**• Please fill out the following to the best of your ability.
• If you have multiple projects and/or codes represented by this case study, note this in the text and then fill out the table at the end using aggregate numbers where appropriate (e.g., total hours used) and maximum values elsewhere (e.g., number of compute cores used per job).
• If needed, include a different table for each major code.
• For reference, review the graph of historical usage at
http://www.nersc.gov/science/requirements-workshops/case-study-faq/**

**• In a question is not applicable to your project, please enter "N/A."**

# Case Study Title: (!! enter a title here !!)

**Principal Investigator:**

**Worksheet Author(s) (if not PI):**

**NERSC Repositories**:

# Project Description

## Overview and Context

**Please give a brief, high-level description of your research and the role that High Performance Computing (HPC) and storage play. (1-3 paragraphs)**

## Objectives for 2017

**What are your project’s goals for 2017? Do not limit your answer to the computational aspect of the project. (1-3 paragraphs)**

# Computational Strategies (now and in 2017)

## Approach

**Give a short, high-level description of your computational problem and your strategies for solving it.**

## Codes and Algorithms

**Please briefly describe the codes you use and algorithms that characterize them (1-2 sentences per).**

# HPC Resources Used Today

## Computational Hours

**How many hours on conventional cores (not GPUs) will your project(s) use at NERSC in 2013?
If you have significant allocations and usage at other sites please describe it here.**

## Parallelism

**How many (conventional) compute cores are typically used for production runs at NERSC today using the codes you described above? (You can give a range.)**

**What is the maximum number of cores that these codes could use today?**

**If the typical number is less than the maximum, briefly explain why fewer than the maximum might be used.**

**Which is more important for your software or project, strong scaling or weak scaling? Why? (Strong: you have a problem of a given size and you'd like to use parallel computing to solve it faster. Weak: you have a problem of a given size and you'd like to use parallel computing to solve a bigger problem in the same time.)**

## Scratch Data

**How much temporary disk space (space that can be purged) do you need to perform your runs?**

## Shared Data

**NERSC provides “project directories,” which are permanent, global, shared storage areas for collaboration. Does your project have a NERSC project directory? If so, what is its name? What is the primary reason you have this space?**

## Archival Data Storage

**How much data do you have stored on the NERSC HPSS data archive in 2013?**

# HPC Requirements in 2017

## Computational Hours Needed

**How many conventional compute hours will your project require in CY 2017? (“Conventional” refers to a machine like Edison, Hopper, Carver or PDSF.)** **Include all hours your project will need to reach the scientific goals you listed in 2.2 above.**

**If you expect to receive significant allocations from sources other than NERSC, please list them here.**

**If you expect to need more compute hours in 2017 than you used at NERSC in 2013, what is the primary factor driving the need for more hours?**

## Parallelism

**How many conventional compute cores will your code(s) typically use in 2017?**

**What do you expect is the maximum that could be used in 2017?**

## I/O

**Does your application have built-in checkpoint/restart?**

**How much data will you need to read and write per run in 2017 (including checkpoint/restart data)?**

**Please estimate your I/O bandwidth requirement (bandwidth = data read or written / time to read or write).**

**What percentage of your total runtime are you willing to devote to I/O?**

## Future Data Needs (Please fill in the blanks)

**In 2017, we expect to need \_\_X\_\_ TB of temporary scratch disk space, \_\_X\_\_ TB of NERSC project space (globally accessible shared data), and \_\_X\_\_ TB of storage on NERSC HPSS. The growth in these requirements relative to 2013 is due primarily to \_\_\_\_\_\_\_\_\_\_\_\_\_\_.**

## Memory Required

**For NERSC to plan for future systems, we need to know your memory requirements. How much memory will your codes require per node? How much aggregate memory will be required?**

## Emerging Technologies and Programming Models

**This is a key part of this review. Please discuss the status of efforts to transition your software to emerging architectures. Please answer the questions below and provide any additional information that will help us understand what needs to done to successfully transition Office of Science codes to run efficiently on next-generation architectures.**

**Does your software have CUDA/OpenCL directives? If so, are they used, and if not, are there plans for this? If not, why not?**

**Does your software run in production now on Titan or other systems using GPUs?**

**Does your software have OpenMP directives now? If so, are they used, and if not, are there plans for this? If not, why not?**

**Does your software run in production now on Mira or Sequoia using threading?**

**Is porting to, and optimizing for, the Intel MIC architecture underway or planned?**

**Have there been, are there now, or are there plans for other funded groups or researchers engaged to help with these activities?**

**If you answered "no" for the questions above, please explain your strategy for exploiting these technologies.**

**What role should NERSC play in the transition to these architectures?**

**What role should DOE, NP, and/or ASCR play in the transition to these architectures?**

**Do you have other comments or concerns about this?**

## Software Applications and Tools

**What HPC software (applications / libraries / tools / compilers / languages / etc) will you need from NERSC in 2017? Be sure to include analytics applications and I/O software**.

## HPC Services

**What NERSC services will you require in 2017? Possibilities include consulting or account support, data analytics and visualization, training, support servers, collaboration tools, web interfaces, federated authentication services, gateways, etc**.

## Additional Data Intensive Needs

**Will you have additional needs we have not considered regarding data? These could be related to workflow, management, transfer, analysis, sharing or access, or visualization**.

**Do you already have a data management plan for your project and does it include archival storage?**

## Additional Data Intensive Needs: Burst Buffer

**NERSC is considering including a "burst buffer" (using NVRAM technology) in future systems to improve I/O and/or visualization by allowing data to be staged (before writing), pre-staged (before reading), or otherwise manipulated on the "fly." Please look at the usage scenarios for possible burst buffer use on http://www.nersc.gov/assets/Trinity--NERSC-8-RFP/Documents/trinity-NERSC8-use-case-v1.2a.pdf and comment on which, if any, of these would be most useful for your work.**

## What Else?

**Are there any other services or facilities you would like NERSC to provide?**

**Do you have present or future concerns you’d like to discuss?**

# Requirements Summary Worksheet

Please try to fill out this worksheet, based on your answers above, to be best of your ability prior to the review.

|  |  |  |
| --- | --- | --- |
|  | **Used at NERSC in 2013** | **Needed at NERSC in 2017** |
| Computational Hours |  |  |
| Typical number of cores\* used for production runs |   |  |
| Maximum number of cores\* that can be used for production runs |  |  |
| Data read and written per run  | TB | TB |
| Maximum I/O bandwidth  | GB/sec | GB/sec |
| Percent of runtime for I/O  |  |  |
| Scratch File System space | TB | TB |
| Shared filesystem space  | TB | TB |
| Archival data | TB | TB |
| Memory per node | GB | GB |
| Aggregate memory | TB | TB |

\* “Conventional cores.” For GPUs and accelerators, please fill out section 4.7.

# Additional Storage and I/O Questions

These questions are optional but your answers will provide additional useful data for NERSC. If you don't know the answer to any of these leave them blank.

**For Scratch data (like Question 5.4)**:

• **Is your I/O more serial or parallel?**

• **Is your I/O more single-node or multiple-node?**

• **Is your I/O more shared (N-to-1) or distributed (N-to-N)?**

•  **Is your I/O more small-file or large-file?**