

Zpracování velkých dat na superpočítačích

Ing. Martin Golasowski, Ph.D.

IT4Innovations



Spolufinancováno
Evropskou unií



MUNI
ICS

Registrační číslo IPs EOOSC-CZ
CZ.02.01.01/00/22_004/0007682

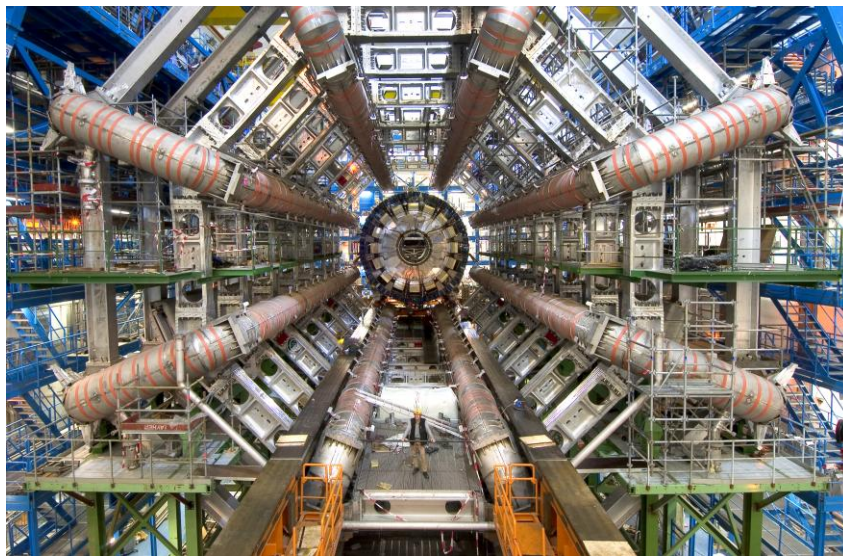
Obsah

- Superpočítače
- Přenos a ukládání dat
- Základy používání superpočítačů
 - Superpočítače LUMI a Karolina
 - Jak získat přístup
 - Jak spustit úlohu přes SSH terminál
- LEXIS Platforma
 - Představení platformy
 - Spouštění úloh na superpočítačích přes webové rozhraní
- Apache Airflow pro zpracování dat
 - Představení nástroje
 - Ukázka instalace a práce s lokální instancí
 - Připojení k superpočítači

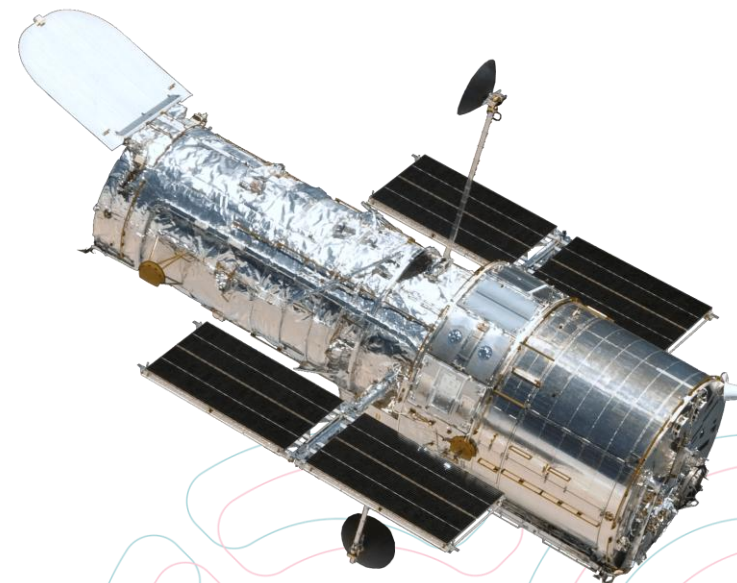
Vědecké nástroje



Mikroskop

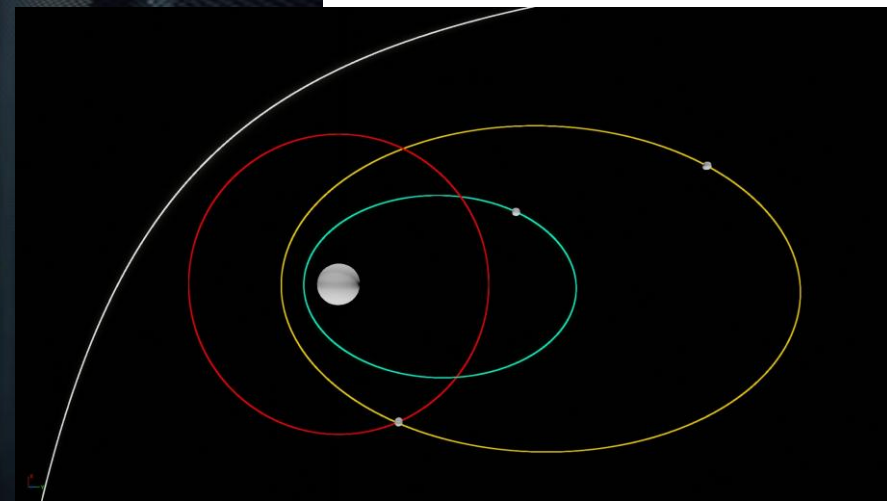


Urychlovač částic



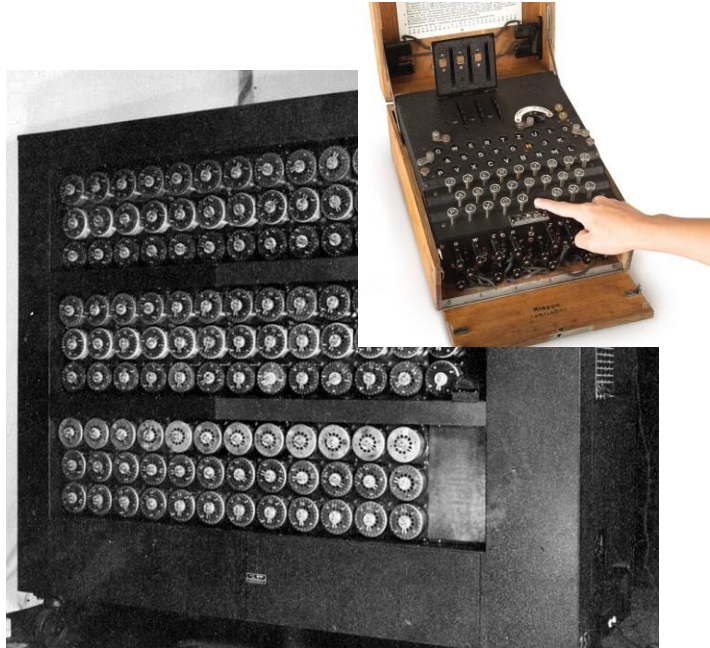
Vesmírný teleskop

Superpočítač

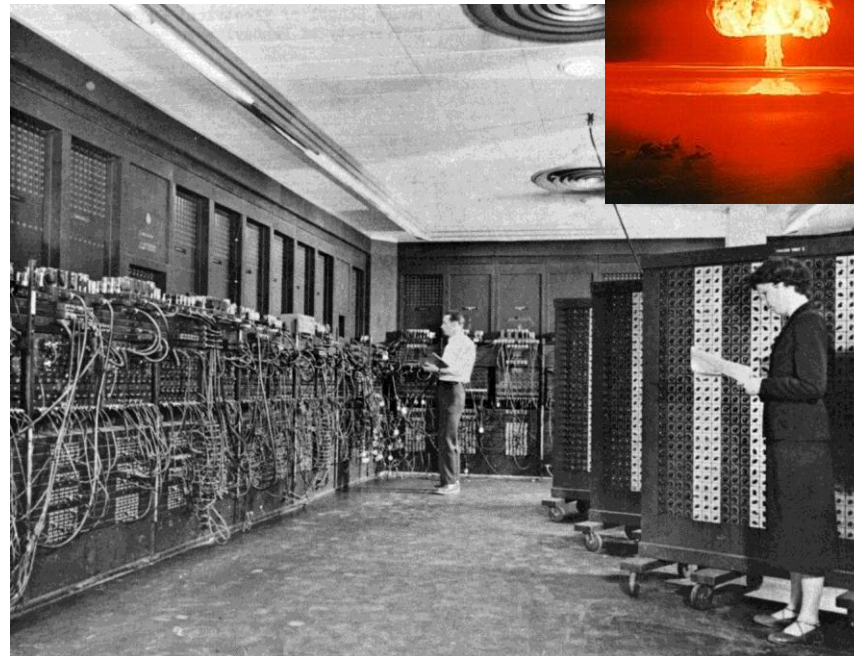


Scéna z filmu Mart'an
Superpočítač PLEIADES v NASA

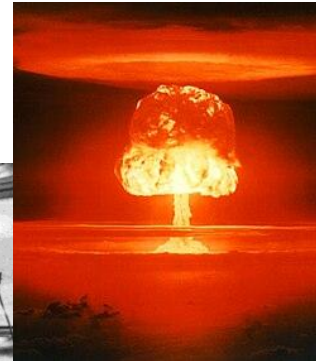
Superpočítače kdysi..



Bombe, UK (1940)
Alan Turing a lámání šifry
ENIGMA



ENIAC, Los Alamos (1946)
John von Neumann
Vodíková bomba
500 FLOPS



Cray-1 a Seymour Cray (1976)
160 MFLOPS

Superpočítače dnes

Výpočetní uzly

CPU

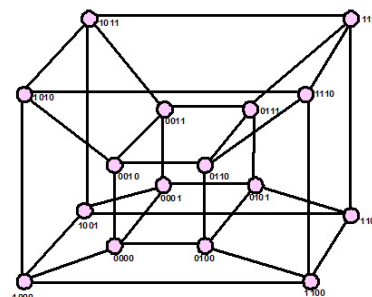
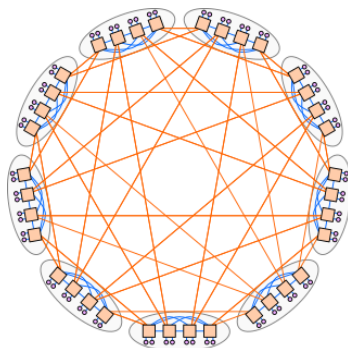
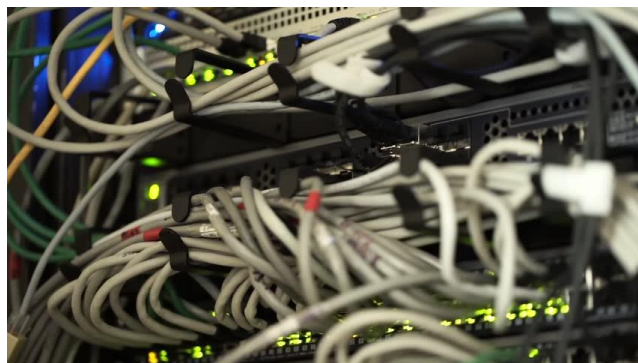
CPU+GPU

FPGA

AI



Síť (Fabric)



Dragonfly a hypercube topologie

Úložiště



Výkon superpočítače

1 FLOP = Floating Point Operation

1 TFLOP = 10^{12} FLOP

1 PFLOP = 10^{15} FLOP

1 EXAFLOP = 10^{18} FLOP

Nejvýkonnější na světě - TOP500.org

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	El Capitan - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581

LUMI - 8. místo na světě, 1. v EU (listopad 2024)

8	LUMI - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107
---	--	-----------	--------	--------	-------

Superpočítače IT4I



Spolufinancováno
Evropskou unií

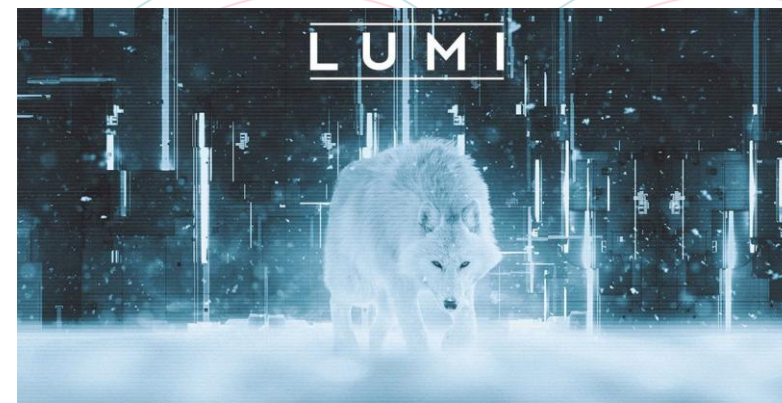


MUNI
ICS

Registrační číslo IPs EOSC-CZ
CZ.02.01.01/00/22_004/0007682

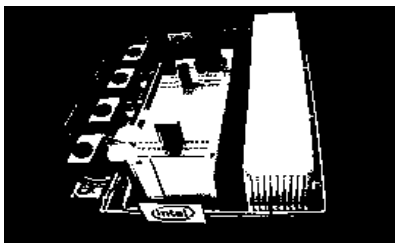
Superpočítač LUMI

- 8. na světě, 500 PFLOPS
- Procesory: AMD EPYC 7763 (362 496 jader)
- Grafické karty: AMD Radeon Instinct MI250X (11 912 karet)
- Síť HPE Cray Slingshot-11
- Umístění: Kajaani, Finsko
- Dostupný pro výzkumníky z ČR
 - <https://www.it4i.cz/pro-uzivatele/jak-ziskat-vypocetni-cas>
- Objektové úložiště LUMI-O: 30 PB přes S3 protokol



Superpočítače v IT4Innovations

Anselm (0.1 PF)



July 2013

DGX-2 (2 PF^{AI})



March 2019

Karolina (15.7 PF , #69)



August 2021

LUMI-Q



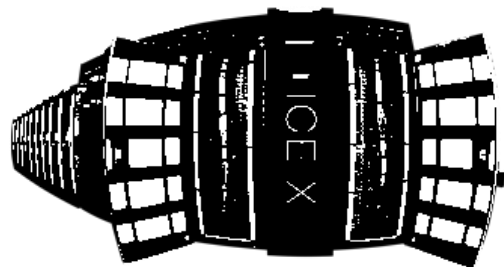
2025

2011

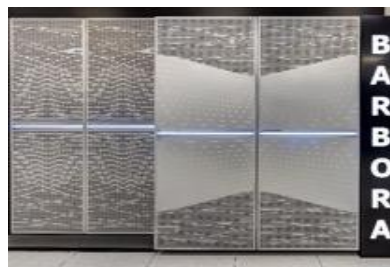
July 2015

October 2019

Autumn 2021



Salomon (2 PF, #40)



Barbora (0.8 PF)



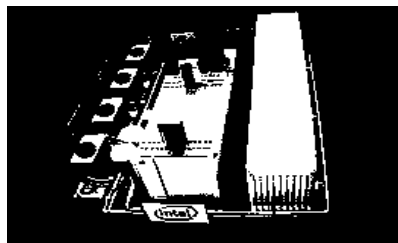
LUMI (500+ PF, #8, in Kajaani, FI)



EuroHPC
Joint Undertaking

Superpočítače v IT4Innovations

Anselm (0.1 PF)



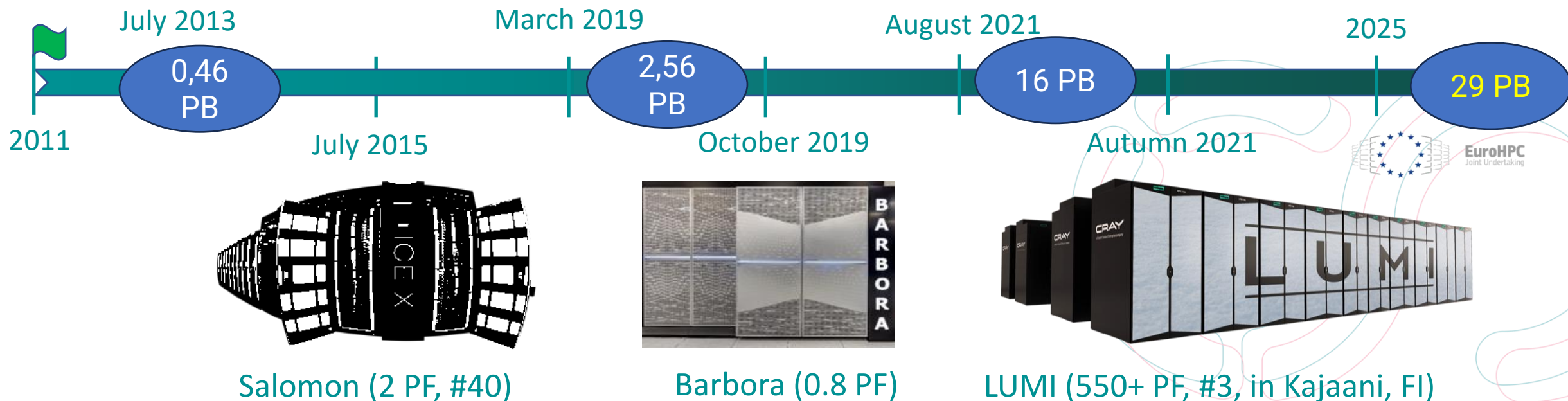
DGX-2 (2 PF^{AI})



Karolina (15.7 PF , #69)

