

V6 basics

What is vantage6

Vantage6 is an open-source software framework for privacy enhancing analysis and decentralized data analysis



Jointly organized by:

Why using vantage6



Open
source with
Apache-2.0
licence



Container
orchestration
for PETs



Extensible
to different
data source
types



Algorithms
in any
language



Other
applications
can connect
using the
API



Graphical
User
Interface



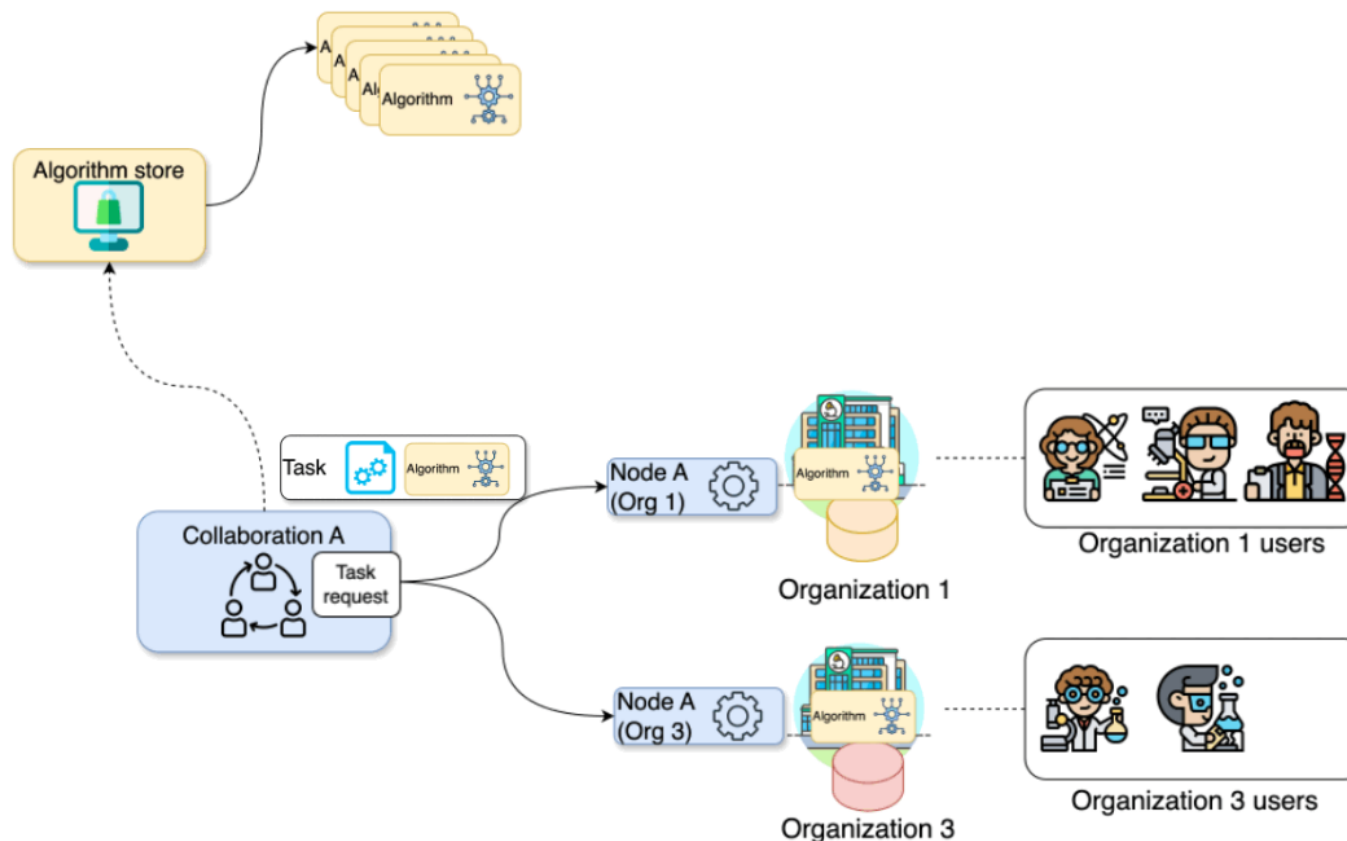
Managing
collaboration
policies



Minimal
network
requirements

Federated analysis projects with vantage6

vantage6 encompasses a project administration system that allows the user to manage permissions and access to the resources, while assuring the protection of the data.

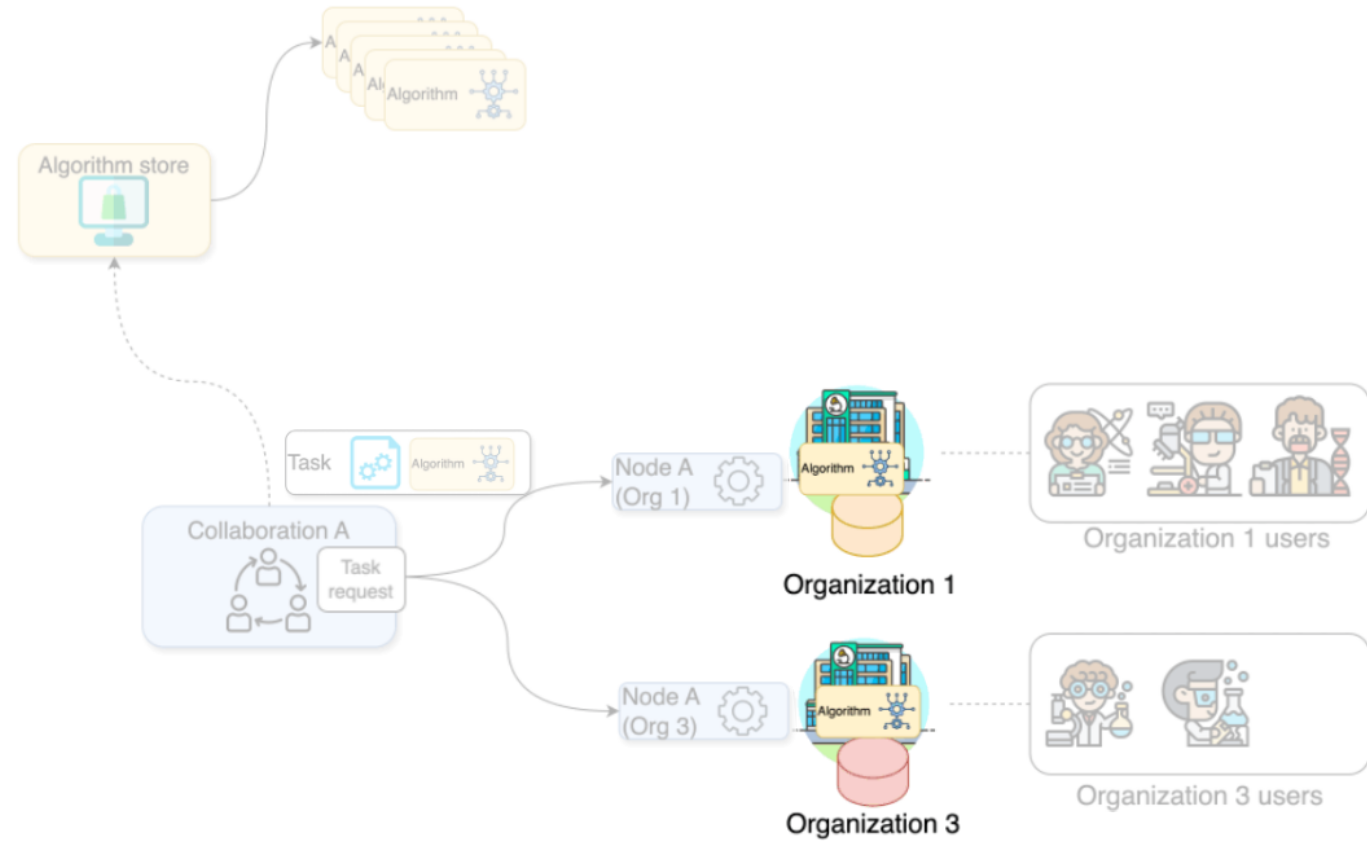


Jointly organized by:

Federated analysis projects with vantage6

Organization

A group of users that share a common goal or interest (e.g., a consortium, an institute, etc.).

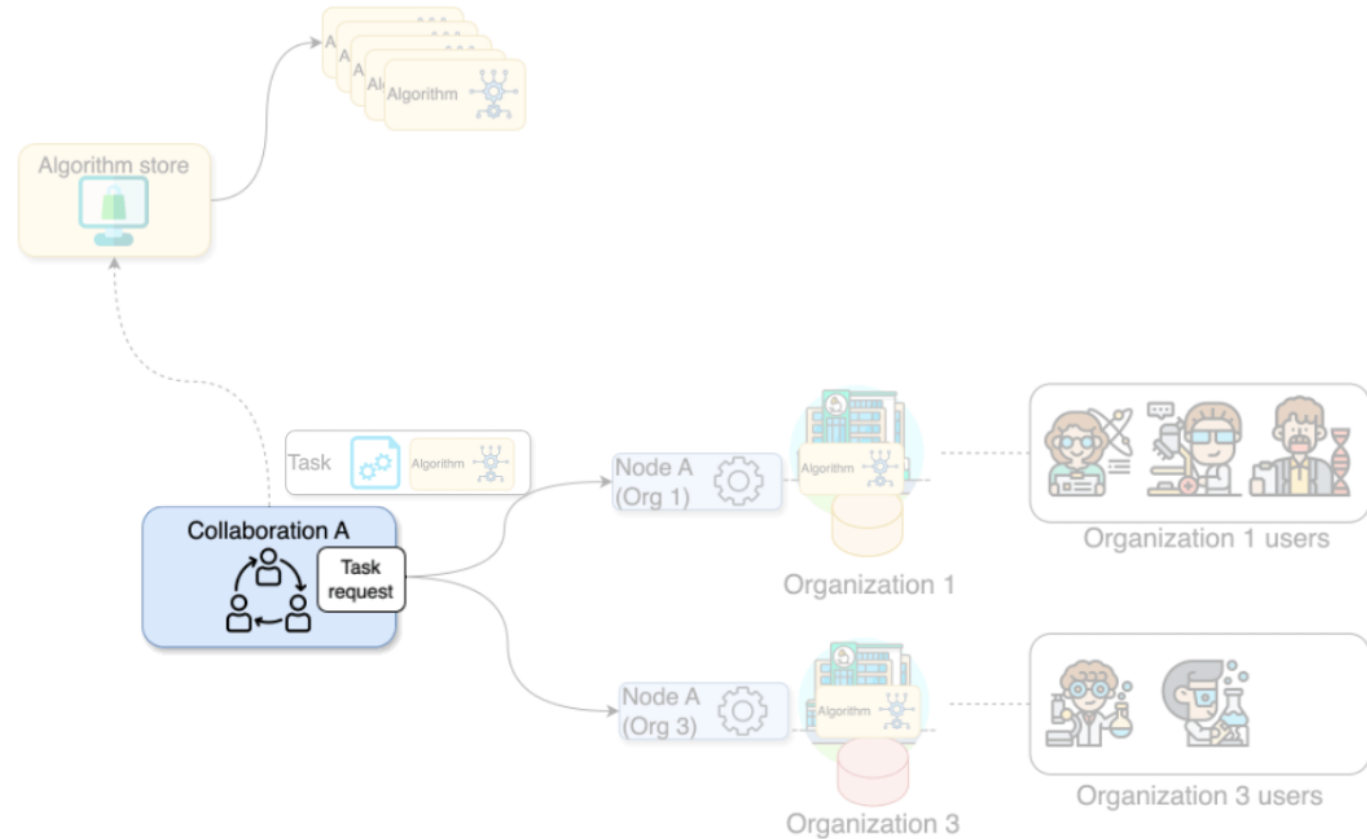


Jointly organized by:

Federated analysis projects with vantage6

Collaboration

One or more organizations working together towards a shared objective.

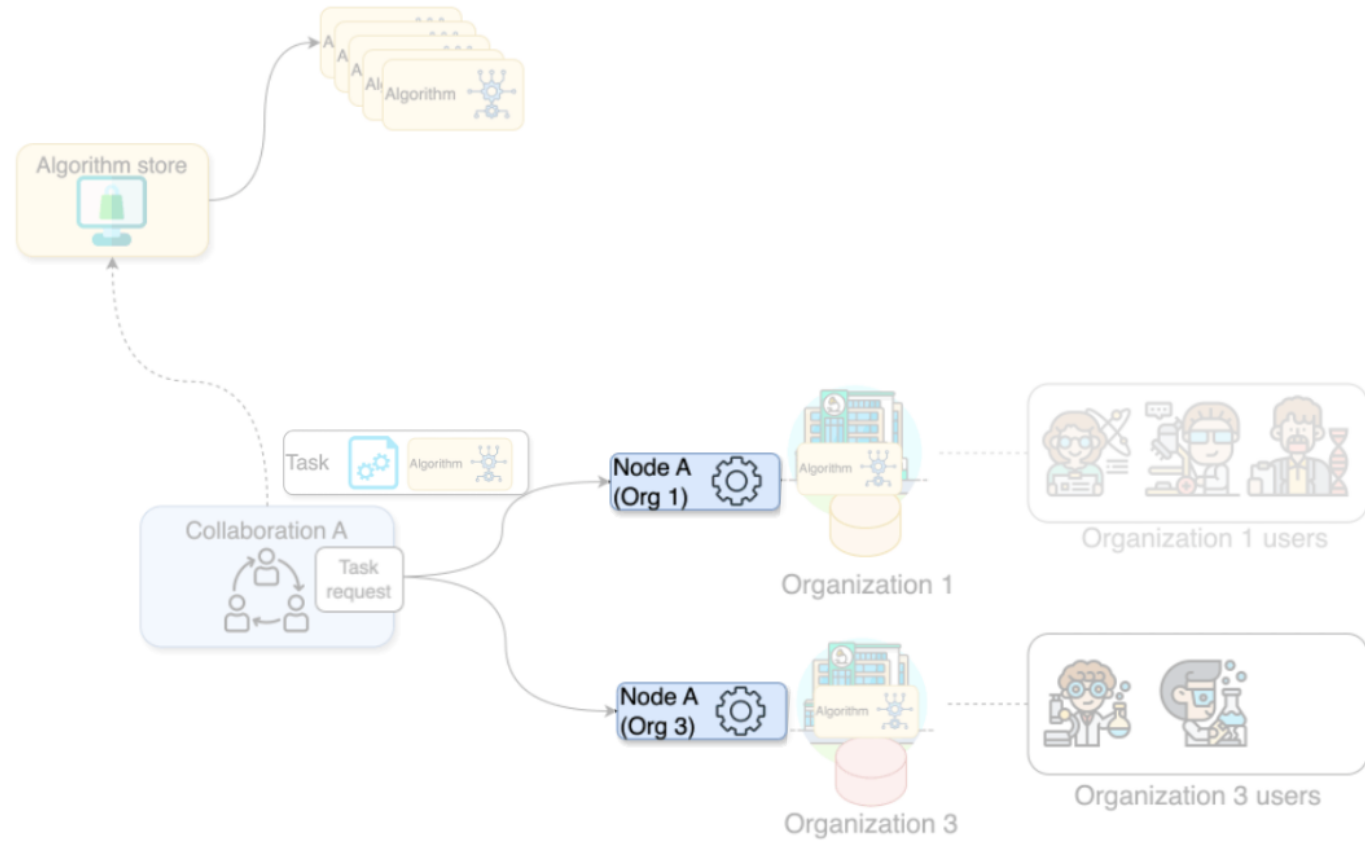


Jointly organized by:

Federated analysis projects with vantage6

Node

Component that accesses the organization data and executes algorithms on it.

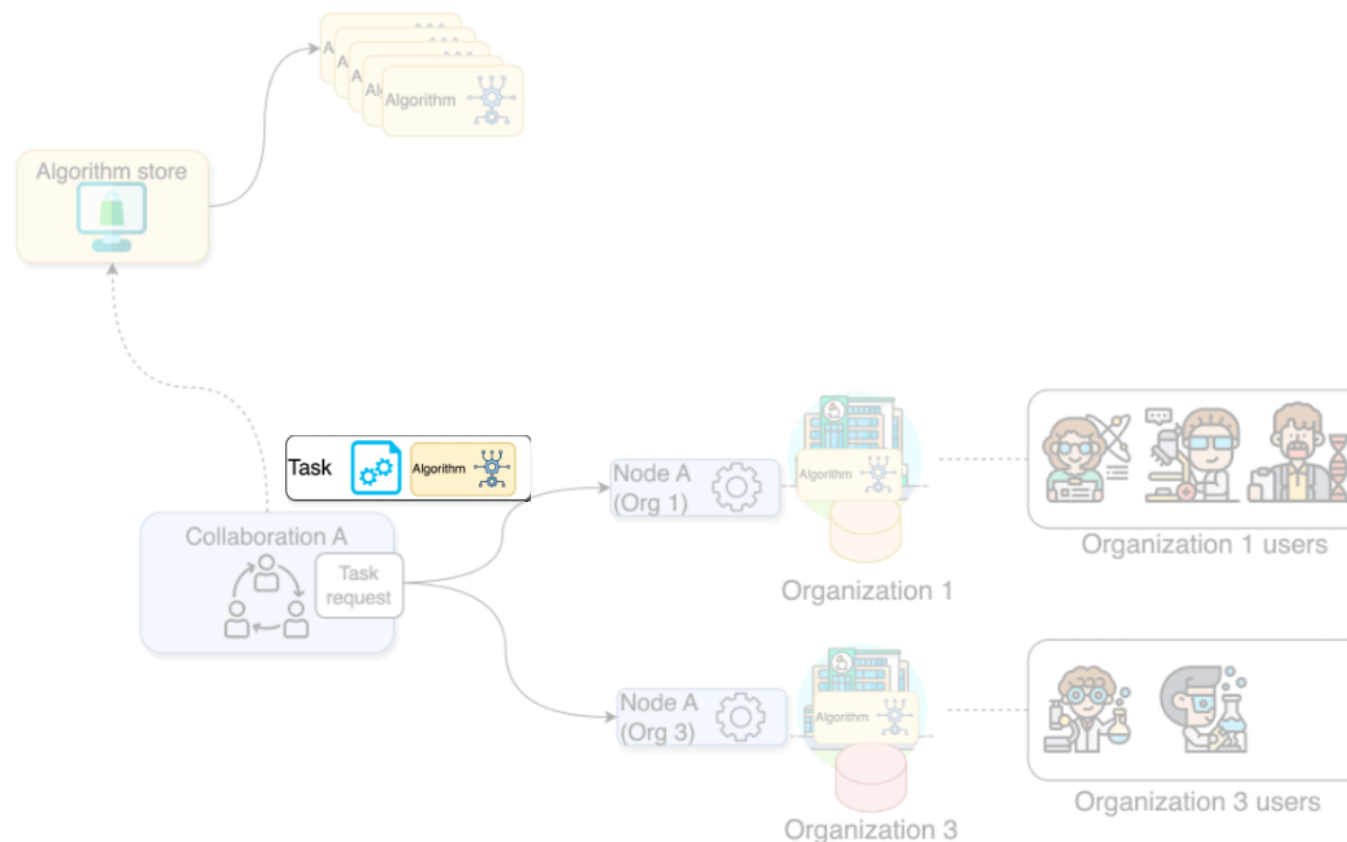


Jointly organized by:

Federated analysis projects with vantage6

Task

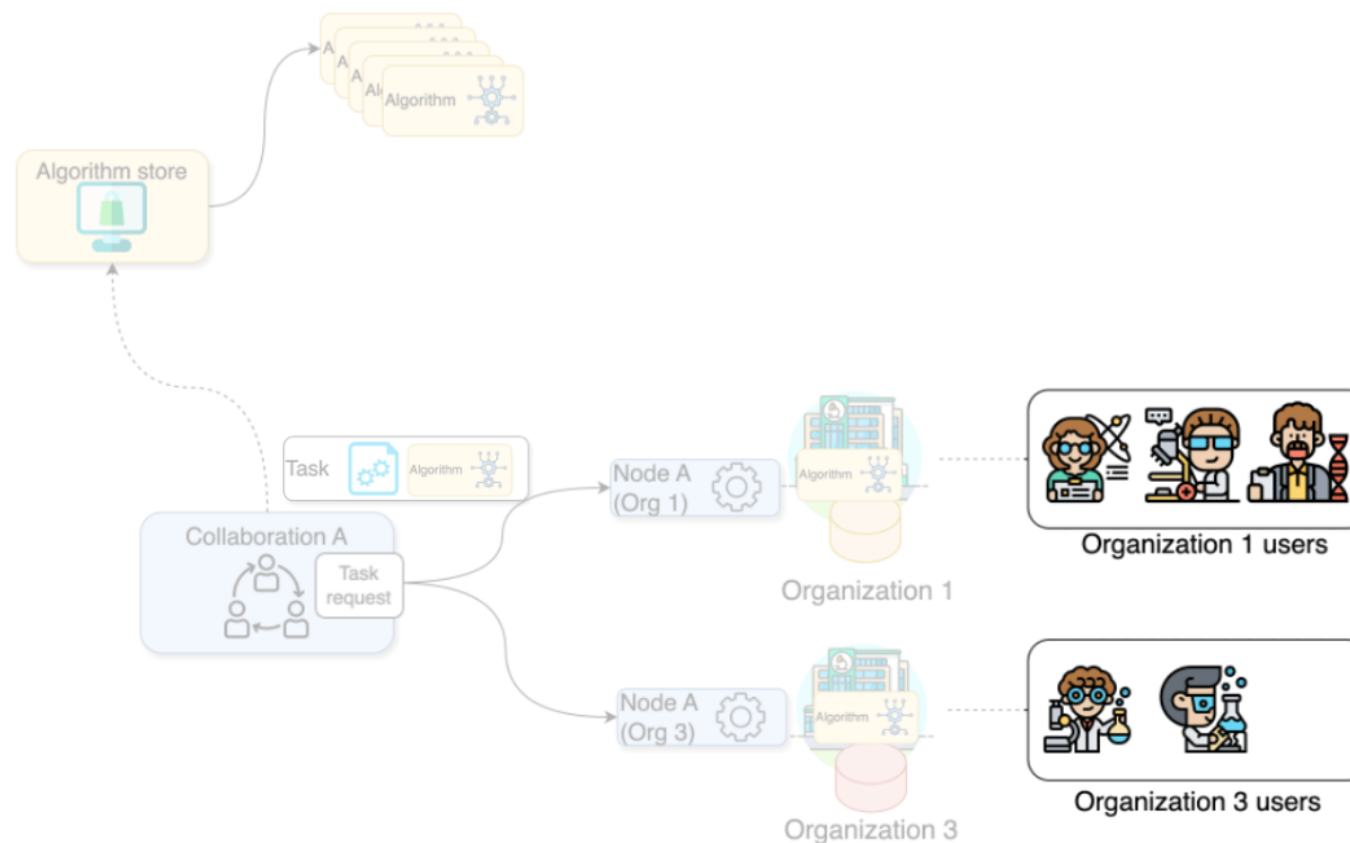
A request for the execution of an algorithm. It is handled by the corresponding organizations' nodes.



Federated analysis projects with vantage6

User

A person that belongs to one organization who can create tasks for one or more organizations within a collaboration.

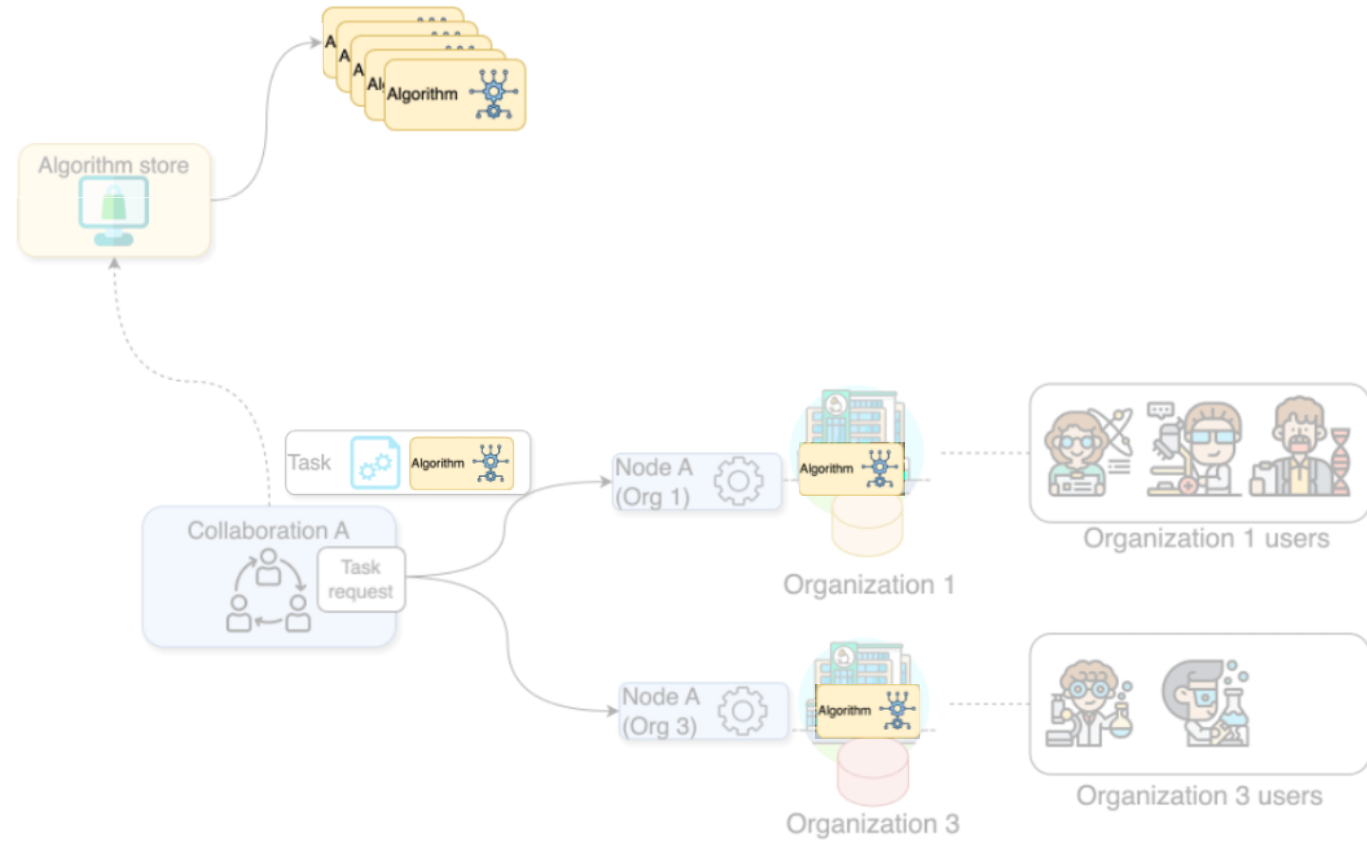


Jointly organized by:

Federated analysis projects with vantage6

Algorithm

A computational model or process which can be securely distributed to nodes for execution.

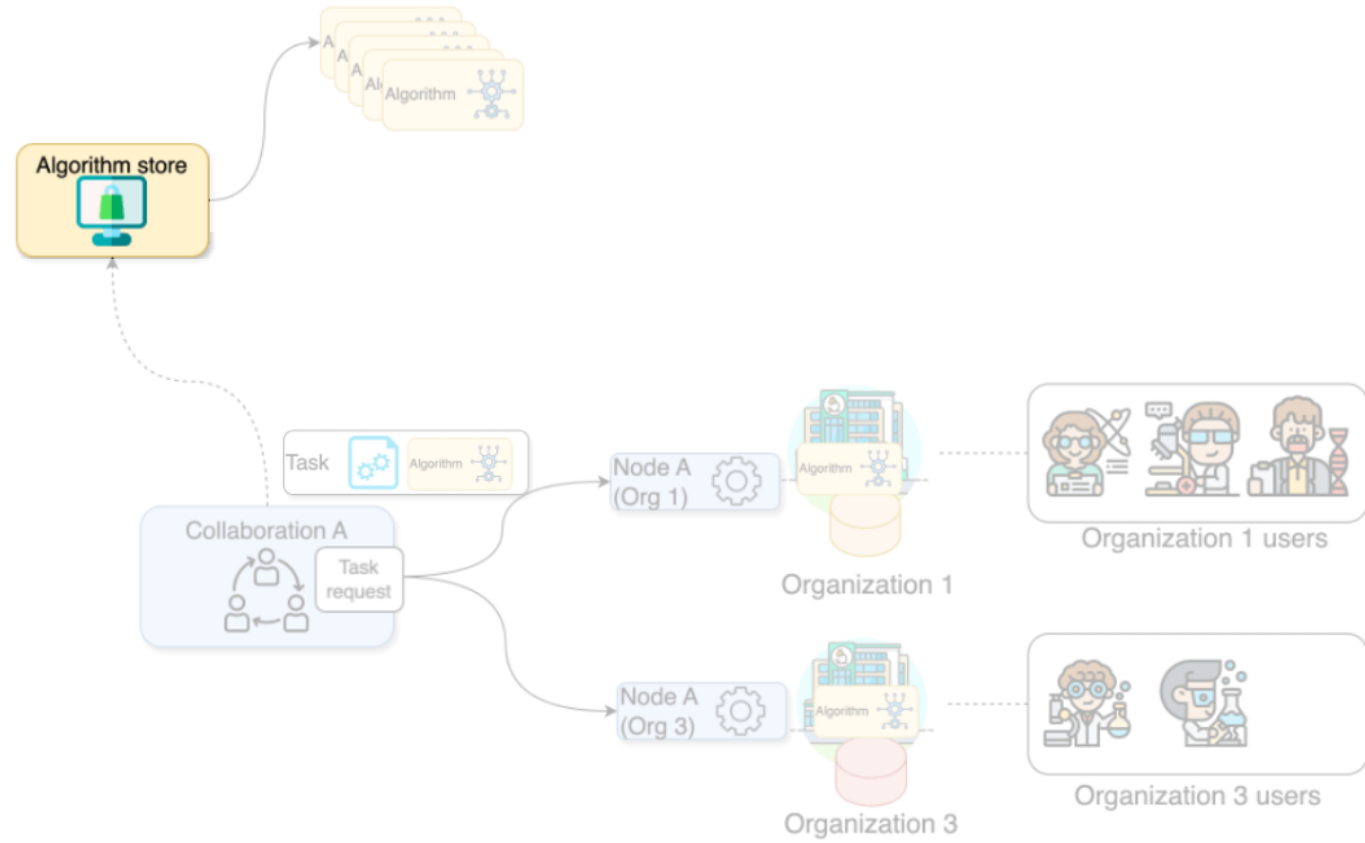


Jointly organized by:

Federated analysis projects with vantage6

Algorithm store

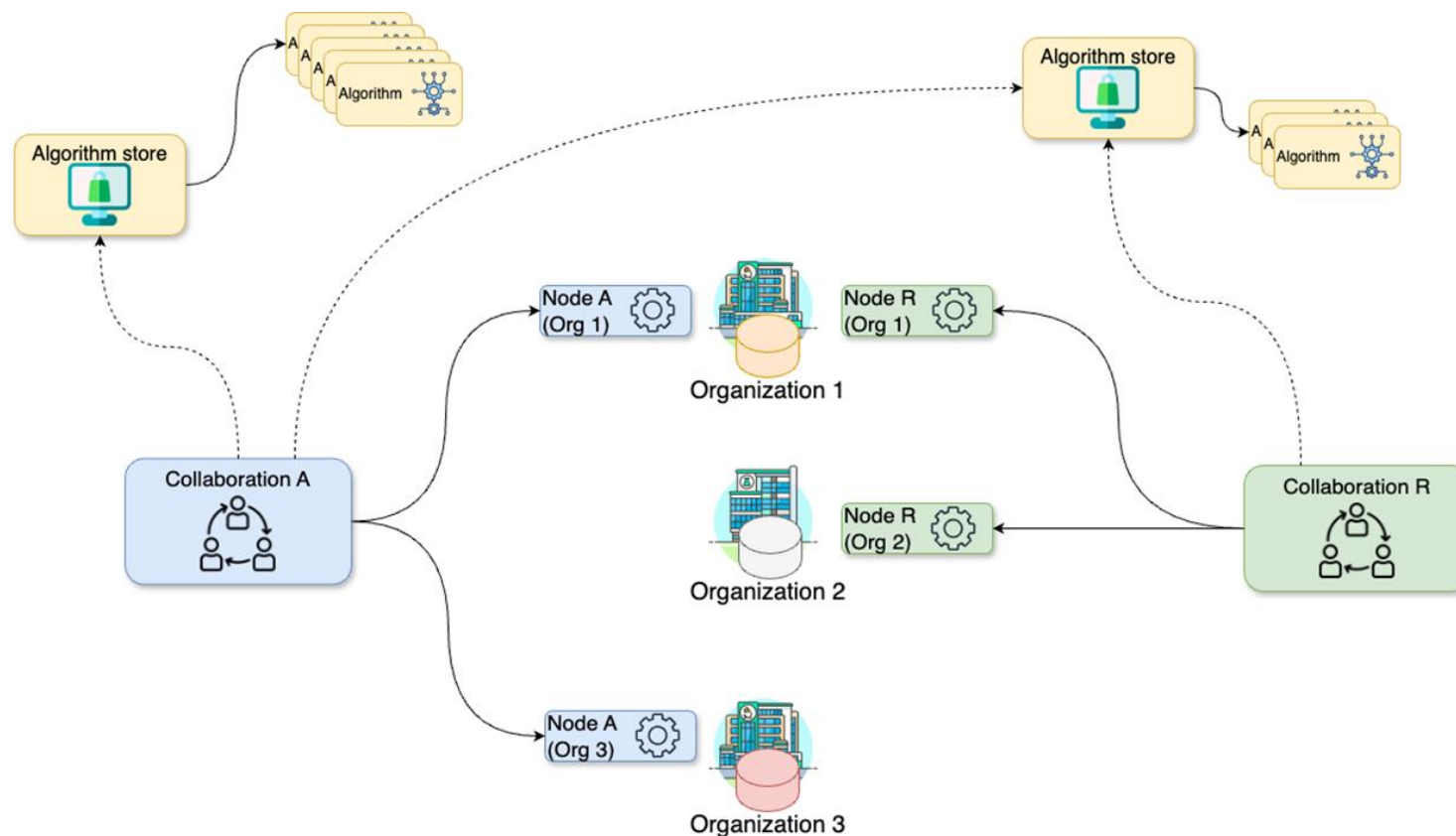
A repository for trusted algorithms within a certain project.



Jointly organized by:

Federated analysis projects with vantage6

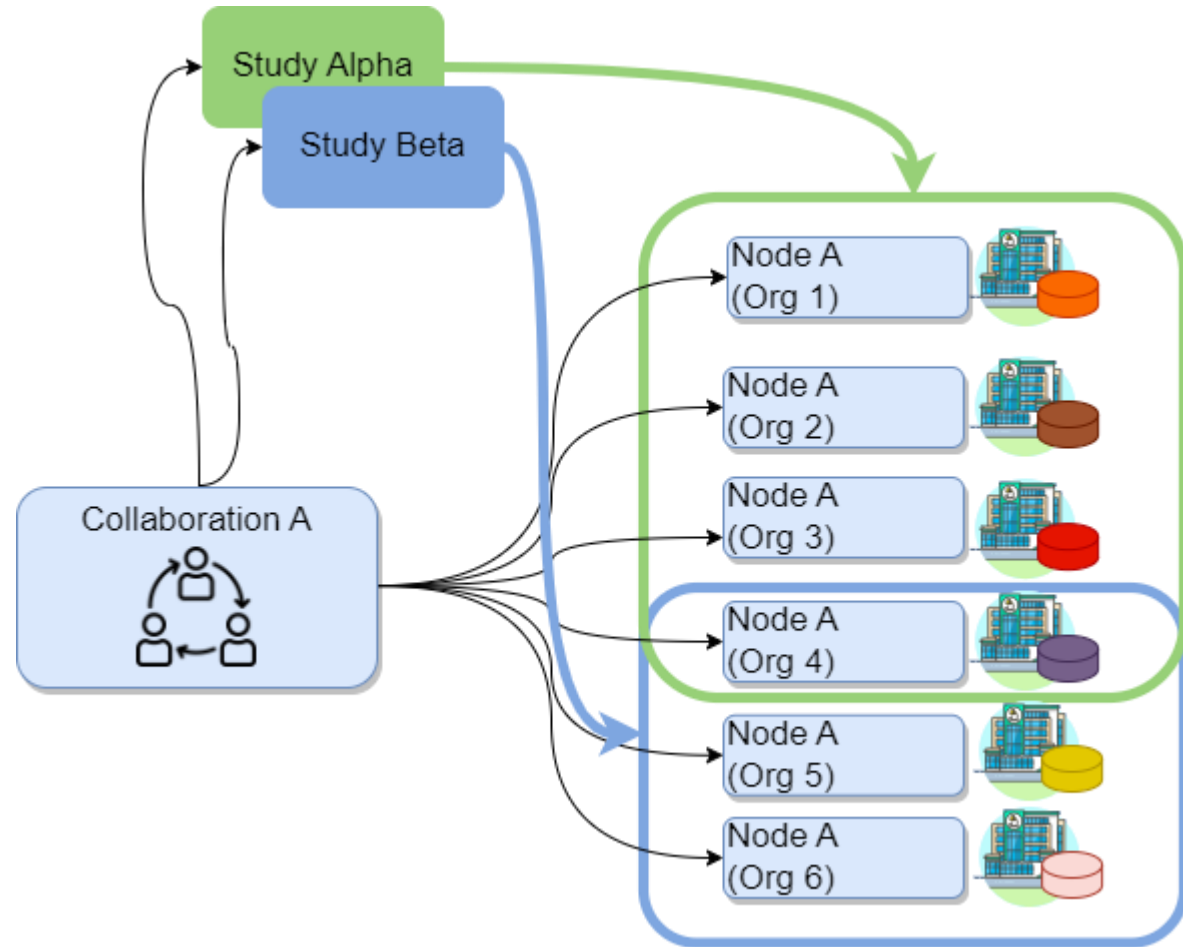
Expanding the scenario:
Organizations can take part to multiple collaborations



