Large Scale Binding Affinity Calculations on Commodity Compute Clouds

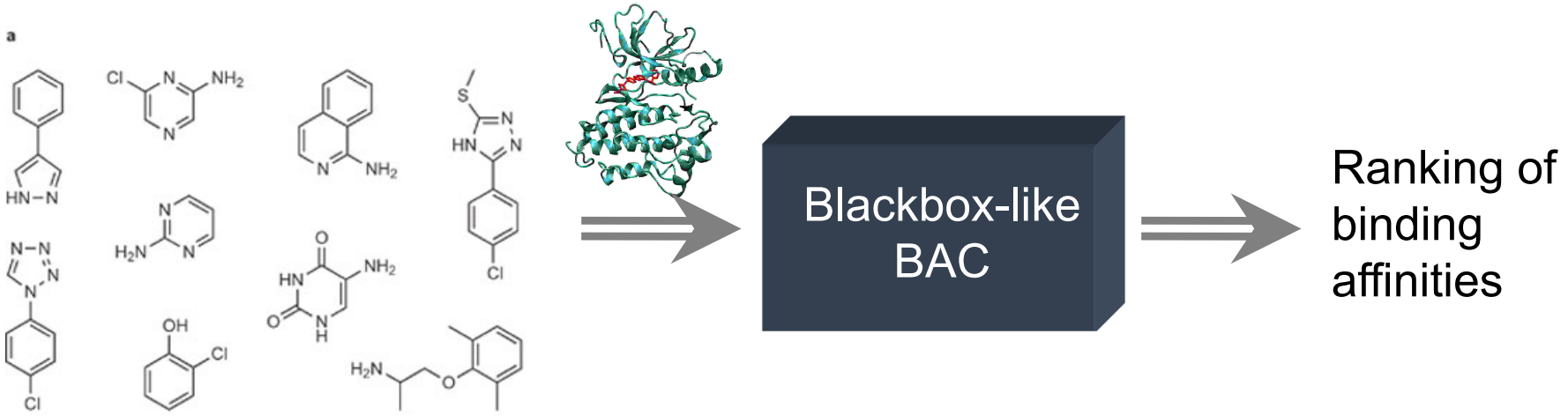
Stefan Zasada

EnsembleMD Ltd

Binding affinity calculator (BAC)

EnsembleMD

BAC can reliably predict binding affinities of compounds with their target proteins, and be used potentially as a drug ranking tool in clinical application or a virtual screening tool in pharmaceutical lead discovery.



*S. K. Sadiq, D. Wright, S. J. Watson, S. J. Zasada, I. Stoica, Ileana, and P. V. Coveney, "Automated Molecular Simulation-Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases", Journal of Chemical Information and Modeling, **48**, (9), 1909-1919, (2008), DOI: 10.1021/ci8000937.*

Free Energy Methodologies

EnsembleMD

ESMACS: Enhanced Sampling of Molecular dynamics with the approximation of Continuum Solvent

- "Absolute" free energies
- Compare diverse ligands
- Statistical error analysis
- MM/PBSA
- Conformational entropy
- 1, 2 and 3 trajec

TIES: Thermodynamic Integration with Enhanced Sampling

- Relative binding affinities
- "Exact"
- Limited range of application
- Ensemble needed for each λ window

