



# Easy Molecular Dynamics on HPC with EXA4MIND

David Číž<sup>1</sup>, Nestor Pylat<sup>1</sup>, Anthony Teye-Adjei<sup>1</sup>, Viktoria Pauw<sup>2</sup>, Vojtěch Mlýnský<sup>3</sup>

<sup>1</sup>IT4Innovations, VSB - Technical University of Ostrava, Czech Republic

<sup>2</sup>Leibniz Supercomputing Centre, Munich, Germany

<sup>3</sup> Institute of Biophysics, Brno, Czech Republic

## Motivation

- Managing Terabytes of trajectory data creates significant barriers to sharing and reproducibility.
- Complex storage landscapes and restrictive network policies currently hinder cross-resource collaboration.
- Traditional random-access I/O patterns over WAN result in severe latency bottlenecks.

## The ADAMS4SIMS Solution

- Automates the extraction and mining of molecular dynamics simulations for systematic improvement.
- Performs HPC-accelerated transposition of raw ASCII data into optimized atom-major HDF5 formats.
- Orchestrates workflows via Apache Airflow with intelligent offloading to the LEXIS Platform and HEAppE.
- Accelerates interactive analysis using a Dask service with smart disk-backed LRU caching.

## Core system

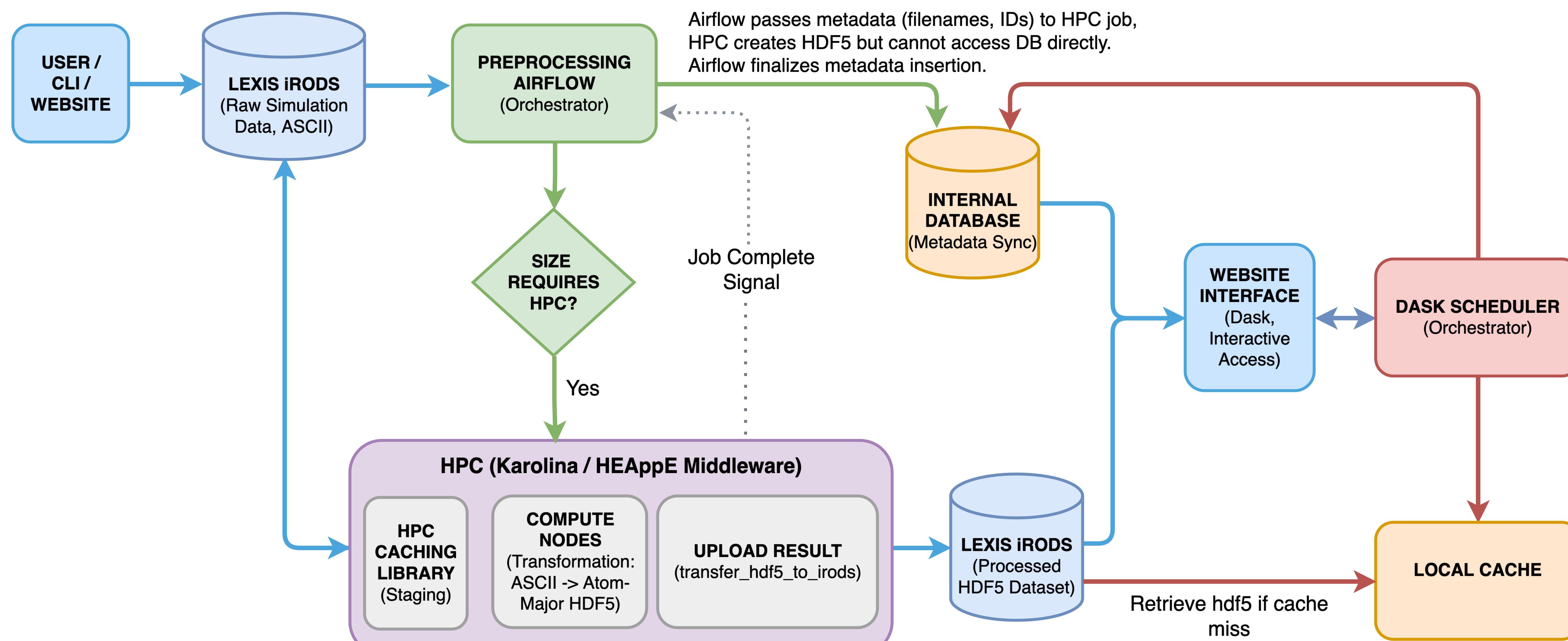


Figure 1: ADAMS4SIMS Processing Architecture and Workflow Orchestration.

## Results

Performance benchmarking across three datasets demonstrates orders-of-magnitude latency improvements for distance, angle, and dihedral calculations. These preliminary results show a significant improvement in time-to-first-result, changing it from minutes or hours to milliseconds.