Federated analysis projects with vantage6

Study

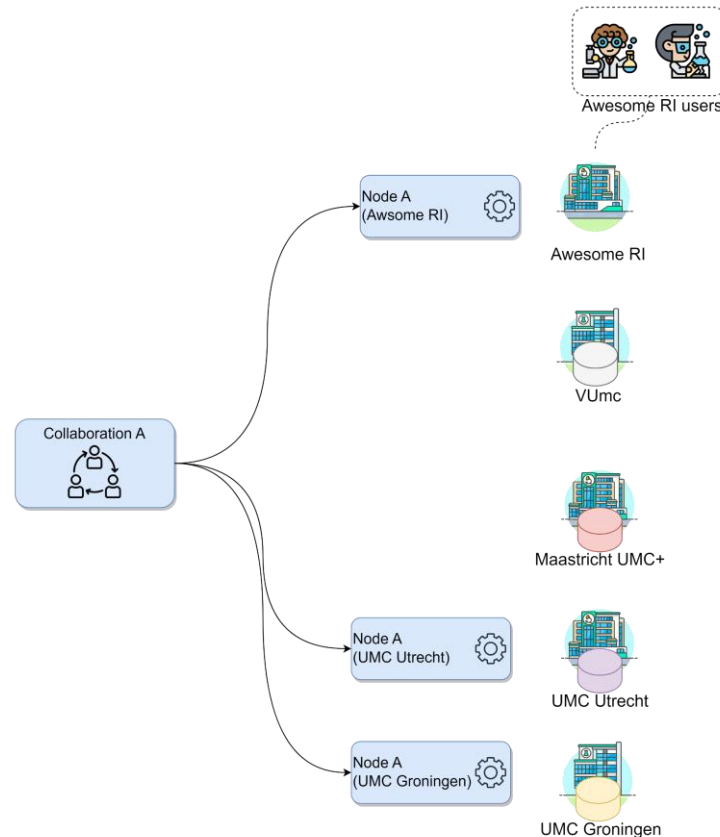
A subset of organizations within a collaboration that are engaged in a specific research question.



Jointly organized by:

Design your own project structure

⚡ CHALLENGE



You, a researcher at the Awesome RI, want to conduct a study across three academic hospitals : VUmc in Amsterdam, Maastricht UMC+, and UMC Utrecht. Consider the following:

- Your research institute has an existing collaboration with UMC Utrecht and UMC Groningen.
- You will be conducting this study with a colleague from your institute. Both of you are already registered on the organization.

How would the concepts described above map to your potential use case?

- 1. Which organizations will you need to add to your collaboration?**
- 2. How many new nodes would you need to set up and on which organizations?**
- 3. How many users would be created?**

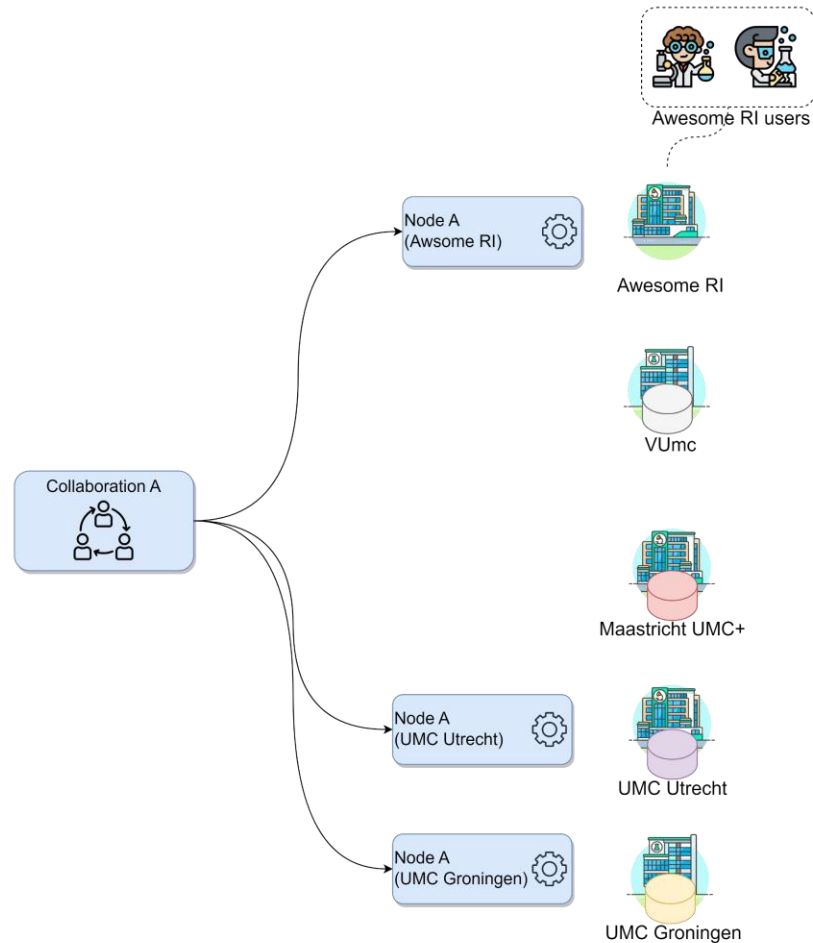
Jointly organized by:

Solution

⚡ SOLUTION

1. Which organizations will you need to add to your collaboration?

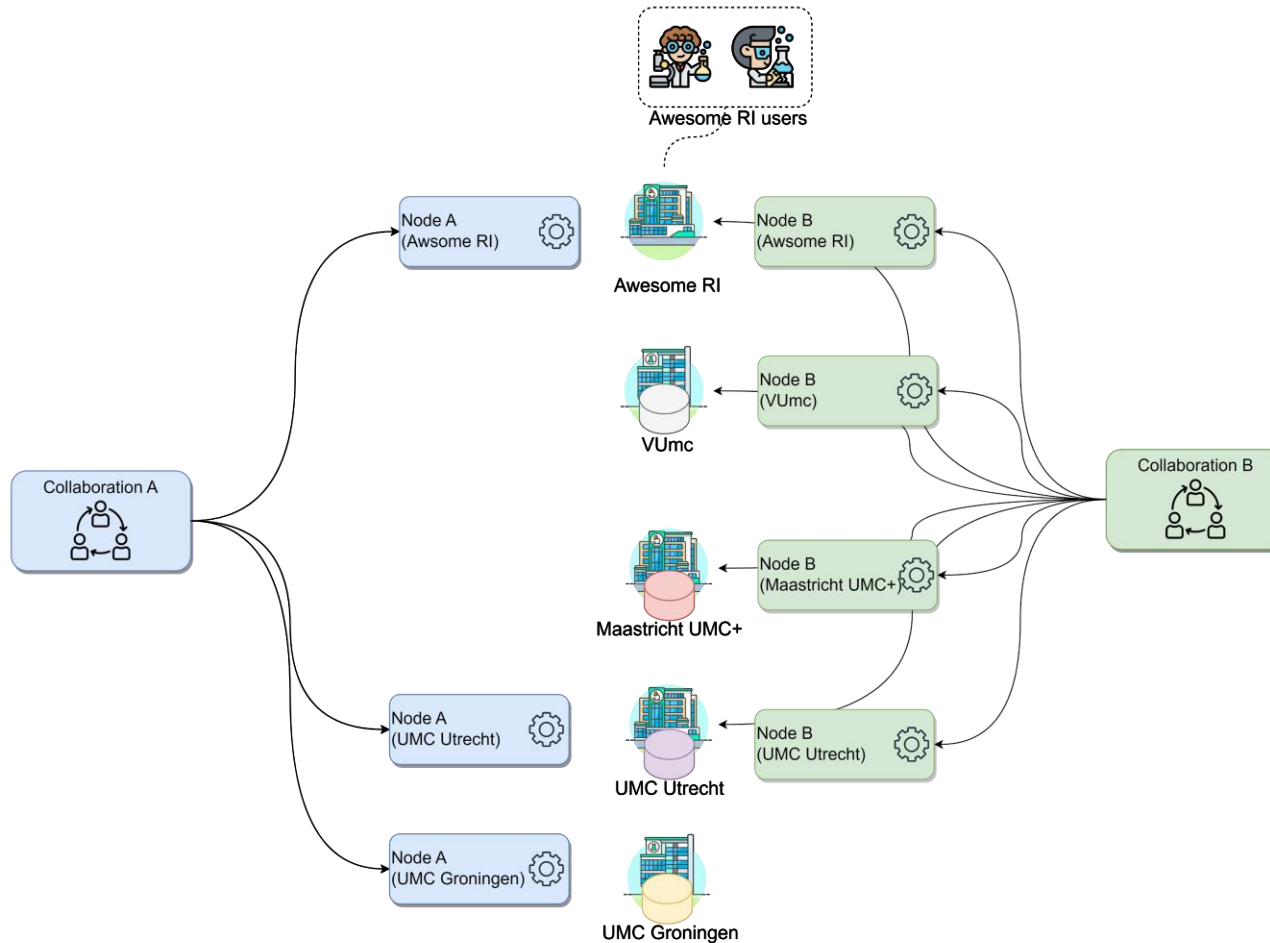
All the academic hospitals as well as your own organization: VUmc, Maastricht UMC+, UMC Utrecht and your research institute.



Jointly organized by:

Solution

⚡ SOLUTION



1. Which organizations will you need to add to your collaboration?

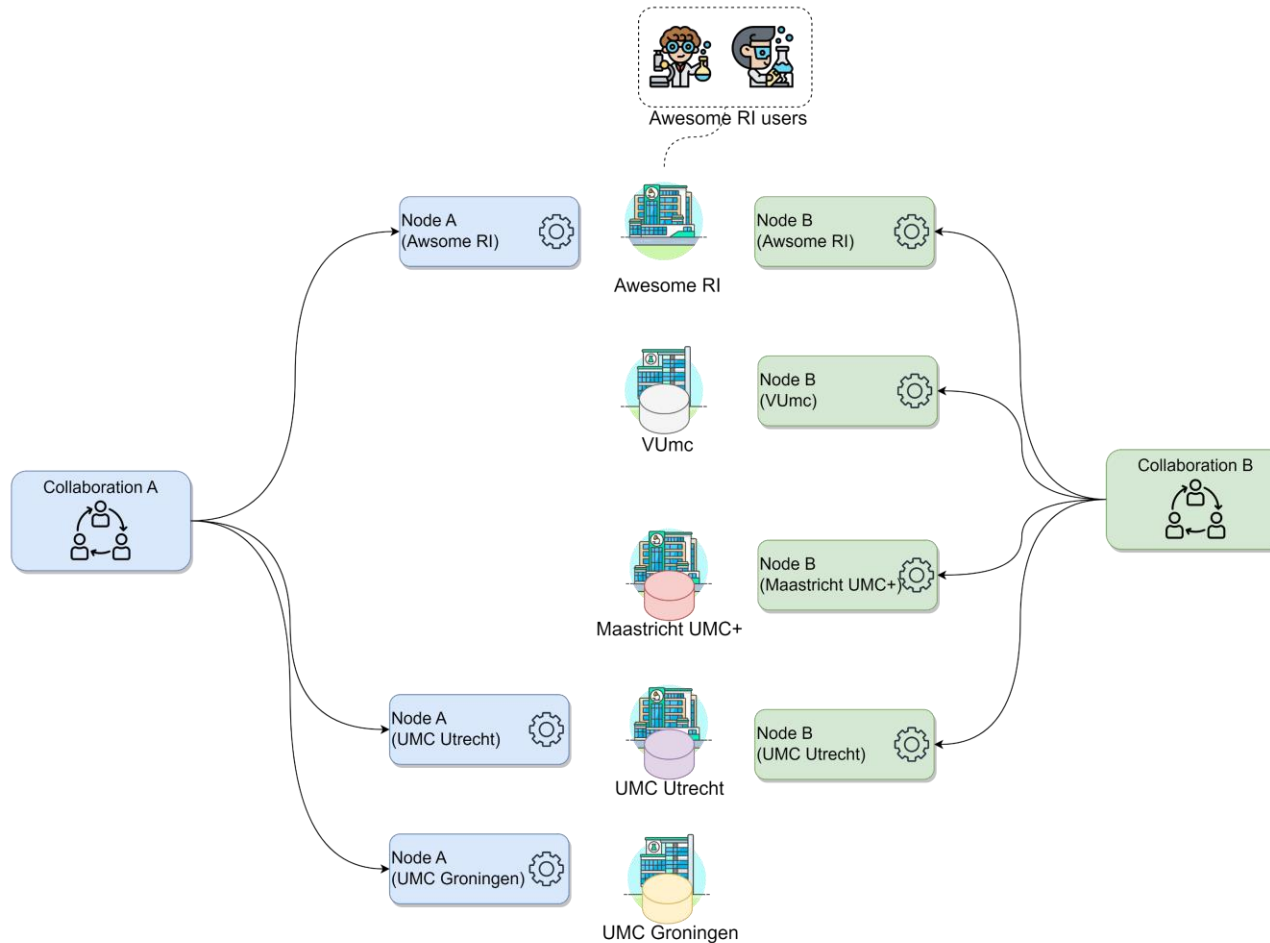
All the academic hospitals as well as your own organization: VUmc, Maastricht UMC+, UMC Utrecht and your research institute.

2. How many new nodes would you need to set up and on which organizations?

One node for every academic organization in the collaboration, so 4.

Solution

⚡ SOLUTION



1. Which organizations will you need to add to your collaboration?

All the academic hospitals as well as your own organization: VUmc, Maastricht UMC+, UMC Utrecht and your research institute.

2. How many new nodes would you need to set up and on which organizations?

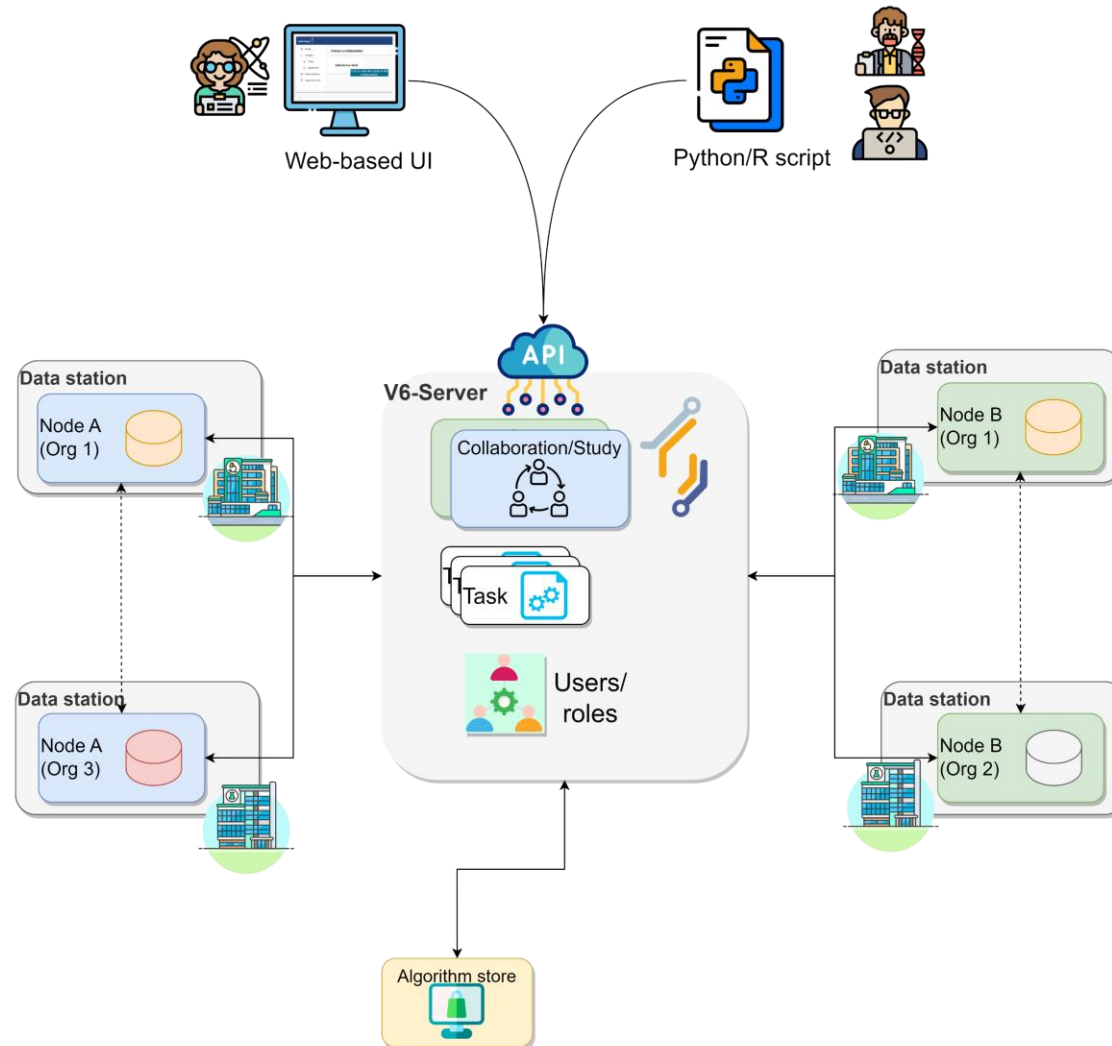
One node for every academic organization in the collaboration, so 4.

3. How many users would be created?

There is no need to create new users, as these are already registered on the organization.

