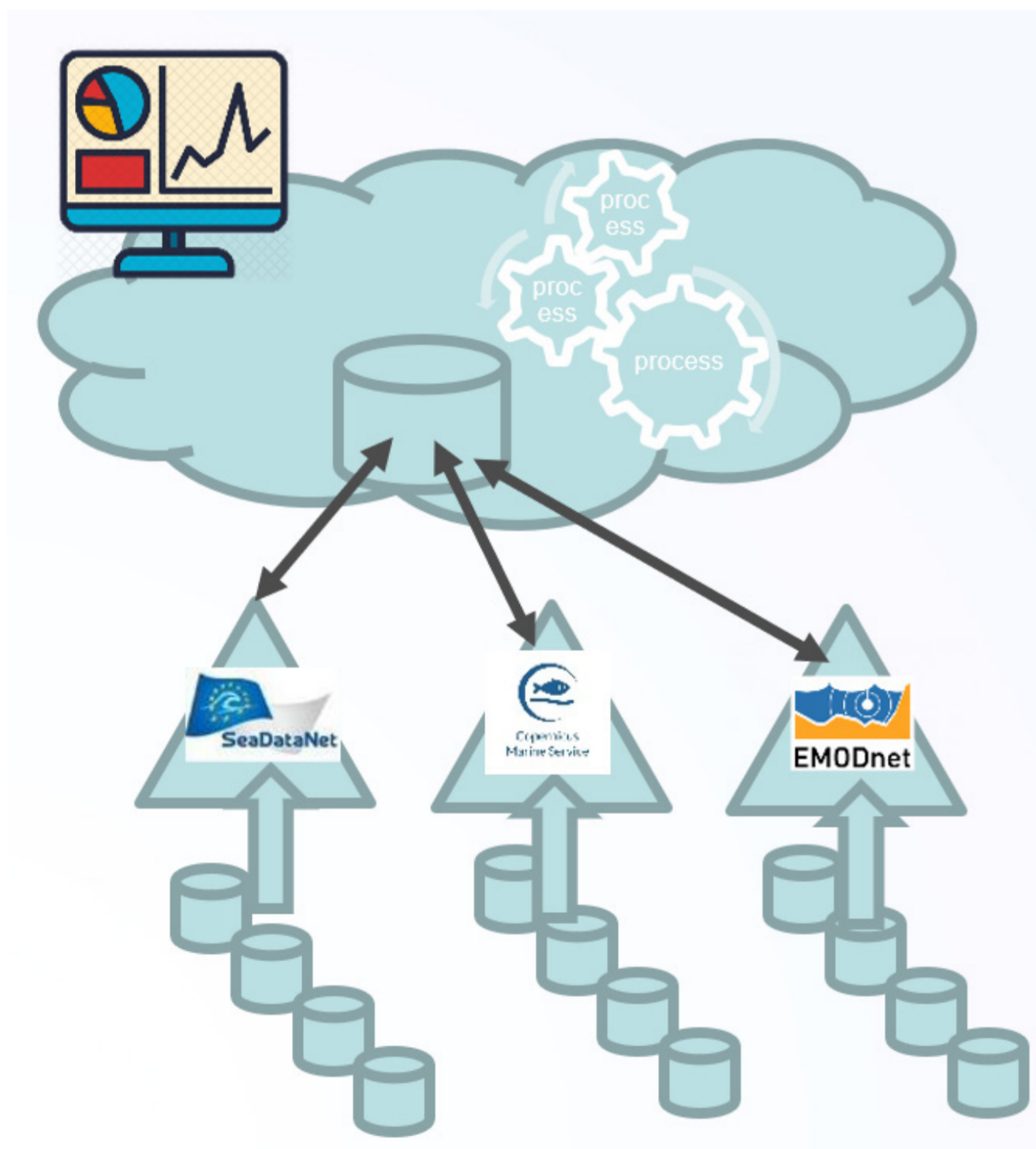


PILOT FOR ACCESSING DISTRIBUTED MARINE DATA USING THE ONEDATA CONCEPT

Current systems:

- Data distributed over data centers and nodes
- Large collections, many files
- Complex workflows for aggregating, storage, accessing, and synchronizing
- Preparing for access by machines in cloud (VRE, EOSC, ..)



sdn-userdesk@seadatanet.org - www.seadatanet.org

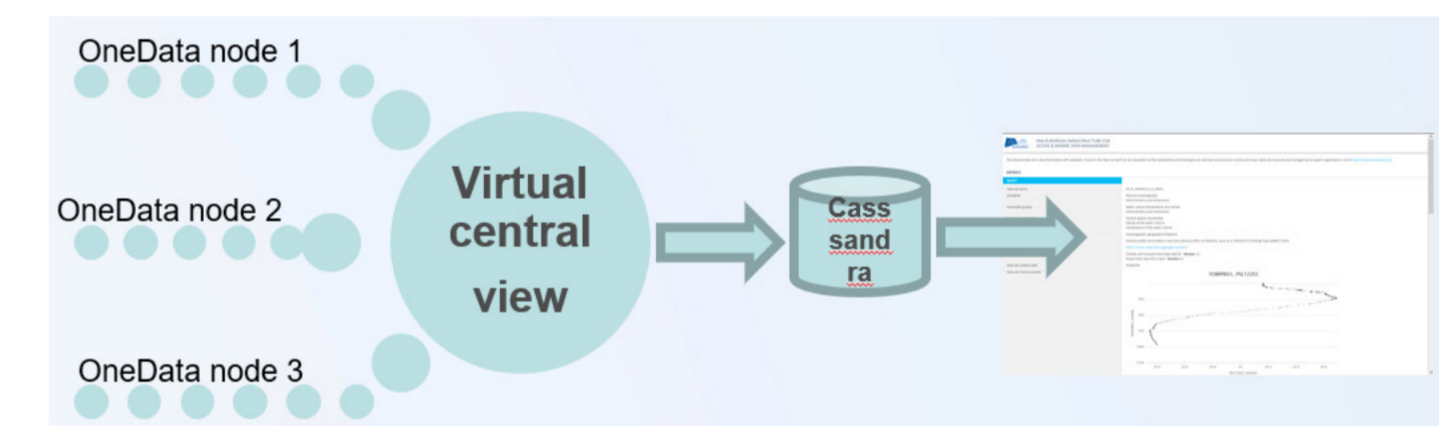
Challenges:

- Managing the access within the aggregators to the distributed data, is now quite complex
- Same for providing access to subsets or collections from aggregator to cloud
- Slow performance of queries and visualisation of datasets and dataset collections
- Technology evolves fast and offer alternatives in the future

=> In this case prototype of EGI's OneData combined with Cassandra

Pilot undertaken:

- Publish datafiles on OneData nodes
- Provides central virtual overview of data on nodes
- Load datafiles into Cassandra for high performance queries
- View the data directly in interface



Lessons learned:

- The OneData system functions very well, offering live and accurate overview of available files on nodes and transfer to cloud.
- OneData could be a very good way to share SeaDataNet collections to external systems, and as a kind of data lake to be used internally for the SDC VRE, or connect to OneData data lakes of other data infrastructures.
- Regarding the performance of visualisation, the Cassandra database querying is fast but not very flexible.
- Visualising the data from Cassandra when set up well to the query, was very responsive.