IMAS Environment in Docker, Demonstration and Discussion

Tomasz Zok (PSNC)

- A **container image** is a self-sufficient, isolated piece of pre-configured software
- A **container** is a running instance of the image
- Container images can be transferred to machines with Linux, Windows or macOS and run in them in exactly the same way
- For flexibility, many organizations package their software products in container images

- Typically, an image starts from a base OS e.g. Ubuntu or CentOS
- Then, the image is built according to a script: install this, copy that, modify the following files, etc.
- The most popular standard for such scripts is called Dockerfile
- Images can be built using Ansible, Buildah or via Docker API (i.e. in almost any programming language)

- The image is instantiated and executed by a container engine
- The most popular is **Docker**:
    - Designed with enterprise uses and microservices in mind
    - Well-supported and actively developed
    - Engine process requires admin privileges
- There is also **uDocker**, a userspace Docker alternative
- For HPCs and scientific computing, **Singularity** is also a viable option

- IMAS is a huge and very complex ecosystem of codes, data and configuration
- Over the years, CPT put a lot of effort to streamline the installation process
- This allowed to build **IMAS Docker** – a container image with full IMAS environment
- Some possible scenarios of how it can be useful:
    - Working with ready-to-use, preconfigured IMAS on a laptop
    - Submitting jobs to computing clusters which do not have IMAS installed
    - Running automatic integration tests of physics codes when new IMAS is released

A full mapping pipeline:

- IMAS Docker hosted on the UDA Server:
  - Reads experimental data and writes them in IDS format
- IMAS Docker hosted on CatalogQT machine:
  - Reads IDS from pulsefiles and populated the SQL database
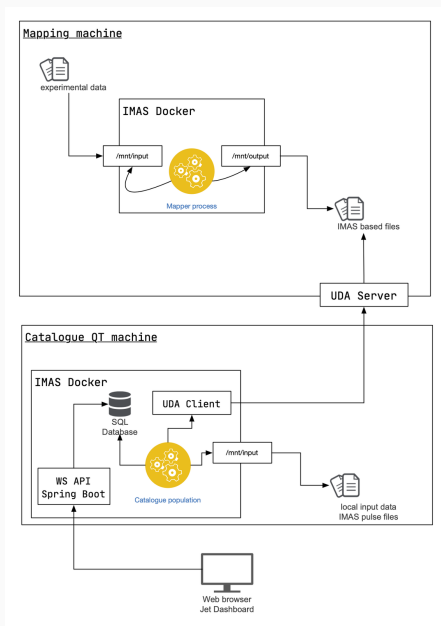
Why Buildah and not plain Dockerfiles?

- Dockerfiles are used by Docker engine, so they share its limitations (process running as root required on the host machine)
- Only Buildah and Singularity allow to build container images root-less, but Singularity adopted its own image format, while Buildah generates standard-compliant image usable by Docker engine, uDocker or Singularity
- Good news: Buildah tools supports Dockerfile script syntax, so we can have both, and we actually do have them! Check out `git.iter.org`, `IMEX/imas-container`, branch: `develop`

Why is the image so big?

- The size of $IMAS_PREFIX on EF Gateway is 2.1 GB
- Efforts so far were focused on building a **development image**, but we wil certainly release a **production image** with stripped down packages to the bare minimum

*Why is this so complex to use? How do I get data to and from the
container?*

- We work on improving the documentation and training materials
- SUMI, developed by Albert Gutierrez, is a tool to simplify
  IMAS-Docker-based batch jobs submission and management
- Maybe in the future, some other specialized software will be created
  (e.g. for data synchronization)