



## WeNMR / HADDOCK

HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven flexible docking approach for the modelling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modelling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.

### Goal

Because of the computational demands of the HADDOCK user group, the availability of easy accessible, affordable and compute resources is important. The portal used by a large worldwide community and availability/reliability of the service is crucial. The Helix Nebula science cloud would provide the opportunity to add these resources in a way that is seamless and invisible to the end user. To reach this outcome, the resources in the Helix Nebula cloud must be made available to the current job scheduler of the HADDOCK project. A version of the HADDOCK portal should be created, of which is portable to all cloud providers in Helix Nebula science cloud.

Long-term objectives include:

- » Service availability independent from a single datacenter or infrastructure.

### Preconditions

1. Batch system installed and preconfigured with the number of cores per node equal to number of queue slots per node (TORQUE/ Maui or SLURM are the preferred schedulers)
2. Since some pre-processing steps involving short calculations and IO operations take place during the submission process, the master node should have some computational resources (e.g. minimum of 4 cores).
3. To test in production mode, sufficient computational resources should be available on the nodes (e.g. bare minimum of 50 cores, best in the range of 100-200)
4. Also ideally with a number of different queue lengths (short up to 4 hours / medium up to 8 hours / long > 8 hours) with a configurable number of slots per queue type
5. NFS mount of the home partition on all worker node VMs, NFS export provided by master VM, password-less access to nodes
6. Sufficient /tmp space on the nodes (e.g. 20 GB) to allow temporary writing during the computations to minimize network traffic.
7. Sufficient space of the /home partition to install software and store data (ideally 1TB)
8. The server should be accessible via standard https port (443), which means that a valid server certificate should be present.

- » Failover scenarios.
- » Long-term sustainability.
- » Larger amount of available compute resources
- » Scalability to efficiently handle growth/ fluctuations in usage of the service

9. A recent version of docker should be installed to facilitate deployment (udocker could be used to run the docker service without elevated privileges)

## The Challenge

The HADDOCK portal is now running on baremetal on two servers at Utrecht University. The front-end of the server is being rewritten in the Flask framework and will be deployed in a modular way via docker-compose, which should make virtualisation/porting easy. The machinery behind the frontend does require sufficient computational resources to perform the actual computation. Ideally this use case would like to virtualize the complete HADDOCK portal on the HNSciCloud infrastructure.

## From the user perspective

Since the load on the server might vary, an elastic cluster configuration would be ideal, in which nodes can be automatically added / removed, dependent of the load on the batch system. This means that the queue configuration should also be dynamically changed when the cluster grows or shrinks.

## Benefits and impact

The main impact of deploying the heavily used HADDOCK portal on a hybrid public-commercial cloud is mainly two-fold:

- 1) Ensuring the continuity of the service, especially if grid computing would become obsolete
- 2) Enabling the provision of a self-contained server which could be exclusively used by interested commercial users (privacy and security are issues here, especially for pharma companies which in most cases are not allowed to use external services)

There are currently more than 10000 users worldwide.

## Procurer sponsoring the use case: SURFsara

SURFsara creates a bridge between research and advanced ICT. We do so with a passion for scientific research in our DNA and with extensive expertise contained in our high-performance infrastructure. This enables us to facilitate scientific research and develop initiatives for the business community.

