



# Experiences with the cloud of the Hungarian Academy of Sciences

Gábor Petó and Peter Kacsuk

[peto.gabor@wigner.mta.hu](mailto:peto.gabor@wigner.mta.hu) (Director of MTA Cloud)

[Kacsuk.Peter@sztaki.mta.hu](mailto:Kacsuk.Peter@sztaki.mta.hu) (Technical director of MTA Cloud)

# Objectives of MTA Cloud



To support those researchers of the Hungarian Academy of Sciences (MTA)

- who want to run
  - Compute-intensive or
  - Data-intensive applications
- who want to use clusters, supercomputers or data centers for accelerating the execution of their applications
- To avoid the costly and time-consuming creation of such computing infrastructures they can immediately use the computing and storage services of MTA Cloud

# MTA Cloud is an alliance of cloud sites



The goal was to build an open multi-cloud architecture to which other clouds could join later.

Therefore MTA Cloud is organized as an alliance of cloud sites. Currently two sites are involved:

- MTA SZTAKI
- MTA Wigner Data Center

Alliance means:

- The user interface to register and authenticate for the users is common
- The applied cloud technology is the same (OpenStack)
- The web page of MTA Cloud is common
- The management of the system is a joint activity of SZTAKI and Wigner Data Center
- The training and user support activities are common

# Current capacity of MTA Cloud



## Wigner site:

800 vCPU

2.5 TB RAM

400 TB HDD

## SZTAKI site:

360 vCPU

832 GB RAM

164 TB HDD

Computing capacity is already fully used but the storage capacity is currently underutilized

# Additional capacity under development



## Wigner site:

### a. Compute extension:

2 x Intel Xeon E5-2660 v2 processor (40 vCPU)

128 GB memory

1 TB disk

**4 nVidia V100 GPU cards**

### b. Storage extension:

2 x Intel Xeon E5-2630 v2 processor

64 GB memory

6 TB disk

## SZTAKI site:

**3** computing nodes (64 vCPUs + 512 GB RAM + **1 GPU**)

1 computing node (40 vCPUs + 128 GB RAM)

**1 GPU computing accelerator**

15x250 GB SSD storage for CEPH

# Services of MTA Cloud



- IaaS resource provision based on the capacity mentioned before
- PaaS services:
  - Occopus cloud orchestrator
  - Data Avenue file transfer service
  - Hadoop cluster
  - Swarm cluster
  - WS-PGRADE/gUSE science gateway and workflow system
  - Flowbster stream-oriented workflow system
- All of them are supported by detailed tutorials at MTA Cloud web page

# Motivations for Occopus



There are many use cases where the goal is to set up a **complete infrastructure or a set of services** called **virtual infrastructures** (VI) on demand in the cloud

