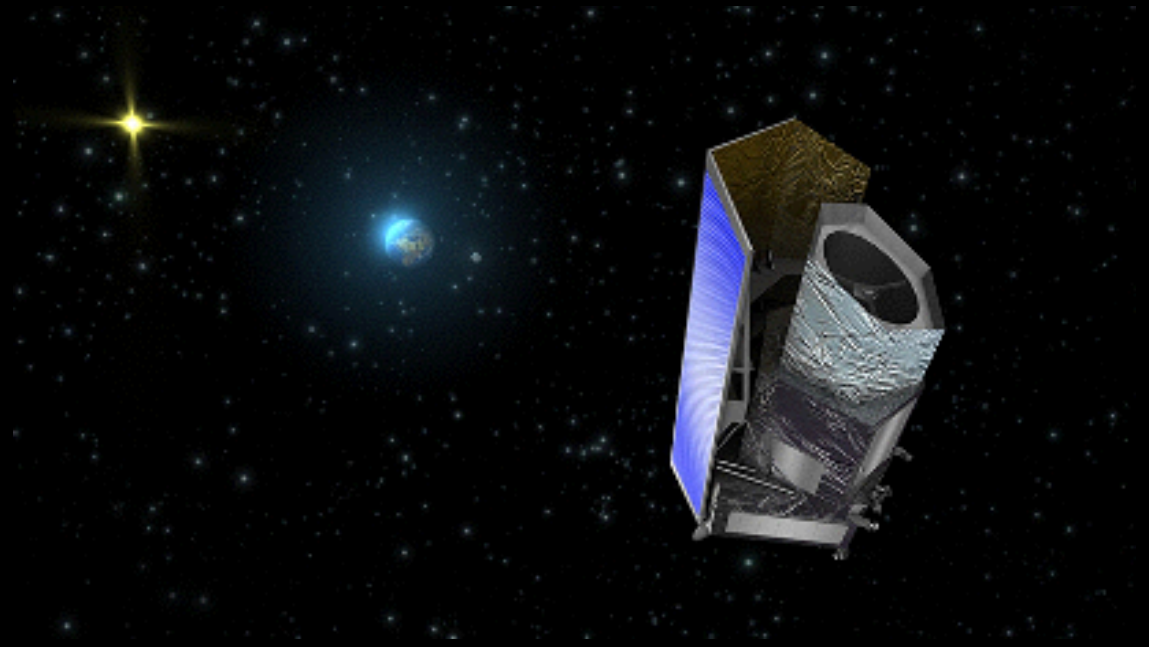
Using CernVM-FS to deploy Euclid processing S/W on Science Data Centres



M. Poncet (CNES)
Q. Le Boulc'h (IN2P3)
M. Holliman (ROE)
On behalf of Euclid EC SGS System Team

ADASS 2016

Outline

- Euclid Project & SGS
- From Euclid pipeline to SGS Architecture
- From source code to processing nodes
- CernVM-FS principles
- Using CernVM-FS on SDCs

- **Euclid Project & SGS**
- From Euclid pipeline to SGS Architecture
- From source code to processing nodes
- CernVM-FS principles
- Using CernVM-FS on SDCs

Euclid project

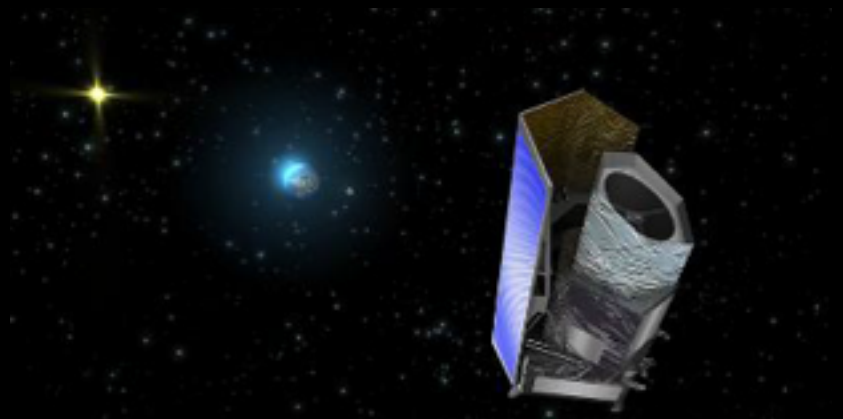
M2 mission in the framework of the **ESA Cosmic Vision Programme**

Euclid mission objective is to map the geometry and understand the nature of the dark Universe **(dark energy and dark matter)**

Actors in the mission: **ESA** and the **Euclid Consortium** (institutes from 15 European countries and USA, funded by their own national Space Agencies)

Euclid Consortium:
15 countries
100+ labs
1200+ members

One of the biggest collaboration!



For more information see :
O1.2, O8.5, P3.5, P8.13, P1.16, P2.13, F

http://sci.esa.int/science-e/www/area/index.cfm?fareaid=102
http://www.euclid-ec.org

Euclid Science Ground Segment (SGS)

Euclid

Ground Station

SDC

MOC
ESOC

DDS

SOC
ESAC

scientific community

Public Archive

EAS

External data
(DES, KiDS, ...)

Project Office

System Team

Simulation
OU-SIM

SDC SDC SDC SDC SDC SDC SDC SDC SDC

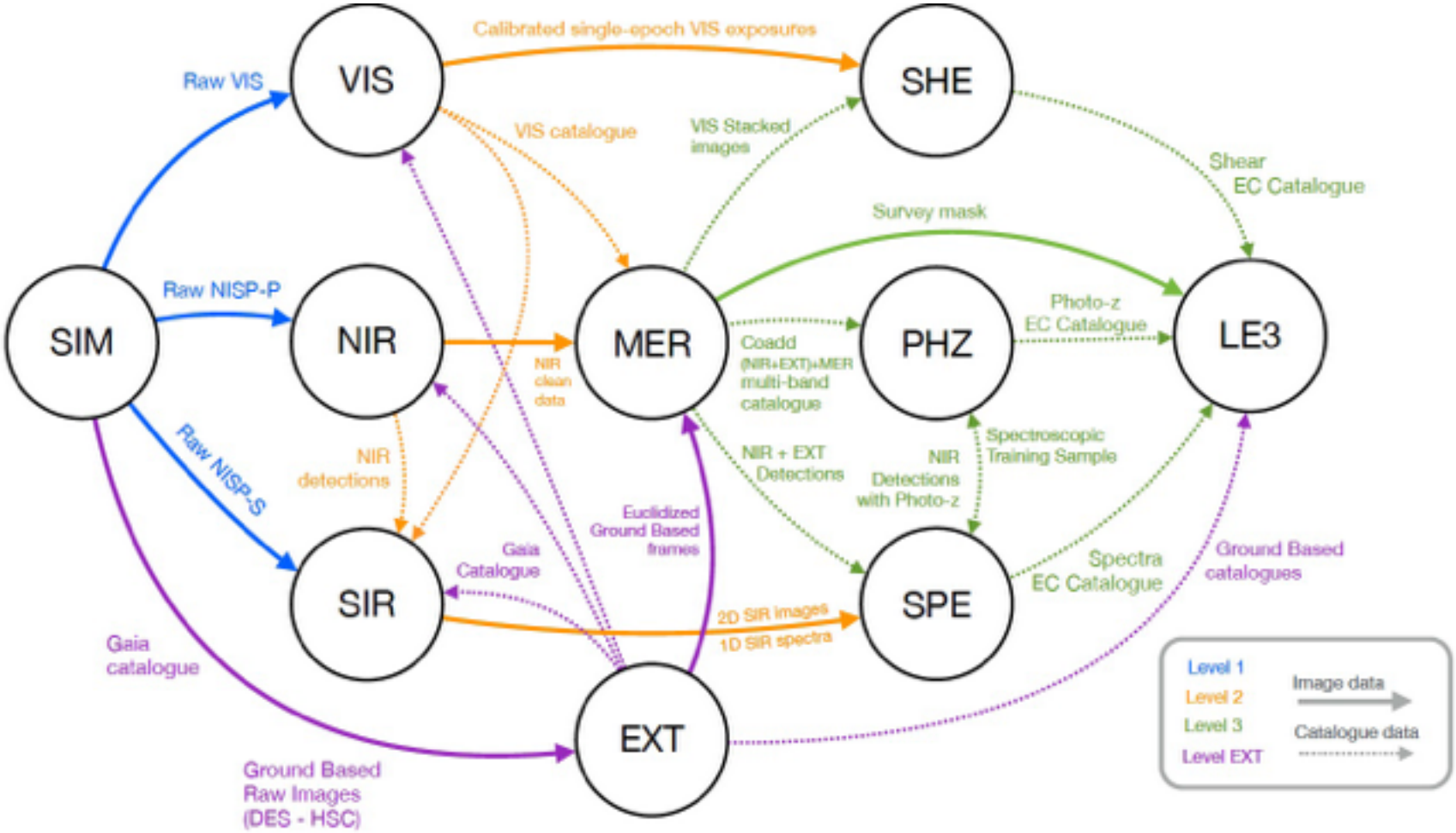
OU-PHZ	OU-SHE	OU-VIS	OU-LE3	OU-SPE	OU-SIR	OU-MER	OU-NIR	OU-EXT
Phot Red Sh.	Morpho & Shear	VIS Imag	Level 3	Spectro Meas	NIR Spectro	Euclidisation	NIR Imag	Ext Data

OU coordinator

OU deputy coordinator

- Euclid Project & SGS
- **From Euclid pipeline to SGS Architecture**
- From source code to processing nodes
- CernVM-FS principles
- Using CernVM-FS on SDCs

Euclid pipeline



From dedicated SDCs to Federated SDCs

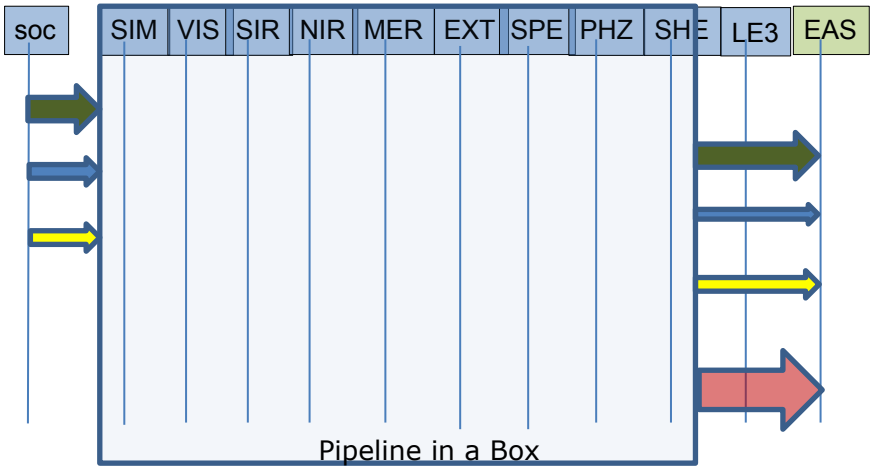
Dedicated SDCs

Unbalanced load

100+ PB in grand total

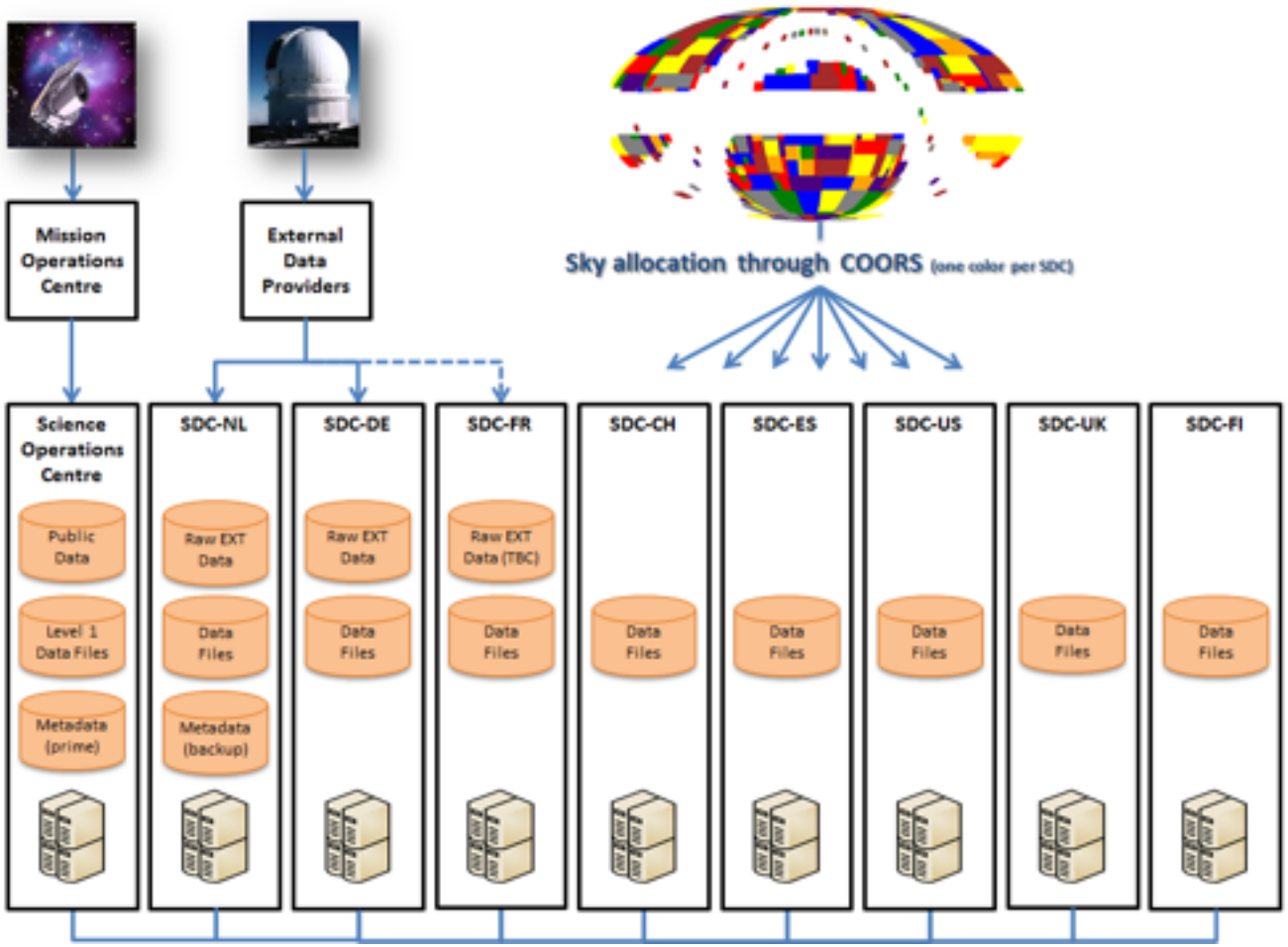
More Data transfer than processing ?