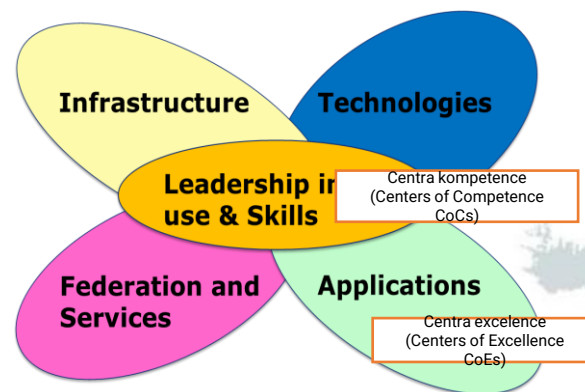


LUMI-Q



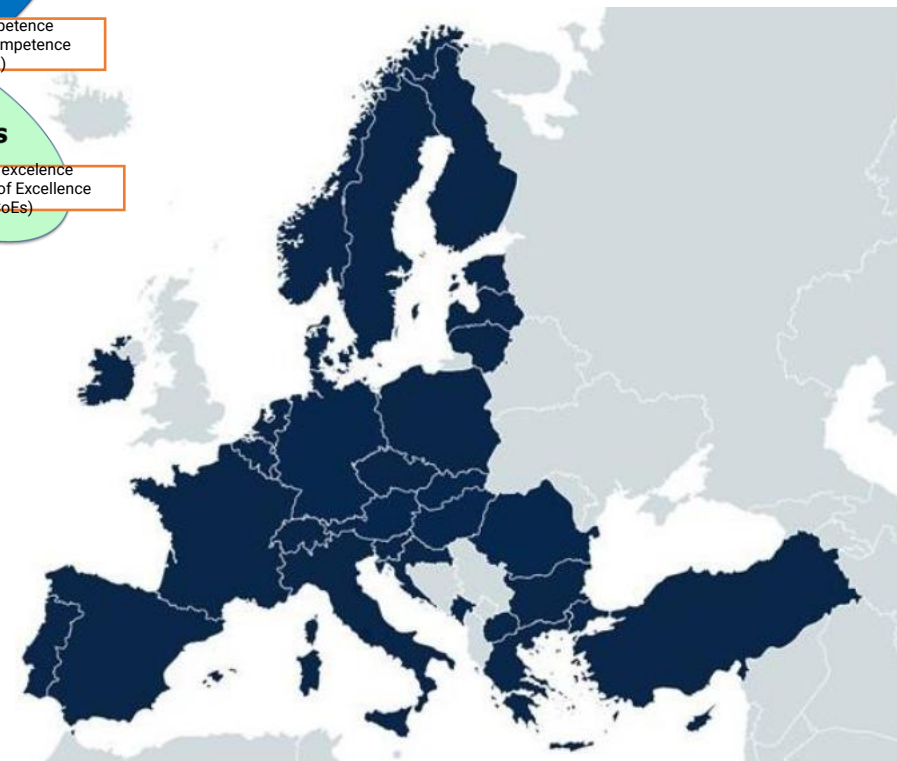
EuroHPC JU (Joint Undertaking)

- Evropské sdružení, které mělo v době vzniku cíl v rámci spolupráce evropských zemí podpořit cestu k vybudování vlastního evropského exascalového superpočítače do roku 2022/2023 a tím významně přispět k digitalizaci evropské společnosti a průmyslu
- Zapojeno 35 evropských států (ČR členem od ledna 2018), EU a 3 soukromé subjekty (ETP4HPC, BDVA a QulC)



Aktuální cíle

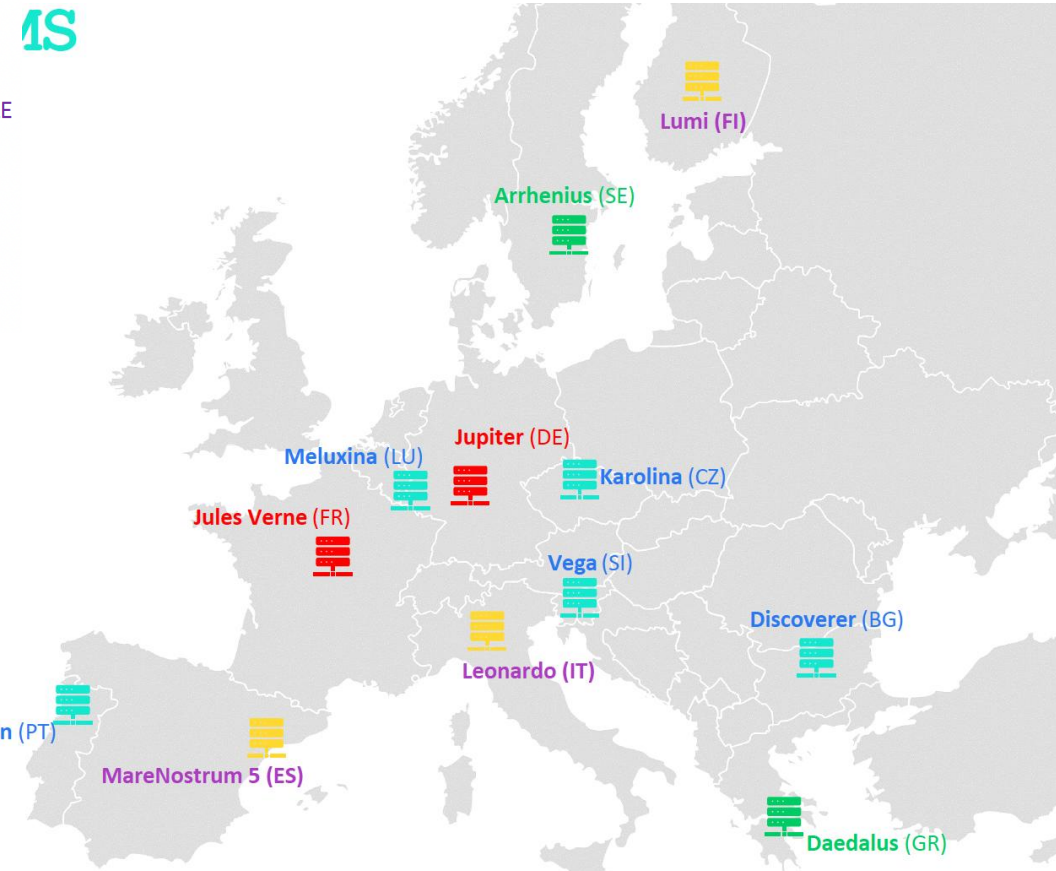
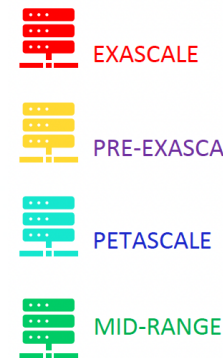
- Vývoj, nasazení a rozšiřování propojeného a bezpečného ekosystému superpočítačů, kvantových počítačů, služeb a datové infrastruktury světové třídy v Evropě.
 - Podpora vývoje a zavádění konkurenceschopného a inovativního superpočítačového systému na základě uživatelských požadavků včetně vývoje aplikací optimalizovaných pro tyto systémy.
 - Rozšíření využití superpočítačové infrastruktury a podpora znalostí a dovedností pro evropskou vědu a průmysl
- <https://eurohpc-ju.europa.eu>



EuroHPC JU – Evropské superpočítače

světové třídy

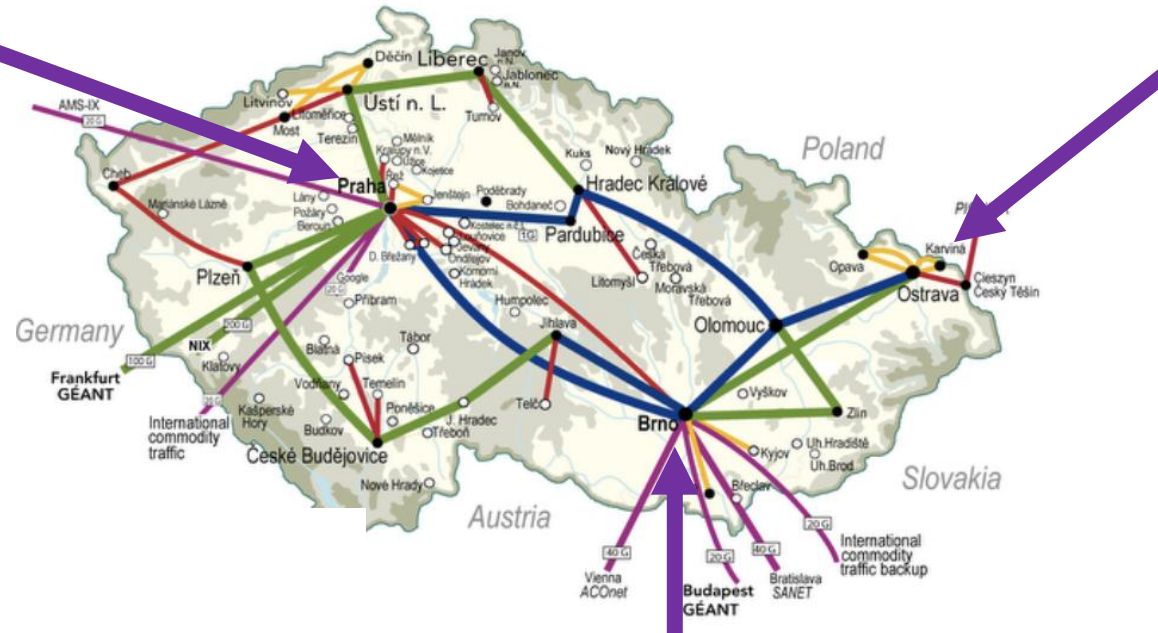
- 2 exa-scale a 2 mid-range superpočítače (2024-2026)
 - Jupiter (Ge)
 - Jules Verne (FR)
 - Arrhenius (SE)
 - Daedalus (GR)
- 3 pre-exascale superpočítače (2021-2023)
 - LUMI - Kajaani (FI)
 - Leonardo - Bologna (IT)
 - MareNostrum5 - Barcelona (ES)
- 5 petascale superpočítačů (2020-2021)
 - MeluXina - Bissen (LU)
 - Karolina - Ostrava (CZ)
 - VEGA - Maribor (SI)
 - Deucalion - Guimaraes (PT)
 - Discoverer - Sofia (BG)
- Podpora HPC ekosystému:
 - Vývoj Evropského procesoru pro HPC a exascale pilotního systému (EPI)
 - HPC centra kompetence (CoC) propagující inovace a vzdělávací aktivity ve všech participujících státech
 - Podpora průmyslu, MSP
- Poskytování výpočetního času
- Přes Open Access výzvy



Česká E-infrastruktura

- Velké národní infrastruktury pro výzkum, experimentální vývoj a inovace podporované Ministerstvem mládeže a tělovýchovy
- Poskytování vysoce kvalitních a dostatečně dimenzovaných ICT služeb pro výzkumnou komunitu

cesnet



cerit
scientific cloud

VSB TECHNICAL
UNIVERSITY
OF OSTRAVA

IT4INNOVATIONS
NATIONAL SUPERCOMPUTING
CENTER



<https://www.e-infra.cz/o-nas/clenove>

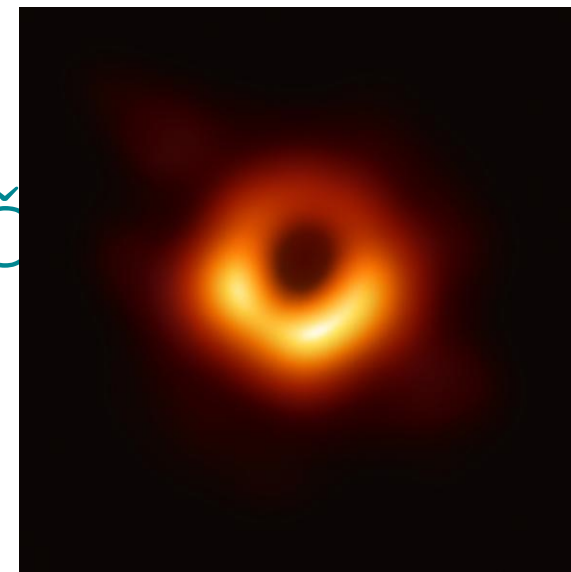
IT4Innovations

- Provoz nejvýkonnějších superpočítačových technologií v České republice
- Provozuje několik superpočítačů
 - Anselm, Salomon (vyřazeny)
 - Barbora
 - DGX-2
 - Karolina (2021)
 - LUMI (2022)
 - LUMI-Q (2025)
- Výzkumné a vývojové centrum se silnými mezinárodními vazbami
 - Zapojení v řadě mezinárodních projektů
- Výzkumné laboratoře
 - Laboratoř pro náročné datové analýzy a simulace
 - Laboratoř vývoje paralelních algoritmů
 - Laboratoř modelování pro nanotechnologie
 - Laboratoř pro kvantové výpočty
 - Laboratoř pro výzkum infrastruktury
- <https://www.it4i.cz>

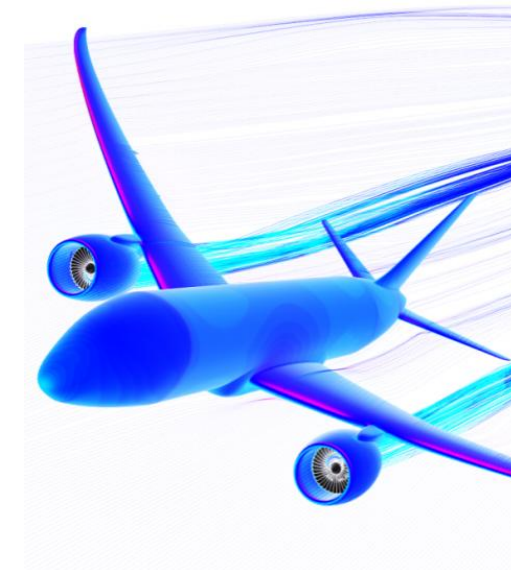
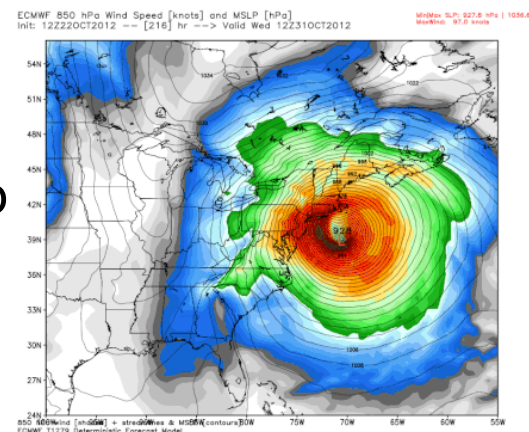


Příklady úloh pro superpočítač

- Střednědobá a dlouhodobá předpověď počasí, modelování klimatických změn
- Bioinformatika, vývoj nových léčiv, analýza DNA sekvencí, modelování proteinů
- Zpracování velkých dat, využití pro strojové učení, umělou inteligenci (např. velké jazykové modely LLM)
- Virtuální realita, rendering videa, zpracování obrazu
- Matematické a fyzikální výpočty, materiálové inženýrství (nové materiály)
- Simulace v různých odvětvích (CFD)
- Digitální dvojčata (digital twins), např. model lidského mozku, model země
- Kvantové počítání, atd.



Černá díra M87*, 5.5 PB dat, Event Horizon Telescope



Využití superpočítačů - paralelizace

Návrh software nebo systému tak, aby fungoval paralelně a mohl využít výpočetní kapacity

Cíl: paralelní využití výpočetních uzlů a jader

- Různé druhy paralelizace

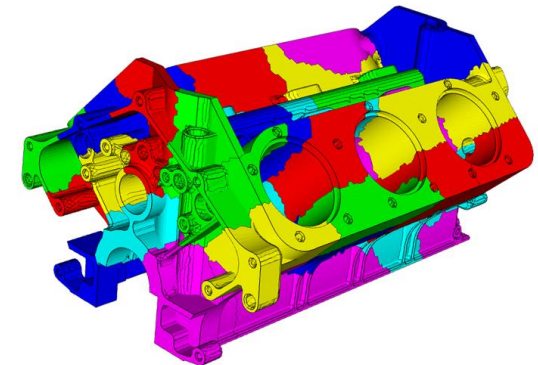
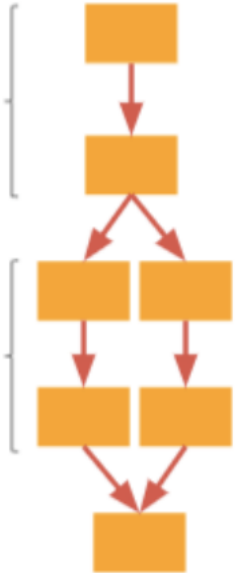
- Z pohledu algoritmu
 - Řešení náročných úloh lze rozdělit na menší úlohy
 - Speciální úprava algoritmů (sekvenční vs paralelní přístup)
- Z pohledu komunikace
 - Práce s velkými daty nebo náročné úlohy vyžadující komunikaci mezi výpočetními uzly
- Z pohledu množství zpracovaných dat
 - Sdílení dat mezi výpočetními uzly (rychlé propojení)

Sekvenční

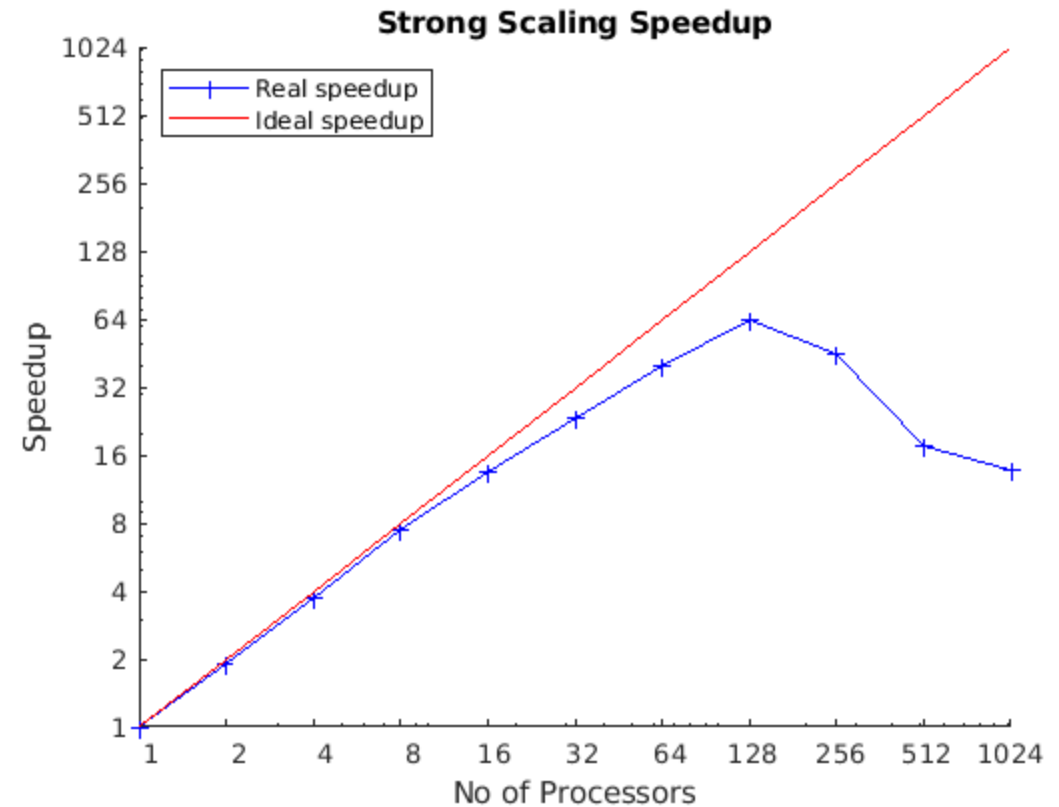
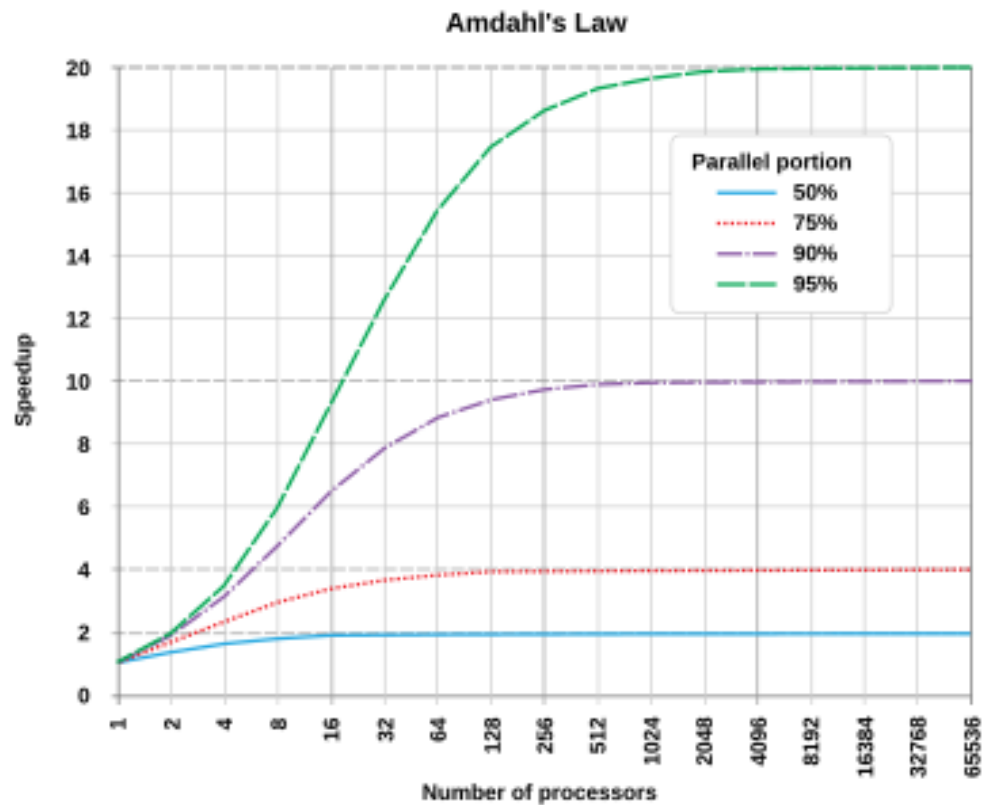


