



**ETH**

Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich  
Institute of Molecular Systems Biology

CloudBroker

# Rosetta in the Cloud

Peter Kunszt

*ETH Zurich:SyBIT & CloudBroker.com*

Lars Malmström

*ETH Zurich:IMSB*



**SyBIT**  
SystemsX.ch  
Biology IT

# Disclaimer

- This is not a science talk
- I am a physicist – please bear with me
- I have VERY limited knowledge of Rosetta

# What did we do

- Lars: Modeling Streptococcus: what is virulence?
- Compare virulent with non-virulent strains
- Identify structurally relevant differences

Number of proteins	1697
Structural domains	832 de-novo, 1440 homology
Number of models	22'720'000
Total CPU hours needed	800'000

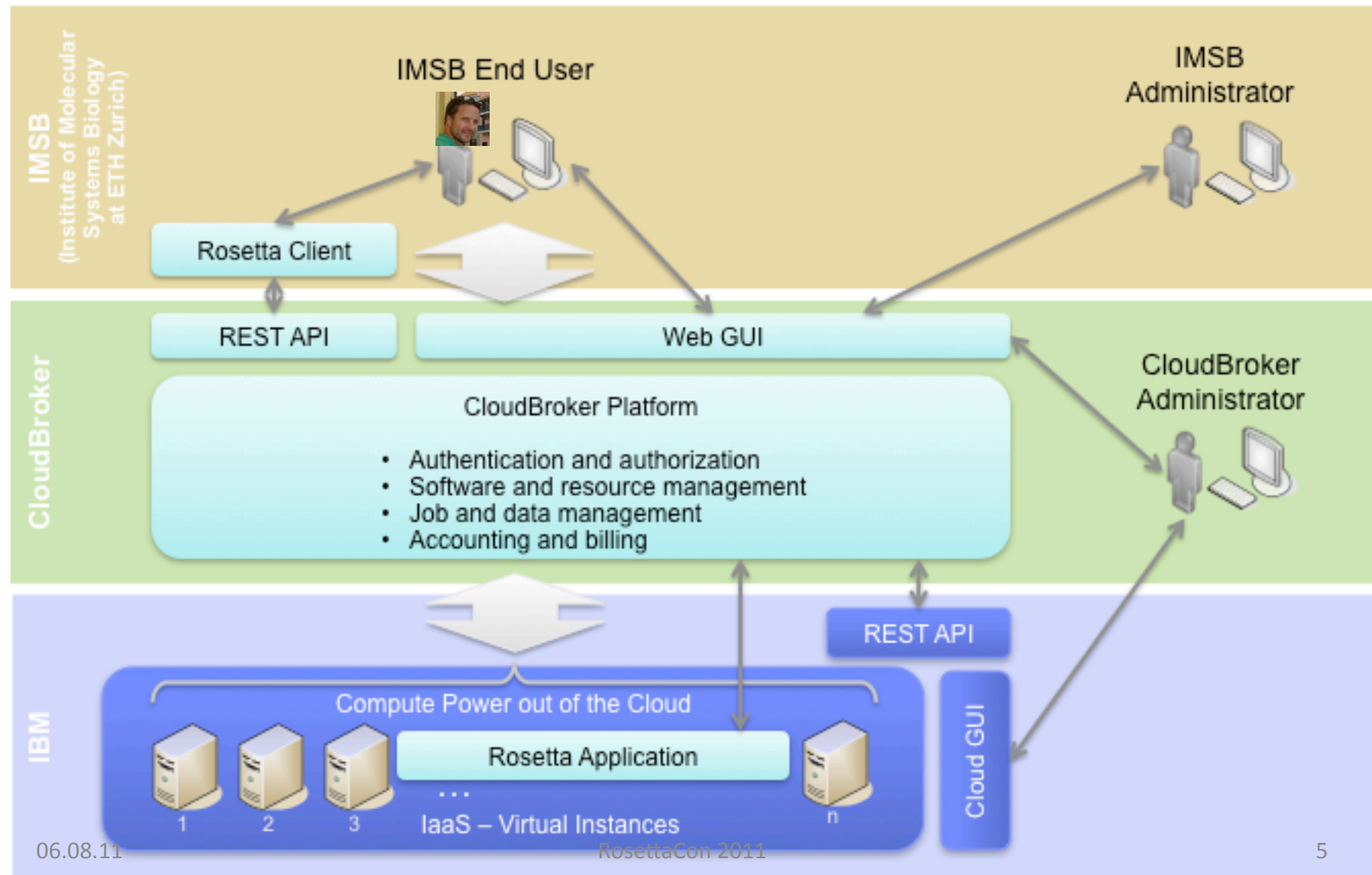
# Components

- The User :