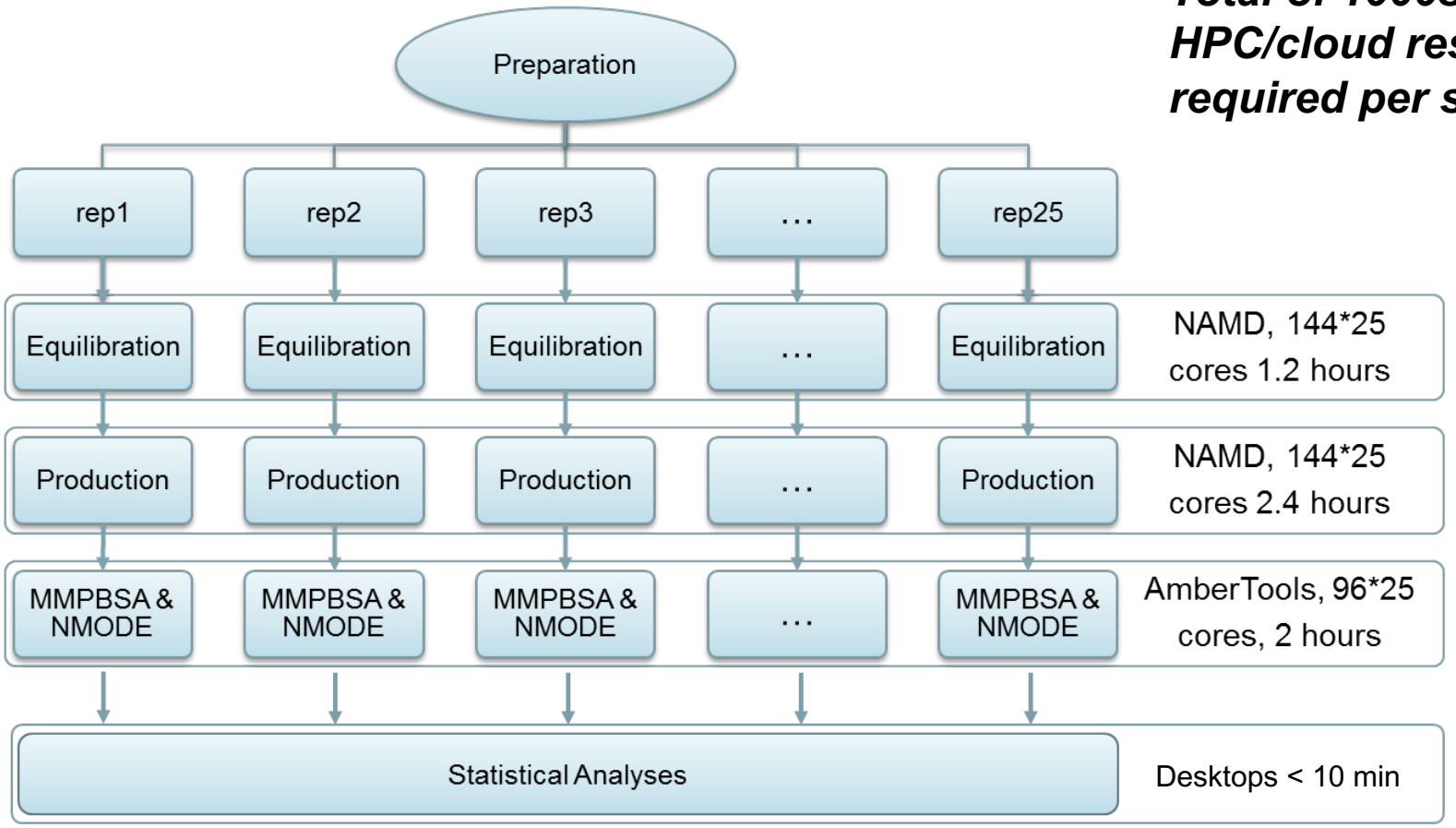
P.V. Coveney & S. Wan, "On the calculation of thermodynamic properties from molecular dynamics", *Phys Chem Chem Phys* **18**, 30236 (2016)

BAC Workflow: ESMACS Method

EnsembleMD

BAC: rapid and accurate binding affinity calculation on timescales relating to clinical decision making on drug selection and to pharmaceutical lead discovery.