Sekvenční

Paralelní



Škálovatelnost – Amdahl's law



EuroHPC JU AI Factories

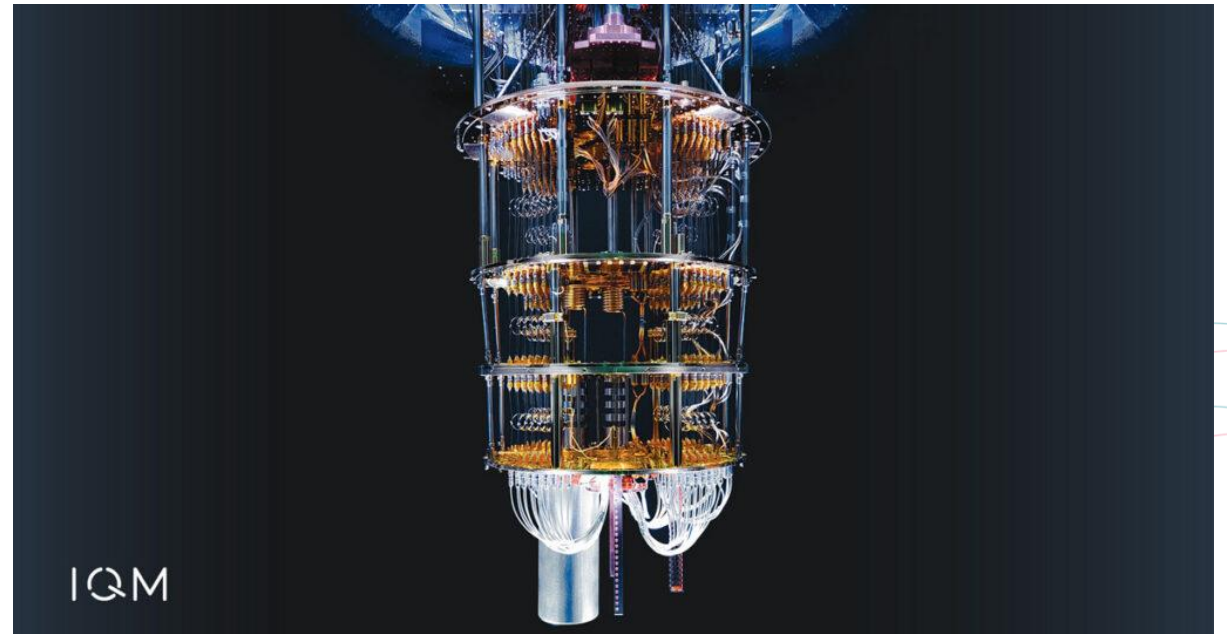
- 5 specializovaných AI superpočítačů napříč Evropou
- Podpora AI pro výzkumné skupiny
- Dostupné pro
 - Veřejné výzkumné organizace
 - AI startupy
 - SMEs



Více na: https://eurohpc-ju.europa.eu/selection-first-seven-ai-factories-drive-europes-leadership-ai-2024-12-10_en

Kvantový počítač VLQ konsorcia LUMI-Q

- Reálný kvantový stroj
- 24 fyzických supravodivých qubitů
- Chlazení na teplotu 10 mK
- Star topologie
- Právě instalovaný v IT4I v Ostravě



Více na: https://eurohpc-ju.europa.eu/selection-first-seven-ai-factories-drive-europes-leadership-ai-2024-12-10_en

Přenášení dat



Spolufinancováno
Evropskou unií

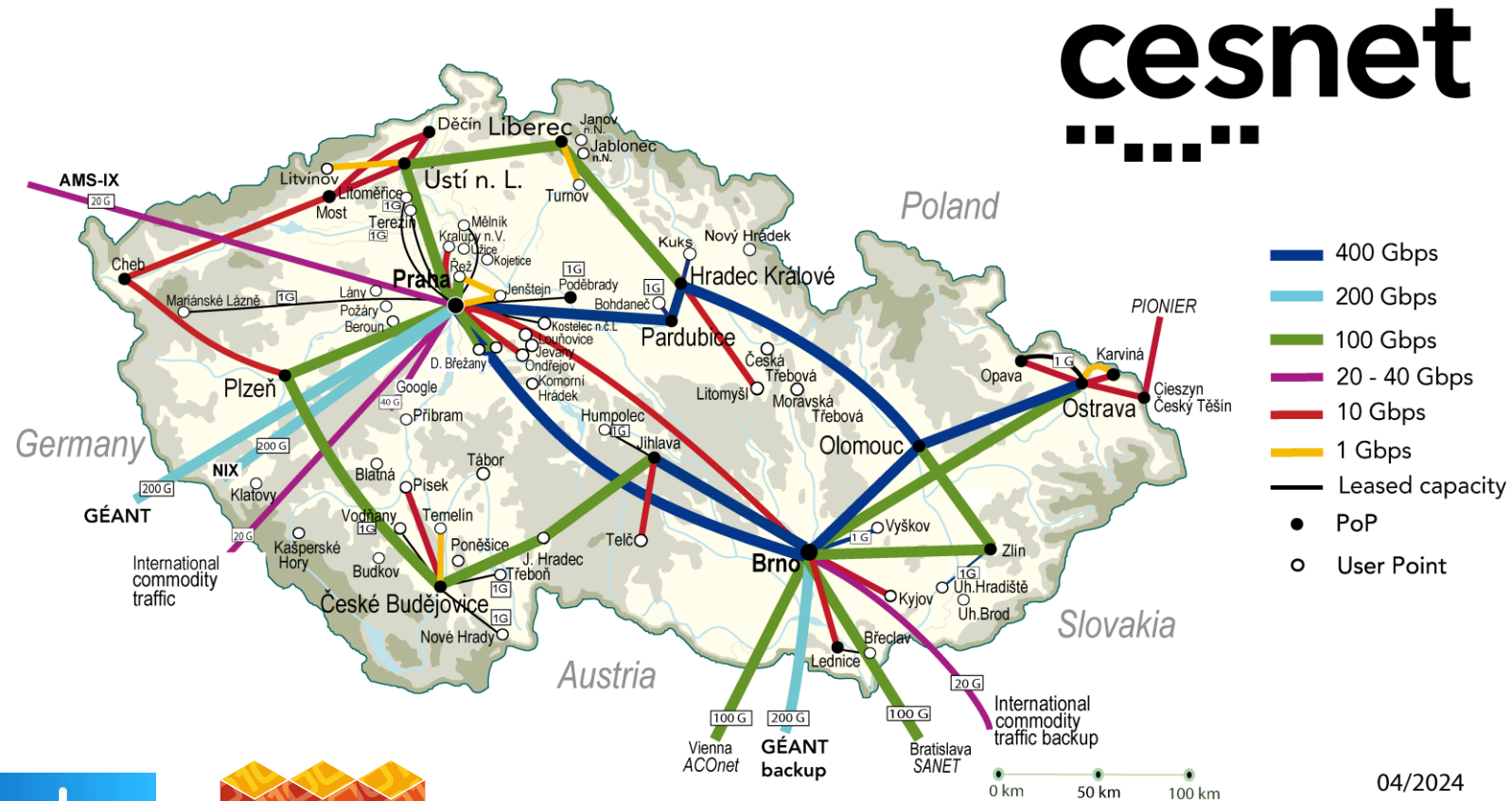


MUNI
ICS

Registrační číslo IPs EOOSC-CZ
CZ.02.01.01/00/22_004/0007682

Přenášení dat – připojení k síti

- Velmi rychlé sítě jsou dostupné
- Důležité je jejich správně využití
- Kabelové připojení v kanceláři 1 Gbit/s max.
- Lokální server až 100 Gbit/s

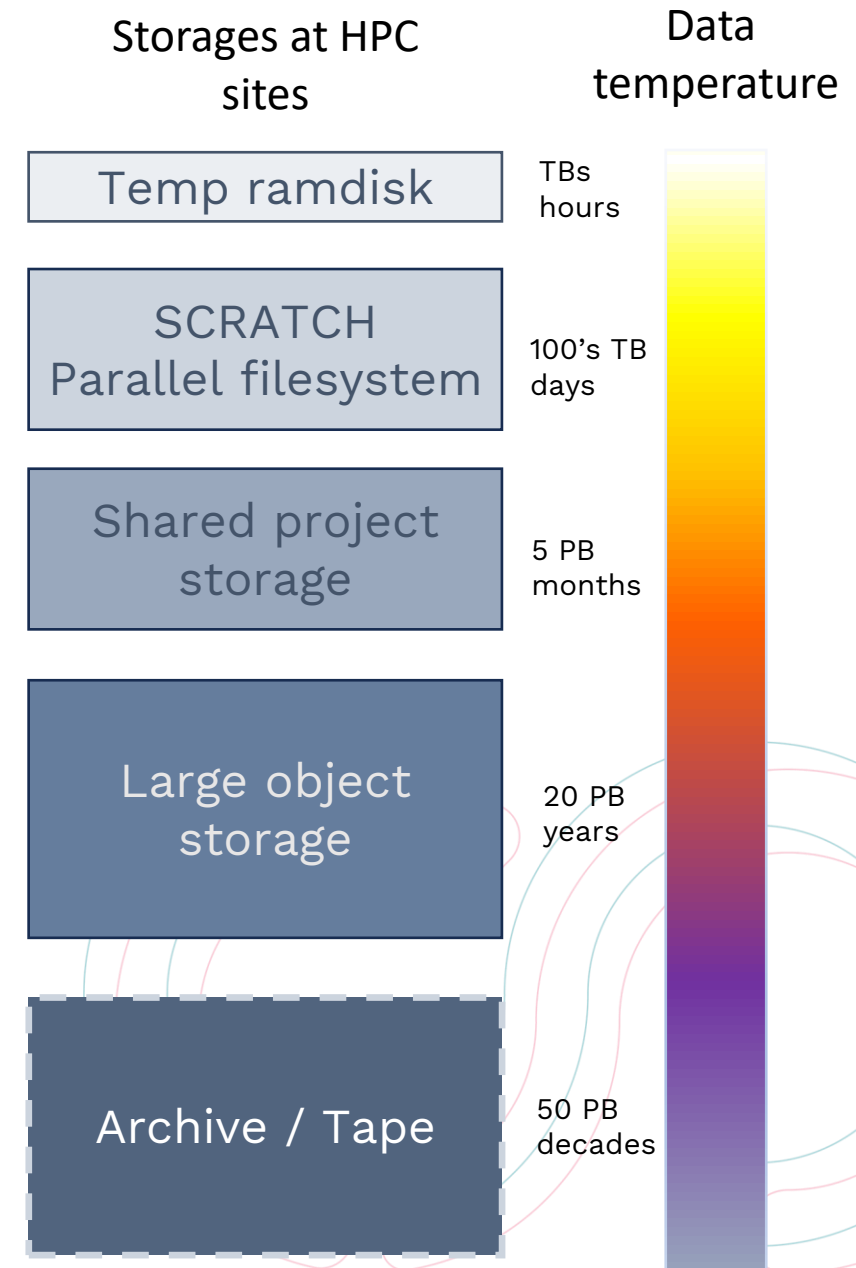


Hugging Face

04/2024

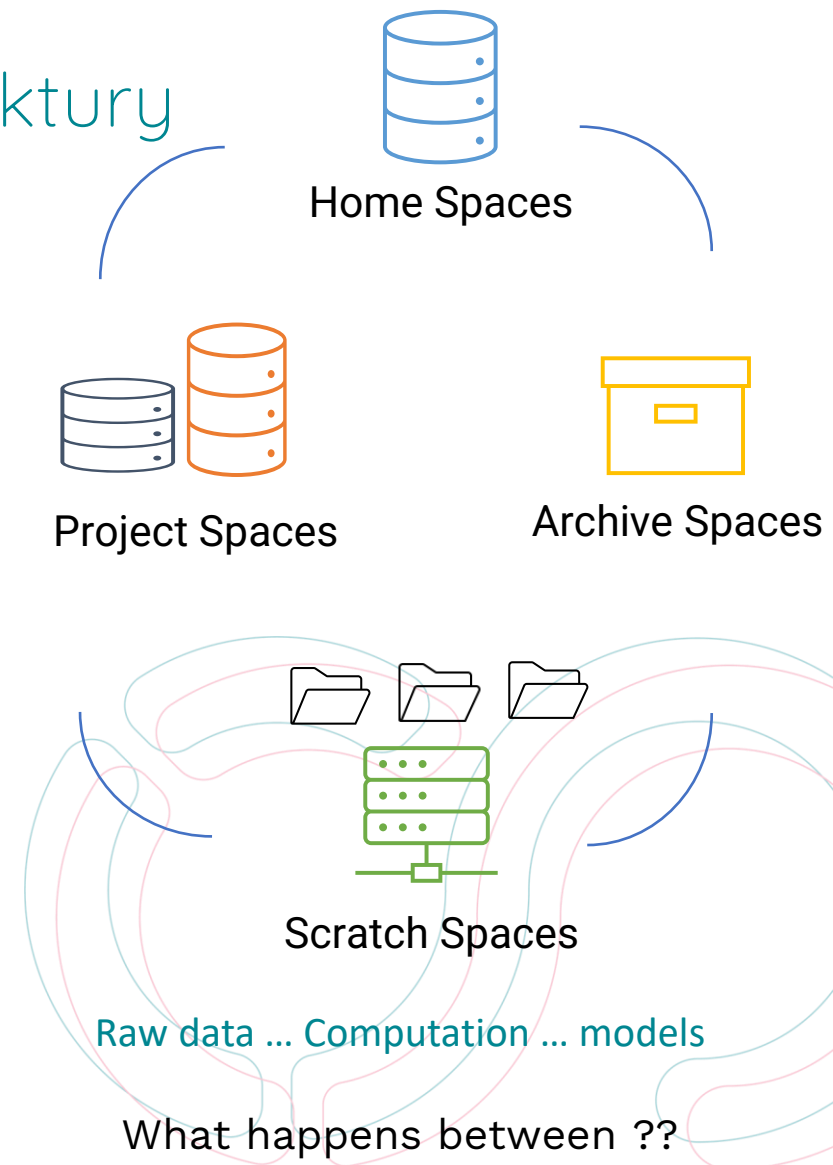
Typy úložišť v superpočítačových centrech

- Typy technologií jsou dané
 - Flash (NVMe, SSD, etc..)
 - HDD
 - Pásky
- Způsob přístupu k nim je standardně přes POSIX (soubory/složky)
 - Nepraktické pro přenos dat přes internet
 - SSH je uživatelsky nepřívětivé
- Objektová úložiště mohou být řešení
 - S3 není standardizované, každý výrobce má svou modifikaci
 - Jedním z řešením je použití abstrakcí jako iRODS
- Archivní média jako např. pásky není možno používat transparentně
- Archivace a obnova velkých souborů i uvnitř datového centra je časově náročná



