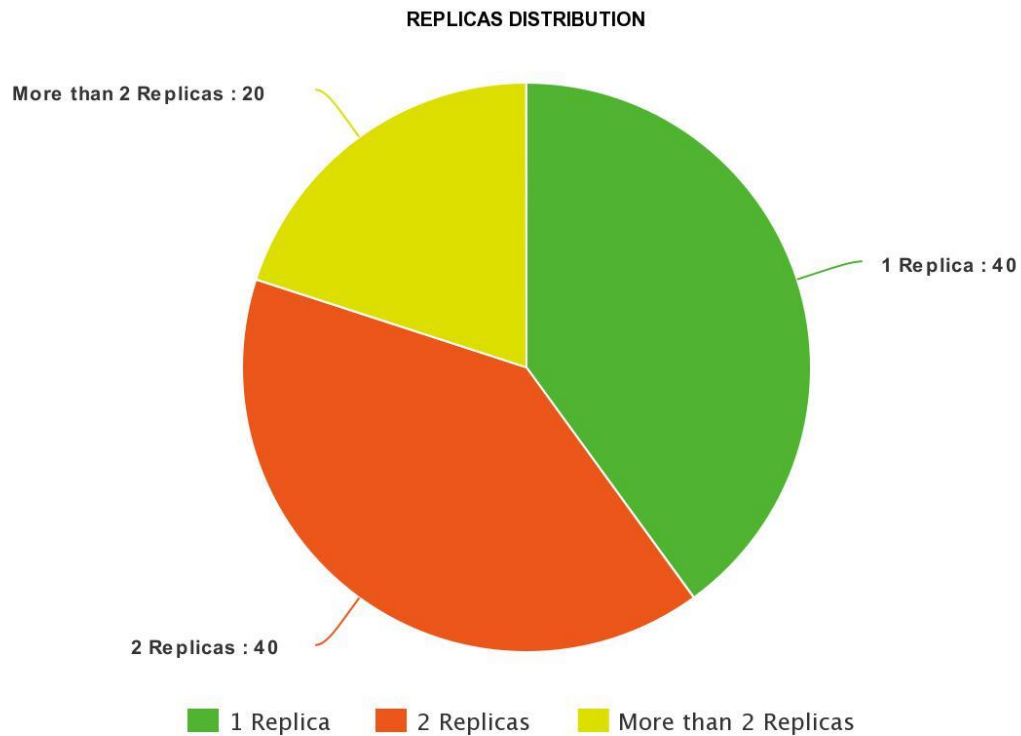


Evaluation of Multi-Source Downloads for FTS

Over HTTPS, GridFTP and XROOT

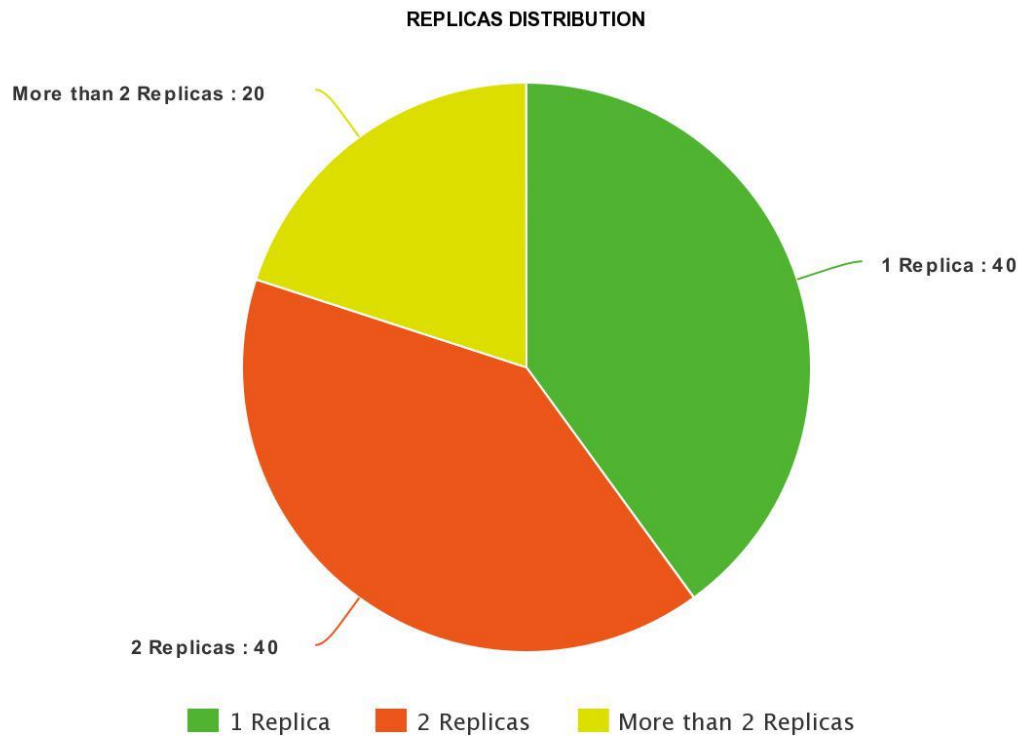
Alin Andrei Grigorean - University Politehnica of Bucharest
Supervisor – Maria Arsuaga-Rios

Current Situation



Current Situation

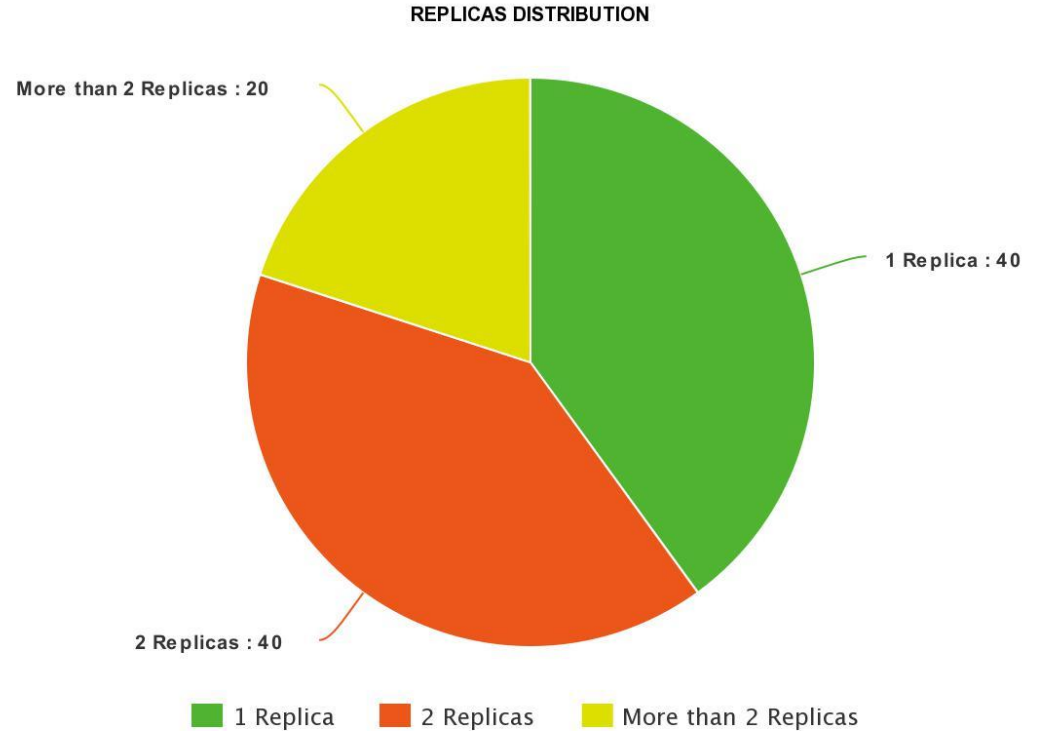
FTS:



Current Situation

FTS:

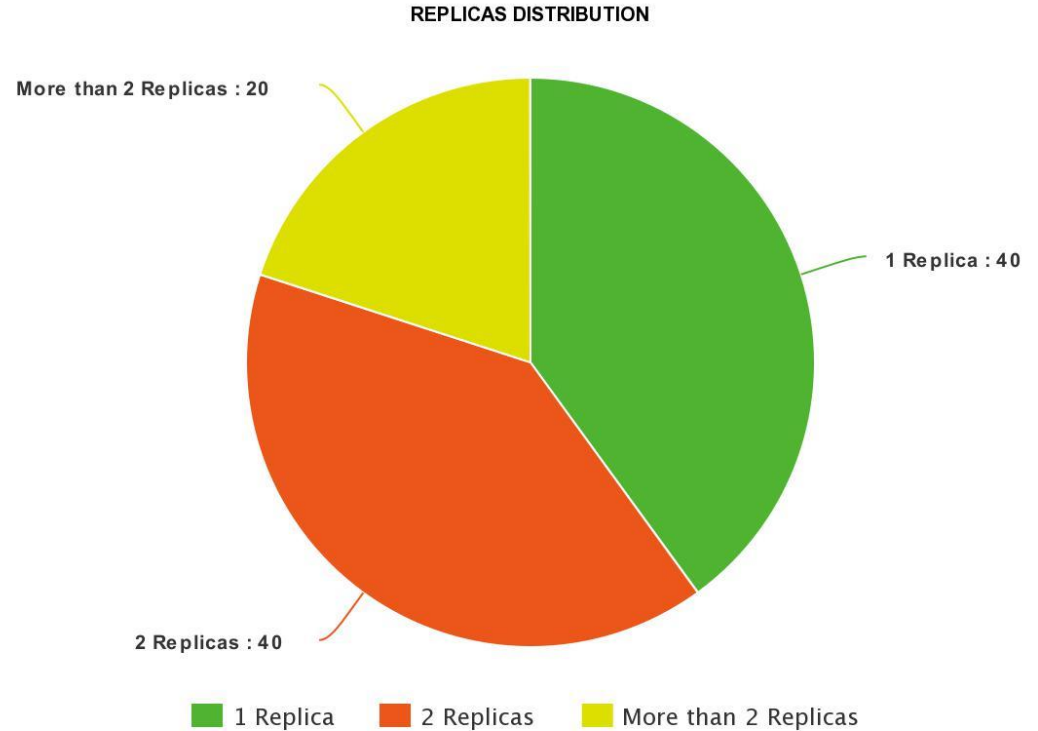
- Transferring from 1 source only



Current Situation

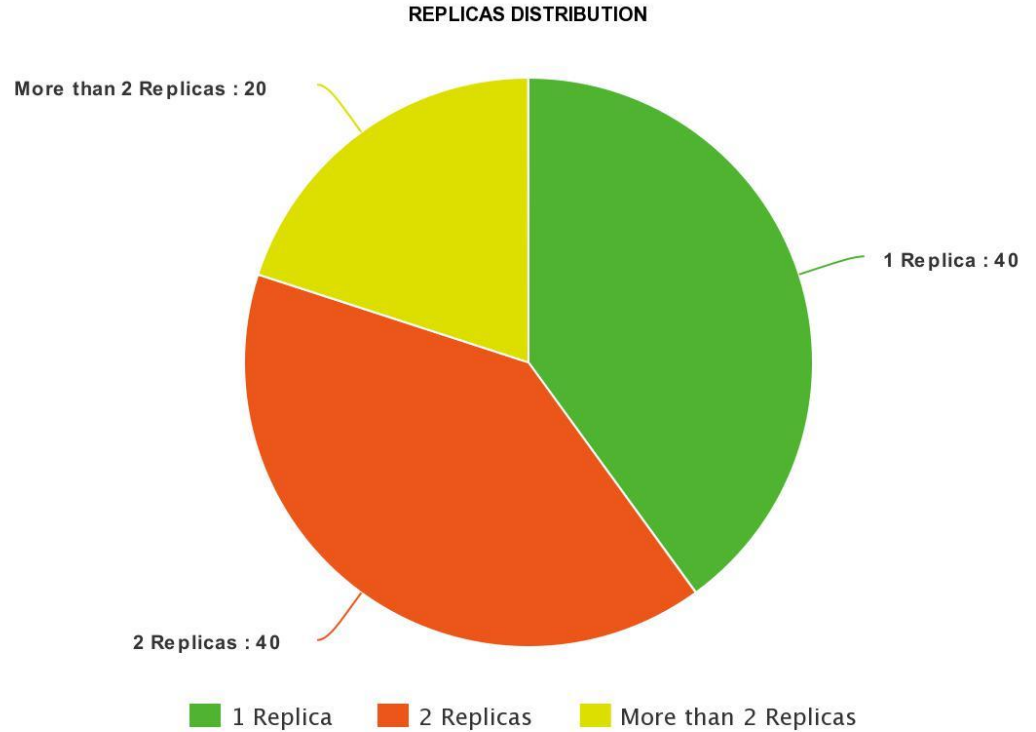
FTS:

- Transferring from 1 source only
- 60% of transfers have 2 or more replicas



Current Situation

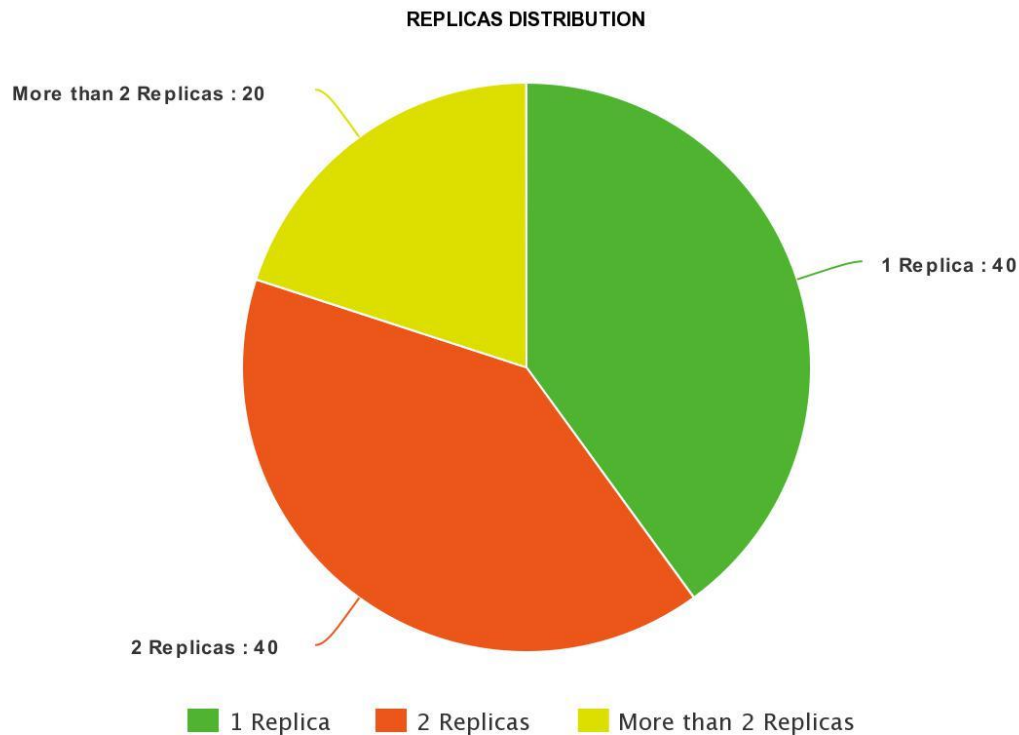
Multi-source downloads with Aria2:



Current Situation

Multi-source downloads with Aria2:

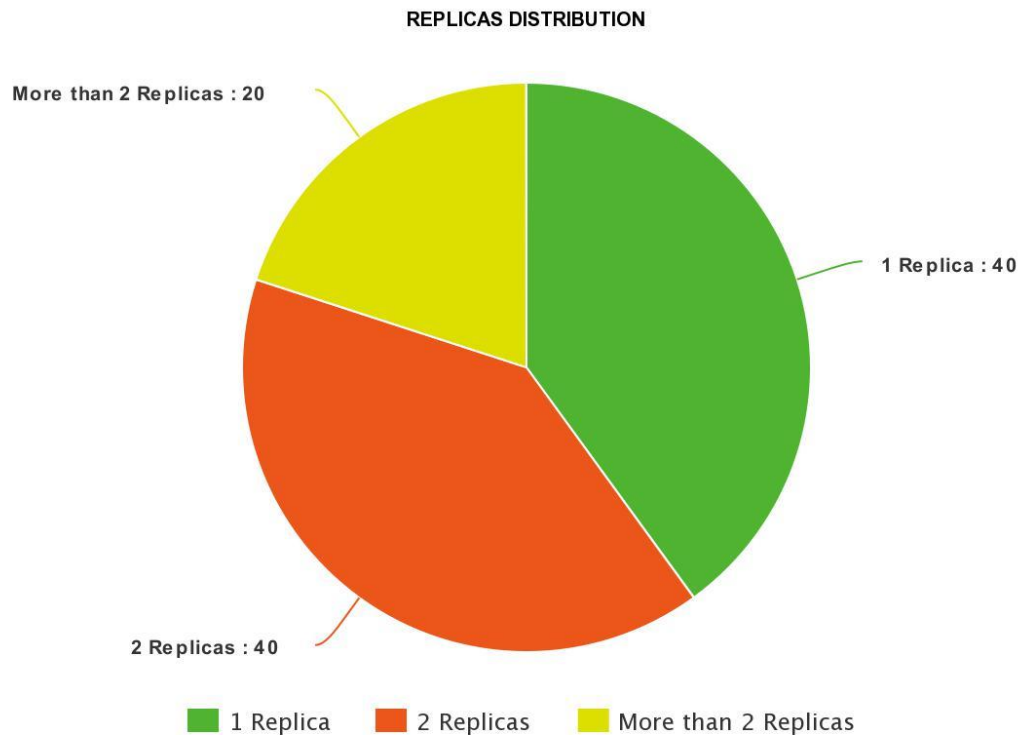
- Downloading as fast as the fastest given source