Jointly organized by:

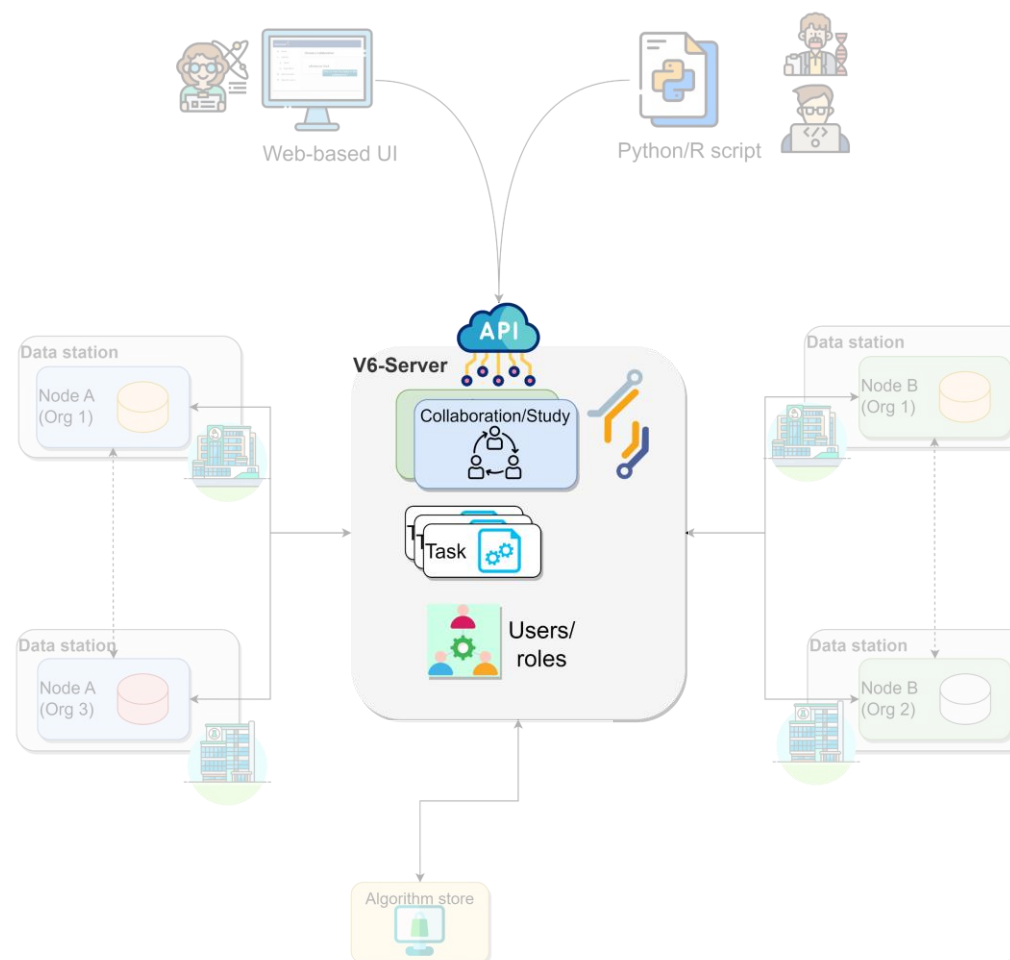
The vantage6 infrastructure



The vantage6 infrastructure

Server

Communication hub between clients and nodes. It also handles authentication and authorization to the system.

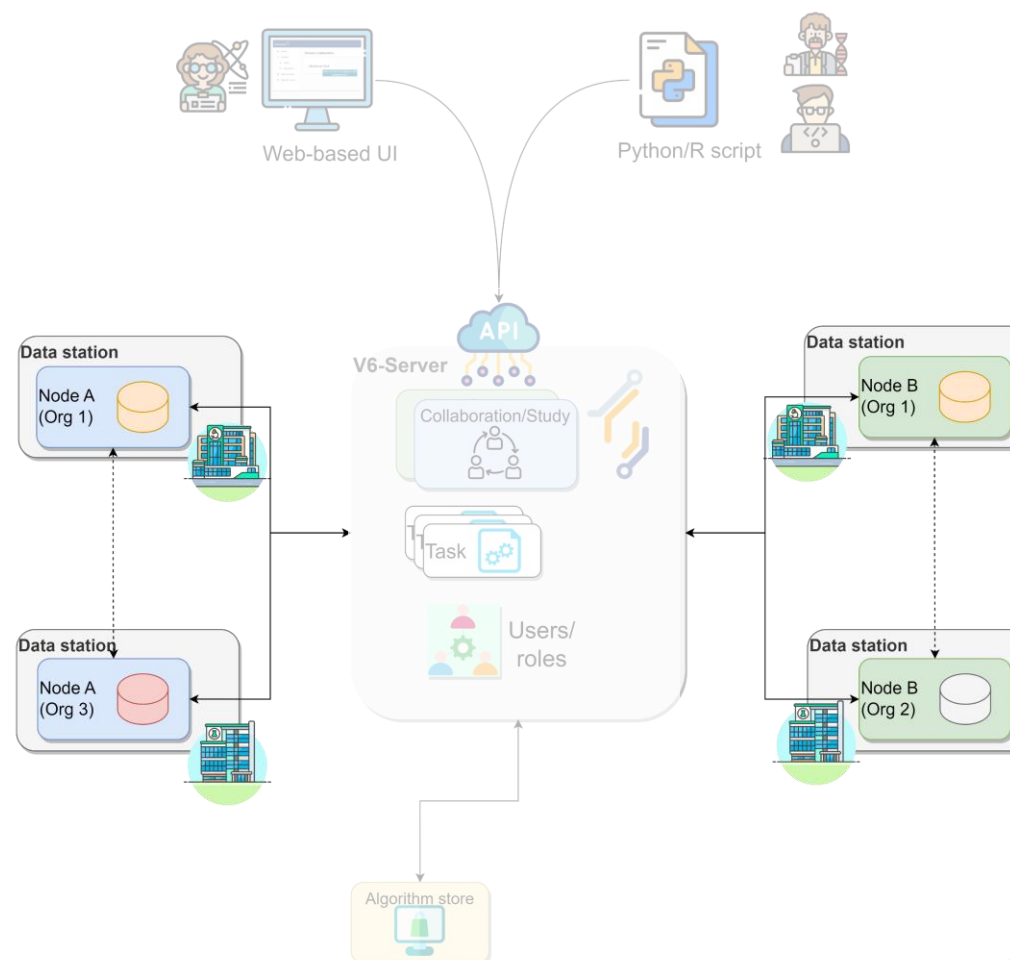


Jointly organized by:

The vantage6 infrastructure

Data station

Hosts the data and the vantage6 node. The vantage6 node executes the allowed algorithms on the local data and sends the results back to the server.



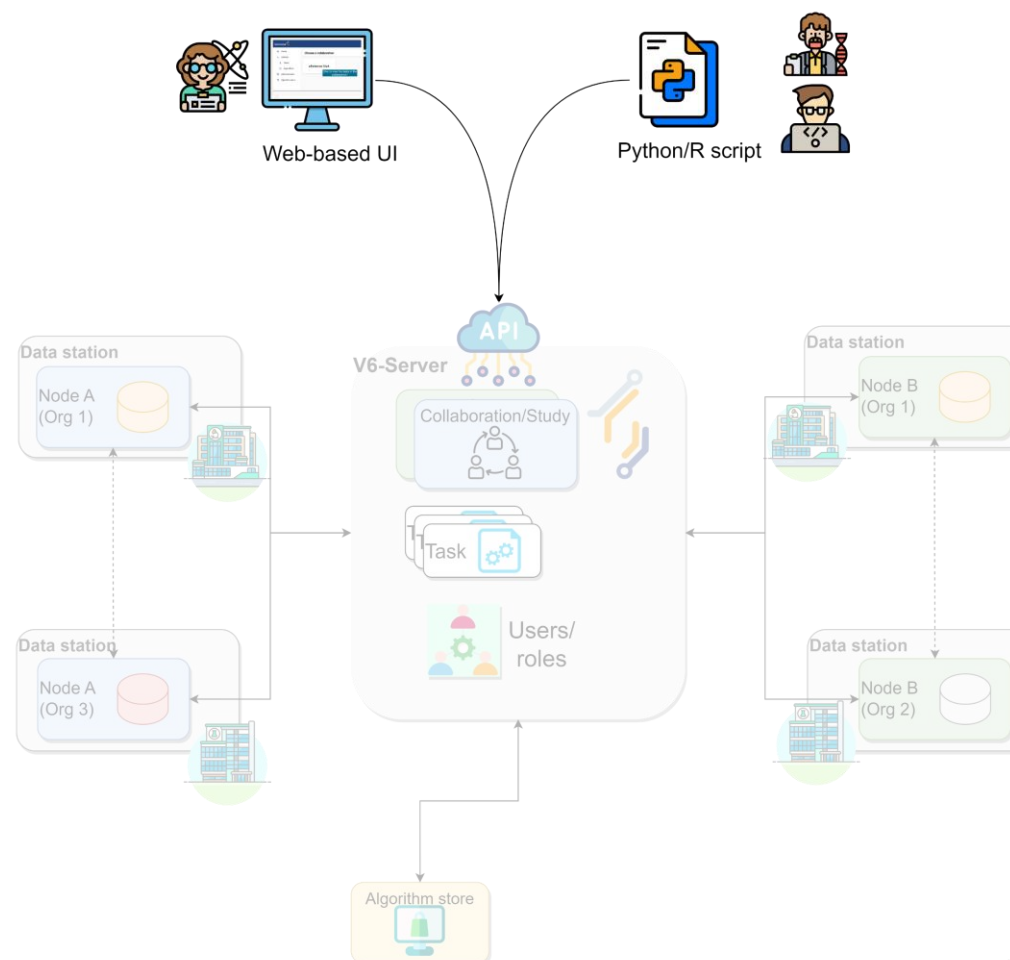
Jointly organized by:

The vantage6 infrastructure

Client

Entity that interacts with the vantage6 server via:

- *API*
- *User interface*
- *Python client*

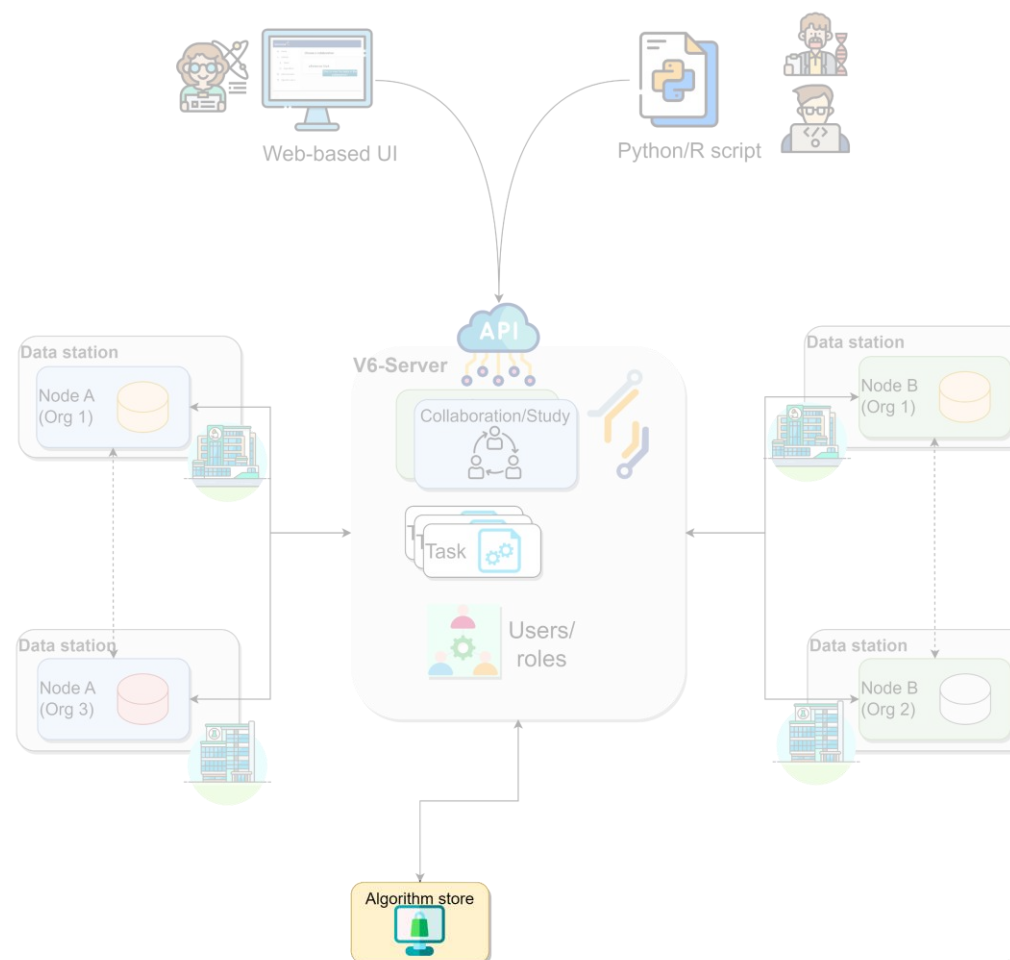


Jointly organized by:

The vantage6 infrastructure

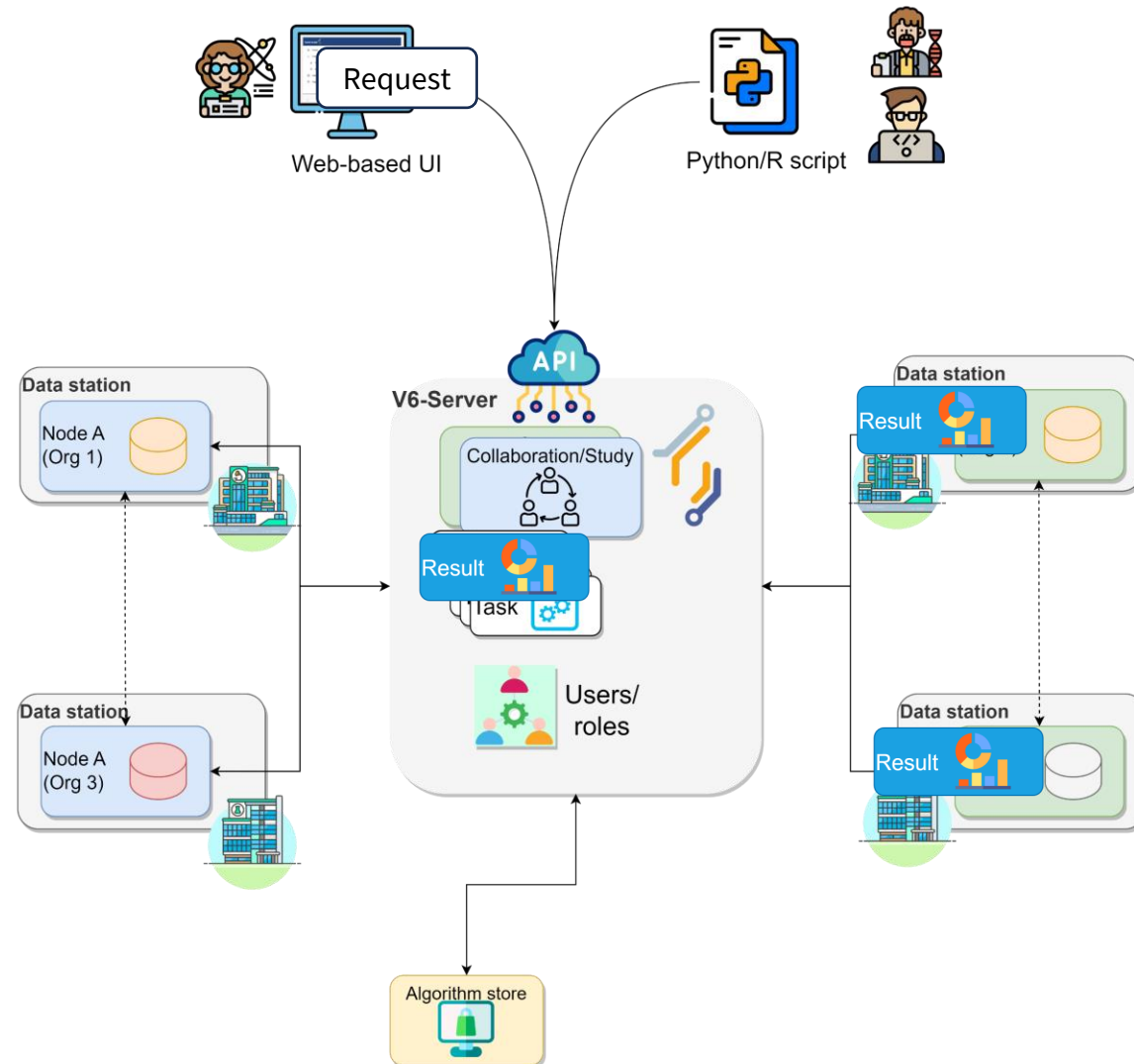
Algorithm store

Repository for trusted algorithms. It allows researchers to explore which algorithms are available and how to run them.