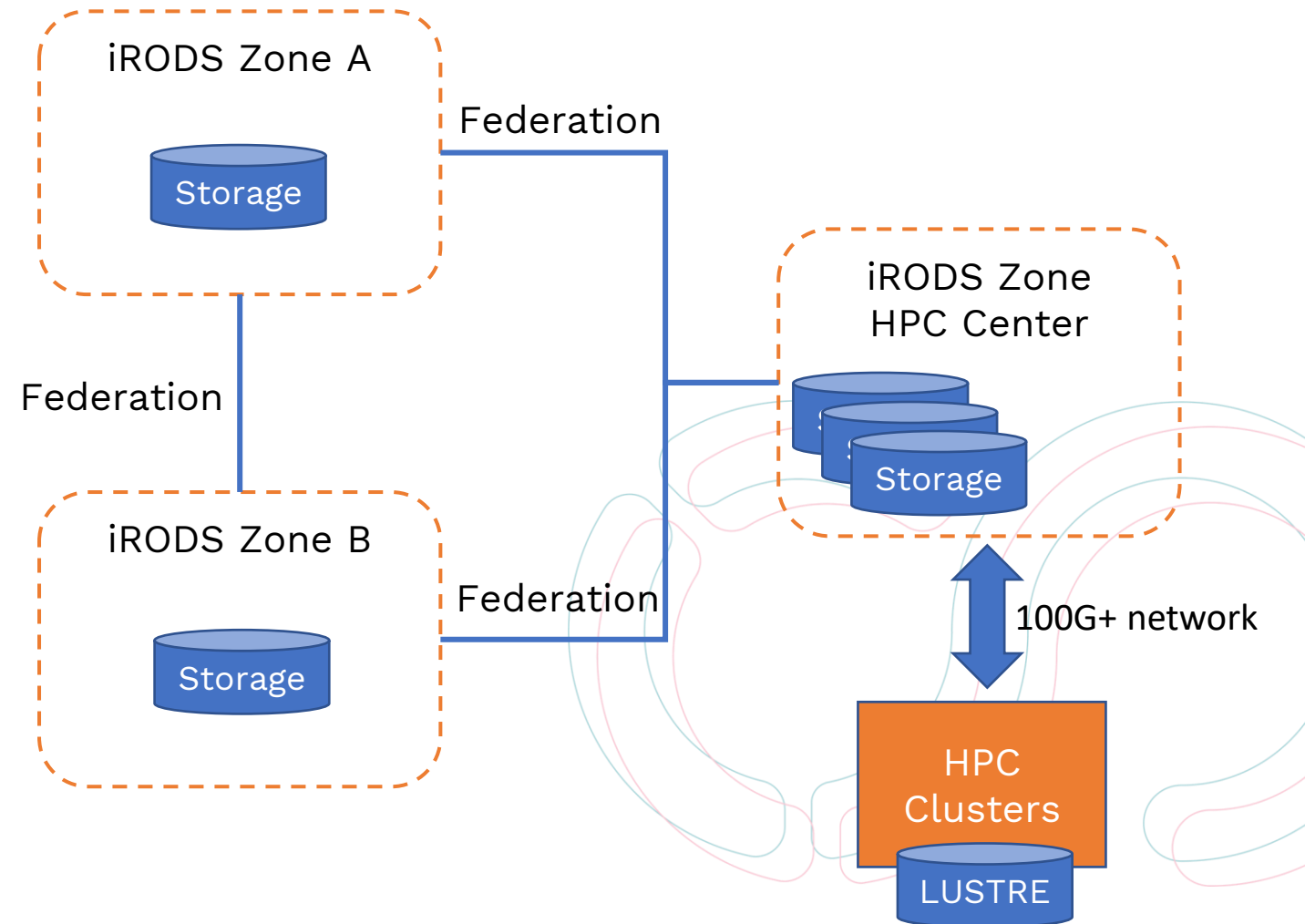
Životní cyklus dat z pohledu výpočetní infrastruktury

- Uživatelský pohled
 - Pokud možno automatizovaný přenos s maximálním využitím linek
 - Součást výpočetního workflow
 - Přístup k datům je “in-kind” v rámci alokace výpočetního času
 - Jak zajistit přístup k velkým datům napříč “projekty”?
- Administrátorský pohled
 - Nutnost trasování původu a použití dat
 - Sběr tzv. provenance údajů pro výsledky
 - Vstupní data – reference
 - Výpočetní prostředí, knihovny, hardware
 - Audit stopa – kdo, kdy, kde
 - Možno využít i pro FAIR



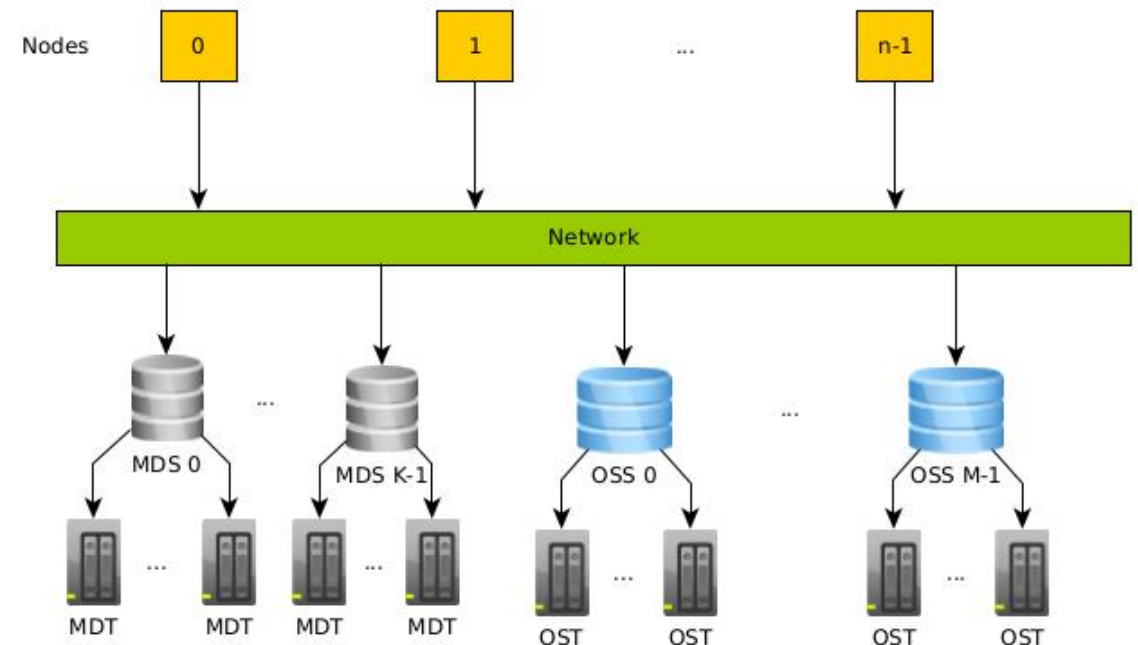
Příklad společného nástroje pro přenášení dat

- Shodný protokol napříč datovými centry
- Přesun dat co nejblíže výpočetním kapacitám
- Autonomie každého centra musí být zachována
- Společné AAI a výkonné sítě



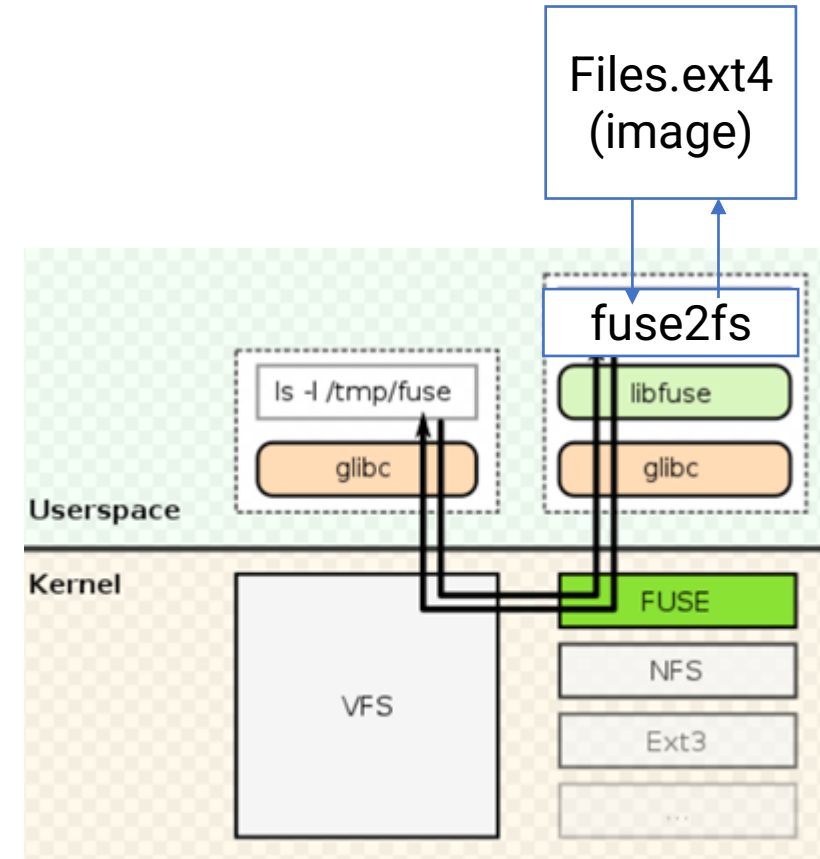
Přístup k datům na superpočítačích

- Velké množství (miliony) souborů může způsobovat problémy
- Sdílený souborový systém (LUSTRE, GPFS)
 - Omezený počet tzv. inode – reprezentace složky/souboru
 - Metadatové servery jsou sdílené napříč uživateli
 - Listování složek apod. může být časově velmi náročné
- Velké soubory rozdělené mezi několik uzlů
 - Je nutné dbát na správnou velikost tzv. stripe
 - Soubor je pak rozdělen na části, které jsou uloženy na nezávislých discích



Přístup k datům na superpočítačích

- Velké množství (miliony) souborů
 - Vlastní souborový systém v jednom souboru
 - Zatěžuje pouze daný výpočetní uzel, ze kterého se přistupuje
 - Nástroje
 - fuse2fs – připojení souborových systémů jako např. ext4
 - archivemount – připojení archivů např. tar, zip, rar..
 - Mnoho dalších FUSE klientů (např. iRODS nebo S3)
- Doporučení
 - Správně strukturovat data do větších bloků
 - Ty ukládat na vhodná úložiště (objektové)
 - Vhodně přenášet objekty na /scratch složku clusteru a přistupovat přes FUSE





Data Management in HPC – selected techniques

- Umělá inteligence není jen trénování sítí...

Training

- Konečné množství dat – **dávková úloha**
- Je nutné přenést data co nejbližší GPU – pro opakované čtení
- Vhodným způsobem ukládat výsledky

Inference

- Dávková úloha i **požadavky**
- Inference jako služba není vhodná pro dávkové plánovače – čeká na požadavky
- Je nutné efektivně nahrávat modely do GPU – vyvarovat se opakovaného nahrávání např. pro každý požadavek

HPC Cluster



HyperQueue



CPU
Partition

GPU
Partition

Efektivní datová paralelizace

Motivace

- Komplexní workflow v rámci jedné dávkové úlohy jsou náročný úkol
 - Úkoly jsou heterogenní a tvoří složité topologie
 - Superpočítače mají různé typy výpočetních uzlů
 - Není vhodné zasílat velké množství velmi krátkých úloh do plánovače typu SLURM (sdílený se všemi uživateli clusteru)

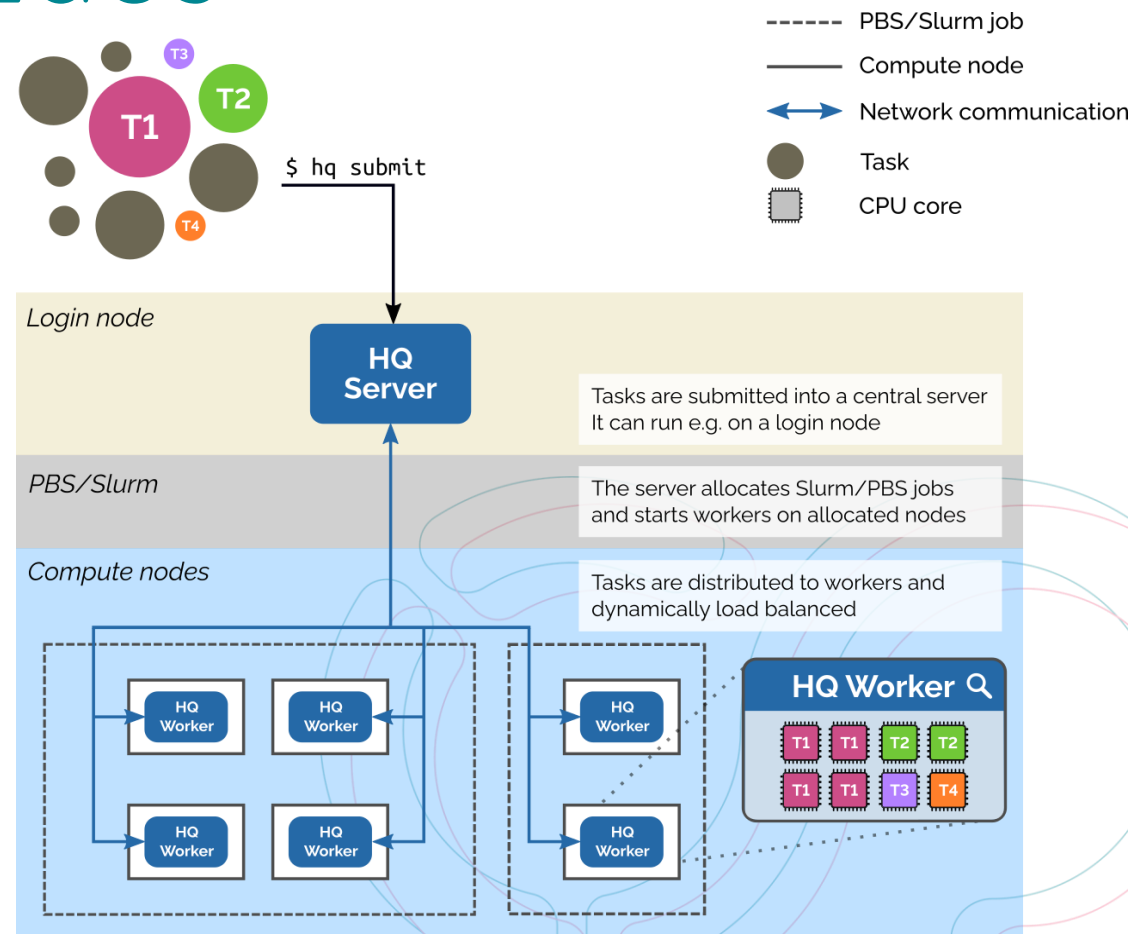
HyperQueue je plánovač optimalizovaný pro běh na HPC clusterech, pro efektivní spouštění složitých workflows s mnoha úkoly

```
#!/bin/bash

while :
do
    # Submit a job and wait for it to complete
    ./hq submit --wait ./compute.sh

    # Read the output of the job
    output=$(./hq job cat last stdout)

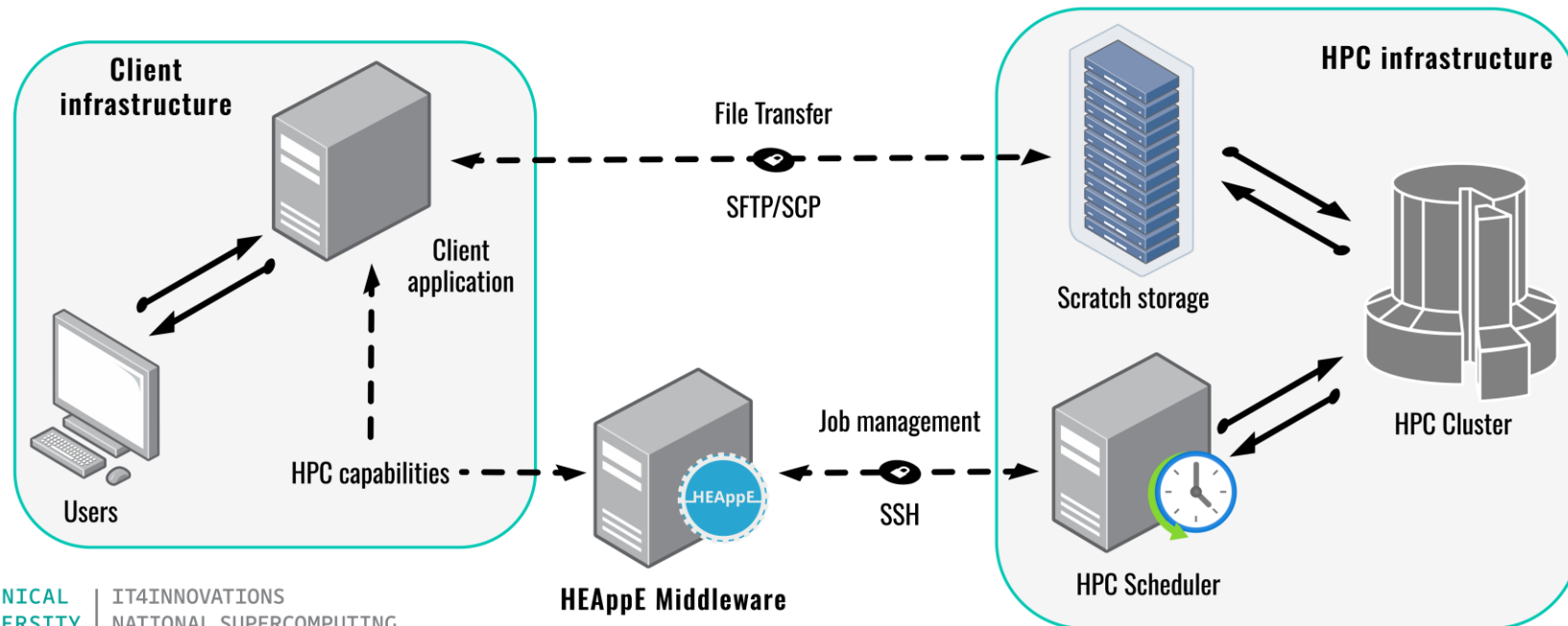
    # Decide if we should end or continue
    if [ "${output}" -eq 0 ]; then
        break
    fi
done
```



Open-source (MIT-licensed) available at GitHub:
<https://github.com/It4innovations/hyperqueue>

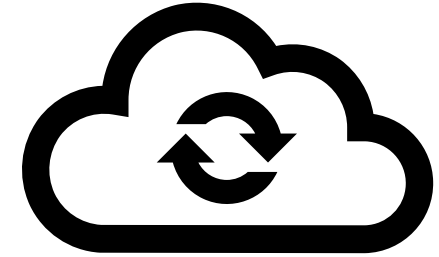
HEAppE as-a-Service přístup

- High-End Application Execution Middleware
 - Přístup k superpočítačům přes REST API
 - Bezpečný a izolovaný přístup
 - Vlastní autentizace & autorizace uživatelů
 - Monitoring spouštěných úloh, včetně přístupu k logům



<https://heappe.eu/>

Cloud computing



- Poskytování IT infrastruktury přes internet, bez nutnosti mít vlastní hardware
- Účtuje se okamžitá spotřeba
- Virtuální instance, kontejnery, sítě
- Zdánlivě nekonečné zdroje – vhodné pro škálování služeb obsluhujících mnoho požadavků
- Populární řešení
 - OpenStack – Virtuální stroje
 - Kubernetes, OpenShift - kontejnery
- Komerční: Amazon, Microsoft, Google
- Pro výzkumníky v ČR:
 - CESNET Metacentrum Cloud
 - IT4I EuroHPC Cloud
- **IaaS** (Infrastructure as a Service) – infrastruktura jako služba
- **PaaS** (Platform as a Service) – platforma jako služba
- **SaaS** (Software as a Service) – software jako služba