Whole pipeline in any SDC on quantum of data



Move the code Not the data paradigm

Move the code not the Data

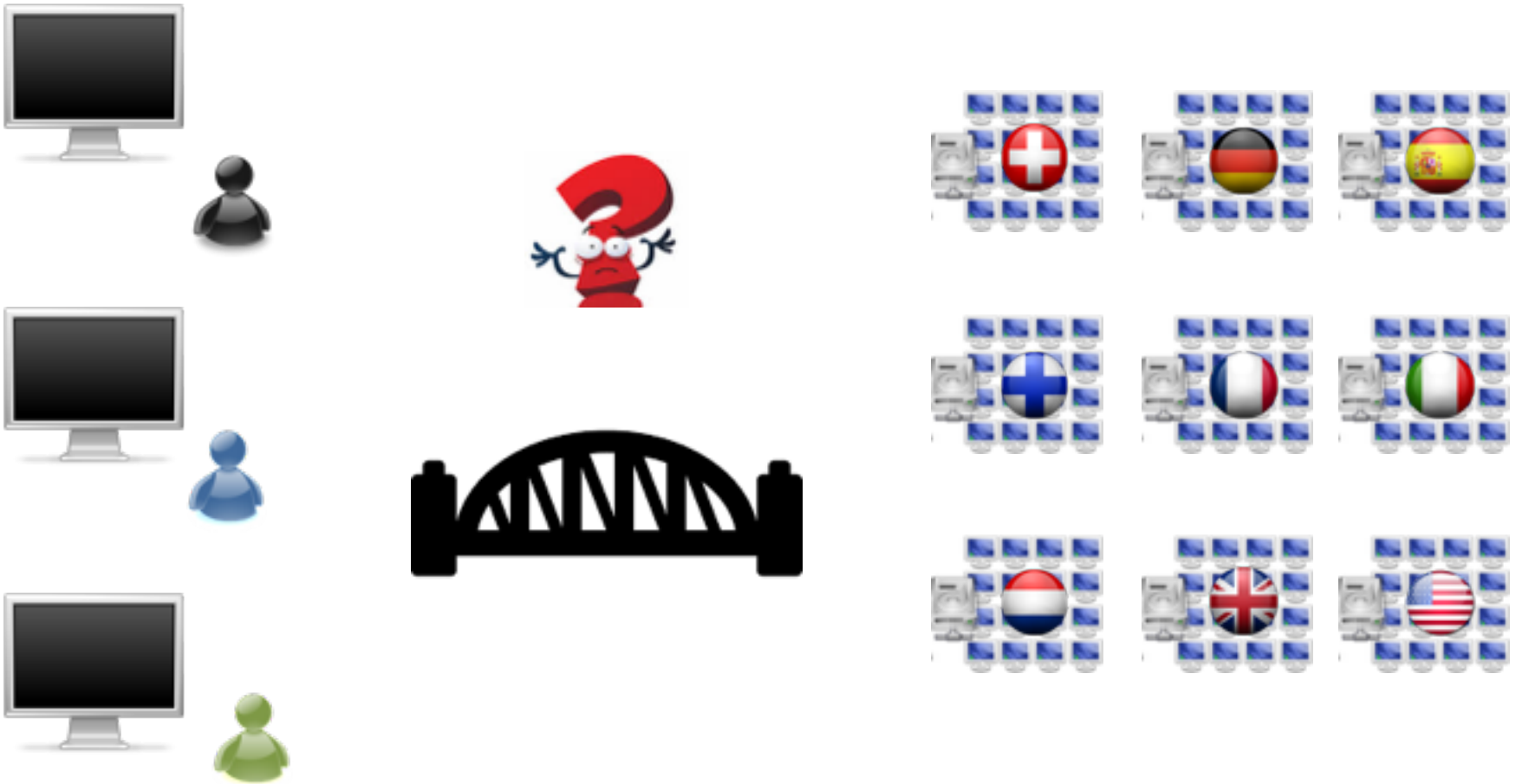


SDCs are both storage and processing nodes

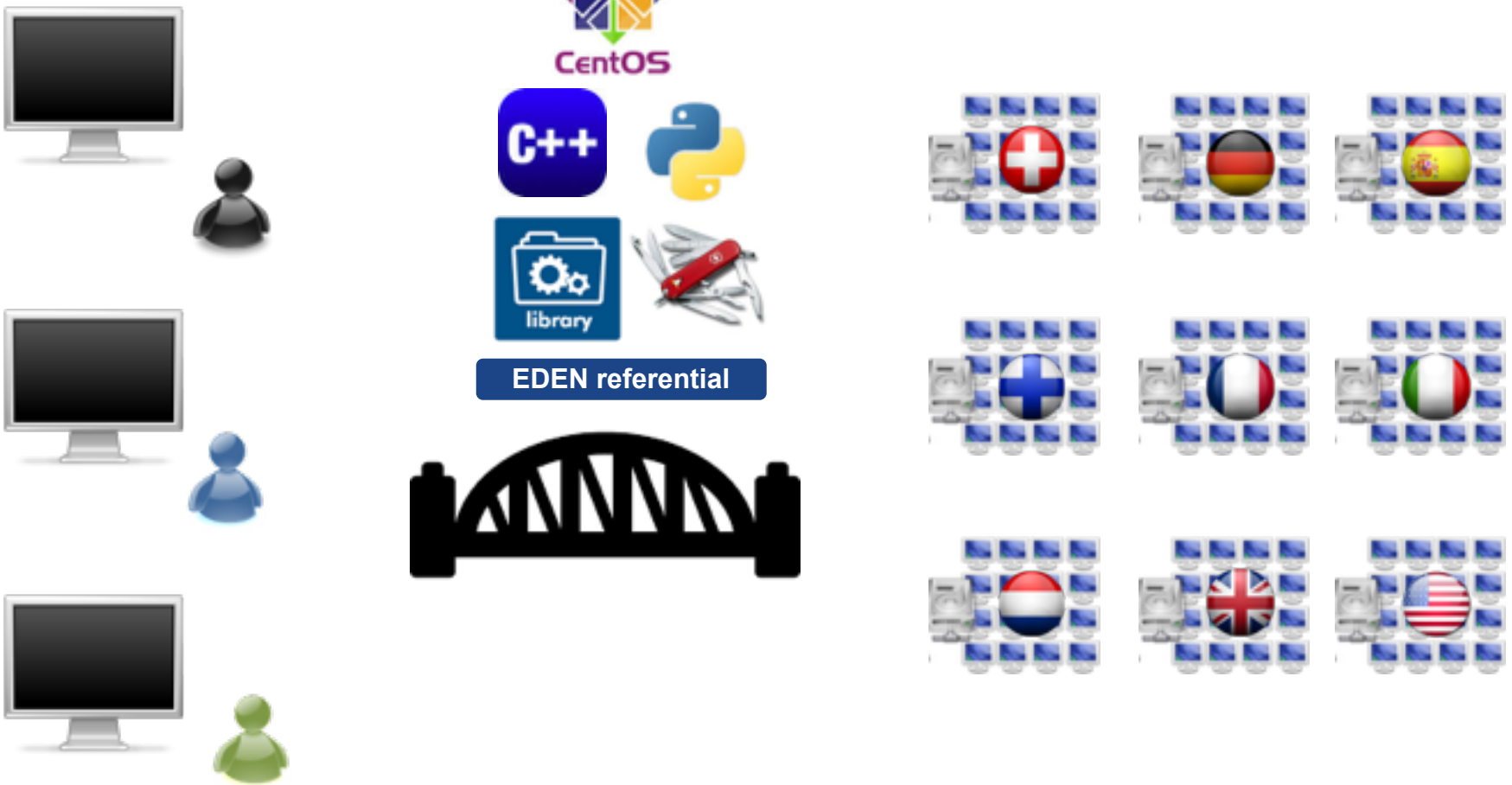
Data storage and processing allocated by sky area

- Euclid Project & SGS
- From Euclid pipeline to SGS Architecture
- **From source code to processing nodes**
- CernVM-FS principles
- Using CernVM-FS on SDCs

From source code to production



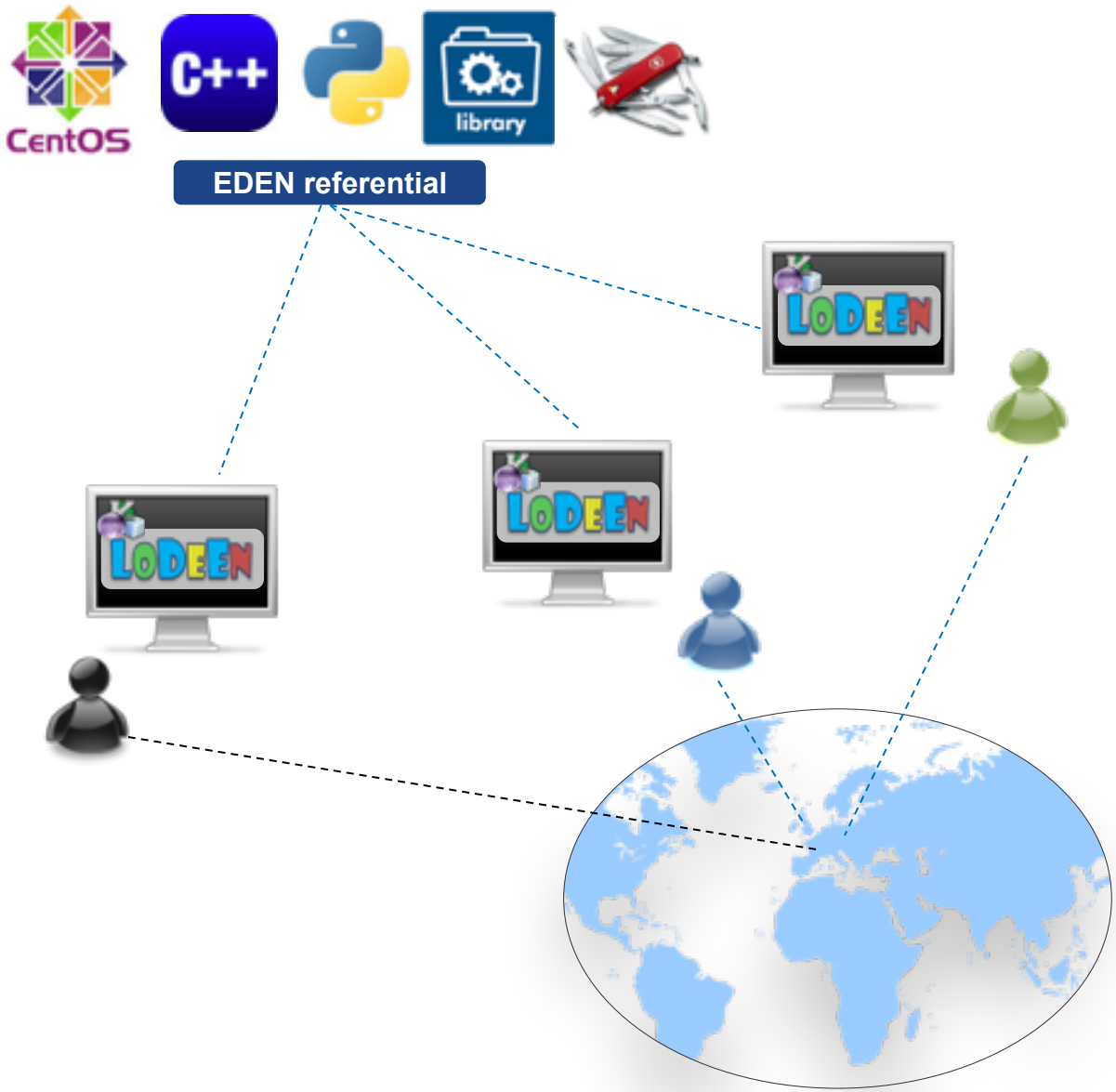
From source code to production



...in collaboration with an international team...

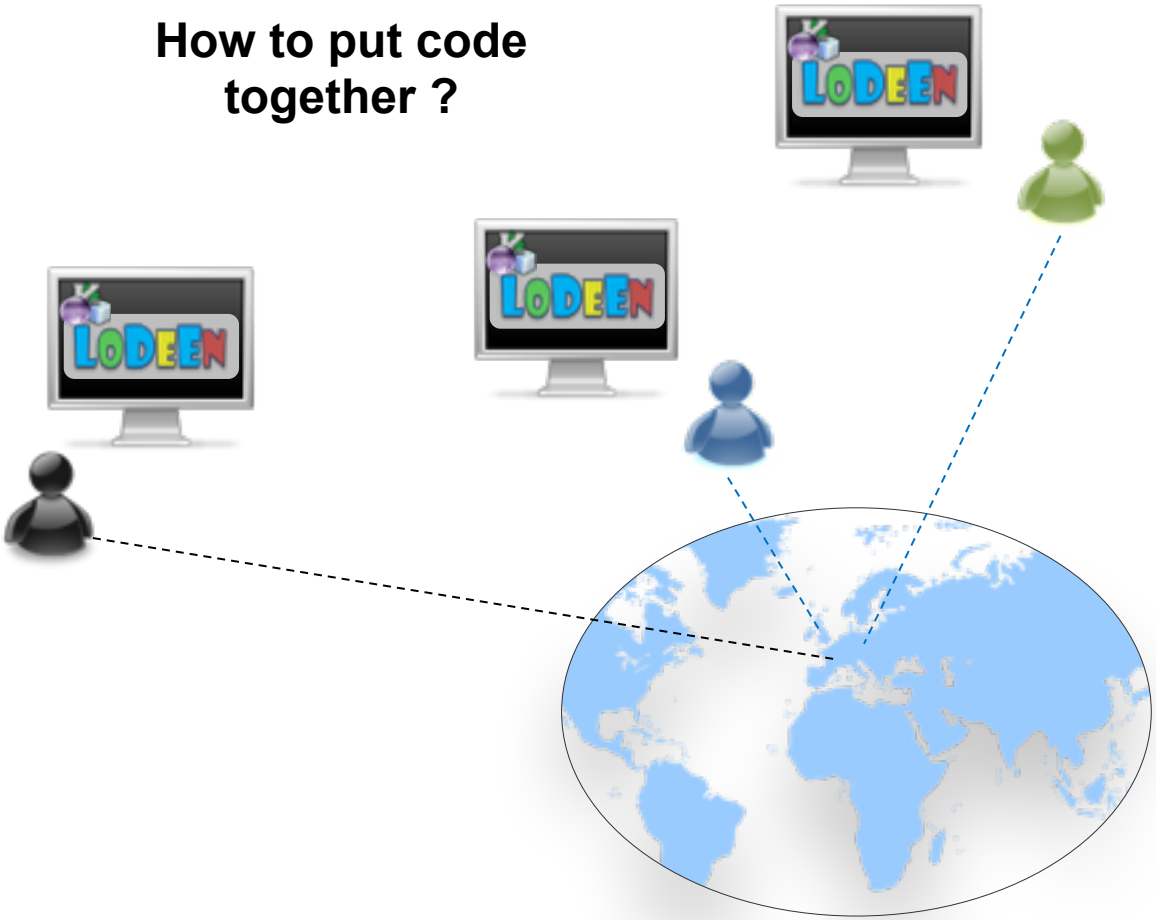
How to federate ?

