eSAAM 2023

on Cloud to Edge Continuum

Enabling Compute and Data Sovereignty with Infrastructure-Level Data Spaces

Jacopo Marino (Politecnico di Torino)

jacopo.marino@polito.it

Oct. 17, 2023

Ludwigsburg, Germany

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

Data is a critical asset in today’s world

Data sovereignty

The control of data flows and the related infrastructure via national jurisdiction

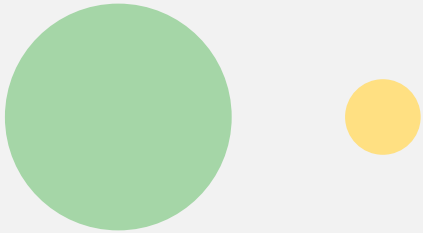
Icon from Flaticon.com

Data gravity

The ability of data to attract applications, services, and other data

Data gravity

The ability of data to attract applications, services, and other data



Data

Application

“A data space is a decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles”

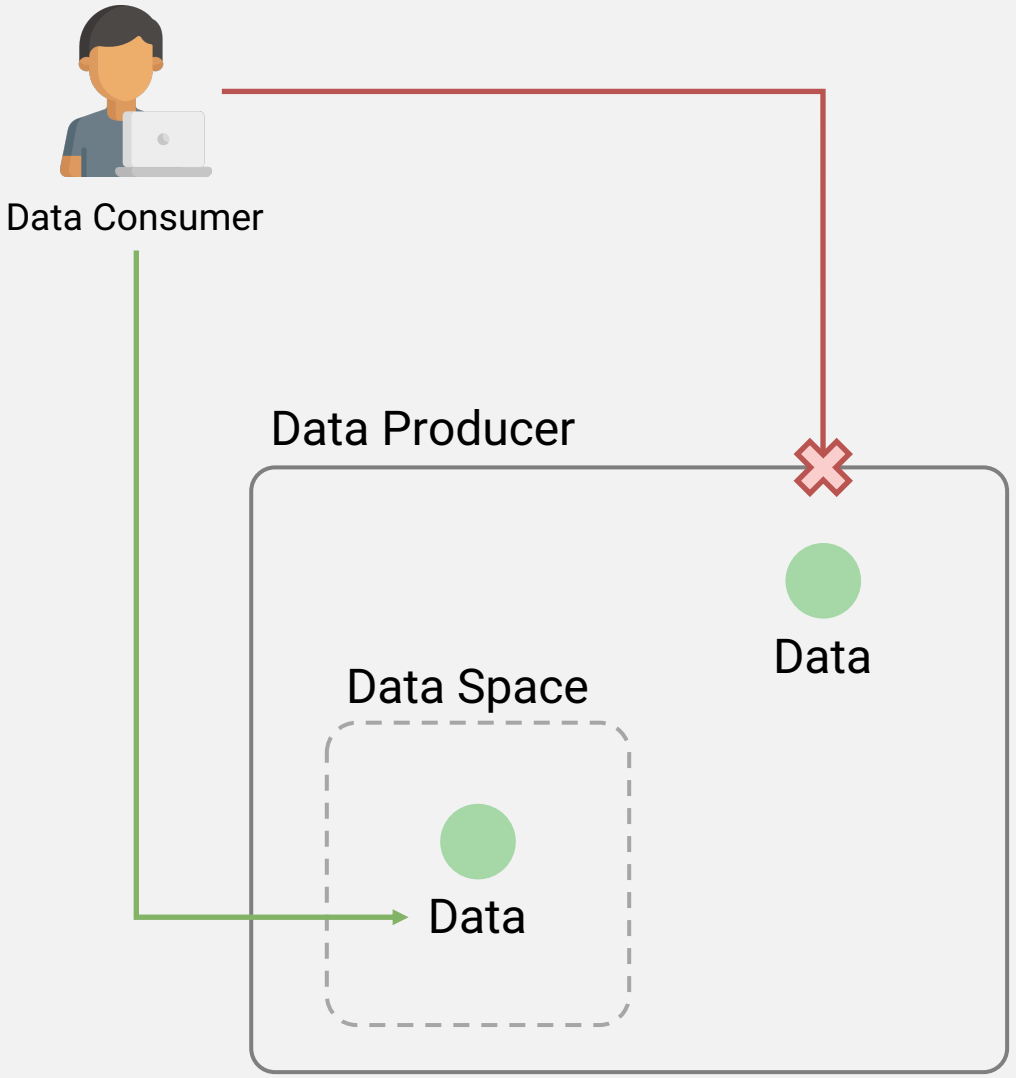
Open DEI project

“A data space is a type of data relationship between trusted partners who adhere to the same high-level standards and guidelines in relation to data storage and sharing within one or many vertical ecosystems (i.e. Health, Infrastructure, Tourism, etc.)”

