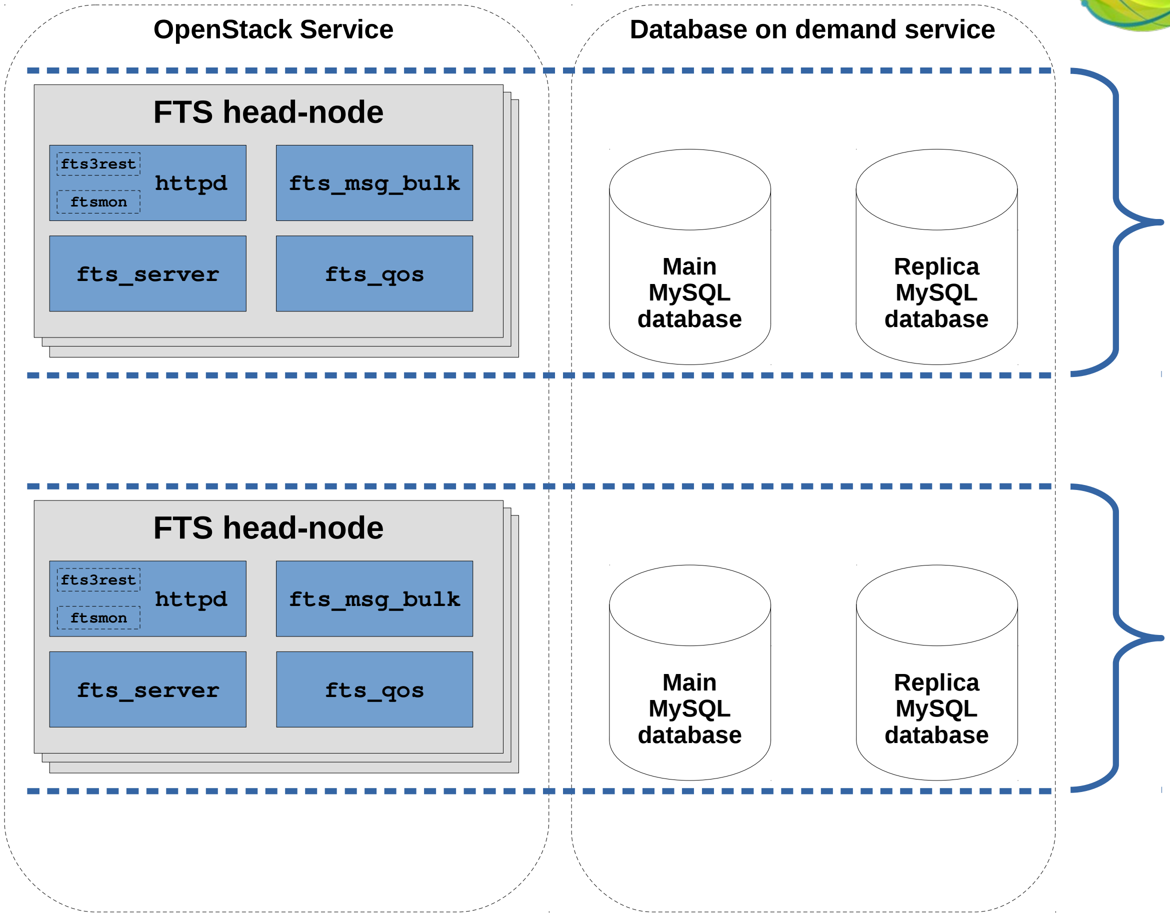
FTS database backup

- Back up important database tables
- Defragment the main databases

The **FTS watchdog** machine sends monitoring information to a Graphite instance provided by the storage group.

The **FTS database backup** machine writes encrypted database dumps of the important FTS database tables to the CERN Public EOS instance.

FTS head-nodes use the CERN Public EOS instance as a disk cache when performing streamed transfers



FTS Instance X
A single FTS instance is made up of several identical head-node machines plus a main database and a replica.