Total of 1000s cores on HPC/cloud resources required per study



~30MB

~15GB

~30GB

~1MB

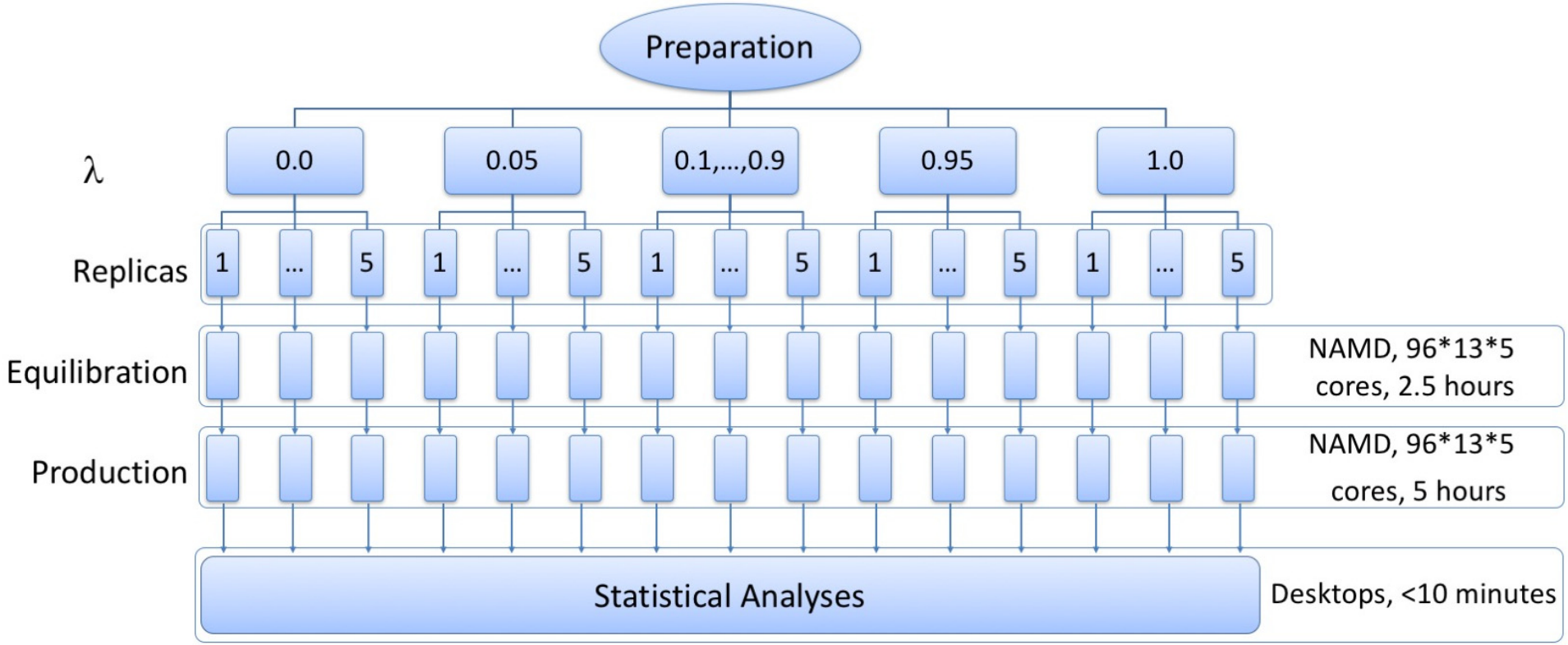
= ~50GB

A. Bhati, S. Wan, D. Wright, P. V. Coveney, "Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration", Journal of Chemical Theory and Computation, DOI: 10.1021/acs.jctc.6b00979.