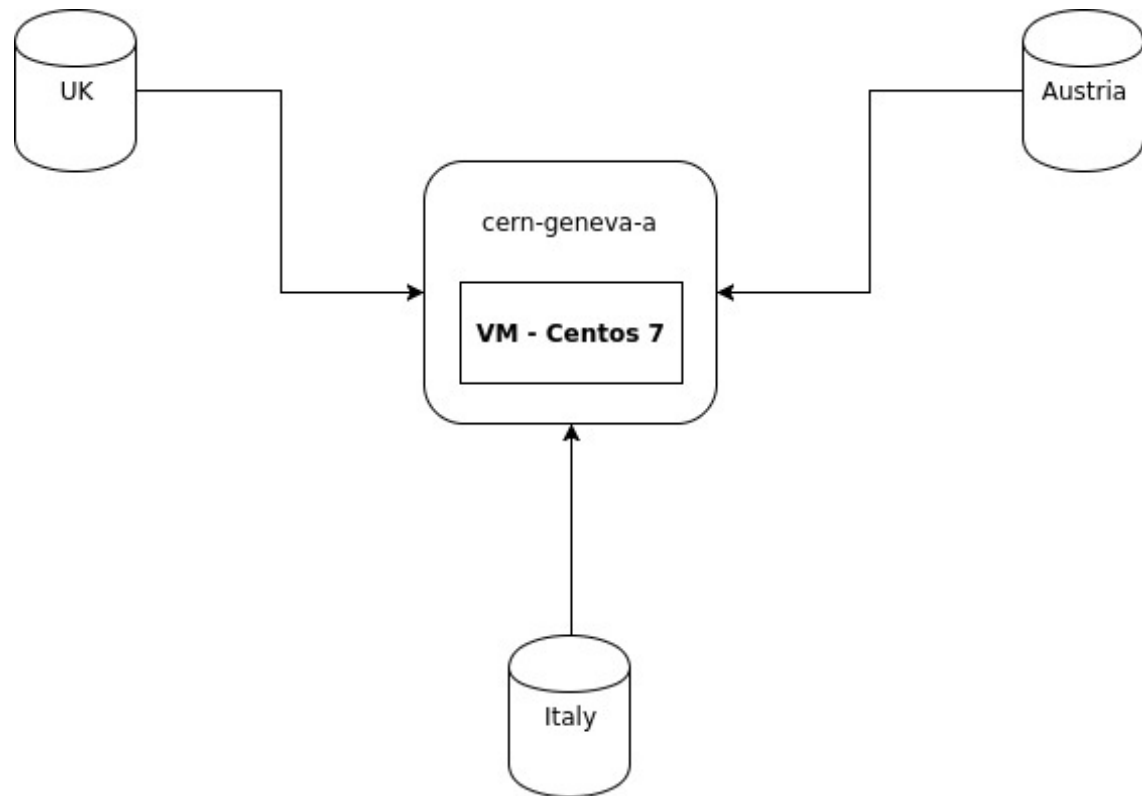
Current Situation

Multi-source downloads with Aria2:

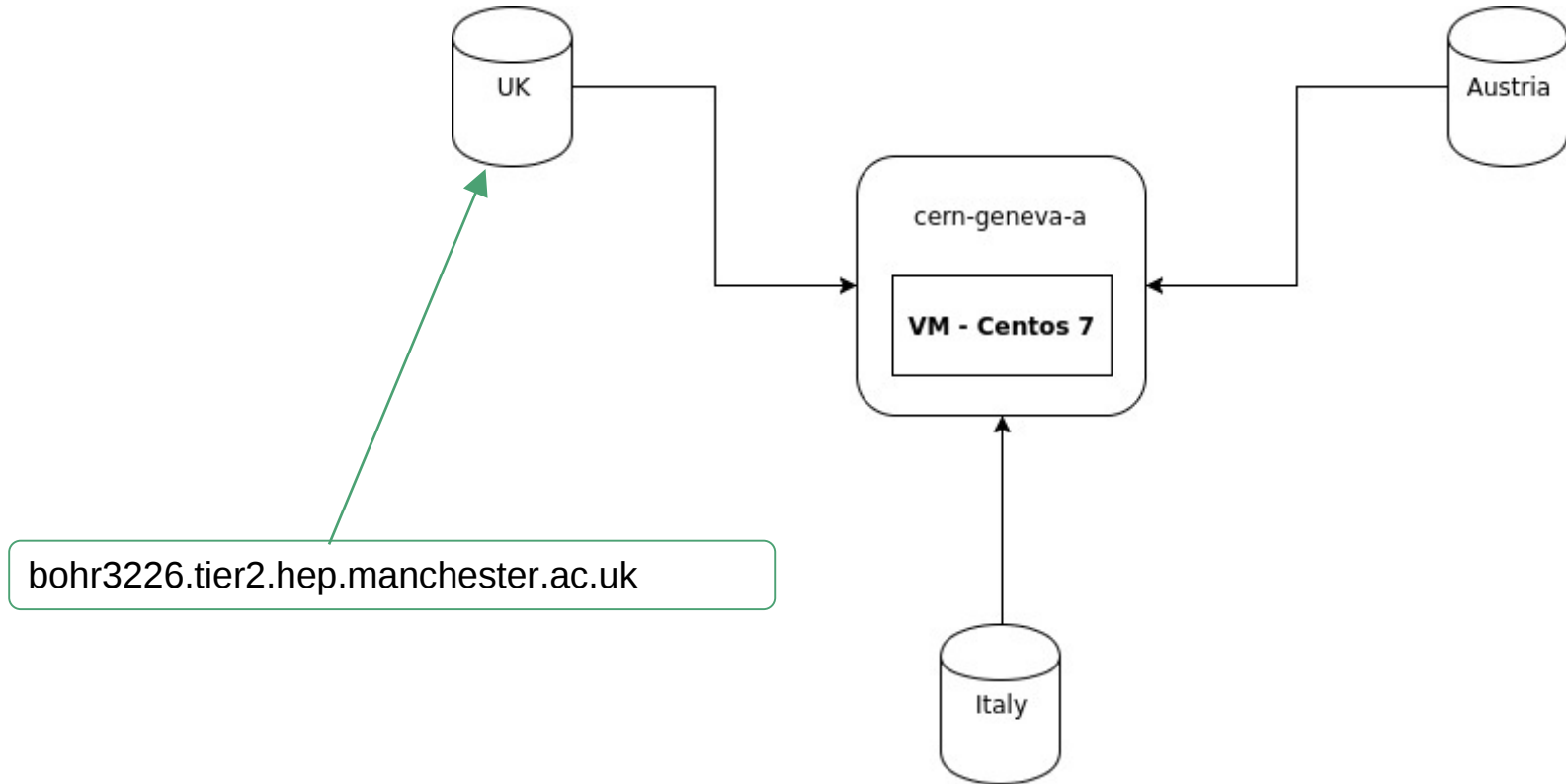
- Downloading as fast as the fastest given source
- More balanced download throughput



