

Analytics for Molecular Dynamics (A4MD)

Towards In Situ Enhanced Adaptive Sampling of Simulation Ensembles

Ian Lumsden, Michael Wyatt, Silvina Caíno-Lores, Michela Taufer



Acknowledgements

Funding and support

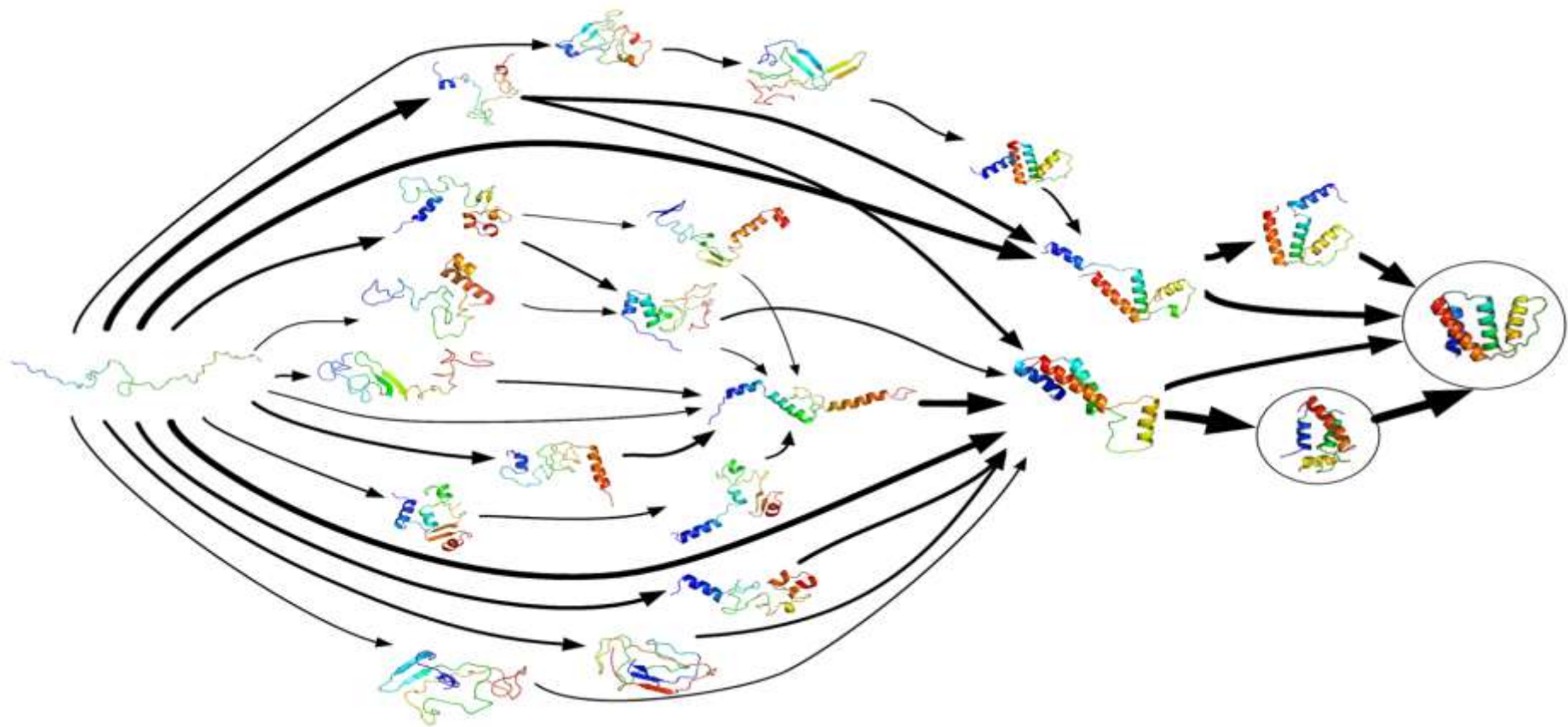
- **IBM** (Global University Program)
- **NSF** (*BIGDATA: IA: Collaborative Research: In Situ Data Analytics for Next Generation Molecular Dynamics Workflows*, NSF award 1741057)

Collaborators

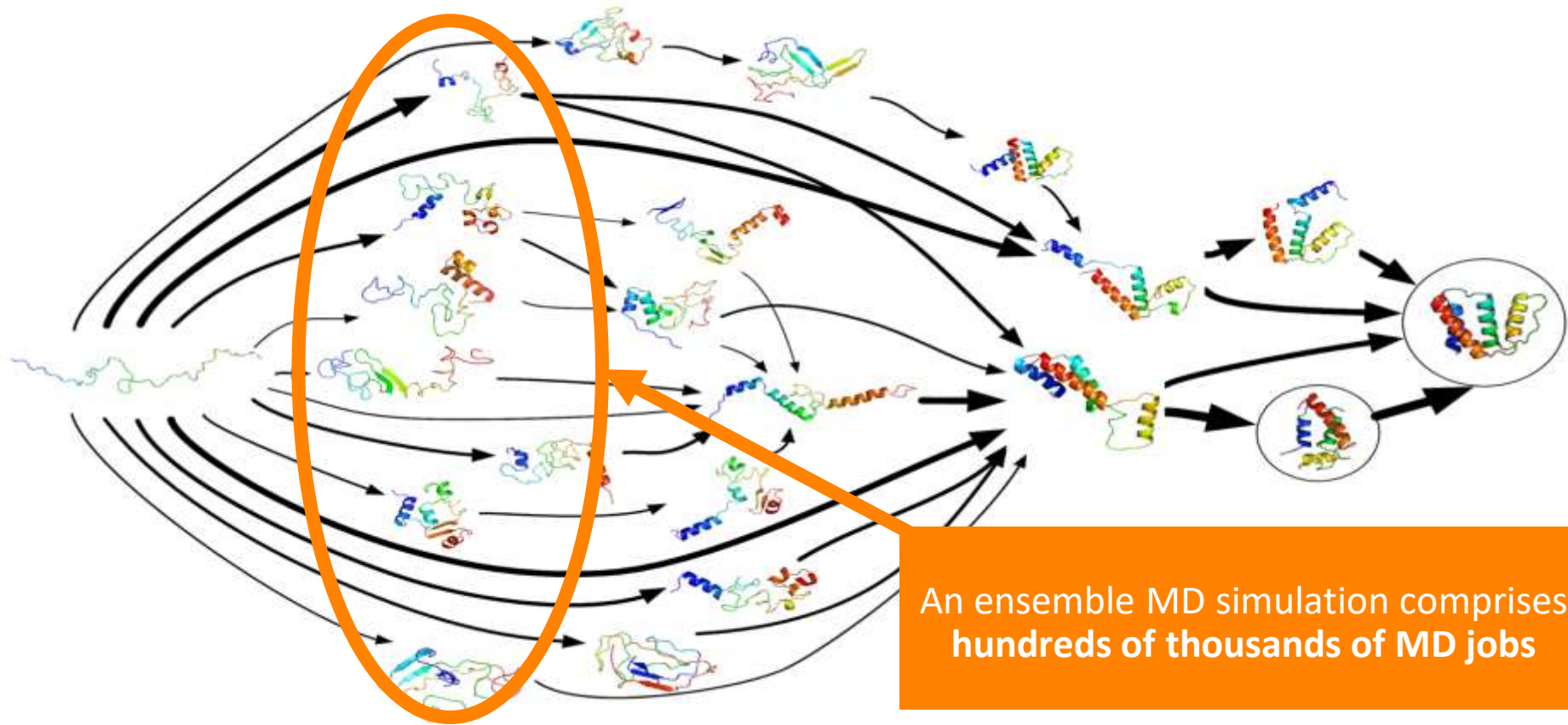
- **Cornell University** (Michel Cuendet, Ekaterina Kots, Harel Weinstein)
- **University of New Mexico** (Hector Alexis Carrillo Cabada, Trilce Estrada)
- **University of Southern California** (Ewa Deelman, Rafael Ferreira da Silva, Tu Mai Anh Do, Loïc Pottier)

And many more!

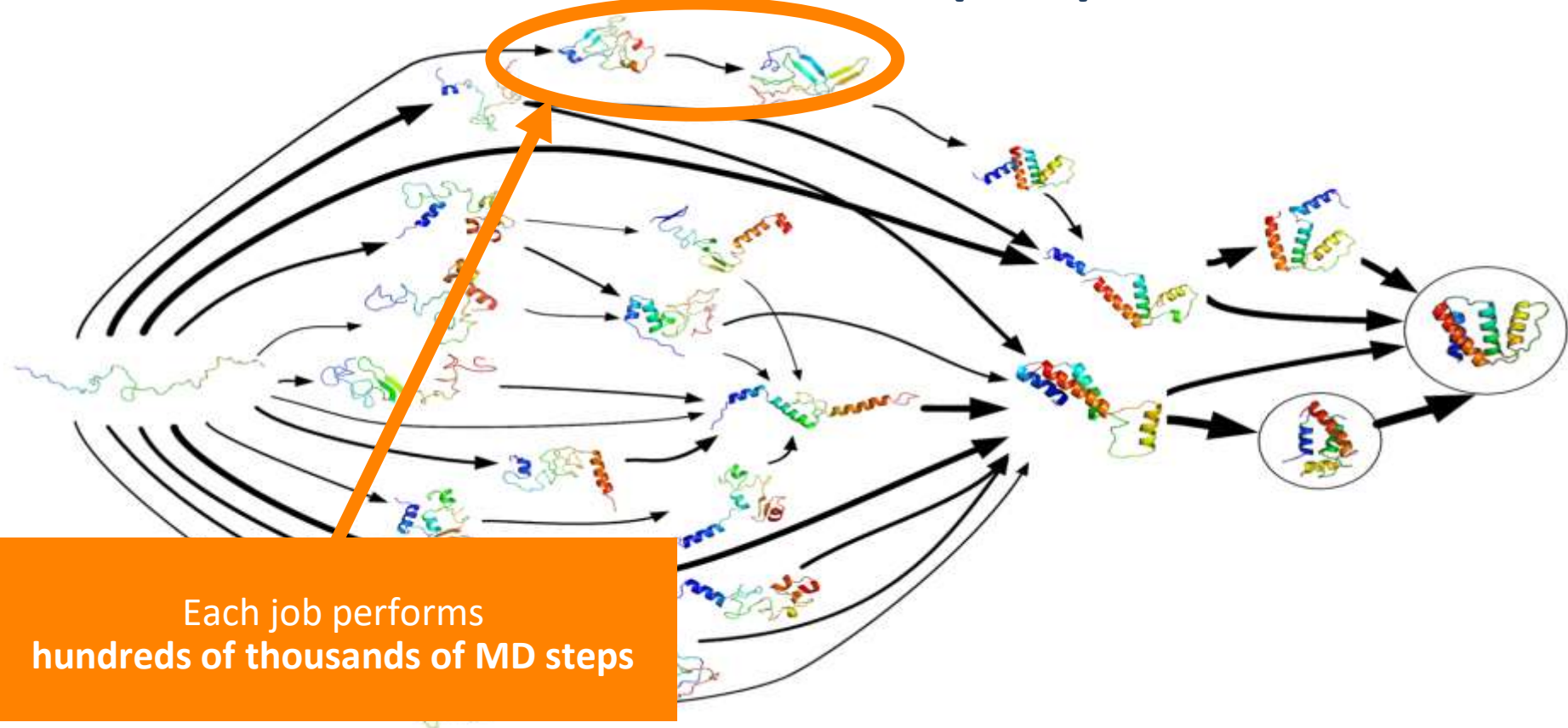
Ensemble Molecular Dynamics (MD) Simulations



Ensemble Molecular Dynamics (MD) Simulations

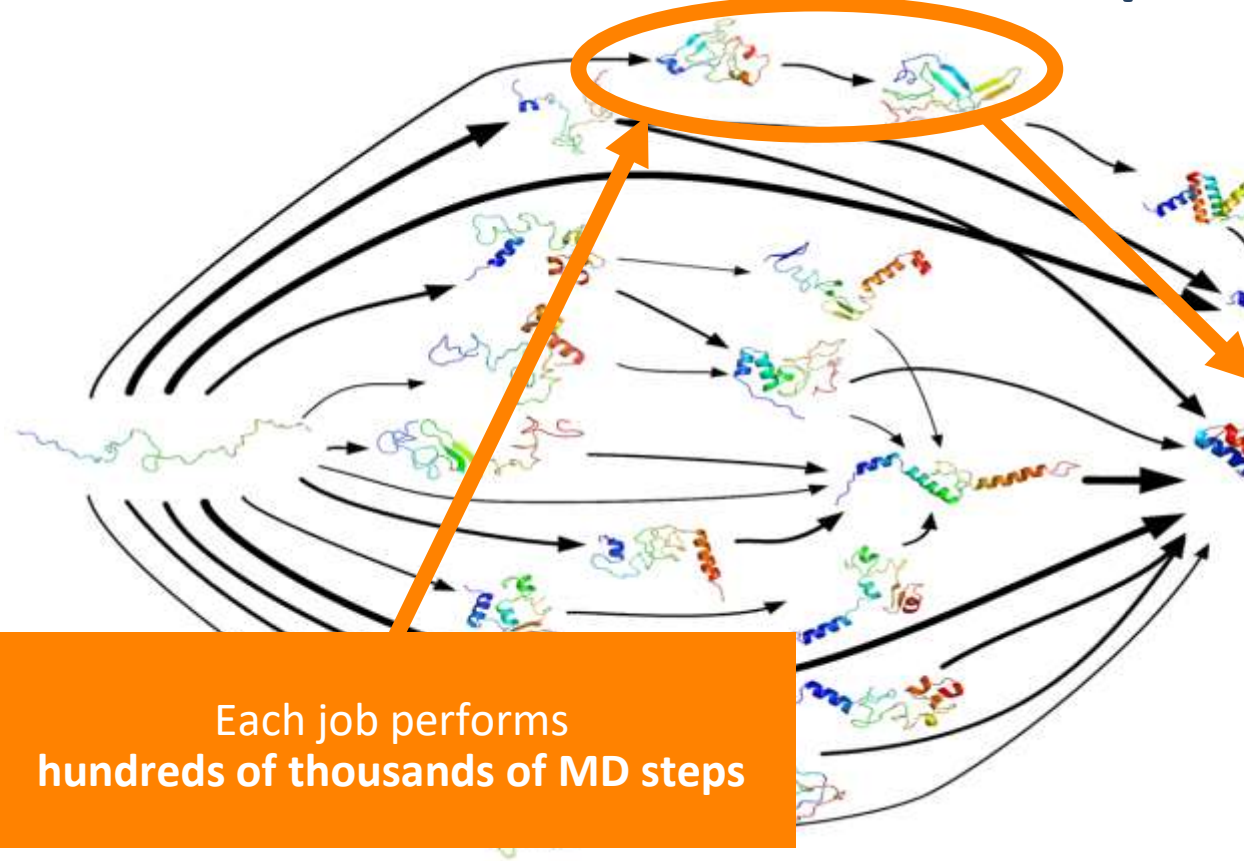


Ensemble Molecular Dynamics (MD) Simulations

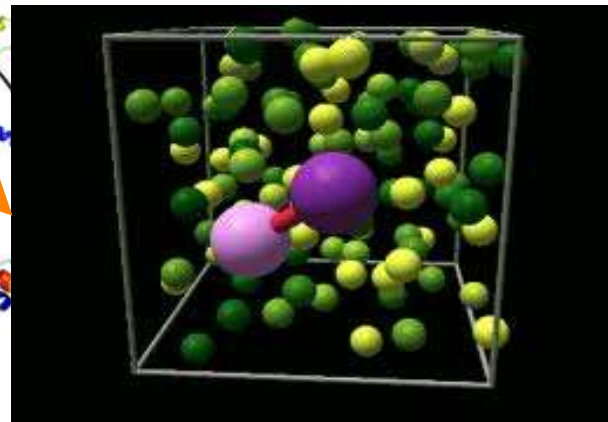


Each job performs
hundreds of thousands of MD steps

Ensemble Molecular Dynamics (MD) Simulations



Each job performs
hundreds of thousands of MD steps



Forces on single atoms

↳ Acceleration

↳ Velocity

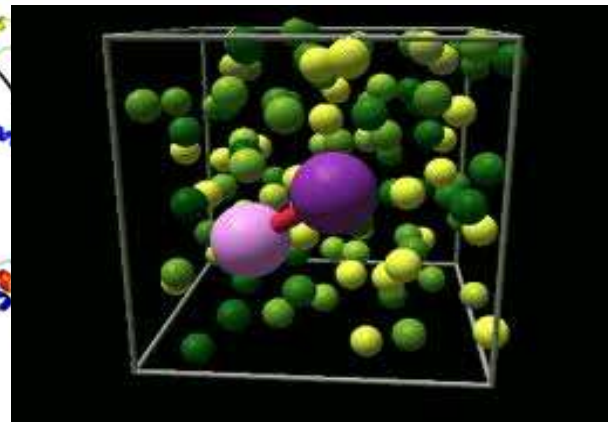
↳

Position

Ensemble Molecular Dynamics (MD) Simulations



Every N steps (*stride*)
a 3D snapshot is stored (*frame*)



Forces on single atoms

↳ Acceleration

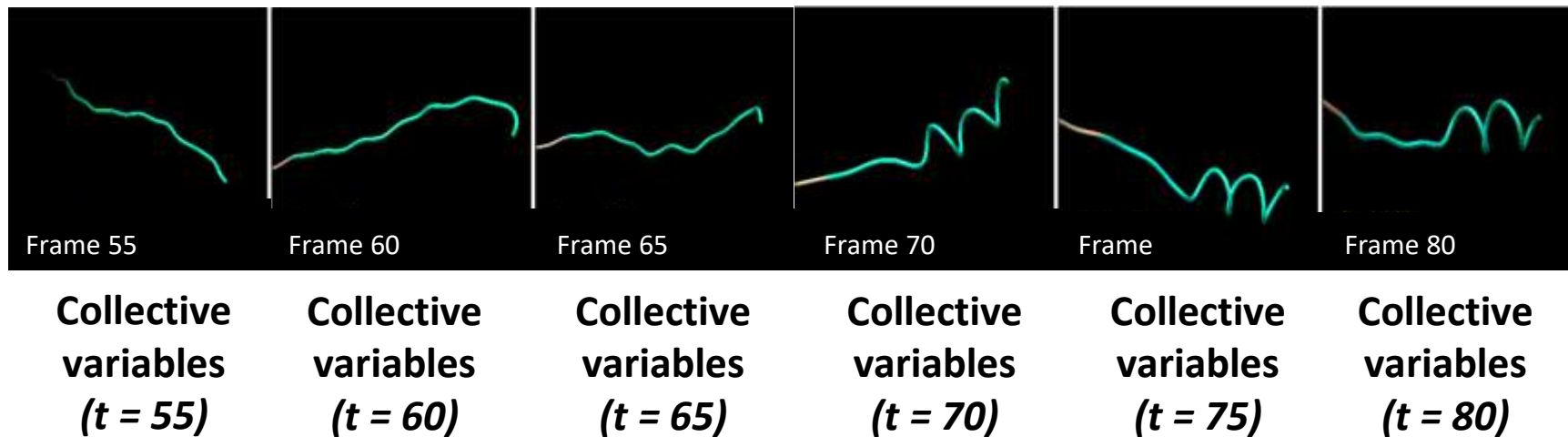
↳ Velocity

↳ Position

Position

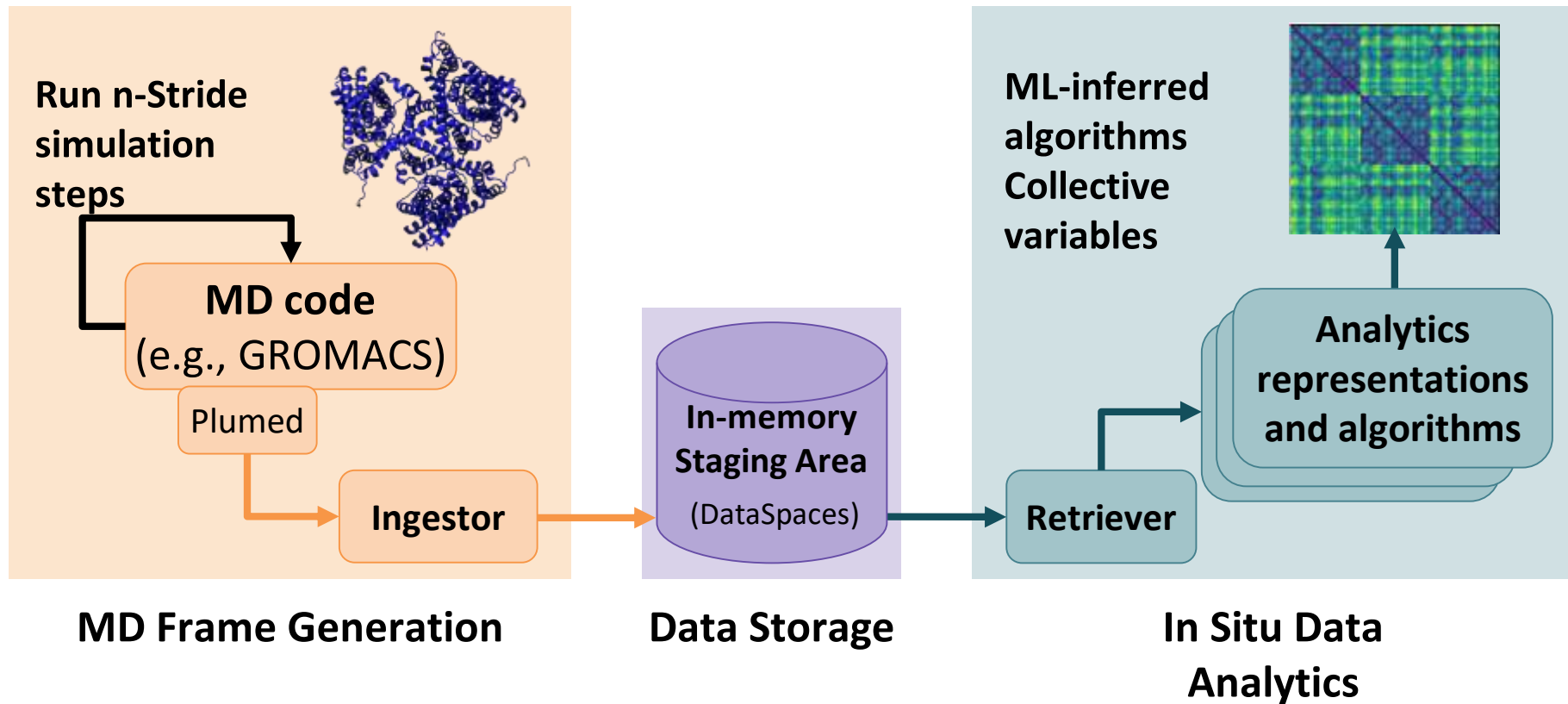
In Situ Analysis of MD Trajectories

Frames of an MD trajectory with a stride of 5 steps:

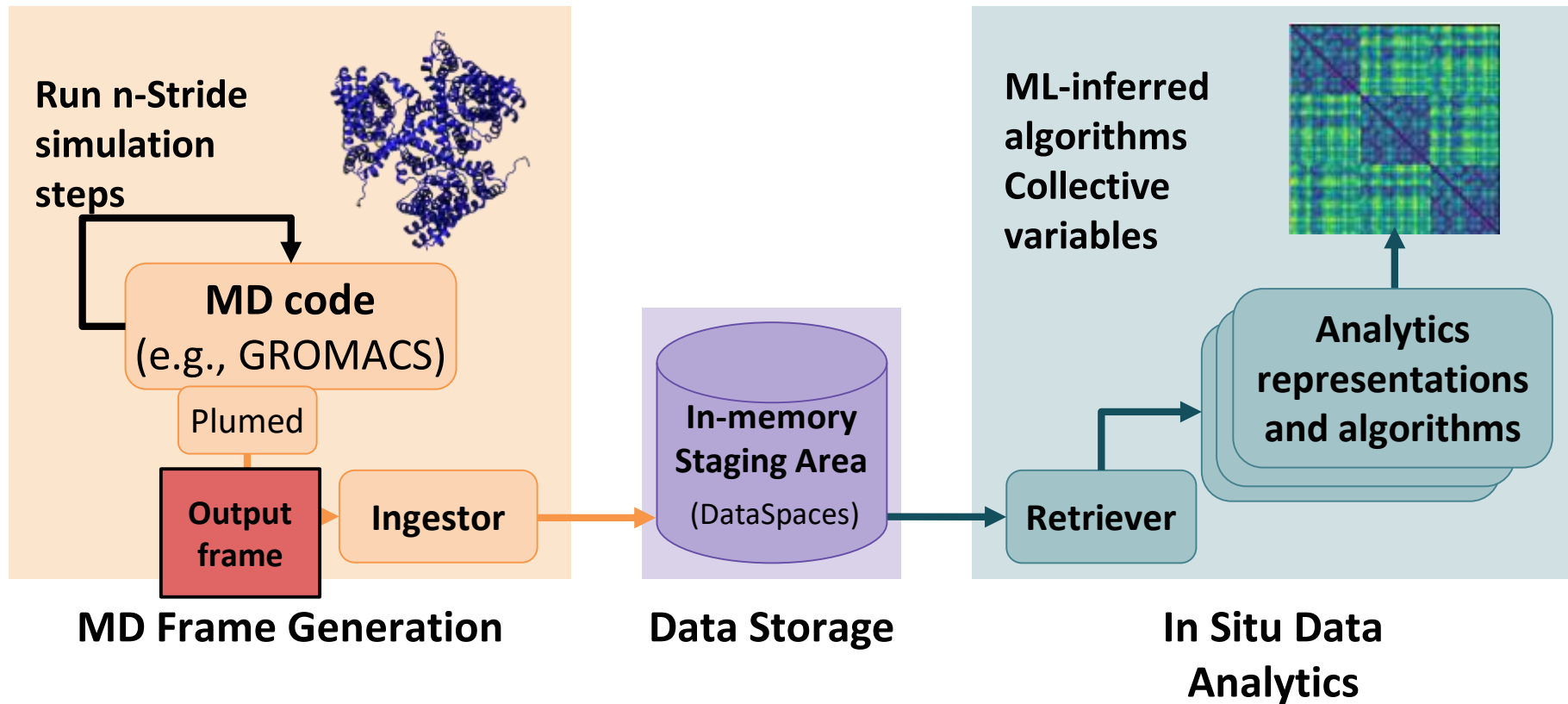


Collective variables serve as proxy for structural and conformational changes

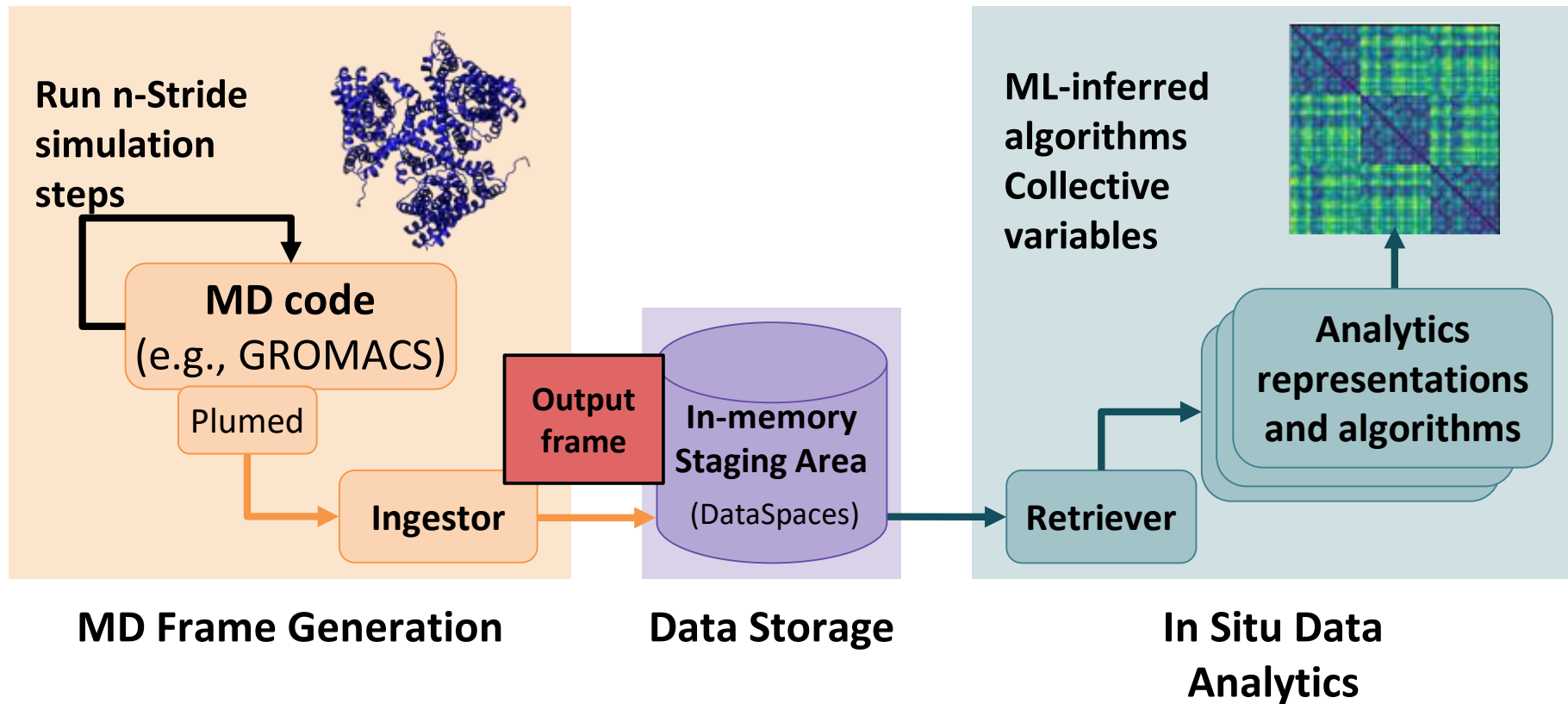
In Situ Analysis of MD Trajectories: A4MD



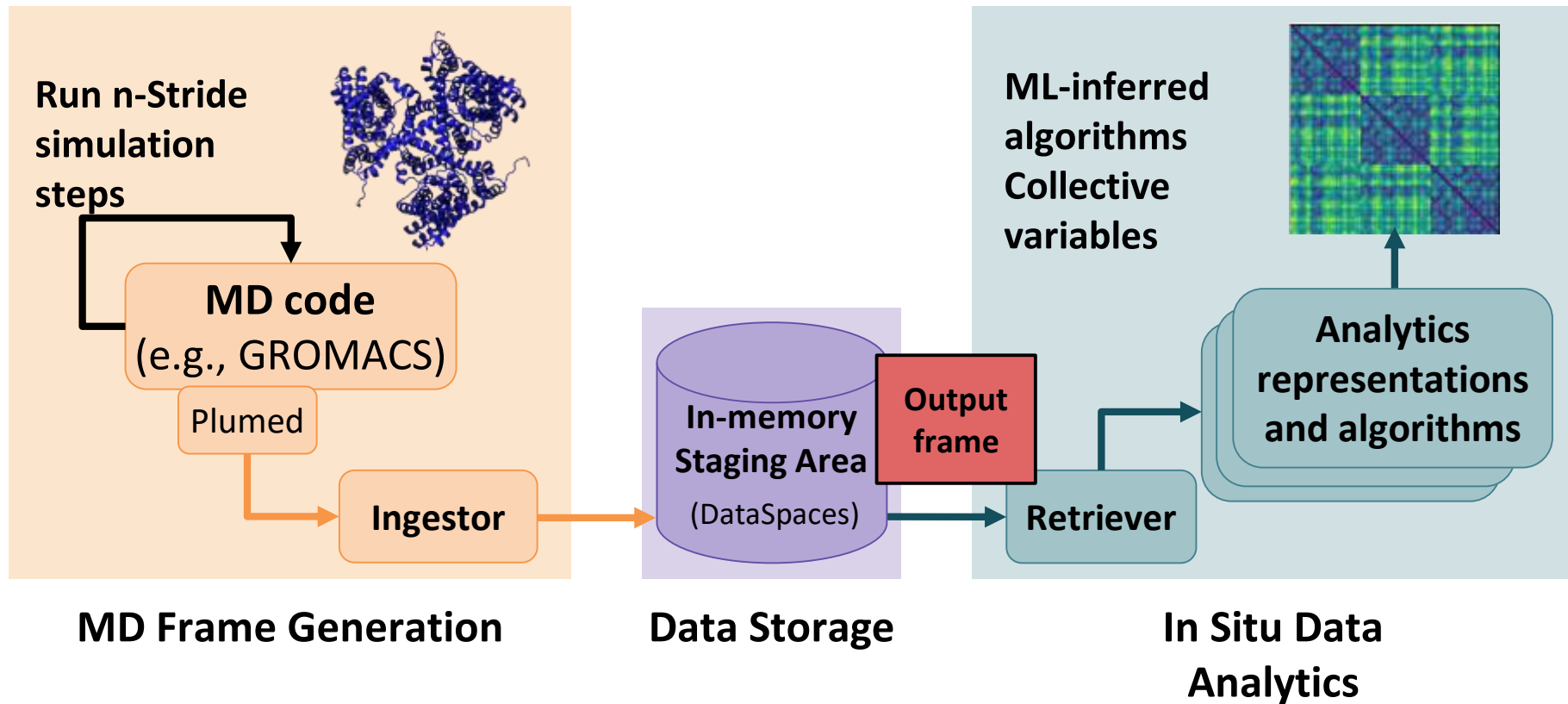
In Situ Analysis of MD Trajectories: A4MD



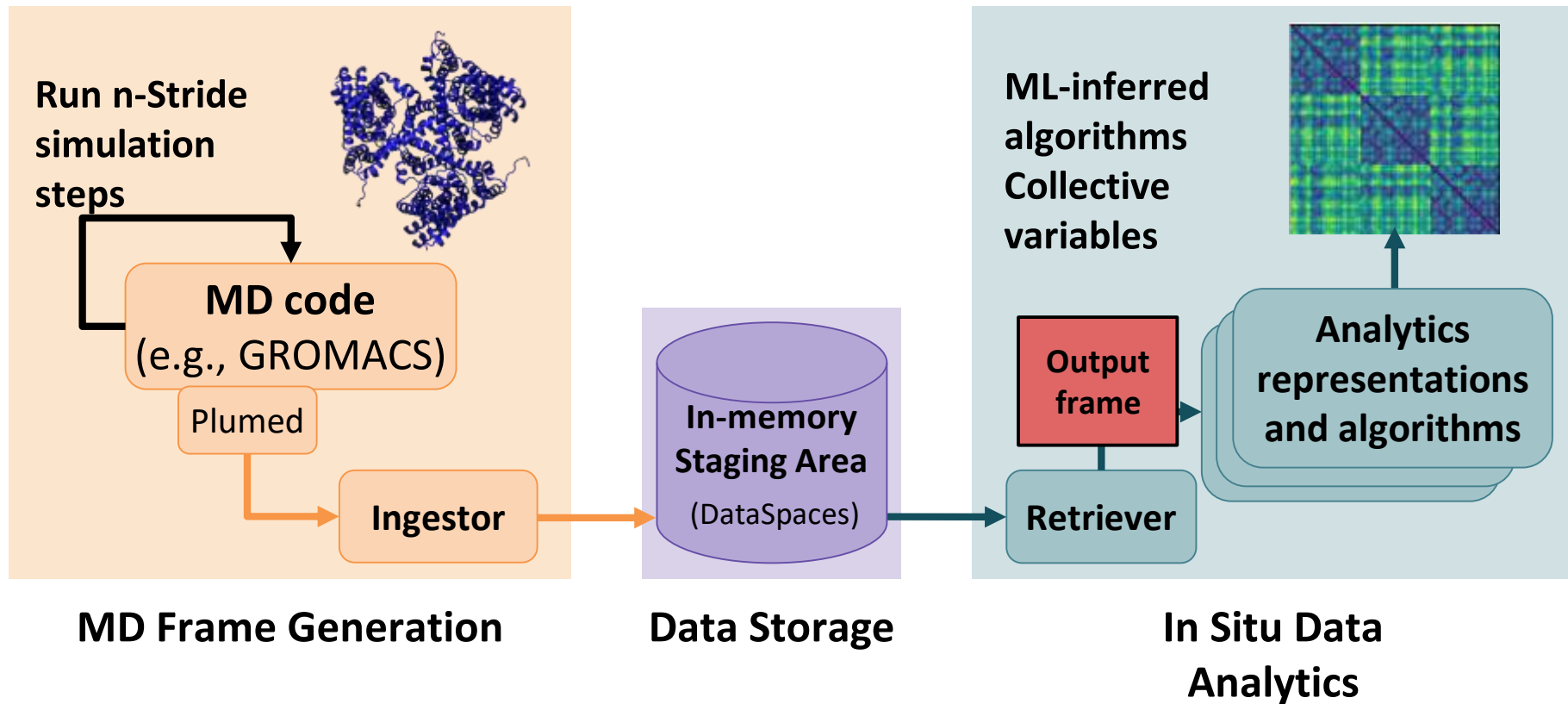
In Situ Analysis of MD Trajectories: A4MD



In Situ Analysis of MD Trajectories: A4MD



In Situ Analysis of MD Trajectories: A4MD



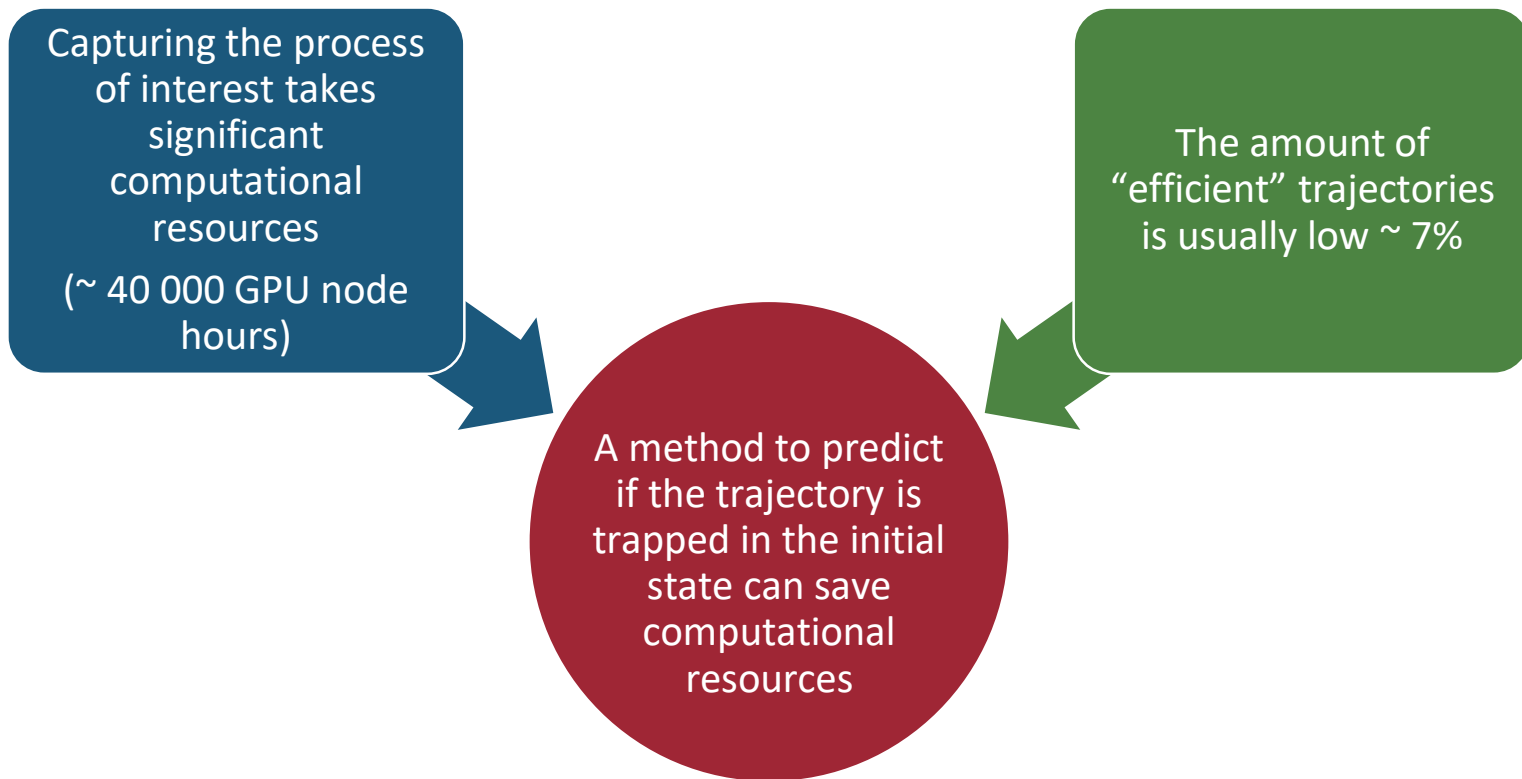
Enhanced Adaptive Sampling (EAS): Motivation

Capturing the process
of interest takes
significant
computational
resources
(~ 40 000 GPU node
hours)

The amount of
“efficient” trajectories
is usually low ~ 7%

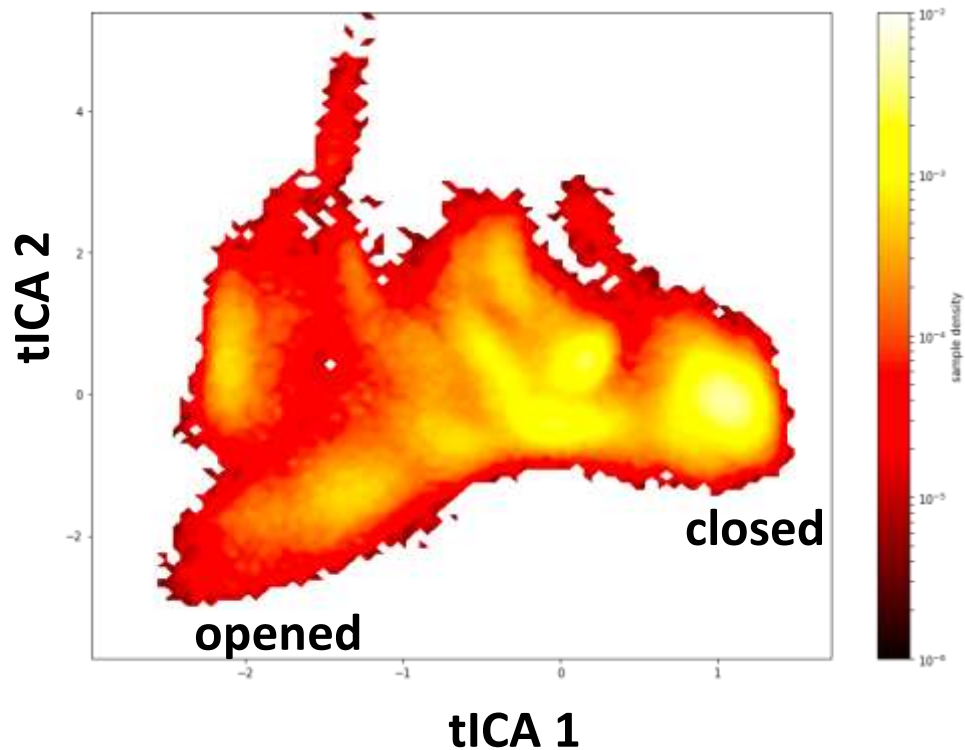
*The numbers presented refer to the BBL system

Enhanced Adaptive Sampling (EAS): Motivation



*The numbers presented refer to the BBL system

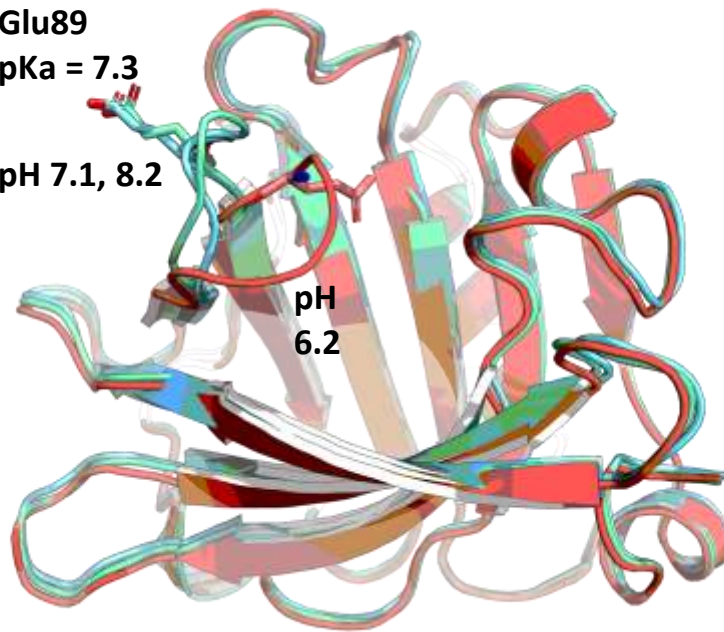
Case Study: Bovine β -Lactoglobulin (BBL) MD



Glu89
pKa = 7.3

pH 7.1, 8.2

pH
6.2

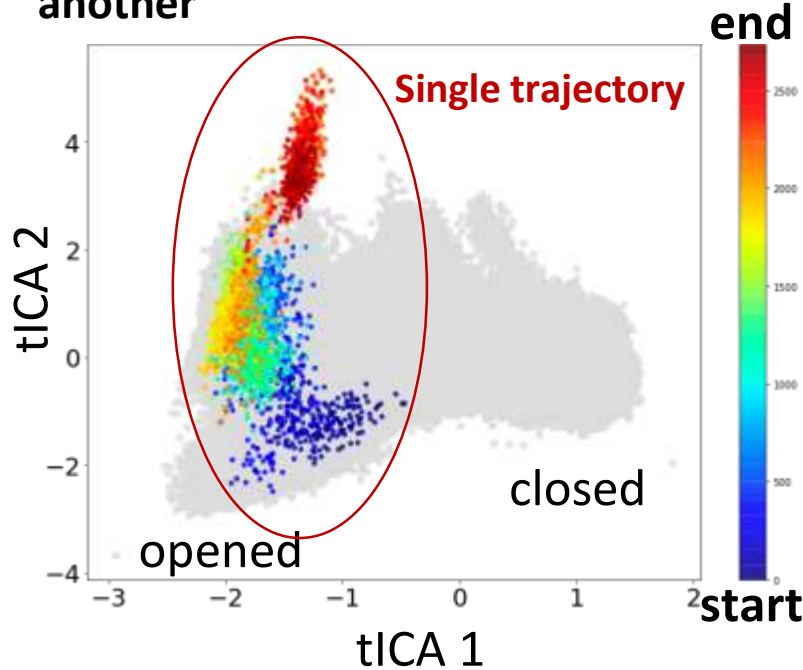


162 residues

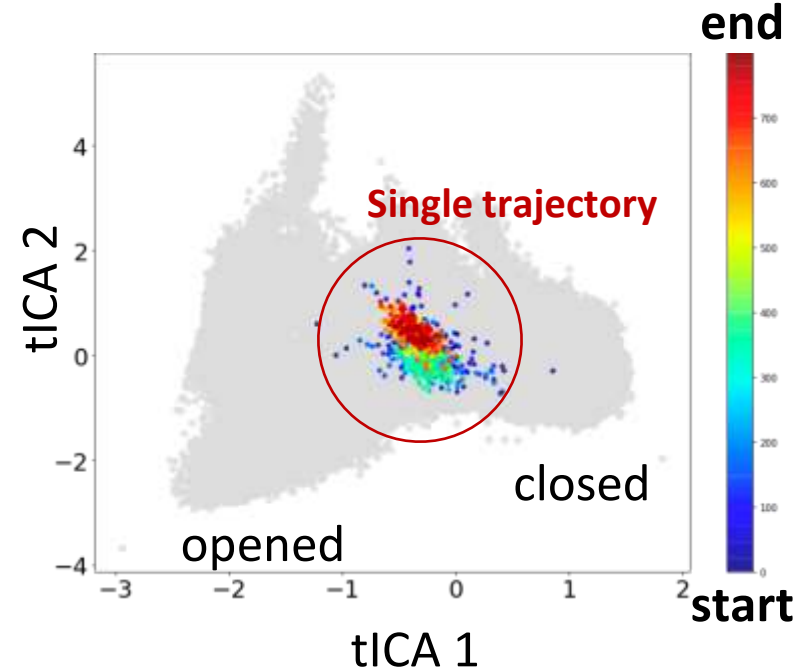
X-ray PDB pH 6.2, 7.1, 8.2

Characterisation of Trajectories

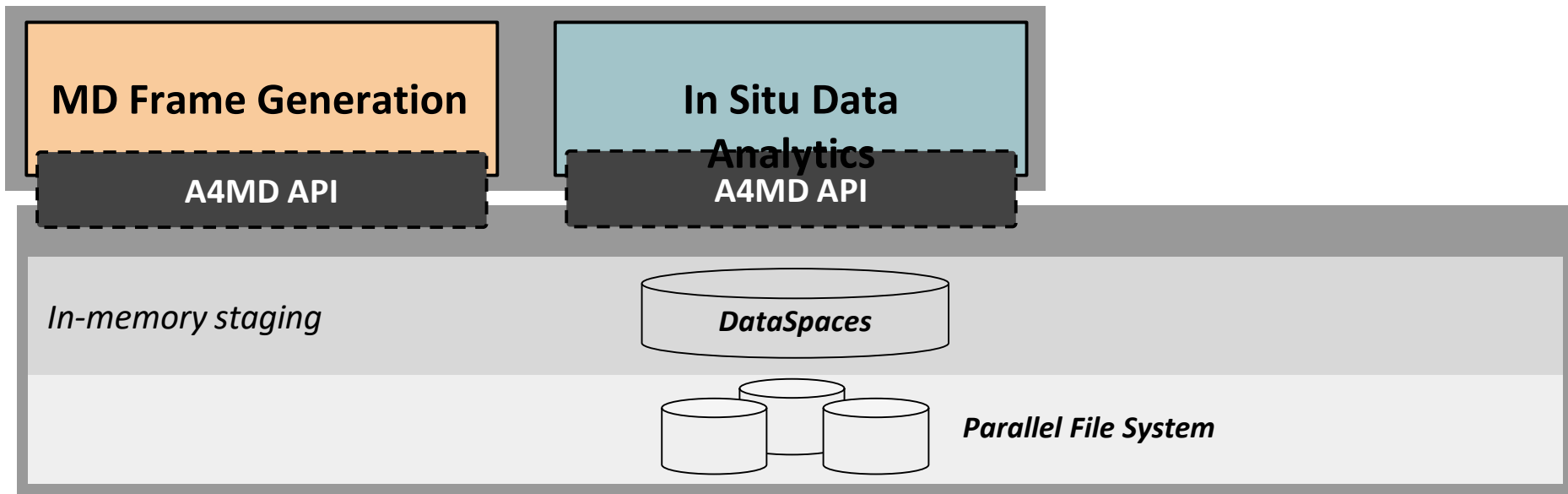
Favorable trajectory: continuous motion from one region of tICA space to another



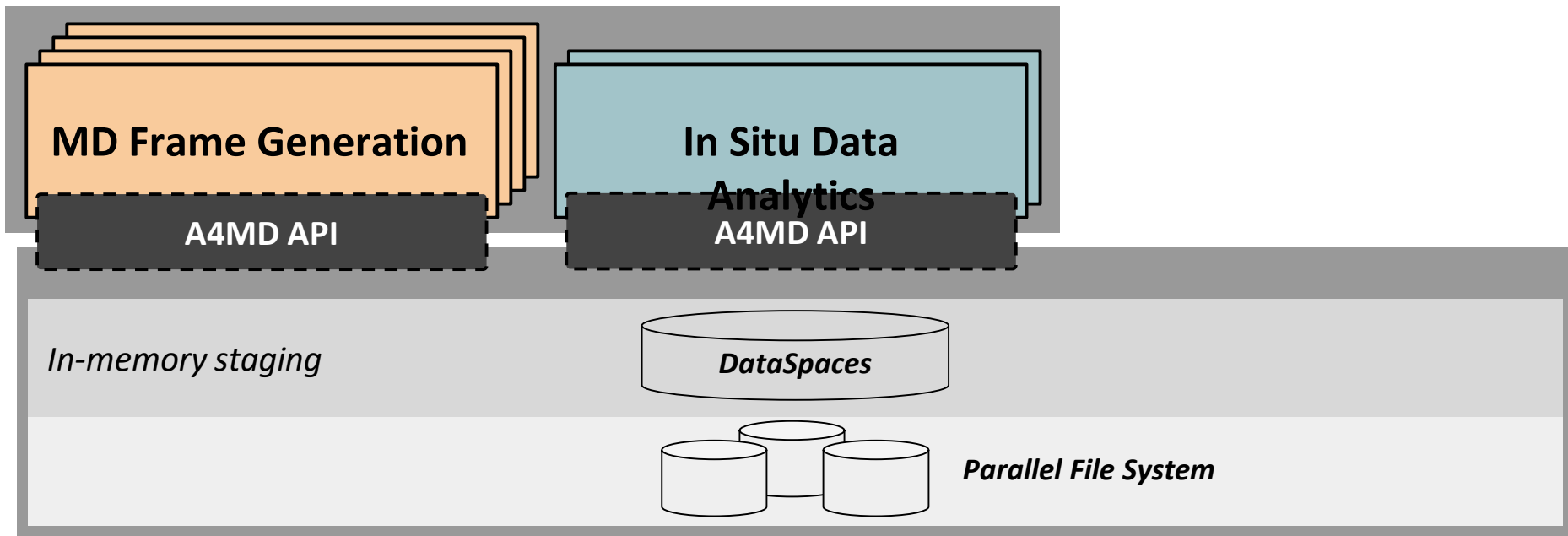
Unfavorable trajectory: does not explore the tICA space



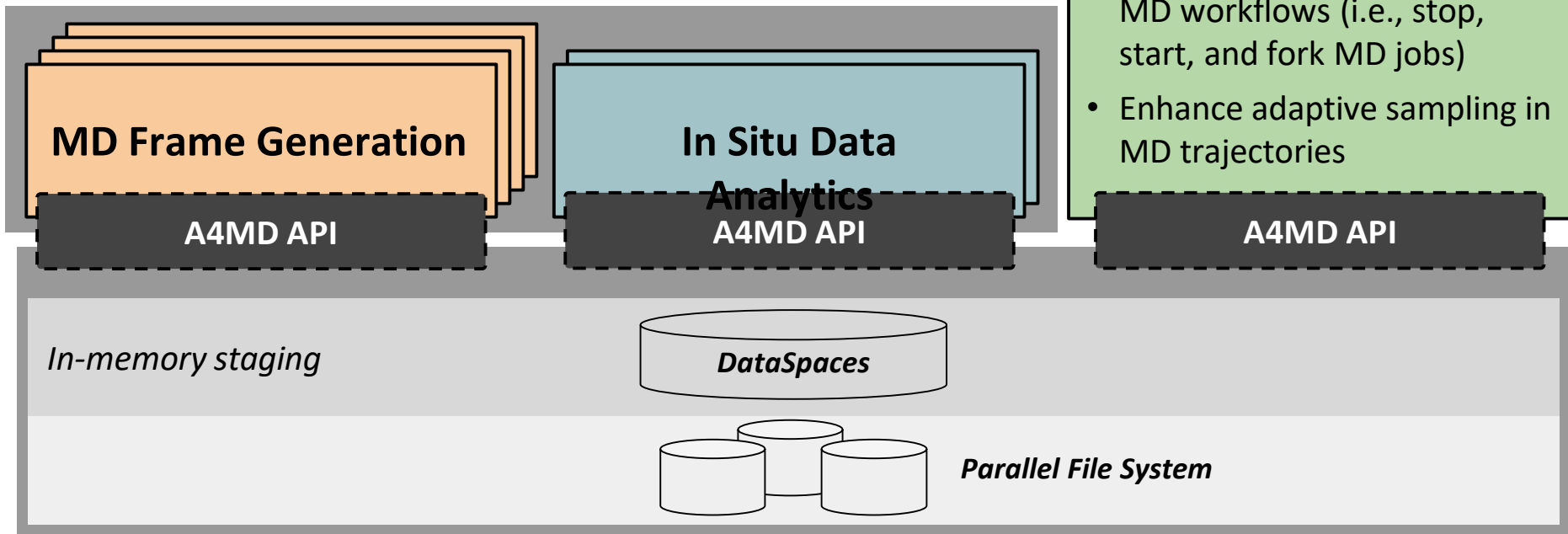
From One Trajectory to Many



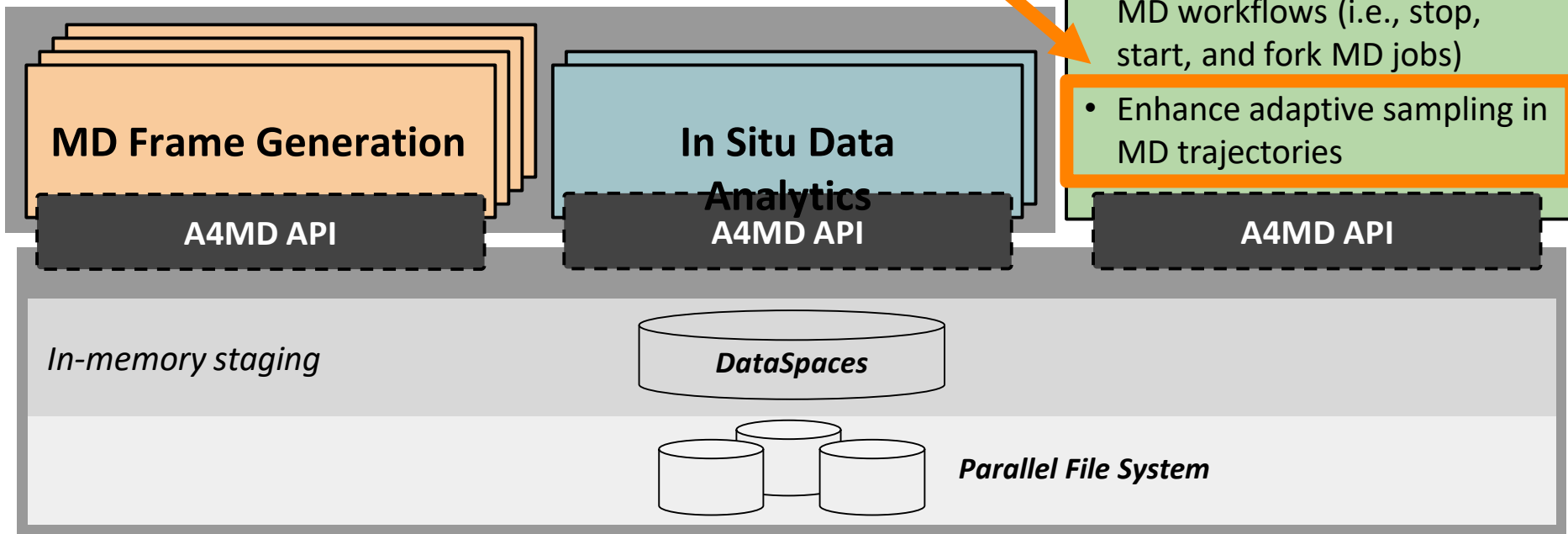
From One Trajectory to Many



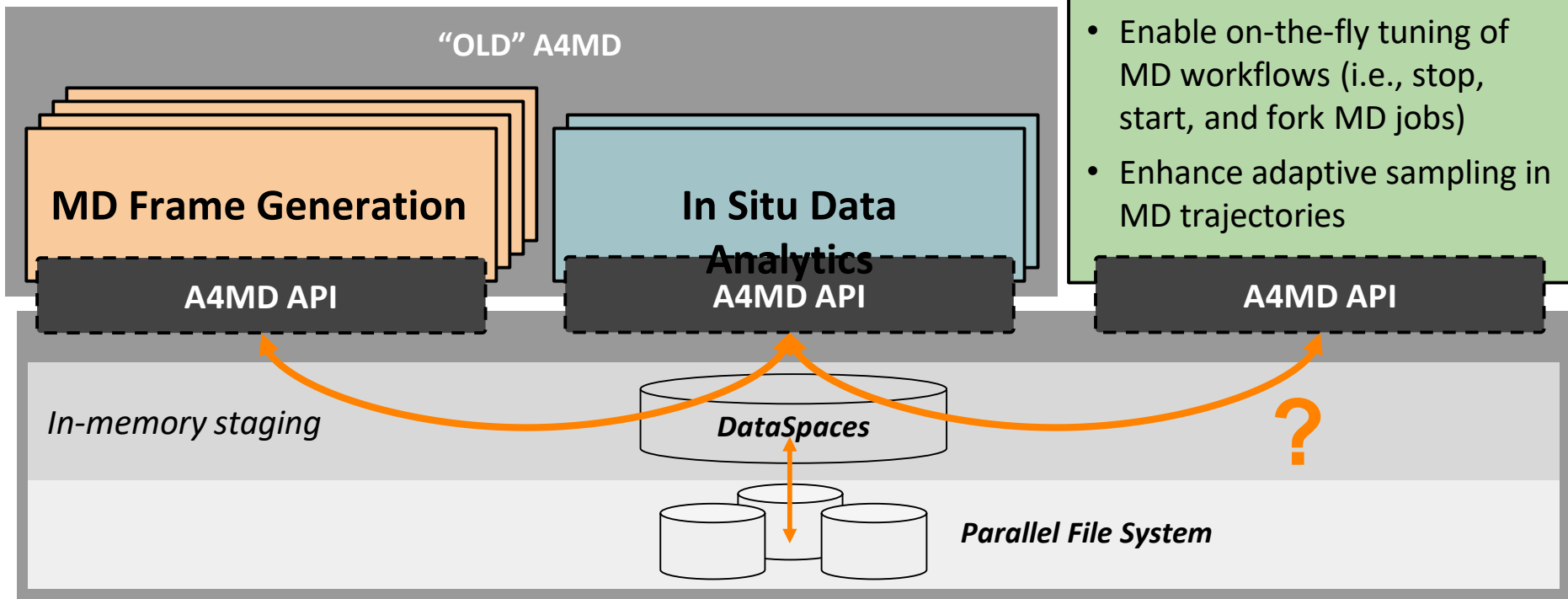
From One Trajectory to Many



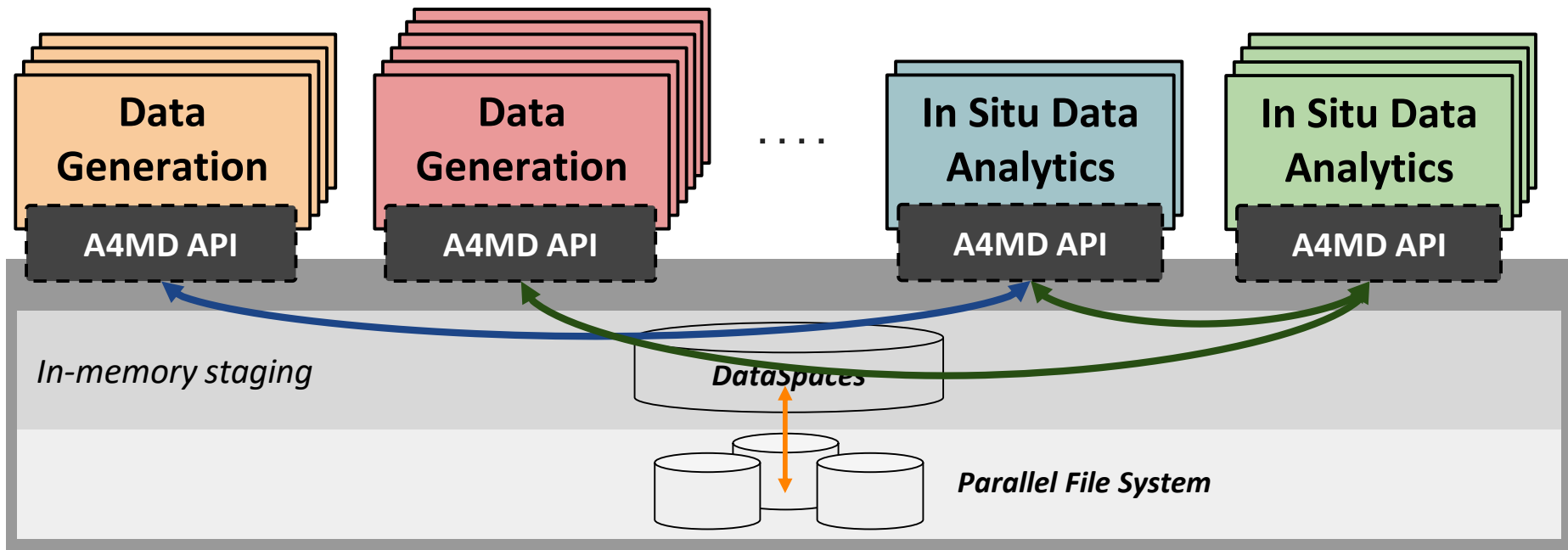
From One Trajectory to Many



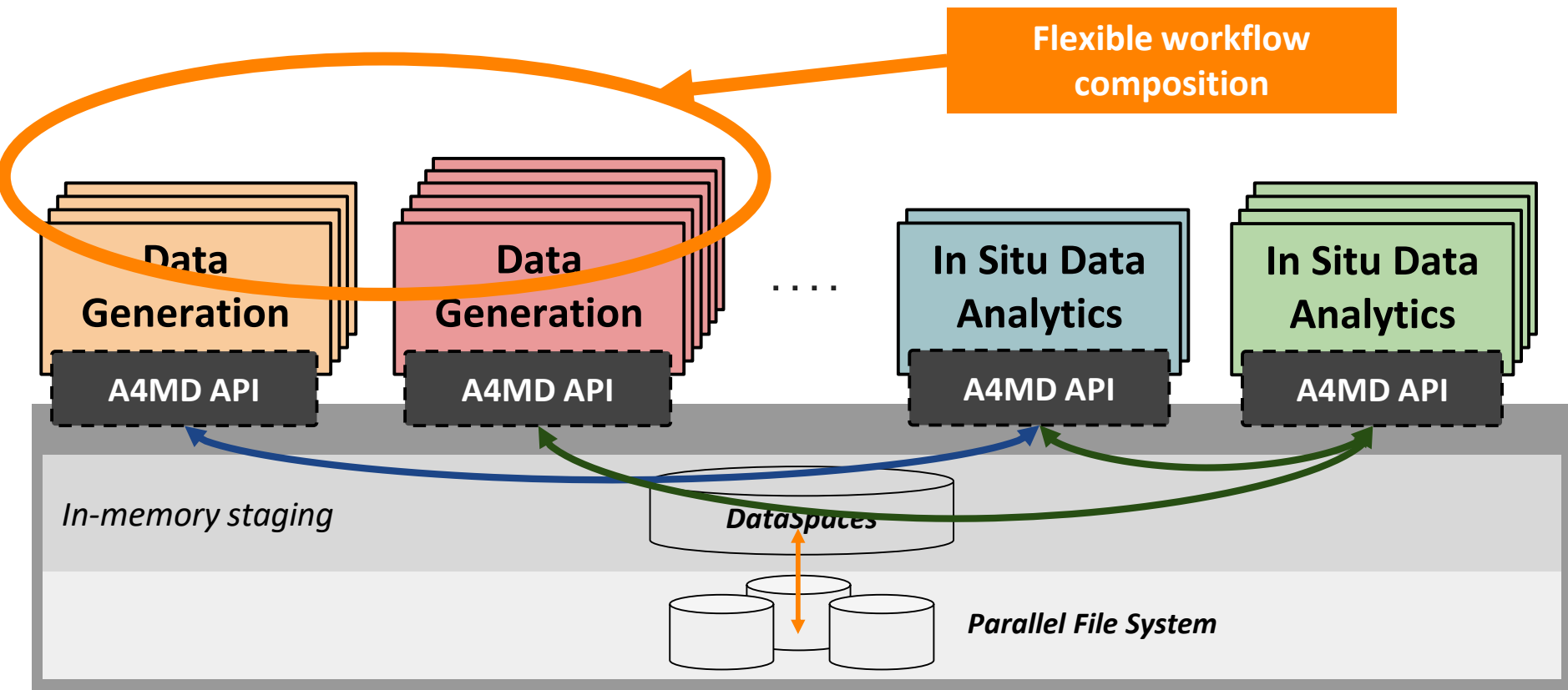
EAS in A4MD



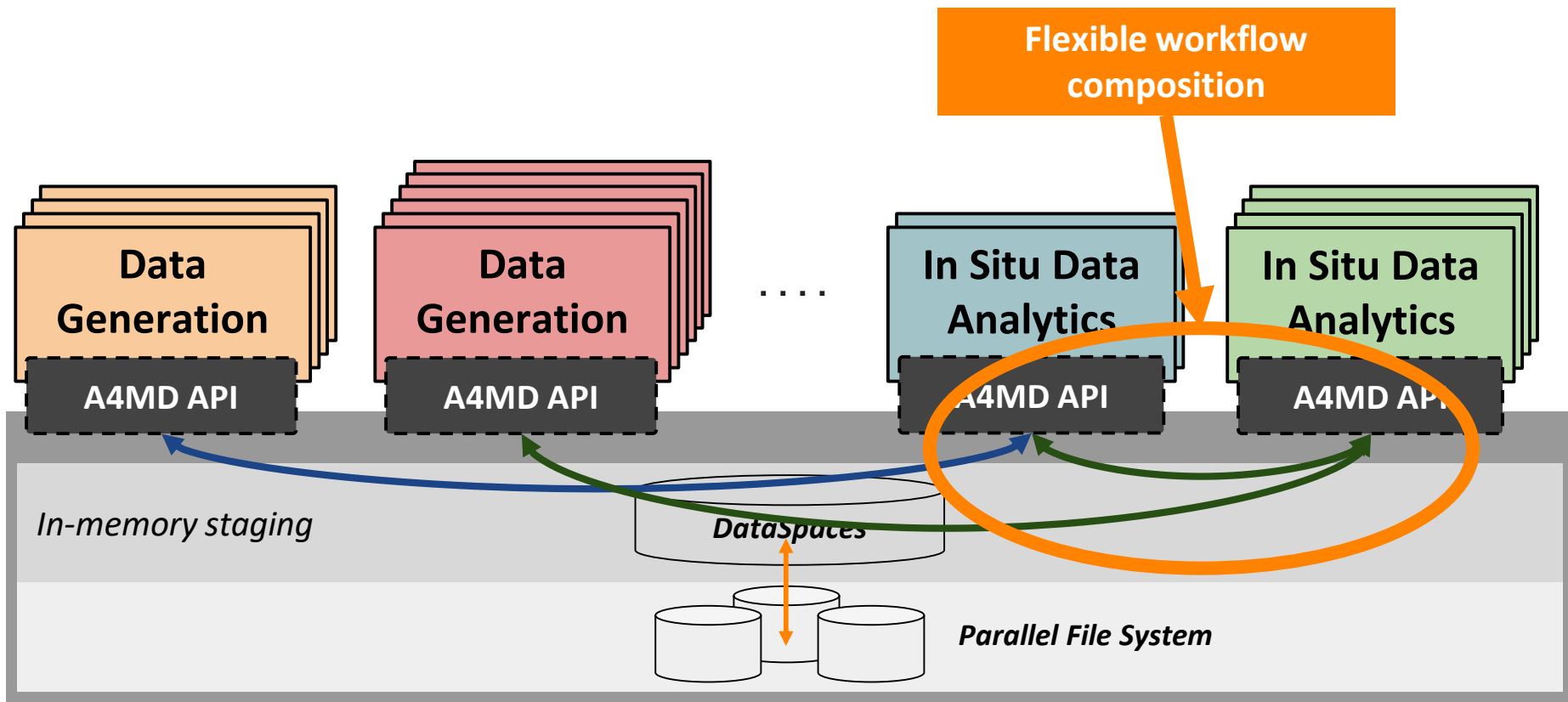
EAS in A4MD



EAS in A4MD



EAS in A4MD



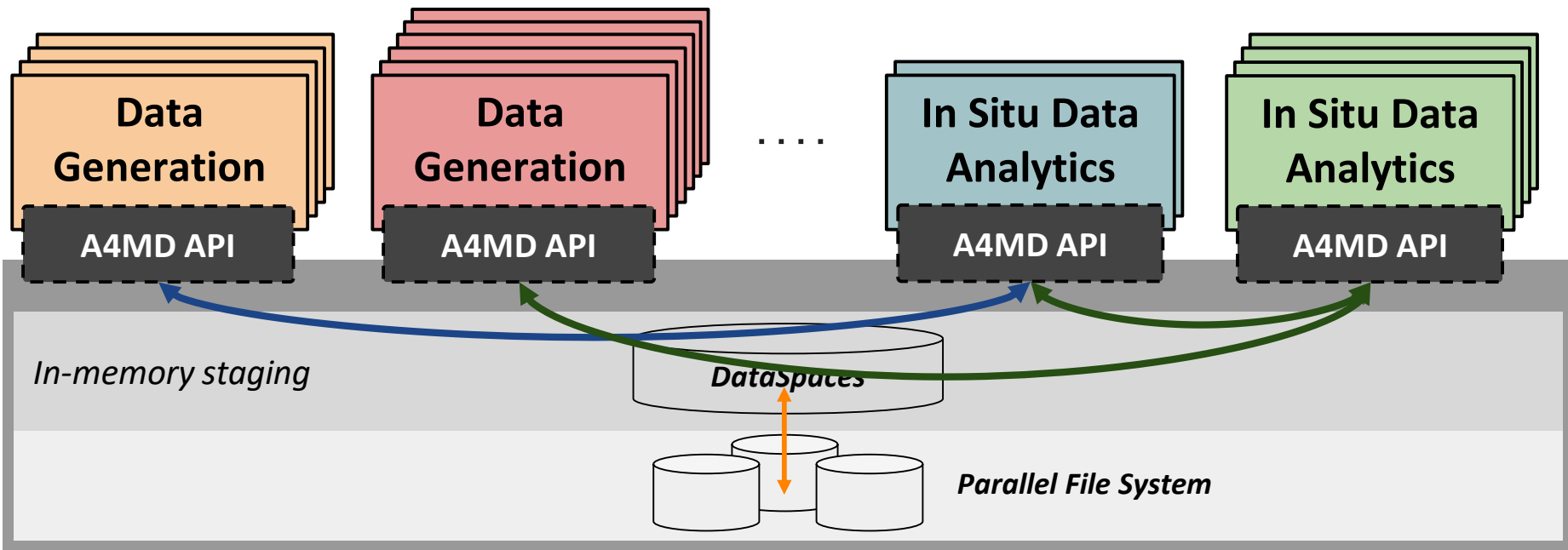
EAS in A4MD: Orchestration

Ensemble generator

From global knowledge, decide which trajectories will be terminated/spawned

Ensemble analyser

Gather all the local data to extract global knowledge about the ensemble



EAS in A4MD: Orchestration

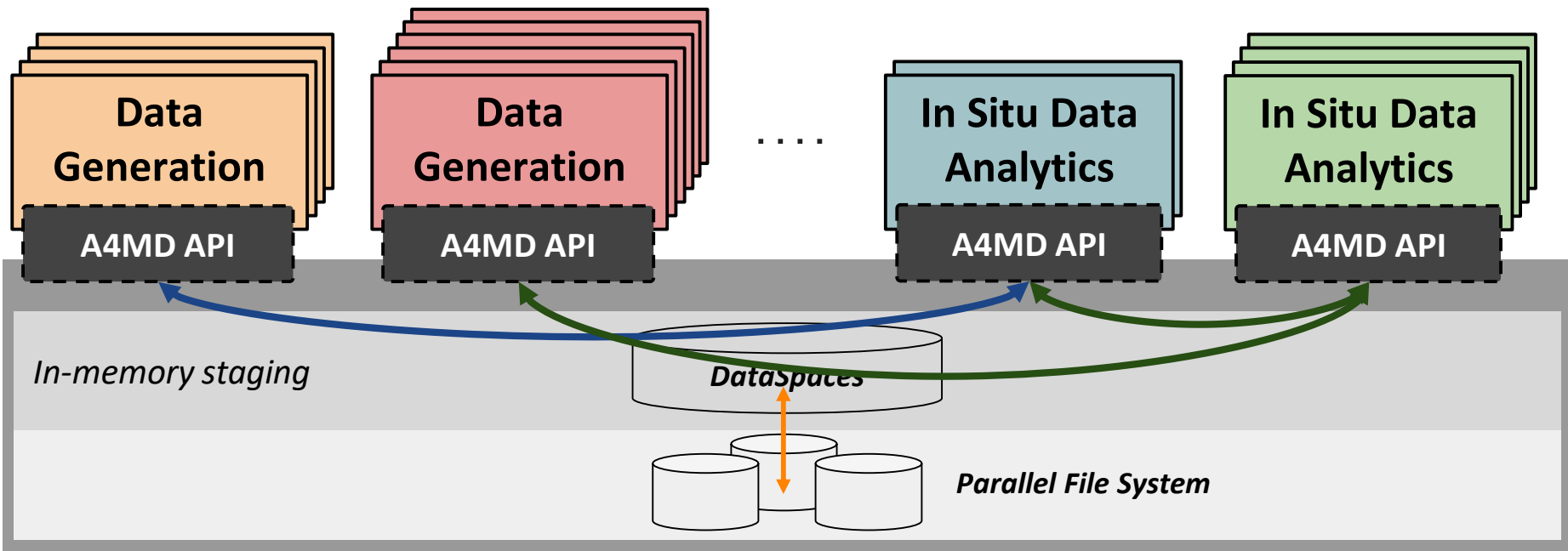
Ensemble generator

From global knowledge, decide which trajectories will be terminated/spawned

Ensemble analyser

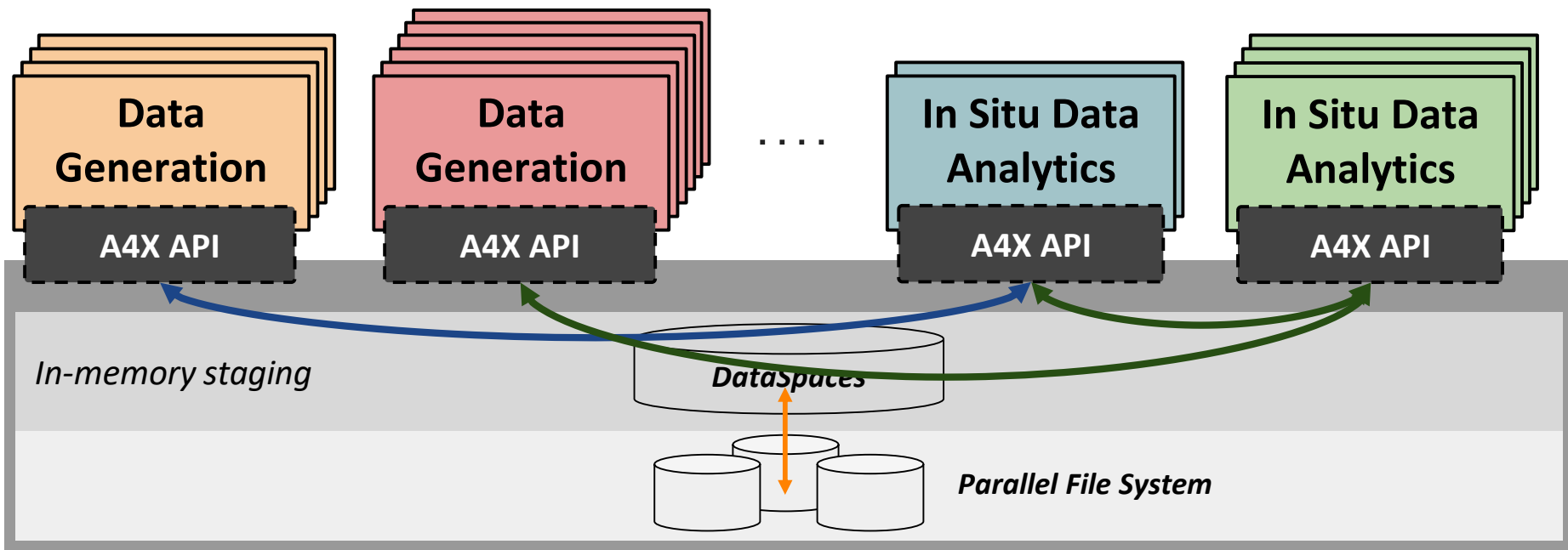
Gather all the local data to extract global knowledge about the ensemble

Ensemble feedback



Generalising the Workflow Engine: A4X

Multiple data sources and analysis modules can exchange information in situ through A4X



Goals for the Next Period

- Introduce features that allow the support of enhanced adaptive sampling in MD ensembles
- Orchestrate the ensemble allowing to stop, start and fork of MD jobs
- Evaluate scalability in a larger environment and with workflows of increased complexity
- Provide mechanisms to collect the intermediary and output data from the simulation and analysis
- Incorporate unsupervised learning techniques for extracting collective variables automatically

Analytics for Molecular Dynamics (A4MD)

Towards In Situ Enhanced Adaptive Sampling of Simulation Ensembles

Ian Lumsden, Michael Wyatt, Silvina Caíno-Lores, Michela Taufer

