MAX DRIVING THE EXASCALE TRANSITION

AiiDA

Description

AiiDA is a computational infrastructure that helps researchers with automating, managing, persisting, sharing and reproducing complex scientific workflows while keeping track of all associated data.

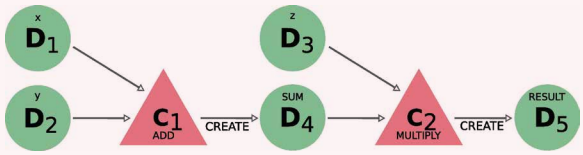
It is distributed as an open-source Python library that provides a framework in which these scientific workflows can be codified to automate both their execution (including adaptation to intermediate outputs and error handling) and the rigorous tracking of the generated data.

Key Features

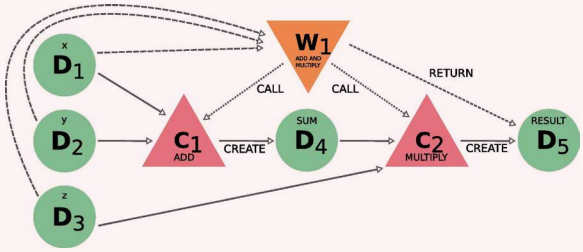
- **Workflows**: AiiDA allows to build and execute complex, auto-documenting workflows linked to multiple codes on local and remote computers.
- **High-throughput**: AiiDA's event-based workflow engine supports tens of thousands of processes per hour with full check-pointing.
- **Data provenance**: AiiDA automatically tracks and records inputs, outputs and metadata of all calculations and workflows in extensive provenance graphs that preserve the full history of all data and calculations.
- **Advanced queries**: AiiDA's query language enables fast graph queries on millions of nodes.
- **Plugin interface**: AiiDA can support via plugins any computational code and data analytics tool, data type, scheduler, connection mode, etc. (over 125 codes and 125 workflows supported in our plugin registry as of July 2022).
- **HPC interface**: AiiDA can seamlessly deal with heterogeneous and remote computing resources; it works with many schedulers out of the box (SLURM, PBS Pro, torque, SGE or LSF).
- **Open science**: AiiDA allows to export both full databases and selected subsets, to be shared with collaborators or made available and browsable online on the Archive and Explore sections of Materials Cloud.
- **Open source**: AiiDA is released under the MIT open-source license.

Provenance

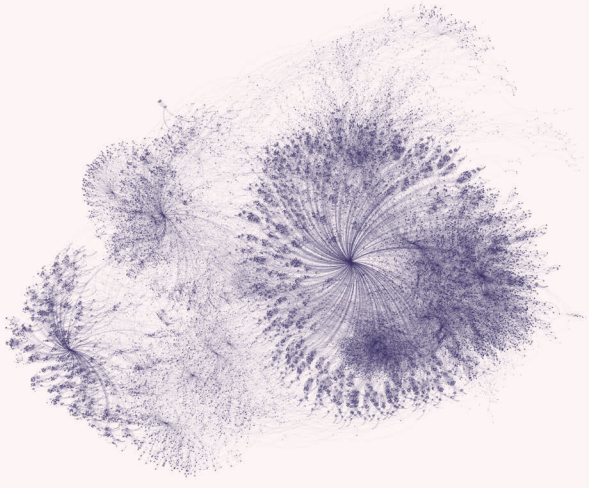
The concept of provenance is one of the key underlying pillars over which AiiDA is built. The code represents the process of conducting research as the act of transforming known pieces of information into new ones (data nodes) through discrete steps. These steps are called "calculations" and contain the recipe to produce the output data nodes with the results from the input data nodes.
A simple example of this can be displayed if we consider a basic procedure of taking three numbers (D1, D2, D3) and first adding two of them together (D4 = D1 + D2) to then multiply the sum by the third one (D5 = D4 x D3).



This node graph is automatically stored in an internal database when running the procedures wrapped around the AiiDA compatible classes. Moreover, AiiDA provides a framework to build logical workflows around these data tracking structures, making them easier to reproduce. They can also be tracked in the provenance graph.



Actual production databases are much more complex. Below there is a visual representation of one for a real research project, consisting of millions of nodes:

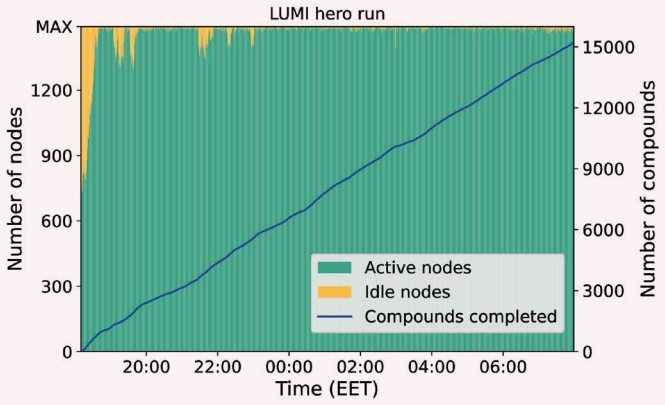


MAX DRIVING THE EXASCALE TRANSITION

AiiDA

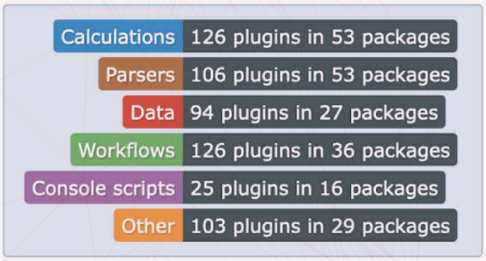
HPC performance

The performance of AiiDA's workflow engine has been demonstrated rigorously on LUMI-C, the CPU partition of one of the most advanced pre-exascale supercomputers in Europe. During a twelve hour time frame, over 55,000 Quantum ESPRESSO calculations were orchestrated by AiiDA, fully utilizing the 196608 AMD Epyc cores of the LUMI-C partition and optimizing the geometry of over 15,000 inorganic compounds at the PBEsol level.



Modularity

The plugin system allows AiiDA users not only to develop their own workflows, code wrappers, data storage types, etc., but also to easily share them in "plugin packages" so that others can then build on top of an existing ecosystem of AiiDA tools.



Available plugins on the AiiDA plugin registry as of July 2022

Queriability

AiiDA stores the most relevant data in formats that are easy to access (for example, as fields in a database). This enables the introduction of a powerful query language that can explore and filter the provenance in a highly efficient manner, using not only the nature of the nodes but also their content.

This query tool called "querybuilder" is accessible to users and can be applied in a variety of situations, such as post-processing of results. Additionally, the functionality can be extended in higher-level layers, such as tools for more advanced graph traversal.

Contact & Support

Website | Github | Plugins | AiiDA docs | AiiDA tutorials | AiiDAlab docs

References

- G. Pizzi et al. Comp. Mat. Sci. 111, 218-230 (2016) — DOI - 10.1016/j.commatsci.2015.09.013
- S.P. Huber et al., Scientific Data 7, 300 (2020) — DOI - 10.1038/s41597-020-00638-4
- M. Uhrin et al., Comp. Mat. Sci. 187 (2021) — DOI - 10.1016/j.commatsci.2020.110086
- A. V. Yakutovich et al., Comp. Mat. Sci. 188, 110165 (2021) — DOI - 10.1016/j.commatsci.2020.110165