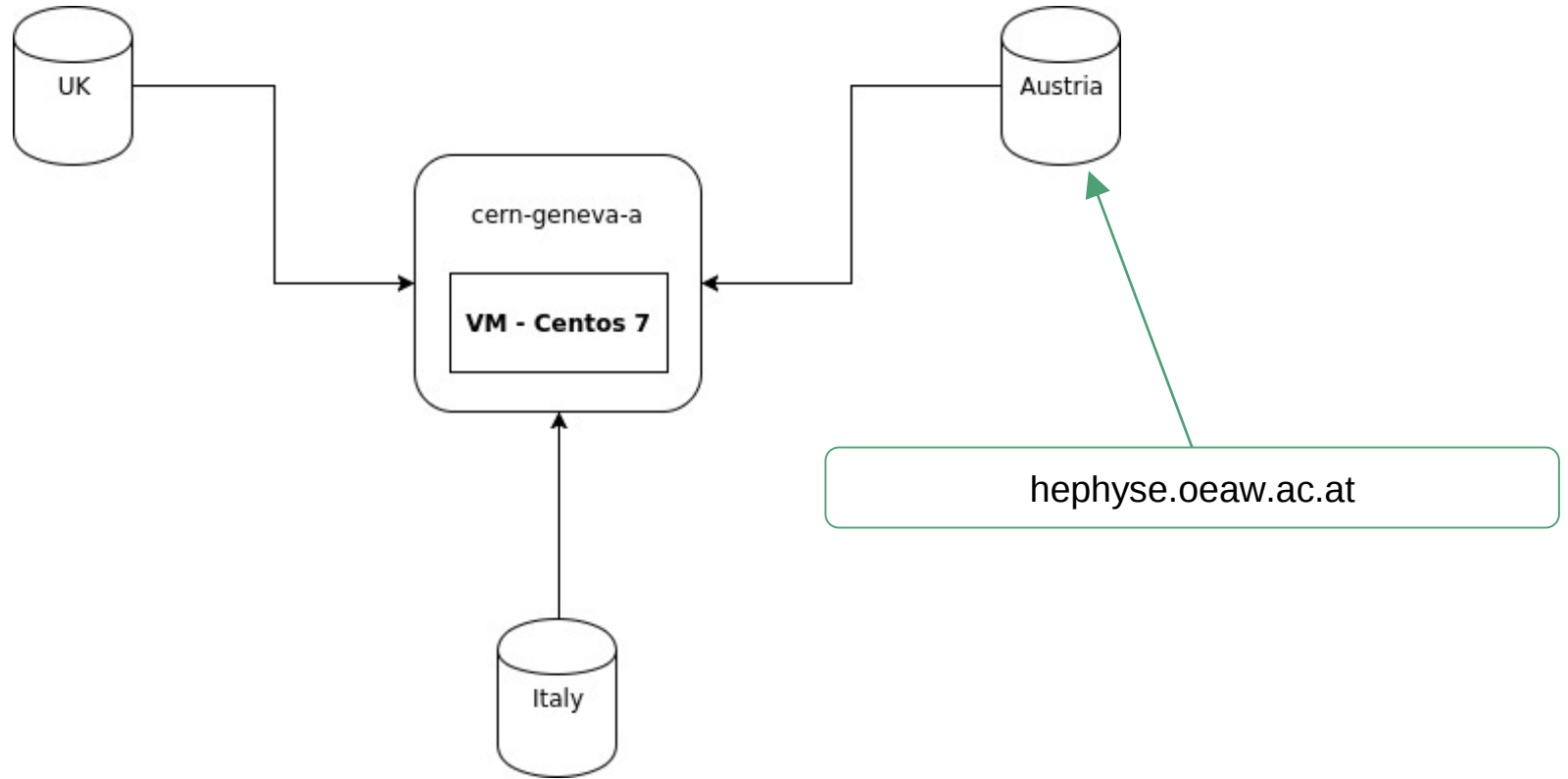
Environment



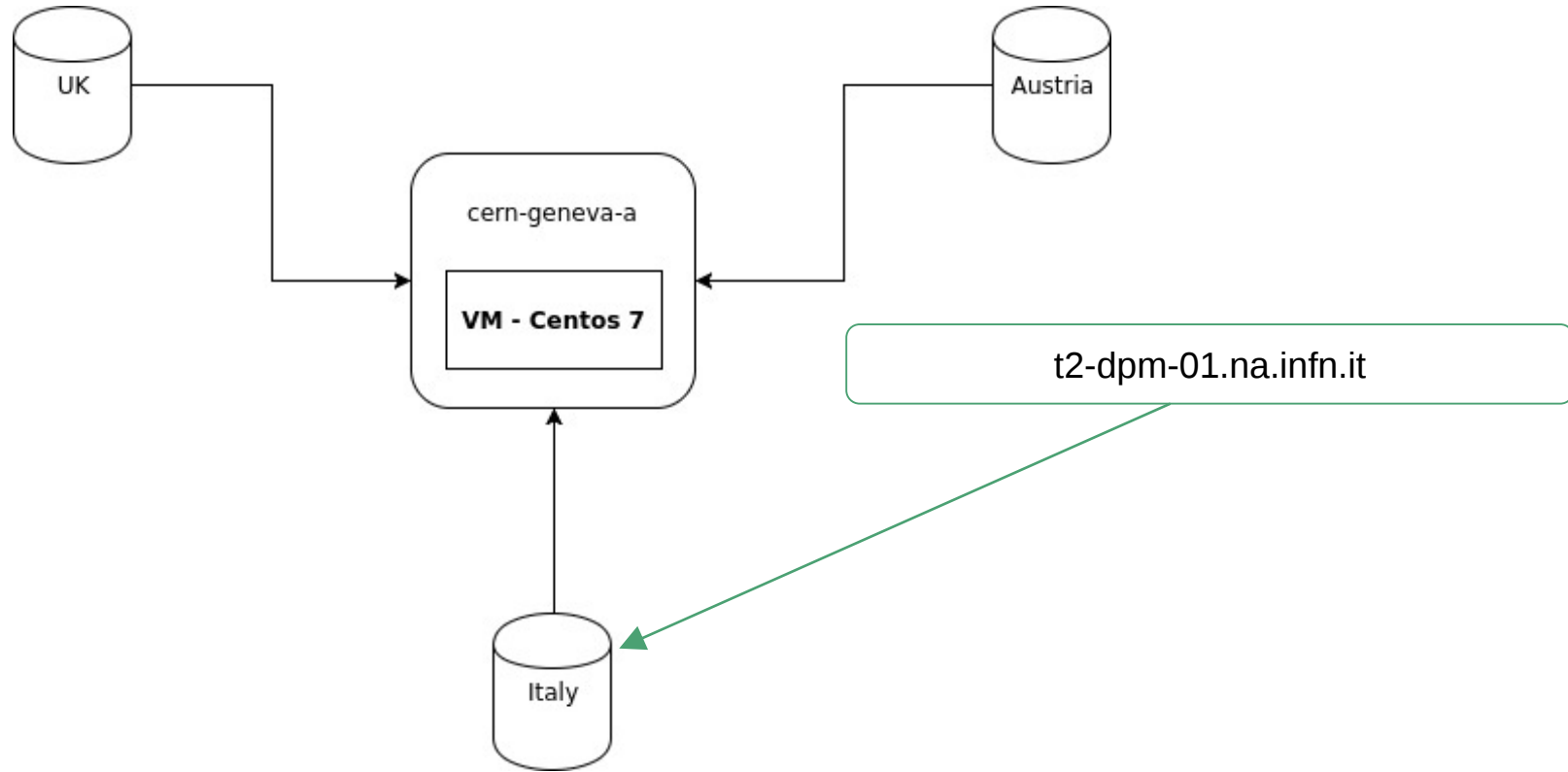
Environment



Environment



Environment

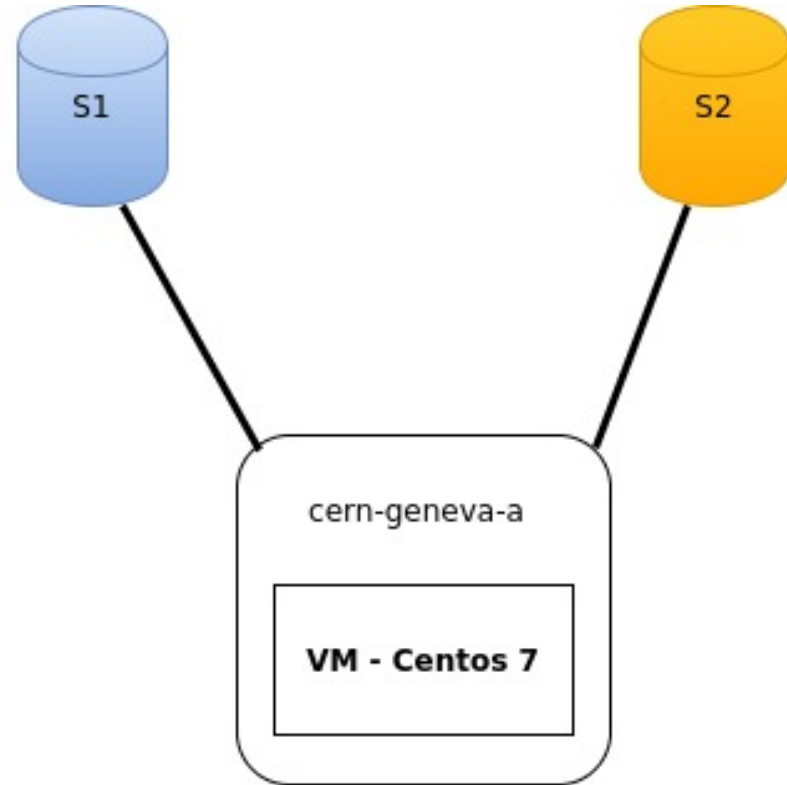


Aria2 - Multi-Source Download Default Configuration

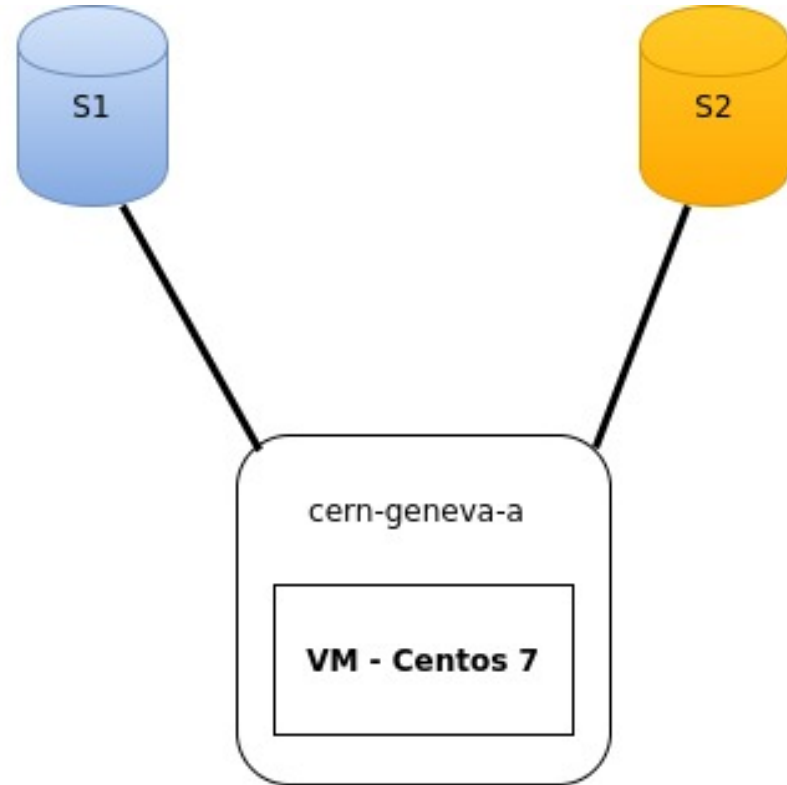
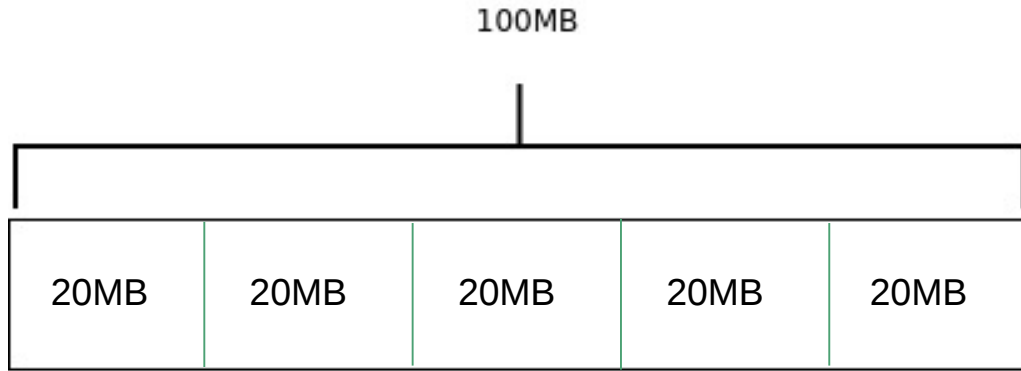
100MB



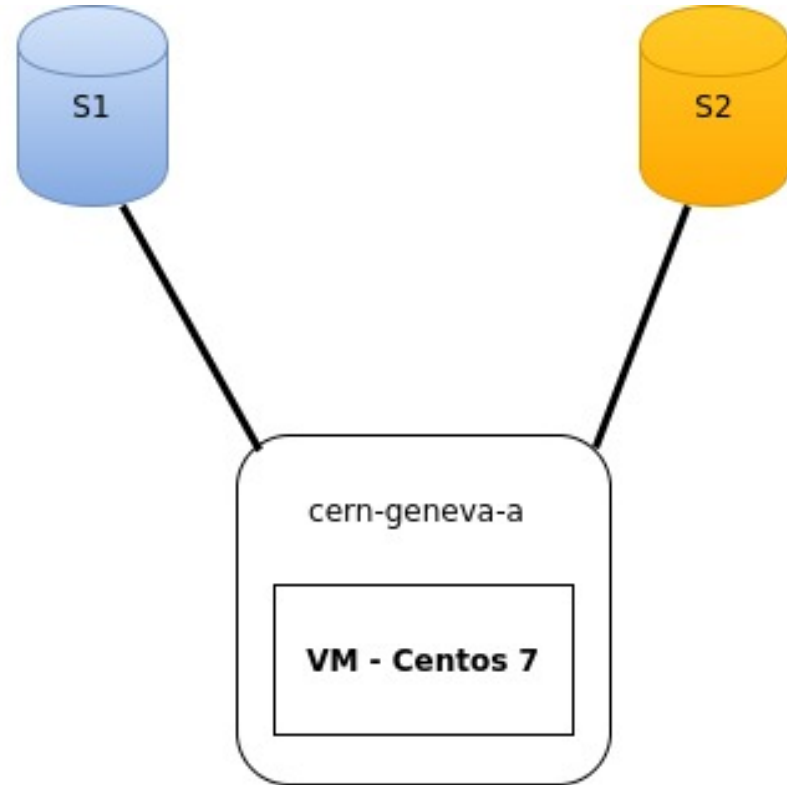
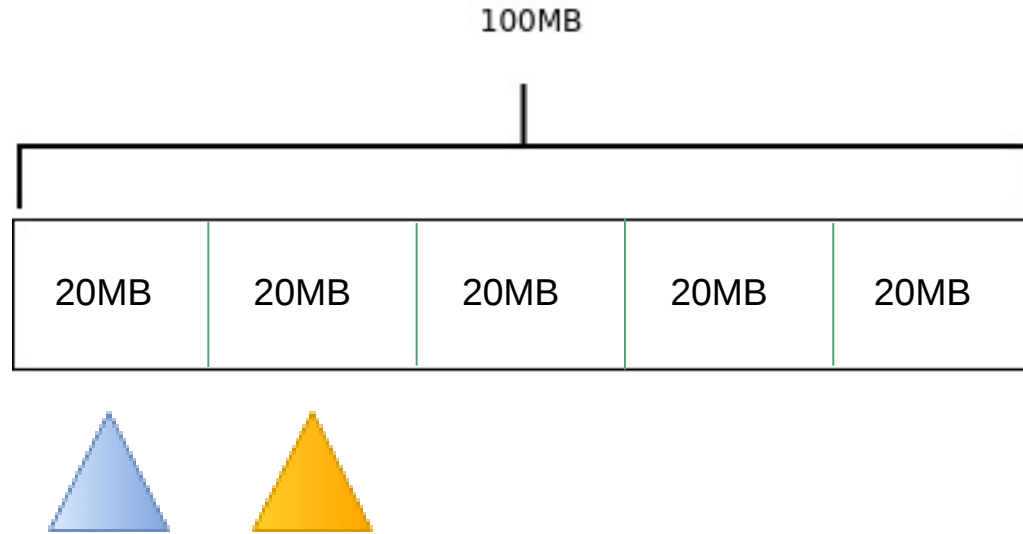
- Maximum 1 connection per Source
- Up to 5 concurrent downloads
- File split into 5 chunks
- Splitting starts from files bigger than 40MB



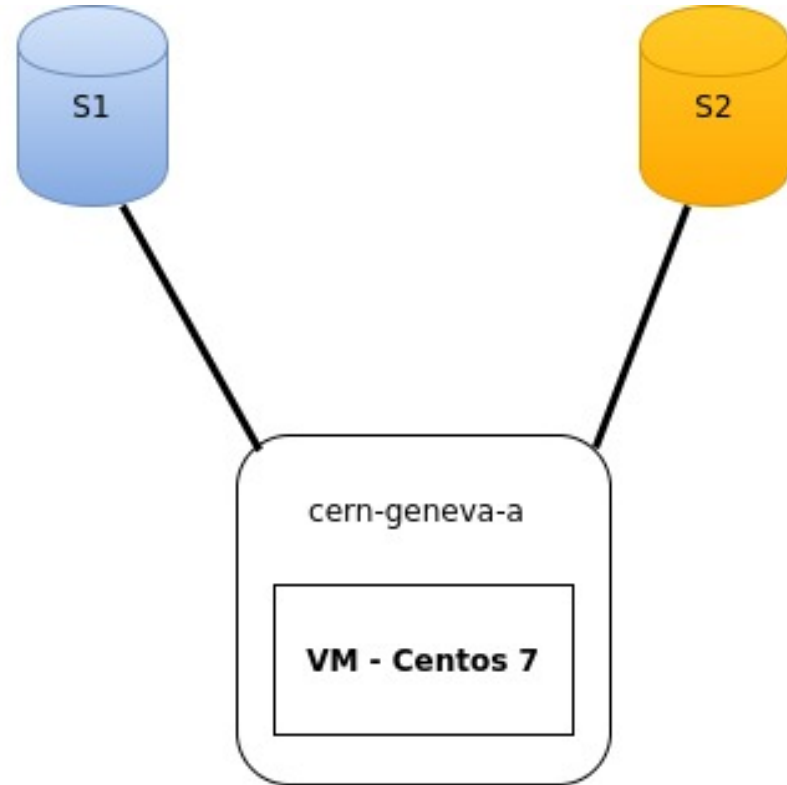
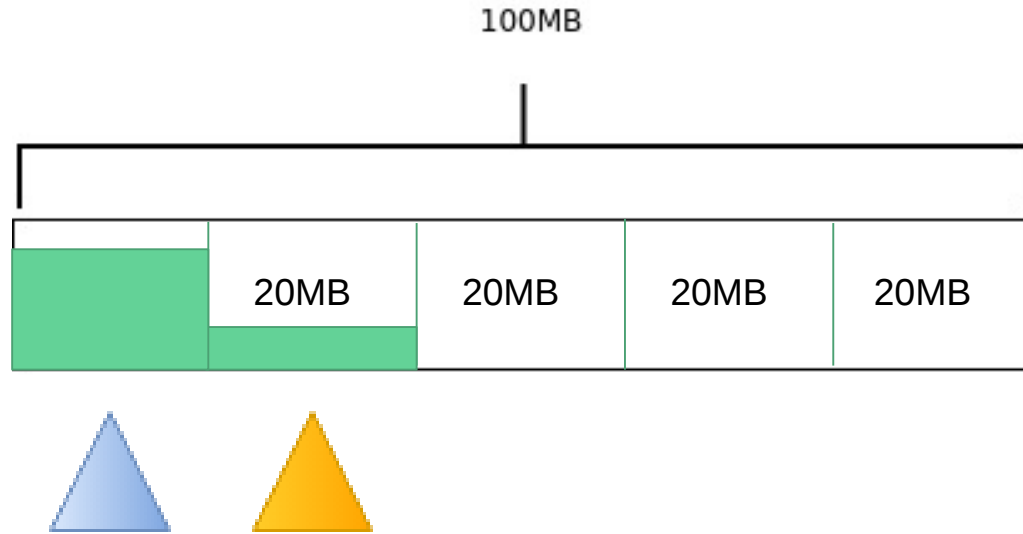
Aria2 - Multi-Source Download Default Configuration



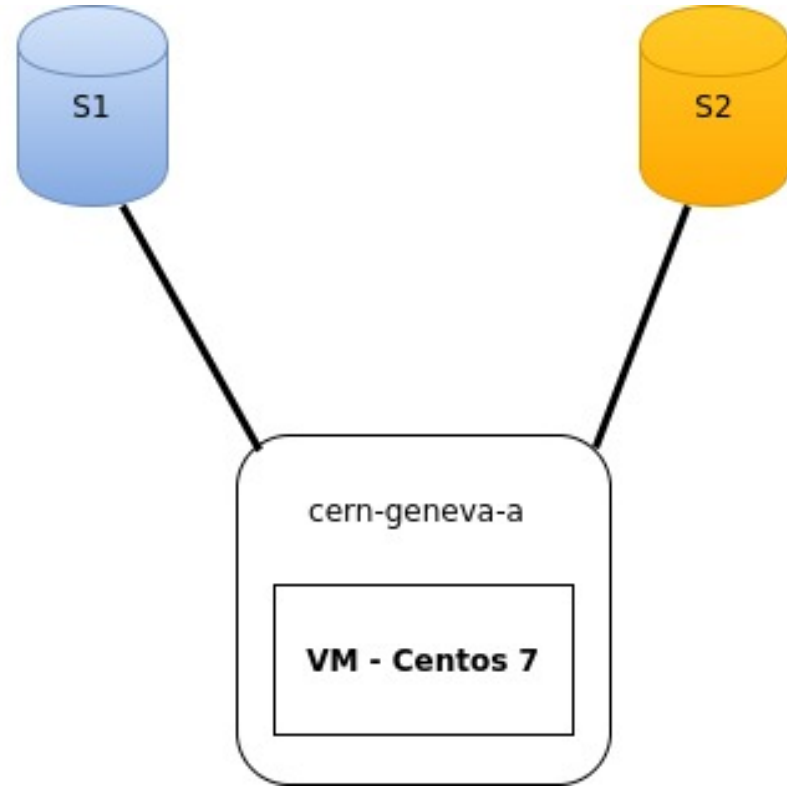
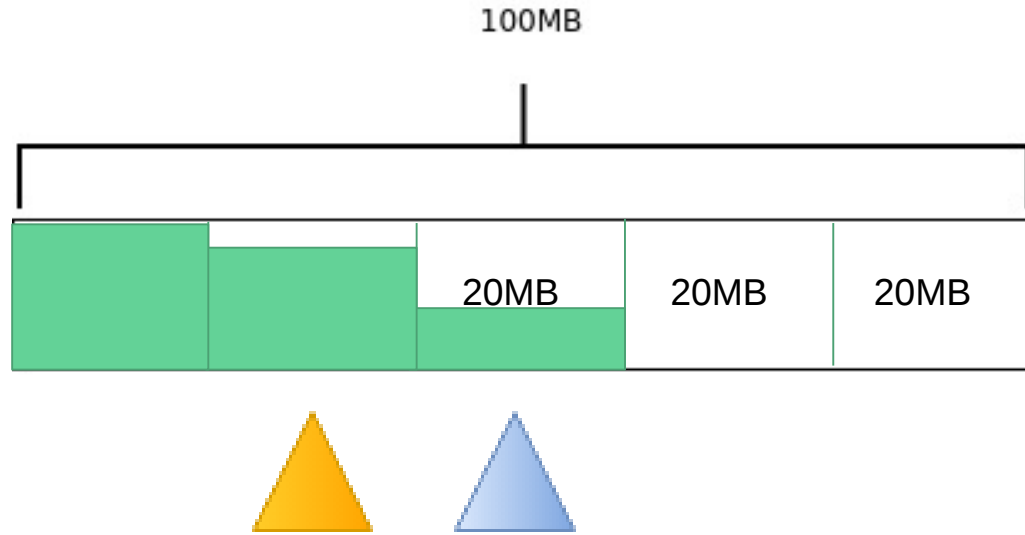
Aria2 - Multi-Source Download Default Configuration