Dataset		Technique		
Atoms	Frames	HDF5	PGSQL	CPPtraj
126	100,000	9.0 ms	1.7 s	4.5 s
17,915	100,000	11.4 ms	448.6 s	641.8 s
20,000	600,000	62.1 ms	993.0 s	N/A*

Table 1: Analysis Latency (HDF5 vs. Traditional Methods)

\*process timed out

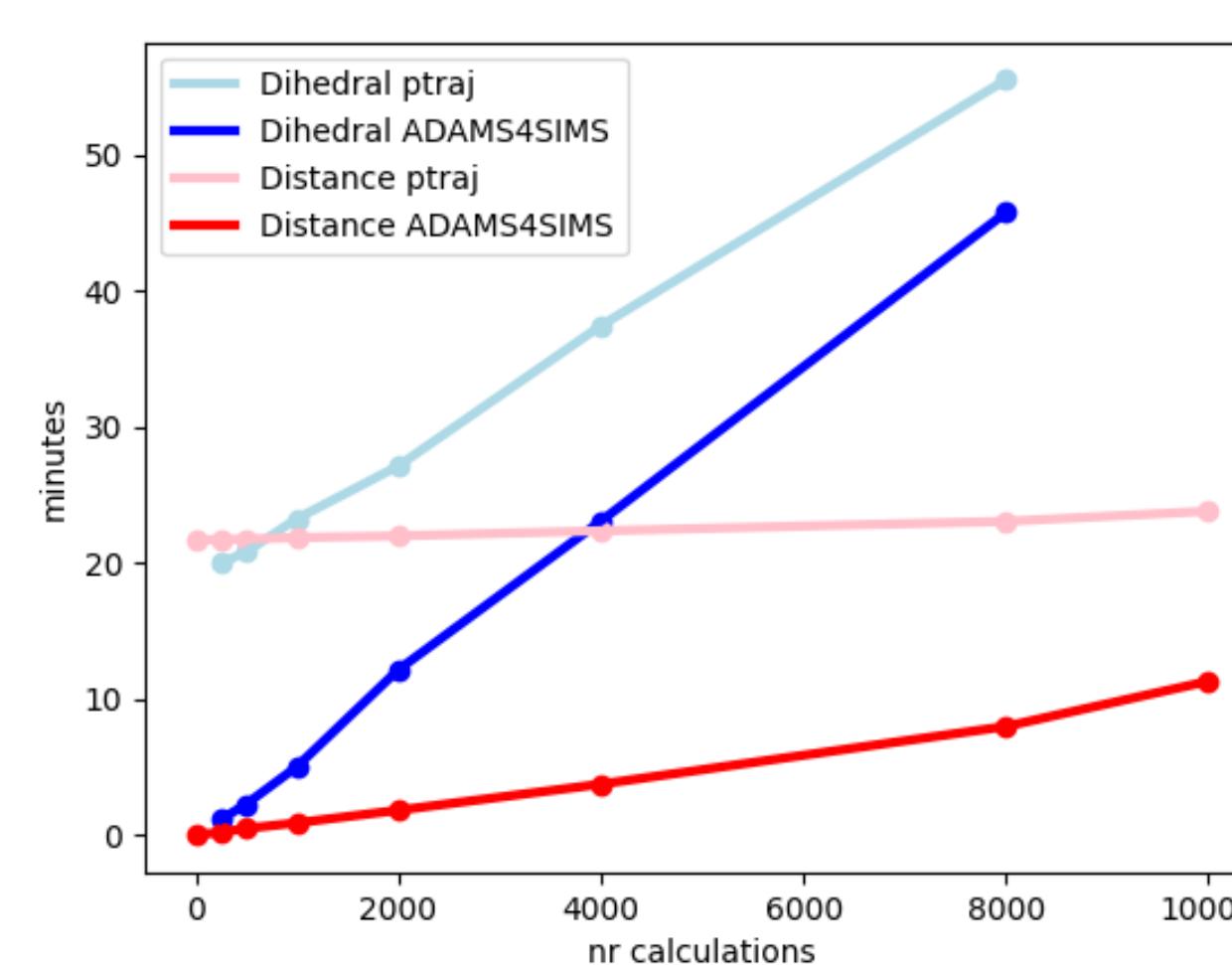


Figure 2: Scaling of HPC (HDF5 vs. Traditional Methods)

## Conclusion & Future Impact

ADAMS4SIMS successfully bridges the gap between HPC storage and interactive web analysis. By combining atom-major transposition with smart caching, we reduced analysis latency from minutes or hours to **milliseconds**.

### Future Directions:

- Democratizing Data:** Enabling public, interactive browsing of TB-scale datasets without downloads.
- AI Readiness:** Establishing a structured foundation for large-scale, data-driven force-field optimization.



<https://heappe.eu>



<https://exa4mind.eu>