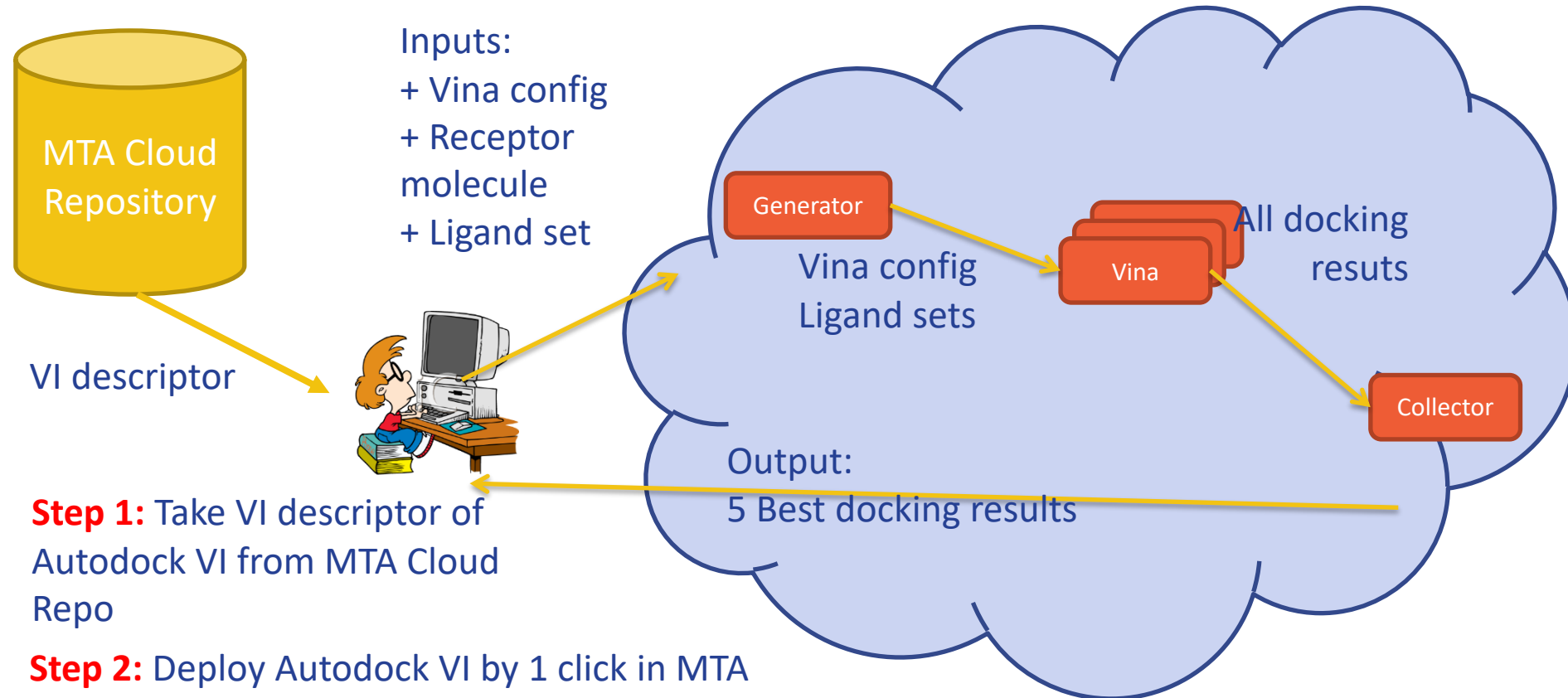
Goal is to enable end-user scientists to easily (ideally by 1 click) deploy such a VI in the cloud either temporarily or for a longer period of time

Example:

Biologist would like to run autodock in-silico simulation between one receptor molecule and 100.000 ligand molecules

Biologist takes the autodock VI from EGI AppDB and by 1 click deploys it in a target cloud and uses it