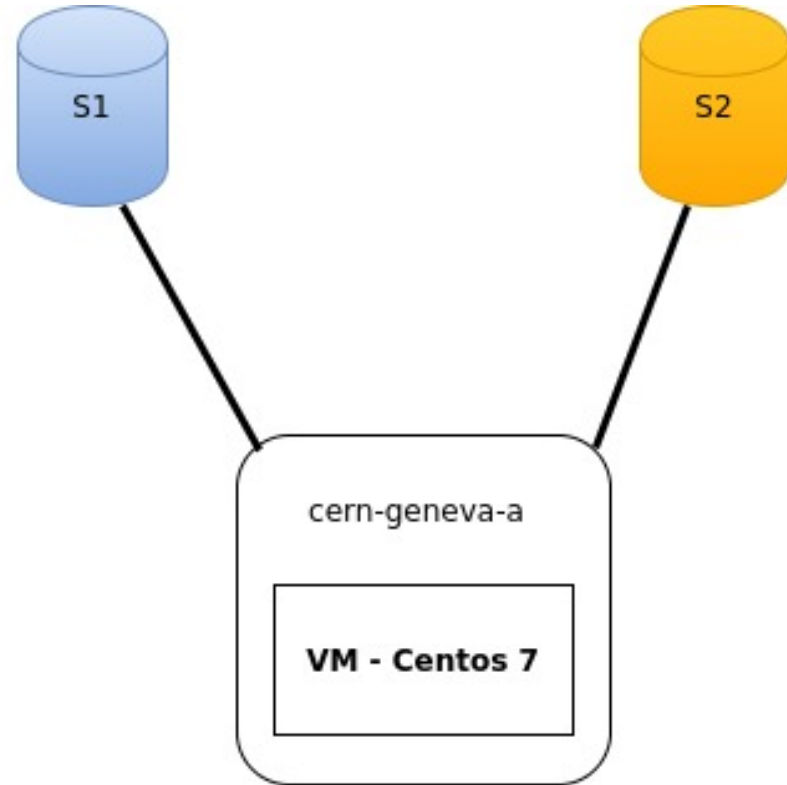
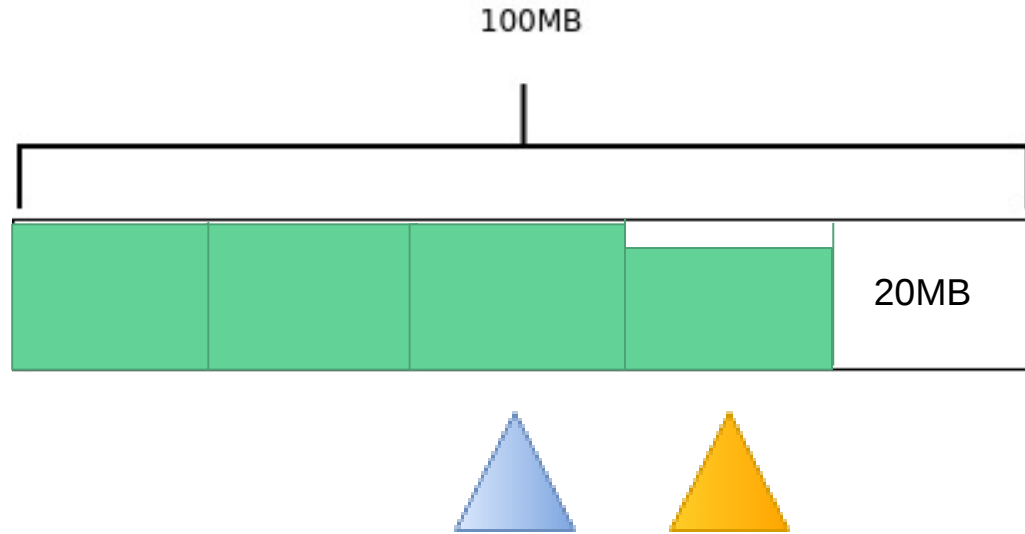
Aria2 - Multi-Source Download Default Configuration



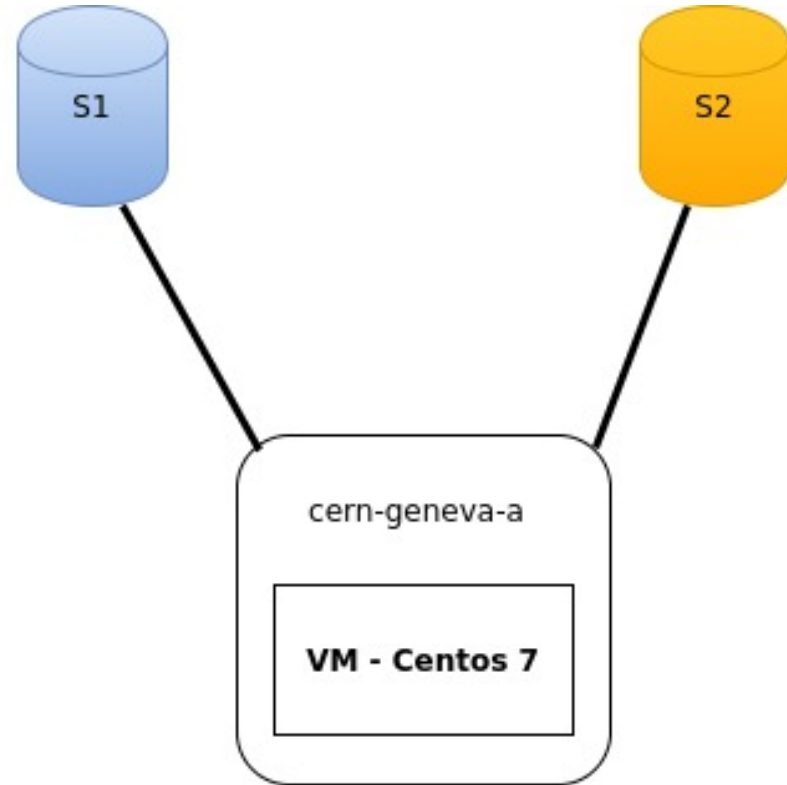
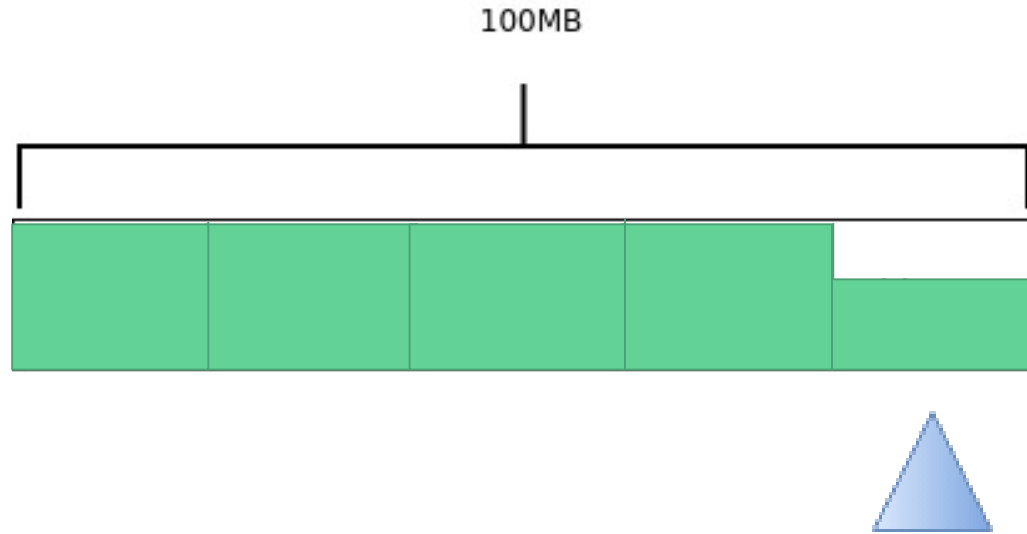
Aria2 - Multi-Source Download Default Configuration



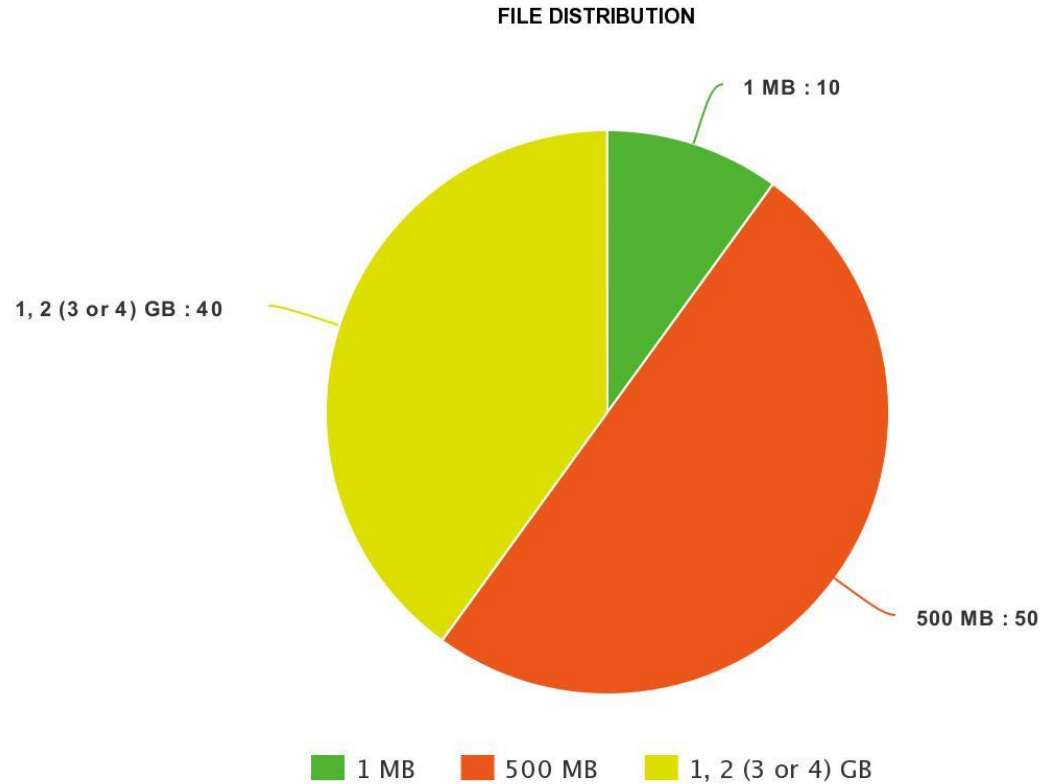
Aria2 - Multi-Source Download Default Configuration



Aria2 - Multi-Source Download Default Configuration



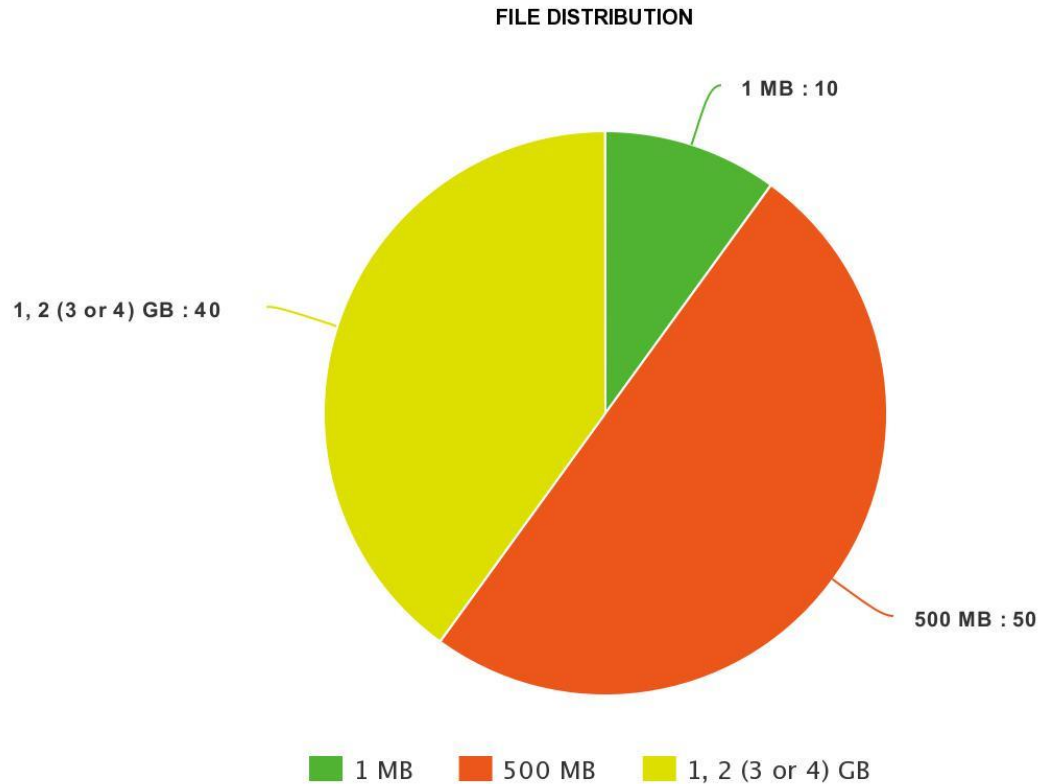
Testing Data



Testing Data

Files used in tests:

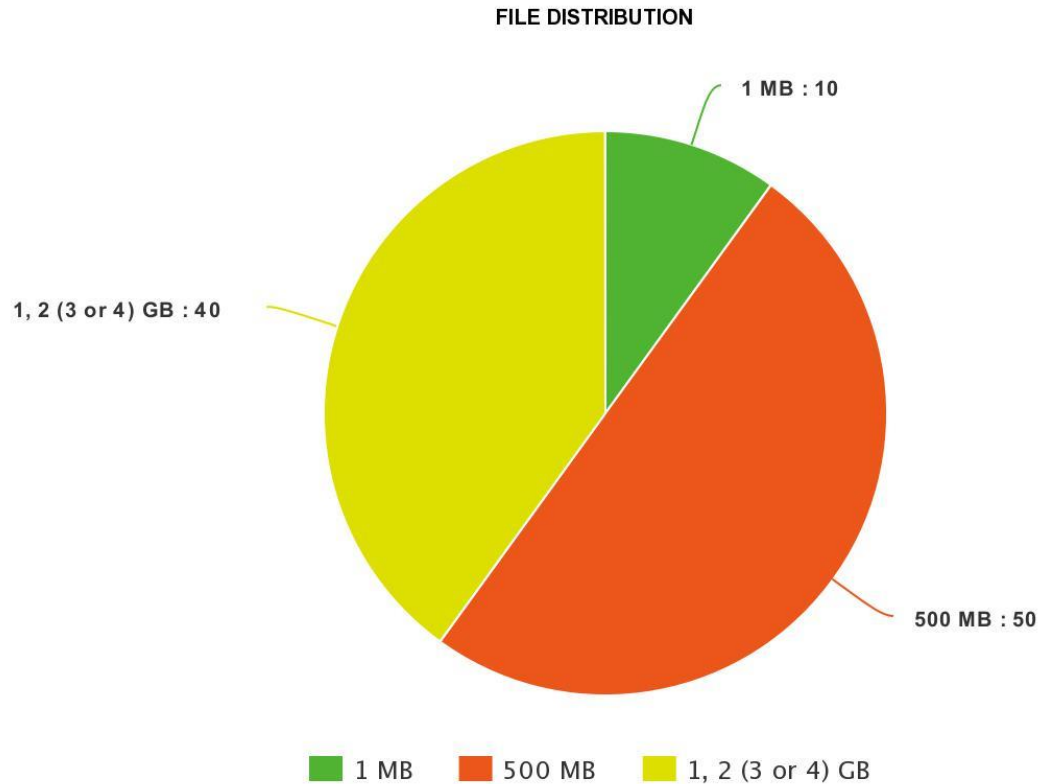
- 1 MB



Testing Data

Files used in tests:

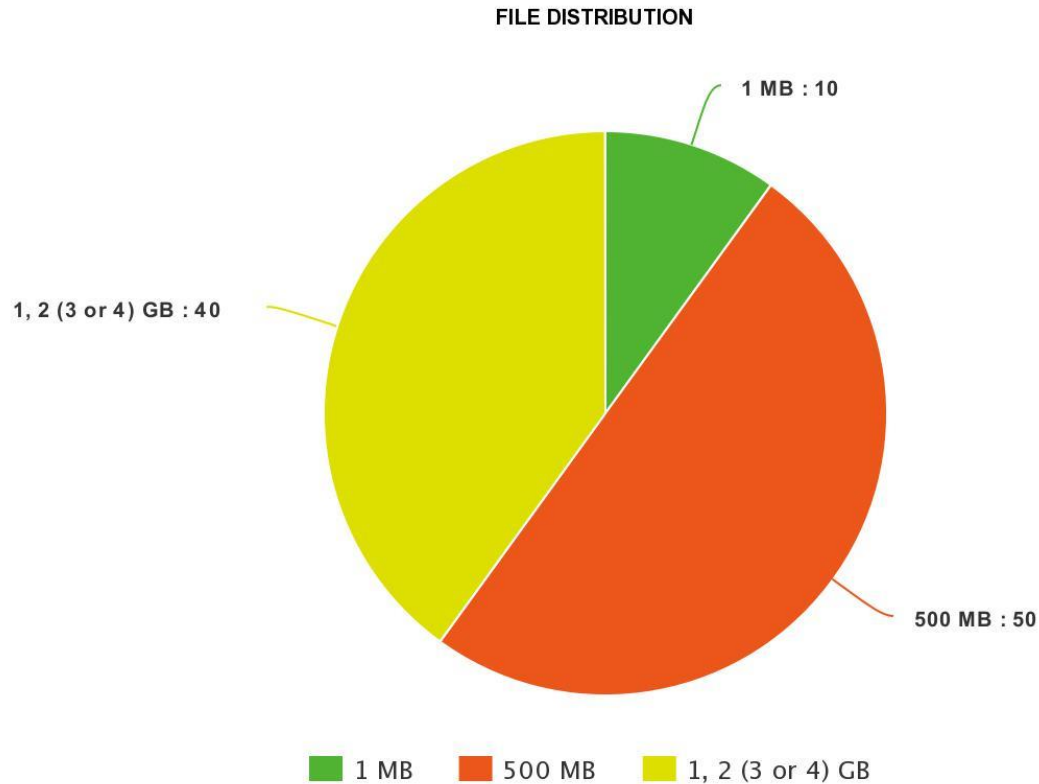
- 1 MB
- 500 MB



Testing Data

Files used in tests:

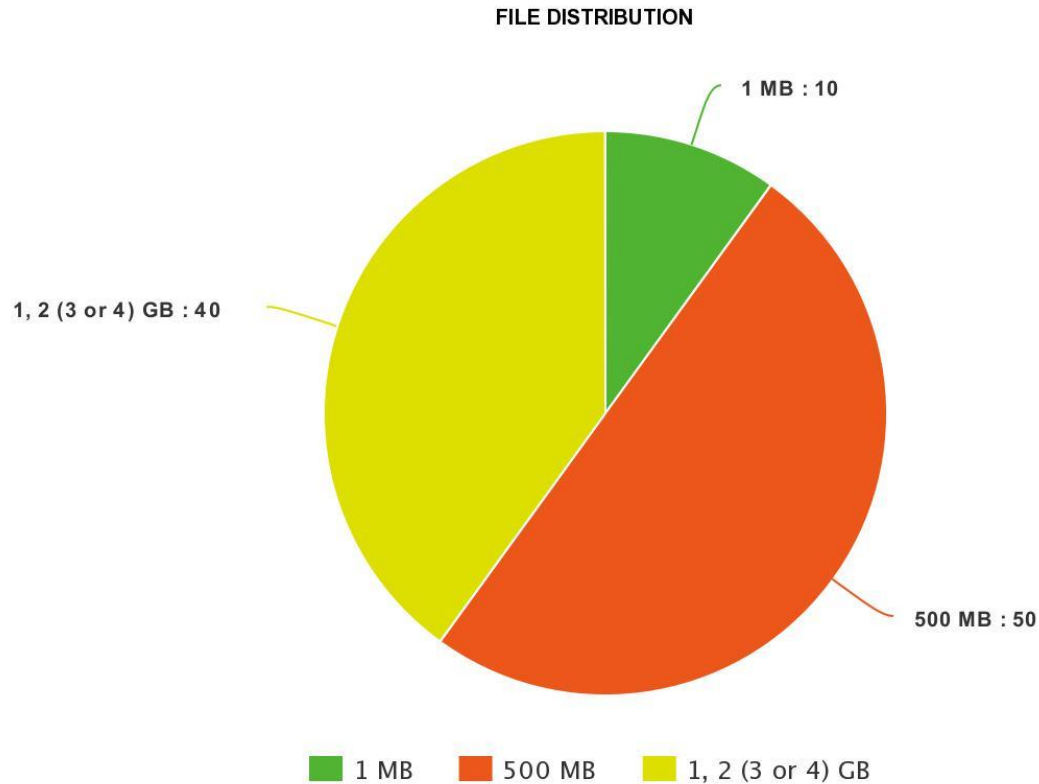
- 1 MB
- 500 MB
- 1 GB



Testing Data

Files used in tests:

- 1 MB
- 500 MB
- 1 GB
- 2 GB



Methodology and Parameters

Step 1

Creating the dteam
proxy needed for
authentication and
endpoint access

Methodology and Parameters



Step 1

Creating the dteam
proxy needed for
authentication and
endpoint access

Step 2

Getting aria2 to
download from
endpoints

Methodology and Parameters

Step 1

Creating the dteam proxy needed for endpoint access

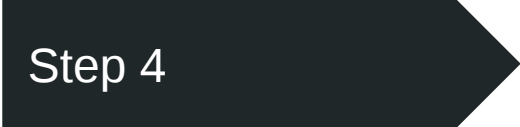
Step 2

Getting aria2 to download from endpoints

Step 3

Testing the endpoints and selecting those that support HTTPS

Methodology and Parameters



Step 4

Isolate endpoints that
also have R/W rights

Methodology and Parameters

Step 4

Isolate endpoints that
also have R/W rights

Step 5

Choose 3 countries
where the endpoints
should be located

Methodology and Parameters

Step 4

Isolate endpoints that
also have R/W rights

Step 5

Choose 3 countries
where the endpoints
should be located

Step 6

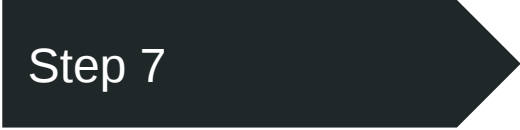
Isolate 3 endpoints that
have support GridFTP
and XROOT too

Methodology and Parameters

Step 7

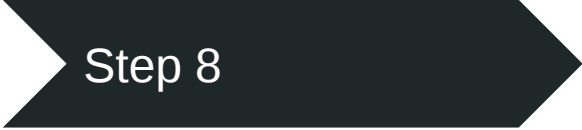
Initialise the 3
endpoints with test
files of different
dimensions

Methodology and Parameters

A dark blue arrow pointing to the right, containing the text 'Step 7' in white.

Step 7

Initialise the 3
endpoints with test
files of different
dimensions

A dark blue arrow pointing to the right, containing the text 'Step 8' in white.

Step 8

Repeatedly
downloading the files
and logging the time,
file size and endpoint
address

Methodology and Parameters

Step 7

Initialise the 3 endpoints with test files of different dimensions

Step 8

Repeatedly downloading the files and logging the time, file size and endpoint address

