Survey Report

Yves Kemp

November 18, 2025

1 Introduction

A survey on NAF usage and users was conducted in October and November 2025.

The platform was the HIFIS limesurvey hosted by DKFZ.

Open: 2.10.2025 - 7.11.2025, with pre-opening for NUC members to comment on questions one week before.

Advertized to the NAF users list as well as the NUC list with the question to advertize on dedicated channels. A reminder was being sent at half time of the survey.

2 Overall statistics

The survey consisted of 25 questions, some multiple choice, some free text.

A total of 115 responses were received, of which 93 complete anwers. Out of the 22 incomplete ones. These are ignored in the evaluation.

These 93 completed responses must be put into perspective:

- More than 300 active accounts have been seen on the NAF login nodes from 13.9.-13.10.2025

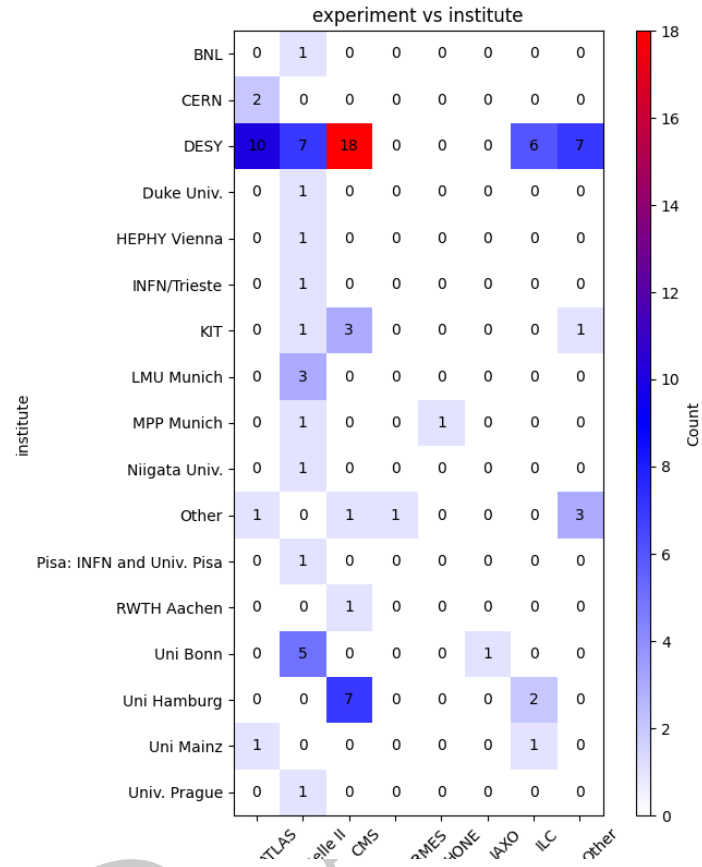- Nearly 1500 users are subscriped to the NAF mailing list

3 Detailed analysis

3.1 About the users, career, experiments, institutes
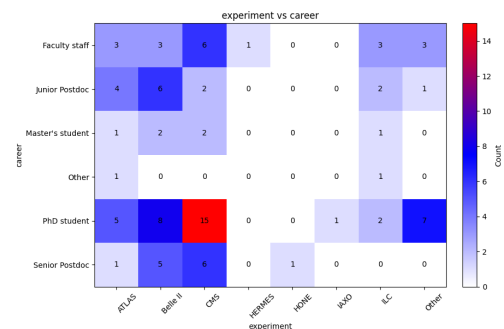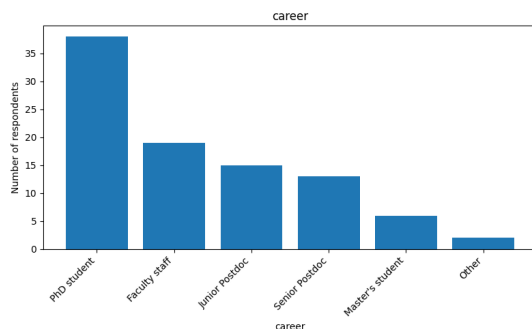
We see mostly PhD students on the NAF, this is more or less independent of the experiment.

A comprehensive view of all institutes, split among the experiments:

experiment vs institute

A summary of the institutes (some institutes with only one participants are summed up), and the list of users per experiment

Top 10 institutes (others aggregated) (Combined)
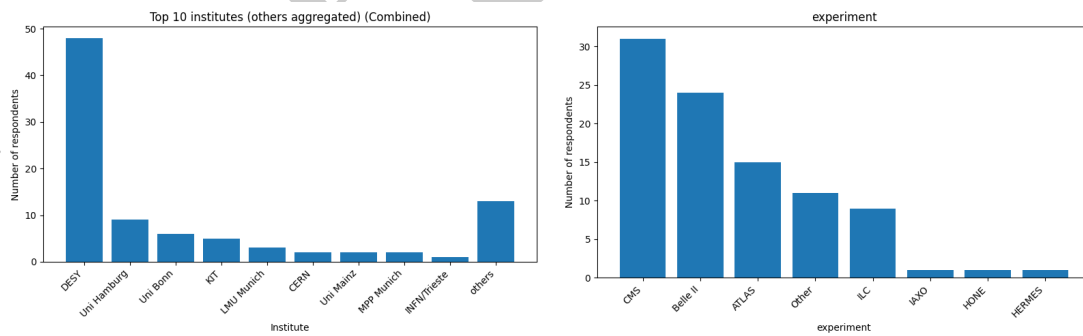
experiment

This institute list is summarized more, and split among experiments:

experiment vs provenience
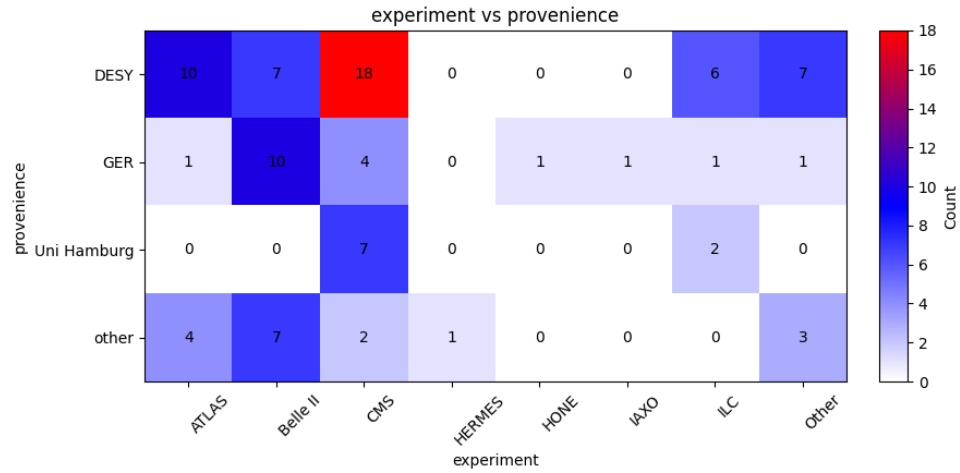
We see from these answers to the survey:

- DESY users are the large majority

- Uni Hamburg users are detailed separately, since they have provided the NAF with resources in the past, and the NAF is their only analysis resource.

- Concerning the Terascale usage of the NAF: There are some answers from German institutes, but mostly from Belle users. The survey did not reach users from other experiments, or these users are not using the NAF.