...in collaboration with an international team...



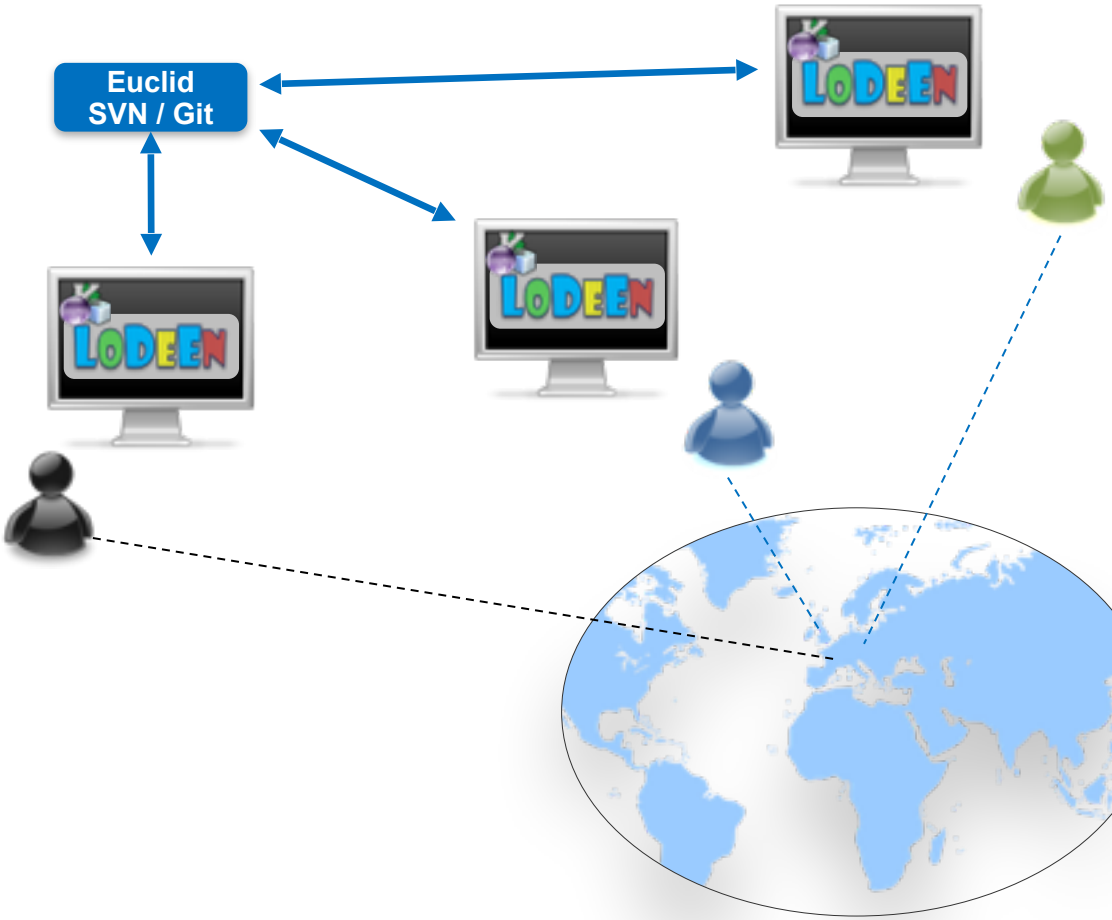
...in collaboration with an international team...

How to put code together ?



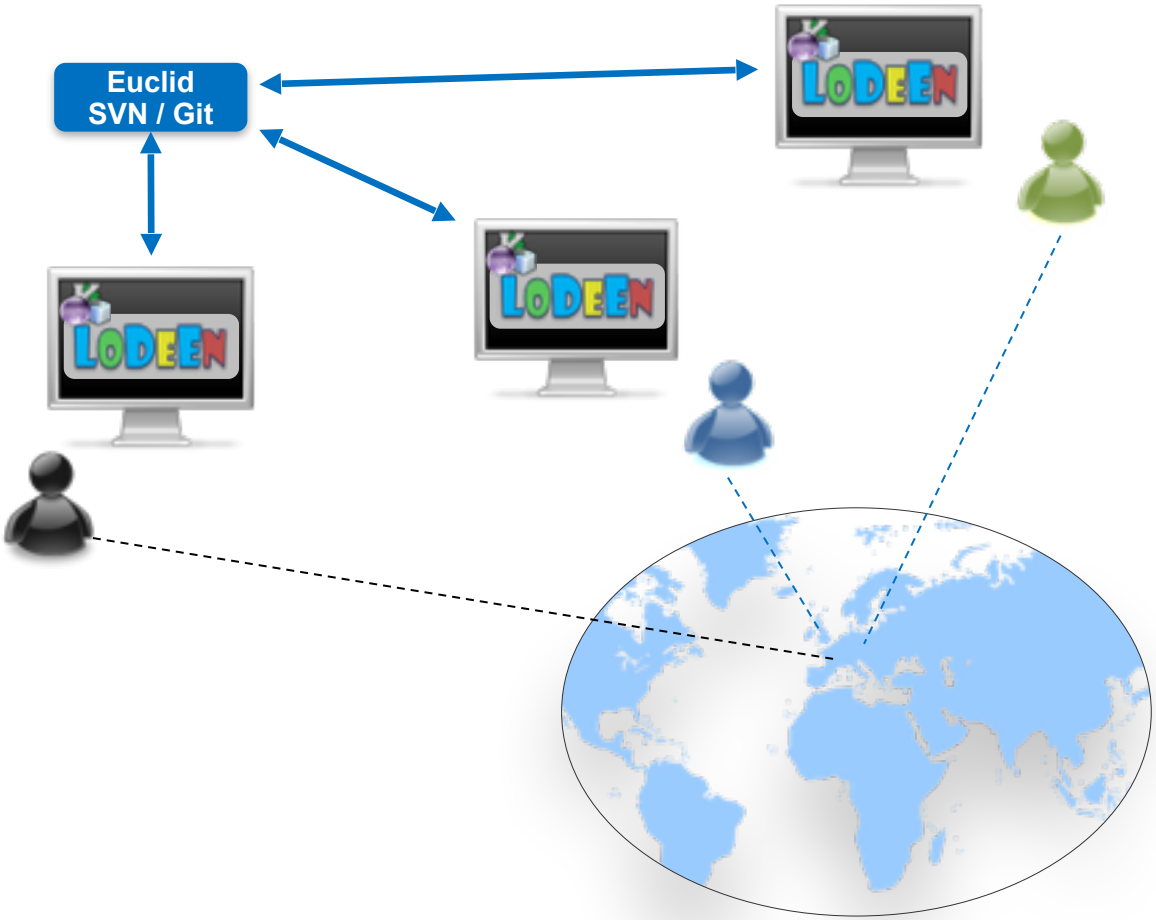
...in collaboration with an international team...

Configuration management



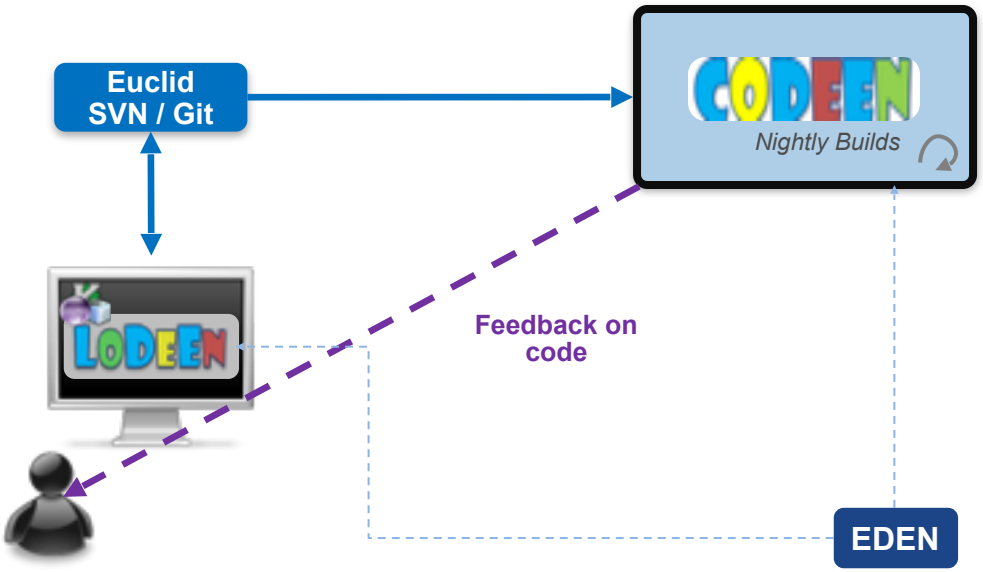
...in collaboration with an international team...

How to build… ?