Step 9

Aggregate the data in order to be able to run test and create charts out of it

Methodology and Parameters

Step 10

Checking if data is
normally distributed
using shapiro test

Methodology and Parameters

Step 10

Checking if data is
normally distributed
using shapiro test

Step 11

Plot the charts
accordingly

Methodology and Parameters

Step 10

Checking if data is
normally distributed
using shapiro test

Step 11

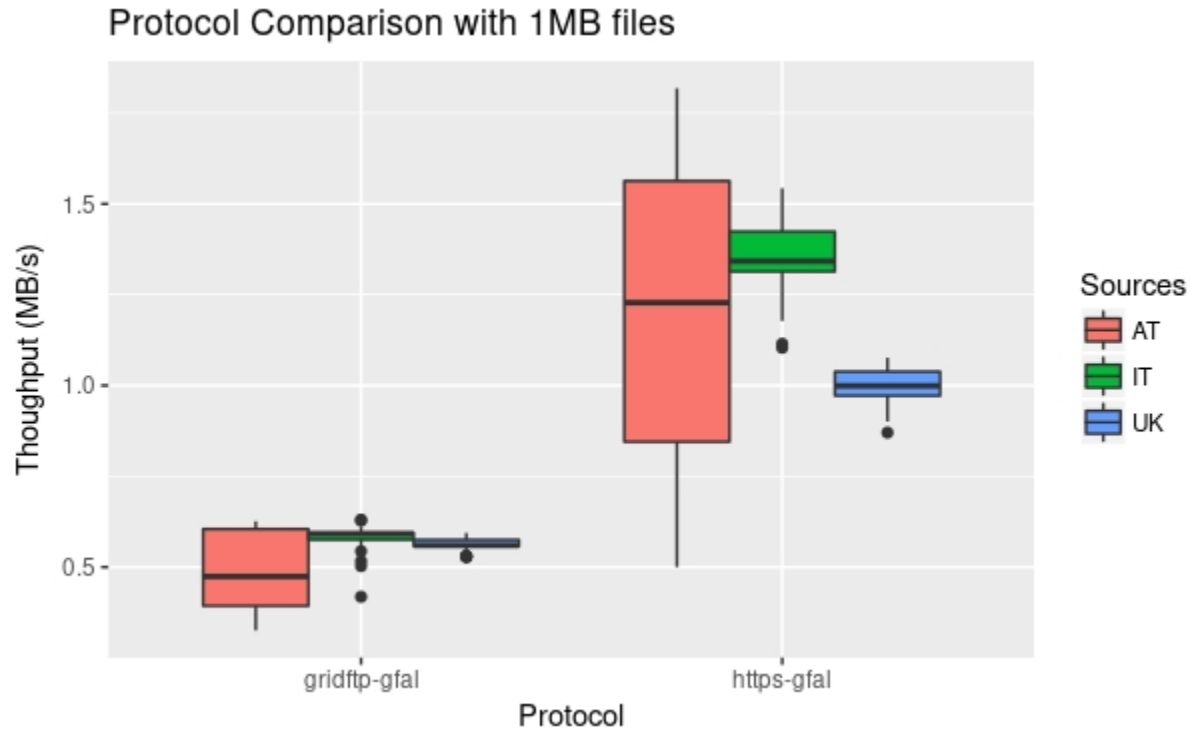
Plot the charts
accordingly

Step 12

Analyse the charts

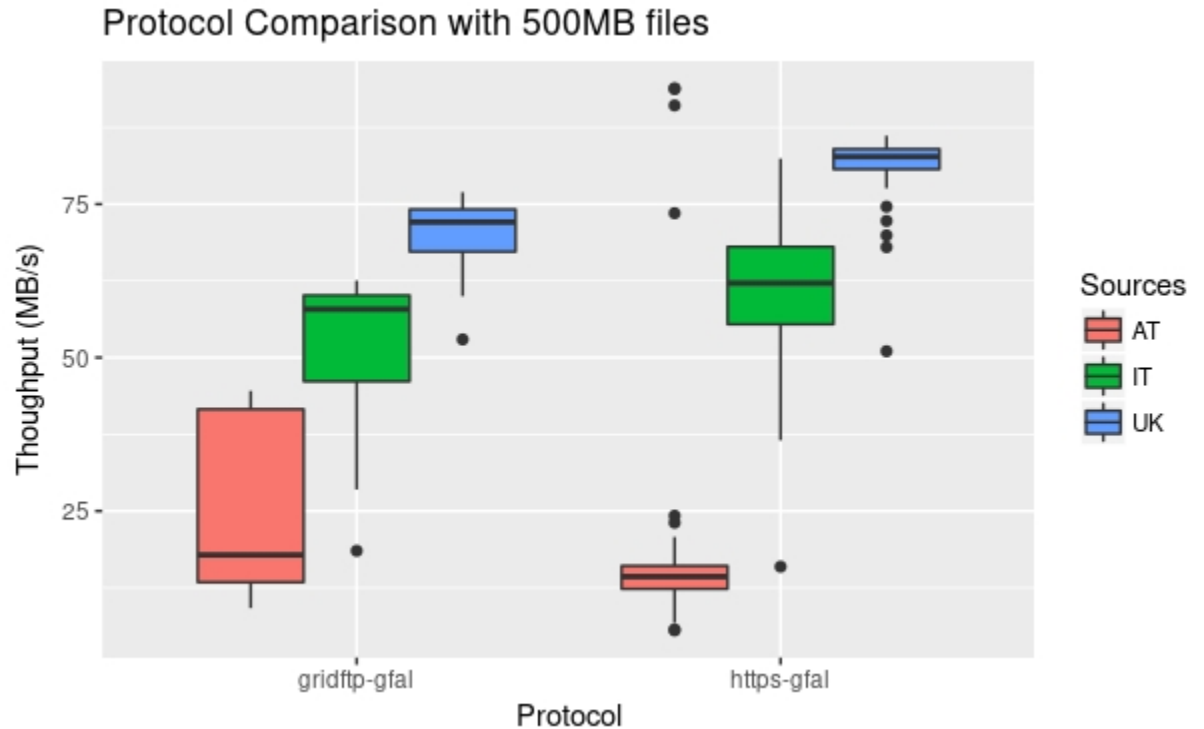
Gfal - HTTPS vs GridFTP

- 1 MB
- 50 Runs
- Gfal with GridFTP
- Gfal with HTTPS



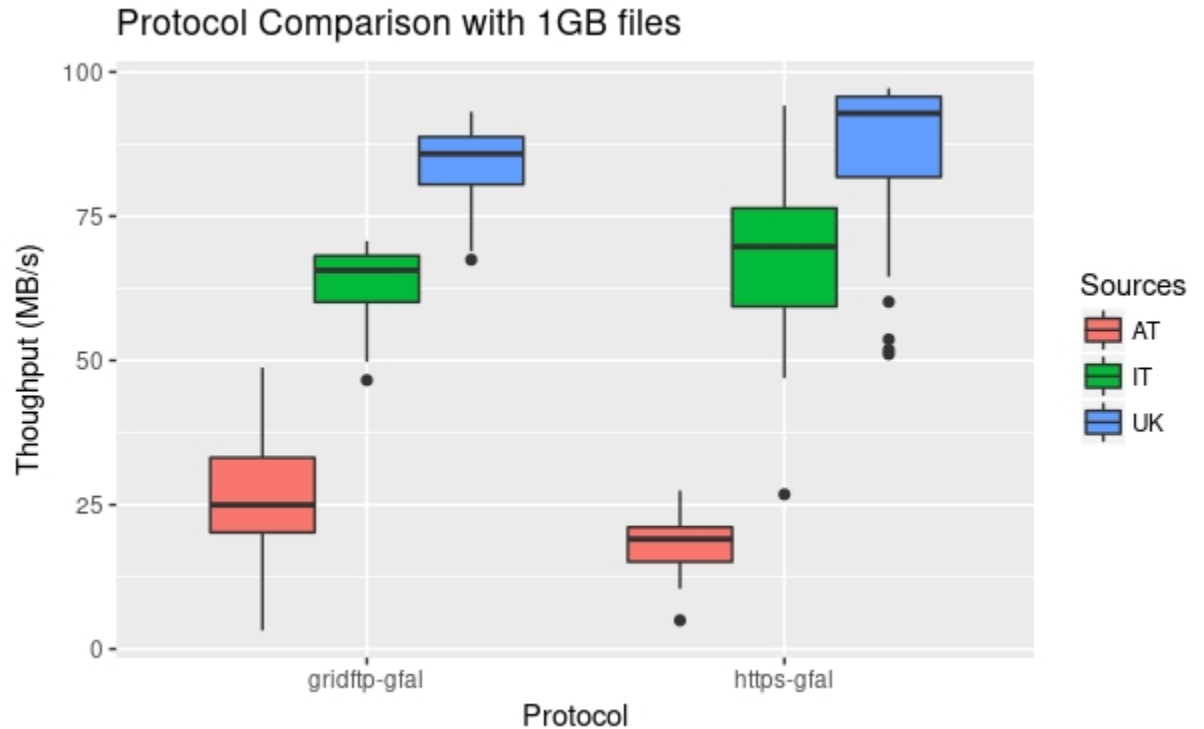
Gfal - HTTPS vs GridFTP

- 500 MB
- 50 Runs
- Gfal with GridFTP
- Gfal with HTTPS



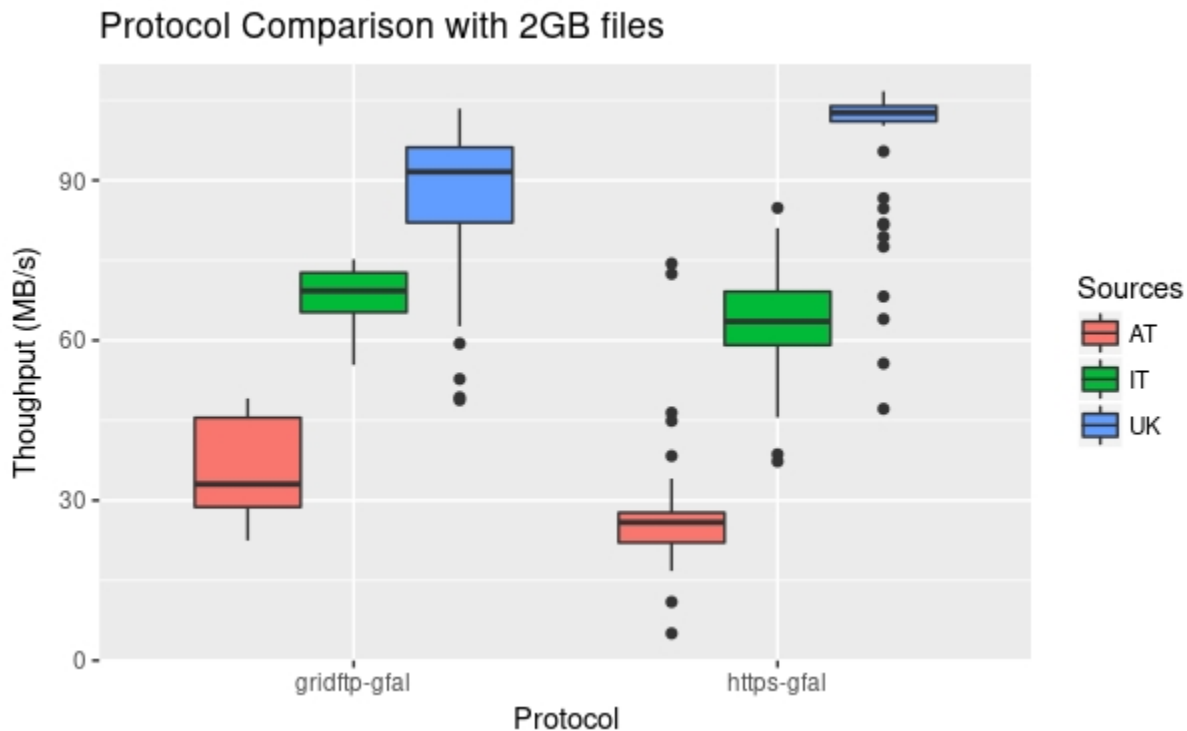
Gfal - HTTPS vs GridFTP

- 1 GB
- 50 Runs
- Gfal with GridFTP
- Gfal with HTTPS



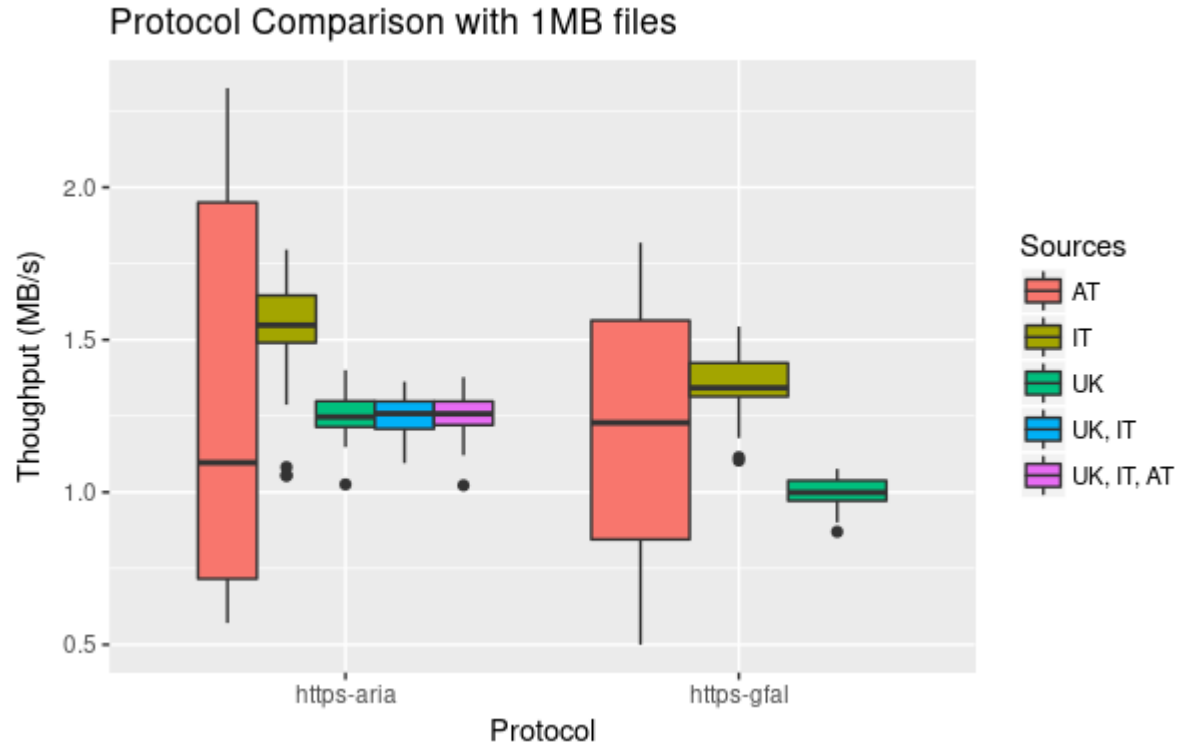
Gfal - HTTPS vs GridFTP

- 2 GB
- 50 Runs
- Gfal with GridFTP
- Gfal with HTTPS



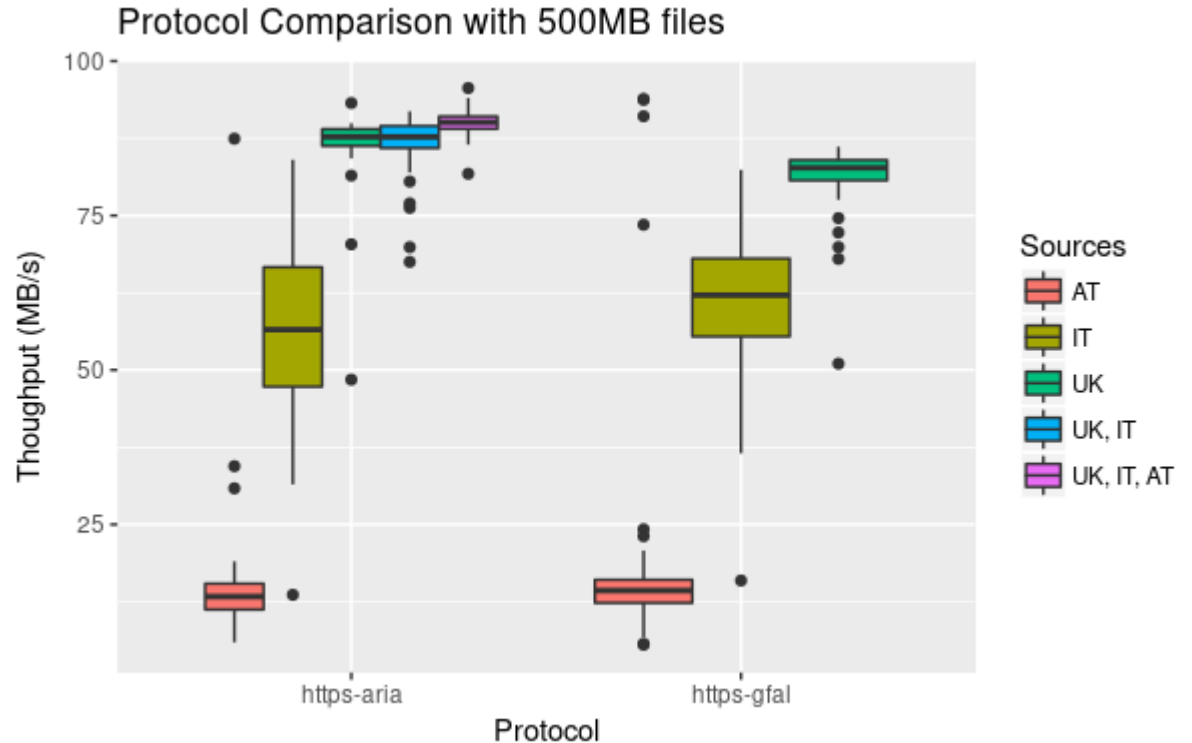
Gfal vs Aria2 - HTTPS

- 1 MB
- 50 Runs
- Aria2 with HTTPS