Gaia-X

Data Spaces

Enable a data producer to offer its data to potential consumers, without giving up on security and data ownership/sovereignty rules, and without affecting the possibility of consumers to read and process arbitrary data.



Icon from Flaticon.com

Application-level data spaces problems

01.

No data gravity

Secure data exchange, but **no support for any computing component** (i.e., pods); hence, in case of a very large quantity of data, all data needs to be transferred from the producer to the consumer

02.

API gateway

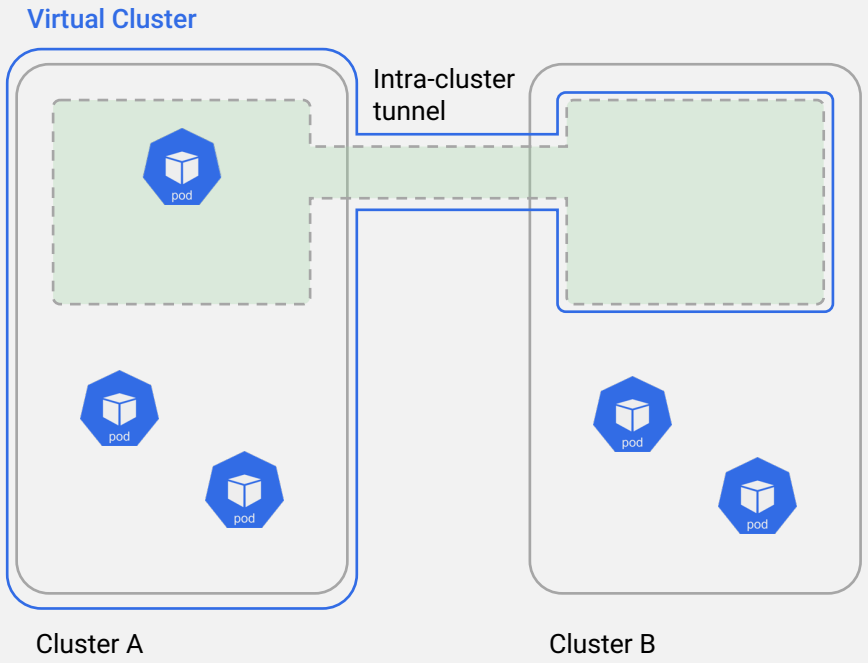
IDSA connector allows access to data from the data consumer, but the latter might prefer having its **own API gateway with its aggregation rules**

03.

Exposed endpoints

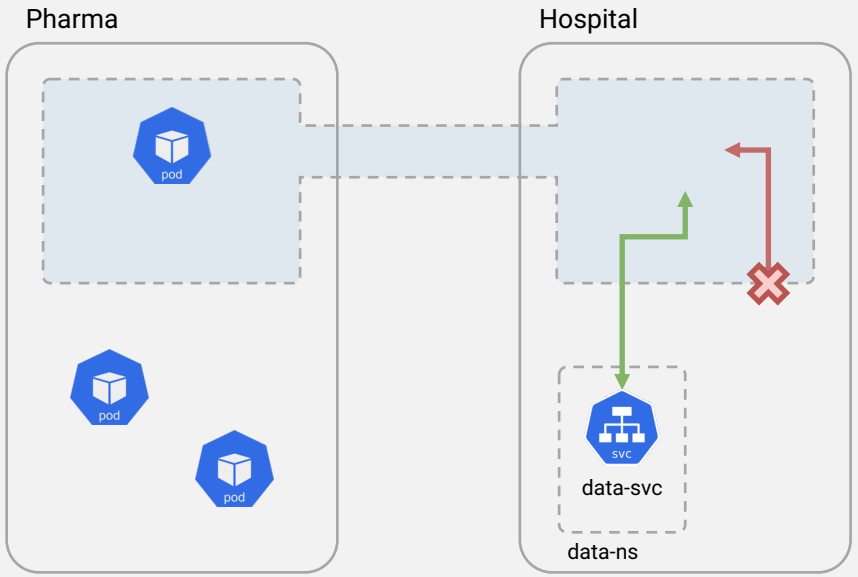
Mutually reachable endpoints are required (e.g., **publicly exposed on the Internet**) to allow connectors to exchange data

