Centrum Kompetencji
w Zakresie Rozproszonych Infrastruktur
Obliczeniowych Typu Gridowrgo – PLGrid Core

Assessment of Docker Containers for the PL-Grid Infrastructure

Bartosz Wilk[1], Daniel Harężlak[1], Marek Kasztelnik[1], and Marian Bubak[1,2]

[1]Academic Computer Center CYFRONET
[2]AGH University of Science and Technology

INNOWACYJNA GOSPODARKA
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO

Insilico experimentation issues

- Heterogenity of computing infrastructures

- Constantly changing cluster configurations

- Job setup script maintenance effort

- Deteriorating reproducibility over time
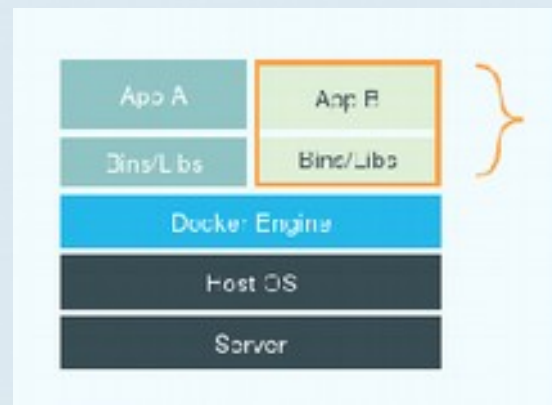
- Development tied to the target infrastructure

Problem Statement

The need for a tool within the PL-Grid Infrastructure to support in-silico experiment development and maintenance.

Introduction to Docker

- Virtualization based on lightweight containers
- Instant container startup
- Container versioning
- Easy container sharing with public and private repositories
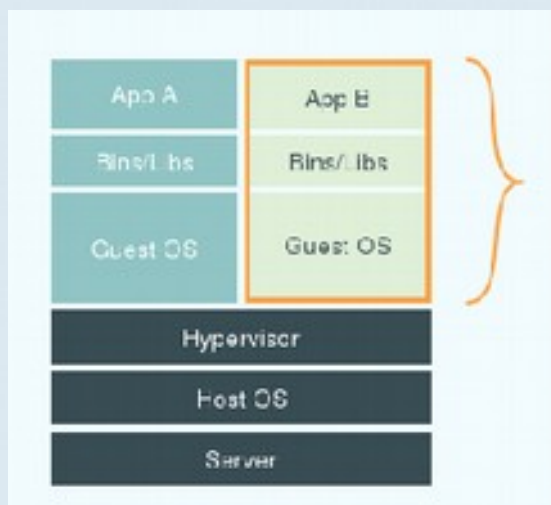
Virtual Machine
  - large images,
  - different cloud stacks.

Docker
 - small images,
  - any Linux will do.