- The ratio between CMS:BELLE:ATLAS answers is approximately identical to the number of logins seen by these respective groups (as recorded outside of the survey)

3.2 About the knowledge of users

Users knowledge

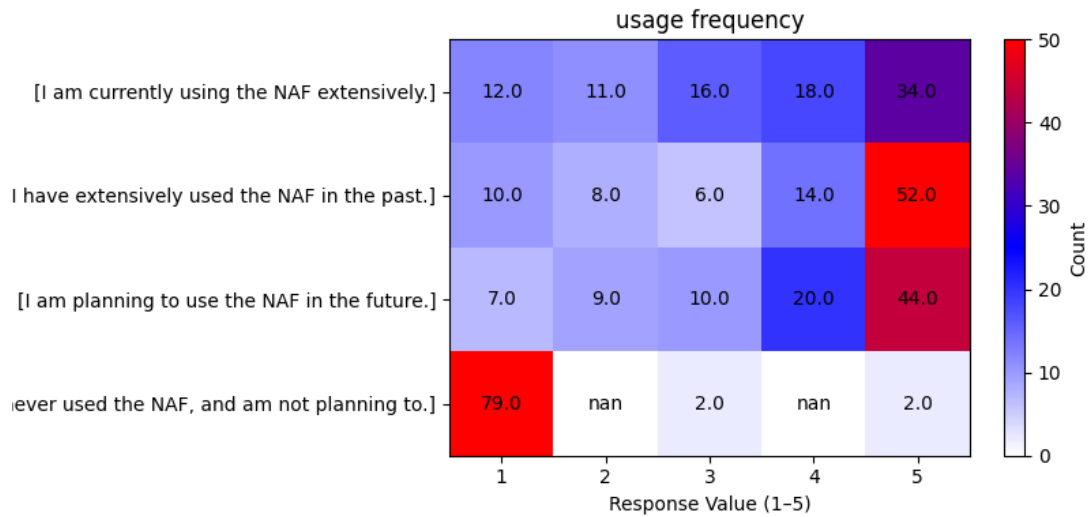|	1	2	3	4	5
[expertise with Linux and CLI?]	nan	2.0	14.0	57.0	20.0
[expertise in C/C++ in general?]	4.0	10.0	21.0	49.0	9.0
[expertise in ROOT (CERNs analysis software toolkit)?]	5.0	18.0	15.0	43.0	12.0
[expertise in Python in general?]	1.0	8.0	12.0	50.0	22.0
[expertise in other data analysis frameworks]	4.0	12.0	20.0	37.0	19.0
[expertise with Jupyter notebooks]	4.0	25.0	16.0	30.0	17.0
[expertise in HTCondor scheduling system?]	9.0	11.0	39.0	28.0	5.0
[expertise in Python based scale-out data analysis tools ?]	37.0	26.0	14.0	9.0	3.0
[expertise working with large scale storage systems ?]	19.0	21.0	25.0	16.0	7.0

Response Value (1–5)

Overall, knowledge of users of standard tools (system, HEP, data analysis) is rather good.
Knowledge of large scale systems and tools is smaller, knowledge level more diverse.

3.3 About the use cases

[expertise in Python based scale-out data analysis tools ?] vs experiment

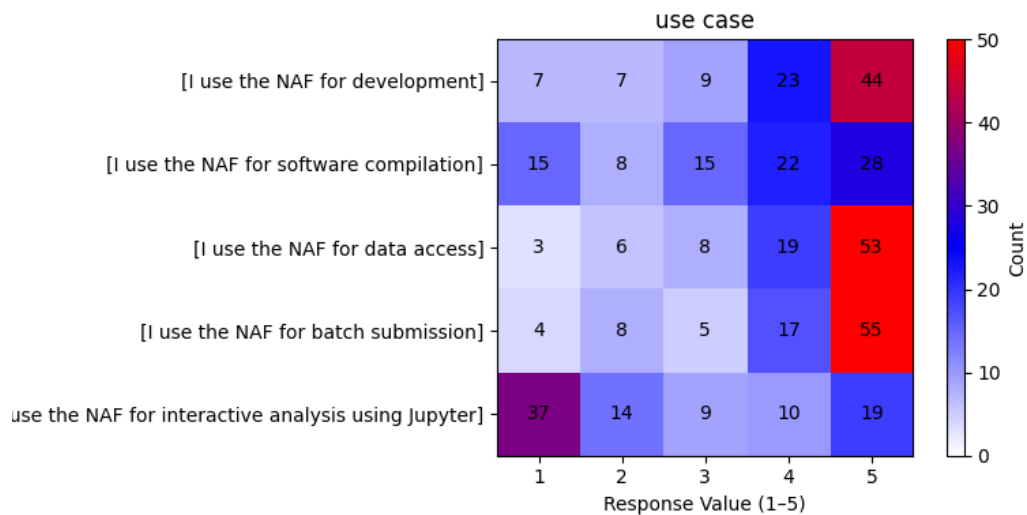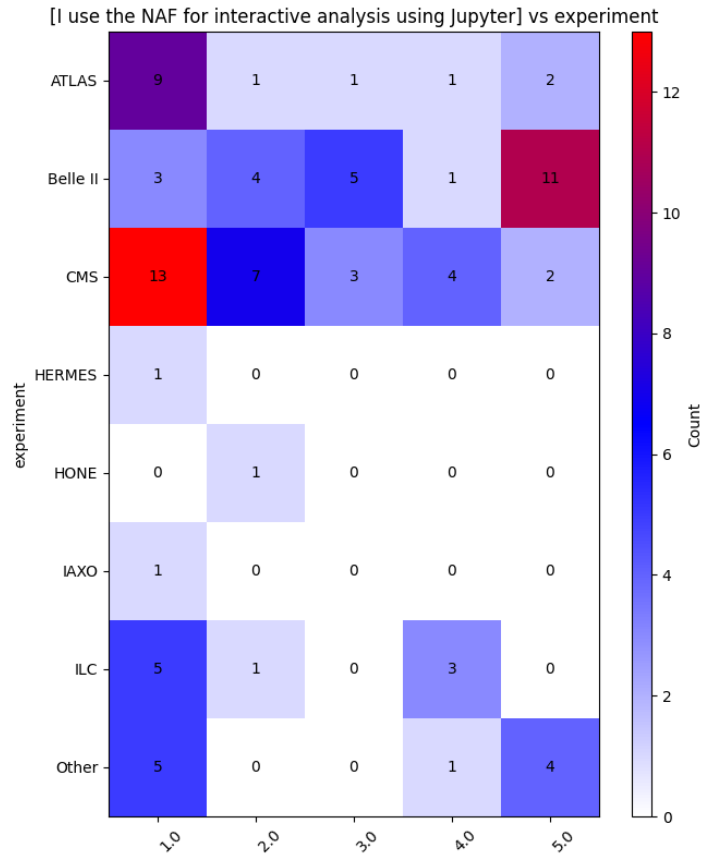| experiment	1.0	2.0	3.0	4.0	5.0
ATLAS	7	2	2	0	0
Belle II	9	10	2	3	0
CMS	7	8	9	4	3
HERMES	1	0	0	0	0
HONE	1	0	0	0	0
IAXO	0	1	0	0	0
ILC	5	3	1	0	0
Other	7	2	0	2	0
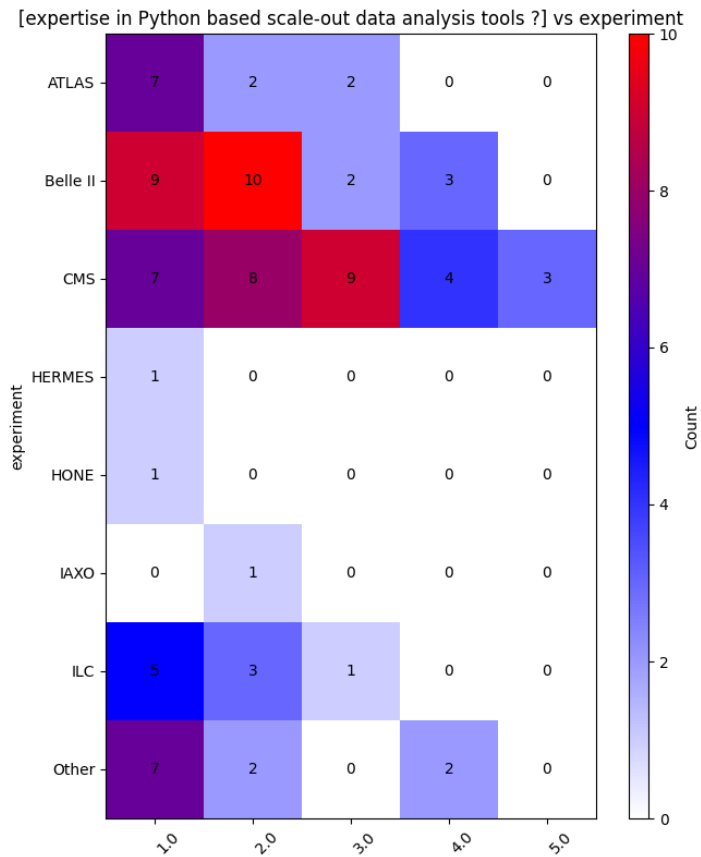
Out of all answers, the expertise in Python scale-out is highest in CMS (though not too high),
and some in Belle.