- The CloudBroker Platform
- IBM SmartCloud Enterprise cloud: sponsored by IBM

# Setup



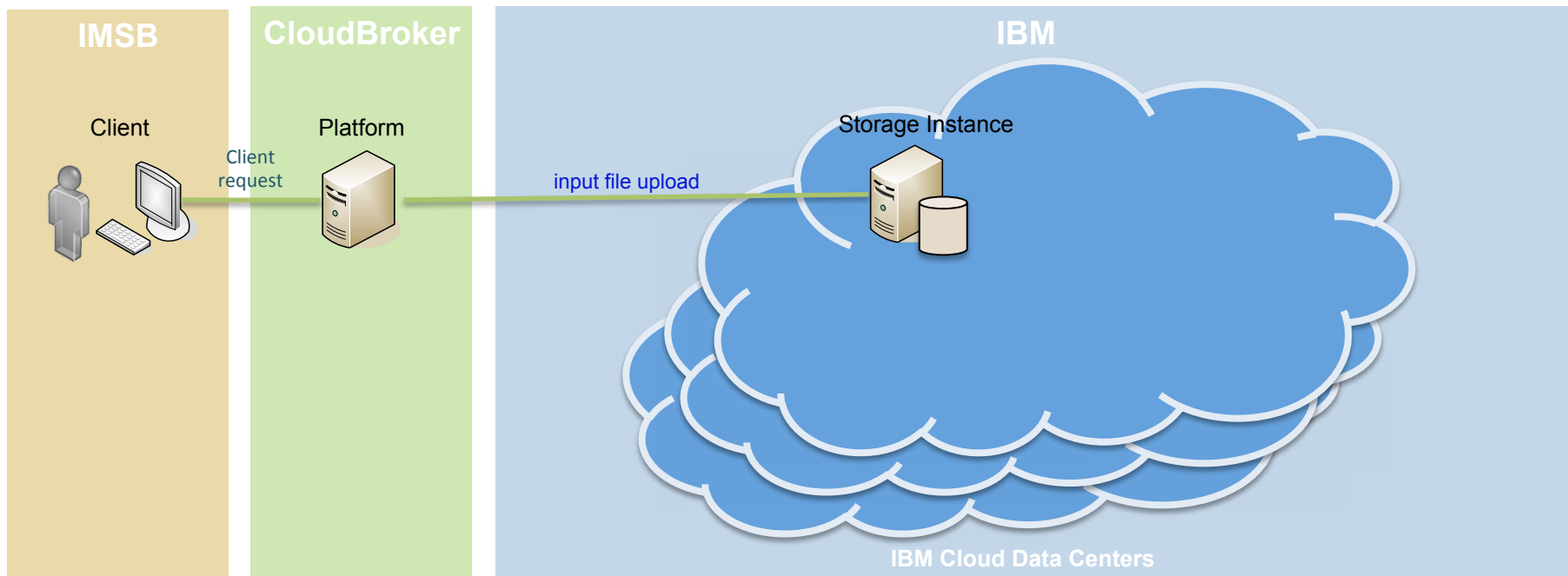
# CloudBroker Rosetta Client Tool

Java executable – takes care of

- Automatically create and submit jobs
- Restart or recover lost jobs
- Monitor jobs (status information)
- Retrieve output at the end

```
java -jar CloudBrokerClient.jar <arguments> :  
- submit [propertyFile] [numberOfModels] [inDir]  
- status [propertyFile] [numberOfModels] [inDir]  
- restart [propertyFile] [numberOfModels] [inDir]  
- output [propertyFile] [nModels] [inDir] [outDir]
```