Superpočítač prakticky



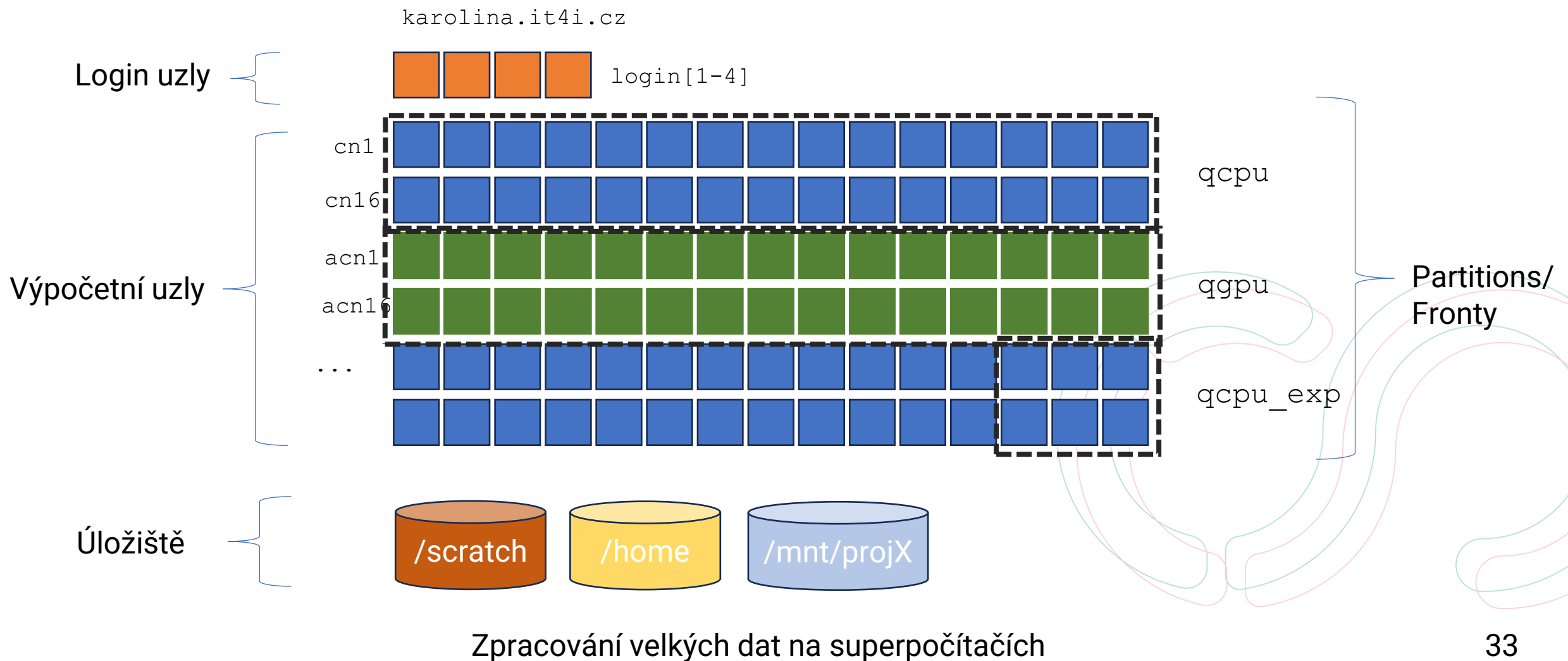
Spolufinancováno
Evropskou unií



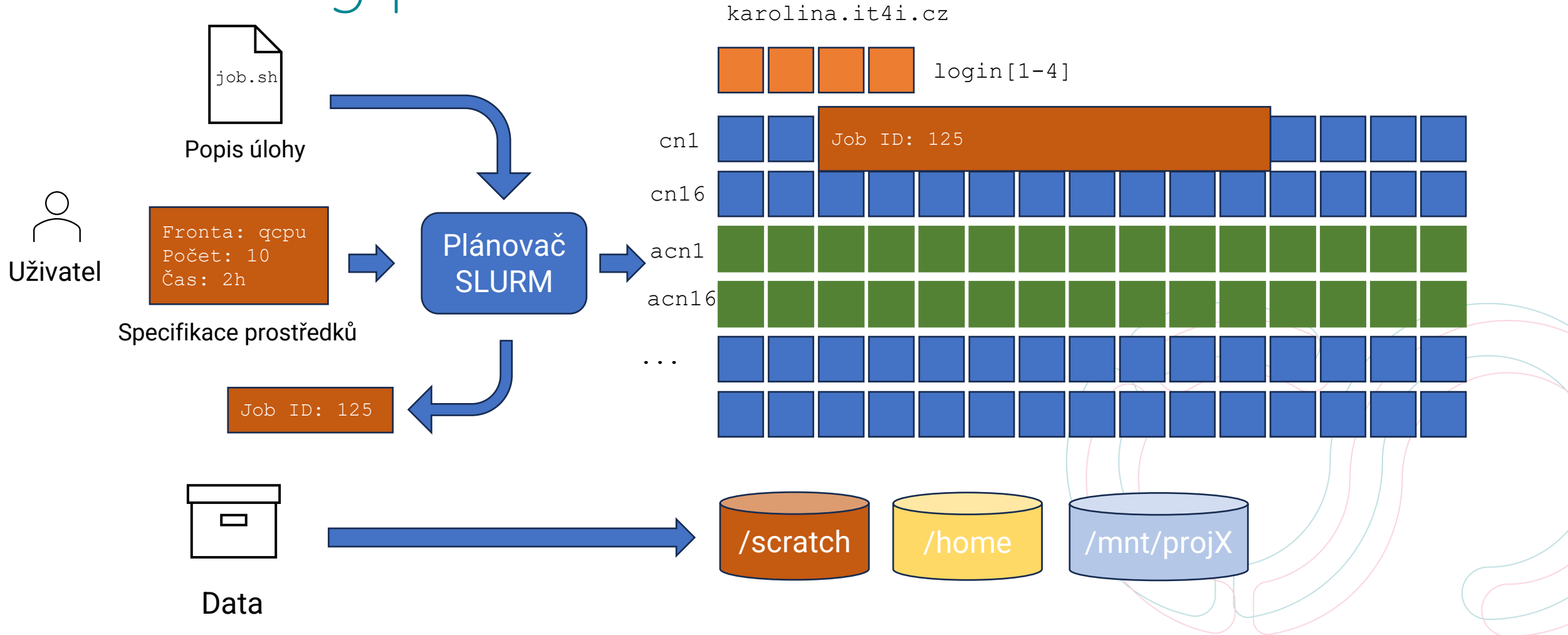
MUNI
ICS

Registrační číslo IPs EOSC-CZ
CZ.02.01.01/00/22_004/0007682

Organizace superpočítače



Dávkový plánovač



Přenos dat

- SSH nástroje (rsync/sftp)
 - Přes SSH server na login uzlech clusteru
 - Vhodné pro menší soubory, není implicitně paralelní, možné paralelizovat ručně
 - Příklad:
 - `rsync -p -v test10G.dat mgolas@karolina.it4i.cz:/home/mgolas/test10G.dat`
- Lokální objektové úložiště S3/iRODS
 - Paralelní přenosové protokoly – nutné použít pro efektivní využití přenosových linek
 - Příklad s S3 úložištěm
 - Vstupní data od uživatele na S3 úložiště
 - `s5cmd --stat --credentials-file ./wp.cfg --endpoint-url https://obj.proj4.it4i.cz cp --sp ~/test10G.dat s3://wp/test10G.dat`
 - Stáhnutí dat na clusteru (login uzel nebo dávková úloha na 1 uzlu)
 - `s5cmd --stat --credentials-file ./wp.cfg --endpoint-url https://obj.proj4.it4i.cz cp --sp s3://wp/test10G.dat ~/test10G.dat`

Jak se měří používání superpočítače?

- Výpočetní čas, který úloha stráví v běhu na clusteru je měřen v tzv. uzlo-hodinách (nodehour)
- 1 uzlohodina = 1 uzel používaný po celou hodinu
- Většina clusterů má exkluzivní alokaci uzlů – nejmenší možná jednotka, kterou uživatel může využít je jeden s nikým nesdílený uzel
- Více na: <https://docs.it4i.cz/general/resource-accounting/>



Jak plánovač určuje prioritu úlohy?

Partition	Project resources	Nodes	Min ncpus	Priority	Authorization	Walltime (def/max)
qcpu	> 0	720	128	2	no	24 / 48h
qcpu_biz	> 0	720	128	3	no	24 / 48h
qcpu_exp	< 150% of allocation	720 max 2 per user	128	4	no	1 / 1h

Fronty plánovače clusteru Karolina

Více informací na: <https://docs.it4i.cz/general/job-priority>

Fair-share priority is calculated based on recent usage of resources. Fair-share priority is calculated per project, i.e. all members of a project share the same fair-share priority. Projects with higher recent usage have a lower fair-share priority than projects with lower or no recent usage.

Job priority is calculated as:

$$10^8 * \text{queue_priority} + 10^7 * \text{fairshare_priority} + 10^5 * \text{job_age}$$

Vzorec pro výpočet priority úlohy

Jak získat výpočetní čas na LUMI nebo Karolině?

Detailní informace zde: <https://www.it4i.cz/pro-uzivatele/jak-ziskat-vypocetni-cas>

- Zaměstnanci výzkumných organizací ČR mají čas dostupný zdarma
- Přístup je excellence based – je nutné požádat o zdroje na základě výzkumného záměru
- Komerční využití je možné taktéž
- Otevřená soutěž IT4I
 - Pravidelné národní výzvy
 - Vyhlašovány 3x ročně
 - Až na 36 měsíců
 - Institucionální a komunitní přístup
 - Rychlý a testovací
 - Tématický
- EuroHPC JU výzvy
 - Regular call (větší zdroje, delší vyhodnocení)
 - Development & Benchmarking (měsíční)

Chcete počítat ihned?

<https://www.it4i.cz/pro-uzivatele/jak-ziskat-vypocetni-cas>

- Podejte žádost do Rychlého a testovacího přístupu – tzv. Fast Track Access (FTA)
- Rozhodnutí max. do 7 dní od podání
- Maximální délka alokace 4 měsíce

UNIVERSITY OF OSTRAVA | NATIONAL SUPERCOMPUTING CENTER

Fast track access

NOTICE! Delete all text in grey. Do not exceed the maximum document size of 3 pages.

Name of the project:

Project duration: Max. 4 months

Number of node hours requested for each platform:

- a. **Barbora CPU:** Max. 1400 NH
- b. **Barbora GPU:** Max. 60 NH
- c. **Barbora FAT:** Max. 10 NH
- d. **DGX-2:** Max. 10 NH
- e. **Karlin GPU:** Max. 500 NH

Popular abstract:

Include a popular abstract in a form which is immediately available for publication on the website or in newspapers etc., outlining the related research, the methods to be used, and the expected impact, in language appropriate for the general public.

Be concise; do not exceed 1500 characters in the abstract.

Technical readiness and computational resources justification:

Technical readiness:

Describe your experience with the requested resources. List the computational methods and software tools, and libraries that you will use and describe their suitability and current limitations to efficiently utilize the requested resources (if possible, provide related references and data for your application's performance).

Computational resources justification:

Justify the requested computational resources. Provide the basis on which the requested resources were estimated.

Platforma LEXIS



Spolufinancováno
Evropskou unií



MUNI
ICS

Registrační číslo IPs EOOSC-CZ
CZ.02.01.01/00/22_004/0007682

How to supercompute?

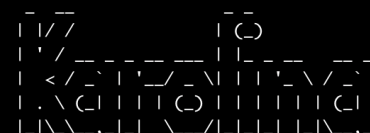
- Pick a supercomputer
- Fill out a request for allocation
- Get it approved
- Set up an account
- Set up a SSH key
- Know how to use Linux terminal
- Login in to the Supercomputer
- Learn how to launch jobs in SLURM
- Learn about modules
- Learn about storages and data transfer
- ... finally compute something

```
GROMACS/4.5.5-gompi-2020c-ORCA-5.0.1
GROMACS/2021.4-foss-cuda-2020b-PLUMED-2.7
GROMACS/2022-foss-cuda-2020b
GROMACS/2023-foss-2022a-CUDA-12.2.0-PLUMED-2.7
GROMACS/2023-foss-2022a-CUDA-12.2.0-PLUMED-2.7
GROMACS/2023.4-gompi-2020c-ORCA-5.0.1-docker
GROMACS/2024.1-foss-2023b

OpenFOAM/v2312-foss-2023a OpenFOAM/9-
...
ABINIT/9.10.3-intel-2022a
ASE/3.22.1-gfbf-2023b
Amber/22.0-foss-2022a-AmberTools-23.3-CUDA
BEEF/0.1.1-intel-2020a
CP2K/9.1-foss-2022a
Critic2/1.1stable-intel-2020b
DFT-D4/3.6.0-intel-2022b-Python-3.10.8
DFT-D4/3.6.0-intel-2022b-Python-3.10.8
Libint/2.7
Molden/7.3
Molpro/mpp
Molpro/mpp
Molpro/mpp
lines 1-27
```



```
Your public key has been saved in the
The key fingerprint is:
SHA256:fFp7L5cfGT0jmWRJ1TvVwHRJCzUR
The key's randomart image is:
+---[RSA 3072]-----+
|      .o+=00|
|      ..++*|
|      .o|
|      ..o|
|      ..o|
```



...running on Rocky 8.X

Public Service Announcement: Apptainer on the Karolina cluster
Posted: (2024-05-10 10:23:47)

Apptainer is now a part of the operating system, you do not need to load the module.

```
$ apptainer --version
apptainer version 1.3.1-1.el8
```

```
Last login: Mon May 27 10:05:28 2024 from 89.24.247.60
[mgol@login2.karolina ~]$
```

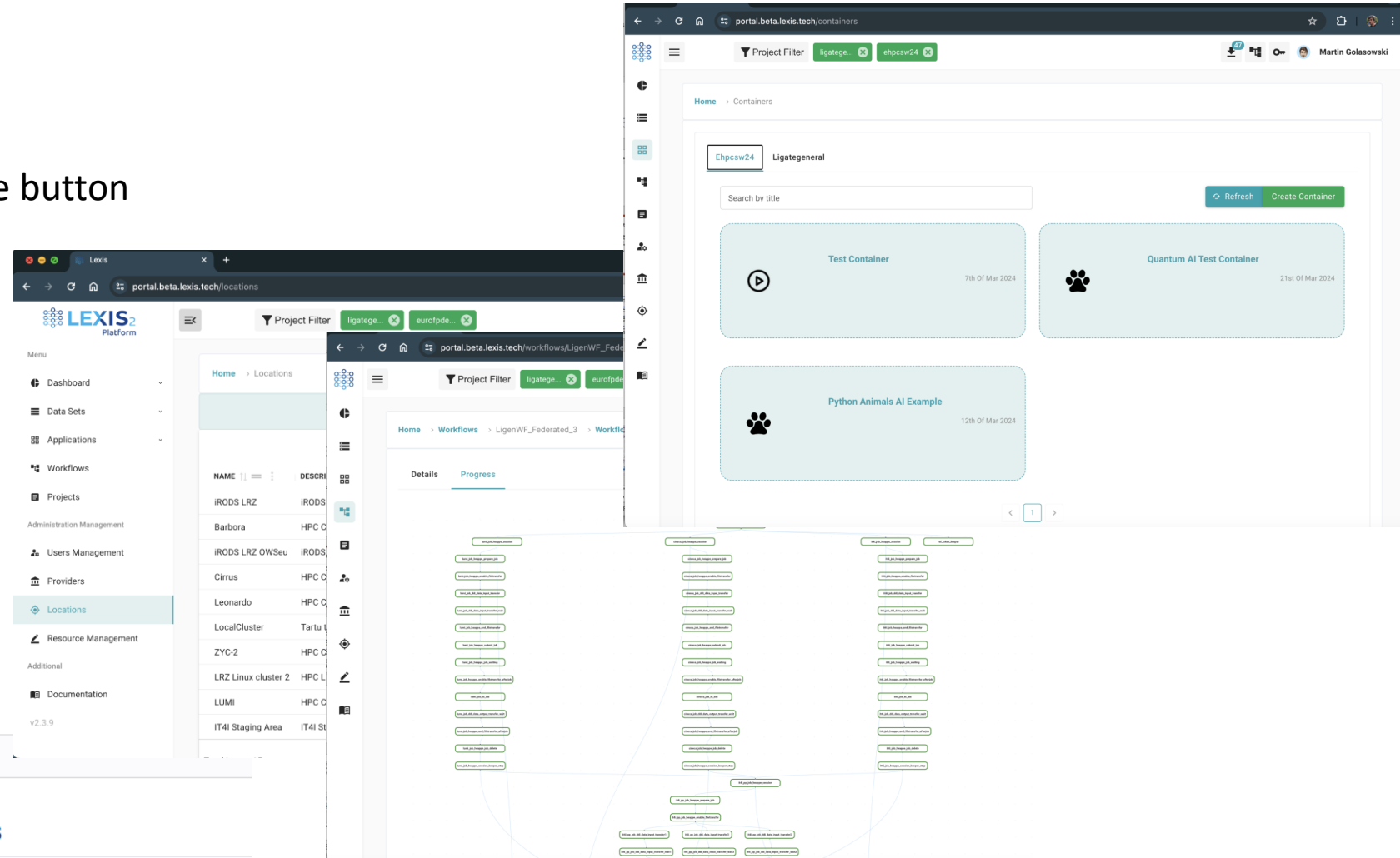
Easy and Safe access

- Let's have a nice web interface
- Allocations at one place
- Across multiple clusters
- Launch applications with a single button
- Get logs
- Manage data in iRODS
- Use common web login
- ... and many more