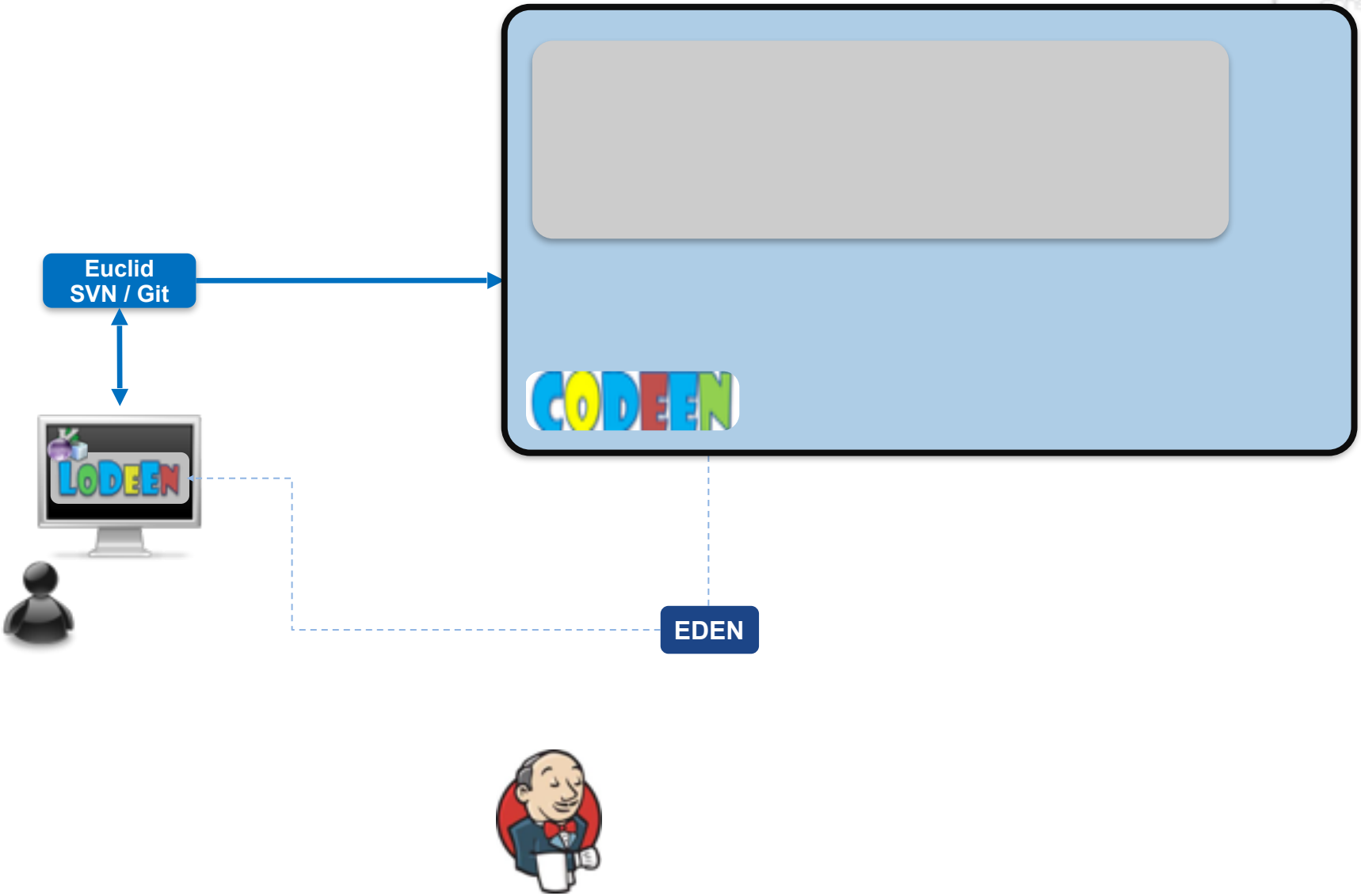


…continuously integrating its code…

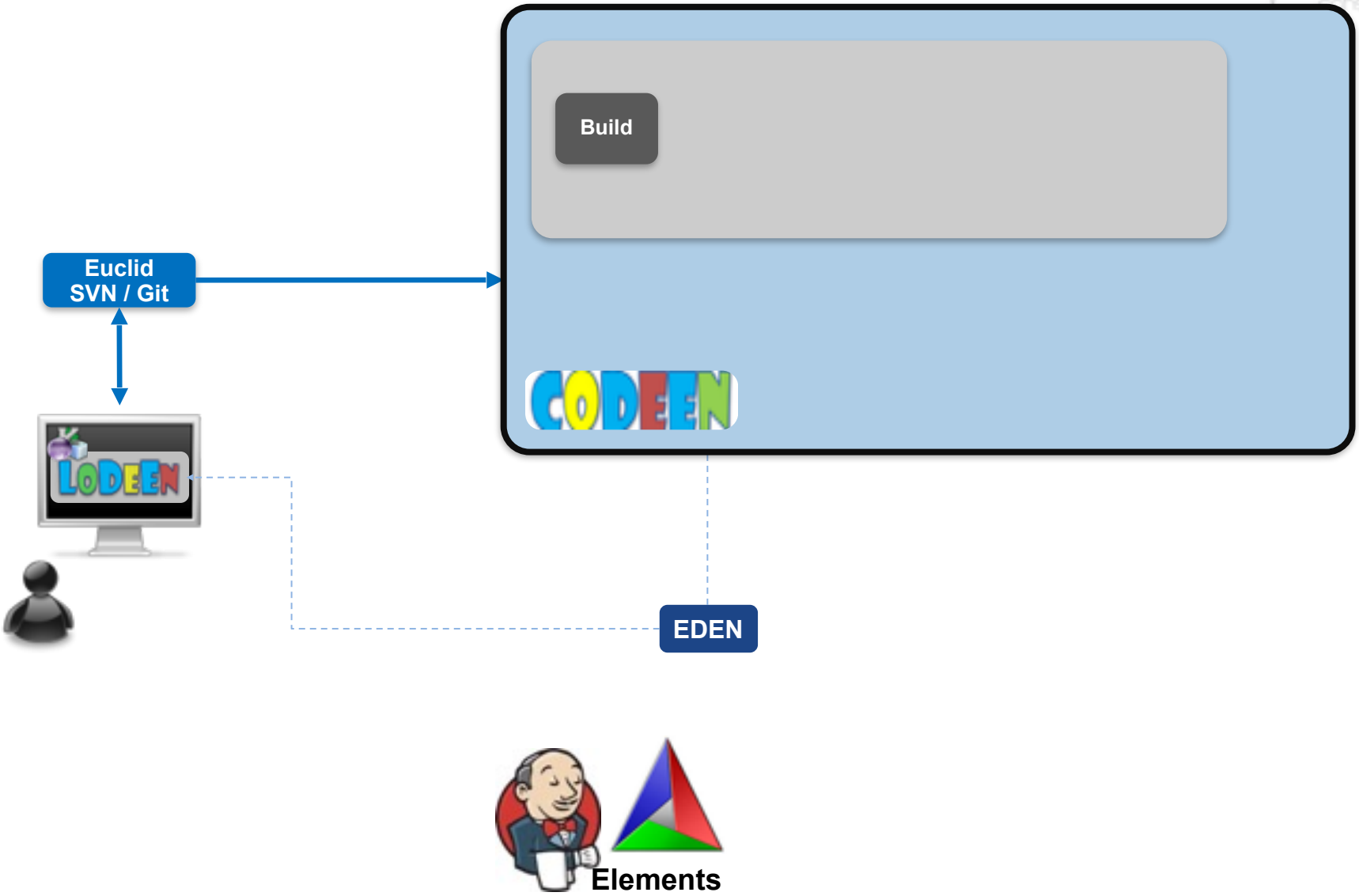
Continuous integration



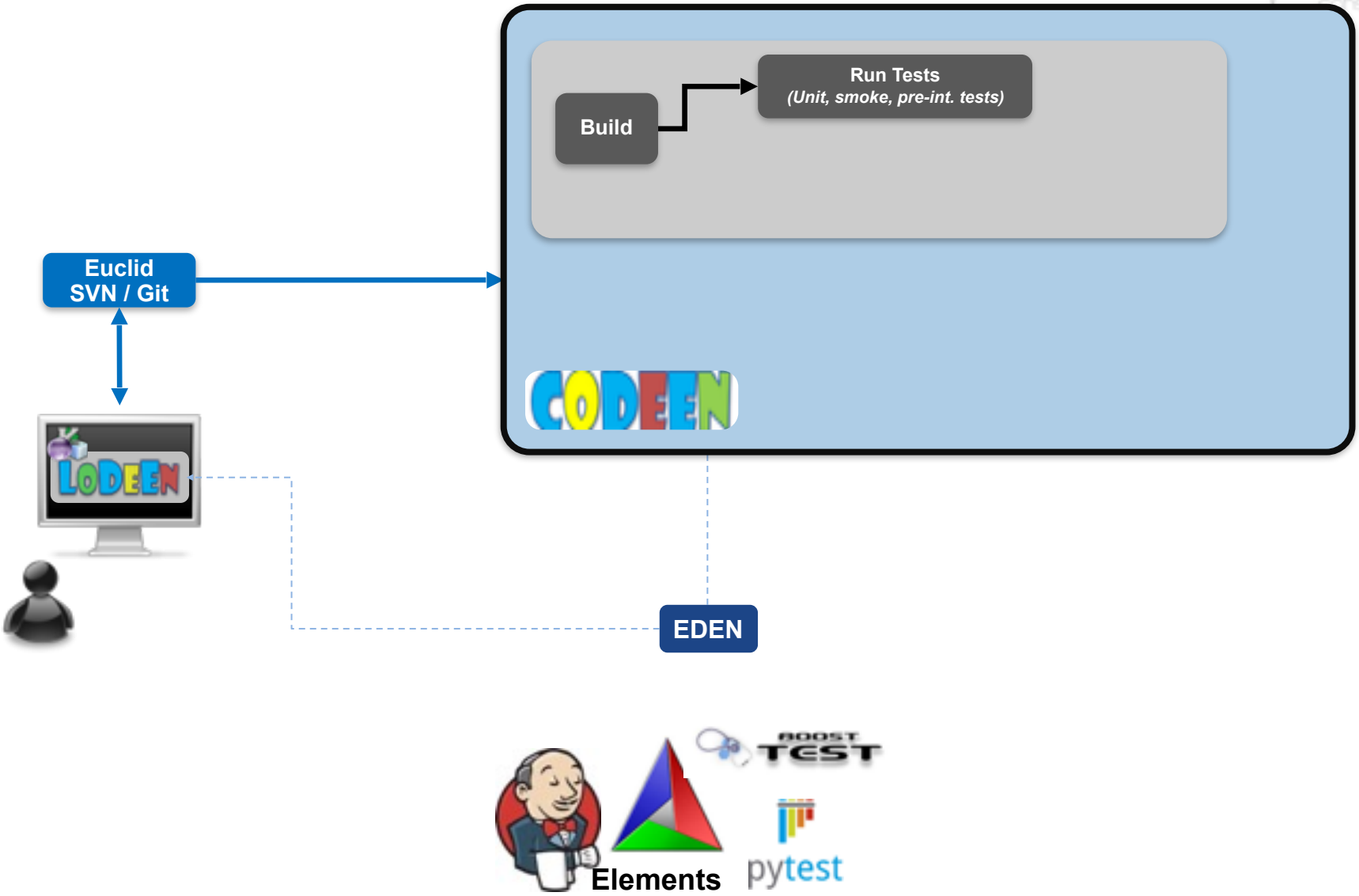
…continuously integrating its code…



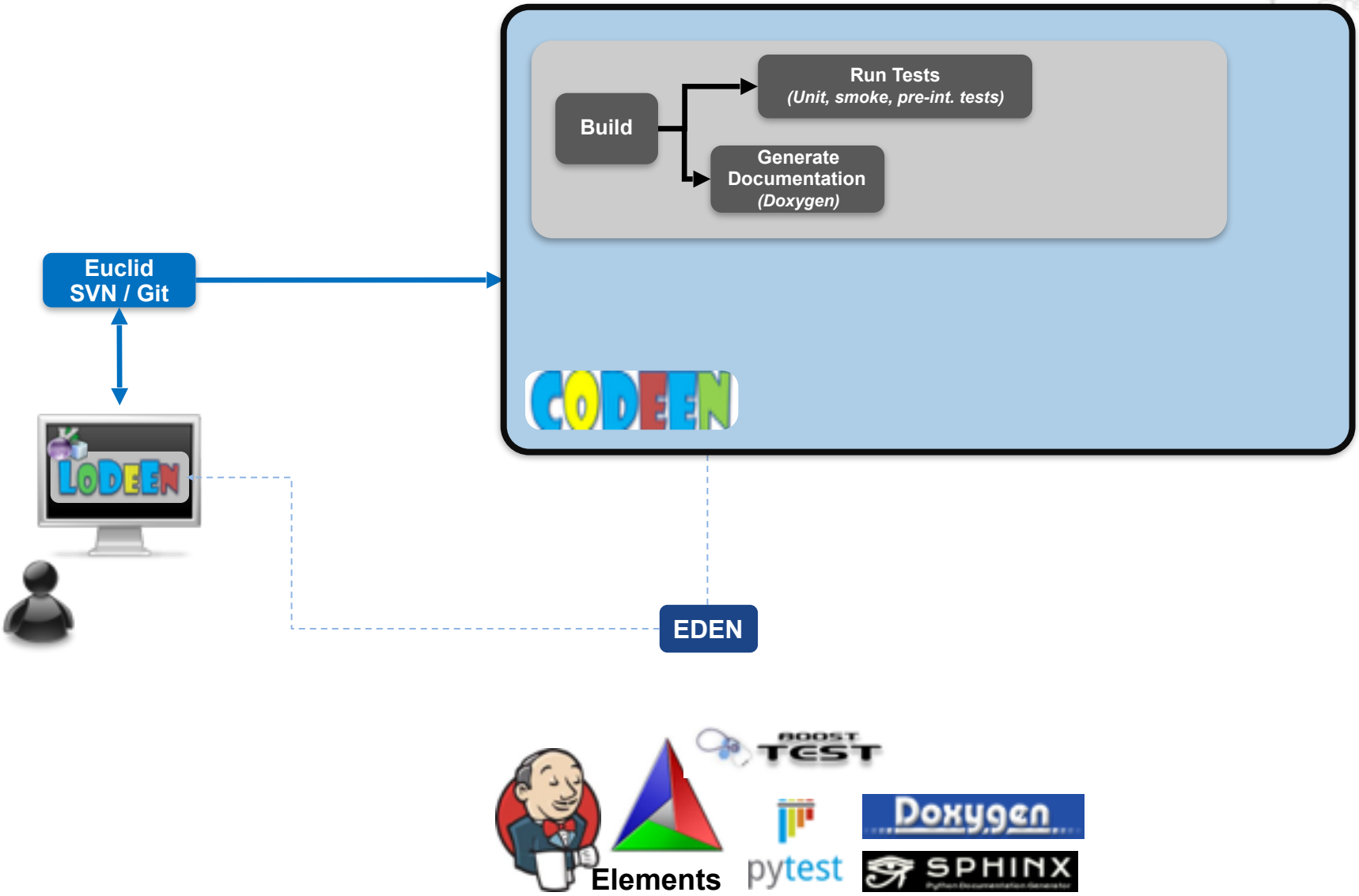
…continuously integrating its code…



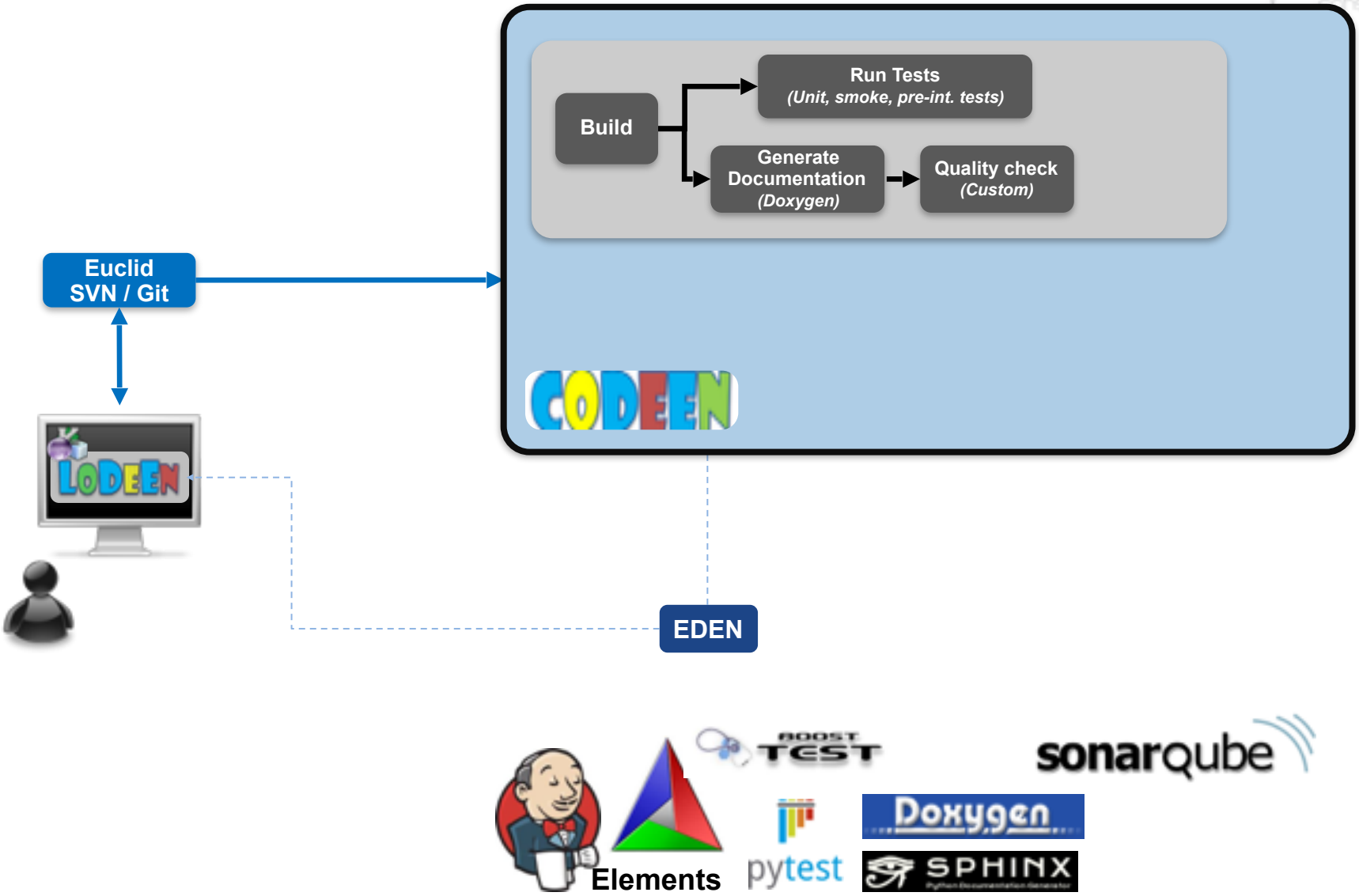
…continuously integrating its code…



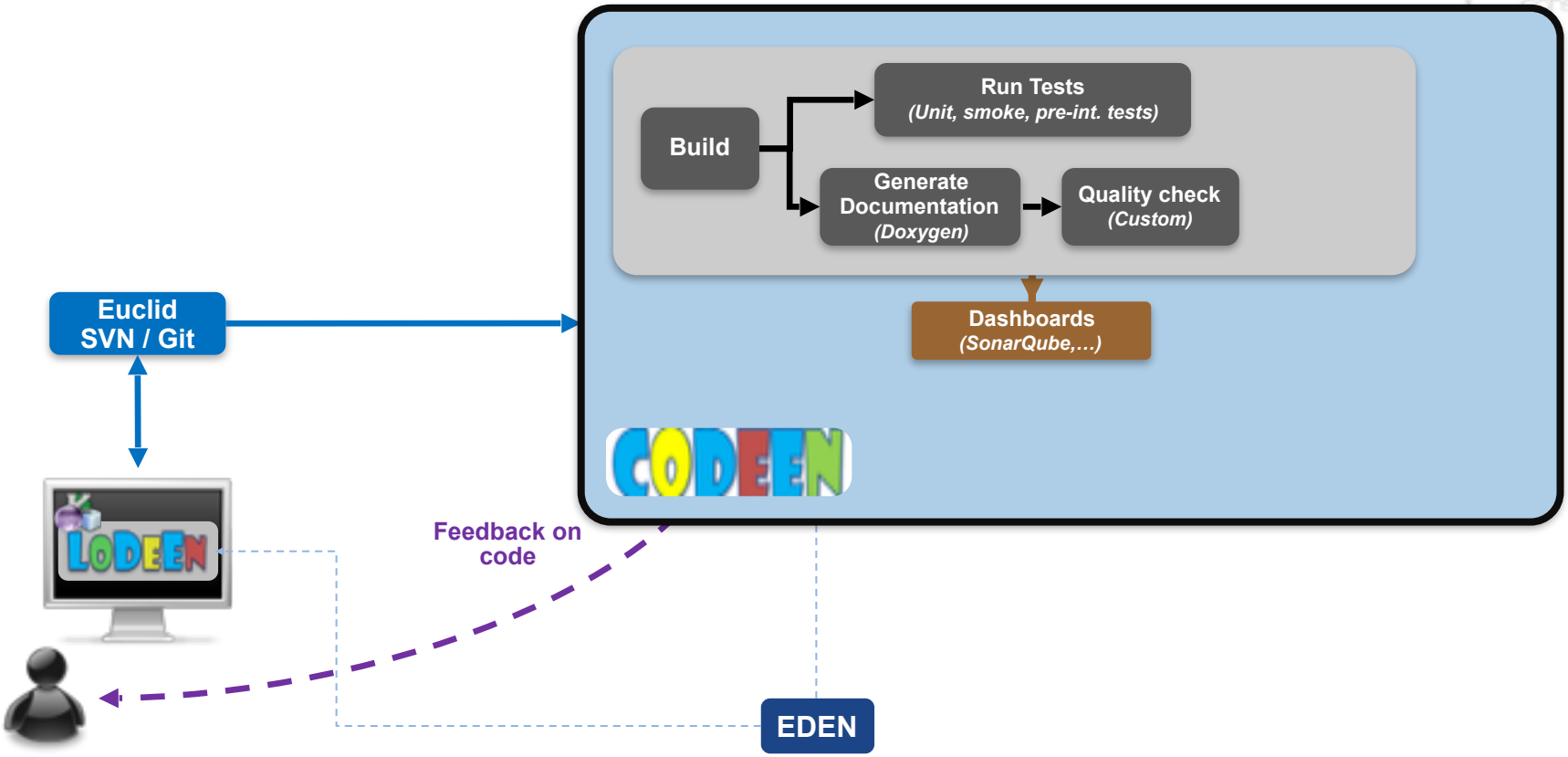
…continuously integrating its code…



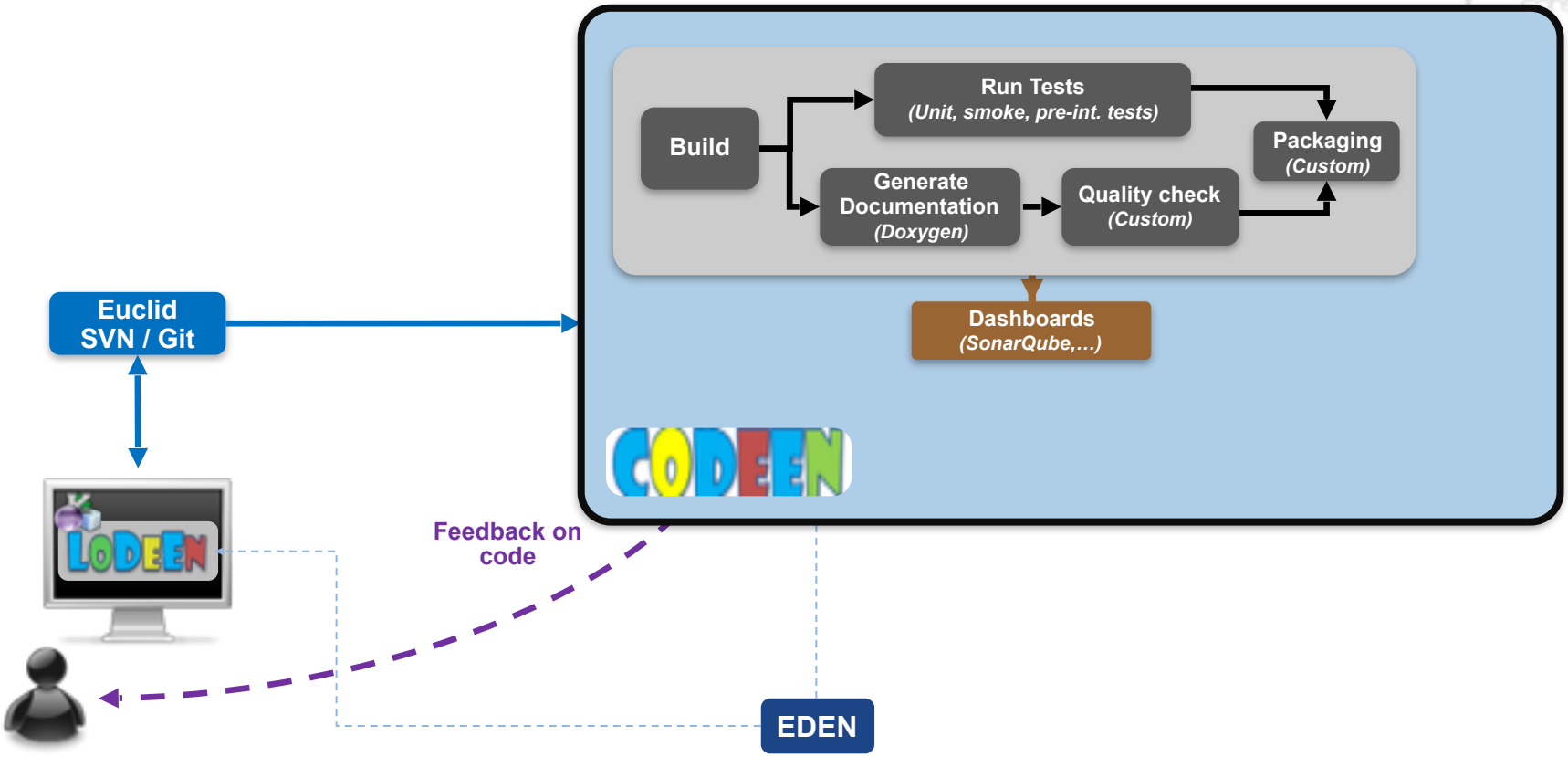
…continuously integrating its code…



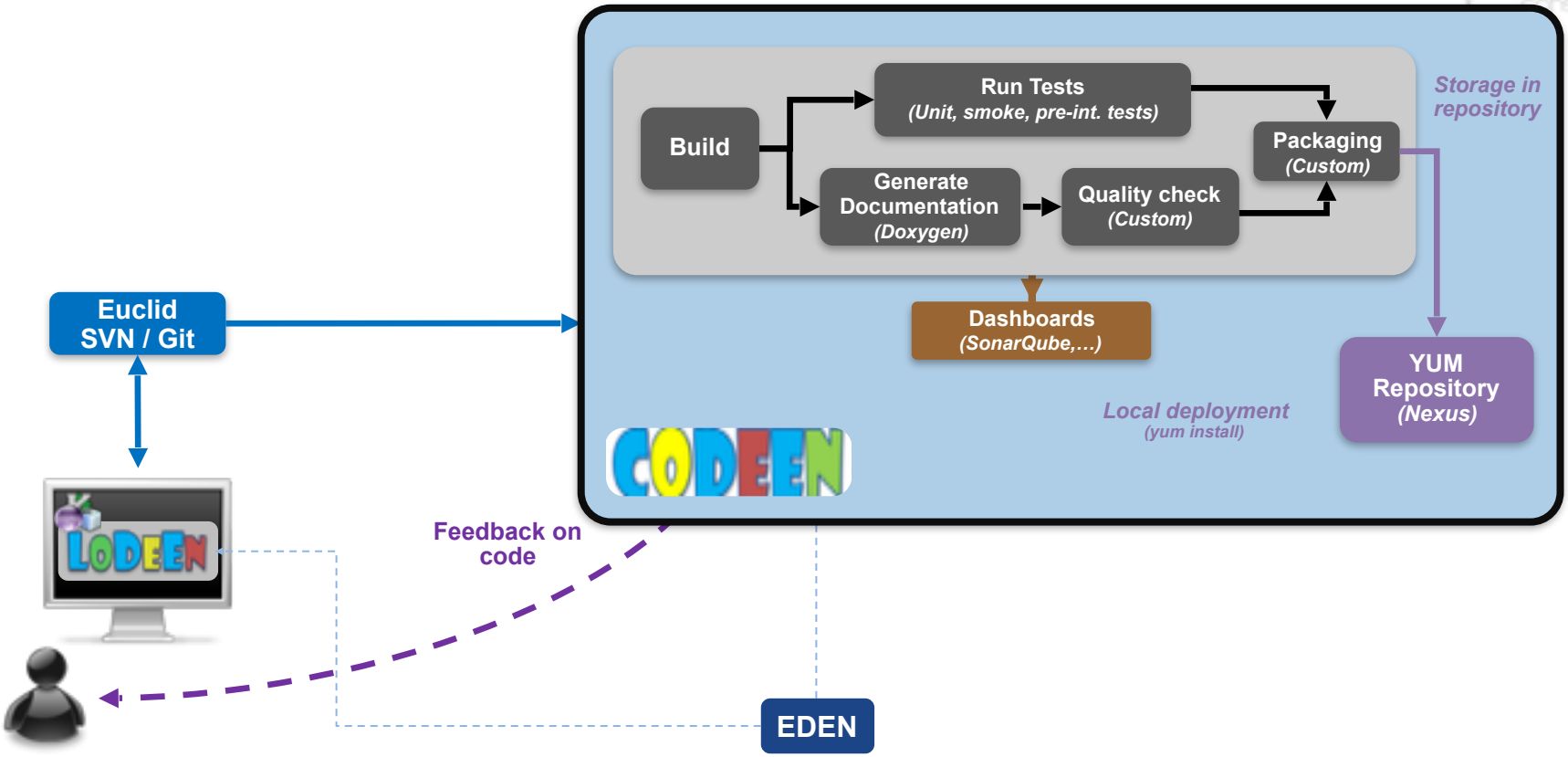
…continuously integrating its code…



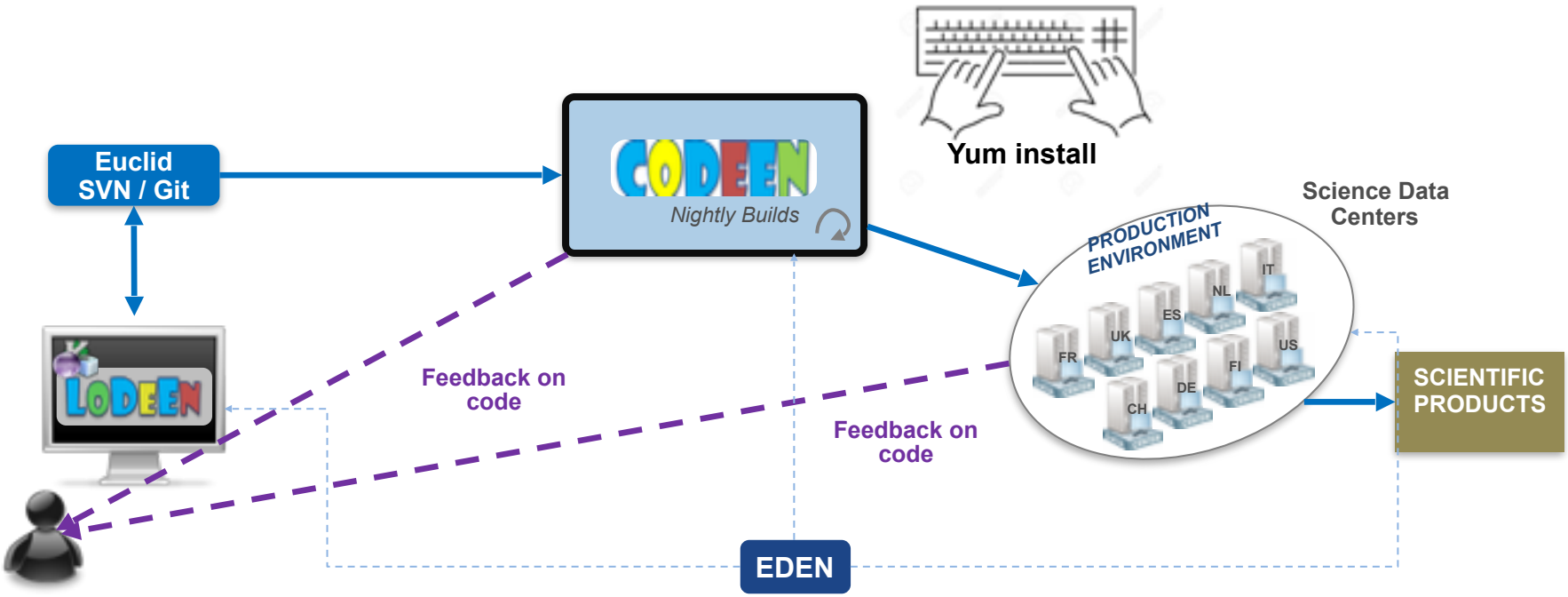
…continuously integrating its code…