4

usage frequency
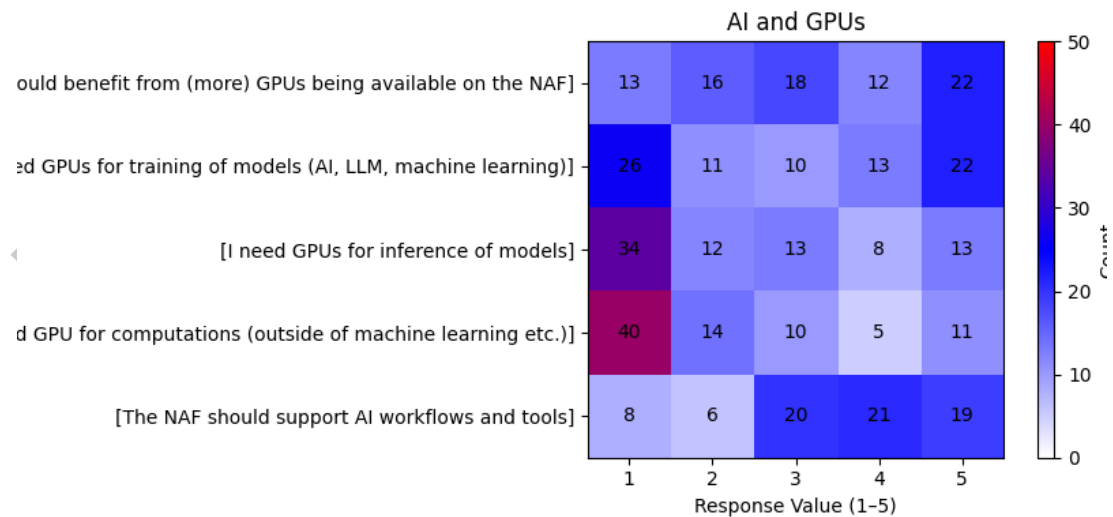
	1	2	3	4	5
[I am currently using the NAF extensively.]	12.0	11.0	16.0	18.0	34.0
I have extensively used the NAF in the past.]	10.0	8.0	6.0	14.0	52.0
[I am planning to use the NAF in the future.]	7.0	9.0	10.0	20.0	44.0
...ever used the NAF, and am not planning to.]	79.0	nan	2.0	nan	2.0

Response Value (1-5)

The answers we received were mostly from frequent users of the NAF.

use case

	1	2	3	4	5
[I use the NAF for development]	7	7	9	23	44
[I use the NAF for software compilation]	15	8	15	22	28
[I use the NAF for data access]	3	6	8	19	53
[I use the NAF for batch submission]	4	8	5	17	55
...use the NAF for interactive analysis using Jupyter]	37	14	9	10	19

Response Value (1-5)

The use cases we designed the NAF for are mostly well reveiwed. The Jupyter usage seems to be divided, one fraction making large use, and a (larger) fraction making no use. Software compilation should be investigated further. A substantial fraction does not use the NAF for compilation. Knowing we had some trouble with compilation speed in the past, users might compile elsewhere - or they do not need to compile since all software is already available to them.
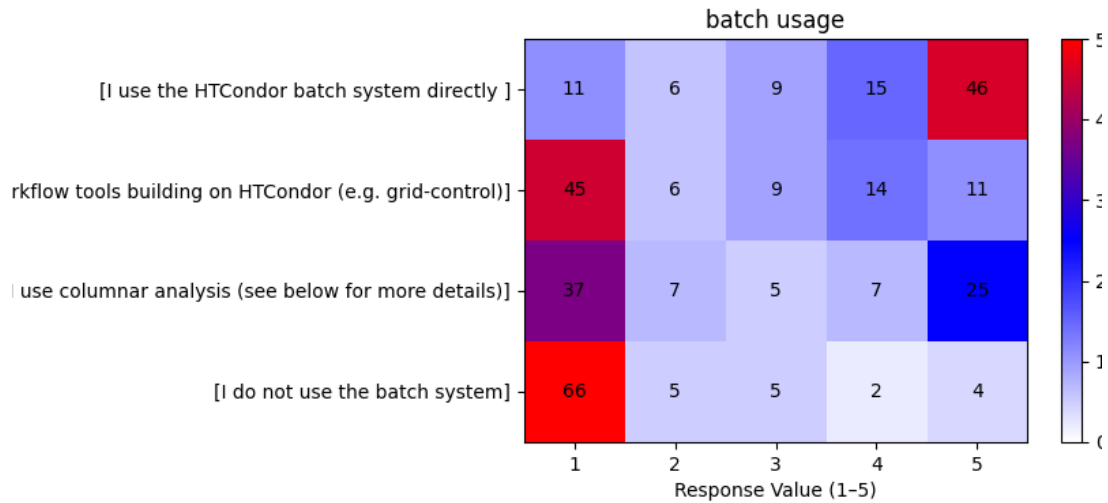
[I use the NAF for interactive analysis using Jupyter] vs experiment

experiment	1.0	2.0	3.0	4.0	5.0
ATLAS	9	1	1	1	2
Belle II	3	4	5	1	11
CMS	13	7	3	4	2
HERMES	1	0	0	0	0
HONE	0	1	0	0	0
IAXO	1	0	0	0	0
ILC	5	1	0	3	0
Other	5	0	0	1	4

A closer look shows that Belle is the major user of Jupyter, whereas Atlas and CMS have not widely adopted its usage.