The main database is used to queue and track transfers and is mission critical. The replica is used by FTS web monitoring.

FTS Instance Y
Another FTS instance.

Legend

▢	Machine
■	Process
▨	Web service
⬚	External service

Database configuration

- **FTS queries benefit a lot from a large RAM cache:**
  - `innodb_buffer_pool_size`

- **It is important to have a long enough log file to record operations that have occurred during an on-line Data Definition Language (DDL) operation:**
  - `innodb_online_alter_log_max_size`

CERN production instances

There are 6 production FTS instances at CERN

Instance	No. of VMs	VCPUs per VM	RAM per VM	Disk space per VM	`innodb_buffer_pool_size`	`innodb_online_alter_log_max_size`
ATLAS	10	16	28.6 GiB	160 GB	80 GiB	12.5 GiB
CMS	10	16	28.6 GiB	160 GB	40 GiB	12.5 GiB
DAQ	5	8	14.2 GiB	80 GB	4 GiB	128 MiB
LHCb	5	16	28.6 GiB	160 GB	12 GiB	128 MiB
Pilot	5	8	14.2 GiB	80 GB	12 GiB	1 GiB
Public	5	8	14.2 GiB	80 GB	4 GiB	1 GiB

Database as a service

- Our in-house database on demand (DBoD) service provides our MySQL databases
- Some FTS use-cases lacked performance
- The performance problems were addressed by:
  - Adding a replica database for long monitoring queries
  - Defragmenting the main database once a week
- We have setup our own replicated database on dedicated hardware but we are currently sticking with DBoD

FTS headnode machines

- **All HTTP transfers use `libcurl` as opposed to `libneon`**

```
/etc/sysconfig/fts-qos:DAVIX_USE_LIBCURL=Y
/etc/sysconfig/fts-server:DAVIX_USE_LIBCURL=Y
```

- **`systemctl` restarts the FTS daemons when they crash**

```
/usr/lib/systemd/system/fts-bringonline.service:Restart=on-failure
/usr/lib/systemd/system/fts-msg-bulk.service:Restart=on-failure
/usr/lib/systemd/system/fts-msg-bulk.service:RestartSec=3
/usr/lib/systemd/system/fts-qos.service:Restart=on-failure
/usr/lib/systemd/system/fts-qos.service:RestartSec=3
/usr/lib/systemd/system/fts-server.service:Restart=on-failure
/usr/lib/systemd/system/fts-server.service:RestartSec=3
```

- **HTTP daemons are restarted every hour to make them read the Certificate Revocation Lists (CRLs)**

```
# crontab -l
...
30 * * * * (/usr/sbin/fetch-crl; /usr/bin/systemctl restart httpd.service) &> /dev/null
```

FTS watchdog machine

- **Poll FTS and send monitoring messages to Graphite**

```
# crontab -l
...
* * * * * /var/fts-watchdog/fts_db_file_states_poll.py --instance XXXX ...
*/5 * * * * /var/fts-watchdog/fts_db_staging_requests_poll.py --instance XXXX ...
*/5 * * * * /var/fts-watchdog/fts_db_nb_connections_poll.py --instance XXXX ...
* * * * * /var/fts-watchdog/fts_db_seconds_behind_main_poll.py --instance XXXX ...
```

Graphite

```
def _pickle_send(self, metrics):
  payload = pickle.dumps(metrics, protocol=2)
  header = struct.pack("!L", len(payload))
  message = header + payload
  conn = socket.create_connection((self.carbon_host, self.carbon_port))
  conn.send(message)
  conn.close()
```

SMS alarm via mail to SMS gateway

```
def _send_mail(self, seconds_behind_main):
    sendmail = subprocess.Popen(['/usr/sbin/sendmail', self.mailing_list],
                                stdin=subprocess.PIPE,
                                stdout=subprocess.PIPE,
                                stderr=subprocess.PIPE)
```

```
KILL 12345;
KILL 67890;
```

```
mysql ... --execute="source /tmp/fts-connection-cleaner-${instance}- ... .query"
```