Source: https://www.docker.com/whatisdocker/

INNOWACYJNA GOSPODARKA
NARODOWA STRATEGIA SPÓJNOŚCI

CYFRONET

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO

# What Docker offers for Insillico Experiment Development

- Homogenity by the use of virtualization

- Once container is saved it never needs maintenance

- Instant environment startup

- Easy to change target infrastructure:
  - develop and test locally using commodity hardware
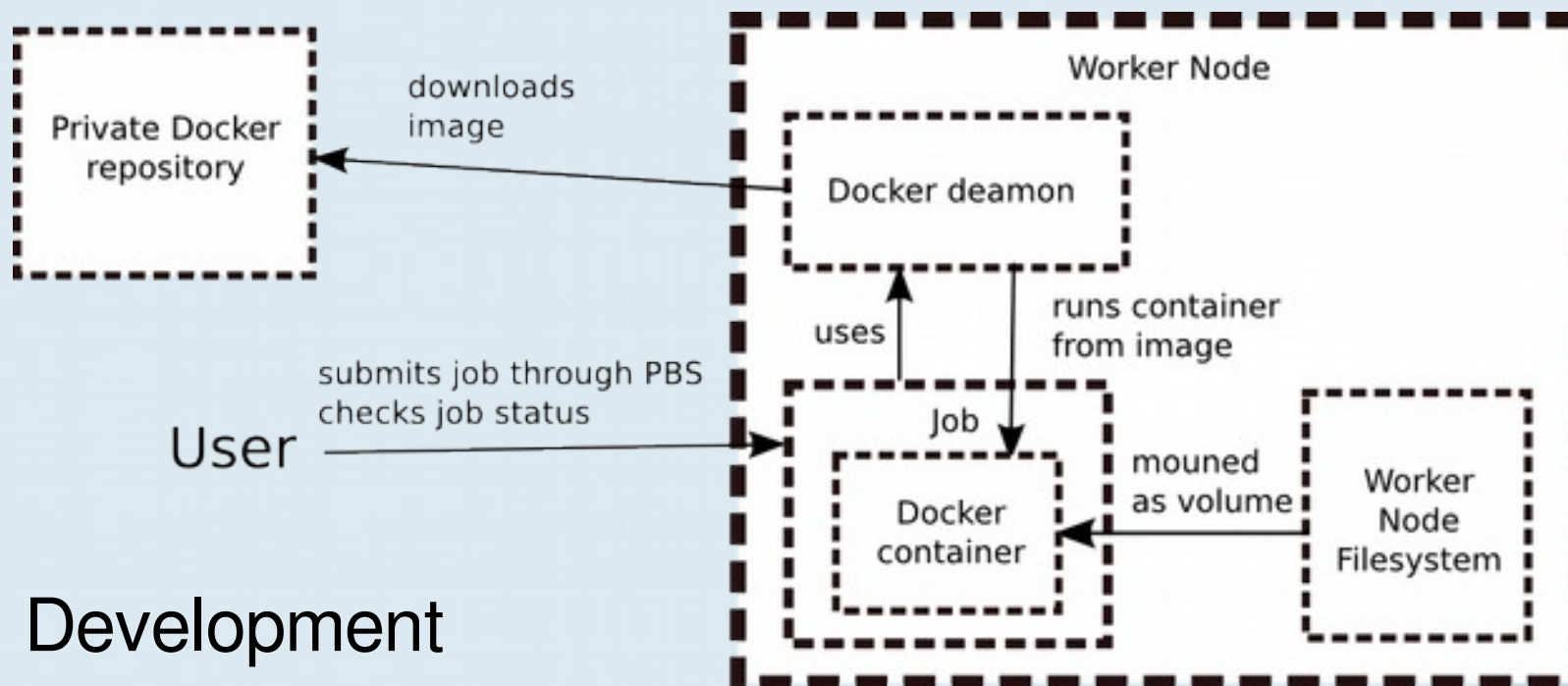 - execute remotely on a cluster
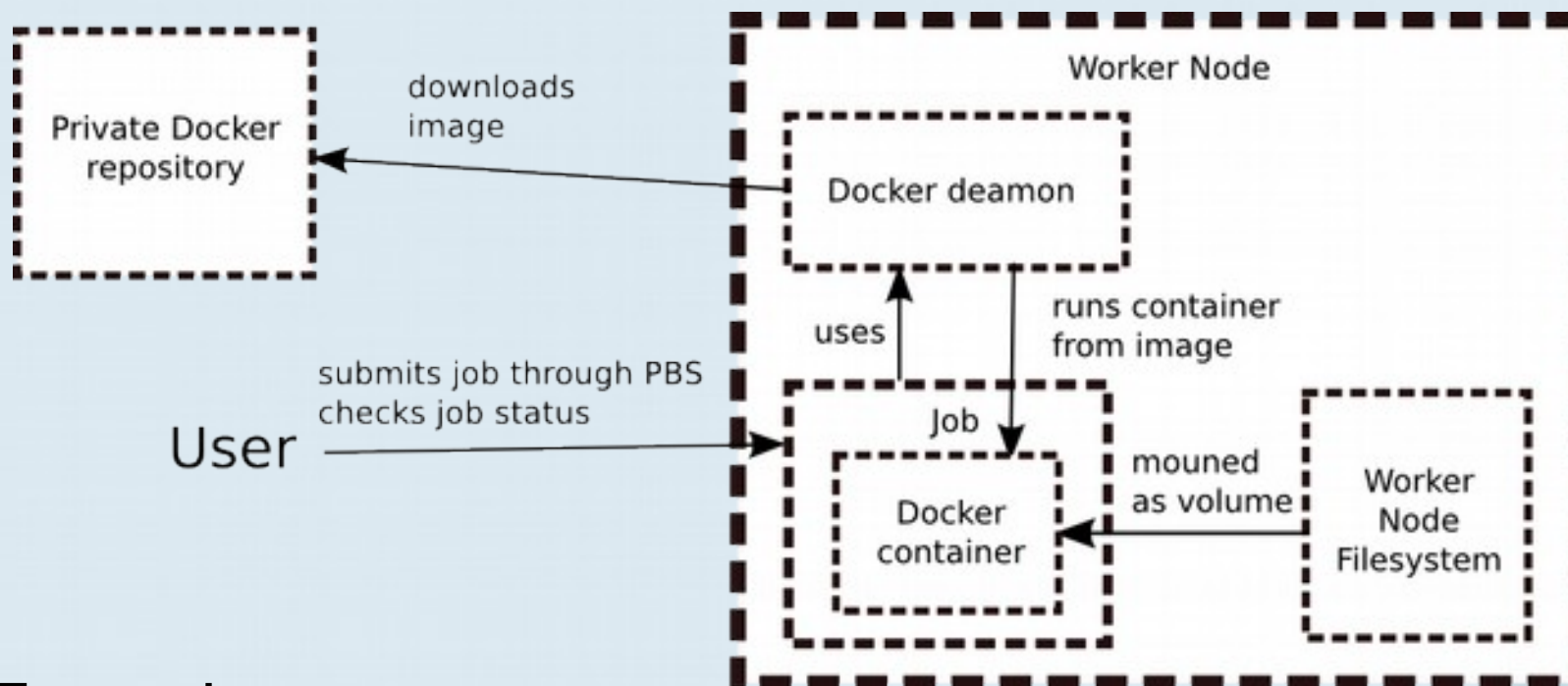
Execution