BAC Workflow: TIES Method

EnsembleMD

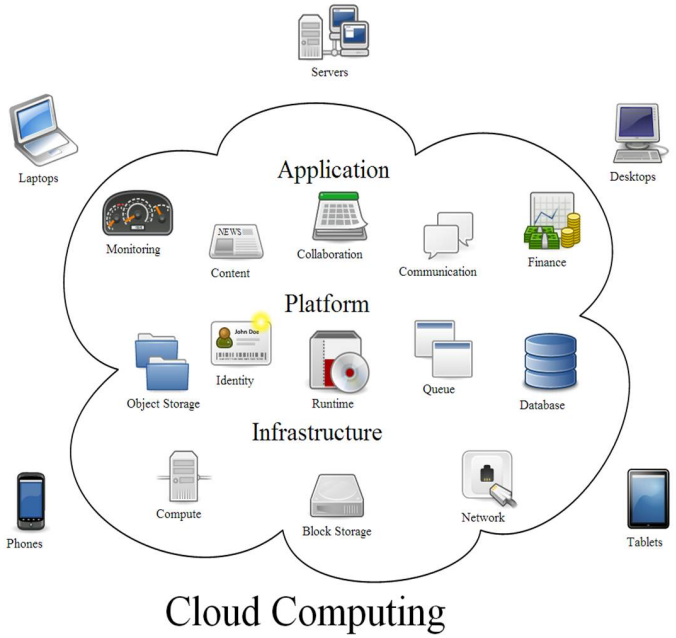
Binding Affinity Calculator (BAC) is a software toolkit which automates the implementation of TIES (and ESMACS) methods for binding affinity calculations



S.K. Sadiq, D. Wright, S.J. Watson, S.J Zasada, I. Stoica, P.V. Coveney, J. Chem. Inf. Model., **2008**, 48, 1909-1919.
A.P. Bhati, S. Wan, D.W. Wright & P.V. Coveney, *J. Chem. Theory Comput.*, **2017**, 13, 210–222.

Commercial Clouds

- Cloud computing is an alternative scheme for running applications on remote resources.
- Access to compute is provided in return for monetary payment.
- *Infrastructure as a Service* (IaaS) clouds provide access to CPU, memory and storage.
- *Software as a Service* (SaaS) clouds provide access to applications.



We've deployed BAC on Azure, AWS and DNAnexus cloud platforms

ufBac

EnsembleMD

- User Friendly BAC makes reliable, repeatable binding affinity calculations available to anyone
- Web interface allows full BAC workflows via simple, user friendly client
- Manages execution of calculations on a range of resources, from HPC to commercial cloud platforms
- ufBac manages complete study execution and data archival on behalf of the user