- **Close idle connections**

```
# crontab -l
...
0 */1 * * * /var/fts-watchdog/fts_db_connections_cleaner.py
```

FTS database backup

- **Backup important database tables**

```
# ls -1 /etc/cron.daily/
backupdb-atlas.sh
backupdb-cms.sh
backupdb-daq.sh
backupdb-lhcb.sh
backupdb-pilot.sh
backupdb-public.sh
```

```
mysqldump --defaults-file=${MYSQL_DEFAULTS_FILE} --single-transaction newfts3atlas \
t_activity_share_config \
t_authz_dn \
t_bad_dns \
t_bad_ses \
t_cloudStorage \
t_cloudStorageUser \
t_config_audit \
t_optimizer \
t_link_config \
t_schema_vers \
t_se \
t_server_config \
t_share_config \
t_stage_req | gpg2 --batch --symmetric --force-mdc --cipher-algo AES256
  --passphrase-file /etc/backupdb/backupdb_gpg_passphrase
  --output ${FTS_BACKUPDB_FILE}
```

Important tables

Encrypted

- **Defragment the main databases**

```
# crontab -l
0 10 * * 1 /usr/bin/ftsdefragdb --vo XXXX ...
```

```
OPTIMIZE NO_WRITE_TO_BINLOG TABLE t_file
OPTIMIZE NO_WRITE_TO_BINLOG TABLE t_job
```

Data privacy

- Privacy Notice: File Transfer Service (PN00048)
  - https://cern.service-now.com/service-portal?id=privacy_policy&se=file-transfer&notice=fts
- Details include:
  - Personal Data we process
  - Personal Data we keep
  - Who at CERN has access
  - Personal Data we may transfer to others

Disaster recovery

- The FTS Virtual Machines are fully Puppetized
- New fully installed Virtual Machines can be created in tens of minutes
- If needed the main database tables can be retrieved from encrypted backups

```
umask 0077
cat /eos/workspace/f/fts/backupdb/2022-08-12_config_fts_atlas_DB.sql.gpg | gpg2 --batch
  --passphrase-file /etc/backupdb/backupdb_gpg_passphrase
  --output 2022-08-12_config_fts_atlas_DB.sql
```

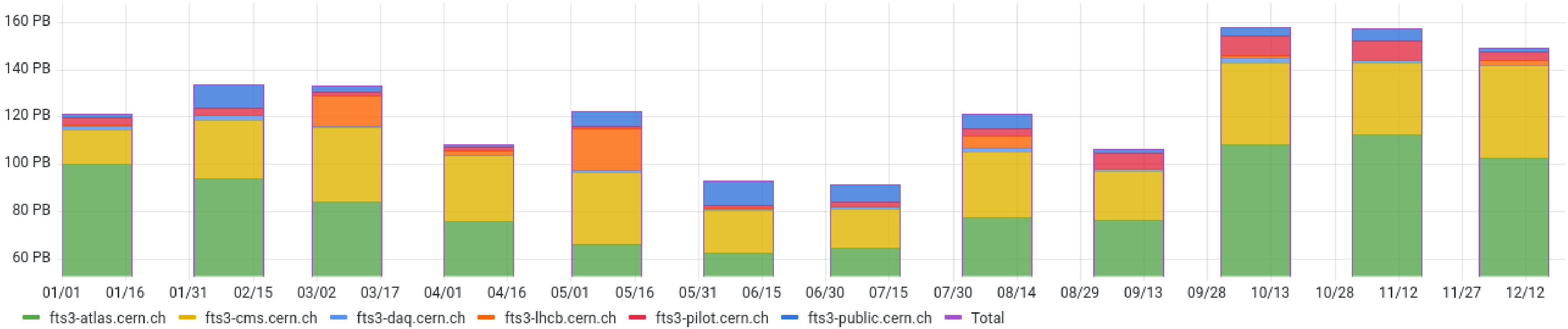
Note:
All queued transfers will be lost but all configurations will be recovered

