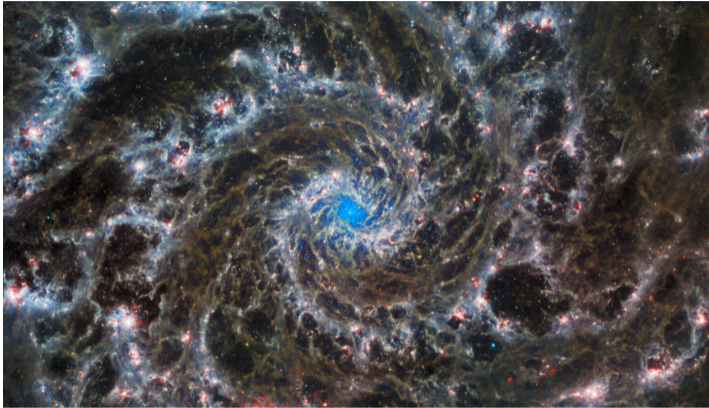


USEGALAXY.CZ

včera, dnes a zítra

**Aleš Křenek, Boris Jurič,
Martin Demko, Miroslav Ruda**

Seminář gridového počítání, Praha, 13.4.2023



ESA/Webb, NASA & CSA, J. Lee and the PHANGS-JWST Team. Acknowledgement: J. Schmidt

Galaxy Using 10.2 MB

Workflow Visualize Shared Data Help User

Tools

Upload Data

- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- Animal Detection on Acoustic Recordings

Tool Exit Code: 0
Job API ID: 4198208dde6a83a6

Job Parameters

Input Parameter	Value
Which tool was used generate logs?	fastqc
Type of FastQC output?	Raw data
FastQC output	<ul style="list-style-type: none"> 5 : FastQC on data 1: RawData 7 : FastQC on data 2: RawData
Report title	Blabla
Custom comment	bubububuj
Use only flat plots (non-interactive images)	False
Output the multiQC plots raw data?	False
Output the multiQC log file?	False

Job Outputs

Tool Outputs	Dataset
MultiQC on : Webpage	9 : MultiQC on data 7 and data 5: Webpage
MultiQC on : Stats	8 : MultiQC on data 7 and data 5: Stats

History

Unnamed history

10.7 MB

- 9 : MultiQC on data 7 and data 5: Webpage
- 8 : MultiQC on data 7 and data 5: Stats
- 7 : FastQC on data 2: RawData
- 6 : FastQC on data 2: Webpage
- 5 : FastQC on data 1: RawData
- 4 : FastQC on data 1: Webpage
- 3 : wildtype.fna
- 2 : mutant_R2.fastq
- 1 : mutant_R1.fastq

Co je to Galaxy

„Galaxy is an open-source platform for FAIR data analysis that enables users to:

- Use tools from various domains (that can be plugged into workflows) through its graphical web interface.
- Run code in interactive environments (RStudio, Jupyter...) along with other tools or workflows.
- Manage data by sharing and publishing results, workflows, and visualizations.
- Ensure reproducibility by capturing the necessary information to repeat and understand data analyses.

The Galaxy Community is actively involved in helping the ecosystem improve and sharing scientific discoveries.”

<http://galaxyproject.org>

Pozitiva a jistoty

- Portfolio dostupných nástrojů
 - de-facto komunitní standard, široká podpora
- Grafické rozhraní
 - „nahraju vstup, zadám parametry, spustím výpočet, stáhnu výsledek“
 - příkazová řádka, použité knihovny atd. jsou běžnému uživateli skryty
 - integrované vizualizační a analytické nástroje
- Provenance dat (F.A.I.**R**.ness)
 - u každého výsledku je zaznamenáno, z jakých vstupů a jakým nástrojem (včetně verze) byl získán

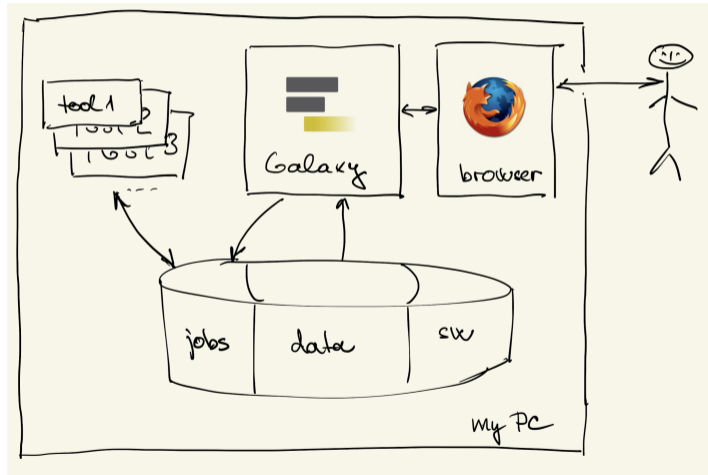
Dostupné instalace

- vlastní
- Galaxy Main (usegalaxy.org, usegalaxy.eu, usegalaxy.org.au)
 - široké portfolio nástrojů, komunitní podpora
 - kvota 250 GB, 6 současně běžících úloh (1 CPU, 8 GB RAM)
- specializované:
 - CZ: Repeat Explorer, UMSA
 - <https://galaxyproject.org/use/>

