CERNBox: User Stories, Proposed Features and Opportunities for Improvements

D. Piparo (CERN) - December 8, 2021

This Talk

- A collection of: user stories / proposals / opportunities for improvement for the CERNBox service
- Input collected from colleagues in CMS
  - Different roles and level of experience: physicist, developer, groups conveners, student/postdoc/professor

Executive summary: **CERNBox...**

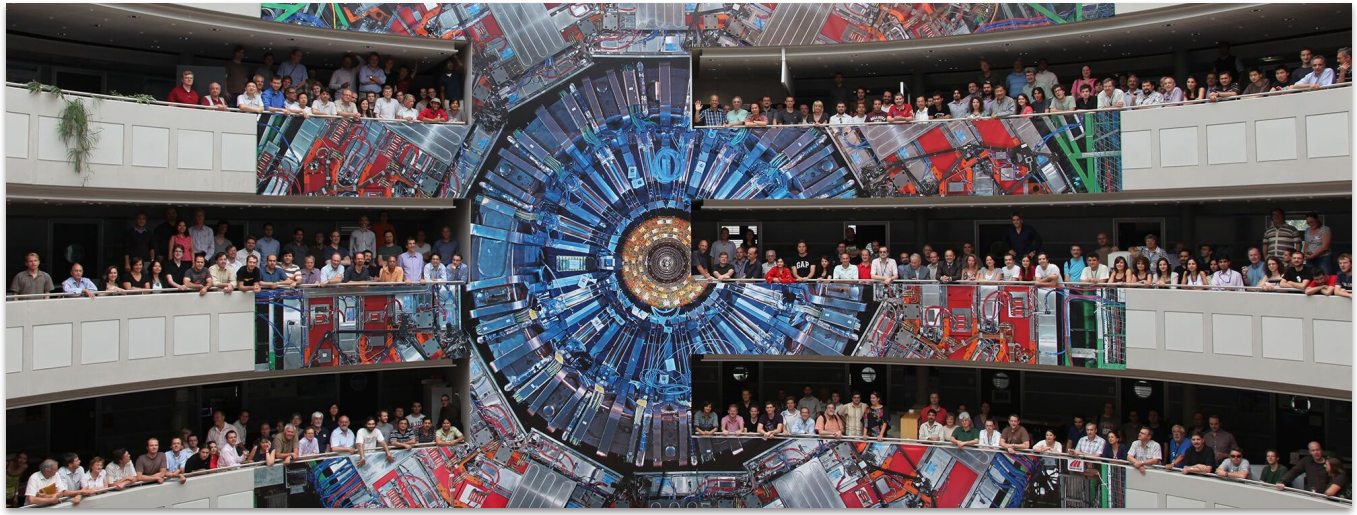
- **... is an integral part of many aspects of everyday work in CMS**
- **... managed to increase the productivity of its users**
- **... provides a unique mix of features not easily available on the market**
- **... can further evolve improving existing features and user experience as well as providing new capabilities**

Introduction

Image: © CERN

Reminder: What Use Cases in an Experiment?

- LHC Experiment: **thousands of collaborators**
  - **Distributed around the globe, not only at CERN**
- Activities: analyse collision data, operate computing distributed infrastructure, develop code, coordinate activities, build/operate detectors, administration, online, offline, financial matters …
- **Basically all possible level of experience in HEP research**
- This result in a **very, very broad set of use cases!**
  - **Hard to capture each one of them**, today we review some



From CMS Website

Appreciated Features

- **Under the control and access of the CERN IT Department**
- **Privacy**: the data is in the CERN data centre, nobody will sell/analyse sensitive data
- Direct access to the **underlying storage and integration with EOS**
- **Custodial disk storage**, and a significant amount of it (1 TB!)
- **Silent and efficient sync**, much improved over the years
  - Clients for all platforms (Android, OSX, Linux…)
- Easiness of **sharing content, protected with passwords or limited access** (groups) as well as expiry date
- User experience of the **web interface greatly improved** in the last 24 months: fast, reactive, clean
- **Integration with other services** or functionality e.g. SWAN, Office Suites

Opportunities for Improvement

- Web interface
  - **Needs to be more stable**, e.g. not automatically reload after visualising content
  - **Pages need to load quickly**, instantly if possible
  - All resources dedicated to make it faster and more responsive are well invested
- Even more privacy: **allow to encrypt directories, easily**
  - Give users the responsibility not to lose their encryption key?

User Stories

Image: © CERN

A Physicist Far Away from CERN

- **User:** data analyser
- **User Story:** I do not work at CERN, but in an institute which is geographically very far from it. I regularly use CERN compute resources for my work (e.g. batch system). I produce a lot of plots, which I need to check to understand if my analysis is going in the right direction. It is not feasible to open a graphic session to lxplus: the latency is far too high. Thanks to CERNBox, I can have a look to my ROOT files through the jsroot plugin and figures on the web interface, optionally deciding to sync the results back to my computer.
- **How CERNBox makes the difference:**
  - Transparent and efficient synchronisation of files
  - Direct access to the underlying storage
  - Web interface

SWAN

- **User:** infrastructure log data analyst
- **User Story:** I need to produce some trends analysis the logs of the infrastructure I am responsible for. I need quite some computing power and use SWAN interfaced to Spark clusters at CERN. Since my home directory in SWAN is on CERNBox, I can share this project via CERNBox with some colleagues so that they can follow the progress made.