# Example: Autodock virtual infrastructure in cloud



**Step 1:** Take VI descriptor of Autodock VI from MTA Cloud Repo

**Step 2:** Deploy Autodock VI by 1 click in MTA Cloud

**Step 3:** Run autodock appl in MTA cloud with 100.000 ligands



# Building Hadoop cluster by Occopus in the MTA Cloud



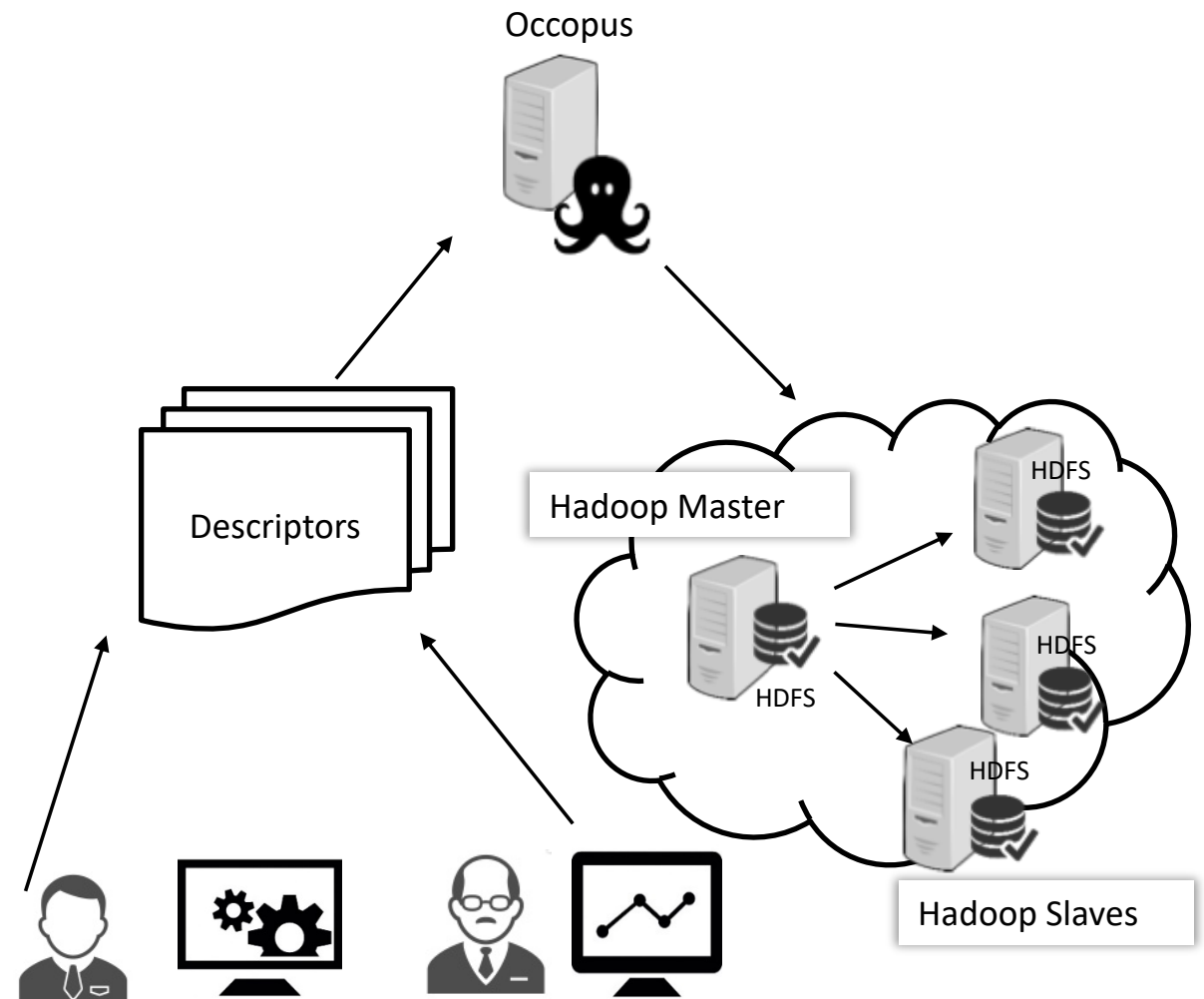
**Step 1:** Deploy Occopus

**Step 2:** Download Hadoop descriptors

**Step 3:** Personalize descriptors

**Step 4:** Deploy Hadoop cluster (based on the personalized descriptors)

**Step 5:** Run MapReduce jobs

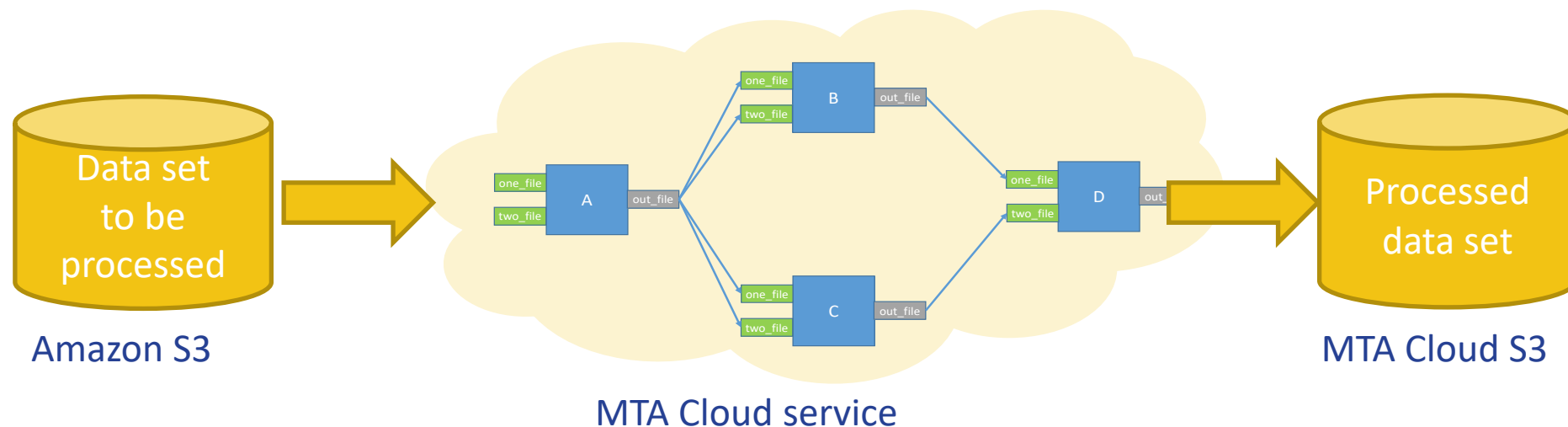


# Concept of Flowbster



The goal of Flowbster is to enable

- The quick deployment of the workflow as a pipeline infrastructure in the cloud by Occopus
- Once the pipeline infrastructure is created in the cloud it is activated and data elements of the data set to be processed flow through the pipeline
- As the data set flows through the pipeline its data elements are processed as defined by the Flowbster workflow



# Usage of MTA Cloud

- Web page lists all the projects running on MTA Cloud
- Currently 52 projects from 17 research institutes

Name of the project

Goal of the project

Research institute

Project director

Újgenerációs szekvenálási (NGS) adatok feldolgozása

Enzimológiai Intézet

Fóthi Ábel

A projekt célja a felhő tesztelése NGS adatok, elsősorban kis RNS szekvenálási adatok tisztítására, elemzésére.

Ultrarövid lézerimpulzussal létrehozott rubídium plazma szimulációja

Részecske- és Magfizikai Intézet

Demeter Gábor