- Two plan recovery strategy:

  1. Try to recover here at CERN during approximately 1 hour
  2. If CERN is still not back then ask experiments to redirect their FTS requests to alternative sites around the World, for example:
     - ATLAS – Use BNL FTS
     - CMS – Use FNAL FTS

Data volume transferred per month during 2022

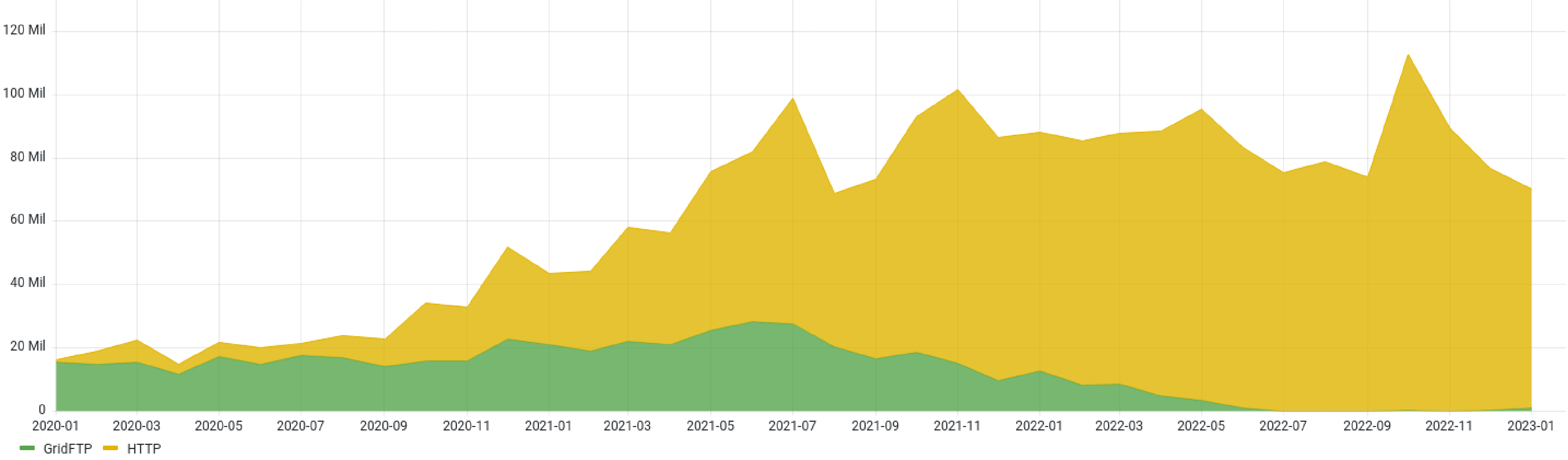
FTS File Transfer Service