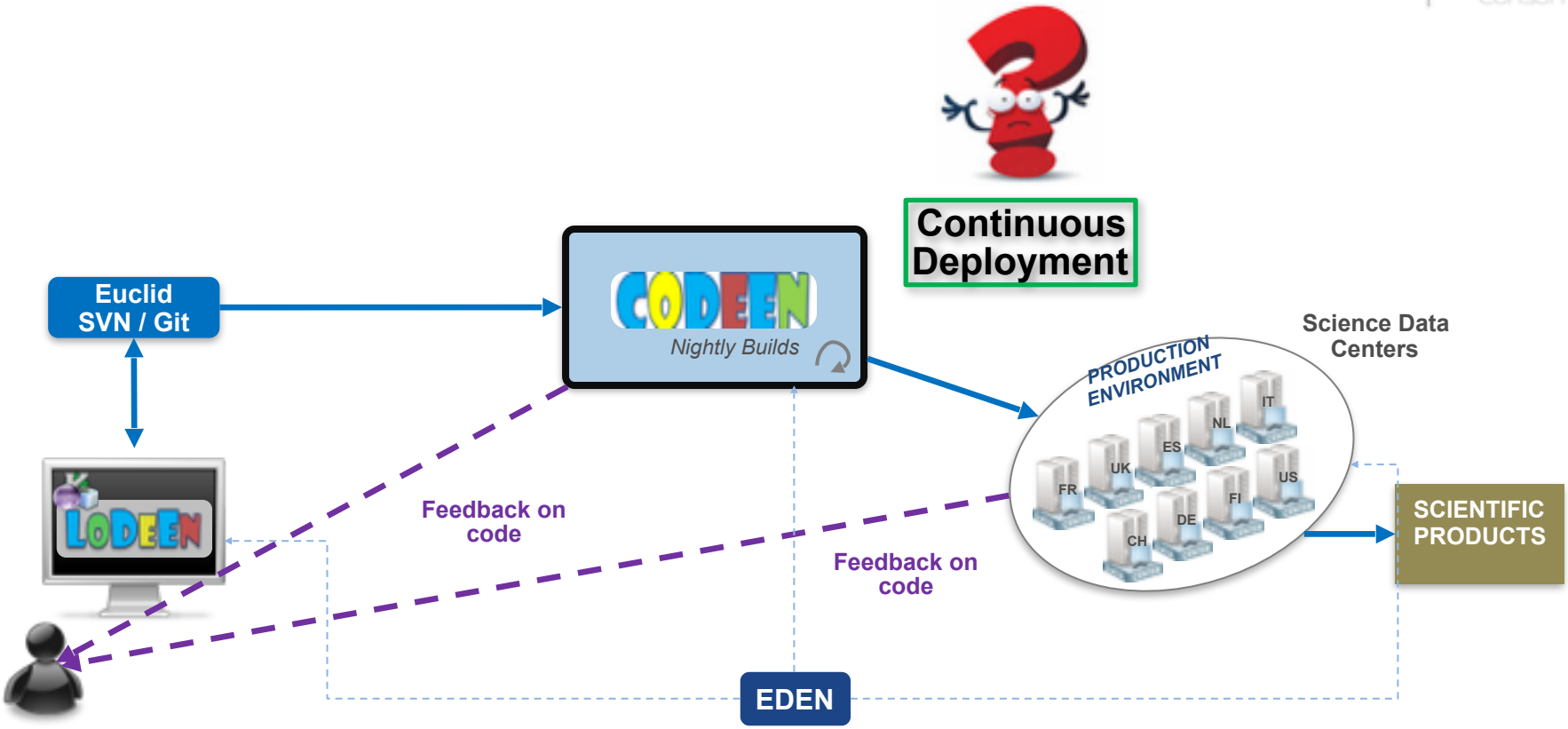
…continuously integrating its code…



…able to quickly deploy it on production…



…able to quickly deploy it on production…



« Dream » Solution

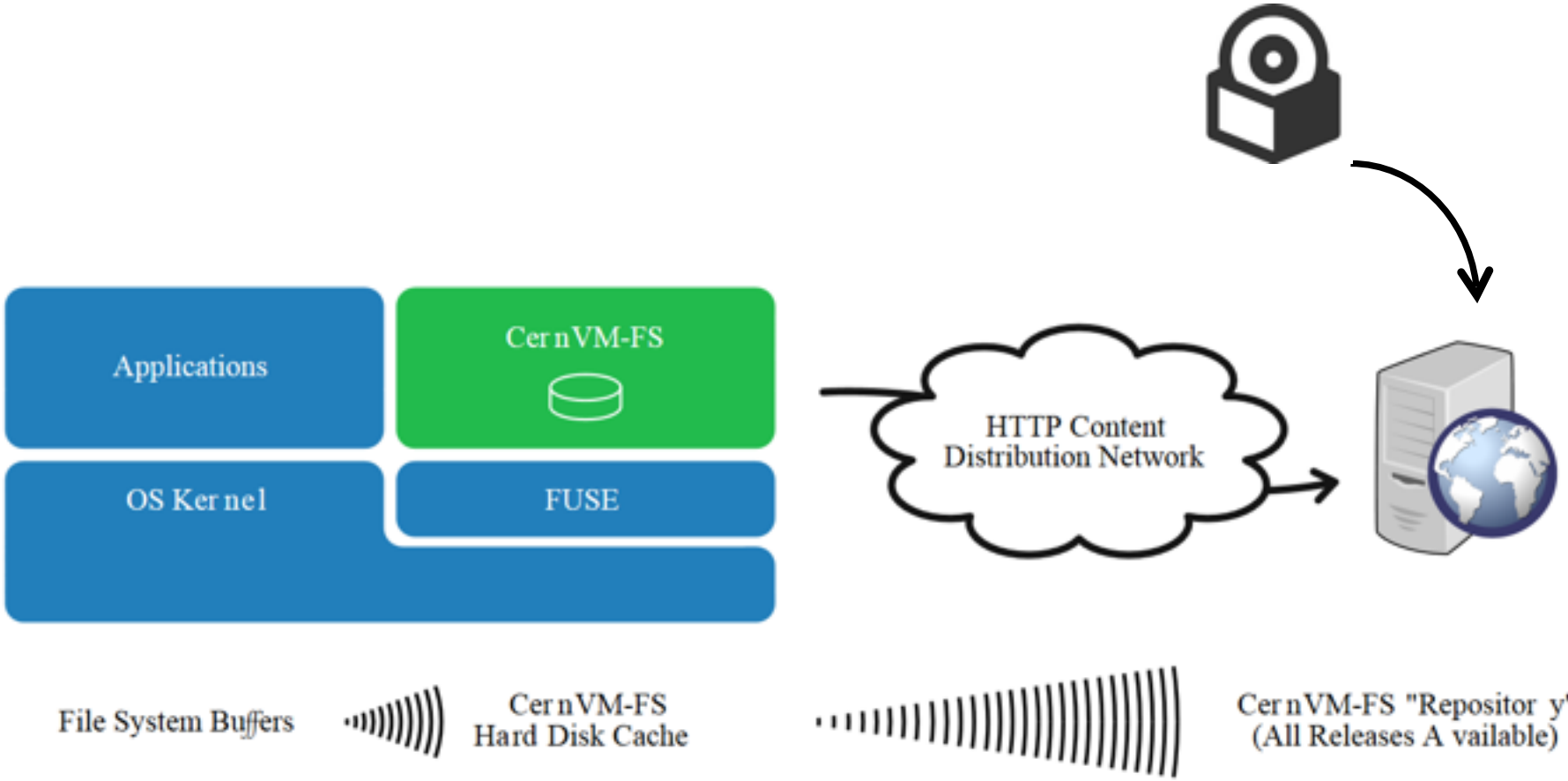
- not (too) intrusive
- Automatic deployment
  - No (less) admin intervention required
- Security friendly
  - No exotic protocol (easy to filter)
  - Outgoing connexion prefered (no incoming)
- Mutiple versions in //
  - EDEN versions
  - Pipeline versions
- Efficient

- Euclid Project & SGS
- From Euclid pipeline to SGS Architecture
- From source code to processing nodes
- **CernVM-FS principles**
- Using CernVM-FS on SDCs

CernVM File System (CernVM-FS)