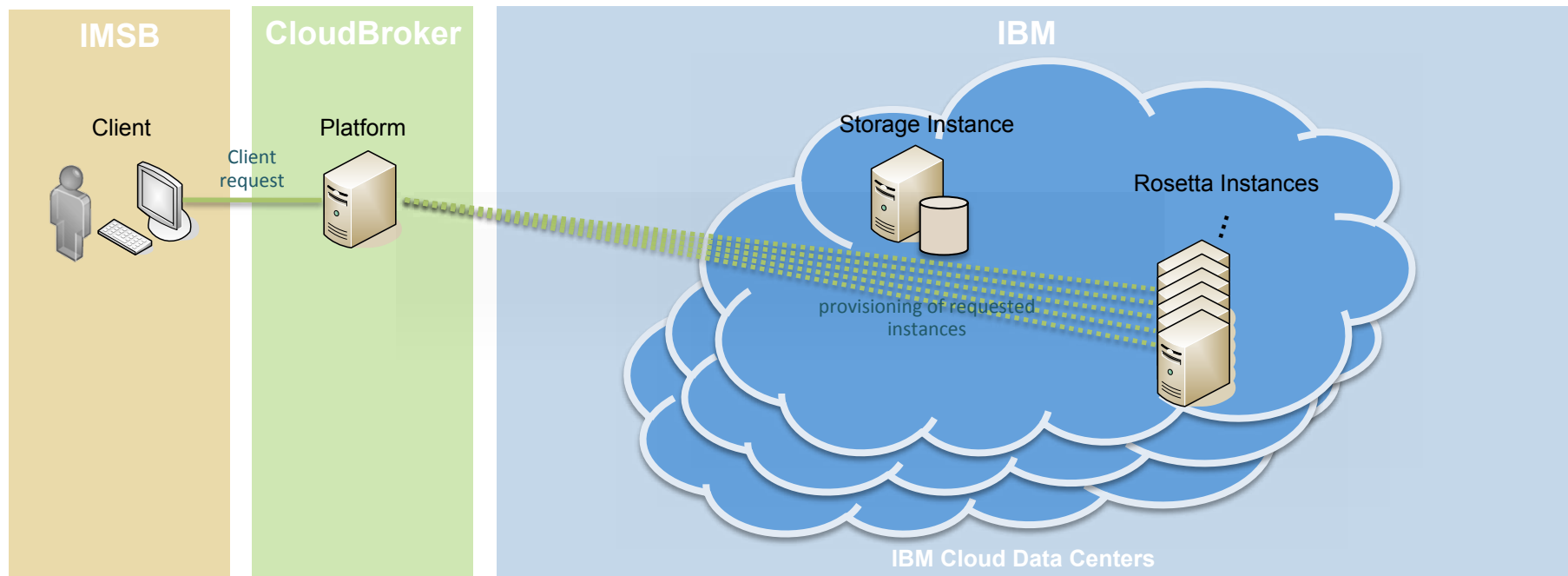
## Rosetta Run– Phase 1: Client prepares jobs



### Activities:

- The client specifies the number of jobs and provides the Rosetta specific input files for the calculation jobs.
- The input file upload to the storage instance is initiated and controlled by the CloudBroker Platform.