Open-source project that enables dynamic and seamless Kubernetes multi-cluster topologies.
It allows to share resources and services coming from different clusters, which are aggregated into a seamless computing continuum called **virtual cluster**.



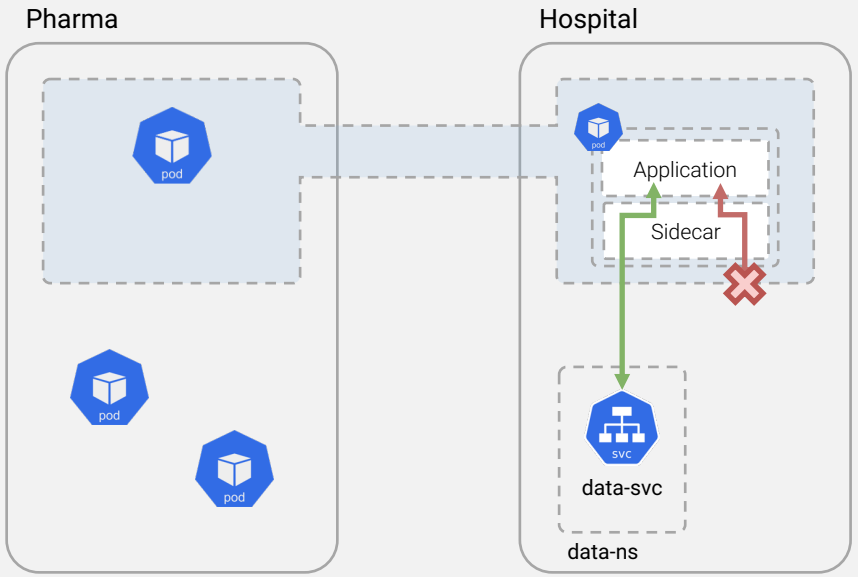
Pharma outsourcing

Pharma **outsources the execution** of its workloads to Hospital, while retaining the ability to govern the offloading process using the capabilities provided by Liqo.



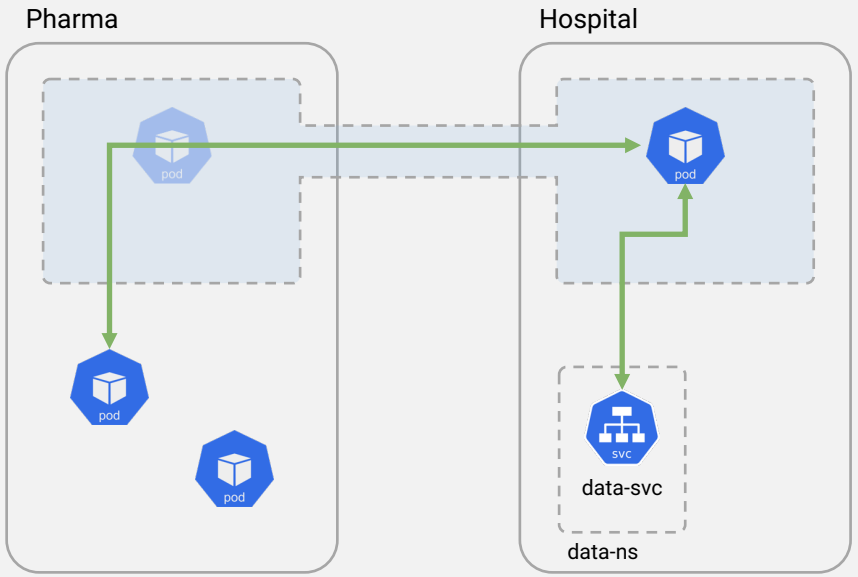
Hospital security measures

During pod offloading there is the addition of an **Init container** and a **Sidecar** to the main application container. The Sidecar acts as a proxy and monitors all the pod's communications, allowing only the desired ones.