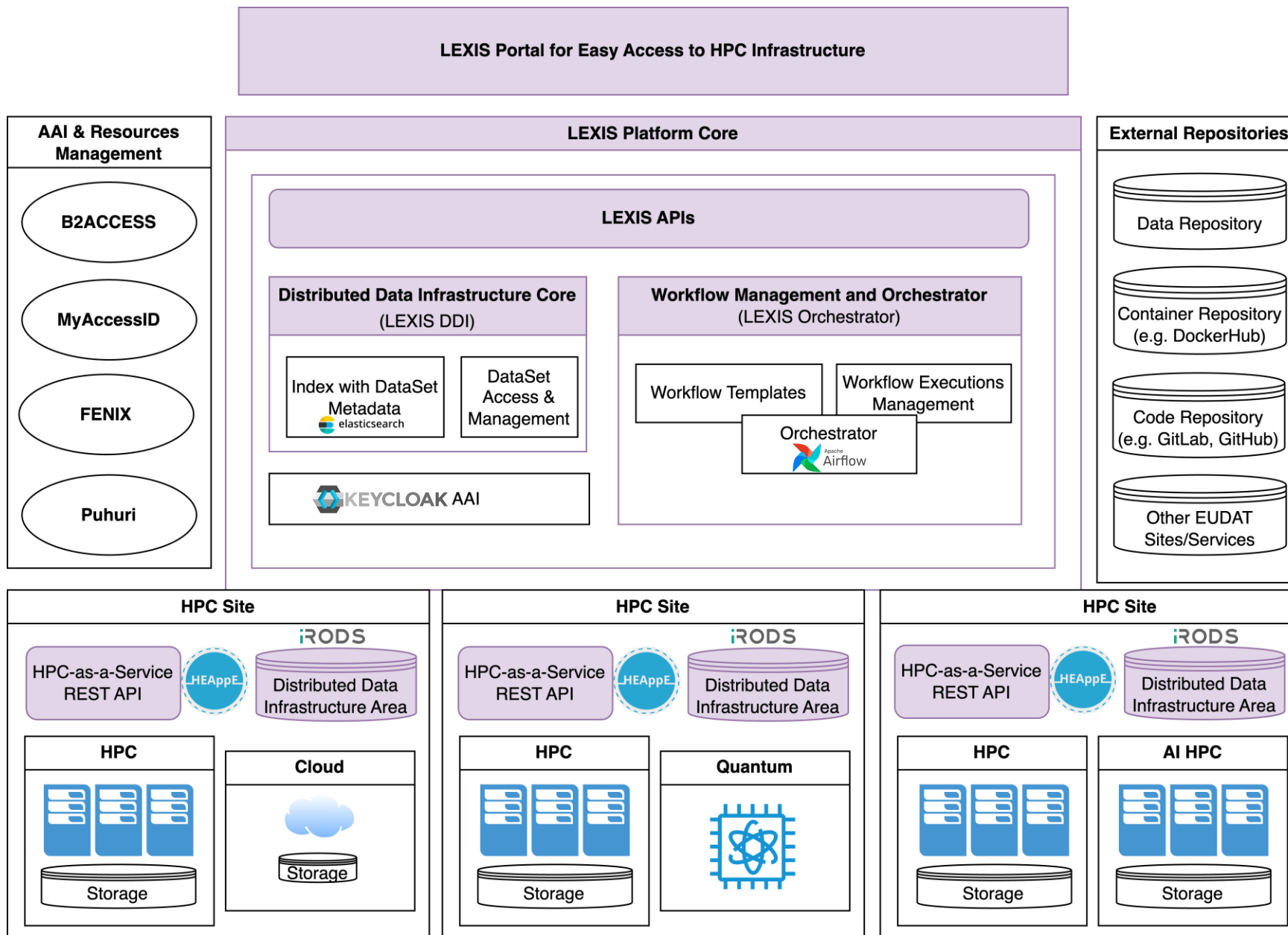
Visit for more:

docs.lexis.tech

opencode.it4i.eu/lexis-platform



LEXIS Platform



- Federation of **European computing centres**
- **Hiding** of technical and **operational differences** across organizations
- **HPC & Cloud** service providers, Data providers
- Unified & distributed **data management**
- Orchestration
- **Federated Authentication & Authorization Infrastructure (AAI)**

```

# README.md

py4lexis

Package py4lexis provides functions to manage datasets within the LEXIS platform. Uploading datasets' files is performed by tuspy
Python package, i.e. by TUS Client.

Install

We recommend to use virtual environment, e.g.:

python -m venv ./venv

Activated virtual environment, type:

python -m pip install py4lexis
```

```

preprocessing_start_date": "2022100500",
preprocessing_dataset_path_geographical_data_path": "project/proj934a1763631a81e

flowExecutionName": "Run3",
flowExecutionConsent": True

def prepare_data():
    prepared = WRF_EXEC_PARAMS
    prepared["inputParameters"]["preprocessing_start_date"] = datetime.datetime.now().str
    prepared["inputParameters"]["preprocessing_start_date"] = '2022120400'
    return prepared

def submit_wf_execution(token):
    res = requests.post(LEXIS_API + WORKFLOW_ID + '/execution',
    json = prepare_data(),
    headers = {'Authorization': 'Bearer ' + token})

    if res.status_code != 201:
        logger.error(res.content)
        return

    return res.json()

```

```
(venv) mgolas@Martins-MacBook-Pro ~/I/L/L/Py4Lexis (develop)> python examples/upload.py
```

Progress: | ██████████-----| 30.0% Uploaded

LEXIS Platform

Selected use-cases

LIGATE Project

Application for molecular docking simulation – private IP by DOMPÉ
LEXIS provides access to workflows with this application running on HPC
Without direct access to the binary or source code



OpenWebSearch.eu

European open web index processed through LEXIS on several HPC locations
(LRZ, IT4I, CSC, DLR)
Public indices made available through the LEXIS Portal



LEXIS Platform

User story - SME DOMPE & LiGen SW



- DOMPE has its own code co-developed with POLIMI & CINECA
- DOMPE LiGen code IP does not allow to share LiGen SW (even binary)
- DOMPE would like to extend its customer base thanks to LEXIS Platform and EuroHPC infrastructure without losing IP
- DOMPE would like to make available LiGen to non-profit academic research and public institutions as a Platform-as-a-Service solution for drug discovery

LEXIS and EXA4MIND Platforms Integration Plan



User interfaces
Web GUI
REST APIs



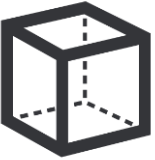
Resource
management



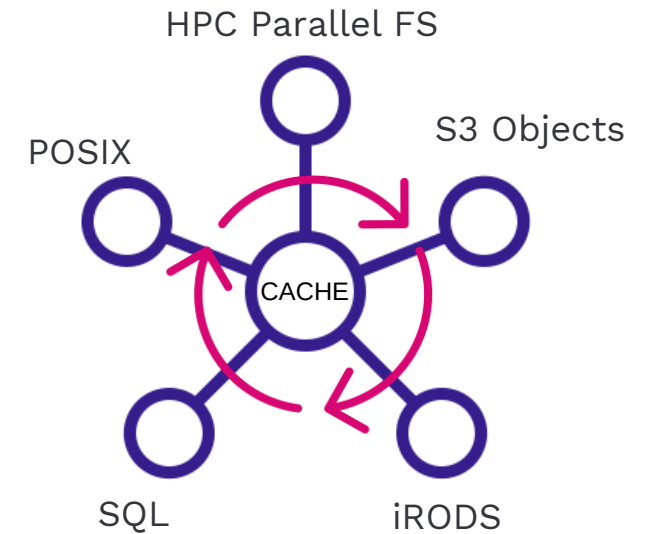
Workflow
orchestration



Metadata &
control APIs



Indexes



AAI & User
management



Monitoring



Smart
scheduling



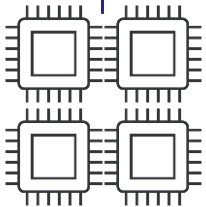
Distributed Data
Infrastructure

Data staging and transfer for Extreme Data

**Supercomputing
Centre
Resources**



Cloud compute

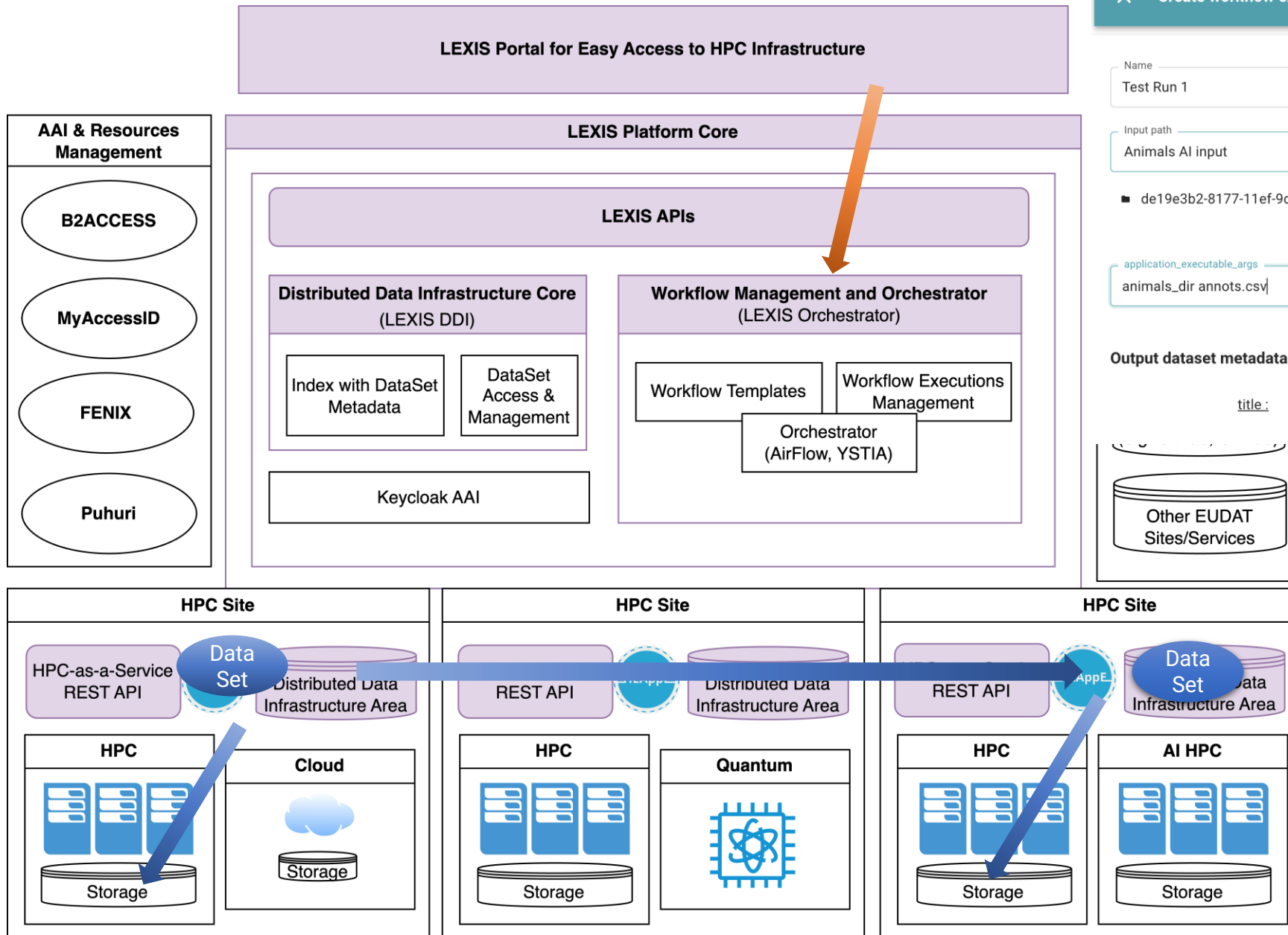


HPC Clusters



Remote storage





×

Create workflow execution

Name

Test Run 1

Input path

Animals AI input

×

▼

■ de19e3b2-8177-11ef-9de3-0242c0a8700a

application_executable_args

animals_dir annots.csv

Output dataset metadata :

title :

Animals AI Video Example 3 output dataset

value

Animals AI Video Example 3 output dataset

Home

>

Workflows

>

Animals_AI_Video_Example_3_Workflow

>

Workflow Execut

Details

Progress

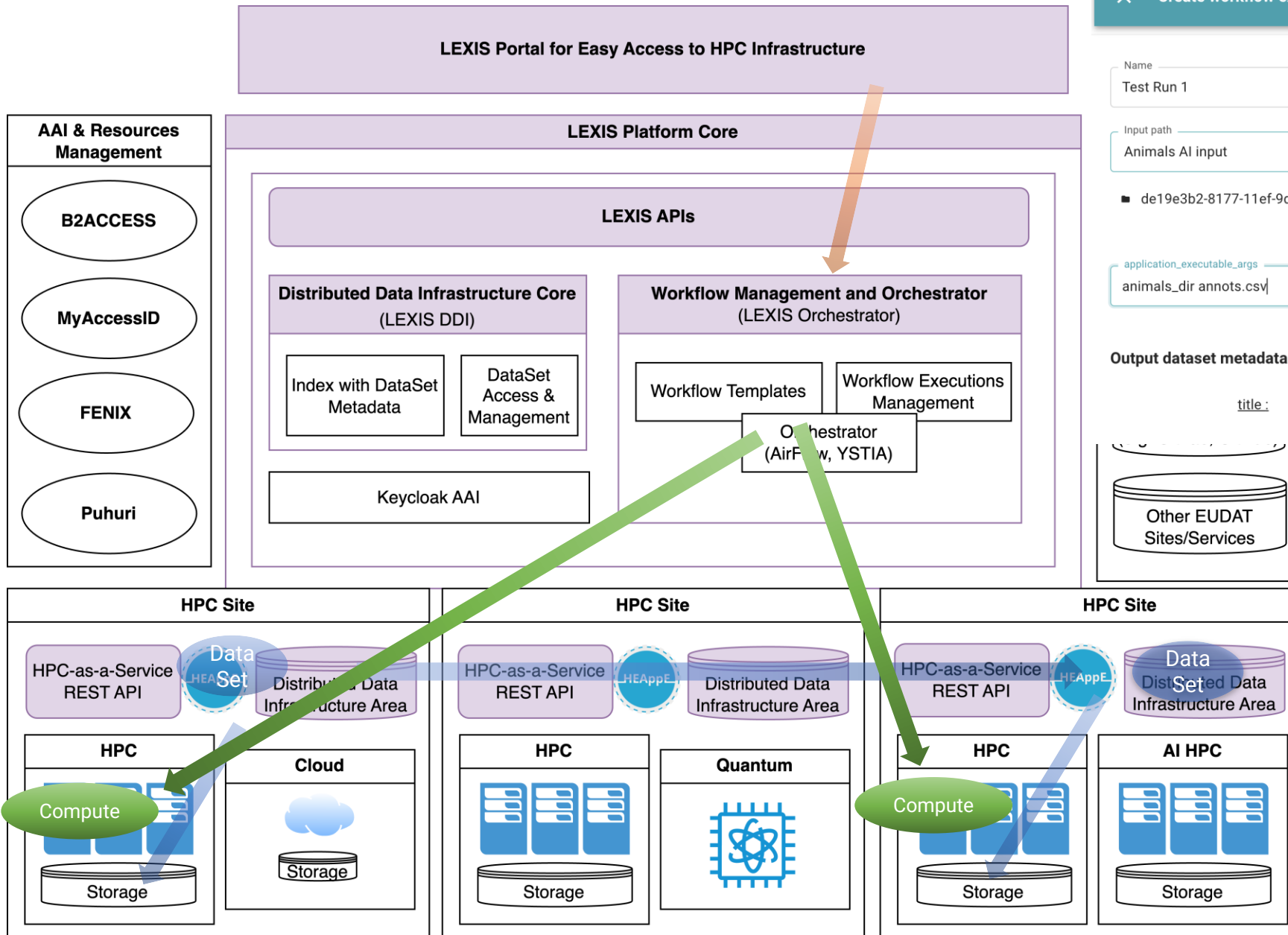
Workflow

job_runner_execution

ref_values_loader

Other EUDAT Sites/Services





Create workflow execution

Name

Test Run 1

Input path

Animals AI input

X ▼

■ de19e3b2-8177-11ef-9de3-0242c0a8700a

application_executable_args

animals_dir annots.csv

Output dataset metadata :

title :

