Efficient and scalable computing cloud access to Big Data EO repositories

– CREODIAS, WEkEO, CODE-DE and other platforms

25 May 2022
Session C5.02 Scalable platform architectures enabling big EO data analytics in cloud-based platform

Michał Bylicki, Monika Krzyżanowska

1. CloudFerro – cloud for EO

2. Challenges and needs come from the users

3. EO – multi platform approach

4. Best practices and lessons learned

CloudFerro

CloudFerro – who we are

- **European, private, technological company** founded in 2015, leveraging over 20 years of experience of its team members in the IT industry

- Provider of **dedicated cloud computing services**

- Delivers and operates cloud platforms for **demanding markets**, such as the European space sector, climate research and science

- Specialized at storing and processing **big data sets**, like multipetabyte repositories of Earth Observation satellite data

- Provides **flexible, open-source** based, tailor made cloud solutions, matching best technical and organizational practices to business requirements

- Achieved **technological autonomy** and guarantees reliability and independence of delivered services thanks to ful control of technology stack

CloudFerro

CloudFerro – company story

CloudFerro was founded

2015 — 8 Employees

2016 — 14 — **0,9 PB of data***

2017 — 24 — **1,5 PB**

2018 — 32 — **4,5 PB**

2019 — 56 — **13 PB**

2020 — 80 — **20 PB**

2021 — 120 — **25 PB**

2022 — >140 — **>28 PB****

* Amount of EO data publicly available for users, | ** May 2022

Challenges and needs come from the users

Policy Makers/Society:

- Strategical autonomy
- Sustainability
- Standardization
- Market competitivity
- Usage promotion

Users:

- Data availability
- Several platforms, clouds and data sources
- Data consistency (temporal and spatial)
- Diverse access methods
- Download times/reliability
- Limited processing capacity
- Tools availability
- Vendor lock-in
- Complex data licensing

Data providers:

- Growing data sizes
- Growing number of datasets
- Downlink bandwidth
- Reach customers
- Keep costs reasonable

CloudFerro

EO Data consumption models

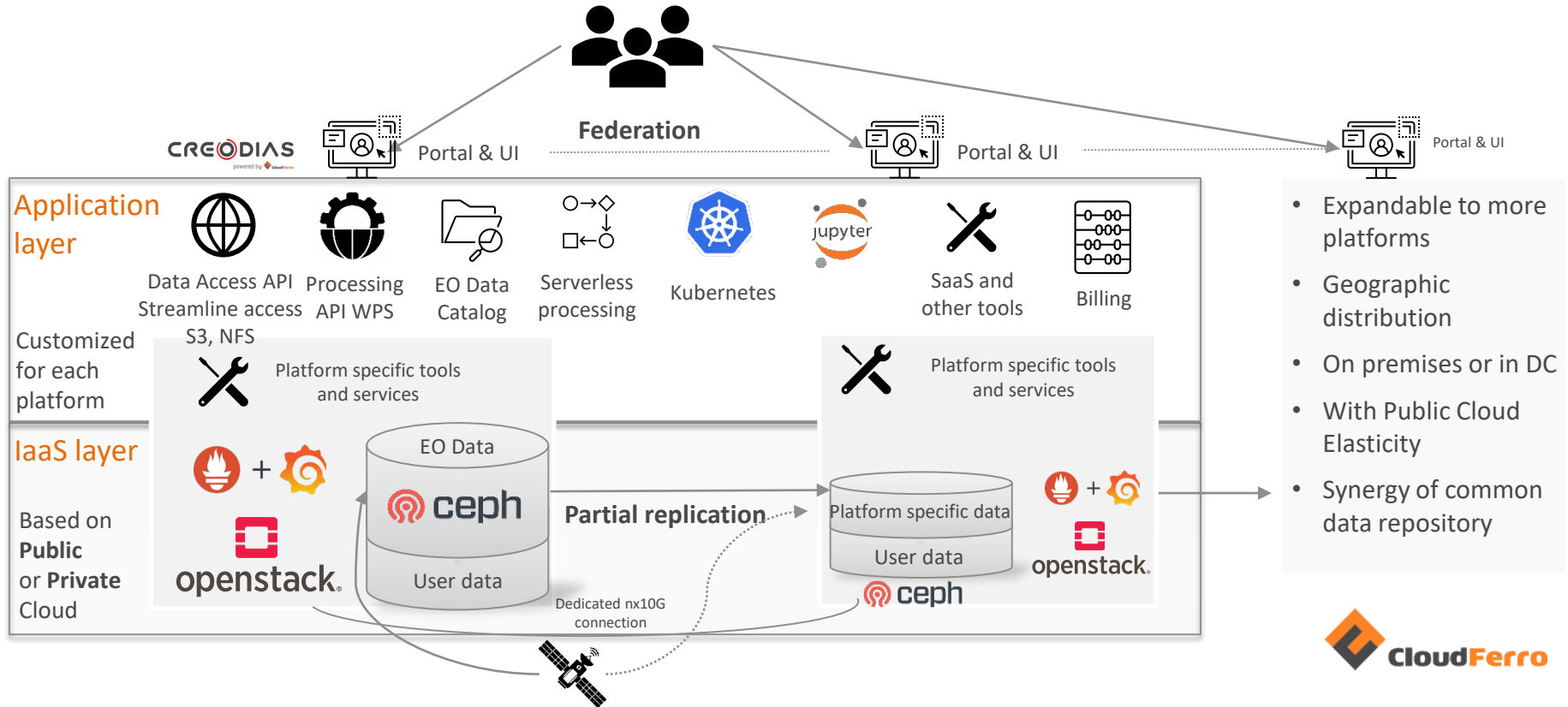Browse & View

Download (or push) & process locally

Process in cloud using IaaS infrastructure

Process in cloud using predefined processors

CloudFerro

Outline of multi-platform approach

Federation

Portal & UI

Portal & UI

Portal & UI

Application layer

Customized for each platform

Data Access API
Streamline access
S3, NFS

Processing API WPS

EO Data Catalog

Serverless processing

Kubernetes

Jupyter

SaaS and other tools

Billing

Platform specific tools and services

Platform specific tools and services

IaaS layer

Based on **Public** or **Private** Cloud

EO Data

ceph

User data

openstack

Partial replication

Platform specific data

User data

ceph

openstack

Dedicated nx10G connection

- Expandable to more platforms
- Geographic distribution
- On premises or in DC
- With Public Cloud Elasticity
- Synergy of common data repository

CloudFerro

Best practices/Lessons learned

Provide lots of scalable bandwidth and processing power

Store (almost) everything you have observed or generated online

Provide harmonized data in standard, easily accessible formats

Federate users and data access

Use and contribute to open-source

Save Energy and resources

CloudFerro

Thank you
for your attention

Michal Bylicki
mbylicki@cloudferro.com

For more information,
please visit: www.cloudferro.com
or link up with us on:

facebook.com/cloudferro   linkedin.com/company/clfr/   twitter.com/CloudFerro