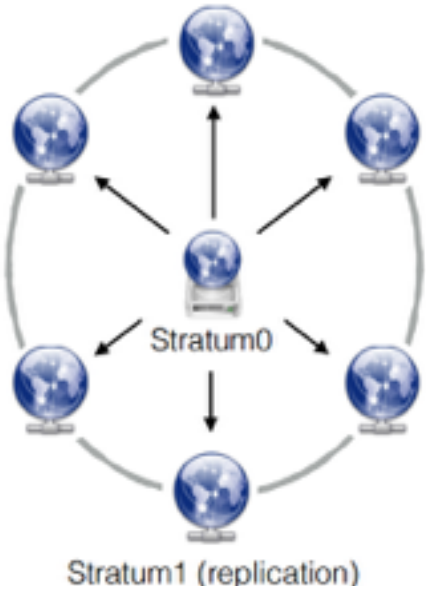
- Developed by CERN (European Organization for Nuclear Research)
- For High Energy Physics (HEP) collaborations
- To deploy software on the worldwide-distributed computing infrastructure:
  - HTTP based
  - Pull mode: get locally only on access
  - Cache hierarchy
  - User space FUSE local read-only mounting point
    - /cvmfs

Cvmfs principles



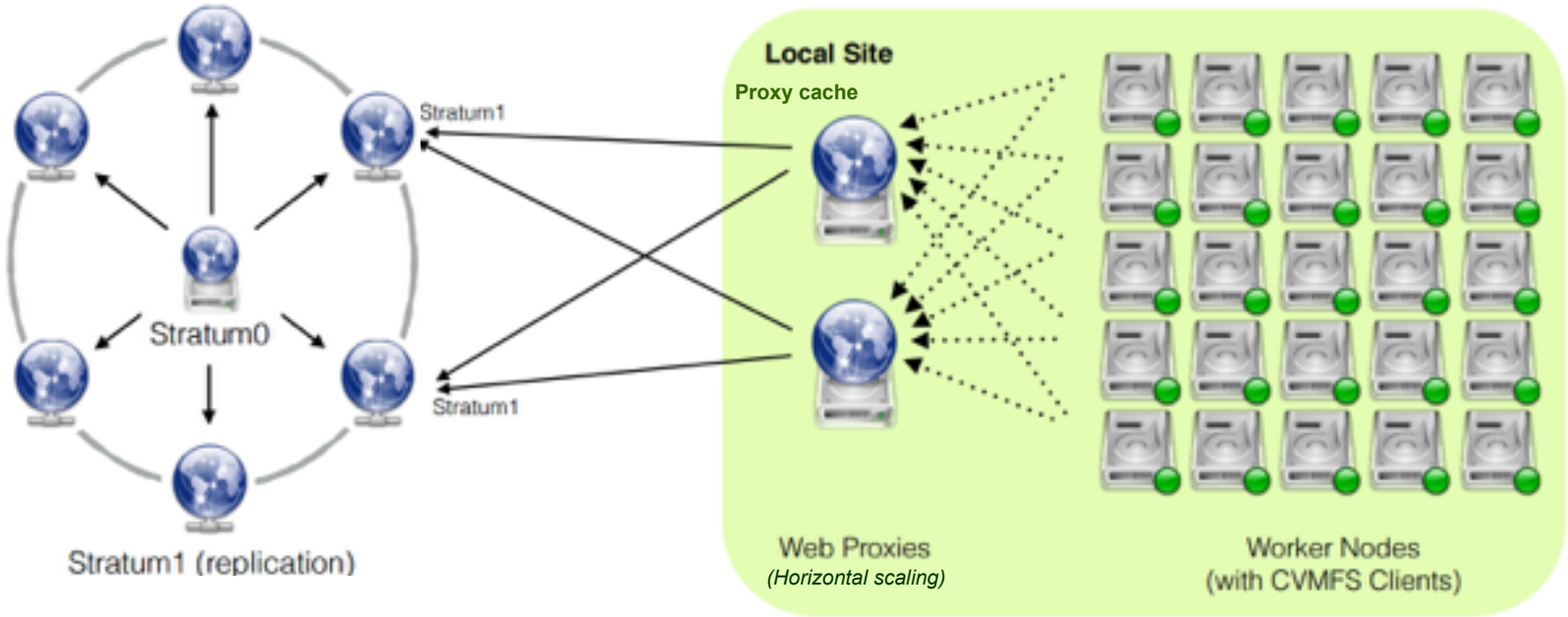
Cvmfs distribution network

- Central repository
- Ring of mirrors: scalability, reliability



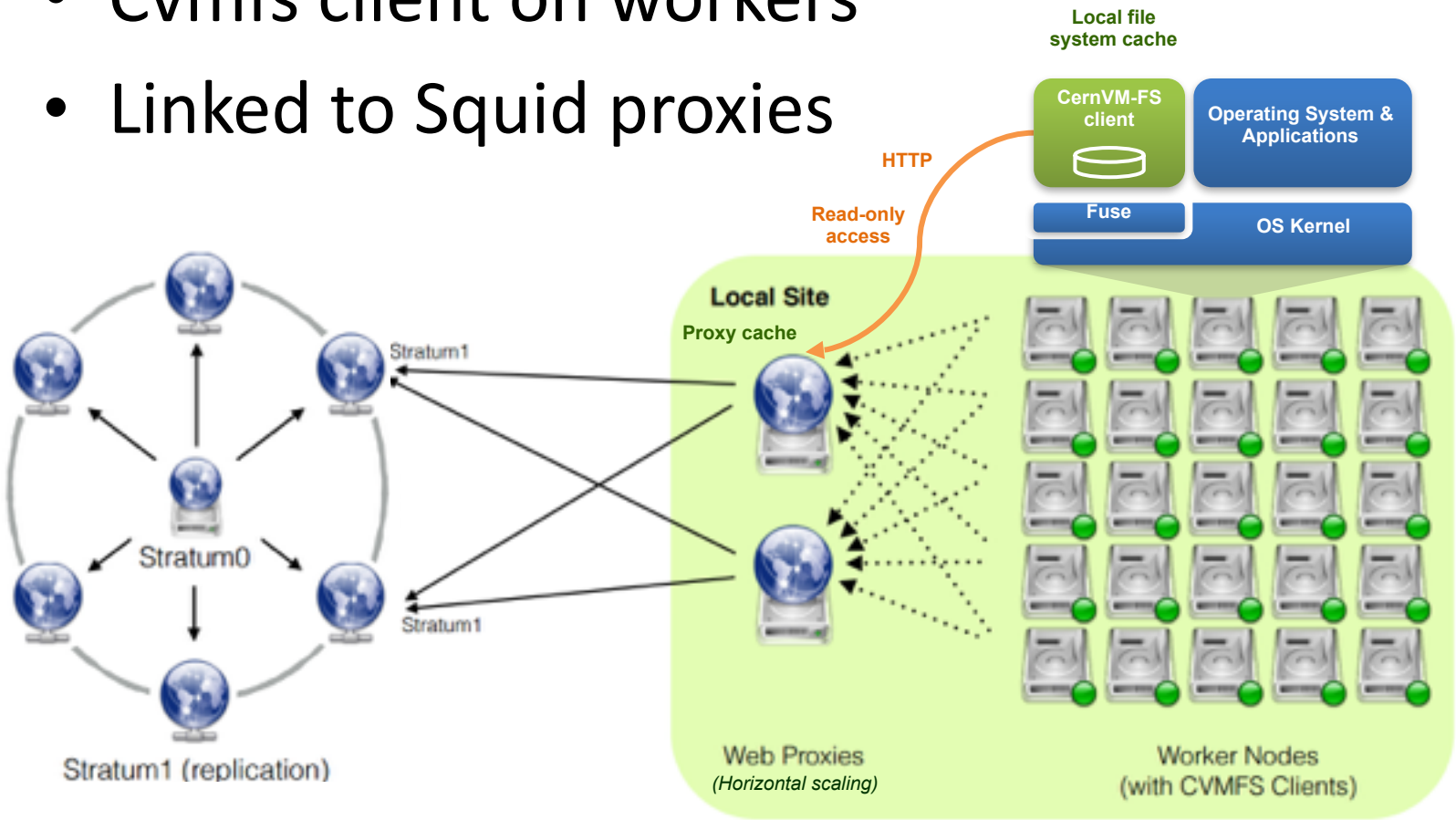
Cvmfs distribution network

- Local Squid proxy: cache, security
- Linked to stratum 1: failover, load balancing