Animals AI Video Example 3 output dataset

Other EUDAT Sites/Services

Home > Workflows > Animals_AI_Video_Example_3_Workflow > Workflow Execut

Details

Progress

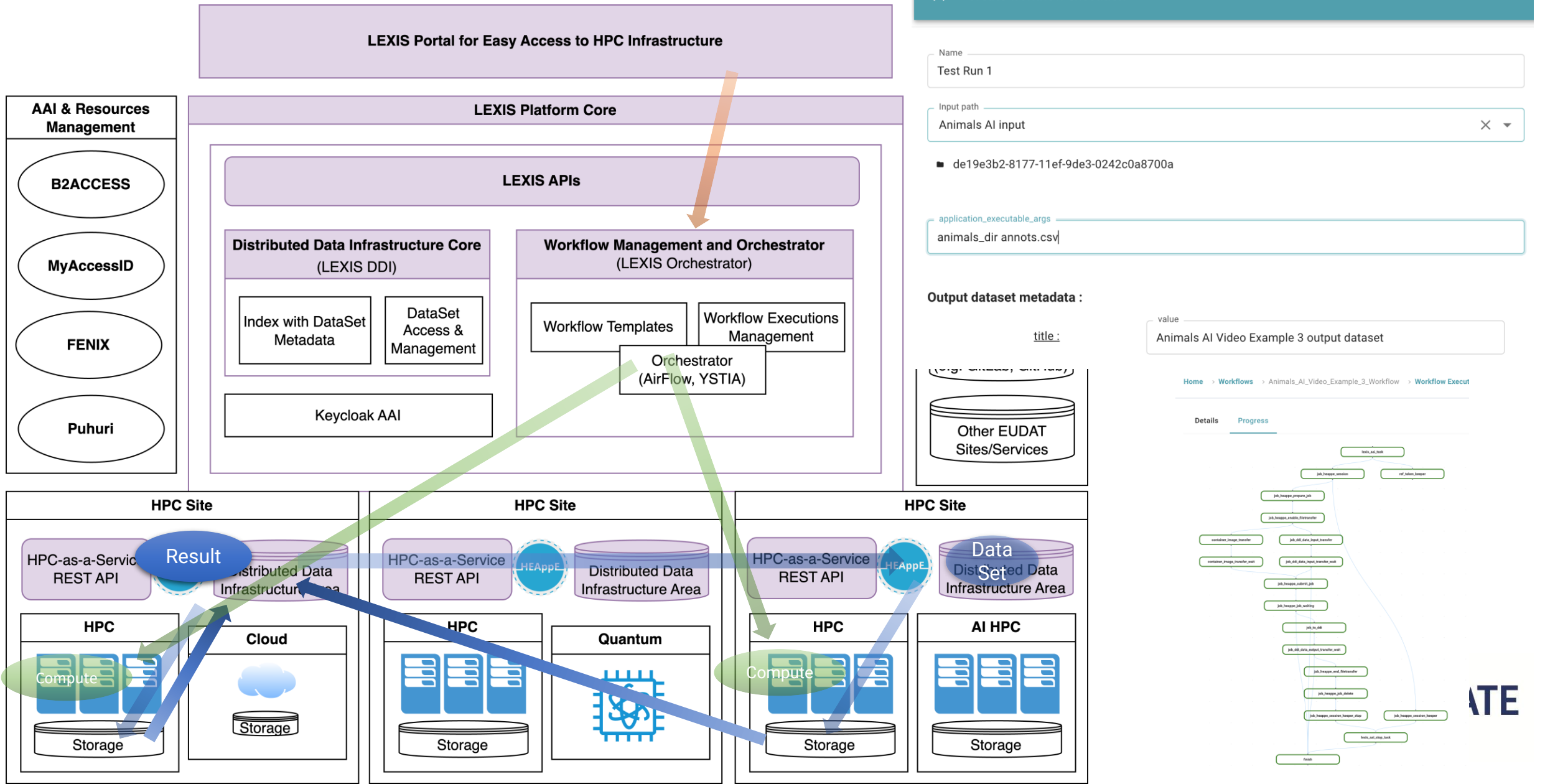
Workflow Execution Progress Diagram

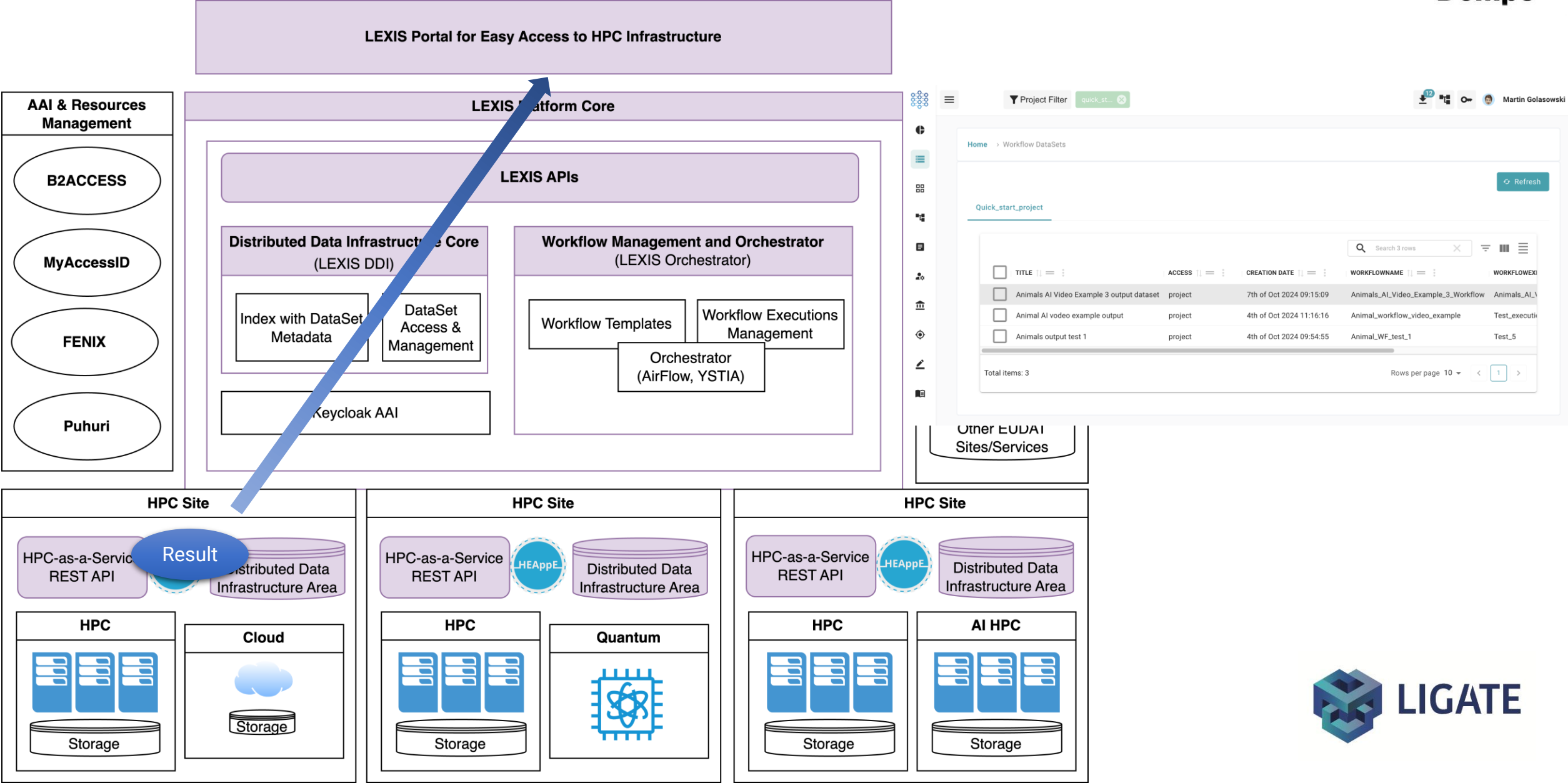
graph TD; A[Workflow Execution] --> B[job_package_execution]; A --> C[ref_data_loader]



LEXIS Platform

User story - Workflow execution





Airflow integration in LEXIS

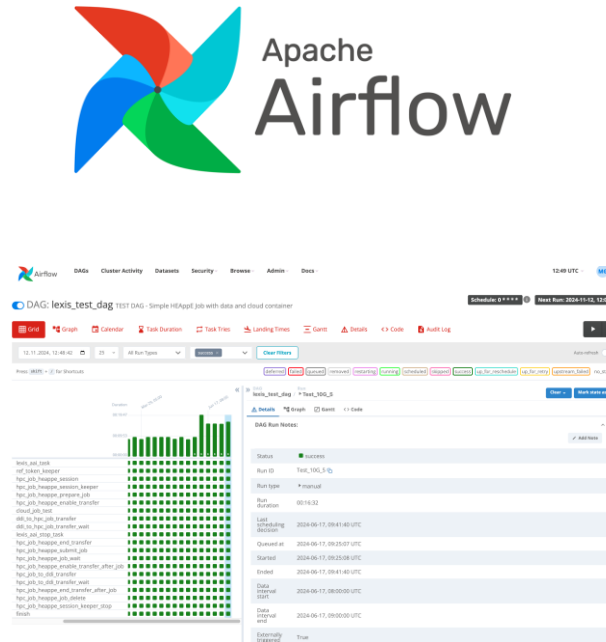
Apptainer Container

HEAppE Template

Jobscripts

CWL

Custom Workflow
DAG



Workflow specification

Integration with LEXIS APIs

LEXIS
Provider

DDI, HEAppE, AAI
Operators

LEXIS API
Plugin

AAI, Container/HEAppE
templates
TOSCA
CWL

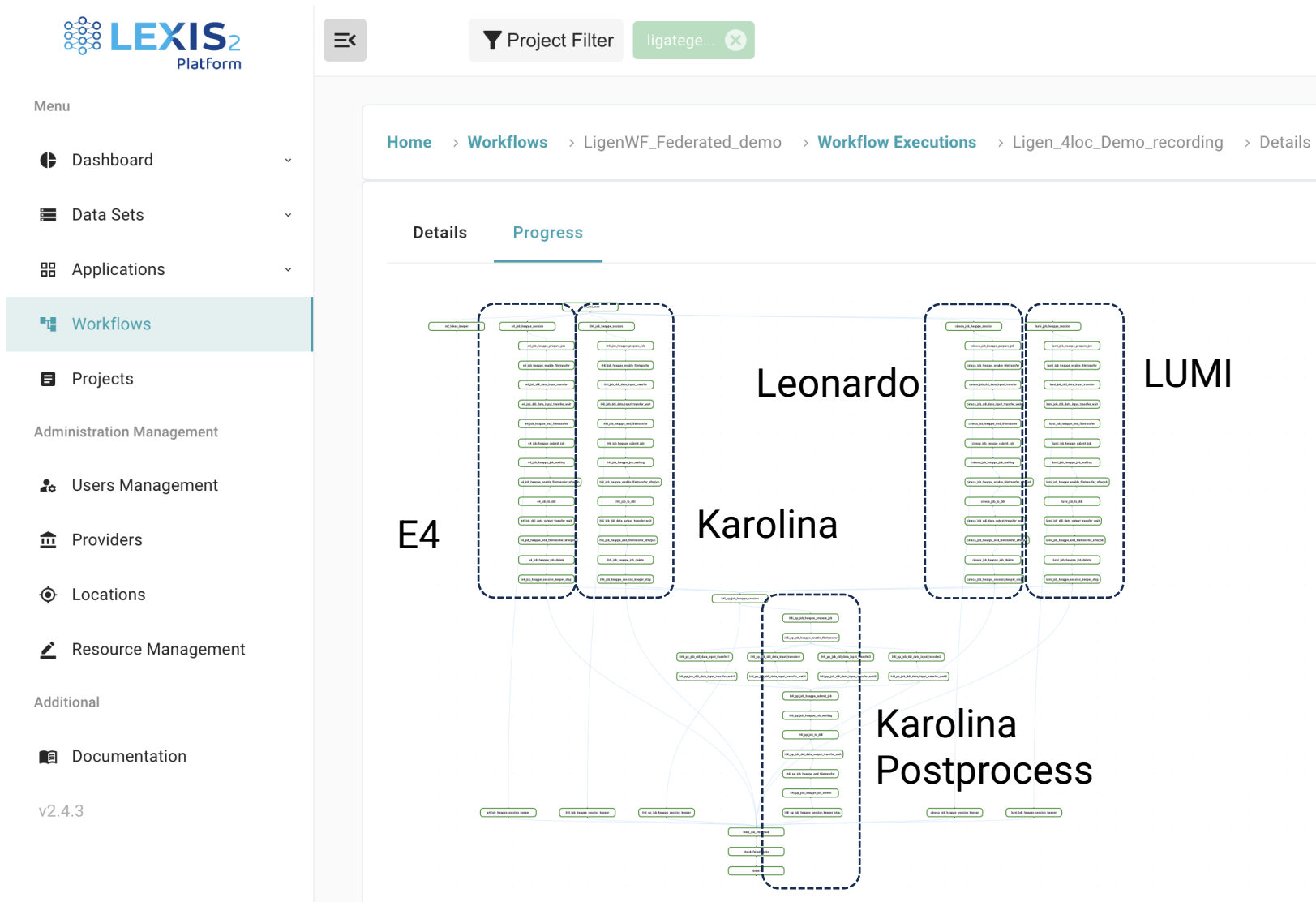
DDI

HEAppE

AAI

UserOrg

Federated execution on 4 HPC clusters at once



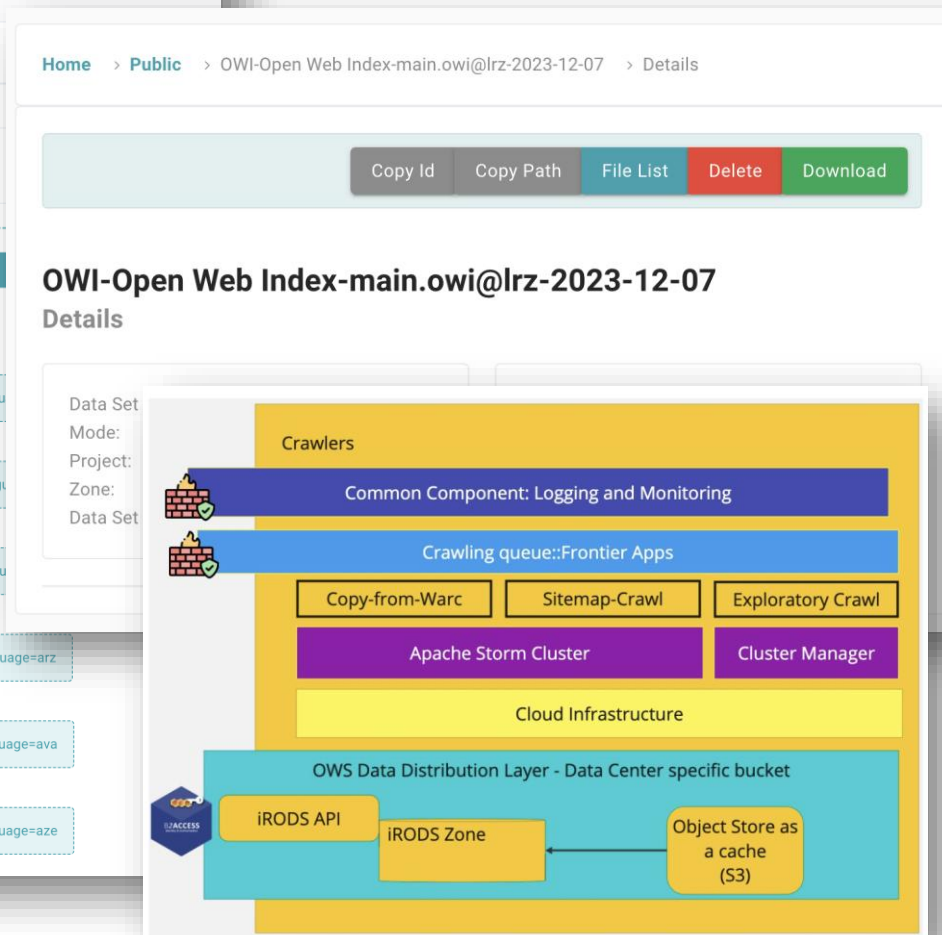
Executed on 4 locations

- LUMI
- Leonardo (CINECA)
- Karolina (IT4I)
- E4 AMD Cluster

OpenWebSearch.eu - Public web indexes



A screenshot of the LEXIS2 Platform web interface. The browser address bar shows the URL: portal.beta.lexis.tech/publicDataSets/e2f40f27-9e47-4464-a7a8-447eb3d7d388/details/fileList. The interface includes a sidebar menu with options like Dashboard, Data Sets, Uploads, Workflows, Public, Applications, Workflows, Projects, and Administration Management. The main content area shows a breadcrumb trail: Home > Public > OWI-Open Web Index-main.owi@lrz-2023-12-07 > Details > File List. Below this, there's a folder structure: OWI-Open Web Index-Main.Owi@Lrz-2023-12-07 > 2023 > 12 > 07. A 'Refresh' button is present. The file list displays various folders named 'language=...' with language codes like aar, abc, abk, afr, aii, aka, amh, amz, ang, arc, arg, ary, ase, asm, ast, ave, aym, azb, and aze.



FLOREON+

- Uses LEXIS workflows for simulations and data processing
- Developed since 2006
- Decision support system for crisis management domain

Hydrology

- Monitoring of hydrological situation and flood forecasting

Traffic

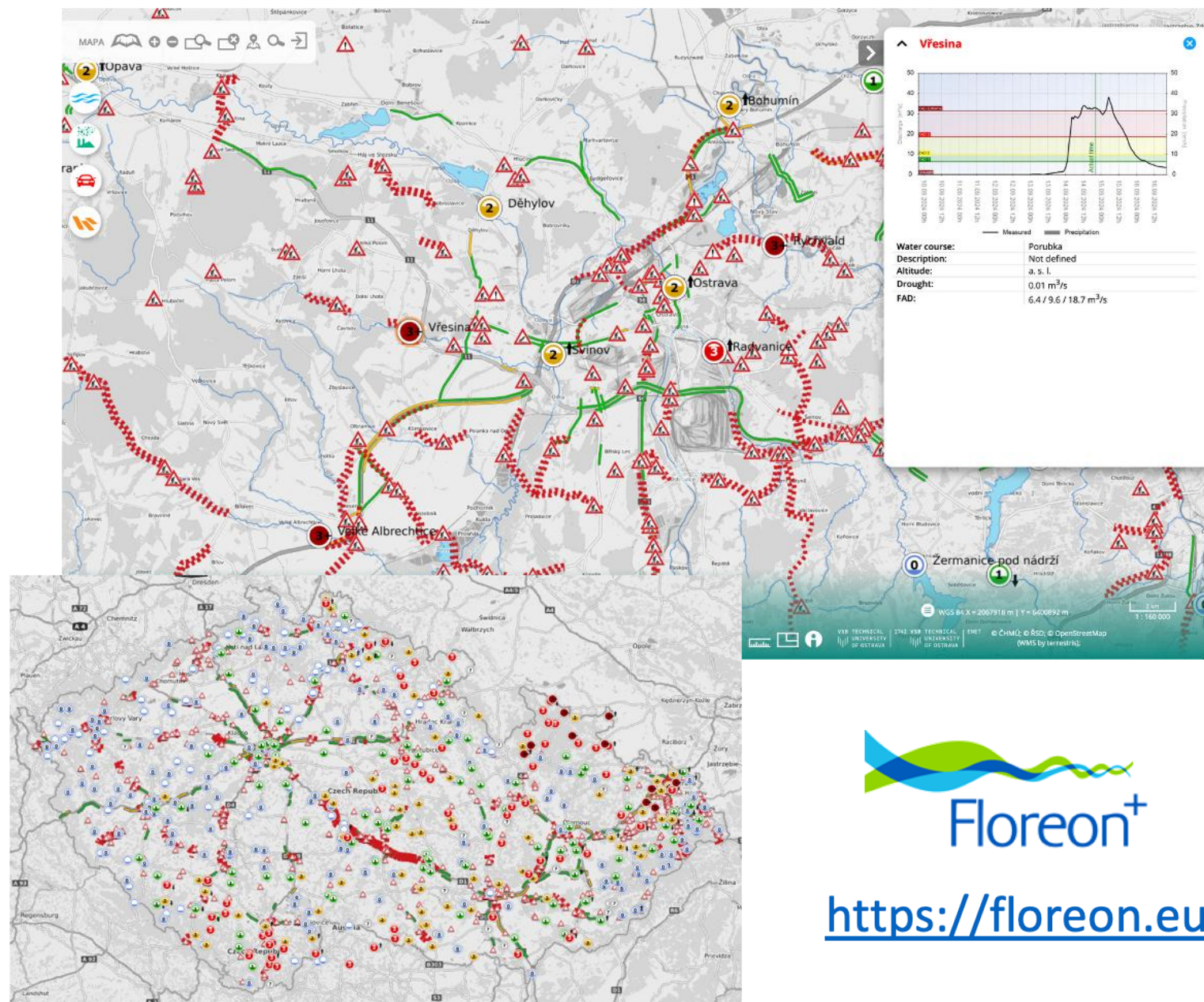
- Monitoring and modelling of traffic situation

Air quality

- Monitoring of air quality and simulation of the spread of hazardous substances

Landslides and subsidence

- Tracking of elevation point movements



<https://floreon.eu>

Hydrology and traffic interaction



LEXIS Platform

Containers & Scripts

portal.beta.lexis.tech/containers

Project Filterligatege...ehpcsw24

Martin Golasowski

Home > Containers

Ehpcsw24Ligategeneral

Search by title

RefreshCreate Container

Test Container7th Of Mar 2024

Quantum AI Test Container21st Of Mar 2024

Python Animals AI Example12th Of Mar 2024

1

Lexisportal.beta.lexis.tech/locations

Project Filterligatege...europde...