## Rosetta Run– Phase 2: Client submits jobs

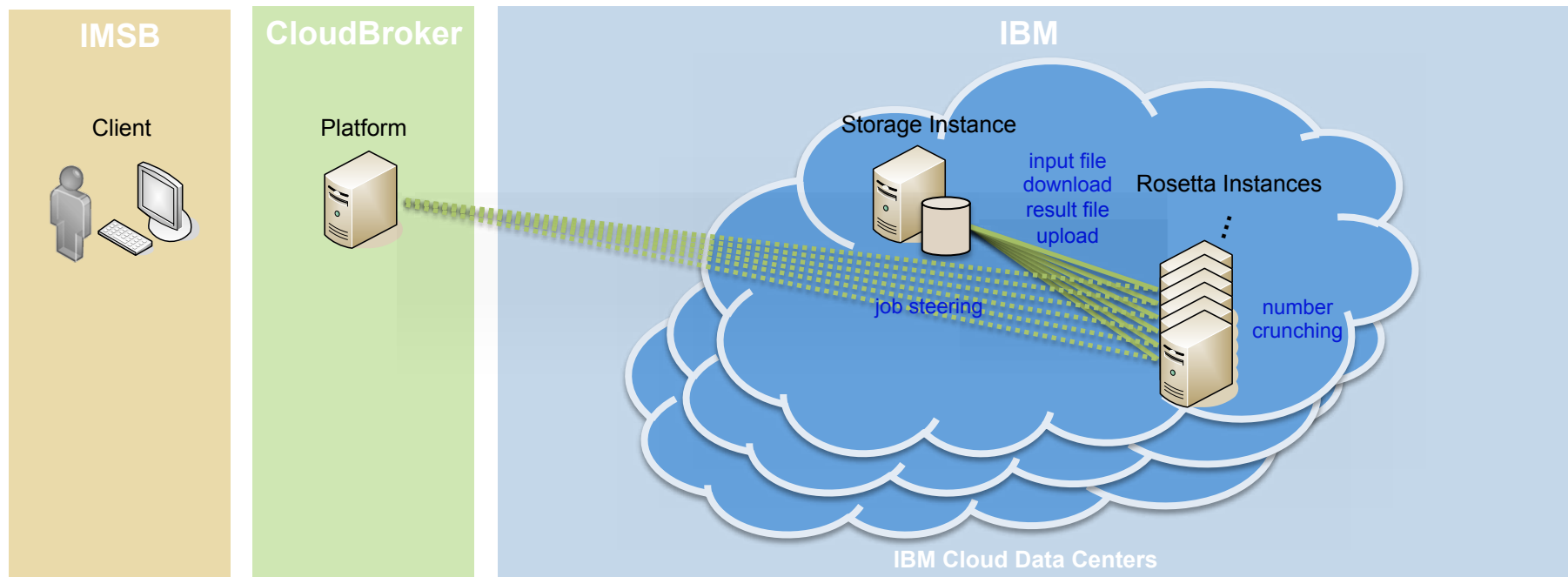


### Activities:

- The CloudBroker platform provisions the cloud instances and starts the jobs
- IBM provides the requested amount of compute power out of the cloud.



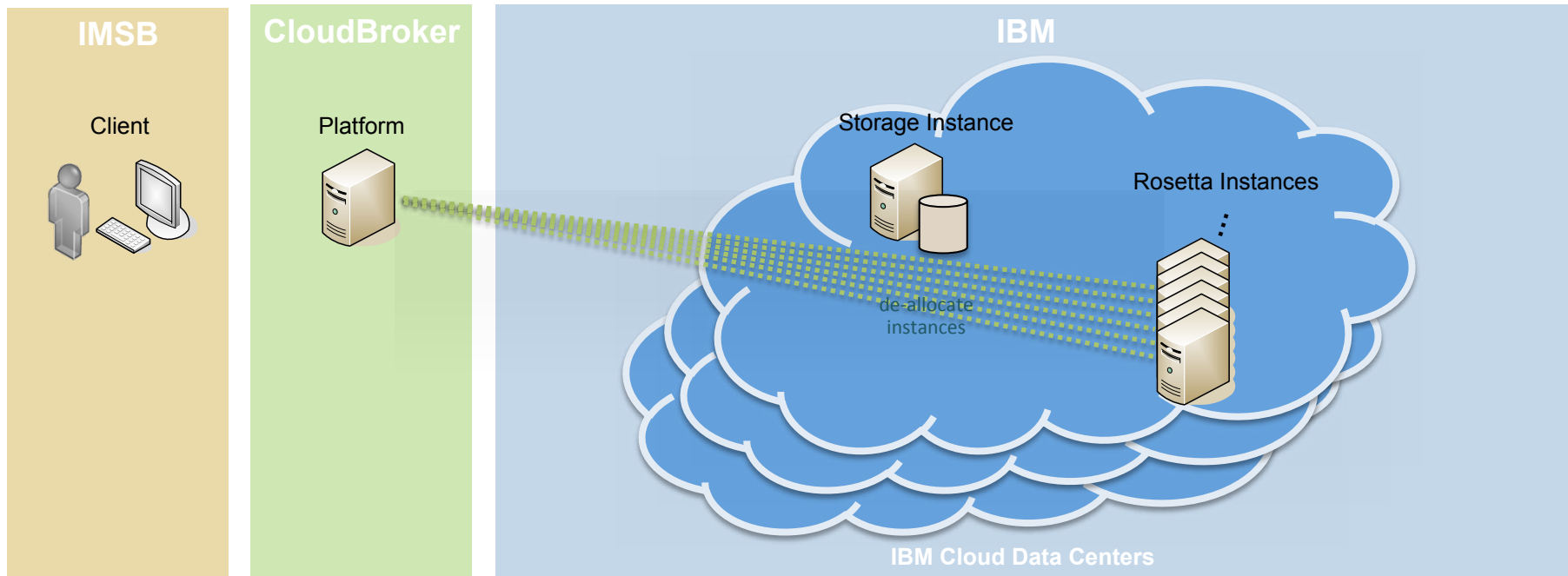
## Rosetta Run – Phase 3: Rosetta instances process the input files