Cloud Deployments

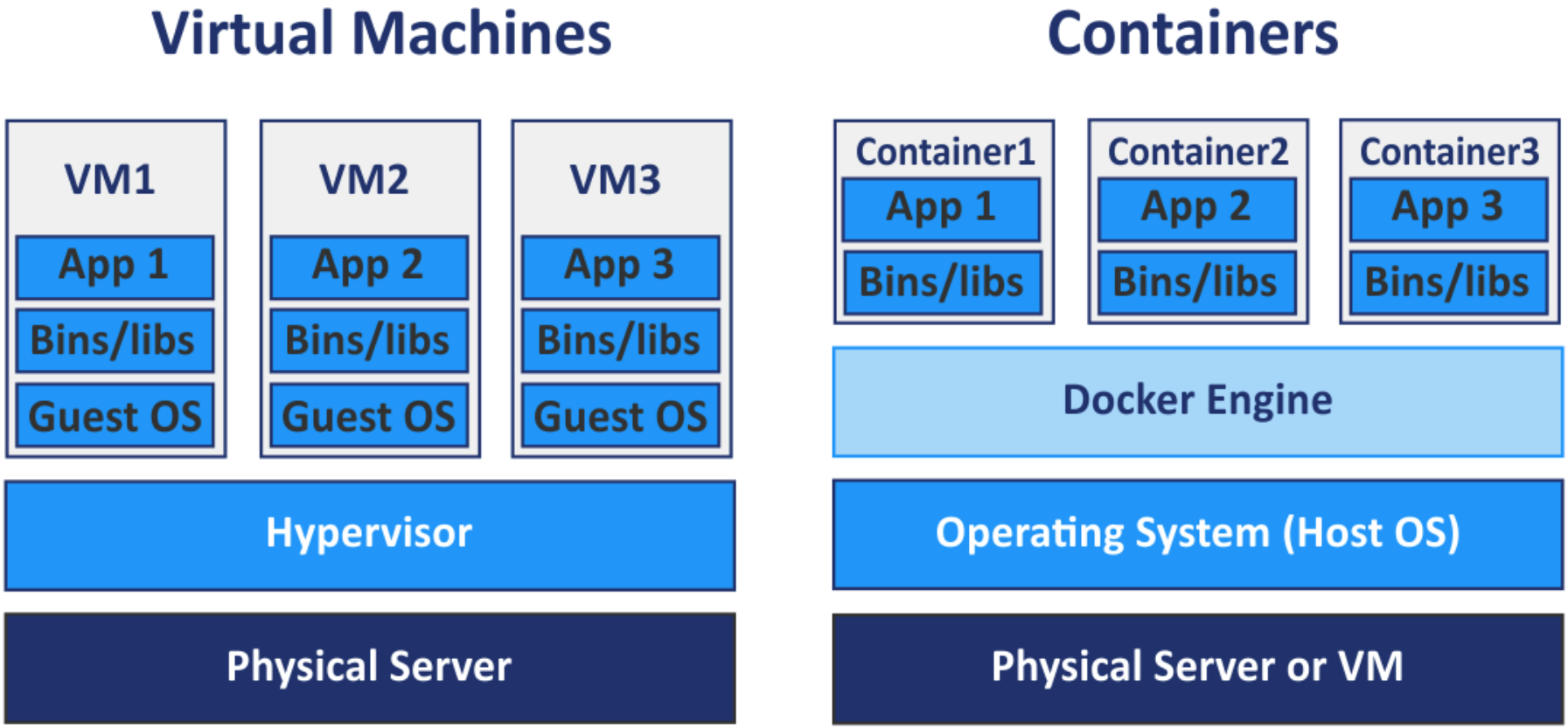
- **Docker** is used to create containers – self container systems containing applications and all dependencies
- Each component of the BAC workflow is encapsulated in a docker container and stored in a **container registry**
- **Kubernetes** is used to deploy GPU and CPU clusters on cloud which run different parts of the workflow