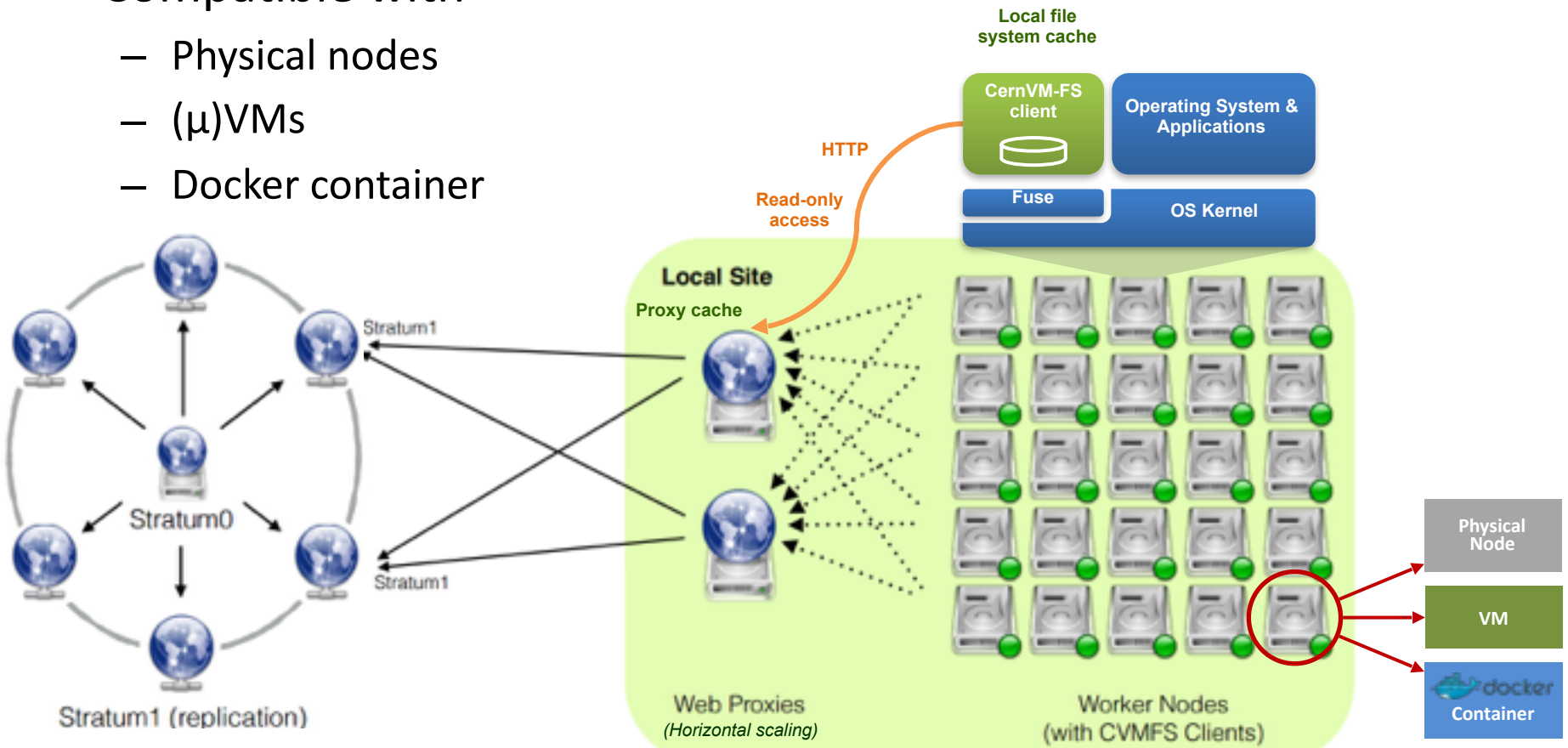
Cvmfs distribution network

- Cvmfs client on workers
- Linked to Squid proxies



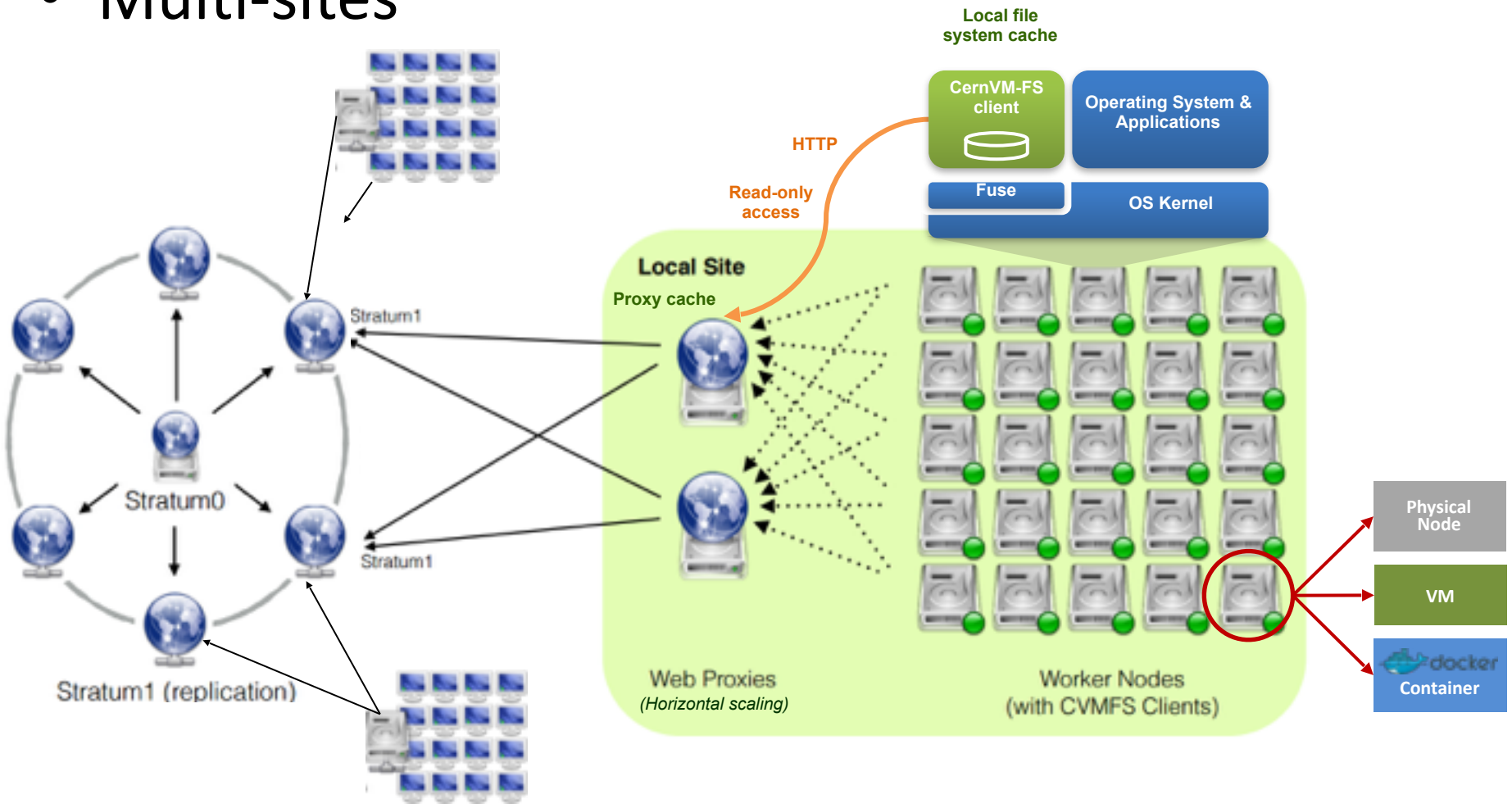
Cvmfs distribution network

- Compatible with
  - Physical nodes
  - (μ)VMs
  - Docker container



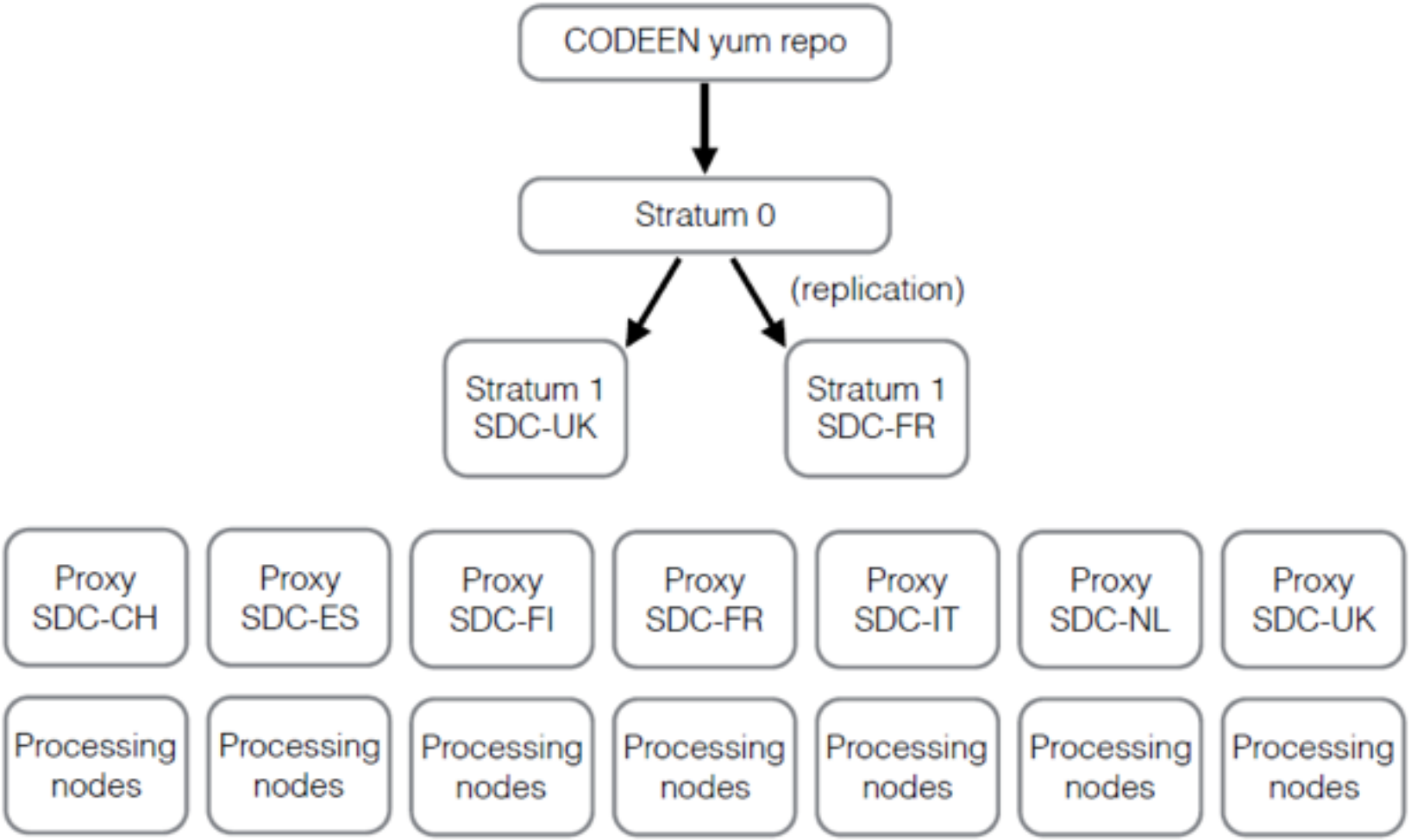
Cvmfs distribution network

- Multi-sites



- Euclid Project & SGS
- From Euclid pipeline to SGS Architecture
- From source code to processing nodes
- CernVM-FS principles
- **Using CernVM-FS on SDCs**

Euclid cvmfs at a glance

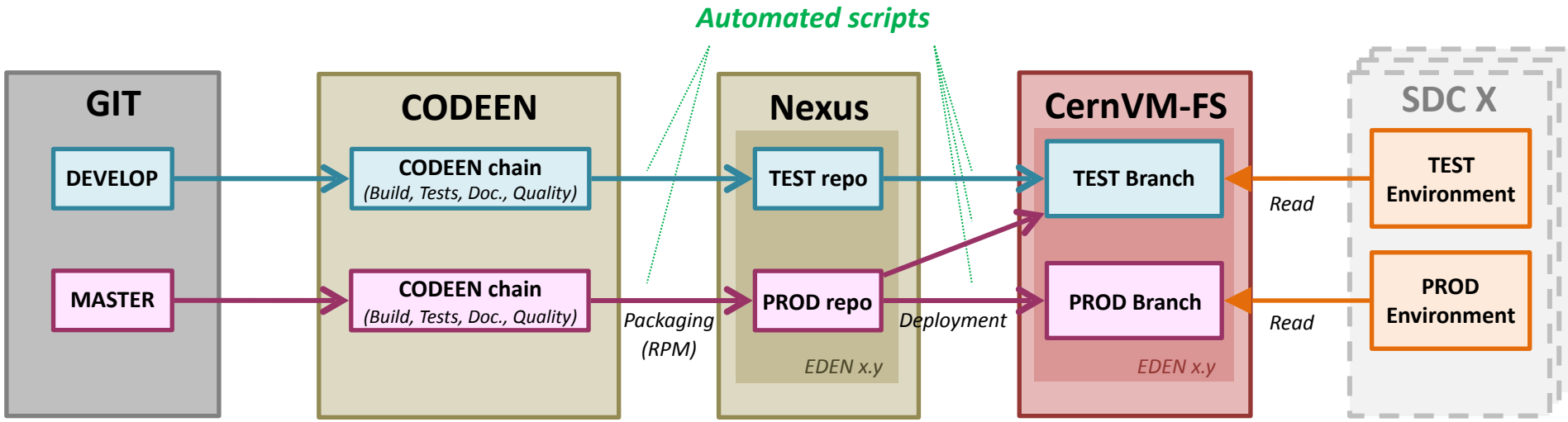


Euclid cvmfs repository structure

```
/cvmfs/euclid.in2p3.fr/
        SL6/
        CentOS7/
                EDEN-1.0/
                EDEN-1.1/
                EDEN-1.2/
                        etc/
                        usr/
                        opt/
                        var/
                EDEN-1.0_env.sh
                EDEN-1.1_env.sh
                EDEN-1.2_env.sh
```

