

# Case Study and INDIGO solution for:

## TRUFA



INDIGO - DataCloud

RIA-653549

Presented by Lara Lloret Iglesias  
IFCA (Spain)

[\*\*\*lloret@ifca.unican.es\*\*\*](mailto:lloret@ifca.unican.es)



Instituto de Física de Cantabria

**INDIGO SUMMIT**

EGI-INDIGO workshop on community application support  
Catania, 10<sup>th</sup> May 2017



INDIGO-DataCloud is co-funded by the  
Horizon 2020 Framework Programme

# TRUFA (<http://trufa.ifca.es/>)



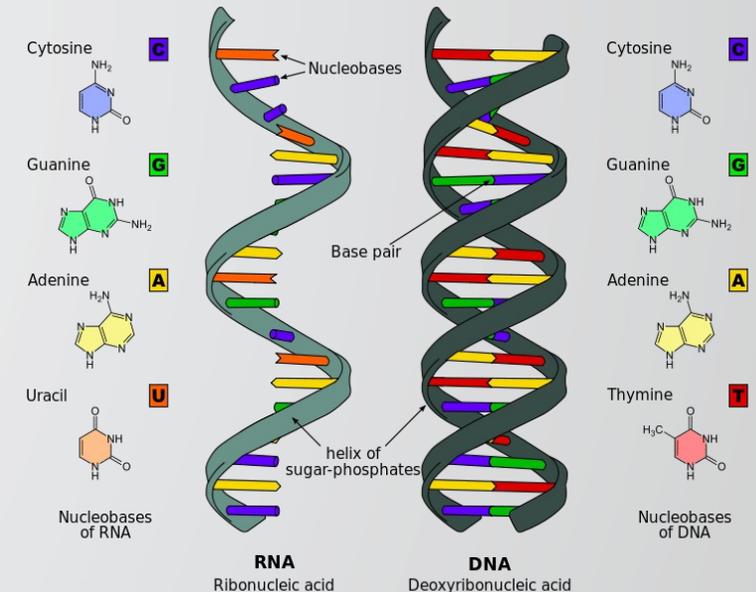
A web-based RNA-seq application

**RNA-seq is used to analyze the continually changing cellular transcriptome**

- *RNA comparison* between healthy and unhealthy organisms can lead to *better understanding* diseases such as *cancer, diabetes, alzheimer...*
- Also useful for *comprehending several biological mechanisms* (i.e venom production, silk production...)

**Application of next-generation sequencing (NGS) methods for transcriptome analysis (RNA-seq) has become increasingly accessible** in recent years and are of great interest to many biological disciplines including, *evolutionary biology, ecology, biomedicine, and computational biology.*

Although virtually any research group can now obtain RNA-seq data, **only a few have the bioinformatics knowledge and computation facilities required for transcriptome analysis.**



# TRUFA (<http://trufa.ifca.es/>)

## A web-based RNA-seq application



**The main advantage of TRUFA is that it allows to gather together the most advanced tools in RNA-seq in a friendly and interactive way**

→ Completely transparent. No informatics knowledges need the user.

**The typical file size used in this kind of analysis goes from 5 GB to 40 GB per file.**

→ Large data both for processing and for storing

**Most of the analysis take more than 3 days to be completed.**

→ A lot of CPU time

**Before starting the project TRUFA was ...**

...running in Altamira. Some jobs were killed due to an excess of CPU time.

...having problems handling such big input files

...not completely up to date with respect to the latest and greatest tools in the market

# TRUFA (<http://trufa.ifca.es/>)

A web-based RNA-seq application



## **Key solutions :**

**Build a new version of Trufa allowing to run in the CLOUD**

- Each of the TRUFA steps (cleaning, assembly, identification and expression) are *implemented within an ubuntu container*
- Dockers are *managed using udocker*
- The user only has to *download the TRUFA docker repository* and can launch the analysis to his favourite cloud *without worrying* about dependencies

**Updating/changing the different applications according to the current state of the art.**

- In contact with biologist (MNCN, UB) using the tool to *cover their needs*
- Other INDIGO solutions that may be included in the future: OneData, Chronos..