AI and GPUs

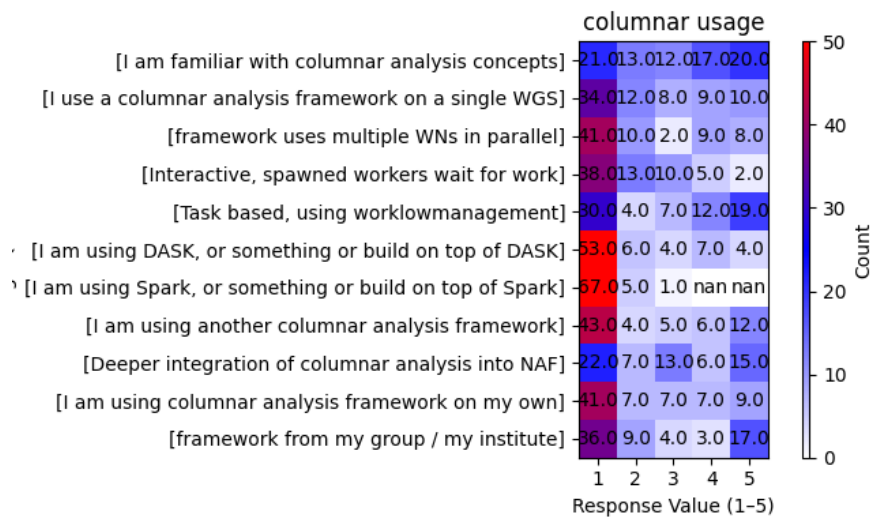| Response Value (1–5)	1	2	3	4	5
ould benefit from (more) GPUs being available on the NAF]	13	16	18	12	22
ed GPUs for training of models (AI, LLM, machine learning)]	26	11	10	13	22
[I need GPUs for inference of models]	34	12	13	8	13
d GPU for computations (outside of machine learning etc.)]	40	14	10	5	11
[The NAF should support AI workflows and tools]	8	6	20	21	19

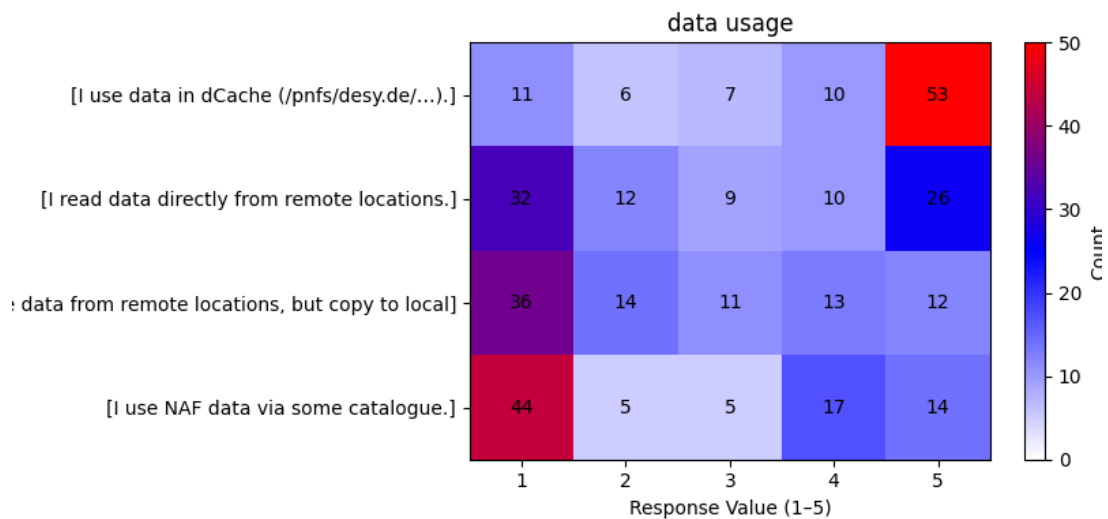Some people will use more GPUs for AI and ML, as well as support for tools and workflows, but the tendancy is relatively small.

The hardware needed is limited, and of the hardware needed, general GPU computing (as opposed to AI/ML loads) is limited.

batch usage

[row]	1	2	3	4	5
[I use the HTCondor batch system directly]	11	6	9	15	46
rkflow tools building on HTCondor (e.g. grid-control)]	45	6	9	14	11
use columnar analysis (see below for more details)]	37	7	5	7	25
[I do not use the batch system]	66	5	5	2	4

Response Value (1–5)

Most people use the batch system, most people use it directly without any further tools.

columnar usage

[row]	1	2	3	4	5
[I am familiar with columnar analysis concepts]	21.0	13.0	12.0	17.0	20.0
[I use a columnar analysis framework on a single WGS]	34.0	12.0	8.0	9.0	10.0
[framework uses multiple WNs in parallel]	41.0	10.0	2.0	9.0	8.0
[Interactive, spawned workers wait for work]	38.0	13.0	10.0	5.0	2.0
[Task based, using worklowmanagement]	30.0	4.0	7.0	12.0	19.0
[I am using DASK, or something or build on top of DASK]	53.0	6.0	4.0	7.0	4.0
[I am using Spark, or something or build on top of Spark]	67.0	5.0	1.0	nan	nan
[I am using another columnar analysis framework]	43.0	4.0	5.0	6.0	12.0
[Deeper integration of columnar analysis into NAF]	22.0	7.0	13.0	6.0	15.0
[I am using columnar analysis framework on my own]	41.0	7.0	7.0	7.0	9.0
[framework from my group / my institute]	36.0	9.0	4.0	3.0	17.0

Response Value (1–5)

Current columnar analysis usage is overall rather small.

data usage

[row]	1	2	3	4	5
[I use data in dCache (/pnfs/desy.de/...).]	11	6	7	10	53
[I read data directly from remote locations.]	32	12	9	10	26
data from remote locations, but copy to local]	36	14	11	13	12
[I use NAF data via some catalogue.]	44	5	5	17	14

Response Value (1–5)

Large scientific data is stored on dCache, which is well used. Usage from remote location seems to be split: Some people use it, some people do not use it at all. Catalogues are not used largely.
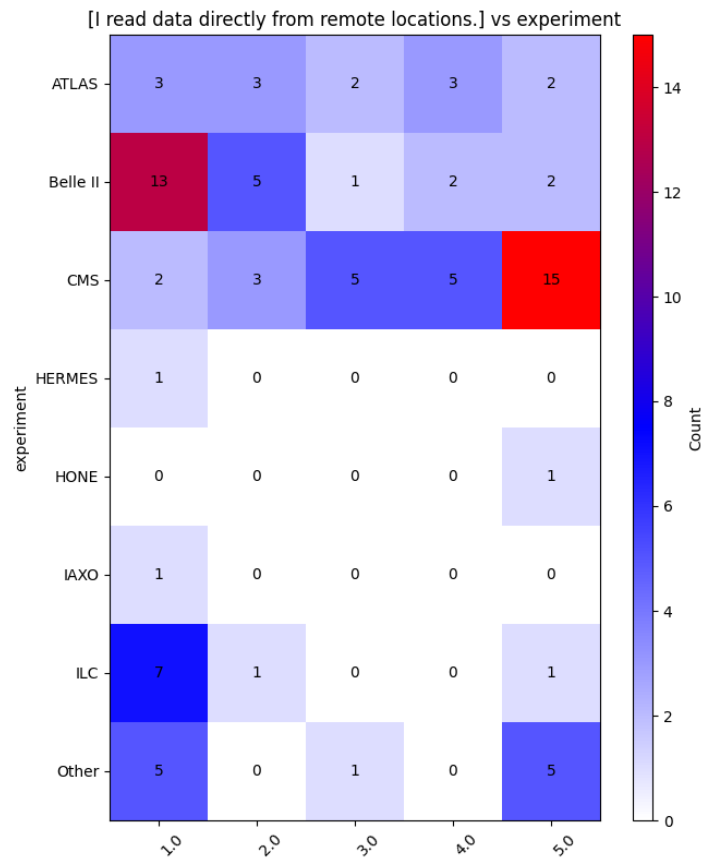
A closer look shows that a large majority of CMS users read data directly from remote.
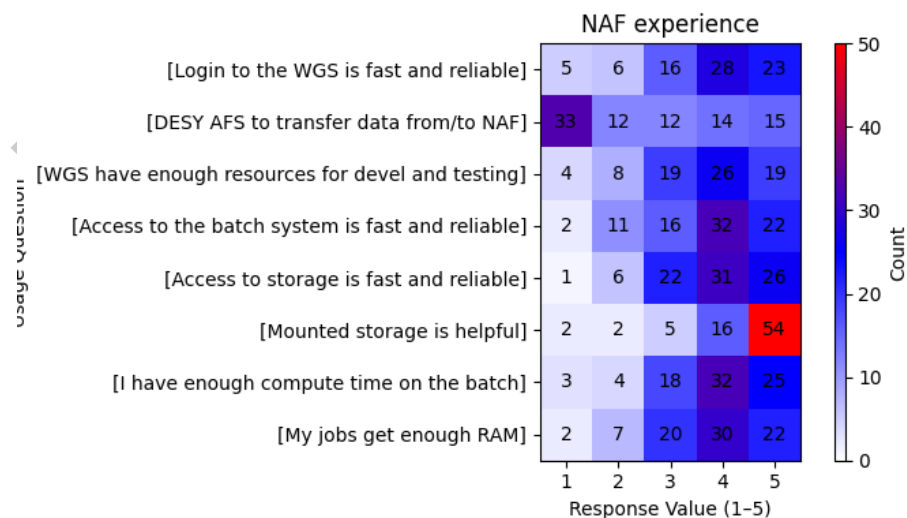
Reliability and overall experience is about "good". This block also contained two storage questions:

- Outside access to/from AFS: about a larger half of the people do not use AFS, the other

smaller half uses it to some degree. The global filesystem nature of AFS seems to be not too important.

- Having data presented as a mounted filesystem (locally on the NAF) is clearly important for all users.
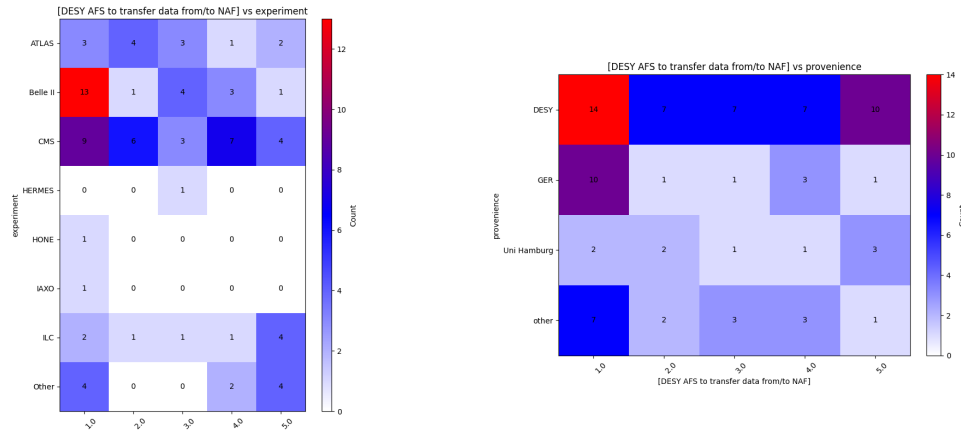
A further look into AFS usage shows that mostly people associated with DESY or to a lesser extent University of Hamburg use its global filesystem functionality. Similarly, Belle II makes the least usage of this global filesystem functionality.
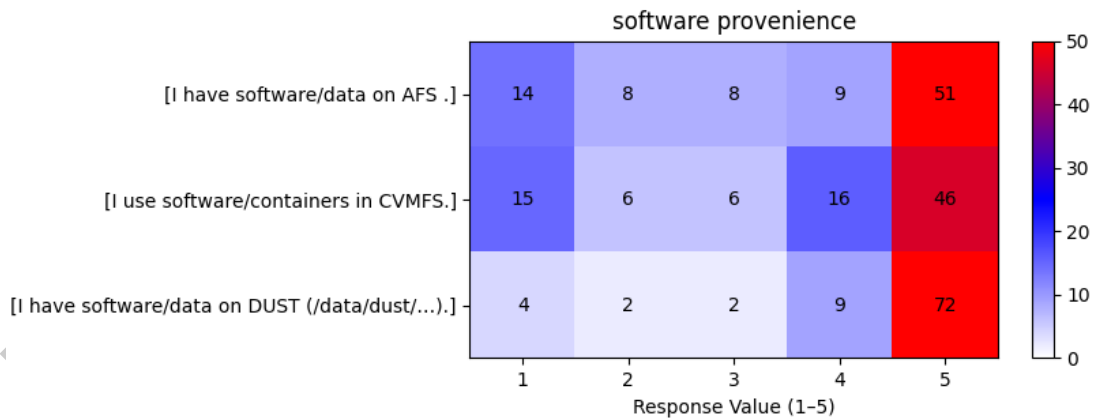
Software is used from AFS, CVMFS, and DUST, with some more software on DUST than the others. Judging from the numbers, most users have several sources for their software.

NAF support and improvements
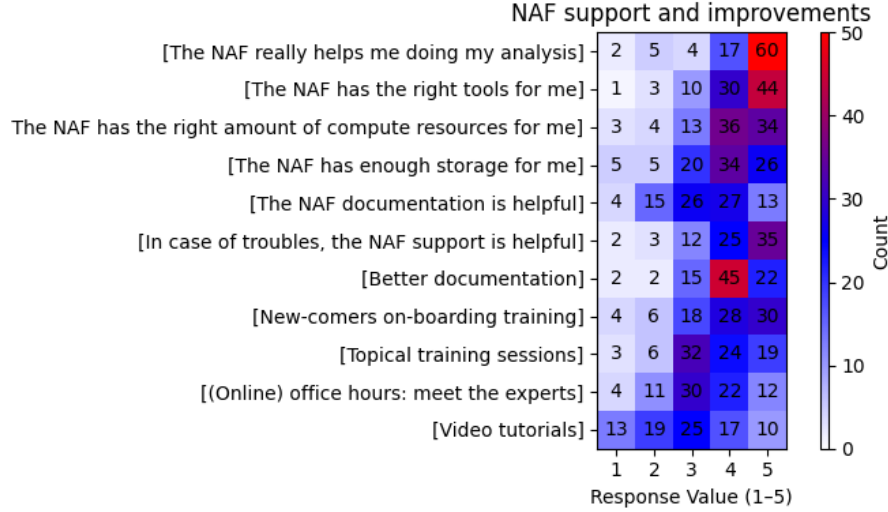
On suggested improvements, some topics come up:

- Overall the NAF is well perceived, and has the right tools

- While compute and storage resources are seen OK, the numbers show us to watch the situation

- Documentation should be improved

- Training events and sessions are neither seen helpful nor not helpful.

3.3.1 Summary of free-text answers

See the full list of full answers in the appendix

AI and LLM: A small number of additions only. The NAF cluster itself does not have many GPUs available, users resort to other resources. Access of GPUs via Maxwell as a NAF project becomes more important. **Further/additional comments:** (selected items)

- One user mentionned Gitlab integration at another institute, which he lacks on the NAF

- Mattermost or slack channel of all users is proposed as a tool to reach out

- Work on (written) documentation - eventually using LLMs

- NAF experience has increased w.r.t last year

Annoying on the NAF: (selected items)

- FastX to the WGS themselves

- storage issues: data loss on dCache, /nfs/ slowniness, CMS dCache performance

- VS Code integration

- mentionning several operational and stability issues

- user quotas within Belle II

- MFA and password regulations

- tmux sessions

- one user took the /nfs to /data change badly

Columnar analysis comments: Users want to use different frameworks. Users miss documentation on our concepts and offerings. **Good on the NAF:** (selected items)

- Stability improved, communication improved, hard work and facility support is appreciated. Also experiment support (Belle II) is appreciated.