- **User:** participant to a course
- **User Story:** I am a student, I followed a training where we worked on some notebooks via the SWAN service. Now the training is over but I have all the files created during the training in my CERNBox and I can continue learning, resuming from where I left.

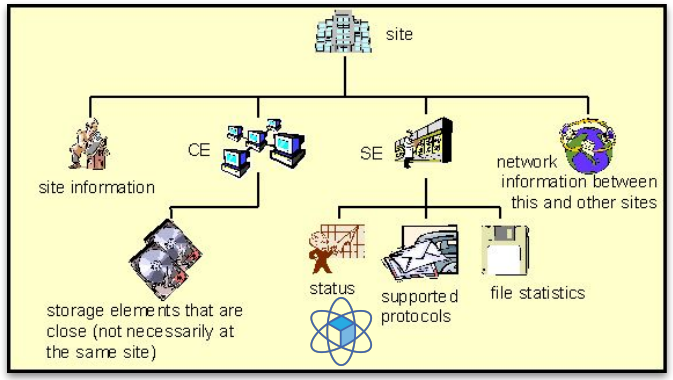
- **How CERNBox makes the difference:**
  - Seamless integration with SWAN
  - Direct access to the underlying storage

CRAB: CMS Remote Analysis Builder

CERNBox: heavily used and trusted by CMS physicists

- **CRAB: Tool for CMS users to submit Grid-jobs**
  - 60-70k cores used by CRAB on the Grid at any point in time
  - Rich mix of jobs: gen/sim/reco/analysis/tests
- Original vision: when a user Grid-job finishes, outputs moved to a site specified by the user, e.g. Tier-2 or Tier-3 ("Storage Elements" or SEs)
- **CRAB allows to copy job outputs to the user's CERNBox**
- **"Fictitious" Tier-3: T3_CH_CERNBOX: extremely handy and popular!**
  - Very well integrated with the existing tools
  - Fits very well CMS NanoAOD data format (1-2 kB/evt)
- Trend observed: CERN disk more and more useful to users
  - Complete set of services and resources to support analysis
  - Common ground for analysis teams
  - **Disk at CERN: a very precious resource…**
- Heads up: CMS sites are migrating their disk storage endpoints to WebDAV
  - To our knowledge eosuserftp.cern.ch is not supporting WebDAV
  - Upgrade necessary to continue to transfer to CERNBox from sites dropping gridftp

gsiftp://cmseos-gridftp.fnal.gov	22381053
gsiftp://eoscmsftp.cern.ch	13241055
srm://dcache-se-cms.desy.de	12322216
gsiftp://eosuserftp.cern.ch	12097125
gsiftp://cms-gridftp.rcac.purdue.edu	9060208



LCG-1 User Guide, 2003

Code Development

- **User:** data analyser, code developer
- **User Story:** During my daily work I develop code. I have a workstation at CERN, a professional laptop and a desktop at home, that I also use for work. CERNBox allows me to conveniently synchronise the files of some of my projects on all these computers. Moreover, it's easy to run the code on CERN resources from home.
- **How CERNBox makes the difference:**
  - Transparent and efficient synchronisation of files
  - Storage integrated with CERN compute resources

Exchange of Sensitive information

- **User:** supervisor recruiting
- **User Story:** I carried out some chats with candidates for a position in my team, together with some other experts in my team. We took notes during the chats, and filled in tables to keep track of how the different criteria were met by the candidates. To complement this information, we added the CVs and letters of the various candidates to the folder. We shared this information among ourselves through CERNBox, sure that nobody but us will access this information.

- **User:** person responsible of a budget
- **User Story:** During my mandate, I am responsible of a budget and use also an spreadsheet to summarise current and past expenditures and to plan the next few years. My successor will have to manage this budget and I shared with her the document through CERNBox. She, and only she, can see what is in there, xcheck numbers, propose potential changes, ask questions, consult it comfortably from her web browser or from a sync'd version on her laptop.

- **How CERNBox makes the difference:**
  - Allow to securely and easily share sensitive information among colleagues
  - Allow to access and edit documents from the browser or the sync'd copy (or copies on multiple devices)
  - Web interface

Suggested New Features and Long Term View

Image: © CERN

Suggested New Features

- **Stable and reliable concurrent editing**
  - Start from text documents, then presentations
- **Share individual files to users/groups**
  - Read-Only or allow to modify
- More control over the shares
  - **Know who accessed a share and when**
  - **"Give up" shares done by others**, e.g. old folders which are not any more needed (avoid long list of shares)
- The (linux) timestamp of a file synchronised from a source to a target gets changed to be the time when the file landed on the CERNbox server or the target.
  - **The timestamp of the original file could be way more useful**
  - Can this be preserved, perhaps opt-in?

Possible Long Term View

- **Even better integration with the CERN Portfolio of Services**
- Some examples follow, there could be more…
- Example 1: Indico
  - Often meetings are accompanied with a sheet for “Live Notes”
  - **“Add live notes on CERNBox” button for event managers?**
  - Minutes editable, commentable or not according to managers’ preferences
  - Expiration date after which the file is made read only, propagating event’s permissions
  - File automatically appearing in the CERNBox of managers
- Example 2: Indico
  - **“Download all contributions of this event to my CERNBox” button**
- Example 3: Overleaf (not really a CERN service, but available to us)
  - Copy all my projects to CERNBox, so that I can work on them offline