Jointly organized by:

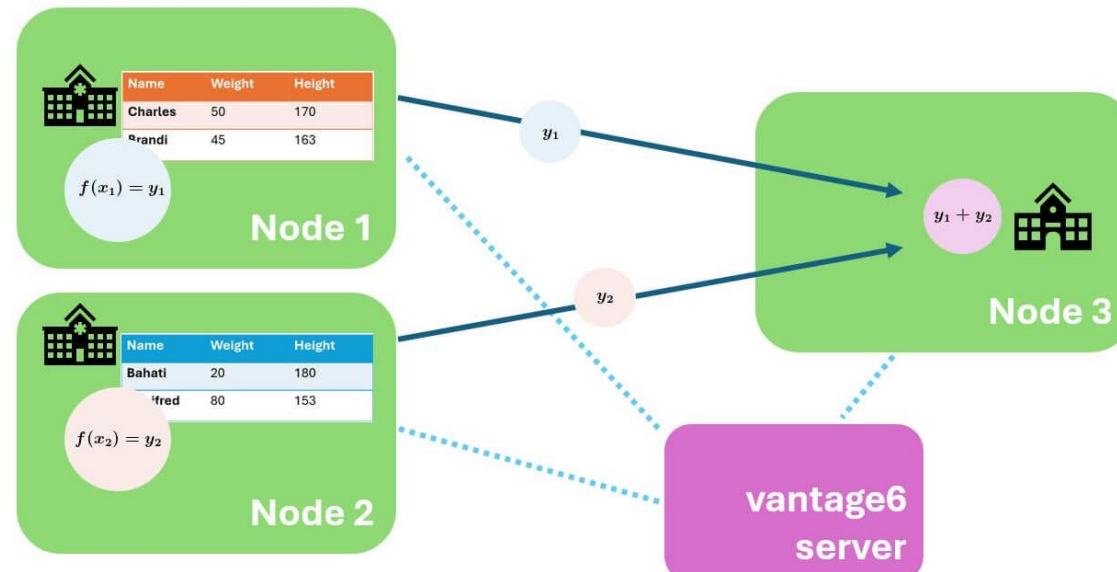
The vantage6 infrastructure



Jointly organized by:

How algorithms run in vantage6

Let us consider the federated sum from chapter 1 again



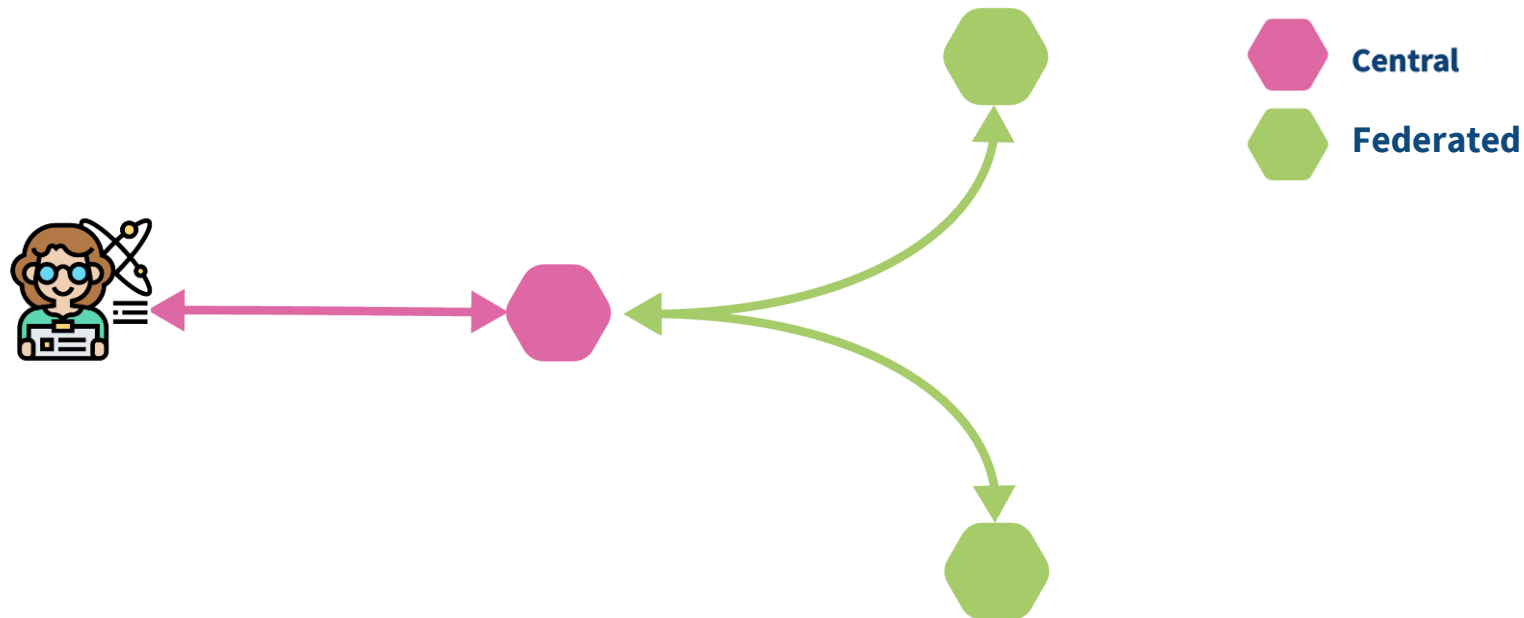
Data sources and the aggregation entity are mapped to nodes.

The vantage6 server is on the side, coordinating the analysis.

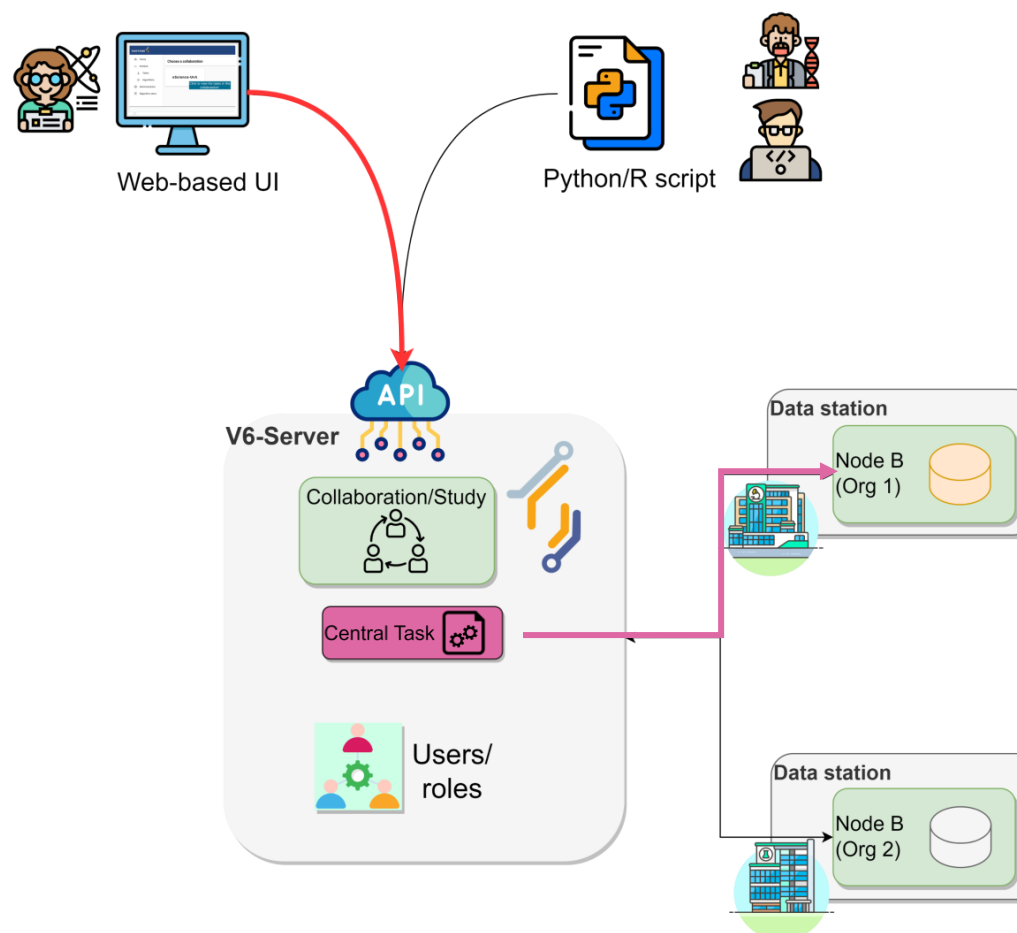
How algorithms run in vantage6

Federated algorithms can be split in a **federated** and a **central** part:

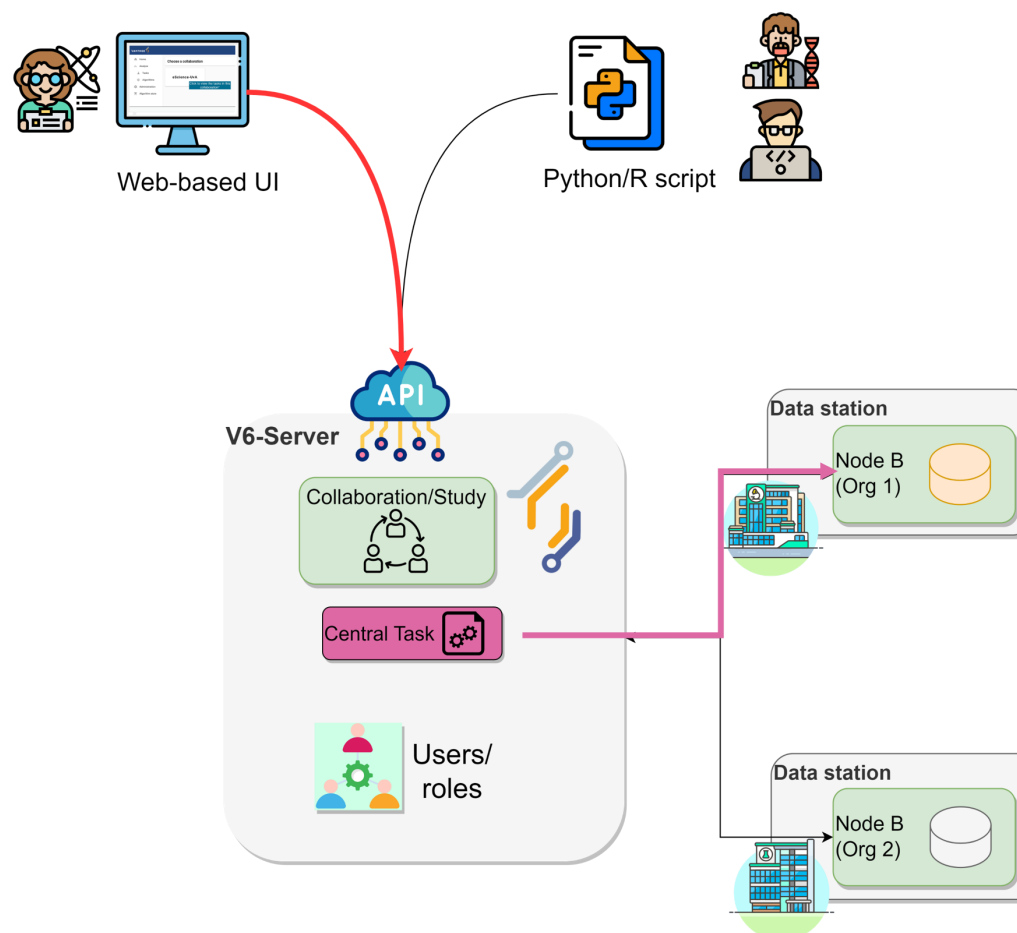
- **Central:** The central part of the algorithm is responsible for orchestration and aggregation of the partial results.
- **Federated:** The partial tasks are executing computations on the local privacy sensitive data.



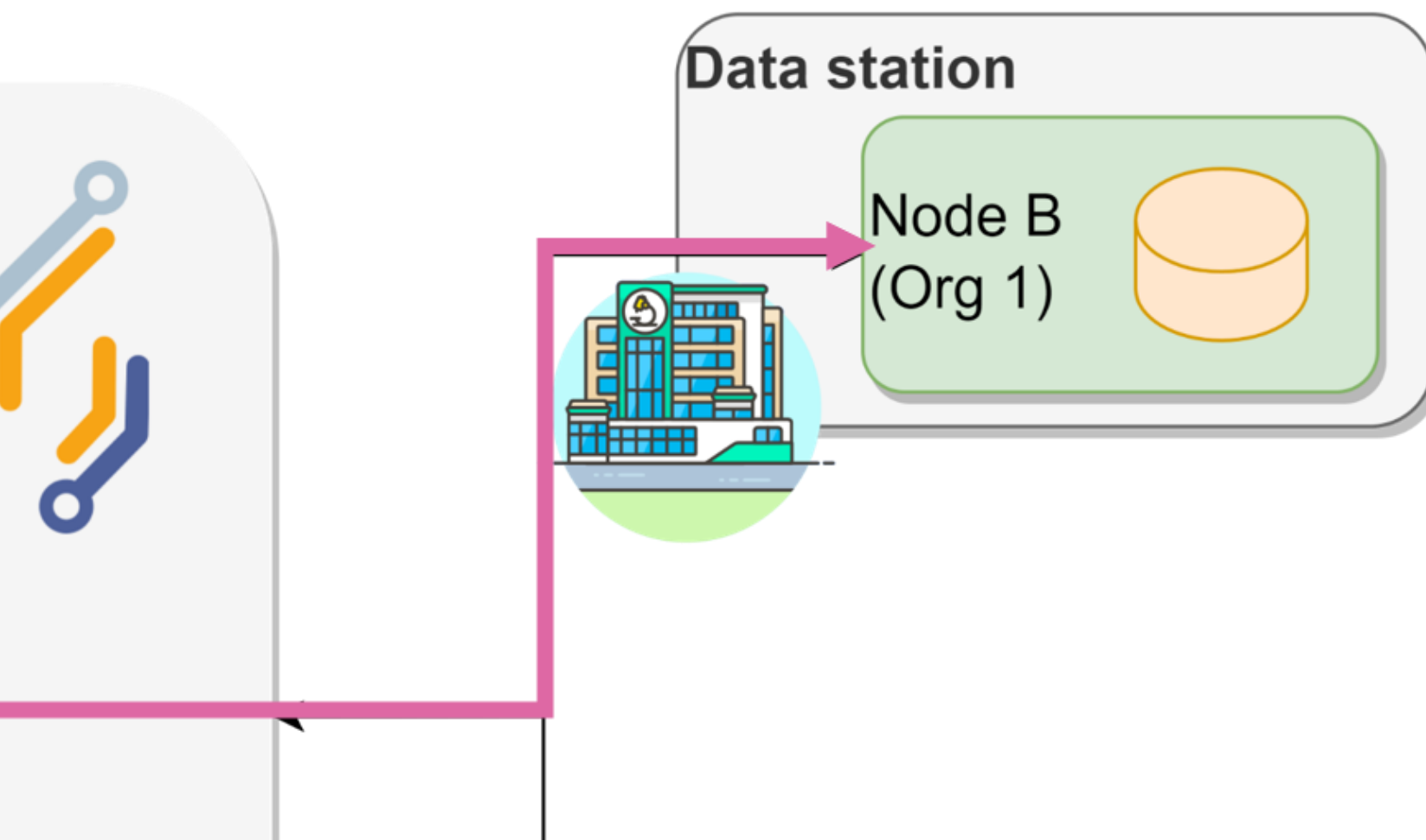
How algorithms run in vantage6



How algorithms run in vantage6



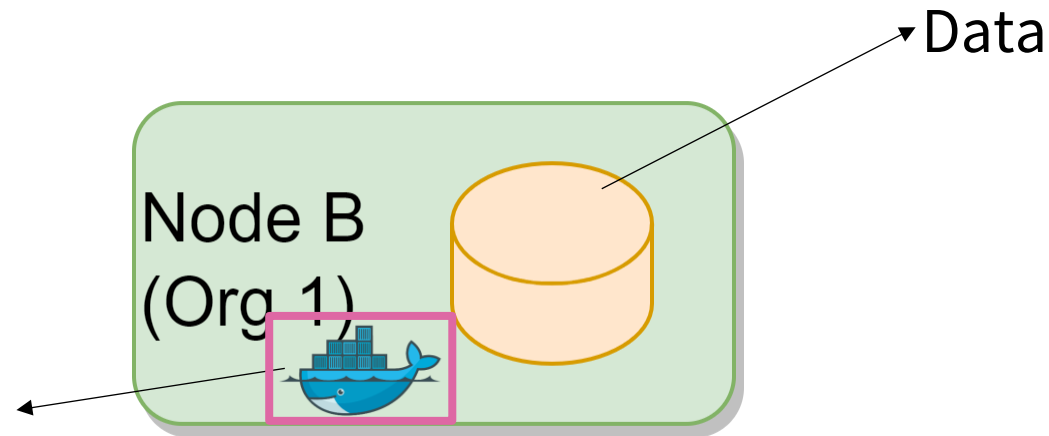
How algorithms run in vantage6



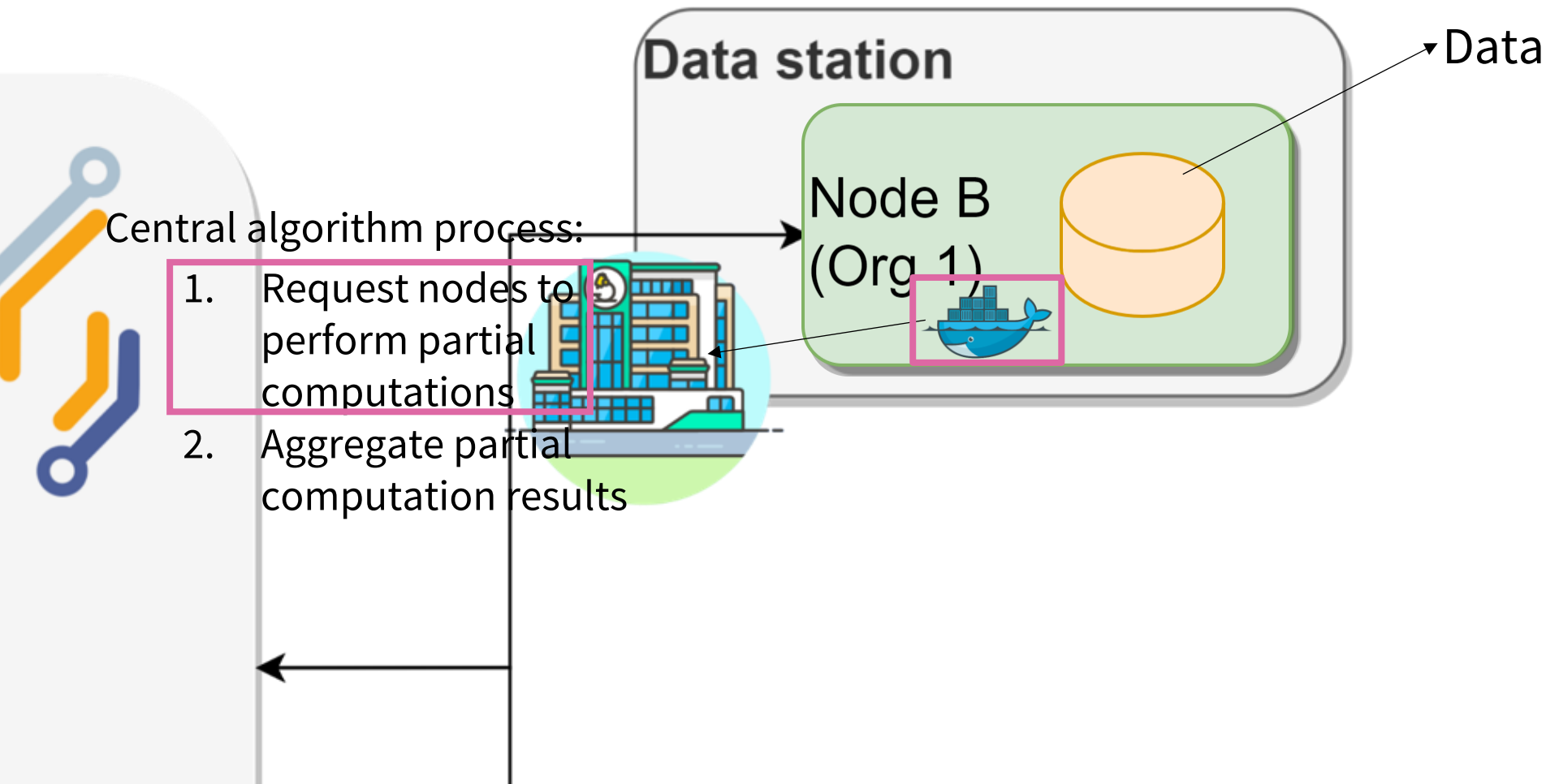
How algorithms run in vantage6

Central algorithm process:

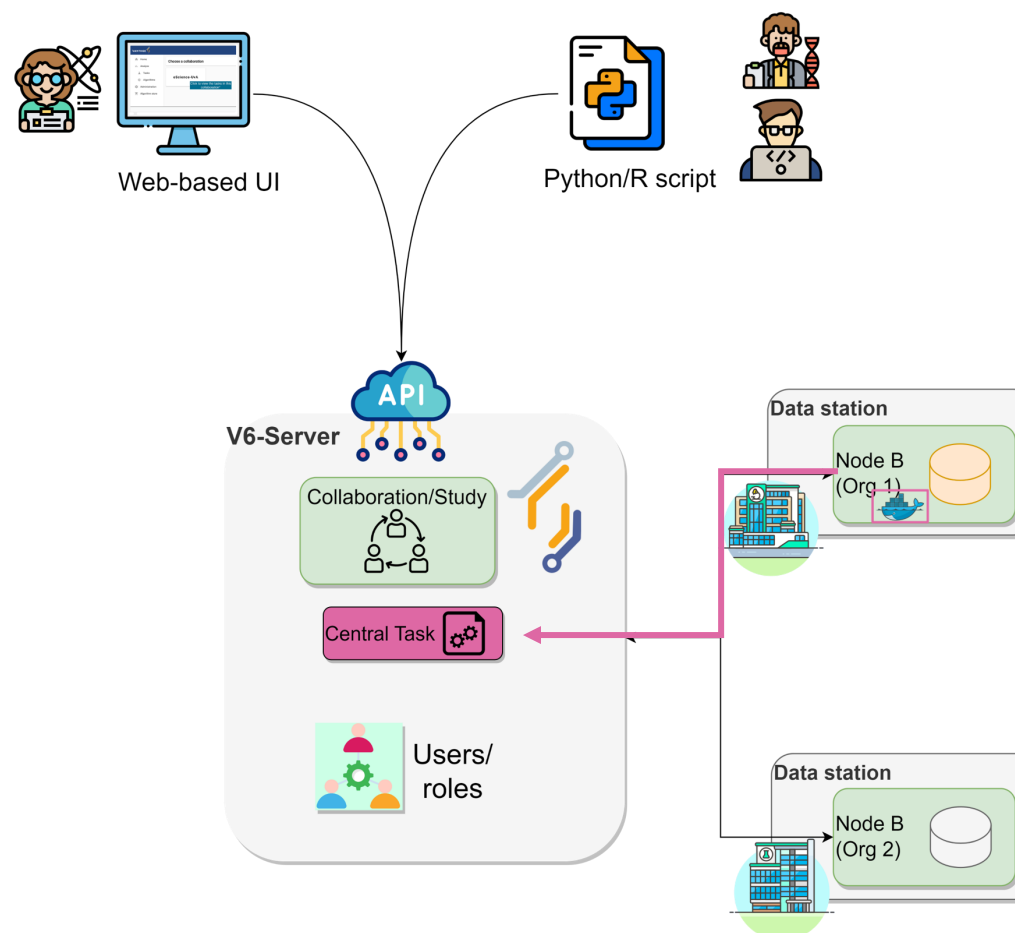
1. Request nodes to perform partial computations
2. Aggregate partial computation results



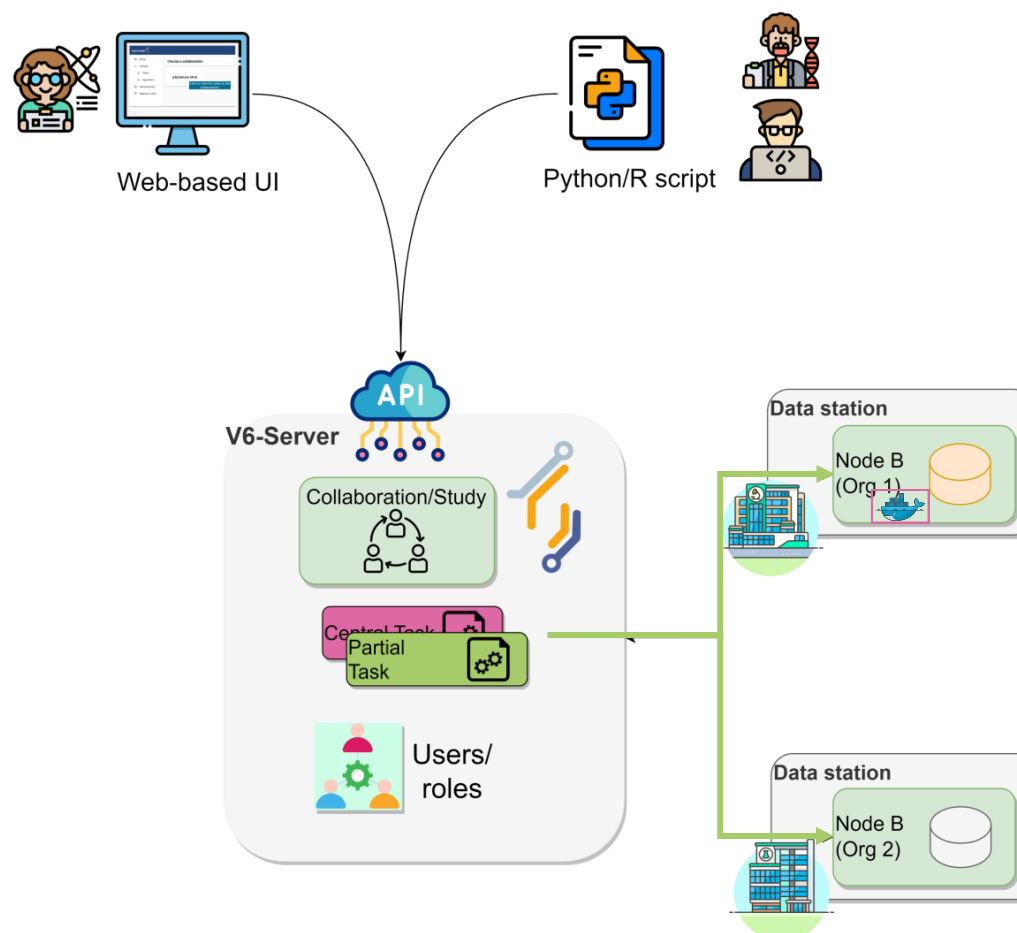
How algorithms run in vantage6



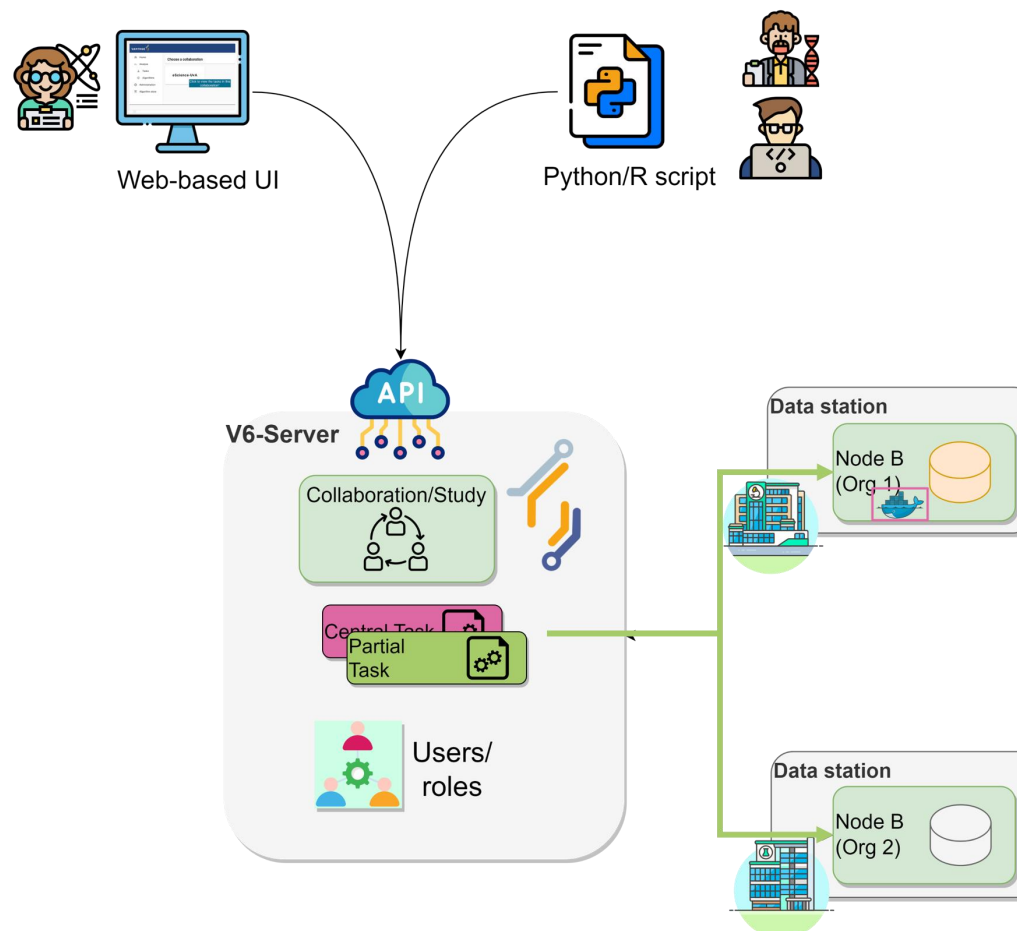
How algorithms run in vantage6



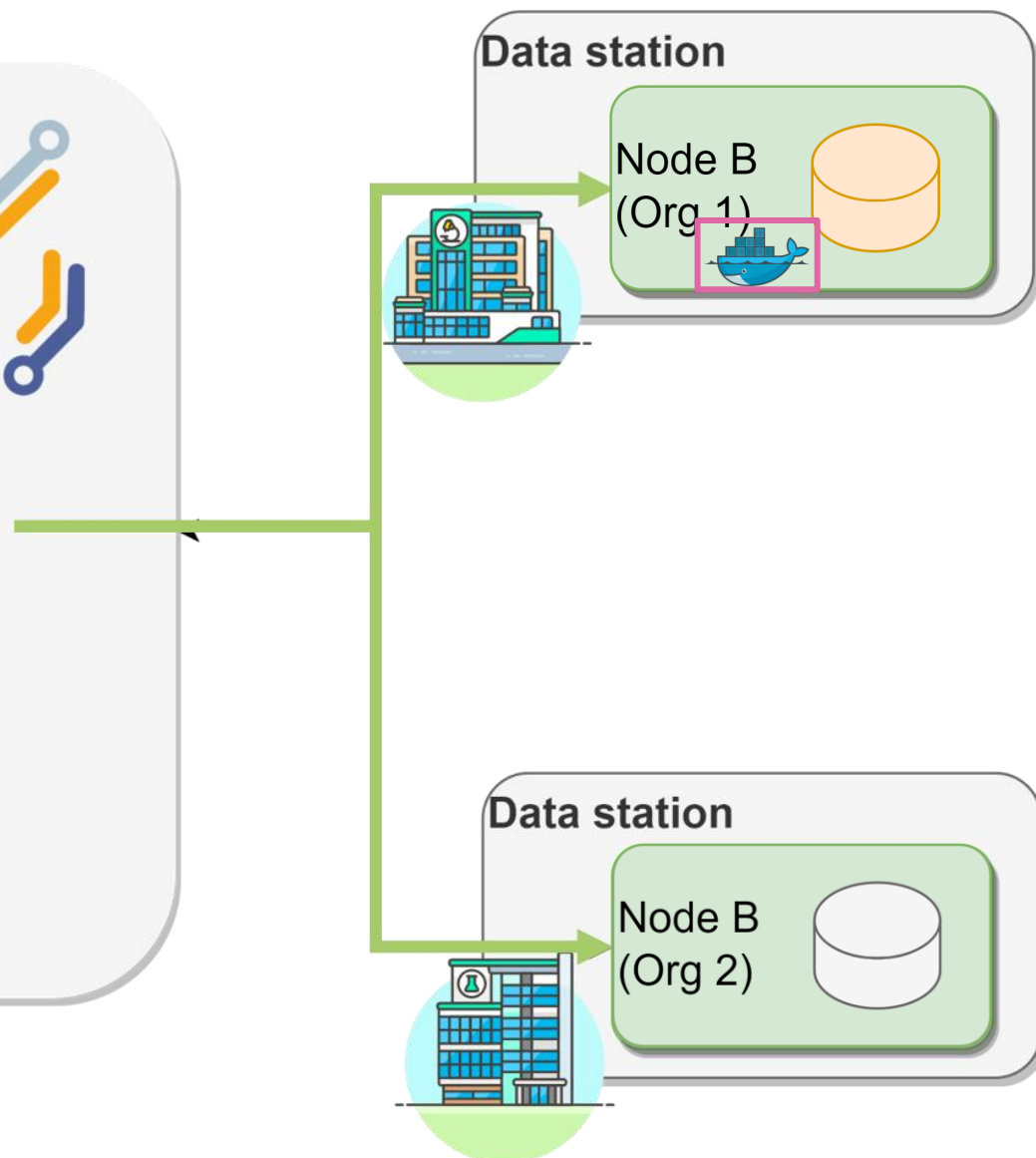
How algorithms run in vantage6



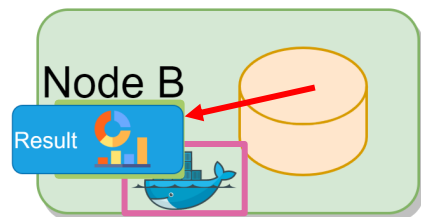
How algorithms run in vantage6



How algorithms run in vantage6

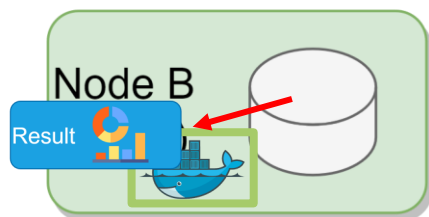


How algorithms run in vantage6

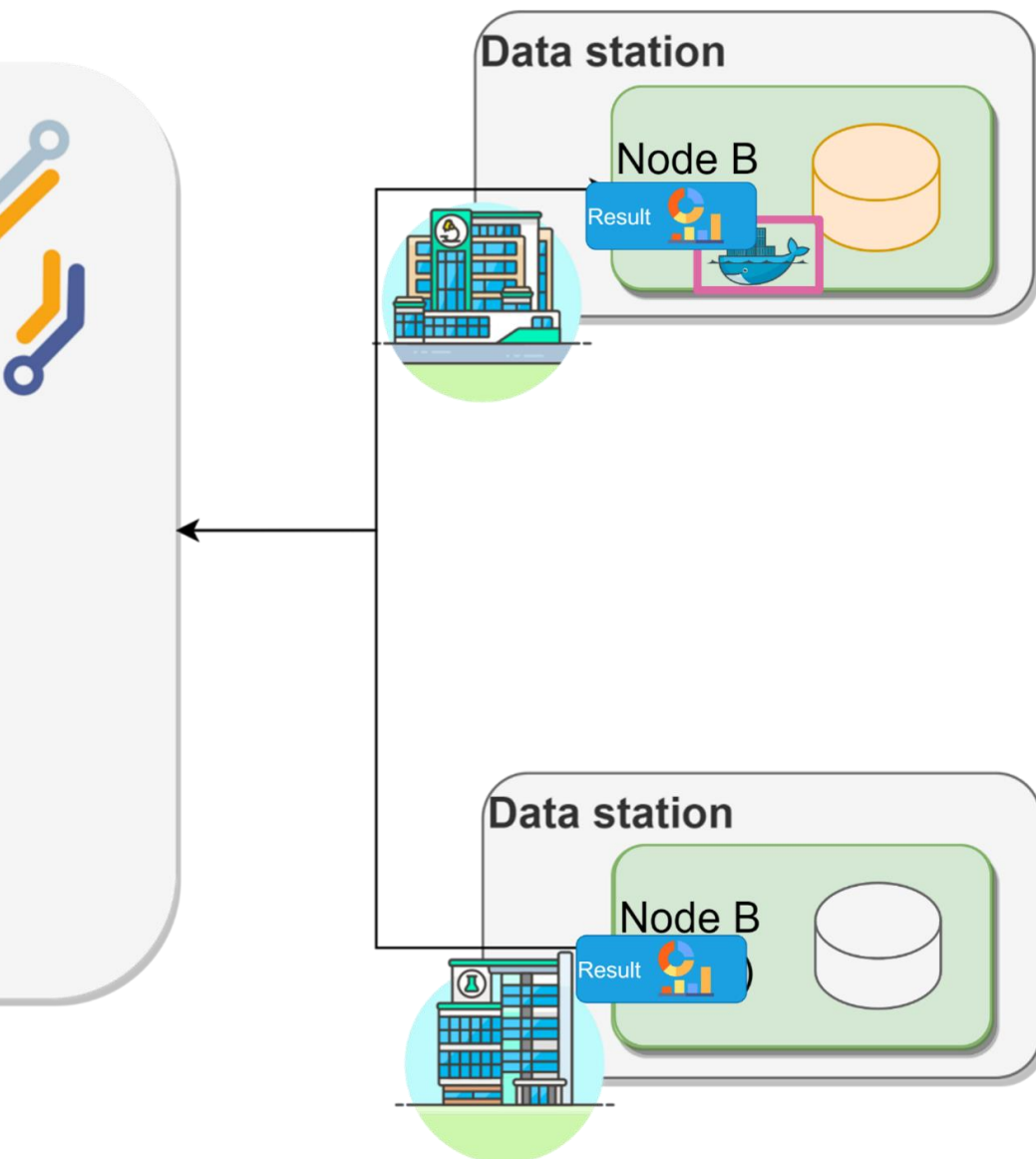


Federated algorithm process:

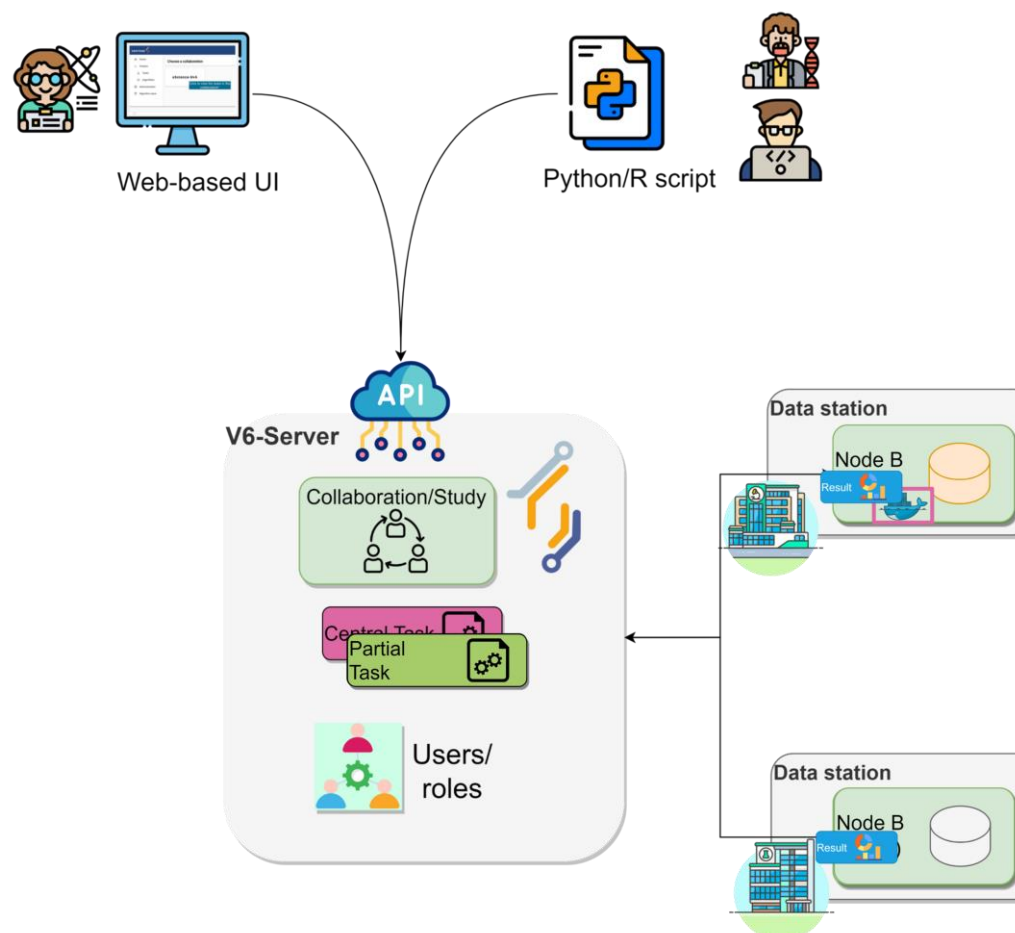
1. Read local data
2. Compute partial results



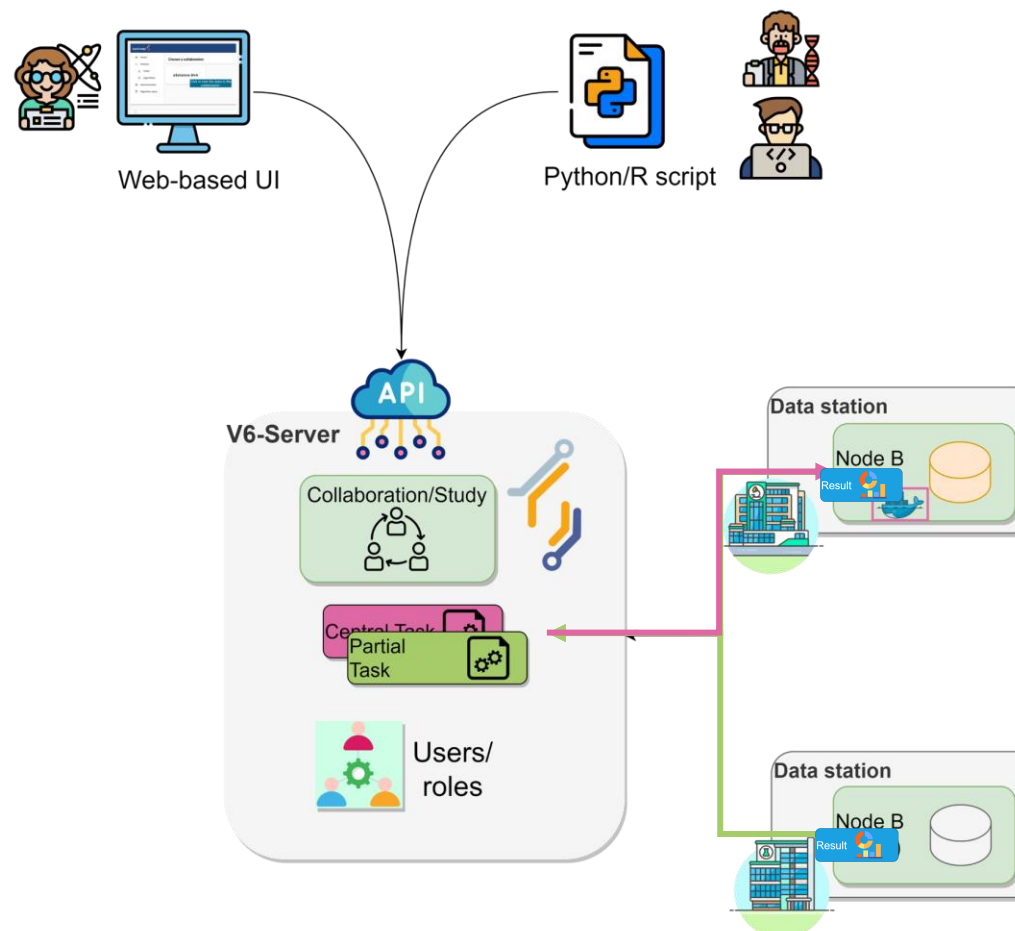
How algorithms run in vantage6



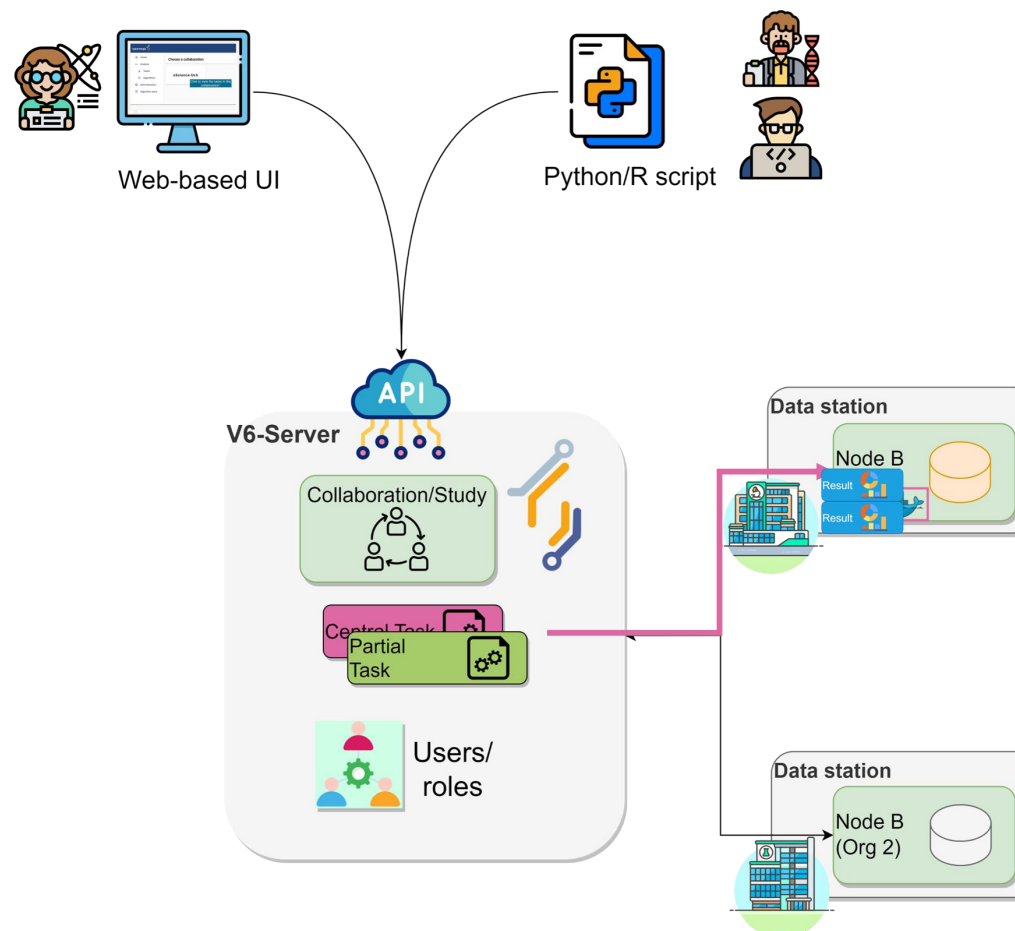
How algorithms run in vantage6



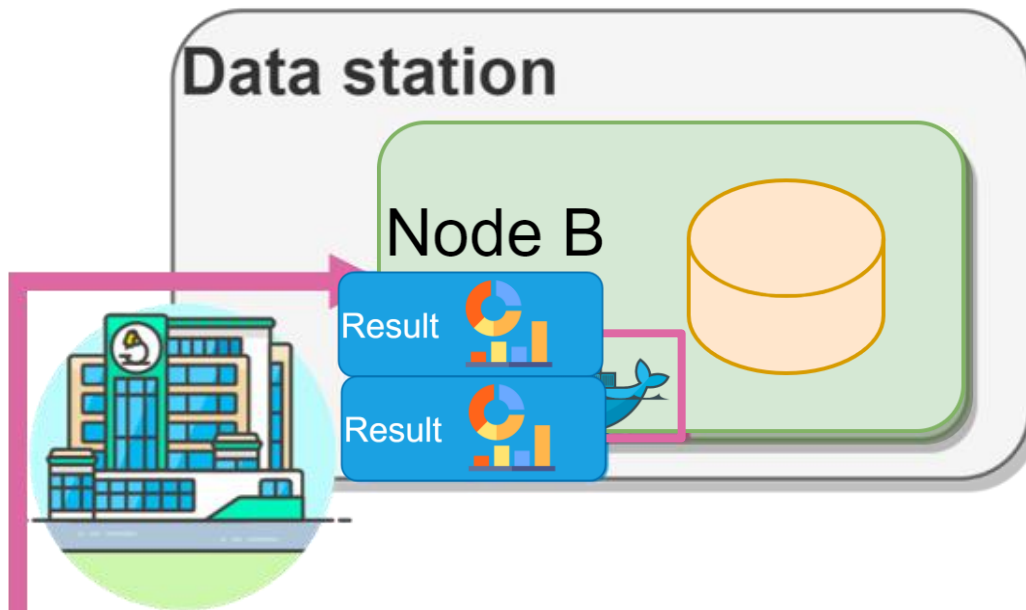
How algorithms run in vantage6



How algorithms run in vantage6



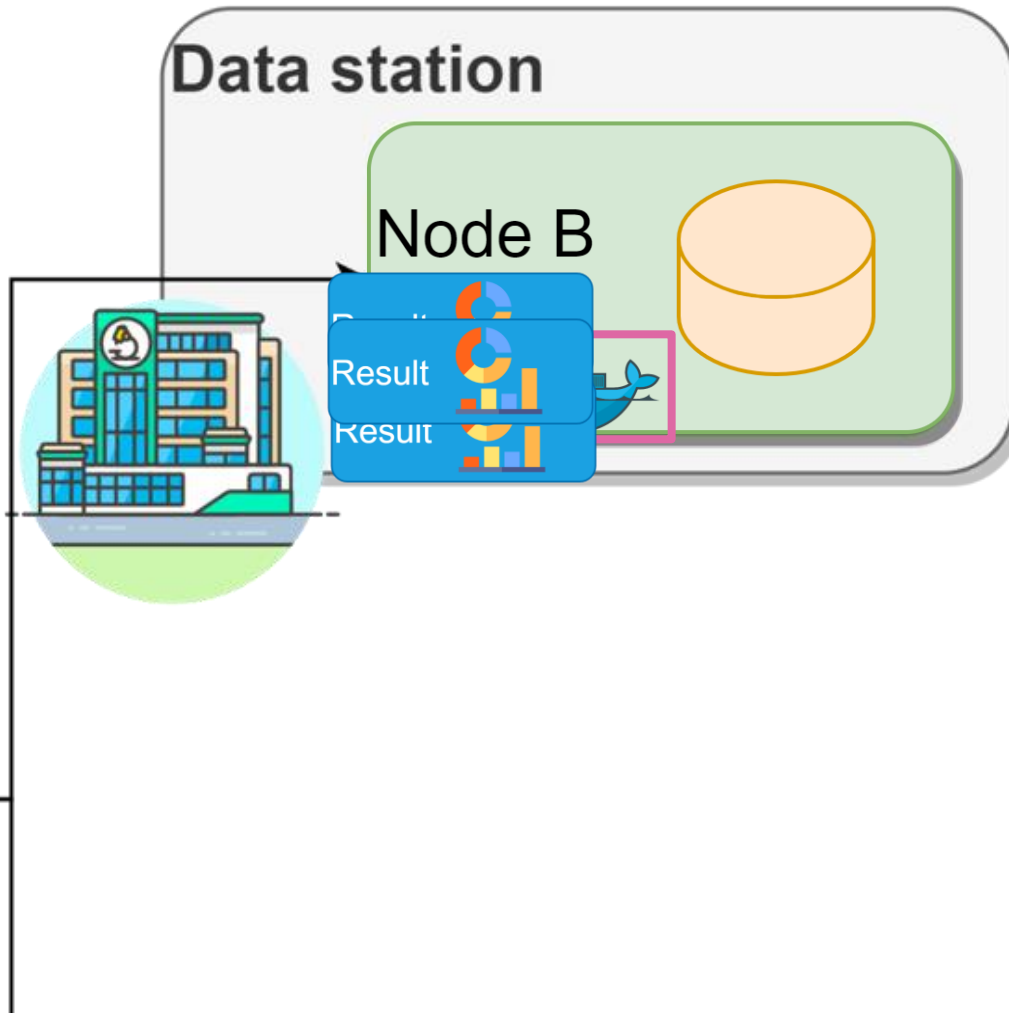
How algorithms run in vantage6



Central algorithm process:

1. Request nodes to perform partial computations
2. Aggregate partial computation results

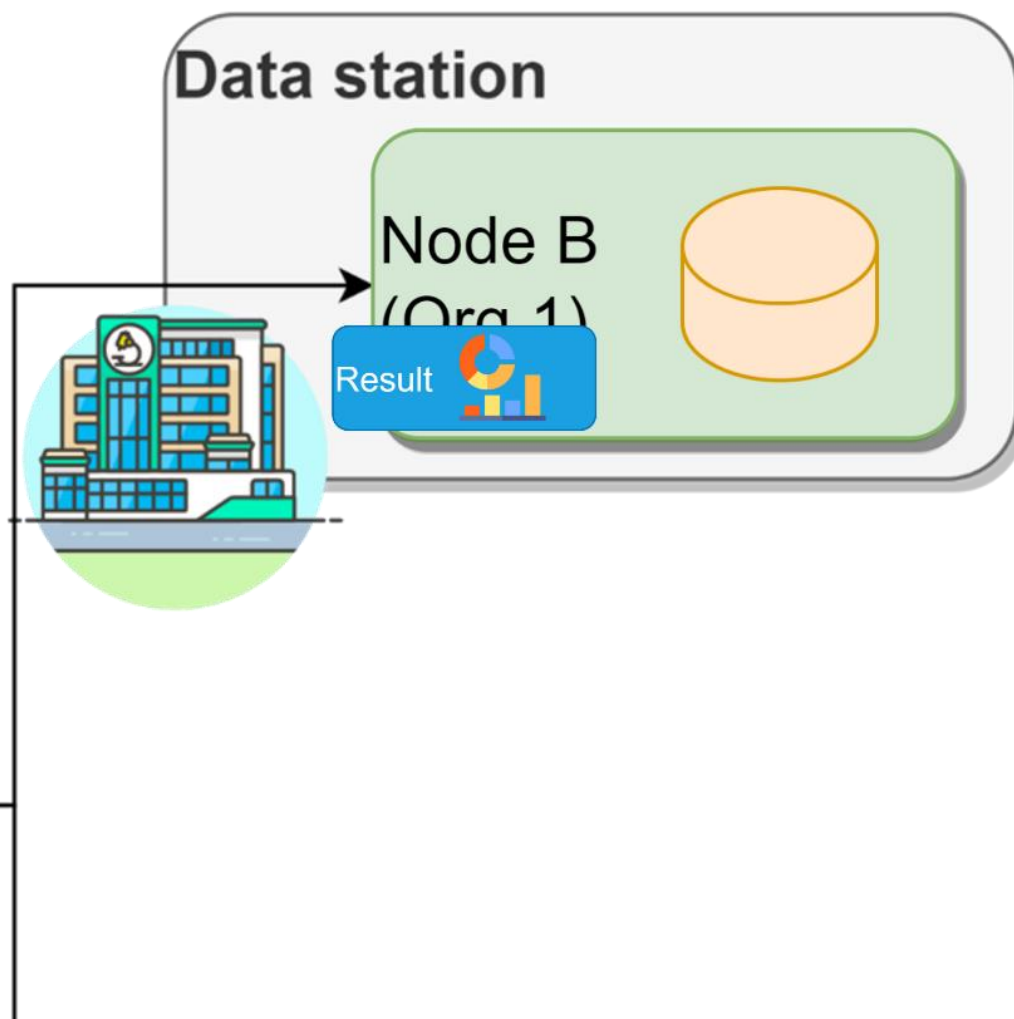
How algorithms run in vantage6



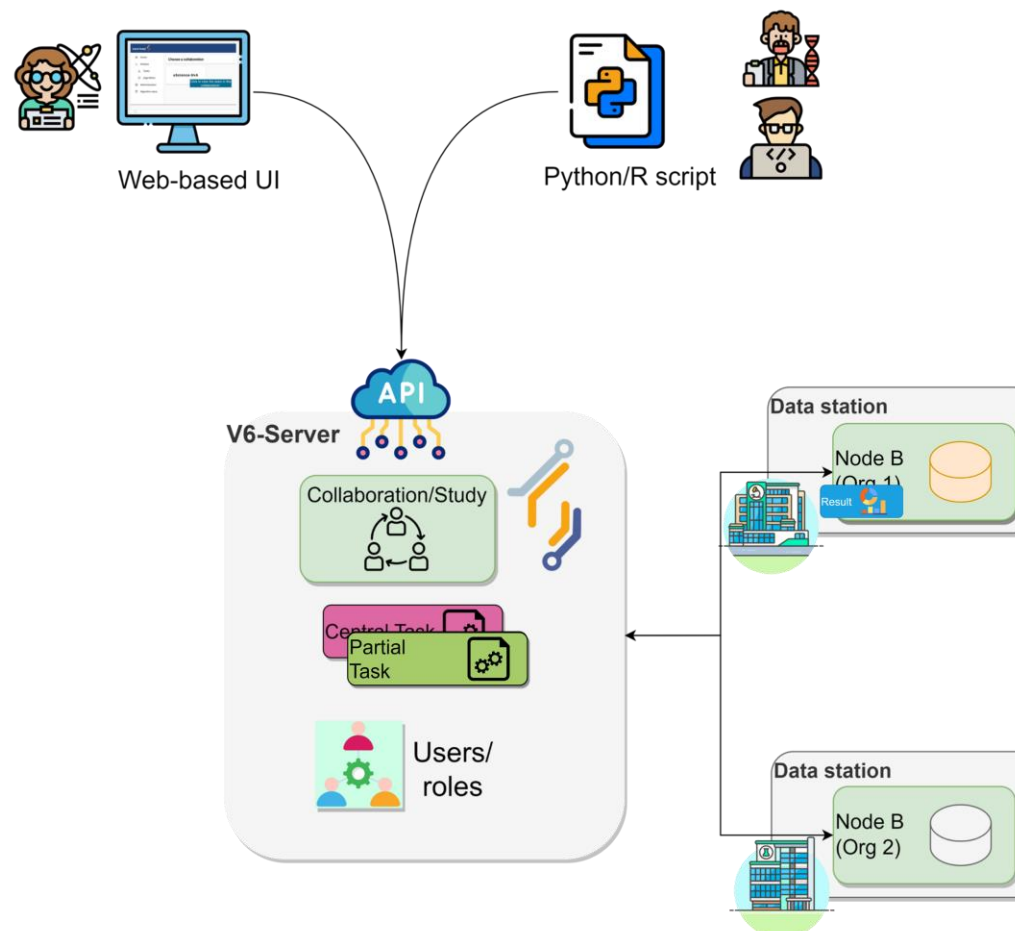
Central algorithm process:

1. Request nodes to perform partial computations
2. Aggregate partial computation results

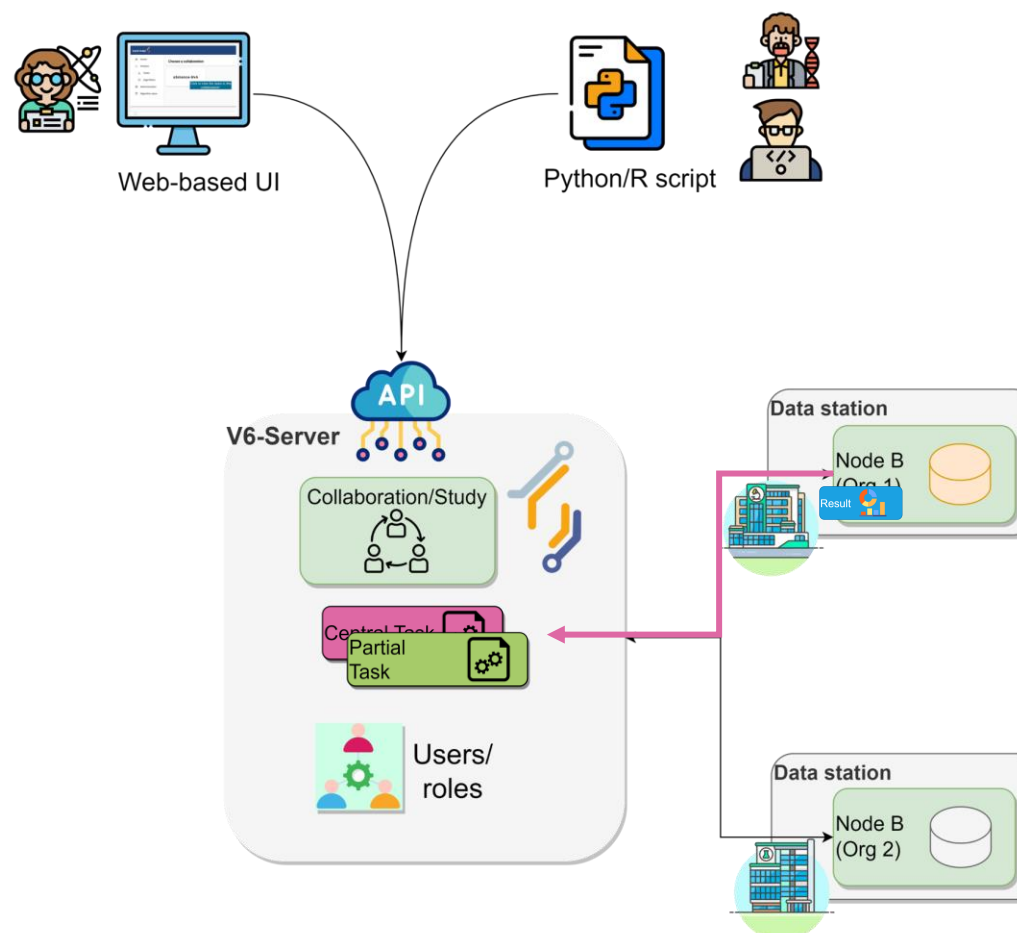
How algorithms run in vantage6



How algorithms run in vantage6



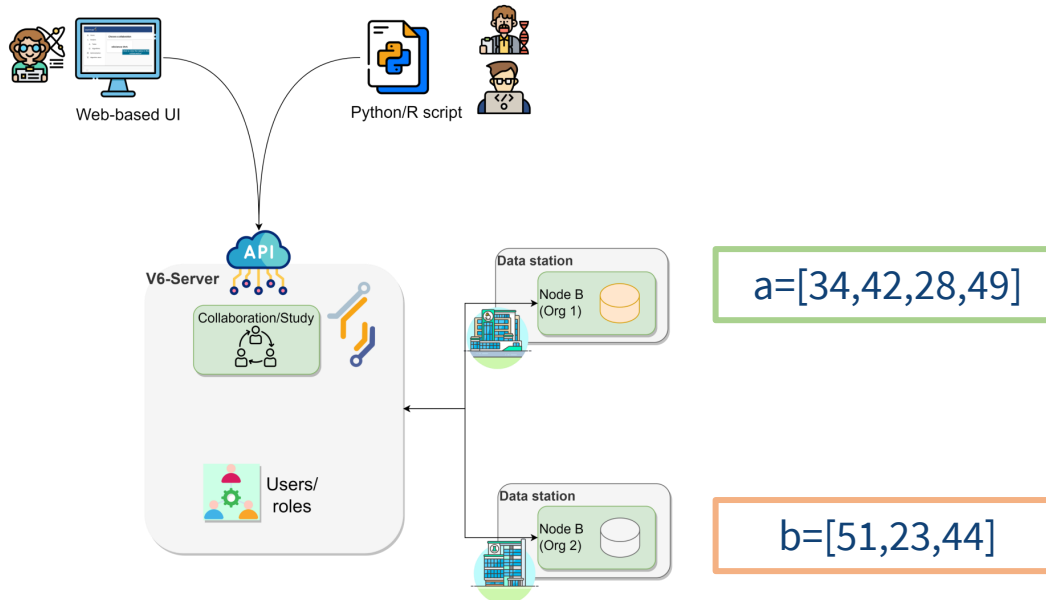
How algorithms run in vantage6



Exercise

⚡ CHALLENGE

Two centers A and B have data regarding the age of a set of patients



Which part of the infrastructure will execute each part of the computation?

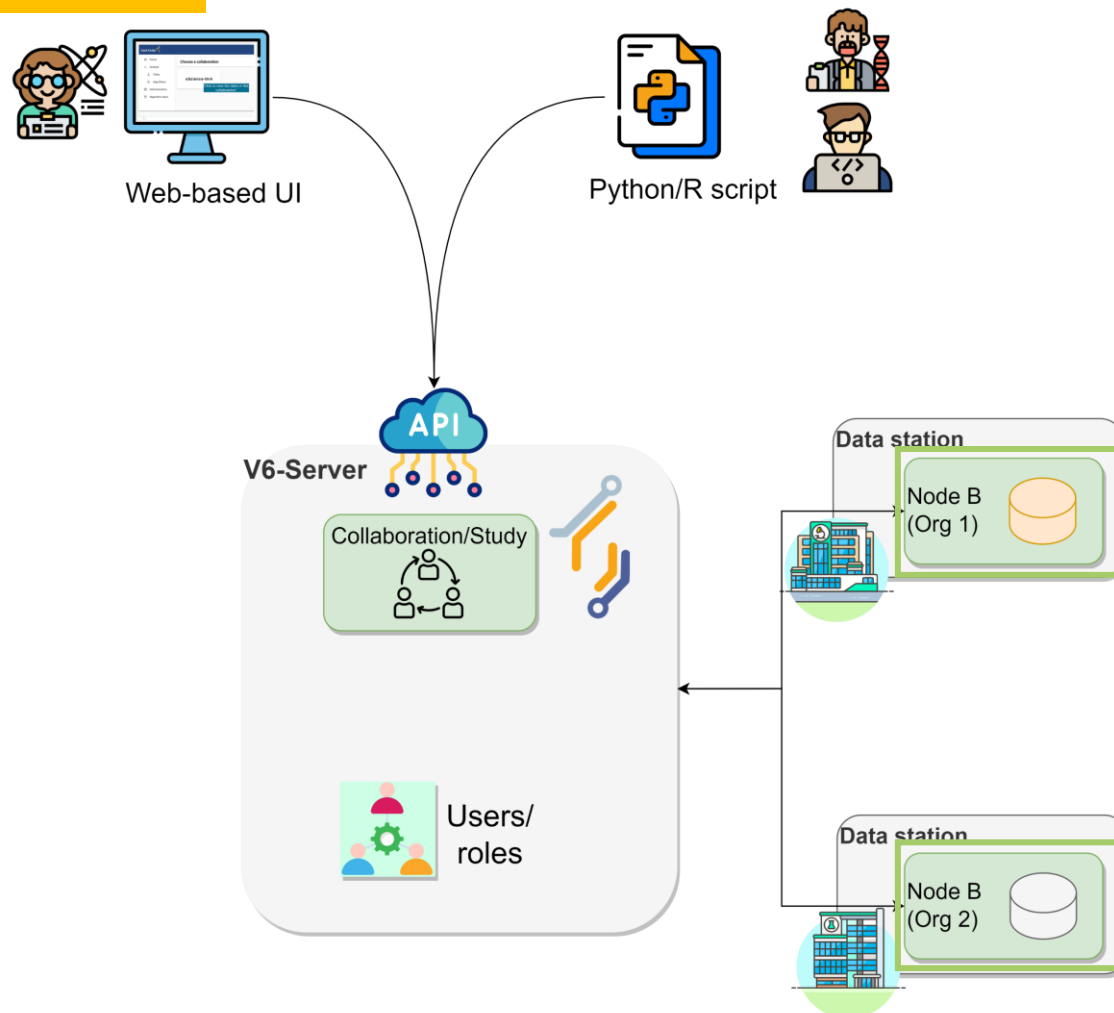
Which is the result returned by the different parts?

They want to compute the average age:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{34 + 42 + 28 + 49 + 51 + 23 + 44}{4 + 3} = \frac{1}{n_a + n_b} \sum_{i=1}^{n_a} a_i + \sum_{i=1}^{n_b} b_i$$

Exercise

⚡ SOLUTION



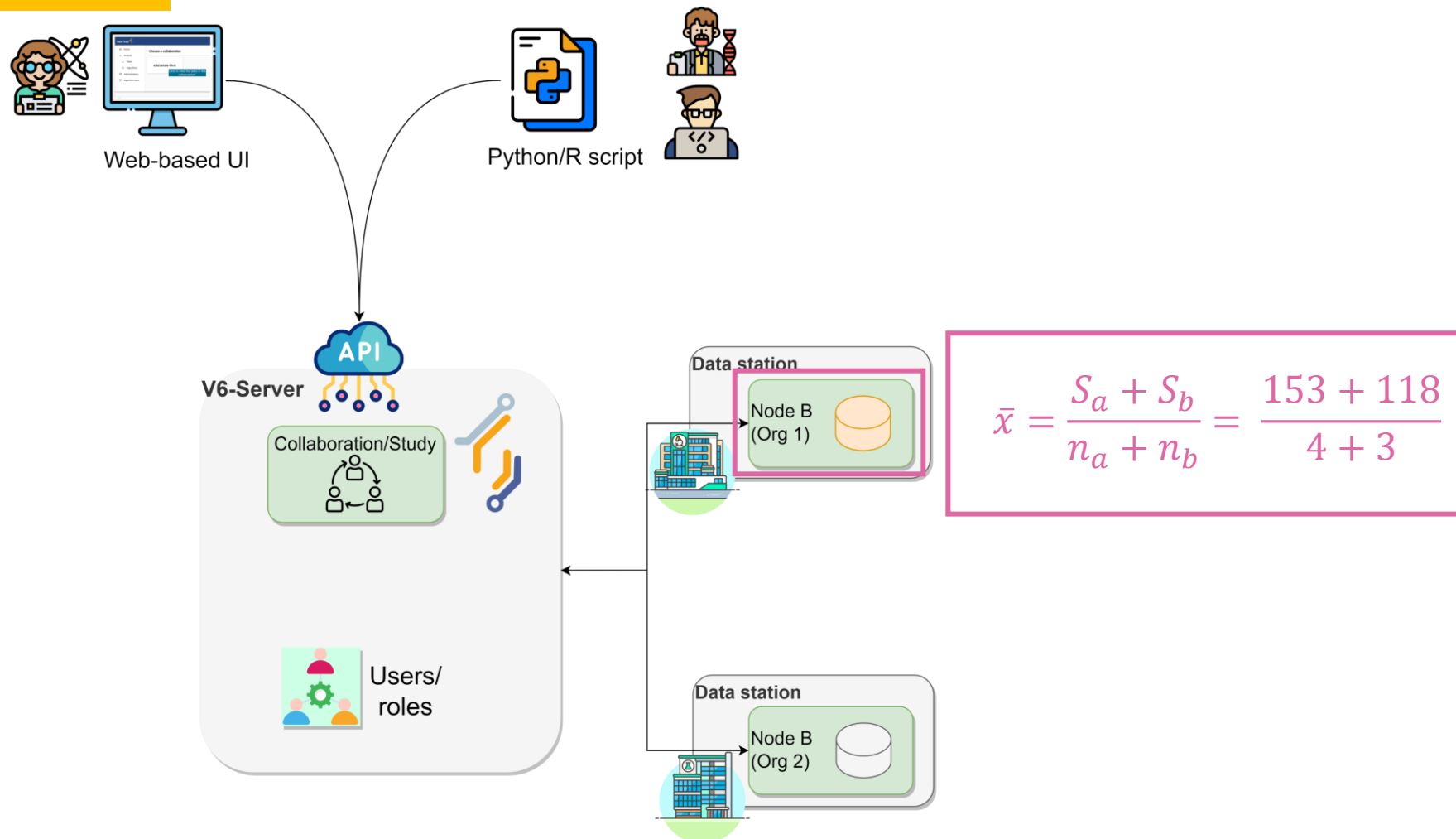
$$S_a = \sum_{i=1}^{n_a} a_i = 34 + 42 + 28 + 49$$
$$n_a = 3$$

$$S_b = \sum_{i=1}^{n_b} b_i = 51 + 23 + 44$$
$$n_b = 3$$

Jointly organized by:

Exercise

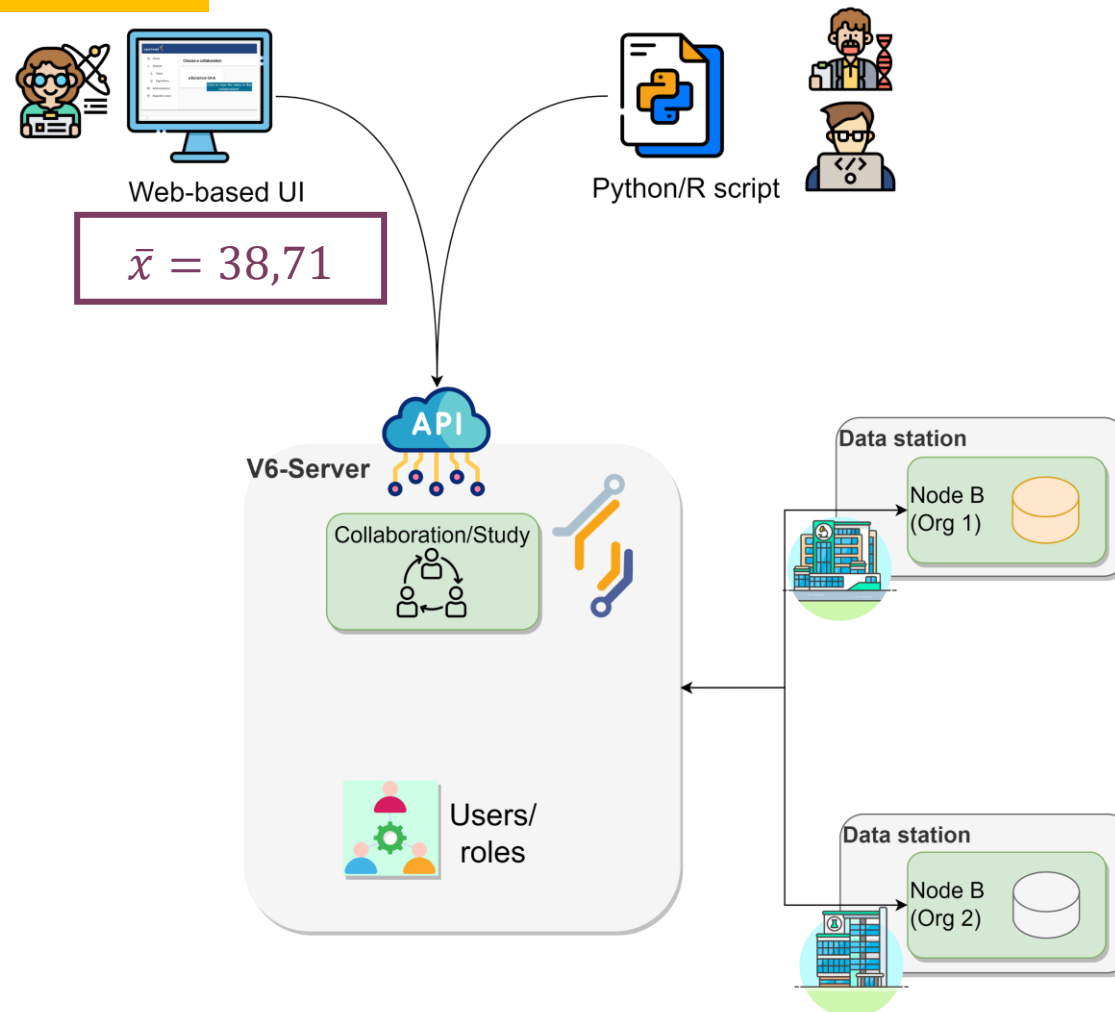
⚡ SOLUTION



Jointly organized by:

Exercise

⚡ SOLUTION



Jointly organized by:

Future developments of vantage6



- **Policies:** we want to extend the current available policies to a more generic policy framework in which any aspect of the vantage6 platform can be controlled by policies. This will maximize the flexibility of the platform and make it easier to adapt to new use cases.
- **Model repository:** Currently, vantage6 is focused on privacy enhancing techniques. Some of these techniques result in a model that can be used to make predictions. We want to extend vantage6 with a model repository in which these models can be stored, shared and used. This will make it easier to reuse models and to compare the performance of different models.