SIRIUS software for ECP

# The goals of your project and its current status

Our long-term vision is to create a sustainable and extensible . Our goal is to reduce the time that solution scientists spend in moving, storing, and processing data.

SIRIUS will support

* List the items

SIRIUS is being created by scientist at ORNL, SNL, Rutgers, UCSC, and Brown University. We also have outside participation from scientist at GT, and Stony Brook.

# Do you release your software as open source?

We will release services and methods for SIRIUS under the completely free BSD license. This license is even more permissive than LGPL. The software will be openly available on GitHub:

# Do you have DOE/NNSA users of your software?

SIRIUS

# Have facilities, vendors, or ISVs picked up your software?

# What is the support model for your software?

# Are there any applications in particular that the outcomes of your project are targeting?

# The specific ties to identified requirements of the applications, other software components?

# Your plan for ensuring that the developed software technologies be mature enough to be part of the software stack on exascale systems expected to be selected in 2019 and installed in 2023?

# What do you feel are the key challenges posed and opportunities offered by exascale systems for your specific area?

# What is the R&D that you would like to carry out within the ECP?

There are plenty of *research* topics where we still need to find out what is the best way to deal with the big data challenges on extreme scale systems:

* *Data Refactoring:* We must reduce the amount of data while retain most of the *information* written to permanent storage. Our approach is to use mathematical and statistical techniques for this and we need to work closely with applications to find the appropriate methods for their data, analysis requirements and data lifecycle management. Also, our idea is to optimize read access to most frequently used data by refactoring data for multiple levels of storage.
* *Data Lifecycle Management:* When we say data, we mean not just the dataset output of an application but also the knowledge gained from it, and the codes and processing workflows and visualizations applied to it to gain that knowledge. We want to *morph* data as it moves to different systems and transforms itself from data to knowledge. We will apply *learning techniques* to pre-fetch, pre-calculate data from semantically described intentions. Also, better data reduction techniques will be needed because of the size of the data.

# What research remains for your project’s outcomes to benefit key DOE applications?

# How would the proposed activities build on the research you have been carrying out with ASCR Research funding?

# What are the proposed activities that you believe would contribute to the ECP?

# Your roadmap/timeline for maturing the software technologies and deploying them on exascale platforms, with a few intermediate milestones or decision points (forks in the roadmap). The timeline is of particular importance in selecting what the ECP will include in the development plans.

**Year 2017**

**Year 2018**

**Year 2019**

**Roadmap beyond 2019**

# Highlight your activities that would help DOE exascale apps achieve ECP performance, efficiency and resilience performance goals on 2023 hardware and system architectures