GridFTP is being phased out

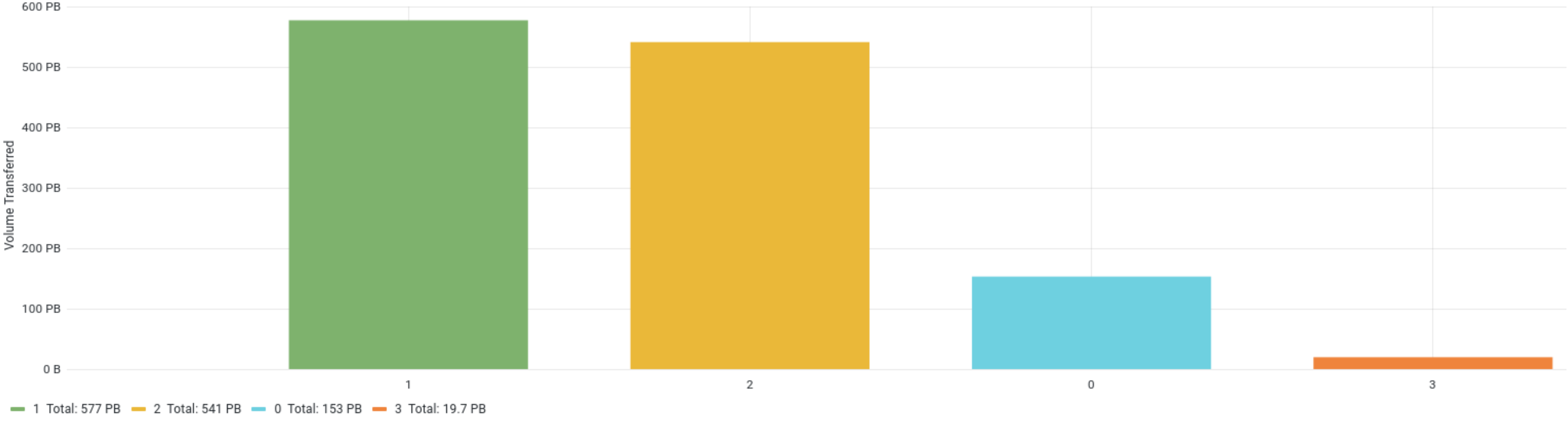
FTS
File Transfer Service

Transfers per month managed by the CERN FTS instances



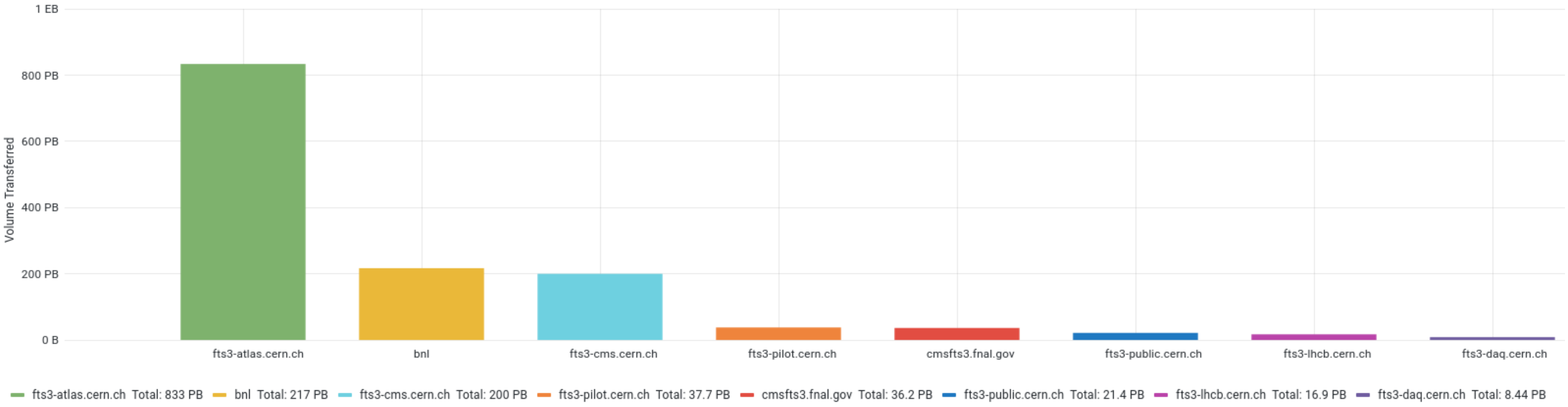
Transfer volume by Tier

Total volume transferred per WLCG tier during 2022 - All FTS sites



Comparison of WLCG instances

Total volume transferred during 2022 - Top 8 WLCG instances



CERN

home.cern