A projekt ultrarövid lézeimpulzusok terjedését vizsgálja rubídiumatom-gőzben. A cél térben kiterjedt, nagyon homogén plazma létrehozásának optimalizálása amelyet azután olyan kísérletekben lehet felhasználni, mint például a CERN-ben folyó AWAKE együttműködés. (<http://awake.web.cern.ch>)

Változócsillagok klasszifikációja

Konkoly Thege Miklós Csillagászati Intézet

Szabó Róbert

A NASA Kepler űrtávcsőve nagy számú változócsillag ultrapontos fényességmérését tette lehetővé. Ezen csillagok automatikus klasszifikációja az eddig használt módszerek újragondolását teszi szükségessé. Ebben projektben a Big Data megközelítés legújabb machine learning algoritmusait alkalmazzuk a Kepler-adatok egy jól körülhatárolt részére, azzal a céllal, hogy új változócsillagokat találjunk és módszereket később más csillagászati adatbázisokra is alkalmazhassuk.

Vírusfehérje molekuladinamikai szimulációja

Mezőgazdasági Intézet

Gellért Ákos

Két rendszeren szeretném kipróbálni az MTA Cloud számítási kapacitását. Az egyik egy növényi vírusfehérje duplaszálú kis RNS-sel alkotott komplexe. Itt egy általunk azonosított fontos szabályzó régió szerepét kívánom vizsgálni. A másik szimuláció pedig a rotavírus belső kapszidrétegét kialakító VP2 fehérje és a vírus belsejében

# Conclusions

- MTA Cloud has been in production since October 2016
- After one year of usage there were more than 40 projects that used nearly the whole set of computing resources
- As a result significant extension of MTA Cloud is on its way
- MTA Cloud is getting more and more popular, currently 52 projects run on MTA Cloud

# Thanks for your attention

[kacsuk.peter@sztaki.mta.hu](mailto:kacsuk.peter@sztaki.mta.hu)