- Gfal with HTTPS



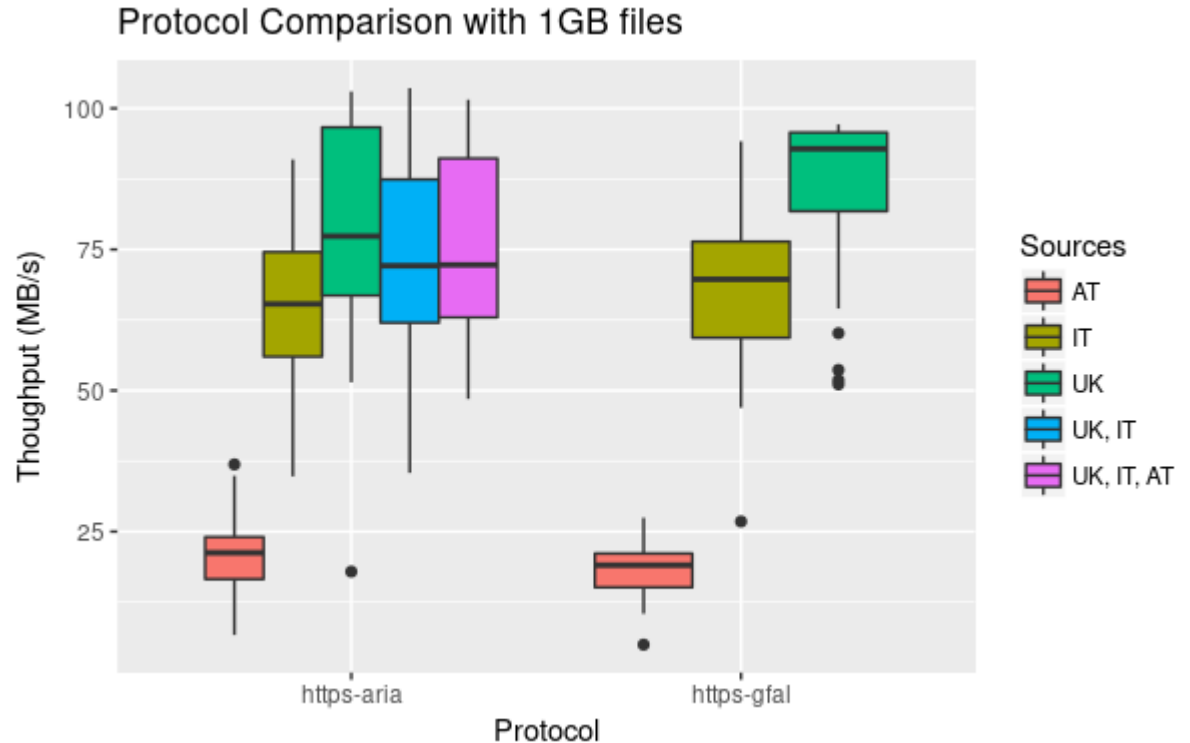
Gfal vs Aria2 - HTTPS

- 500 MB
- 50 Runs
- Aria2 with HTTPS
- Gfal with HTTPS



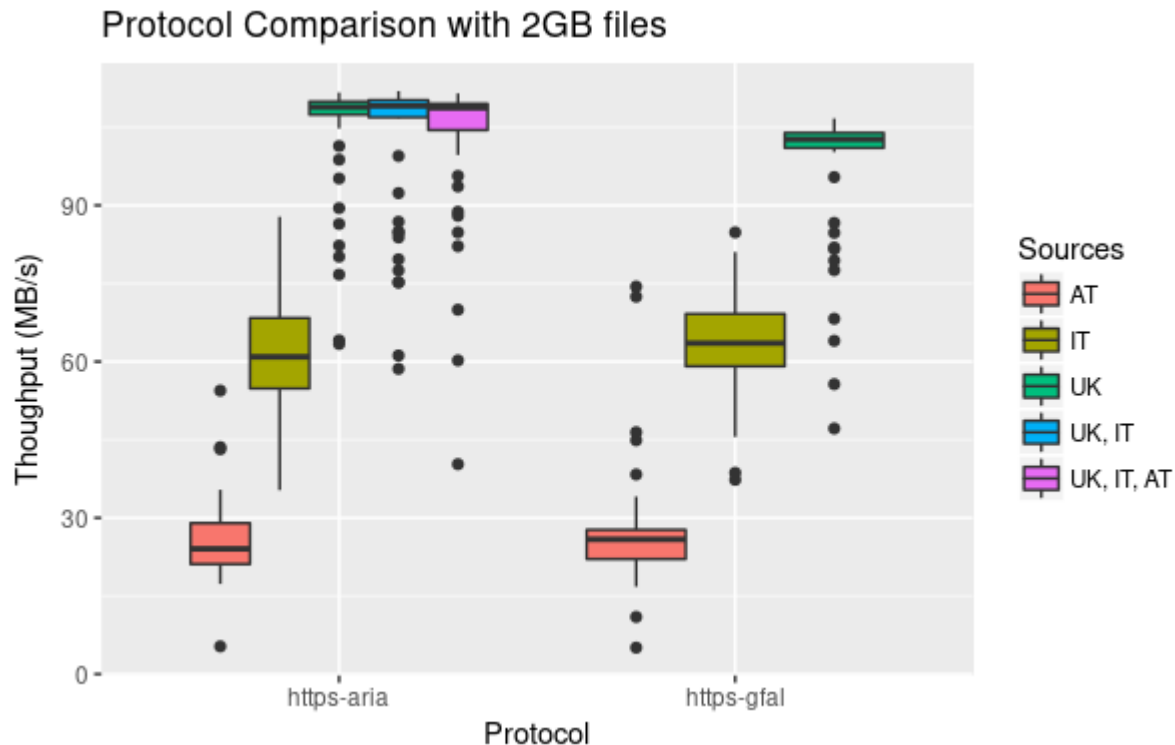
Gfal vs Aria2 - HTTPS

- 1 GB
- 50 Runs
- Aria2 with HTTPS
- Gfal with HTTPS



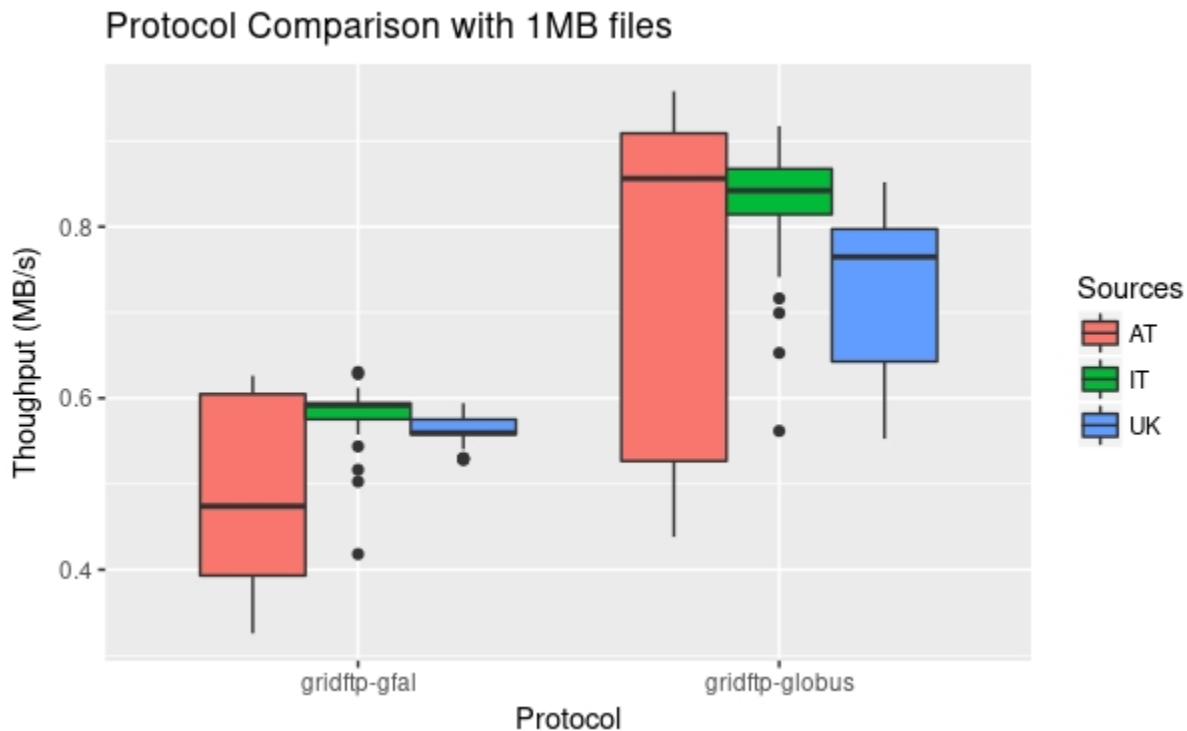
Gfal vs Aria2 - HTTPS

- 2 GB
- 50 Runs
- Aria2 with HTTPS
- Gfal with HTTPS



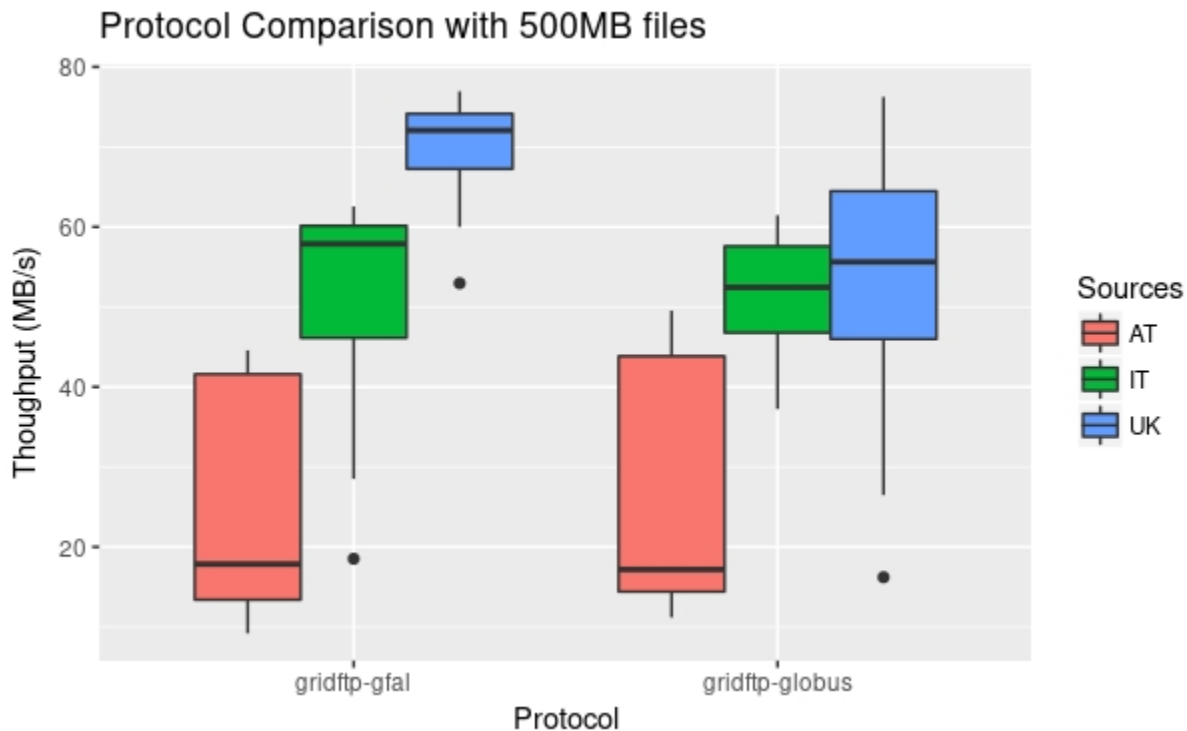
Globus vs Gfal - GridFTP

- 1 MB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP



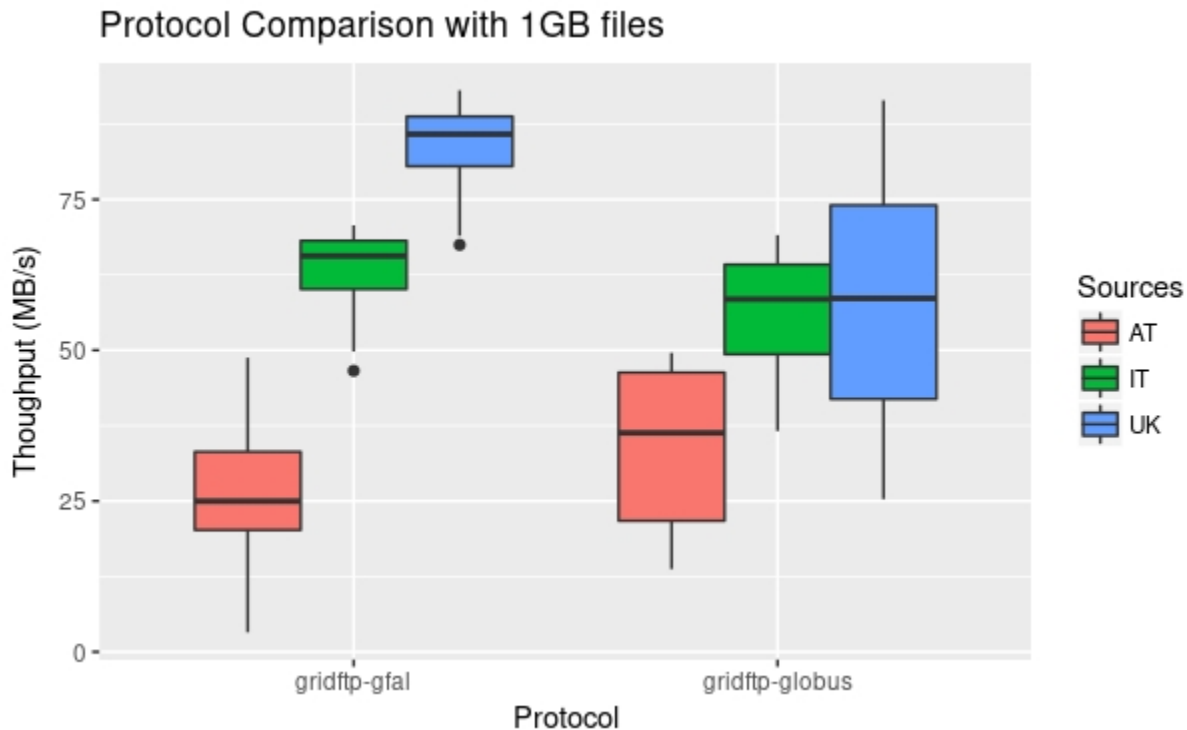
Globus vs Gfal - GridFTP

- 500 MB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP



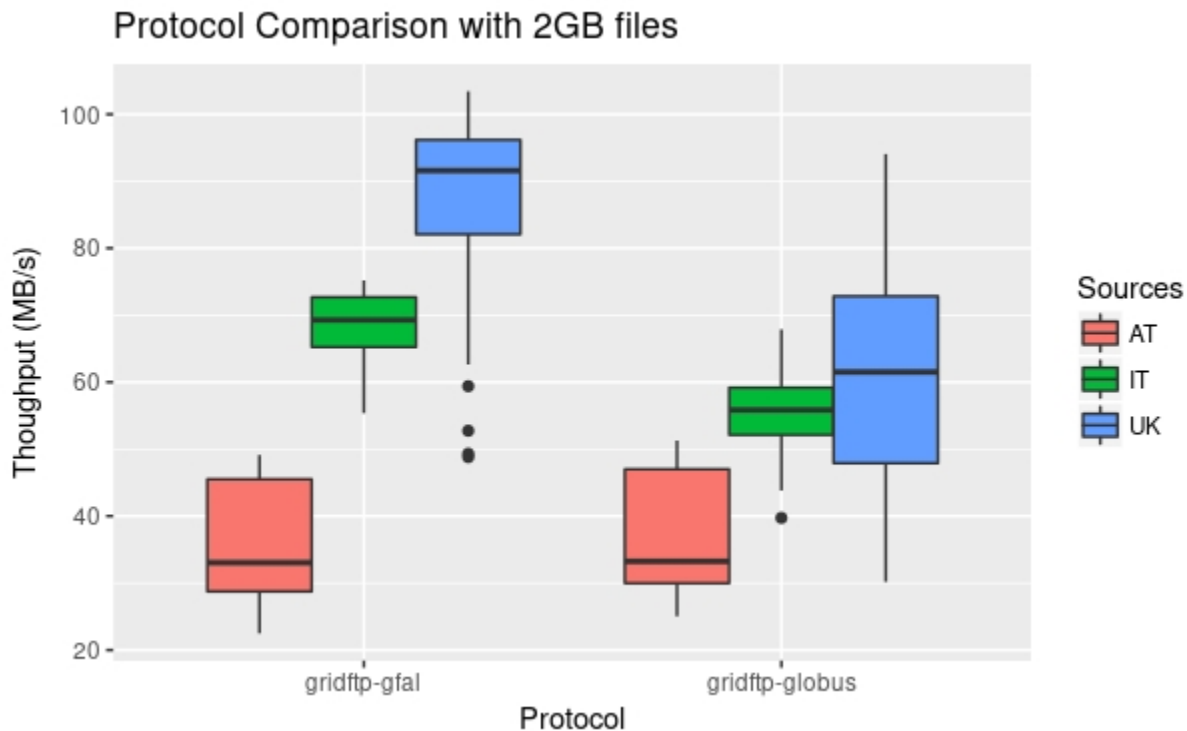
Globus vs Gfal - GridFTP

- 1 GB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP



Globus vs Gfal - GridFTP

- 2 GB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP



Configuration

Aria2

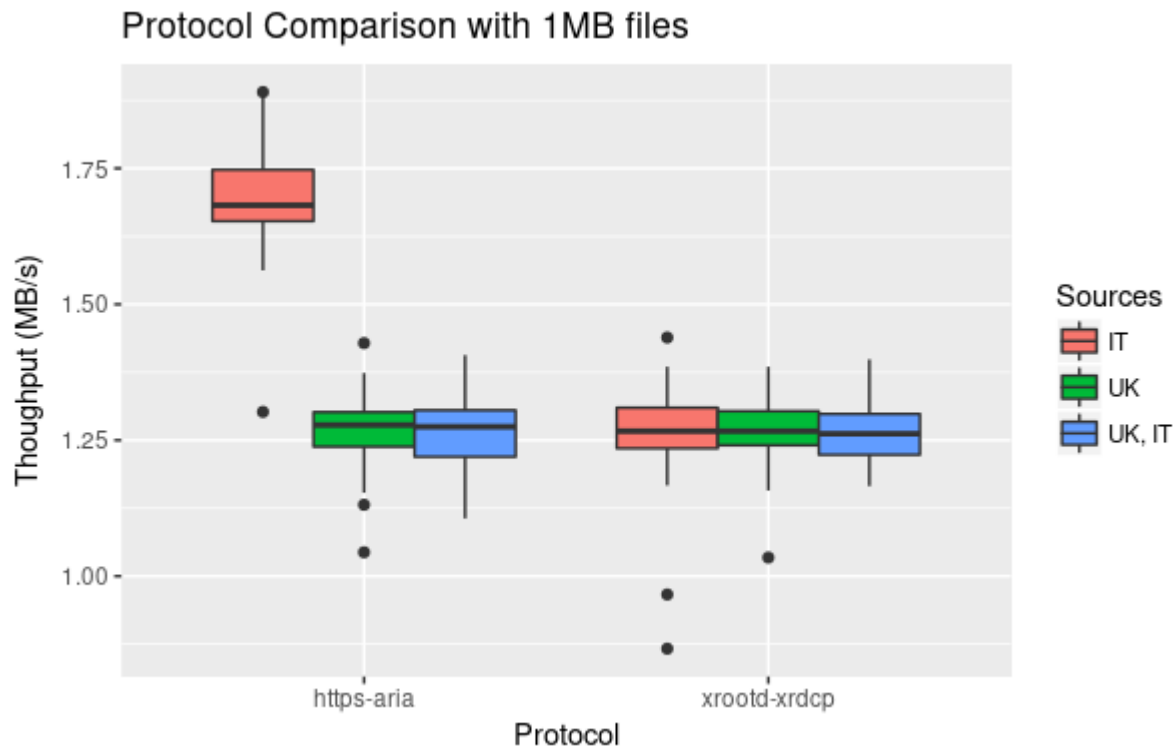
- 4 connection per Source
- 2 concurrent downloads
- File split into 128MB chunks and 125GB chunks for the 500MB files

xrootd

- 4 connection per Source
- 2 concurrent downloads
- File split into 128MB chunks