Martin Golasowski

Home > Locations

Edit ProviderCreate Location

Search 15 rows

NAME	DESCRIPTION	ENDPOINT	TYPE	PROVIDERDESCRIPTION	ACTIONS
iRODS LRZ	iRODS at LRZ	lexis-lrzicat.srv.lrz.de	Storage	Leibniz-Rechenzentrum Compute Centre	Edit
Barbora	HPC Cluster Barbora	barbora.it4i.cz	HPC	IT4Innovations Computing Centre	Edit
iRODS LRZ OWSeu	iRODS Zone OWSLRZZONE	sikplrzs-ows-icat.srv.mwn.de	Storage	Leibniz-Rechenzentrum Compute Centre	Edit
Cirrus	HPC Cluster Cirrus	login.cirrus.ac.uk	HPC	EPCC Compute Centre	Edit
Leonardo	HPC Cluster Leonardo	login.leonardo.cineca.it	HPC	CINECA Computing Centre	Edit
LocalCluster	Tartu testing instance	https://heappe-1.cloud.ut.ee	HPC	Puhuri testing provider	Edit
ZYC-2	HPC Cluster Alveo	10.12.0.11	HPC	IBM Computing Centre	Edit
LRZ Linux cluster 2	HPC Linux Cluster CoolMUC-2	lxlogin4.lrz.de	HPC	Leibniz-Rechenzentrum Compute Centre	Edit
LUMI	HPC Cluster LUMI	lumi.csc.fi	HPC	LUMI Computing Centre	Edit
IT4I Staging Area	IT4I Staging Area	staging	Storage	IT4Innovations Computing Centre	Edit

LEXIS Platform Containers & Scripts



LEXIS₂
Platform

Menu

Dashboard

Data Sets

Applications

Containers

HPC Applications

User Workflows

Custom HPC Jobs

Workflows

Projects

Administration Management

Users Management

Providers

Locations

Resource Management

Additional

Documentation

v2.4.2

Project Filter

openwebs...

11

Martin Golasowski

Home > Custom HPC Jobs > EESSI Demo - estimate PI using R > Workflow Create

1 Workflow name and project

jobscript.sh

```
1 source /cvmfs/software.eessi.io/versions/2023.06/init/bash
2
3 ml spider R
4
5 ml R/4.2.2-foss-2022b
6
7 R --version
8
9 mkdir ./output
10
11 Rscript -e "
12 # Set the number of points to simulate
13 num_points <- 100000
14
15 # Generate random points
16 x <- runif(num_points)
```

Edit

Name

Compute PI in R

Project Short Name

europdemo

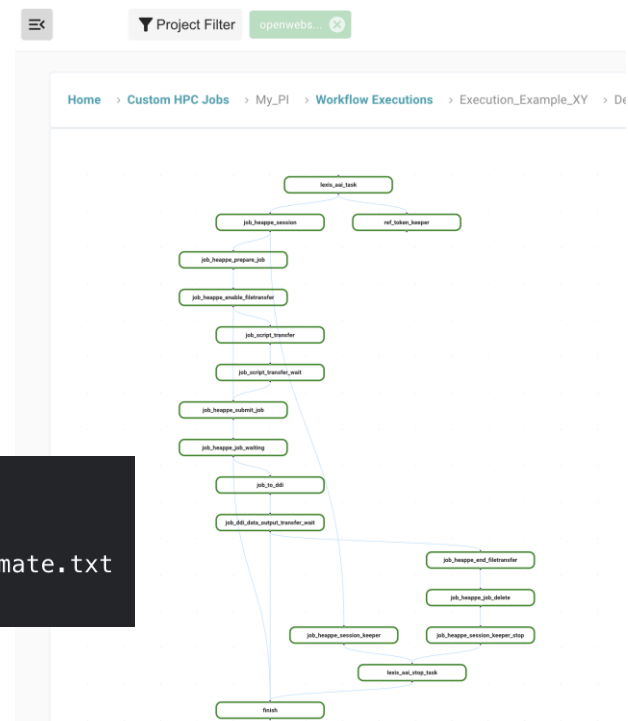
Description

Jobscrip showcase

Continue

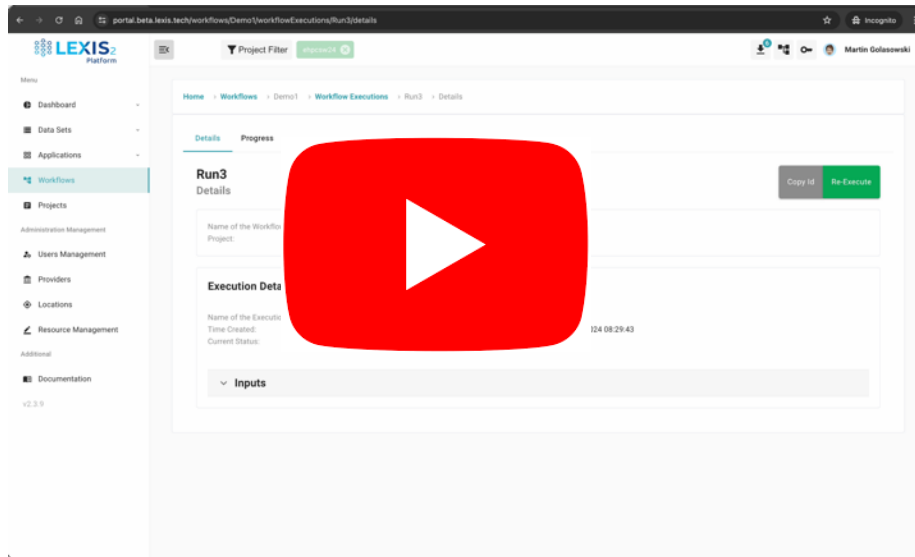
Back

- Run custom R script
- Use common EESSI HPC module
- Workflow & output in GUI



```
188} INFO - https://www.gnu.org/licenses/.
188} INFO - Estimated Pi: 3.1408
188} INFO - Pi estimate has been written to output/pi_estimate.txt
- Success criteria met. Exiting.
```

AI workflow demo with Python in Apptainer



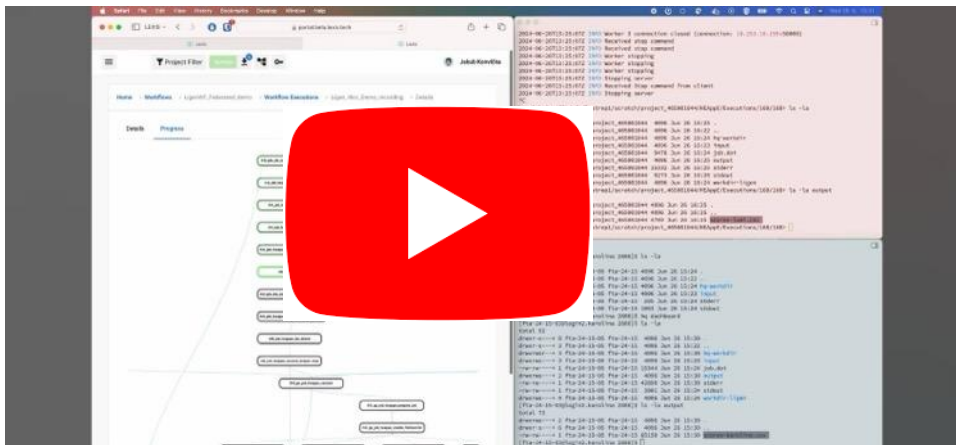
[Check it out!](#)

www.lexis.tech

docs.lexis.tech

portal.lexis.tech

Urgent computing federated workflow execution



[Check it out!](#)

Apache Airflow

Data orchestration tool



Spolufinancováno
Evropskou unií



MUNI
ICS

Registrační číslo IPs EOOSC-CZ
CZ.02.01.01/00/22_004/0007682



Data orchestration

Data **orchestration** is the process of taking siloed data from multiple data storage locations, **combining** and **organizing** it, and making it **available** for data analysis tools. It can also include tasks like provisioning resources and monitoring.

Data Pipelines, Workflows, Jobs

~~CRON + Scripts~~



Apache Airflow

- Started in October 2014 by Maxime Beauchemin at Airbnb
- Apache top-level project since 2019
- **Python** based data orchestrator
- Workflow defined as Python code
- Integrations with many APIs, DBs, services...
- Easy to deploy and use
- Nice web UI
- Complete REST API



Extract step



- Run **long** SQL query
- Stage result to a POSIX file

Transform step



- Run long running HPC job
- Simulation / Analysis

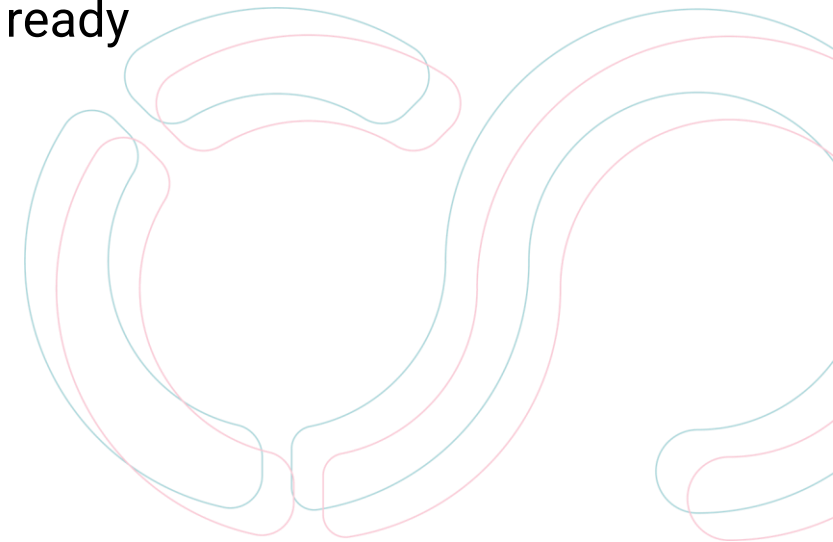
Load step



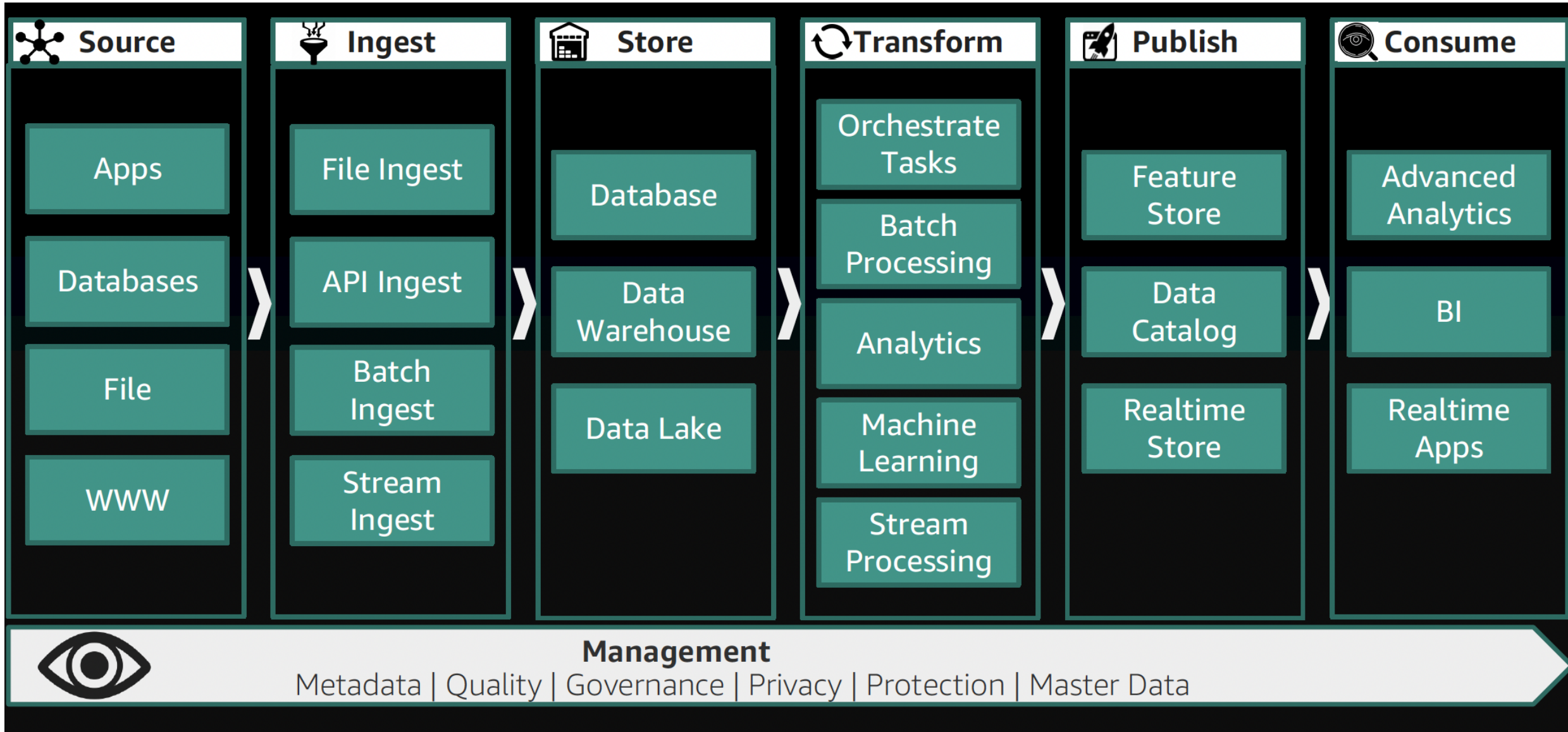
- Stage results back
- Store in SQL
- Create new Dataset

Motivation

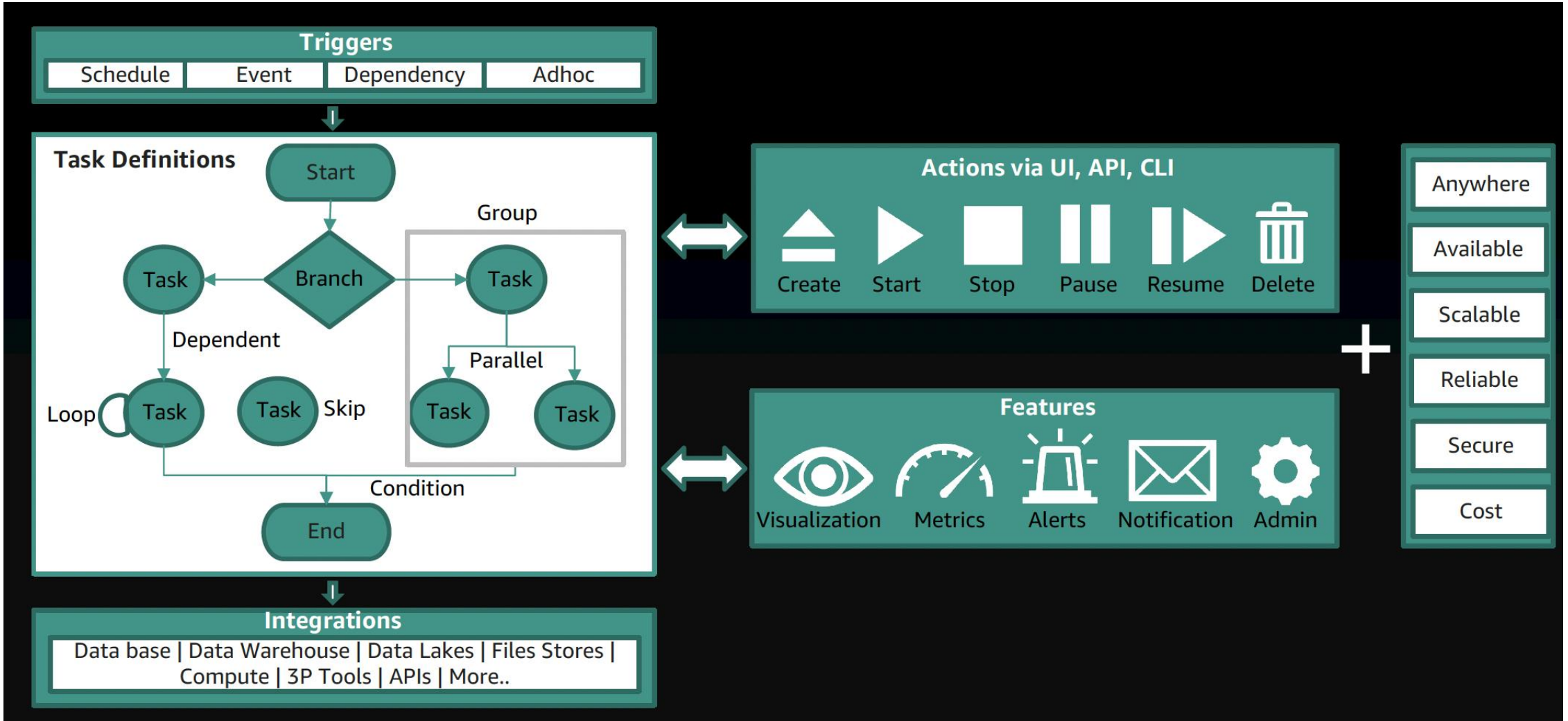
- An online store needs to have a daily report of its customer demography and sold products
 - Multiple live databases (SQL, Elastic..)
 - Proprietary payment platform sends daily report in CSV
 - Boss needs a live dashboard and a nice Power Point for daily meetings
 - Operators will get an e-mail and Slack message when the report is ready
- Daily index of the internet for up-to-date search engine
 - Internet is being crawled daily to TBs of files
 - Data needs to be filtered, URLs identified, normalised...
 - Index has to be uploaded daily to a dedicated server (GBs)



What data pipeline should do...



What the **tool** for the data pipeline should do...



Děkuji za pozornost.

Kontakt: Martin Golasowski <martin.golasowski@vsb.cz>



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Registrační číslo IPs EOSC-CZ
CZ.02.01.01/00/22_004/0007682