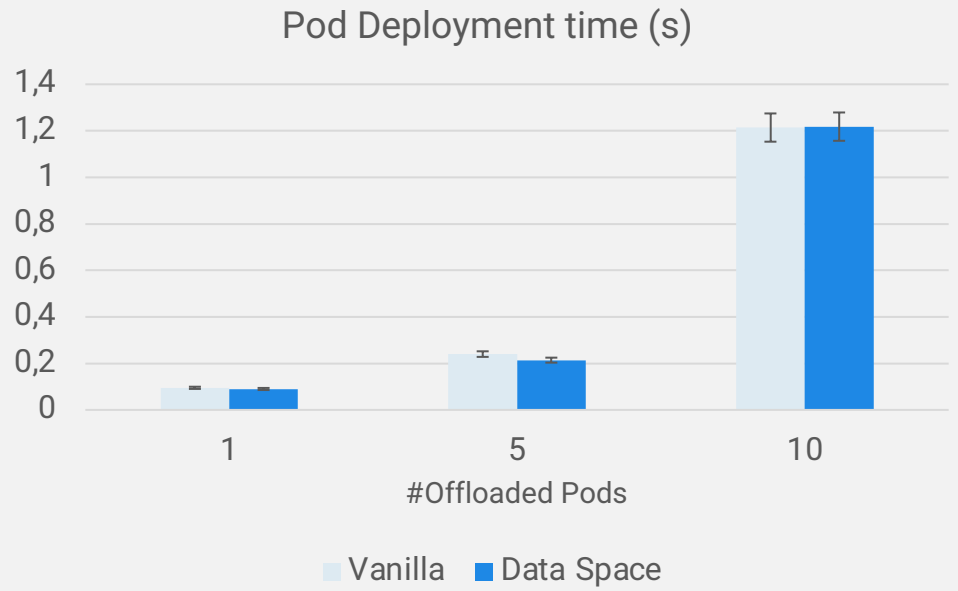
Experimental scenario

Pharma cluster has a single pod containing the processing logic, which is offloaded to the Hospital cluster