- Resources are vast. General setup with WGS and Grid complementarity is appreciated.

What would improve NAF experience (selected items)

- Trainings - maybe in form of video tutorials.

- Involvement of experiment contact persons is crucial for the NAF.

- Tape backup in case of storage failures

- Working on documentation - including a check-list for users when things do not work.

- VS-code

- Marimo Notebooks -¿ Integration with Jupyter Hamburg

- Larger DUST space (-¿ Belle II internal communication issue). Requesting DUST via some interface

- Development on LXPlus since integrated with CERNbox and syncing with private laptop -¿ NAF?

- Stability: batch, tmux

Missing in NAF setup (selected items)

- tools helping with scheduling scripts

- Backup of important data

- DASK and gitlab syncing

- Interactive GPU availability, more GPUs in general (with large RAM)

- DESY Sync&Share access

- Larger Jupyter runtimes, higher stability

- submit jobs from jobs and from CERN containers

Additional Batch Usage CMS crab, ColumnFlow, parsl, b2luigi

Additional software or batch usage using xrootd to CERN, using Webdav for dCache

Editors: There is a wide variety of tools. VI (or variants) seem to still have the majority in 31 answers, while VS Code is mentionned in 25 answers. Emacs is used by 12 users, and 5 use Jupyter. 8 other editors are mentionned.

4 Conclusions

We conducted a survey of NAF users and their usage.

Generally, the NAF is well perceived by its users.

It turns out, however, that (except for Belle), the majority of answers stems from DESY users. About only one third of active users during the survey period have answered the survey - which themselves are rougly one fifth of registeres NAF users.

A more detailed list of future work on and changes to the NAF can be compiled out of the discussion with expert users on these results.

The following shows the complete free-text answers. They have been summarized in the document.

A Comments on: Additional AI, LLM, machine learning, GPU usage:

- ILC : Given my use case in-process parallelization would be the obvious next step to increase performance of the software package, once it is back up to scratch (C++98/11 -¿ current). Then it could well profit from AI, too, seeing it is closer to hardware, which is less explored, so far.

- CMS : GPU availability and memory sizes will be the main bottleneck of my project, so I would benefit greatly if the NAF expanded in that direction.

- Other : columnar frameworks on GPU (e.g. polars)

- CMS : Maxwell

- CMS : I have access to another cluster which targets ML usage with multiple GPUs

- CMS : I use DESY maxwell for that

B Comments on: Any further comments you want to share with us?

- ILC : I already asked colleagues to participate in the survey. Often people complain, but never escalate their problems to the admins or NAF-support, but just revert to suboptimal

workarounds or plain silent suffering. thus, I think this survey is a good opportunity to get a picture of the current situation.

- CMS : Keep doing what you do, the NAF is overall a great platform for scaling the deployment of my tools. For development I prefer staying with my institute's resources, which are more flexible (gitlab integrations) and overall more convenient. Cheers!

- Belle II : thank you for your hard work! i love naf ¡3

- Belle II : Regarding the Columnar Analysis Section, I think my replies are a bit unreliable because I am unfamiliar with the topic. However I am always using the vectorized numpy/panas/polars operation, which might count as Columnar Analysis? Also I use luigi or b2luigi for all sorts of tasks, not just columnar. I now only submit to HTCondor through b2luigi.

- ILC : "How can we better reach you, and enhance your NAF productivity?" Maybe a mattermost/slack channel for live Q&A and online support from experts? It is more interactive than emails/tickets for small Q&As. I have met a few examples where people use it, to name a few:MOGON HPC, CometML

- CMS : Please no video tutorials. Documentation in written form is very much preferred!

- CMS : On the last page, good point = ease of access
  Overall: thank you!

- CMS : Good job, keep it up!

- ILC : Mostly use Maxwell, but the naf is an excellent backup for when maxwell doesn't work/has maintenance.

- CMS : The documentation/support could be enhanced with some LLM since many questions are hidden in the docs.

- CMS : -

- CMS : Thank you for the ongoing work on NAF! I feel it is, on balance, a good facility

- CMS : I see your effort! It's been a much better experience this year than last year already and I assume some degree of annoyance is unavoidable in operations at this scale

- CMS : Thank you for listening to users via this survey.

C Comments on: Anything else in the NAF setup that you find [annoying]

- ILC : I don't know about the current situation, but it used to be preferrable to have FastX access directly on the WGS and not via the graphical login hosts because of the GID change. Some graphical tools, like event displays requiring the GID of the WGS had to

be run via ssh -X or more complicated tricks. That's why we had that enabled on naf-ilc at some point with good success and probably sufficient resources, but I don't know how many people used it that way.

- CMS : Sometimes we can't connect to some nodes on NAF, without getting any information about it

- Other : repeated user dcache storage failures over the last years, leading to unrecoverable user data loss

- CMS : Login, certificates shadow expiring, having to source keychains etc

- ATLAS : Sometimes submitted HTCondor job fail to locate the build/ directory to setup themselves correctly but this does not trigger an error until full requested time of the job is exausted. The only way to notice this is check either the log files or see the RAM/disk usage

- CMS : Some worker nodes have issues accessing data or software on /data/dust (sometimes in the order of 10% of my jobs) and all jobs started by them fail.

- Belle II : VS Code sometimes difficult to setup. Frequent disconnects, for me were fixed by installing vscode-server on /tmp/username, but for other colleagues this did not work. Also for VS Code, sometimes we can't open our files (some file permission error appear). Sometimes we can open them, but if we open a folder/workspace then we get the file perimission error. Very difficult to reproduce and troubleshoot. I think less issues happen with VS Code installations on linux that on Windows installations.

- ATLAS : newer cuda would be benefical for easier use of machine learning

- ILC : 1. Sometimes old files stored on /pnfs are not accecible and require a dedicated email to get them staged for access. 2. When submitting O(1000) almost-identical 10-minute jobs on HTCondor, there are always those 5-10 jobs that can successully run for 12 hours, despite the rest has finished within 10 minutes or so. It is frustrating, because I often either have to cancel those jobs and glightly sacrafice my full reproducible data sample. Or to trace back the failed jobs and their input everytime, and re-submit them manually. Which becomes annoying enough to not bother.

- Belle II : sometimes my kernel just die

- Belle II : the management of submitted jobs on the Belle II nodes is not good, very regularly a single user jams the entire system with a large number of jobs

- CMS : /nfs filesystem is quite slow

- Belle II : Every now and then, trying to start a server at naf-jhub.desy.de fails randomly (but works right after). Not a dealbreaker but can be annoying.

- CMS : Stability of T2_DE_DESY storage is not great

- CMS : Lack of transparent communication in case of problems

- CMS : People using batch system while ending up few k jobs in hold

- Other : tmux is broken (reconnect to a session causes that I cannot access my files/storage anymore) and it is an extremly fundamental tool especially with such an unreliable internet connection

- CMS : Stability issues (downtime of WGS or condor, loss of files) frequent enough to errode trust in NAF

- ATLAS : Having to input the 2FA every time. Changing my passwort every 180 days with way too many strict requirements (including a maximum length?!)