### Activities:

- The CloudBroker Platform initiates and controls the job input file download from the storage instance to the appropriate Rosetta compute instances.
- The CloudBroker Platform manages the processing of the Rosetta jobs on the compute instances via its internal queuing system. Upon client request, new Rosetta jobs can be prepared and submitted, and intermediate or final results can be downloaded while other jobs are running.
- The CloudBroker Platform initiates and controls the job result file upload from the Rosetta compute instances to the appropriate storage instance.

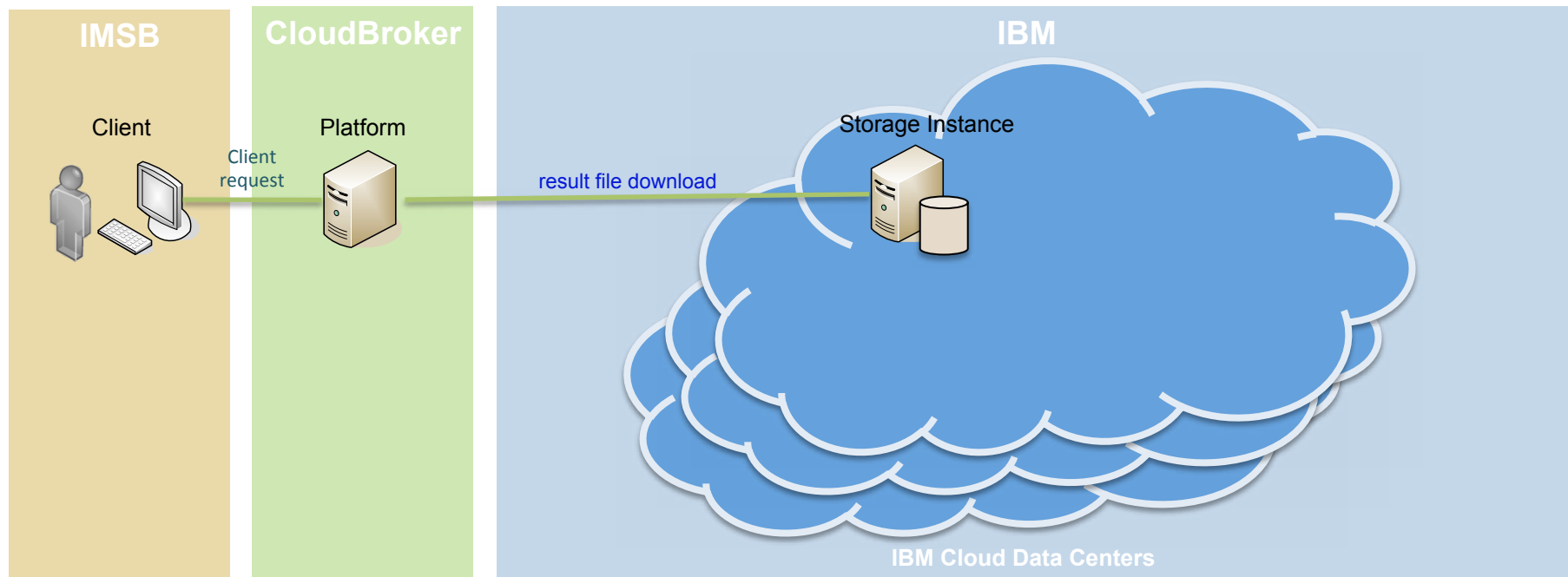
## Rosetta Run – Phase 4: Release of allocated Rosetta instances



### Activities:

- The CloudBroker Platform deletes the input and result files and de-allocates the Rosetta instances. This process happens dynamically while the job queue empties.