- Multiple versions in //
- Any dependency along (yum install – installroot)
- Switch through Env. Variables

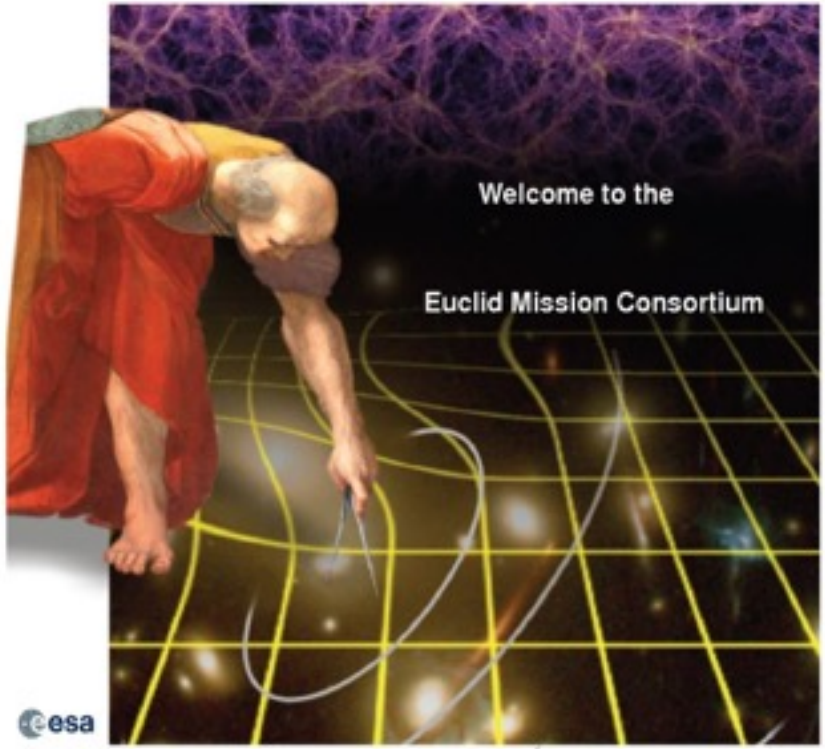
Euclid cvmfs & continuous deployment



- Both Test and Production branches
- 15~mn latency between
  - Installation on stratum 0
  - Availability on any SDC

Conclusions

- Tested through a technical SGS challenge
- Adopted on the whole Euclid SGS
- Used for ongoing Euclid Scientific SGS challenge #3
- Automation scripts almost ready



Thank you for your attention

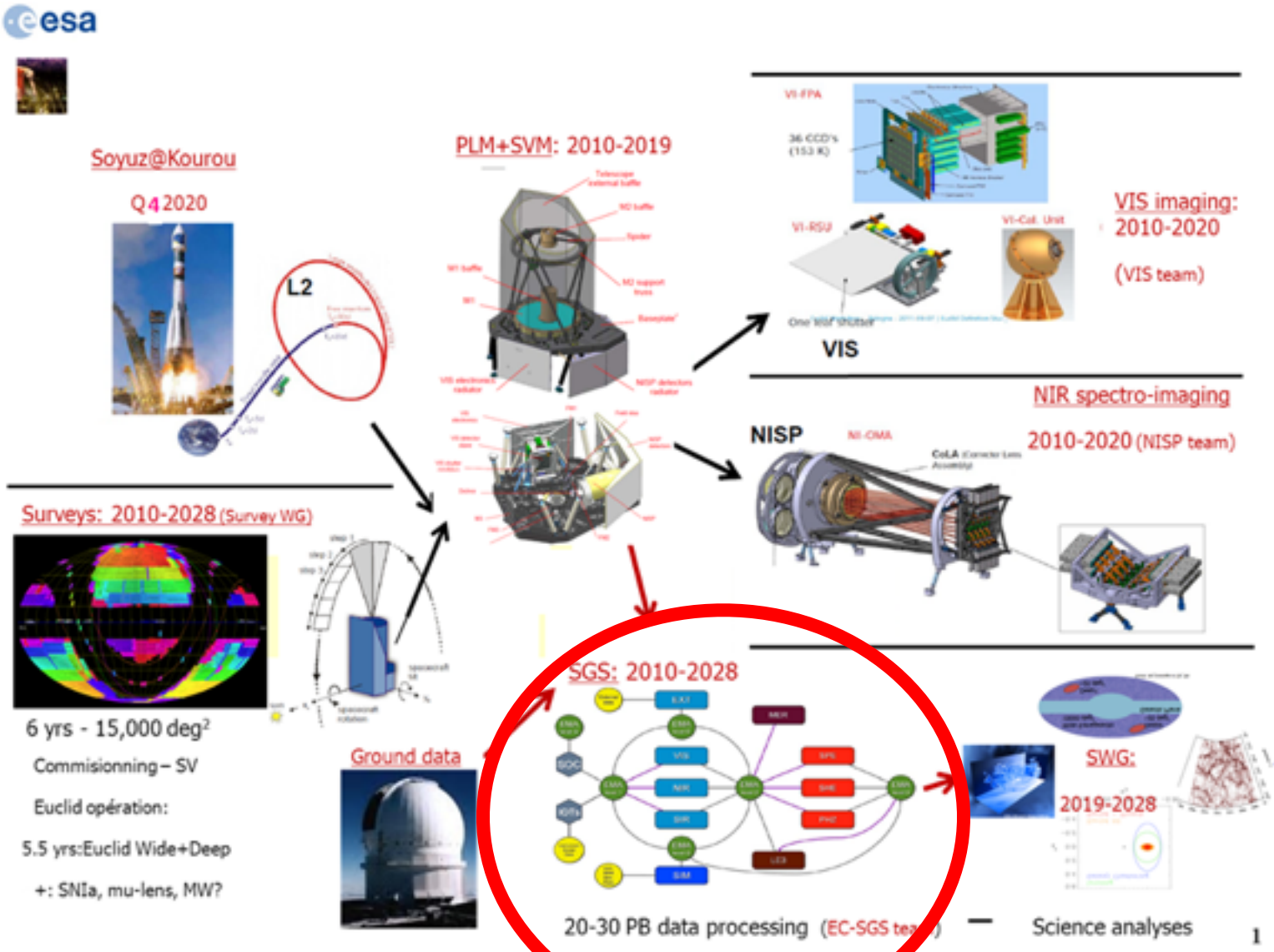
Maurice.Poncet@cnes.fr

Acknowledgments: authors are indebted to all the indididuals participating in the Euclid SGS developement inside ESA and EC, too numerous to be listed here

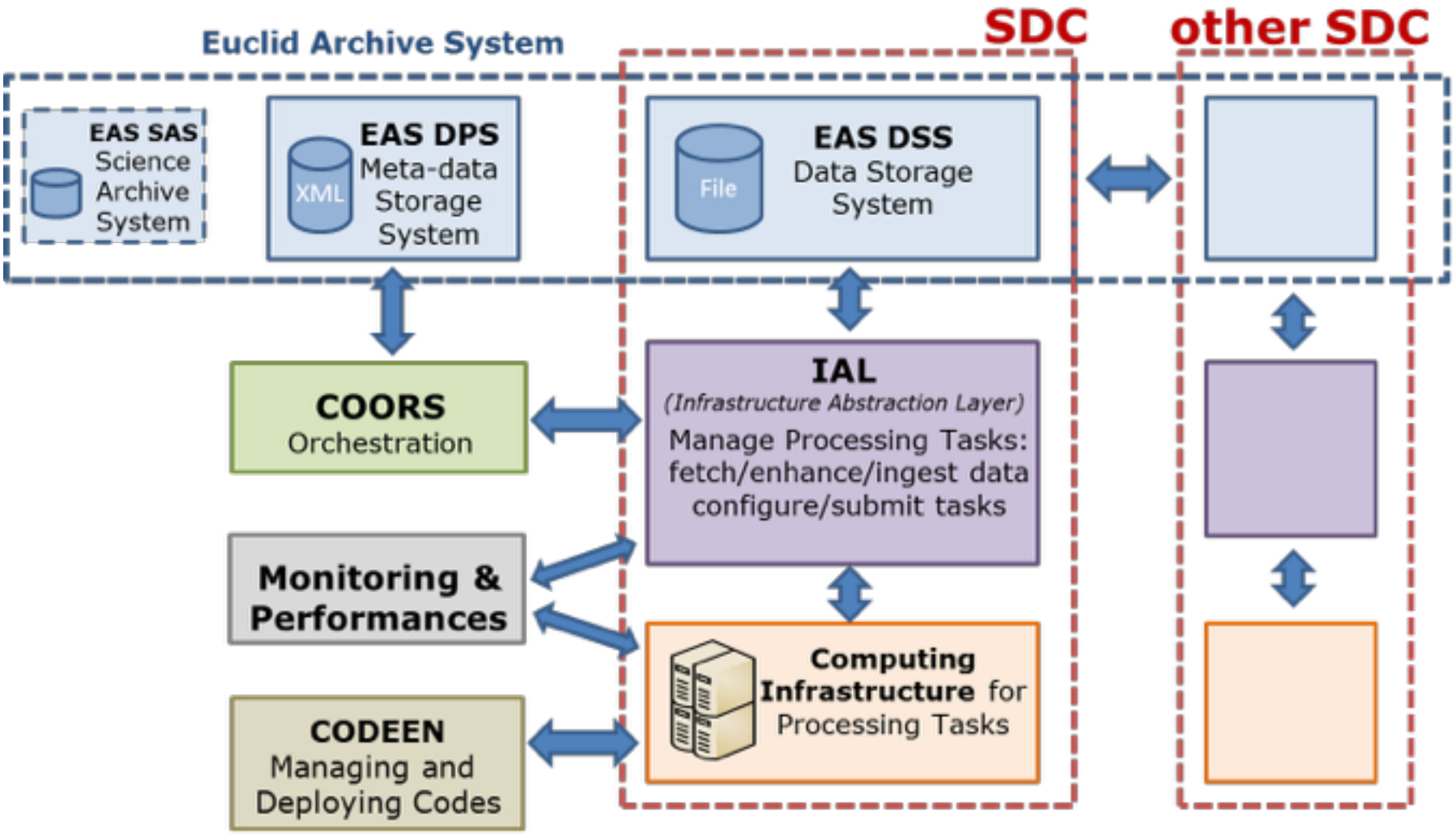
Questions

BACKUP SLIDES

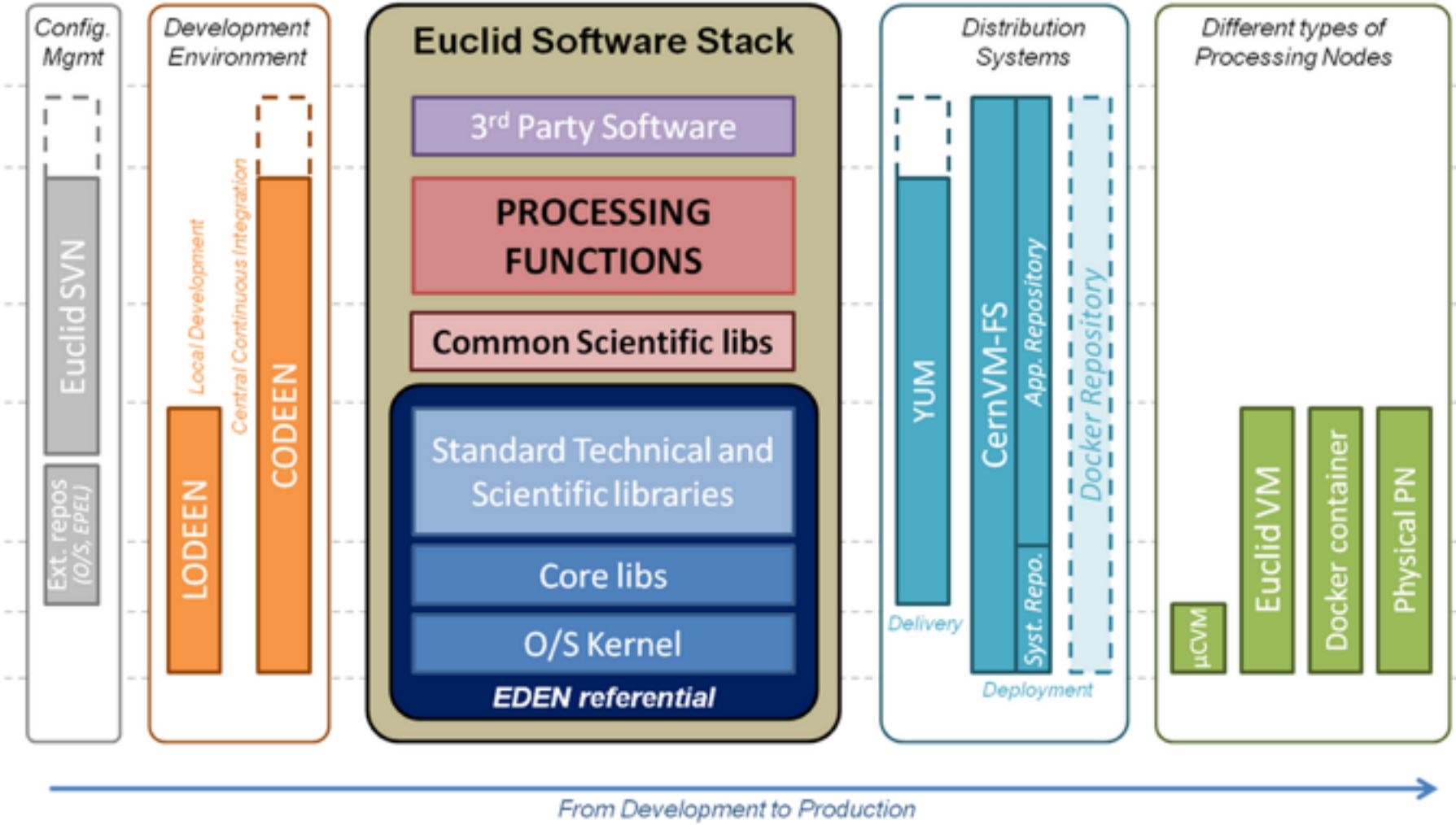
Euclid at a glance



Euclid SGS Components



Euclid tools & software stack



Euclid Software stack