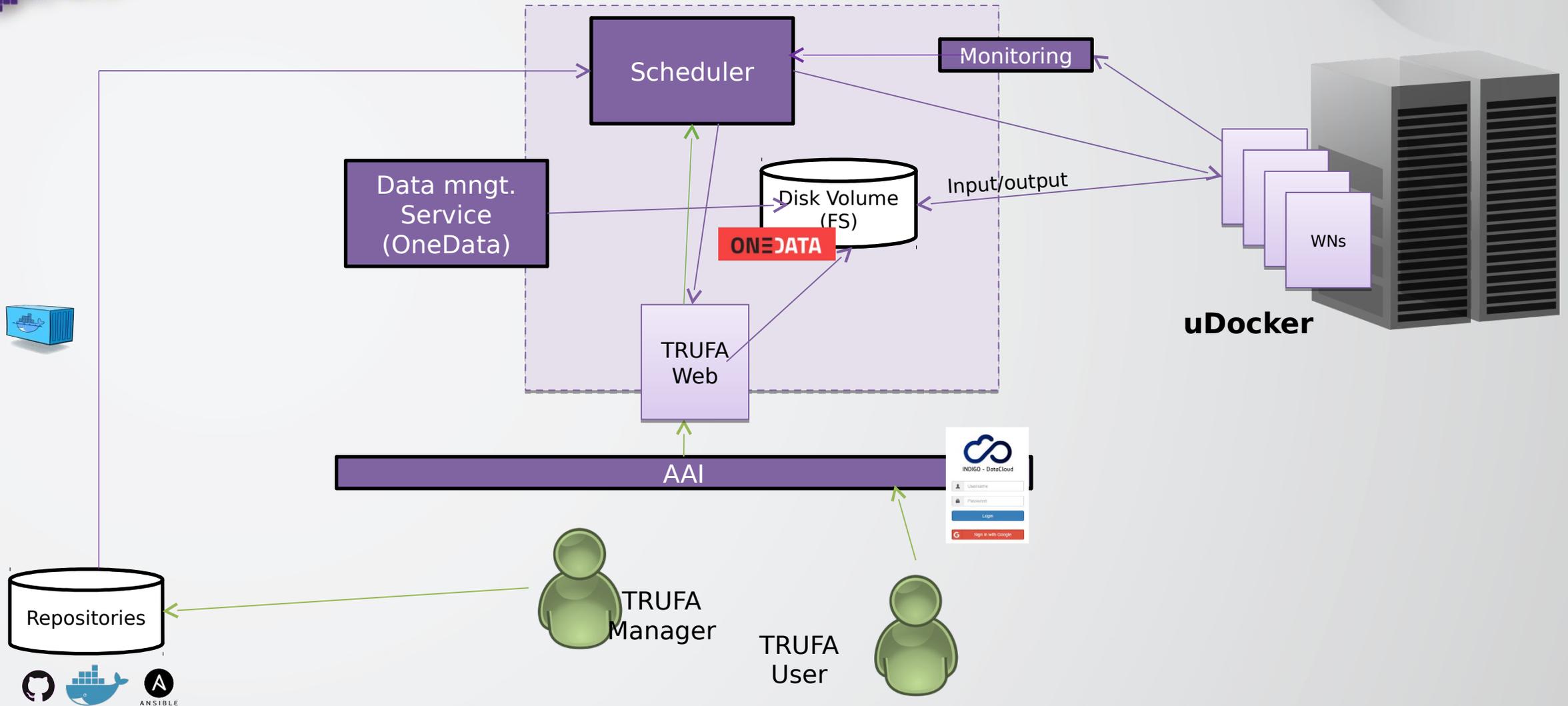
**Already tested in IFCA and working smoothly.**

# How an INDIGO solution is key for us:



- ***Udocker is a key component in our solution:***
  - ***It is used to manage and execute the different containers from TRUFA***
  - ***Before, TRUFA was very difficult to export due to the many packages to be installed and configured all of them with dependencies wrt your operating systems, batch system, etc...***
  - ***With this new approach one can just download the TRUFA docker and run transparently anywhere, including the CLOUD.***

# TRUFA with uDocker



# We are ready to share our experience!



## <http://trufa.ifca.es/>

Kornobis, Cabellos, Aguilar, Frias-Lopez, Rozas, Marco, Zardoya, Lloret



[Link to Trufa Demo](#)

<https://www.indigo-datacloud.eu>

**Better Software for Better Science.**