## Rosetta Run – Phase 5: Client downloads the result files

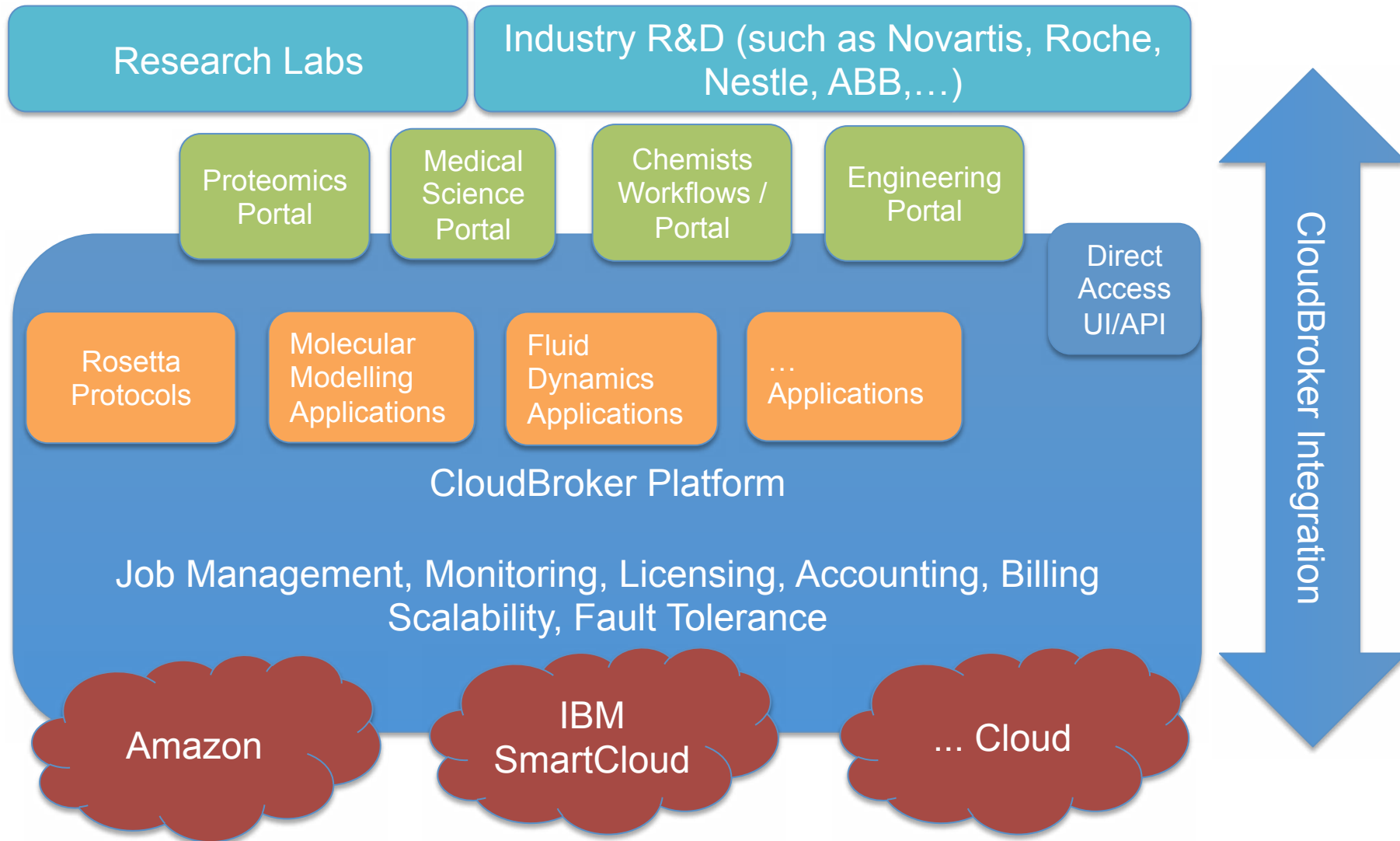


### Activities:

- The Client can download the calculated result files from the block storage attached to the storage instance in the IBM Cloud at regular intervals.