- CMS : often these interactions get bottlenecked by scheduler/storage issues

- ATLAS : stuck working nodes

- ATLAS : a bit too high frequency of issues with dCache storage servers

- ATLAS : but the unreliability in the past few years has made me stop using it almost completely wherever possible, especially jupyterhub. A workflow that should just take 10 mins to run on a normal cluster took me 3 hours due to constant issues with jupyterhub and ssh

- ATLAS : Having to email admins directly to get user and group space on DUST

- CMS : getting screen/tmux sessions to survive a logout does not work properly

- Belle II : often downtimes/problems

- CMS : the change of working area path from /nfs/dust/ to /data/dust/

D Comments on: Additional columnar analysis usage:

- ILC : Can one learn about such concepts in the NAF documentation? From my experience fresh users are often completely inexperienced with advanced concepts of scientific computing, having only used small workgroup clusters at universities or local PCs. I need to give general introductions to all those possiilities.

- CMS : I would like to use DASK for my code. For now I use DASK python commands to build DAG, but they are still executed in a regular HTCondor Job.

- CMS : I use PocketCoffea (A columnar analysis framework in CMS based on coffea)

- Other : polars

- CMS : Parsl

- CMS : Mainly ML/DL?AI work on Maxwell

- CMS : I'm a columnar developer, and advice/discussions with IT are useful (some already in progress)

- ATLAS : Clarification: I use framework based on RDataFrame, and therefore rely on parallelising the work on multiple WNs manually via HTCondor

E Comments on: naf.good

- Other : usage of the naf workgroup servers fur multiple small but essential daily tasks, including afs

- CMS : Large ressources and fast job executions

- CMS : Since the restructuring of the NAF and storage (mostly /data/dust) which caused some disruptions on our side, the performance and reliability really improved. I really appreciate all the work that's happening in the background there to make this as smooth as possible for users every day :)

- ILC : HTCondor cluster computing works like a charm! Thanks to everybody who contributes to make it possible!

- Belle II : Andreas help! Data access, well maintained

- CMS : recently the communication with users is much improved. great job!

- Other : NAF's up time is great!

- CMS : Many experts can be directly contacted

- CMS : Pool of resources usually quite good, don't feel held back by computing

- ATLAS : Having access to grid data, a batch system, and local storage all accessible from the WNs is a huge benefit.

- CMS : vast ressources and usually easy and quick interactions

- ATLAS : the available computing resources are actually above what I consider average (CERN lxplus), so kudos for that!

- ATLAS : It used to work reliably before 2022 and it used to be my main place I did analysis

- ATLAS : Large scratch disks and good batch system

- HONE : support, and everything!

- Belle II : enviromental impact report

- CMS : CMS support desk is very helpful

F Comments on: What else would improve your NAF experience? (Or: If you never used the NAF: What would make you use it?)

- ILC : Trainings are helpful, but are asynchronous, such that newcomers or problem-experiencers may need to wait some time to progress that way. Those could be provided as recorded video tutorials as well. NAF as a gateway for an experimental community (more for small communities like LUXE, etc.) is a great idea. However, that requires those communities to formally take care of contact persons, assembling software suits, storage and data life-cycle strategies, etc. It may be helpful to compile a list of points to be clarified for the NAF operators for such/any usage scenario of the NAF. There are good "role-model" NAFs that provide stereotypical solutions for various usage communities and have already struggled through these questions.

- CMS : The email system takes a while to get responses from. And if I have been asked to update my analysis for a talk in one or two days, and naf is down or condor is not working, I would need faster help. I would like a mattermost channel or similar for naf users - when there's some experts to help and comment, but also other helpers could help or add some +1 that they have the same problem.

- Other : Keep afs! Provide on-demand (or general) tape backup for important user data to limit impact of storage failures.

- CMS : It's awesome that I always get prompt and professional replies when I send emails to the NAF experts. In general, NAF has provided a reliable and good resource for my physics analysis over the past few years. Although the current documentation is probably not straightforward enough for people with limited computing experience.

- CMS : A central documentation about everything there is without being technical and a short list of things to keep in mind when setting up your project. The latter also applies to things that have to be regularly checked such as ssh tokens and keys, the kerberos certificates and the grid certificat. Some better documentation about these topics would help us tremendously.

- Belle II : Unfortunately there is no "back" in the questionnaire. I think I wrote a bit to much on the previous page with the Good, Annoying and Missing NAF questions. I will write again just to be safe.

  _____ VS Code

  There should be a detailed guide on properly setting up the most popular code editor (VS Code at this time) to work on NAF through ssh. As I explained in the previous page "Annoying" section, I used to have the following issues:

  1) VS Code ssh connection would not work no matter how many times I tried 2) VS Code ssh connection would not work for a few times, then connecting normally 3) After connection, can't open files. This would be solved by deleting vscode-server folder, but

then the extensions need to be reinstalled 4) If VS Code window not closed properly (internet disconnect, laptop force shutdown, etc), consistently run into problem 3). If window closed properly everything normal.

3 and 4 seem to have been solved by installing vscode-server in /tmp/username

5) From a colleague asking my help: They were facing problem 1. After installing vscode-server in /tmp/username, they had the following issue: Can access and open files in the connected vscode client by ctr+O but when opening a folder/workspace, they would get (NoPermissions(FileSystemError): An unknown error occured. Please consult the log for more details) we could not find this log anywhere. I shared all my remote-ssh vscode settings with them but they got the same error again

6) After connecting, opening a folder leads to a workspace does not exist error. Clicking on the folder icon gives another pop-up: "Unable to open ¡directory¿. Unable to read file 'vscode-remote://ssh-remote+naf/afs/desy.de/user/u/uname/PATH/TO/DIRECTORY' (NoPermissions(FileSystemError): Error: EACCES: permission denied, stat /afs/desy.de/user/u/uname/PATH

I am currently struggling with problem 6 which I had not encountered in months, while having another window successfully connected to another directory on NAF.

We both had windows 11 with VS Code latest version installed.

Honestly, I don't know how related to NAF this is, but if we find a consistent setup it would really improve my experience which is why I mention it here.

_____ Marimo notebooks https://marimo.io

I already contacted the NAF support to ask about this but here is a summary:

marimo notebooks are native python notebooks (just .py files) and can also be run as simple python scripts. They offer nice integrations for environment isolation and reproducibility, interactive ata visualization and AI assistant integration (code-completion, agent, and chat).

To activate the AI support nodejs is needed which is easy to install on a per user basis, but it would help if it is available by default.

Now to the main point: Launching notebooks on the login nodes is to be avoided, which is why for jupyter notebooks there is naf-jhub. What should we do with marimo notebooks? 1) Try to integrate them to naf-jhub. The jupyter-marimo-proxy package might help 2) Launch the notebook on a HTCondor interactive job and port forward it. I tried that but could not connect.

- Belle II : [what would have helped me when I was starting]: training for newcomers regarding good practices in coding in general, focusing on (but not necessarily limited to) efficient (in terms of computing resources and storage capacity) handling large datasets and analysis workflow

- CMS : /nfs access speed (listing directories, stats of a large number of files)

- CMS : PHP support for DESY hosted home pages (at desy.de/~username)

- Belle II : Please give us more than 1TB of personal space. With the data volumes we deal with this is not a lot of storage and often requires me to delete some old files in the hope that I will not need them again. I then find myself having to recreate them, wasting naf resources.

- Other : I never used NAF until I had to store massive datasets on DUST. Before, I was happier with setting up my own python/jupyter environment and having my own GEANT4 compilation.

