

Argonne Leadership Computing Facility

an Office of Science user facility



ABOUT COMPUTING RESOURCES SCIENCE AT ALCF PROJECTS NEWS & EVENTS USER SUPPORT USER GUIDES AURORA

Science at ALCF

- ▼ Allocation Programs
 - ▼ Early Science Program
 - Data and Learning Call For Proposals
 - Data/Learning Proposal Instructions
 - INCITE Program
 - ALCC Program
 - Director's Discretionary (DD) Program
 - ALCF Data Science Program

[INCITE 2019](#)

[ALCC 2018-2019 Projects](#)

[ADSP projects](#)

[Aurora ESP projects](#)

[Theta ESP Projects](#)

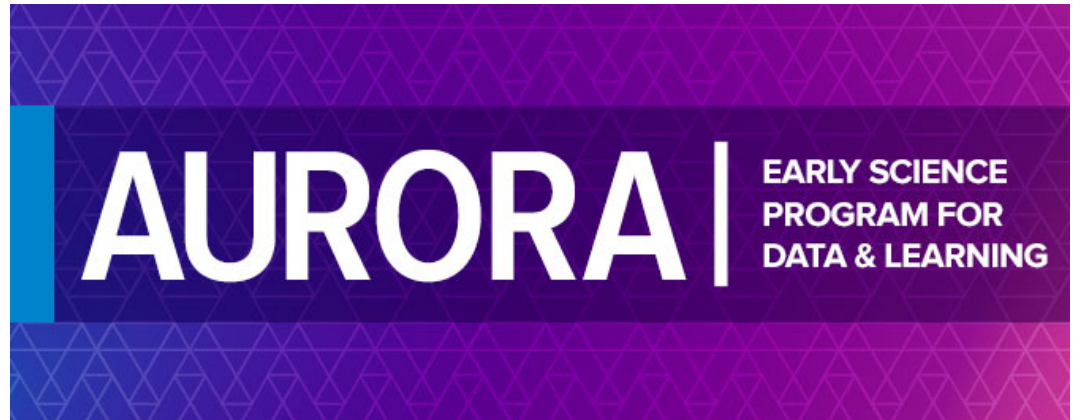
[Publications](#)

[ALCF Tech Reports](#)

[Industry Collaborations](#)

[All Projects](#)

ALCF Aurora 2021 Early Science Program: Data and Learning Call For Proposals



Aurora 2021 Data and Learning Call For Proposals Is Now Closed.

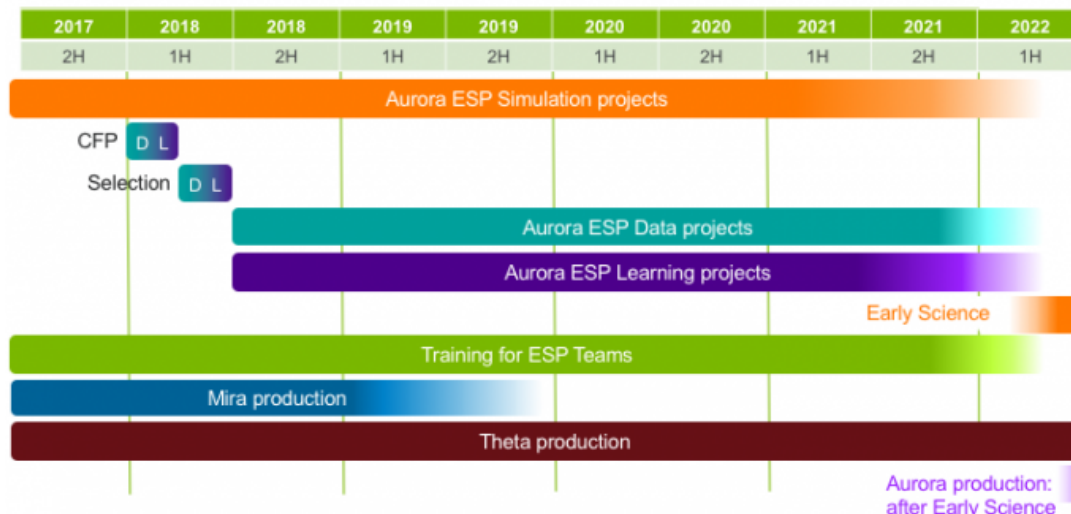
In 2021, the Argonne Leadership Computing Facility (ALCF) will deploy Aurora, a new Intel-Cray system. Aurora, will be capable of over 1 exaflops. It is expected to have over 50,000 nodes and over 5 petabytes of total memory, including high bandwidth memory. The Aurora architecture will enable scientific discoveries using simulation, data and learning. The detailed architecture of Aurora is protected by RSNDA (Restricted Secret Nondisclosure Agreement).