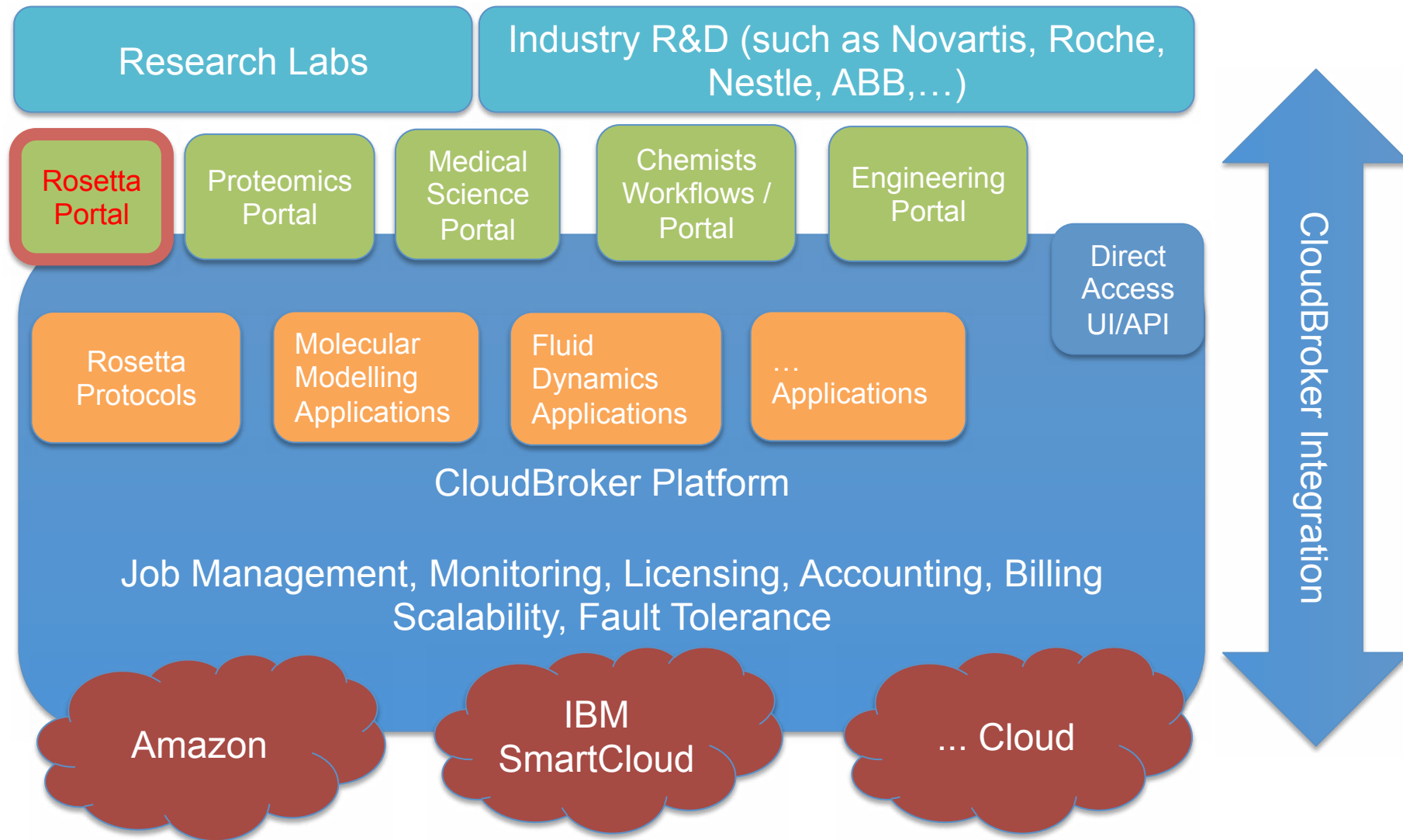
# CloudBroker Platform:

High Performance Applications on the Cloud ready to use for Life Science and Engineering Industry R&D



# CloudBroker Platform:

High Performance Applications on the Cloud ready to use for Life Science and Engineering Industry R&D