- CMS : The NAF design was single core plus batch. Nowadays, having access to multicore for fast analysis would be more appropriate.

- CMS : I prefer to use lxplus/SWAN lately for development, since it is integrated with CERNbox and thus allows syncing with my private laptop. I would love NAF DUST to be able to synch with the (a) desy cloud!

- CMS : The batch submission is not reliable. During my last usage, the jobs were stuck in hold without any obvious reason. Later, it was discovered that the number of available nodes went down because of an internal error. The cases of slow processing of the jobs are becoming more and more frequent. This made me switch almost completely to lxplus, while using pnfs only for storage of the samples.

- CMS : Need to gather major framework developers and have some discussions on usages...

- Other : if tmux would work, that would be a game changer. Reconnecting to a tmux session after having a broken pipe, is not really possible. Within the session I cannot access my directories anymore and get errors when I wanna access my files. I tried very often to renew my kerberos tickets, that is not the issue.

- CMS : It feels like NAF support's first response is often to blame the issues on a user in the group, but is sometimes incorrect, or doesn't properly follow up with this user to educate them

- ATLAS : Finally unifying the documentation following the confluence security issues would be amazing.

- CMS : stability - I understand it's mostly accidental misuse by my colleagues but we would greatly benefit from more stability maybe even so if it comes in form of more rigorous constraints

- ATLAS : A bit larger quota on the AFS system. I use AFS practically only for storing source code, but with the amount of projects (some of which are large code bases, such as ATLAS athena framework) that I work on, the 16GB can really be not enough. I use DUST as secondary storage for software, but the lack of backup redundancy of AFS is a bit concerning.

- ATLAS : Honestly, just making it work and making it reliable such that I can actually start using it again. I've migrated everything to CERN resources because I cannot afford

not being able to do my work a day per week due to either critical stuff randomly being down, or the jupyterhub just acting up

- ATLAS : Being able to request user and group DUST space via an interface

- CMS : Introductory level documentation and simple example condor submit scripts.

G Comments on: Anything else in the NAF setup that you find [missing]

- ILC : Is there a (block)graphical tool to configure and administrate scheduling scripts, asking and explaining mandatory and optional fields? I think especially for beginners and occasional users it is easier to get started with such a tool than with examples, reference documentation and plain/structured? text files. That would lower the threshold and initial problems with batch system usage and thus hopefully lead to a more efficient usage of the NAF components (WGS, batch) and .

- Other : tape backup of important data storage in active use (could be on request)

- CMS : DASK and better gitlab sync (e.g. storing artefacts, having customs runners, more permissions)

- ATLAS : due to requirements on CUDA compute capabilities, many of my jobs rely on using only Tesla V100 GPUs (not 1080Ti/Tesla P100). However, requesting a V100 in a job often takes too long. Better availability of e.g. an interactive condor job with a V100 would be good.

- Belle II : I have had an issue with a marimo notebook (on NAF), where the LLM request (from NAF to Github/gpt-o4) seem to be blocked by the network. Consider supporting marimo notebooks on jhub, create a marimohub or allow users to start their jupyter/marimo notebooks on condor interactive jobs and then allow port forwarding and ssh connections there so that we don't overload the login nodes. For Belle II, it is not obvious how to properly create a fully working kernel from a b2venv, please see related issue https://gitlab.desy.de/belle2/software/tools/-/issues/19

- ATLAS : conda would be appreciated

- ILC : Strong GPUs (A100) for AI research would be nice, but there is Maxwell anyway.

- Belle II : more GPU RAM

- ILC : direct access to desy sync and share

- Other : jupyter servers are often unreliable, break or are slow in processing. It's max. 22h runtime is annoying and inconvenient.

- CMS : Missing possibility to submit jobs from jobs

- CMS : Bigger disk quota for LHC run3 analyses is needed!!! Possibly need to increase ram limit for short queues (and install more rams)

- ATLAS : Setting up support for Yubikeys would be nice.

- CMS : some easy way to use htcondor commands within CERN container environments (such as the cmssw-el7 cmsenv envs)

- ATLAS : reliability

- CMS : better stability for Visual Studio Code ssh connection

H   Comments on: Which is the main editor or IDE you use when working with the NAF:

- ILC : Started with Eclipse, then because of upgrade problems and long interruptions just kate. I could imagine others (VSCode). I'm happy with running it over FastX instead of using built-in remote components. But I know from other users in our group there is a wide range of preferences.

- CMS : Visual Studio Code

- Other : emacs

- CMS : VIM because I cannot be bothered to enter my 2FA token every 30 minutes when using VSCode. I push my code to gitlab then pull at the NAF and execute it there.

- CMS : VSCode

- Belle II : vim

- Belle II : emacs

- Belle II : VS Code

- Other : nedit

- ATLAS : geany or VScode

- Belle II : visual studio code

- Belle II : vim

- CMS : VS Code

- CMS : VSCode (but not using remote connections; I mount /afs via Auristor)

- CMS : VS Code

- Belle II : VS Code

- ATLAS : Visual Studio Code

- ILC : VS Code

- Belle II : PyCharm

- CMS : VSCode

- Belle II : vim

- Belle II : jupyter notebook environment, command line, file manager

- CMS : VSCode

- Belle II : JupyterHub

- Belle II : emacs

- Other : Neovim

- ILC : emacs

- Belle II : vim

- Belle II : NAF Jupyter Hub

- CMS : emacs

- CMS : emacs

- Belle II : NeoVim and Helix

- ILC : vim

- Belle II : Visual Studio Code

- Belle II : vim

- Other : jupyter, vim

- CMS : Visual Studio Code

- CMS : neovim

- ILC : emacs

- CMS : nano

- CMS : VS Code

- ATLAS : emacs

- ILC : emacs

- CMS : vim

- CMS : atom, vim

22

- CMS : VSCode

- Other : vscode & vim

- CMS : nedit - nivarna text editor (GUI plain text)

- Belle II : vim

- ATLAS : I would like to use VSCode, but due to issues with setting it up mainly use straight-up vim at the moment.

- CMS : vim

- CMS : VisualStudio

- ATLAS : vi

- ATLAS : vim/neovim

- ILC : emacs

- ATLAS : VS Code

- ILC : Jupyter notebooks

- ATLAS : vim

- ATLAS : VSCode

- CMS : emacs

- ATLAS : vim

- ILC : VS Code

- Belle II : working locally on IDE and have automated syncronization with the naf via rsync

- HERMES : emacs -nw

- CMS : Visual Studio Code

I Comments on: Additional batch usage:

- ILC : I tried using interactive and non-interactive batch sessions for compiling larger software packages, but that gave a rather bumpy workflow, ading additional distracting oncerns.

- CMS : I also use crab submit commands

- CMS : I use ColumnFlow (god spare me the suffering).

- CMS : I used parsl executor

23

- Belle II : b2luigi

- Belle II : I submit jobs via b2luigi

- Belle II : I use mainly b2luigi as a workflow tool.

- ILC : luigi job pipelining with luigi analysis workflows (law)

J   Comments on: Additional software or data usage:

- ILC : My main use case for ILC is to maintain a currently outdated software package for reconstruction and analysis (as an official but very side topic). Thus only ever small test-runs are done and advanced storage concepts are not a concern at that stage. I'd tend to go to such tools, once it needs to scale.

- Other : multiple; the naf workgroup servers are my main "working horse" for almost everything I do except for web and powerpoint applications.

- CMS : I also use xrootd to access datasets stored remotely at CERN or other T2 sites

- Belle II : webdav to SE

- CMS : I use gfal2, its python bindings, and convenience software on top (e.g. law remote targets).