For the Aurora Early Science Program (ESP) Data and Learning call, we will select 10 new projects, all chosen competitively based on this call for proposals. With this CFP, we are specifically targeting applications in the areas of Data and Learning. These should have strong aspects of data science (Big Data, data-intensive computing, experimental/observational/simulation data analytics, etc.) and/or machine learning (deep learning, neural networks, discovery of patterns and reduced models for scientific data and/or simulation modeling, etc.). Cross-cutting proposals targeting the convergence of simulation, data and learning are very much encouraged.

With Aurora being a dramatically bigger and faster machine than Theta or Mira, the three months of pre-production Early Science time will be a large and valuable allocation of core-hours, with the potential for truly unprecedented computational science—as well as being the United States' first exascale system. ALCF will fully fund 10 postdoctoral appointees for Aurora ESP—one for each selected project.

Below is a rough timeline for the Aurora ESP. The rows labeled “A21 ESP projects” denote the central effort of the projects: developing, porting, and tuning code for the target system:

AURORA ESP TIMELINE (NOTIONAL)



Relationship with existing Aurora ESP projects

Before the plan to shift from a 2018-delivery (180 petaflops) Aurora based on third-generation Intel® Xeon Phi™ processors to the 2021-delivery (1 exaflops) Aurora, we had already selected 10 Aurora ESP projects. These will continue, and serve as the Simulation based projects targeting Aurora 2021 (A21). This is reflected in the topmost, red bar in the timeline figure. An important part of the shift to A21 is the shift from primarily traditional, simulation-based computing at ALCF to expanded scope including data-centric and machine/deep learning projects. We now refer to Simulation, Data, and Learning as the “three pillars” of leadership computing going forward. This call for proposals is for projects in the Data and Learning pillars.

Proposing to Unknown Hardware and Exascale

Details we can provide about the A21 system in this call are very few. The speed and scale of A21 will be vastly greater than today’s systems, or systems on the near-term horizon. We realize that this poses a substantial challenge to proposal authors, especially in the areas of Data and Learning, where there is limited history of applications running at leadership scale; and we will take this into account when evaluating proposals. We believe that the development and optimization for current large-scale computers such as ALCF’s Theta, NERSC’s Cori, OLCF’s Titan, or OLCF’s forthcoming Summit system, will form a solid basis to develop and optimize for Aurora 2021.

Once ESP projects have been selected, project teams will sign RSNDAs and learn about the Aurora architecture. They will have access to a sequence of simulators, compilers, precursor hardware testbeds, and some of the earliest-available processors in the form of white boxes. This development ecosystem, together with training and assistance from ALCF and the system vendors, should allow the teams to achieve a high level of readiness by the time the Aurora hardware arrives in 2021.

Some Guidance About Aurora for Proposal Authors

- Nodes will have both high single thread core performance and the ability to get exceptional performance when there is concurrency of modest scale in the code.
- The architecture is optimized to support codes with sections of fine grain concurrency (~100 lines of code in a FOR loop for example) separated by serial section of code. The degree of fine grain concurrency (number of iterations of loop for example) that will be needed to fully exploit the performance opportunities is moderate. In the ~1000 range for most applications.
- Independence of these loops is ideal but not required for correctness although dependencies that restrict the number of things that can be done in parallel will likely impact performance.
- There is no limit on the number of such loops and the overhead of starting and ending loops is very low.
- Serial code (within an MPI rank) will execute very efficiently and the ratio of the performance of the serial to parallel capabilities is a moderate ratio of around 10X, allowing for code that has not been entirely reworked to still perform well.
- OpenMP 5 will likely contain the constructs necessary to guide the compiler to get optimal performance.
- The compute performance of the nodes will rise in a manner similar to the memory bandwidth so the ratio of memory BW to compute performance will not be significantly different than systems were a few

years ago. A bit better in fact than they have been recently.