- Job is submitted to a worker node through PBS
- Docker image is downloaded from the image repository
- Docker deamon starts container from an image
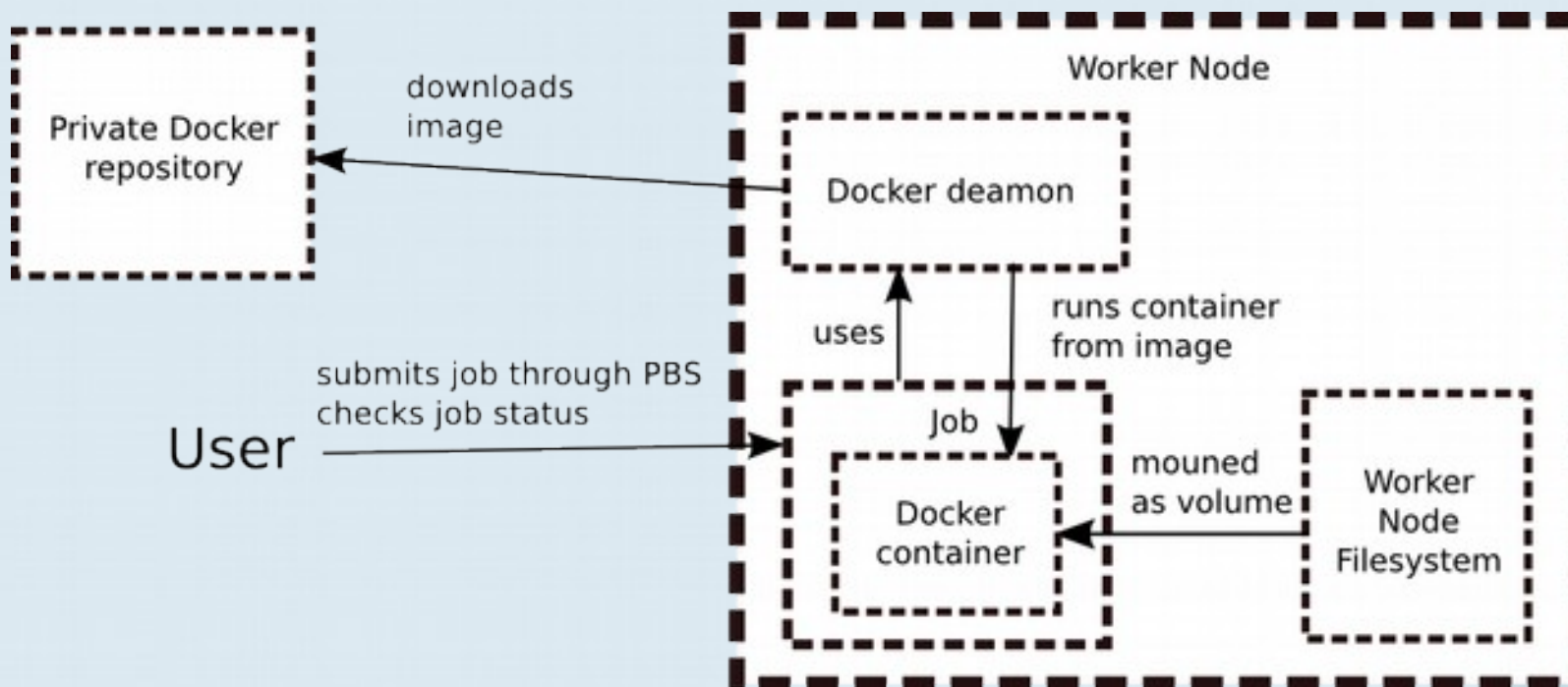
Data

- Stored on the worker node's filesystem
- Shared with Docker container as a mounted file resource

Development

- Development and testing can be performed on a user's machine
- When container is ready it is turned to an image and uploaded to the repository
- Private docker image repository provides security of users images and speeds up image retrieval

Summary and Issues to be Addressed

Conclusions

- Usage of Docker enables to overcome computing infrastructure heterogeneity and configuration variability
- Storage of virtualized execution environment enables to avoid deteriorating experiment reproducibility
- Execution environment can be developed and tested on a local machine which facilitates its preparation

To be addressed

- PL-Grid deployment security
- Access right management

- Contact: **b.wilk@cyfronet.pl**