  → **this allows BAC to be easily ported between clouds**

Docker

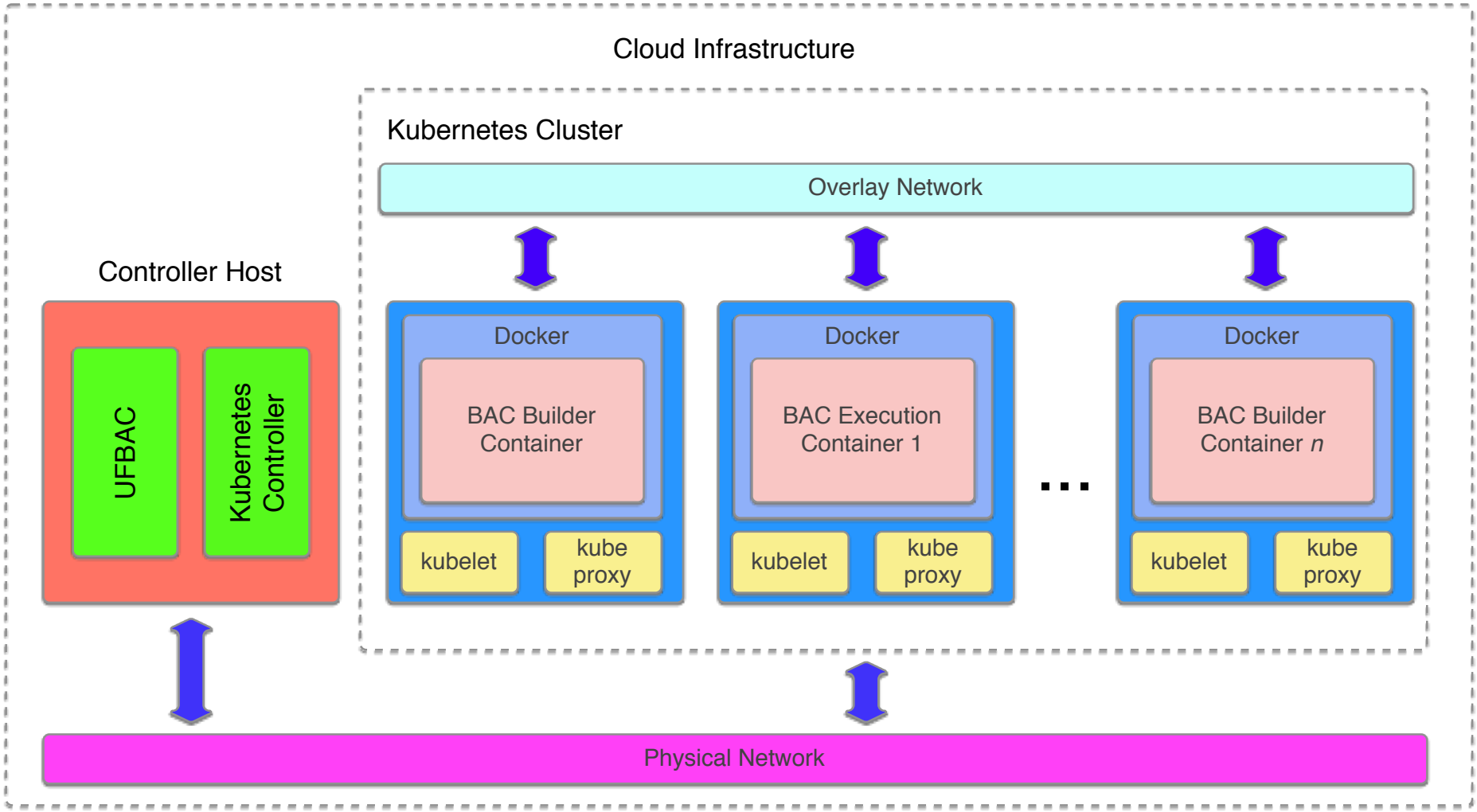
EnsembleMD

Docker is a set of platform-as-a-service products that use OS-level virtualization to deliver software in packages called containers. Containers are isolated from one another and bundle their own software, libraries and configuration files



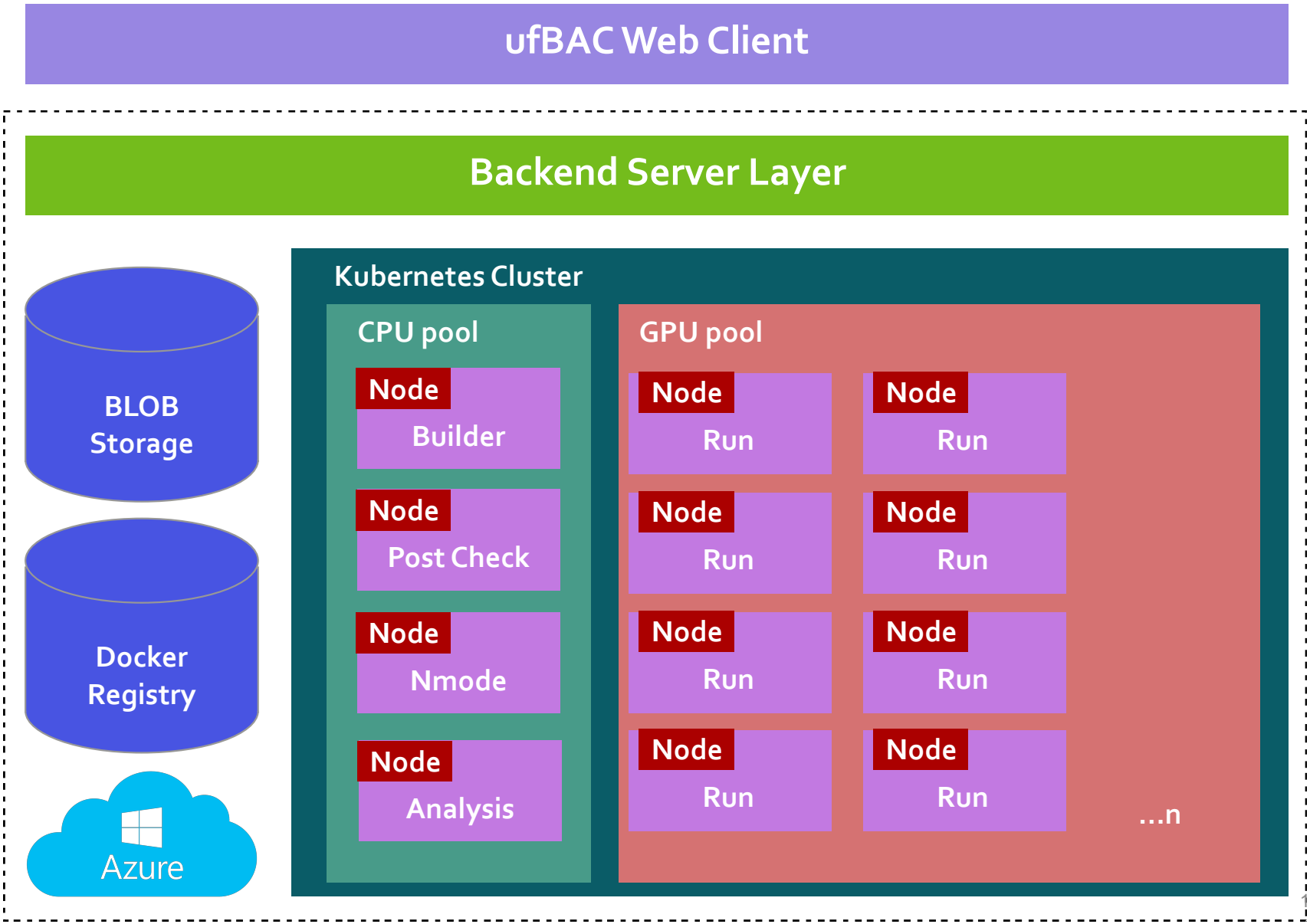
BAC on Kubernetes

EnsembleMD

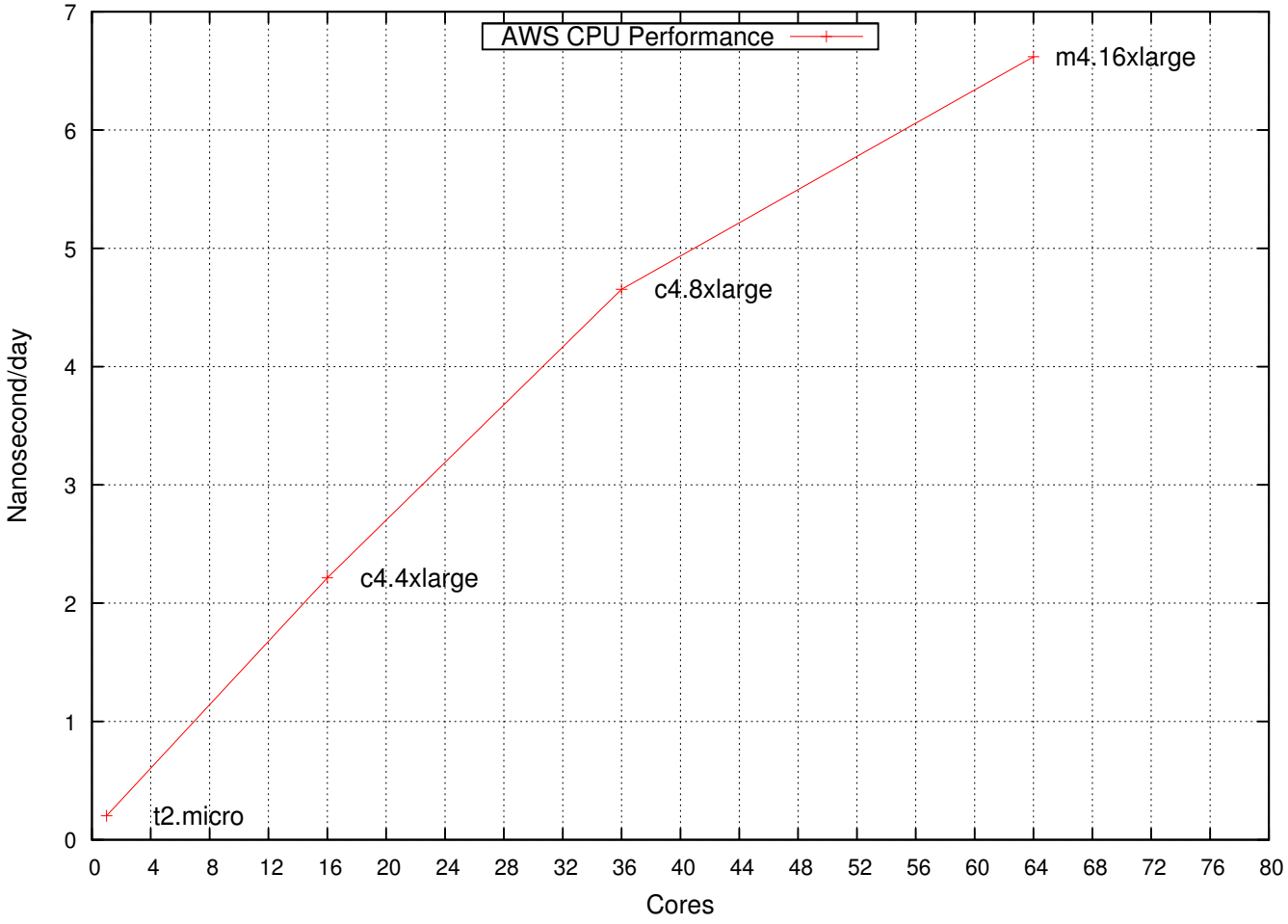


Cloud Architecture

EnsembleMD



Cloud Scaling Performance



Instance	Size
t2.micro	1 CPU/1 GB RAM
c4.4xlarge	16 CPU/30 GB RAM
c4.8xlarge	36 CPU/60 GB RAM
m4.16xlarge	64 CPU/256 GB RAM

Cloud Deployment v HPC Investment

EnsembleMD

- Cloud is ‘elastic’ – should always have enough resource to run your workflow.
- Public HPC runs on first come first served queue
- Cloud companies have invested in security architectures beyond those provided by research HPC organizations.
- Cloud gives access to resources at scale to ‘try things out’.
- However, in production cloud is currently expensive. Can make sense to invest in HPC locally for large scale use.
- Learning curve is steep – you’re on your own and typically need to install all software and libraries

Conclusions

- Reliable, accurate and reproducible binding affinity ranking can be obtained only by ensemble simulations.
- Drug ranking is performed by a novel approach to free energy estimation.
- **Scalable approach**: results in hours.
- ufBAC hides the complexity of the workflow in a simple, web accessible application

Future for pharma in the balance

- Invest in HPC vs use cloud solutions.