- The memory capacity will not grow as fast as the compute performance so getting more performance through concurrency from the same capacity will be a key strategy to exploit the future architectures. While this capacity is not growing fast compared to current machines it will have the characteristic that the memory will all be high performance alleviating some of the concerns of managing multiple levels of memory and data movement explicitly.
- The memory in a node will be coherent and all compute will be first class citizens and will have equal access to all resources, memory and fabric etc.
- The fabric BW will be increasing similar to the compute performance for local communication patterns although global communication bandwidth will likely not increase as fast as compute performance.

Scope of Proposal Topics

As guidance on subject areas for proposals to this call, consider these suggestions. This is not meant to be a comprehensive list, just a source of ideas and orientation:

DATA	LEARNING
<ul style="list-style-type: none">• Experimental/observational data<ul style="list-style-type: none">◦ Image analysis◦ Multidimensional structure discovery• Complex and interactive workflows• On-demand HPC• Persistent data techniques<ul style="list-style-type: none">◦ Object store◦ Databases• Streaming/real-time data• Uncertainty quantification• Statistical methods• Graph analytics	<ul style="list-style-type: none">• Deep learning• Machine learning steering simulations<ul style="list-style-type: none">◦ Parameter scans◦ Materials design◦ Observational signatures• Data-driven models and refinement for science using ML/DL• Hyperparameter optimization• Pattern recognition• Reduced model derivations• Bridging gaps in theory

Cross-Cutting Proposals

Cross-cutting proposals targeting the convergence of simulation, data and learning are very much encouraged.

Support for Aurora ESP Projects

Postdocs

ALCF will directly support efforts on the ESP projects by hiring postdocs to work with some of the project teams. We will fund up to 10 dedicated postdocs, who will be employed by ALCF, but work directly with project investigators on efforts needed to prepare for Early Science runs. Some of these may split time between the ALCF and project-investigator locations. Generally, a postdoc will work on only one Aurora ESP project.

ALCF Staff

We will assign one ALCF staff scientist to each ESP project. This person will spend a fraction of his/her time collaborating with the ESP project and mentoring the postdoc on computational aspects of the project.

Collaboration with Intel & Cray

Another type of support for the ESP projects will be access to applications experts from the Aurora vendors—Intel and Cray. These expert consultants will be made available through various avenues to assist with ESP code porting and tuning.

Training and Hardware Access

ALCF and Intel/Cray will offer training targeted toward the ESP projects. This will include a virtual kick-off workshop about the Aurora hardware and programming the system, and a hands-on workshop as soon as we have sufficient Aurora-generation hardware to support testing and debugging of project applications. Before then, we will also offer access to advanced Aurora simulators and hardware as described under “Proposing to Unknown Hardware and Exascale” above, and allocations on our production system Theta for

ESP development work that does not depend strongly on having the new hardware (e.g., new algorithms, new physics modules, basic introduction of threads). Where appropriate, we will offer joint training opportunities with OLCF, in support of as much portability as possible among our future systems and OLCF's IBM/NVIDIA-based Summit system.

Aurora's Early Science Period

The Early Science period on Aurora is a span of about three months, beginning after system acceptance, but before turnover to production users. During this time, projects selected for the Aurora ESP will have dedicated access to the full system, with large allocations of time in support of their major scientific run campaigns. Access will continue for the remainder of a year, but will be shared with the production users.

Aurora ESP Proposal Process

Proposals for the Aurora ESP must have a plan for the science to be accomplished, and a description of what application development will occur throughout the duration of the project. In addition, each selected project's home institution must pursue an appropriate Non-Disclosure Agreement (NDA) with Intel/Cray for access to needed information on the next-generation architecture. ALCF will provide instructions for obtaining the NDAs to the selected project teams.