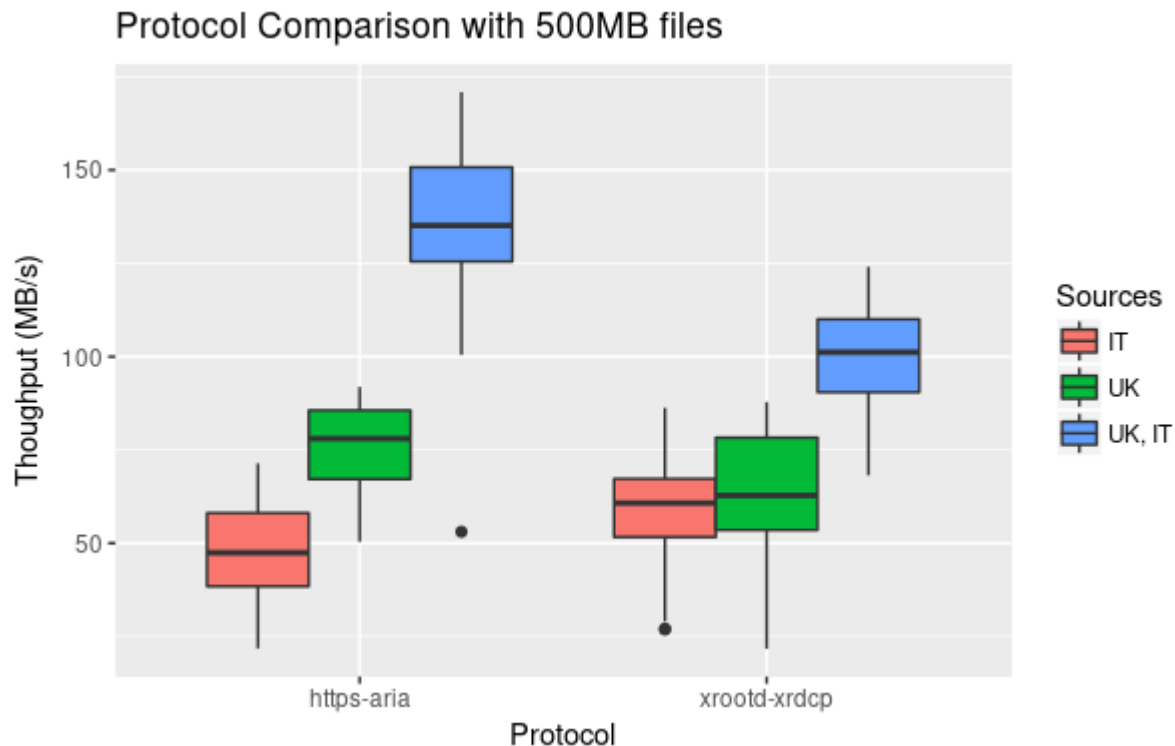
Aria2 vs xrootd - HTTPS - First results

- 1 MB
- 50 Runs
- Aria2 with HTTPS
- xrootd with HTTPS



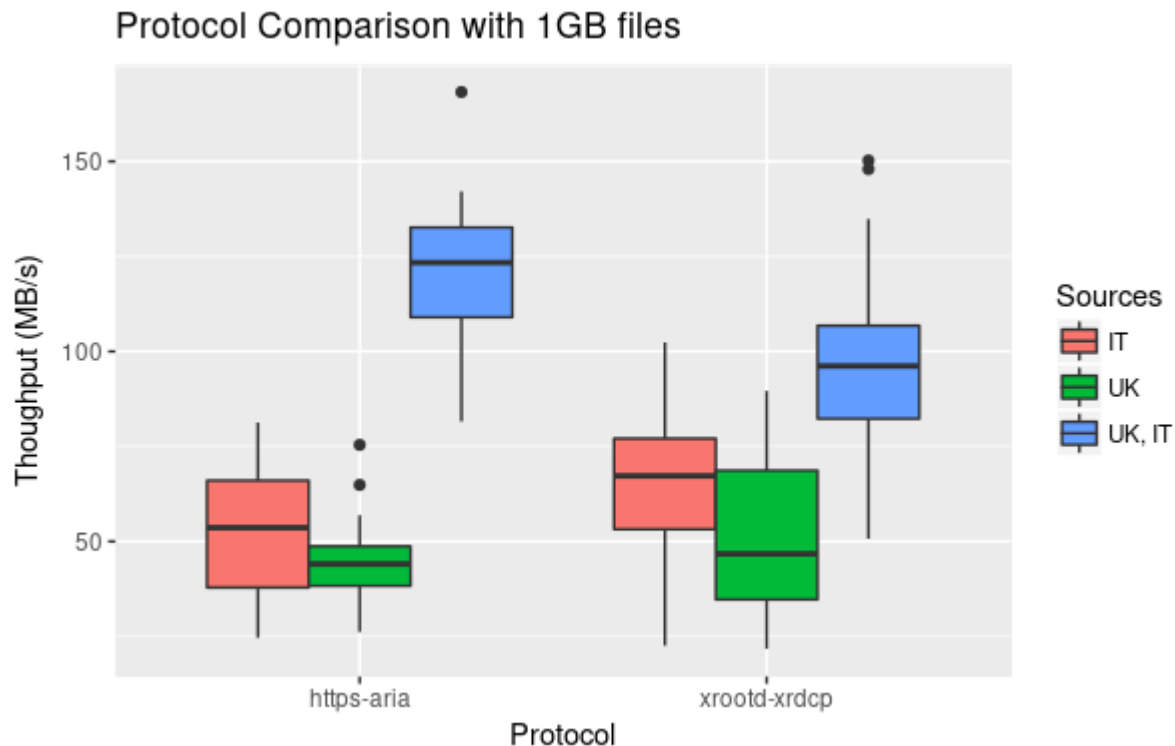
Aria2 vs xrootd - HTTPS - First results

- 500 MB
- 50 Runs
- Aria2 with HTTPS
- xrootd with HTTPS



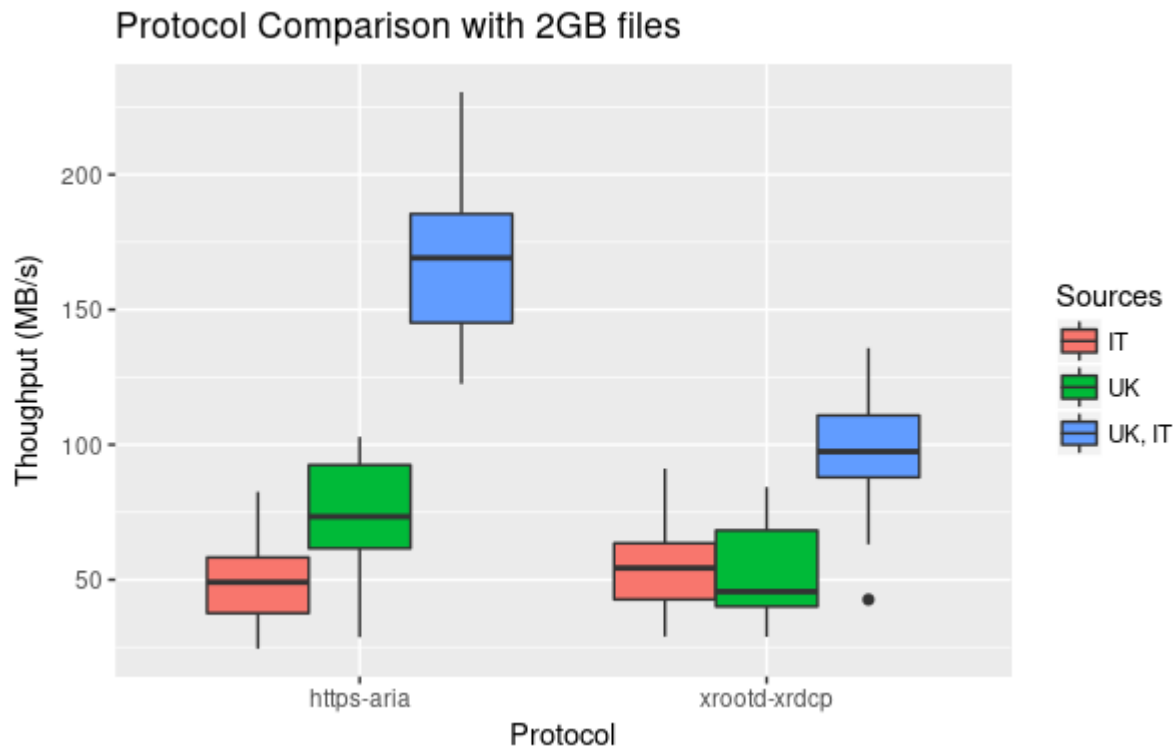
Aria2 vs xrootd - HTTPS - First results

- 1 GB
- 50 Runs
- Aria2 with HTTPS
- xrootd with HTTPS



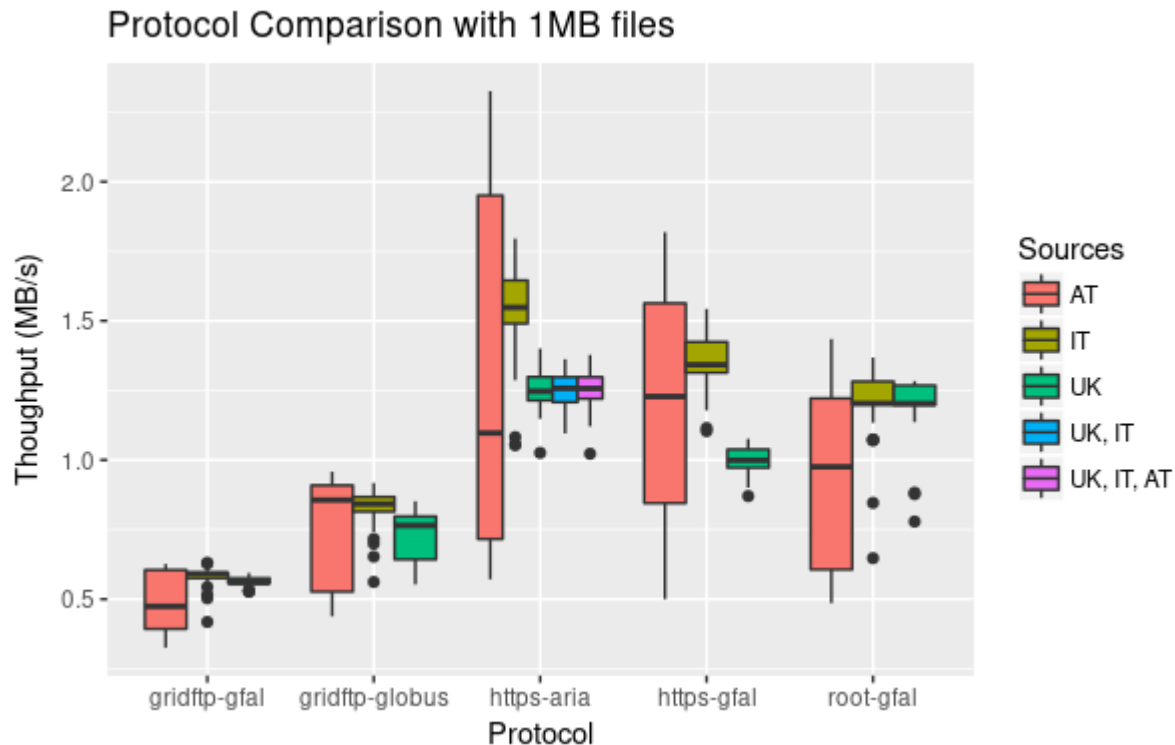
Aria2 vs xrootd - HTTPS - First results

- 2 GB
- 50 Runs
- Aria2 with HTTPS
- xrootd with HTTPS



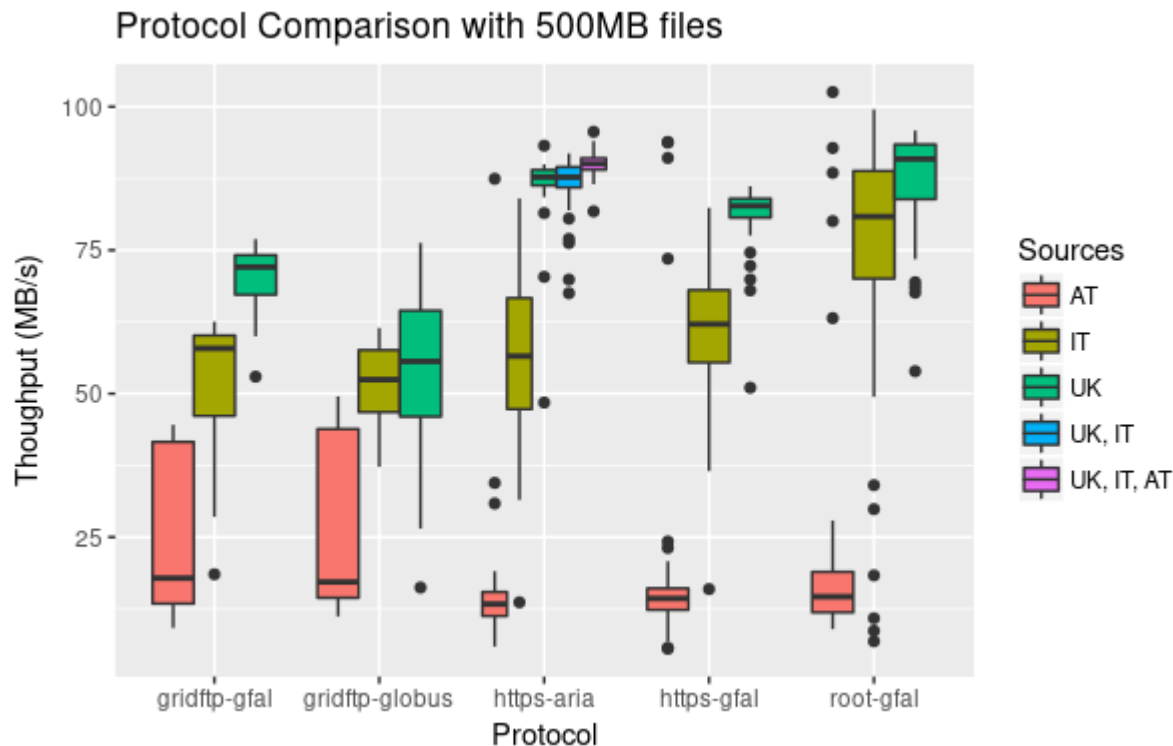
Full Comparison

- 1 MB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP
- Aria2 with HTTPS
- Gfal with HTTPS
- Gfal with XROOT



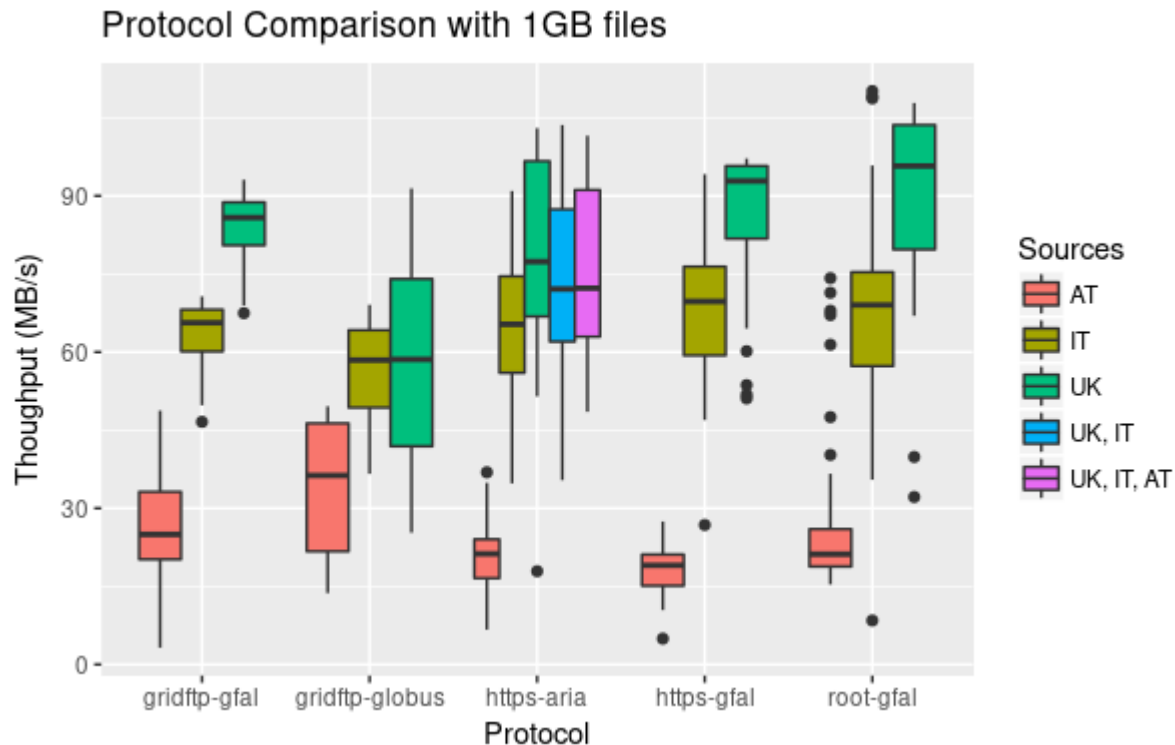
Full Comparison

- 500 MB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP
- Aria2 with HTTPS
- Gfal with HTTPS
- Gfal with XROOT



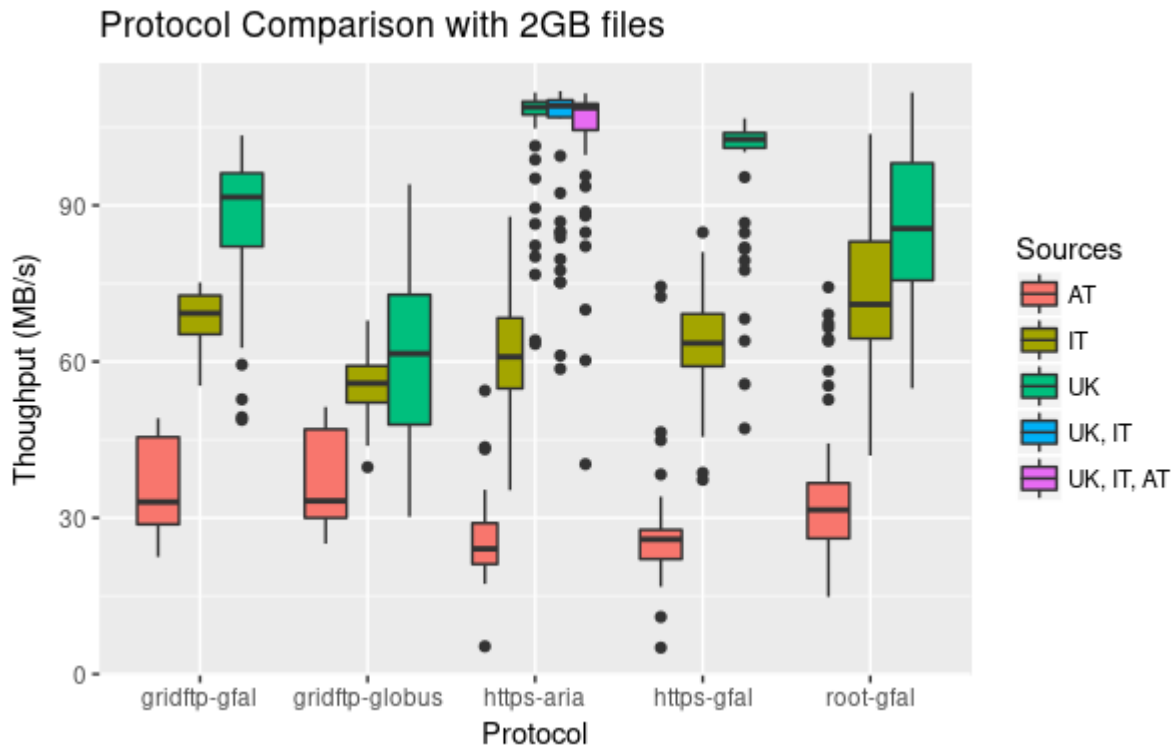
Full Comparison

- 1 GB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP
- Aria2 with HTTPS
- Gfal with HTTPS
- Gfal with XROOT



Full Comparison

- 2 GB
- 50 Runs
- Gfal with GridFTP
- Globus with GridFTP
- Aria2 with HTTPS
- Gfal with HTTPS
- Gfal with XROOT



What I learnt!

Aria2

- Working with aria2c
- Compiling aria with CA bundle
- Use aria2c for downloading from single or multiple sources

What I learnt!

Aria2

- Working with aria2c
- Compiling aria with CA bundle
- Use aria2c for downloading from single or multiple sources

Globus

- Use globus for downloading files from endpoint
- Using globus with dteam proxy

What I learnt!

Aria2

- Working with aria2c
- Compiling aria with CA bundle
- Use aria2c for downloading from single or multiple sources

Globus

- Use globus for downloading files from endpoint
- Using globus with dteam proxy

Gfal

- Use gfal for listing, copying, deleting, creating directories and downloading files

What I learnt!



R

- Environment set-up
- Syntax and semantics
- How to create a plot using ggplot
- How to select specific data out of a table

What I learnt!

R

- Environment set-up
- Syntax and semantics
- How to create a plot using ggplot
- How to select specific data out of a table

XROOTD

- Using xrootd client xrdcp for downloading
- Using xrdcp with multi-source download

What I learnt!

R

- Environment set-up
- Syntax and semantics
- How to create a plot using ggplot
- How to select specific data out of a table

XROOTD

- Using xrootd client xrdcp for downloading
- Using xrdcp with multi-source download

Data normality

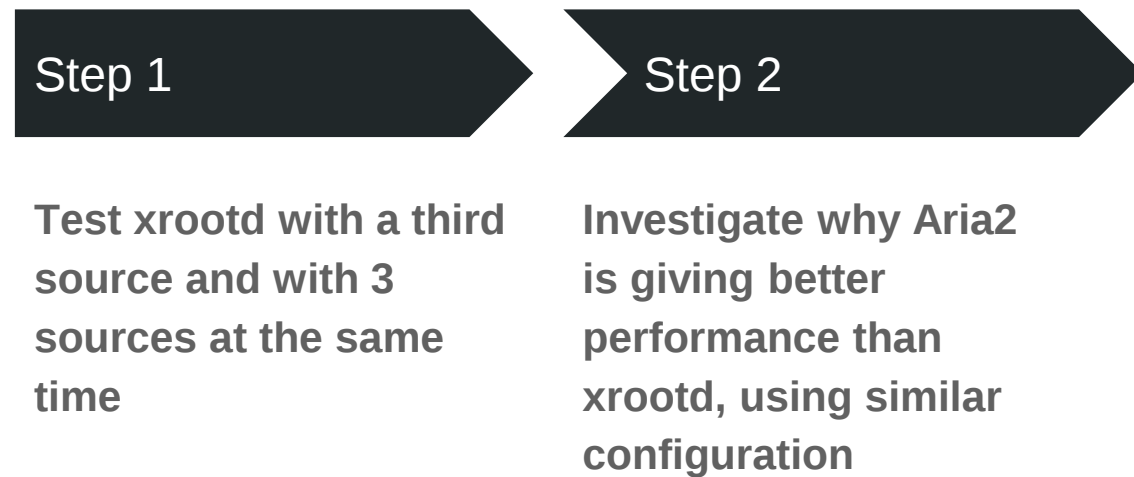
- Testing data to see if it follows a normal distribution using R and Shapiro–Wilk test

Next Steps

Step 1

Test xrootd with a third source and with 3 sources at the same time

Next Steps



Step 1

Test xrootd with a third source and with 3 sources at the same time

Step 2

Investigate why Aria2 is giving better performance than xrootd, using similar configuration

Next Steps

Step 1

Test xrootd with a third source and with 3 sources at the same time

Step 2

Investigate why Aria2 is giving better performance than xrootd, using similar configuration

Step 3

Run the tests for 4 sources, adding a really distant source

Conclusion

Conclusion

- **HTTPS and GridFTP using Gfal provide similar throughput**

Conclusion

- **HTTPS and GridFTP using Gfal provide similar throughput**
- **Downloading from multiple sources with Aria2 gives better throughput or similar to the fastest source with less variance**

Conclusion

- **HTTPS and GridFTP using Gfal provide similar throughput**
- **Downloading from multiple sources with Aria2 gives better throughput or similar to the fastest source with less variance**
- **Using multiple source downloads eliminates the trial and error part of finding the best source**

Conclusion

- **HTTPS and GridFTP using Gfal provide similar throughput**
- **Downloading from multiple sources with Aria2 gives better throughput or similar to the fastest source with less variance**
- **Using multiple source downloads eliminates the trial and error part of finding the best source**
- **When downloading from multiple sources Aria2 isn't able to split files smaller than 2MB, neither xrootd by default**

Conclusion

- **HTTPS and GridFTP using Gfal provide similar throughput**
- **Downloading from multiple sources with Aria2 gives better throughput or similar to the fastest source with less variance**
- **Using multiple source downloads eliminates the trial and error part of finding the best source**
- **When downloading from multiple sources Aria2 isn't able to split files smaller than 2MB, neither xrootd by default**
- **With similar configuration as xrootd, Aria2 is giving better performance for 2 sources**

Questions?