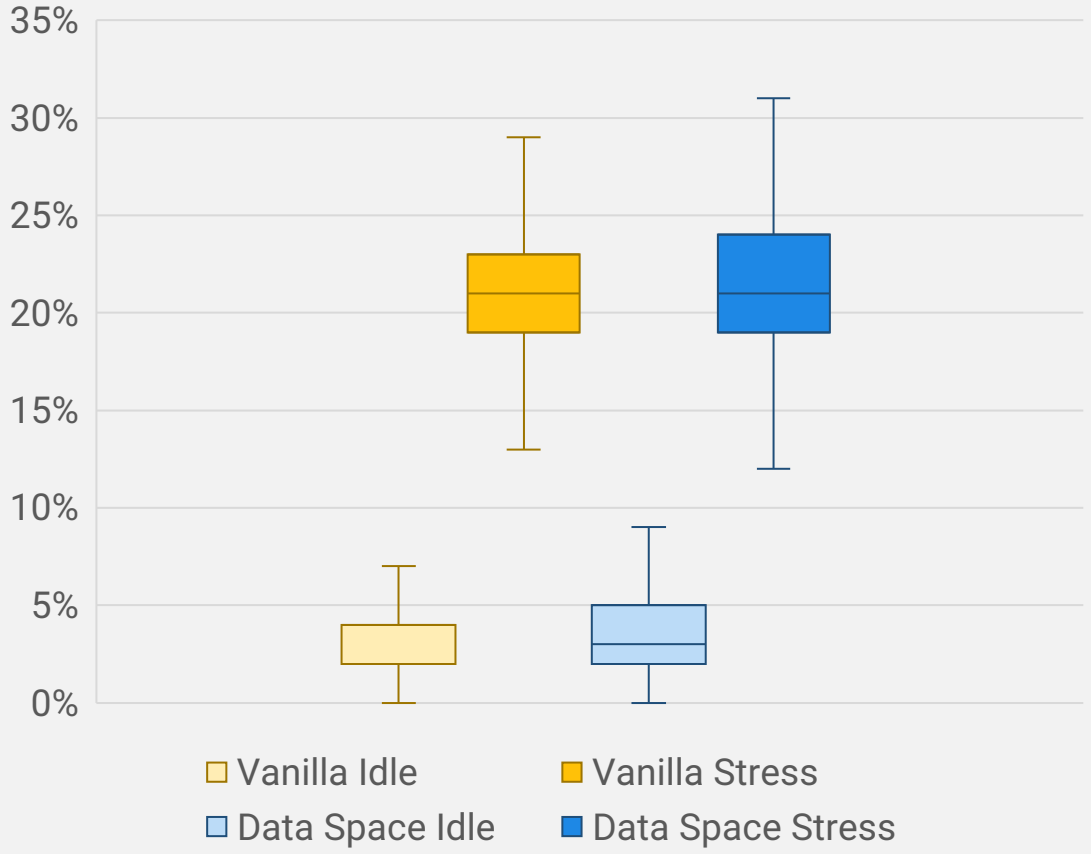
Data Space creation time

The standard, or vanilla, execution of Liqo requires the establishment of a point-to-point connection between the clusters and then the offloading of a namespace



Resource consumption

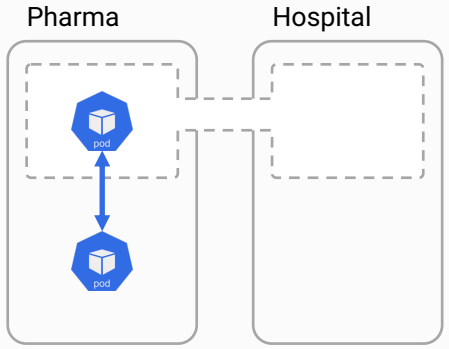
Metric for evaluating the computational demands placed upon a system during data transfer events



Latency scenarios

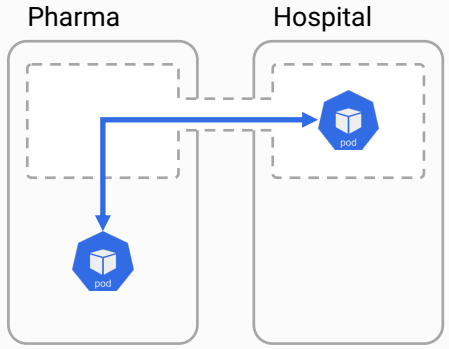
01.

Local



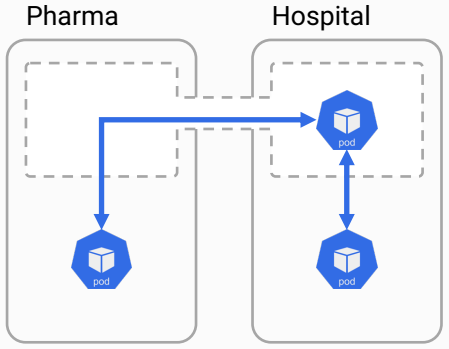
02.

