

Title: Request for Compute Infrastructure to Support the Strategies and Techniques for Analyzing Microbial Population Structures (STAMPS) at the Marine Biological Laboratory in Woods Hole, MA, USA, 2019

Abstract:

Currently, our society as a whole is generating data much faster than it is developing the skills to work with it, as it takes time for our educational infrastructure to catch up with what is needed. This leaves many biologists struggling to develop bioinformatics skills that are rapidly becoming essential, with little or no guidance or expertise available to them. The STAMPS course has been a yearly event at the Marine Biological Laboratory (MBL) in Woods Hole, MA, USA for over a decade, helping nearly 1,000 learners establish a foundation in bioinformatics over the years. In the past, the course has been typically run using the MBL computational infrastructure, but this has been problematic for several reasons. A few of the faculty involved with STAMPS have been and are also involved with other ~2-week long workshops that have utilized Jetstream to great success (C. Titus Brown, and the submitting PI for this request, Michael D. Lee). So we are submitting this education request to hopefully support this year's STAMPS course, taking place over 22-July to 1-August-2019 (11 days). The course will be hosting about 60 attendees and 21 instructors. All materials developed and presented at the course are openly available.

Syllabus:

<https://mblstamps.github.io/>

Materials from last year:

<https://github.com/mblstamps/stamps2018/wiki#schedule>

In-progress schedule for this year:

<https://github.com/mblstamps/stamps2019/wiki#schedule-in-progress>

Keywords: Microbial ecology, Bioinformatics, Training

Principal investigator: Michael D. Lee

Field of science: Microbial ecology and evolution

Required Documents: PI CV (2 pages limit)

Resource Justification:

We are requesting 120 m.medium instances with 6 cores, 16 GB RAM, and 130 GB VM space each for each instructor and student for 2 weeks, plus some “hot spares” and backup machines for when start-up doesn't work for new tutorials. The total request is for 259,200 service units (6 cores * 24 hrs/day * 15 days * 120 instances). To accommodate large size data files, an additional 100 GB of storage volumes are requested for each instance. Persistent storage beyond the duration of the course (22-July to 1-August-2019) is not required.

We are budgeting for the full time period because students sometimes don't shut down their machines, but we will teach them how to shut down and start up new machines as part of the course.

These calculations are based on running a similarly sized course for eight previous years with approximately 360 students total on AWS cloud services and Jetstream.

Resources: IU/TACC (Jetstream)