The [Proposal Instructions](#) should include everything needed to submit a proposal, including a document template. These are, roughly, a simplified version of an INCITE proposal. Please direct any questions to earlyscience@alcf.anl.gov. As part of a DOE user facility collaboration on application readiness, the proposal form will ask for disclosure of participation in OLCF's CAAR program (the equivalent of ESP at the other LCF center), and the Exascale Computing project.

Evaluation of Proposals

ALCF, with the help of internal and external science-domain experts, will evaluate proposals on the strength of

- Potential impact of proposed Early Science run campaign
- Proposed runs that are far beyond what can be done with today's machines such as Theta, *and appropriate for exascale*
- Computational readiness
- Data scale readiness: Description of the data requirements and plans to realize the science at these scales.
- Scaling to hundreds of thousands of cores (50% of Theta, for example). We realize that today's data-centric and machine-learning/deep-learning based applications may be less advanced in scaling than simulation applications, and will take that account in evaluation.
- An existing or reasonably well-planned implementation to make use of thread concurrency and SIMD-like parallelism on Aurora. We realize that today's data-centric and machine-learning/deep-learning based applications may be less advanced in these metrics compared with simulation applications, and that your plan here may hinge on optimized libraries/frameworks. Clear elucidation of status and expectations is a plus here.
- Appropriateness of development team: are the expertise and person-hours proposed likely to succeed—yielding a code ready to run the proposed science on Aurora starting in the Early Science period
- Diversity of science domains and algorithms; we want the ESP projects to be a reasonable sample set of the science and algorithms in production at the LCF site

Aurora ESP Data & Learning Proposal Timeline (2018)

- **January 10, 2018** | Call for Proposals issued
- **April 8, 2018** | Call for Proposals closes (*now closed*)
- **End of June 2018** | Aurora ESP Awards Announced
- **July 2018** | Kick-Off Workshop for ESP Teams (virtual/webinar)

Outcomes and Expectations of ESP Projects

We expect ESP projects to share best practices and lessons learned—the fruits of your labors developing, porting, and optimizing your applications for Aurora—in technical reports and in presentations to the community. ALCF will organize a community workshop or webinar series after the end of the Early Science period. We also expect projects to share and publish their scientific results in appropriate venues, acknowledging the ALCF and its Early Science Program.

We expect project participants to partner with ALCF and Intel/Cray to assess robustness and correctness of the new hardware and software—to help identify the root cause of problems and find potential workarounds.

Our intent is for Aurora ESP proposals to be relatively simple and short—a stripped-down version of an INCITE proposal. The sections of the proposal are

1. PI and co-PI information
2. Project Summary
 - Executive Summary
 - Benefit to Community
 - Science Summary
 - Application Summary
3. Estimate of Resources Required
4. Participation in Other Applications-Readiness Programs
5. Project Team Members, Research Funding

Submission

- Submission deadline: April 8, 2018 before midnight in any time zone (Anywhere On Earth)
- We are using the EasyChair system for proposal submission. You'll need to create an account if you don't have one already, and login to the [Aurora ESP EasyChair website](#).
- Prepare your proposal using the instructions below
- Submit as a single PDF document, by using EasyChair to upload. You may resubmit with revisions as needed up until the deadline.

Please direct any questions to earlyscience@alcf.anl.gov. If needed, contact Tim Williams at 630-252-1154.

Login

Follow us

[Facebook](#)
[YouTube](#)
[Flickr](#)
[RSS](#)

About ALCF

[Overview](#)
[JLSE](#)
[History](#)
[Staff Directory](#)
▶ [Careers](#)
[Visiting Us](#)
[Contact Us](#)

For Users

[Acknowledgement Policy](#)
[New User Guide](#)
[Mira/Theta/Cetus/Vesta](#)
[Cooley](#)
[Machine Status](#)
[Facility Policies](#)

Subscribe to our monthly newsletter

Argonne Leadership Computing Facility

Argonne National Laboratory
9700 South Cass Avenue
Building 240
Argonne, IL 60439