Remote

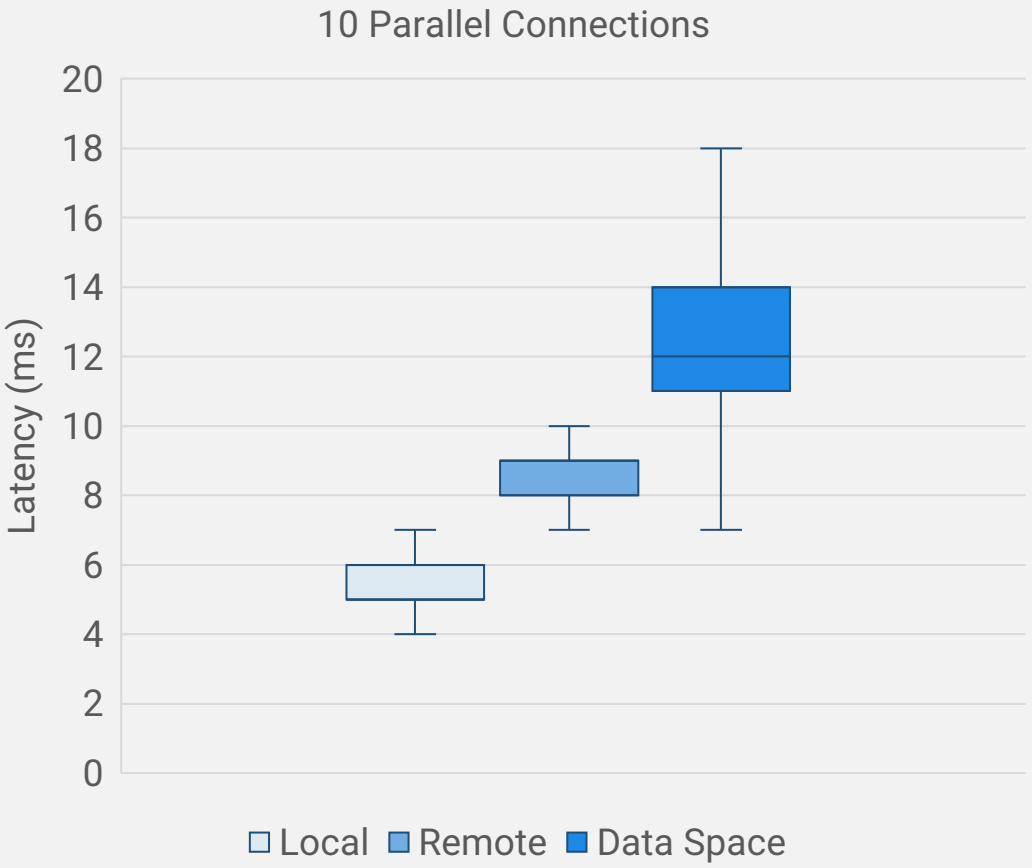


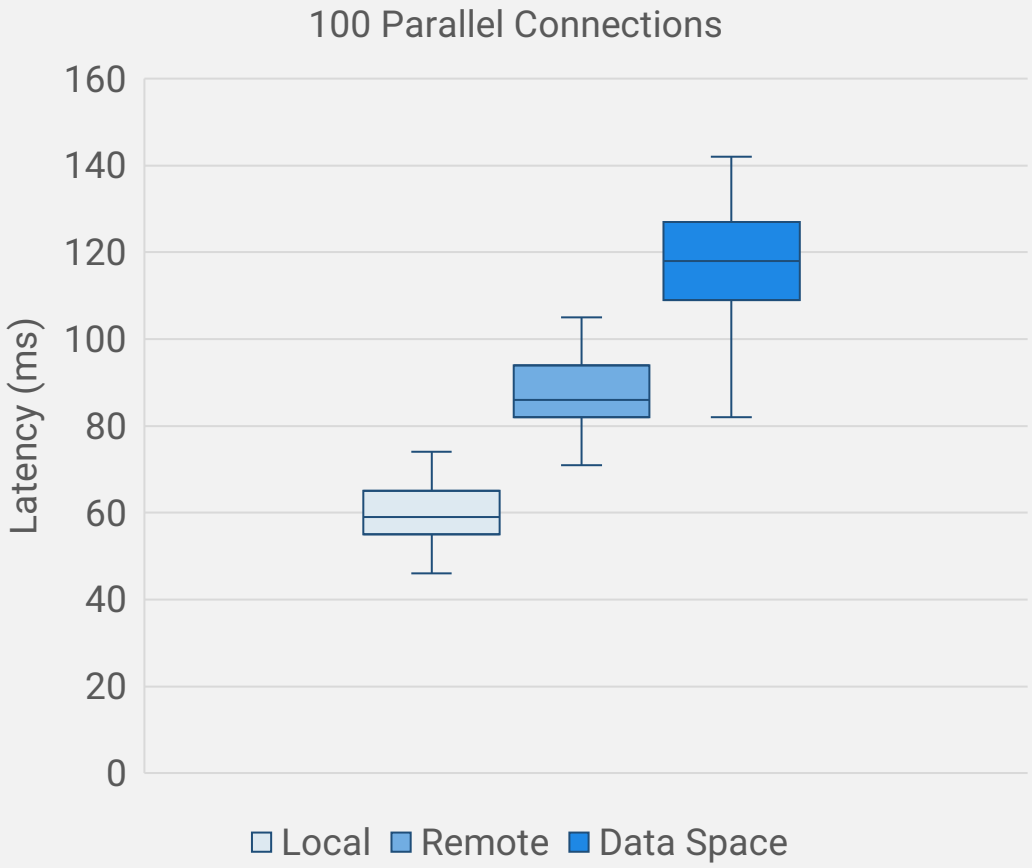
03.