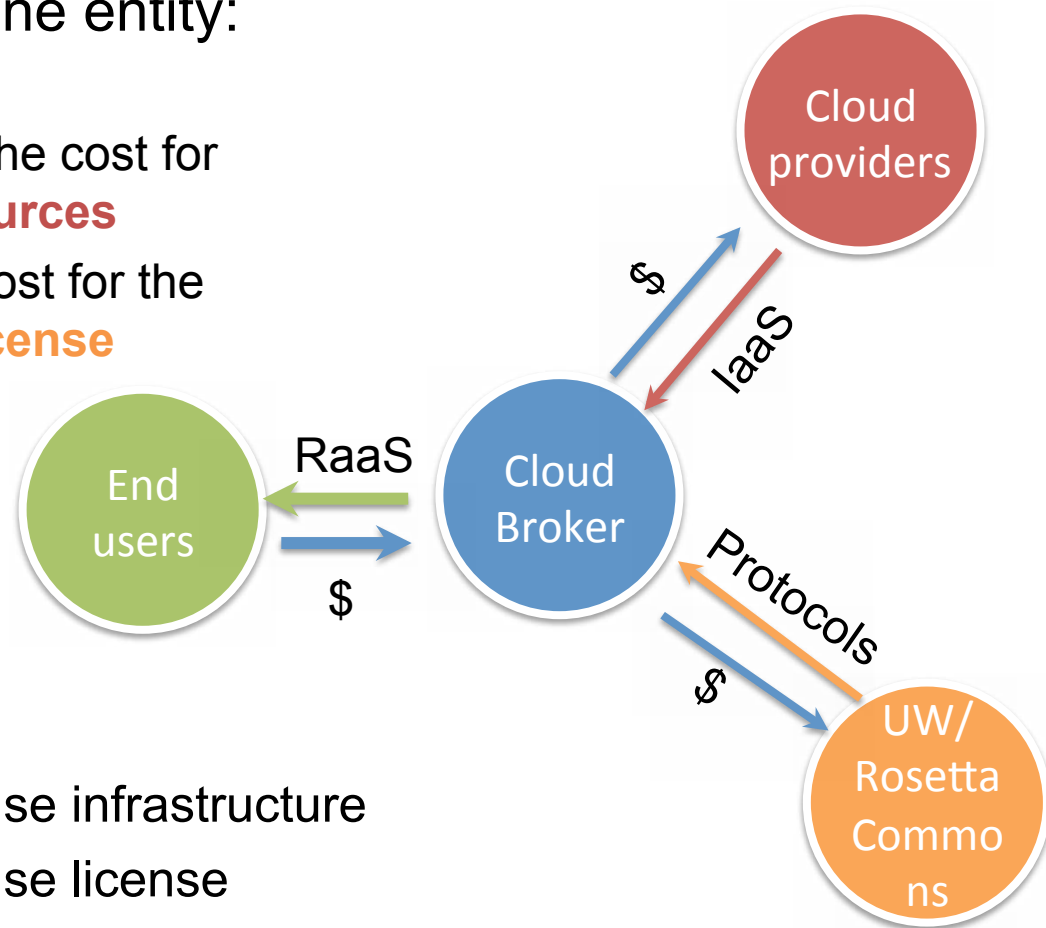
# Results

Structural domains modeled on the Cloud	249 out of 832 de-novo
Number parallel Cloud CPU resources	up to 1008
Net CPU hours processed	~ 249'000
Number of jobs	~ 36'000
Number of models	~ 2'300'000

# Who gets what money

End users just deal with one entity:  
CloudBroker

- CloudBroker manages the cost for commercial **cloud resources**
- CloudBroker adds the cost for the commercial **Rosetta License**



Users pay for  
**Rosetta as a Service**

- Transparent pay-per-use infrastructure
- Transparent pay-per-use license
- CloudBroker surcharge

# Summary

- Huge simplification in the usage of Cloud Resources
- No installation and configuration just set up inputs and run the client
- Capture computing environment from which a particular paper was created, have that available on demand



# Conclusion: Great Potential

- *As a User:* I can read a paper and try out the protocol **immediately** using my own data or just rerun the result of the paper
- *As an Author:* If I publish a protocol on the cloud I probably will get more citations and **recognition** for my work because people can use it easily
- **Lower barrier** to usage of Rosetta by academia and industry massively
- Commercial Cloud licensing can be very easily set up, potentially more income for Rosetta Commons  
😊

# What Next

- Make it possible to set up several protocols as an 'APP' each
  - Create several Rosetta Cloud Apps from well established publications, also as templates for others
  - Create dedicated Rosetta Gateway using the CloudBroker platform
  - See that it works also with university resources
- We will create a dedicated portal – SCIBUS project, starting October 2011

# Acknowledgements

## CloudBroker

- Wibke Sudholt
- Nicola Fantini
- Andrey Sereda

## IBM Switzerland

- Stefan Ruckstuhl
- Marcel Lautenschlager
- Roland Reifler