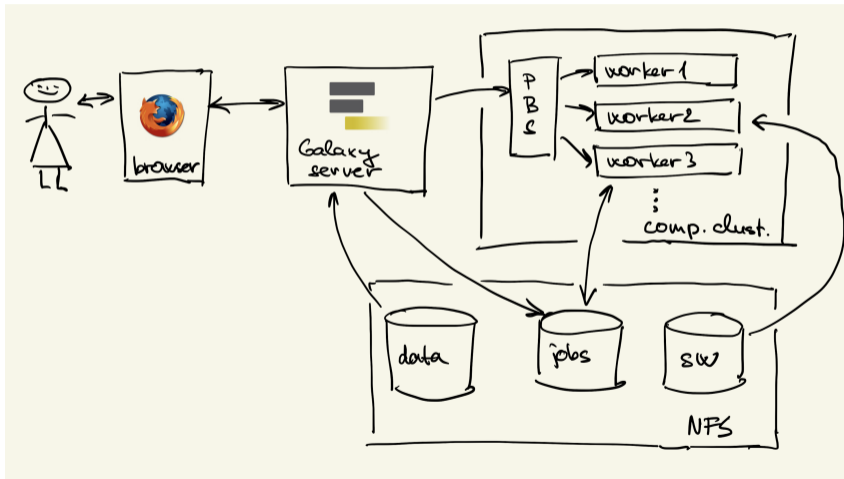
Dostupné instalace

- vlastní
- Galaxy Main (usegalaxy.org, usegalaxy.eu, usegalaxy.org.au)
 - široké portfolio nástrojů, komunitní podpora
 - kvota 250 GB, 6 současně běžících úloh (1 CPU, 8 GB RAM)
- specializované:
 - CZ: Repeat Explorer, UMSA
 - <https://galaxyproject.org/use/>
- **usegalaxy.cz**
 - prostředí co nejpodobnější Galaxy Main
 - přístup ke zdrojům e-Infra CZ (pro jejich legitimní uživatele)

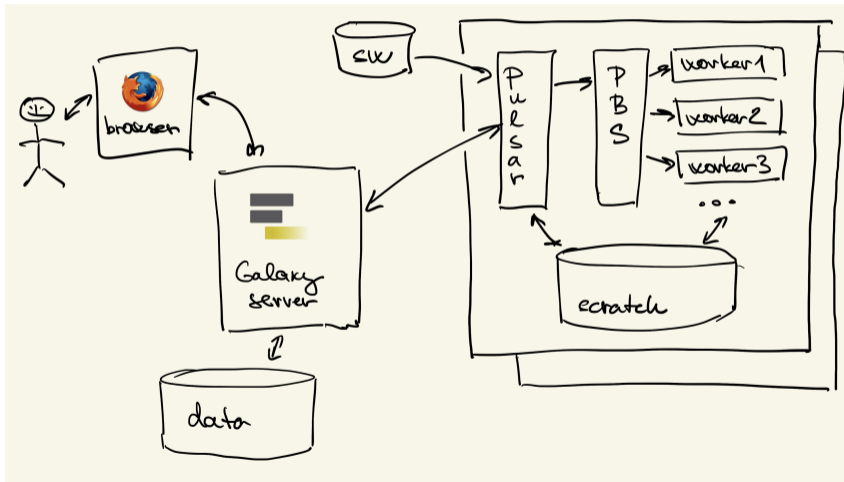
Architektura - vlastní instalace



Architektura – dedikovaný cluster



Architektura - Pulsar



galaxy.metacentrum.cz → usegalaxy.cz

- Stávající instalace
 - galaxy.metacentrum.cz, repeatexplorer-elixir.cerit-sc.cz
 - hybridní architektura „dedikovaný cluster“ bez dedikovaného clusteru
 - na hranici výkonových možností
- Nové instalace
 - usegalaxy.cz, nový repeatexplorer
 - důsledně architektura Pulsaru
 - větší průchodnost a stabilita
 - flexibilita konfigurace

- Clustery Elmo, Elwe (cca. 3000 jader CPU)
- Pilotní přístup na stroje s GPU (AlphaFold)
- Není definitivní, bude se doladovat podle potřeby

A co kontejnery?

- TL;DR: Ano

A co kontejnery?

- TL;DR: Ano
- Galaxy dovoluje nástroje spouštět z uživatelských kontejnerů
 - Docker, Singularity
- Pulsar (v naší aktuální konfiguraci) spouští nástroje v Singularity důsledně

Projekt EuroScienceGateway

- <https://galaxyproject.org/projects/esg/>
 - Accessible e-Infrastructure resources...
 - Support the varieties of analysis types and diverse usage patterns...
 - The application of FAIR principles to workflows...
 - Adoption ... by researchers in diverse scientific disciplines
- Pulsar jako základní stavební prvek celoevropské infrastruktury
- Výhledově přímé napojení na usegalaxy.eu

Pilotní provoz

- usegalaxy.cz je připravena k použití
 - přecházíme z provizorního jména usegalaxy.cerit-sc.cz
- Sada nástrojů kopíruje Galaxy Main (usegalaxy.eu)
- Autentizace uživatelů identitami LifeScience RI
 - diskutujeme přechod na silnější autorizaci e-Infra CZ
- dostatek výpočetní a úložné kapacity
 - využití monitorujeme, podle potřeby a možností přidáváme

Pilotní provoz

- usegalaxy.cz je připravena k použití
 - přecházíme z provizorního jména usegalaxy.cerit-sc.cz
- Sada nástrojů kopíruje Galaxy Main (usegalaxy.eu)
- Autentizace uživatelů identitami LifeScience RI
 - diskutujeme přechod na silnější autorizaci e-Infra CZ
- dostatek výpočetní a úložné kapacity
 - využití monitorujeme, podle potřeby a možností přidáváme
- **Vítáme odvážné pilotní uživatele**
 - jen s vaší pomocí službu vyladíme
 - obraťte se na regalaxy@rt.cesnet.cz