Data Space



Latency

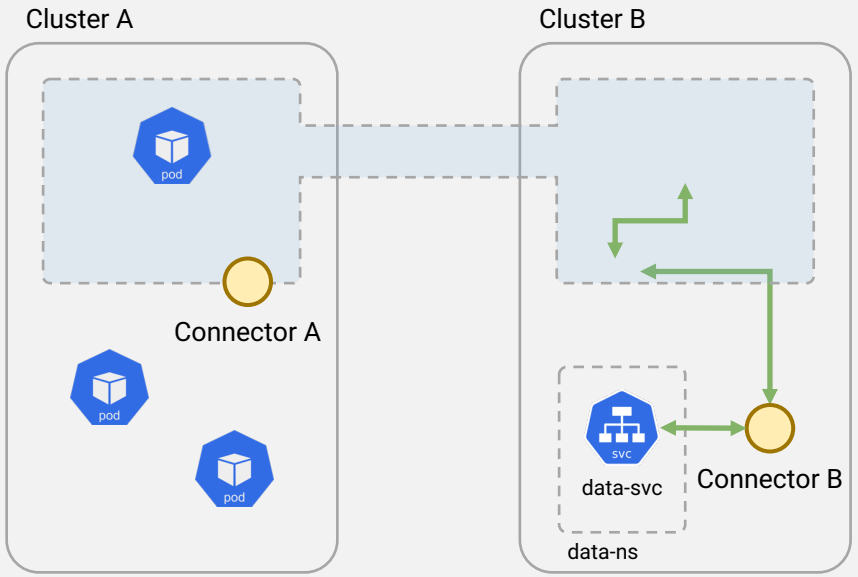

10 Parallel Connections
Latency (ms)
20
18
16
14
12
10
8
6
4
2
0
Local
Remote
Data Space


100 Parallel Connections
Latency (ms)
160
140
120
100
80
60
40
20
0
Local
Remote
Data Space

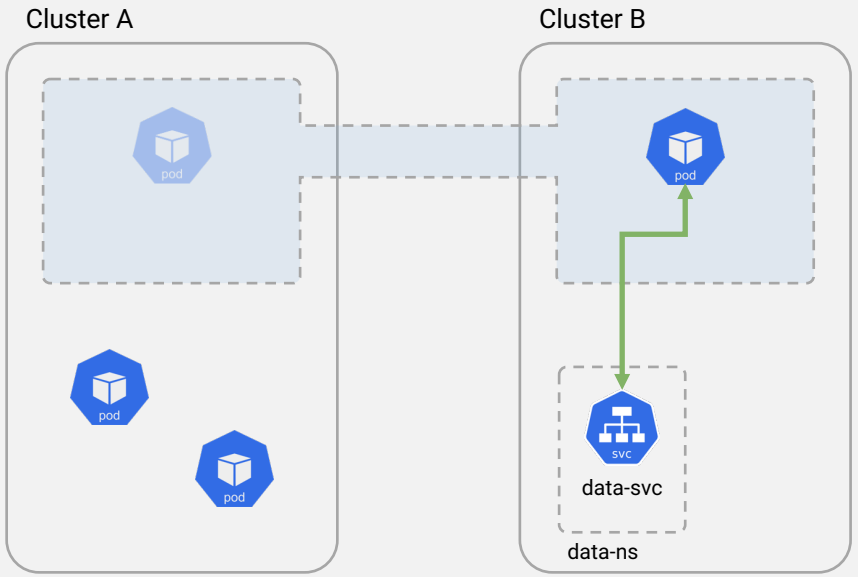
How can we integrate our solution into Gaia-X?

Proposed integration strategy with Gaia-X

In infrastructure-level data spaces, **connectors are just pods** that can be executed on either the local or remote cluster.
Liqo can be used to enhance the pure data exchange capabilities of current data space solutions, enabling the **remote deployment of processing instances**.



Liqo-enhanced Data Gravity in Gaia-X



eSAAM 2023

on Cloud to Edge Continuum

Thank you!

Sponsored by:

Organized by: