Critical Techniques, Technologies, and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)

A 2018 Solicitation

Webinar: March 20, 2018

Chaitanya Baru and Sylvia Spengler

Office of the Assistant Director, Information and Intelligent Systems

Directorate for Computer & Information Science & Engineering

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Purpose of this Webinar

- Orient potential proposers
- Summarize the BIGDATA Program
- Answer questions
- Help improve the quality of proposals

Webinar Outline

- BIGDATA Program description
- Overview of solicitation (NSF 18-539)
- Questions from the community

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

NSF BIGDATA

- Program description
- Participating NSF organizations
- Directorate specific priorities
- Cloud option

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

BIGDATA Program

- Seeks novel approaches in computer science, statistics, computational science, and mathematics leading towards the further development of the interdisciplinary field of *data science.*
- Also seeks innovative applications in domain science, including social and behavioral sciences, education, physical sciences, and engineering, where data science and the availability of big data are creating new opportunities for research and insights not previously possible.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

BIGDATA Program

- Invites two categories of proposals:
  - **Foundations (BIGDATA: F):** those developing or studying fundamental theories, techniques, methodologies, and technologies of broad applicability to big data problems, motivated by specific data challenges and requirements; and
  - **Innovative Applications (BIGDATA: IA):** those engaged in *translational* activities that employ new big data techniques, methodologies, and technologies to address and solve problems in specific application domains. Projects in this category must be collaborative, involving researchers from domain disciplines and one or more methodological disciplines, e.g., computer science, statistics, mathematics, simulation and modeling, etc.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

BIGDATA Program

- Proposals are expected to be well motivated by specific big data problems in one or more science and engineering research domains.
- All proposals are expected to clearly articulate the big data aspect(s) that motivate the research.
- *Innovative Applications* proposals must provide clear examples of the impacts of the big data techniques, technologies and methodologies on applications in one or more domains.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

National and NSF-wide Priorities

The BIGDATA solicitation is responsive to national and NSF-wide objectives, as well as priorities in science areas across the Foundation. This solicitation welcomes proposals that are responsive to the objectives of:

Harnessing the Data Revolution (HDR) aimed at fundamental data science research, research data cyberinfrastructure, and the development of a 21st century data-capable workforce.

More information about HDR can be found at the following web pages:

https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf

https://www.nsf.gov/news/news_summ.jsp?cntn_id=244678&org=CISE&from=news

Participating NSF Organizations

Directorate for Computer & Information Science & Engineering (CISE)

- Office of the Assistant Director
  - Chaitanya Baru (Senior Advisor for Data Science)
- Division of Information and Intelligent Systems
  - Sylvia Spengler (lead Program Director)
  - James French
  - Maria Zemankova
  - Aidong Zhang
- Division of Computing and Communication Foundations
  - Almadena Chtchelkanova
  - Rahul T. Shah
- Division of Computer and Network Systems
  - David Corman
  - Ralph Wachter

Directorate for Education & Human Resources (EHR)

- Division of Research on Learning in Formal and Informal Settings
  - John C. Cherniavsky

Directorate for Engineering (ENG)

- Division of Chemical, Bioengineering, Environmental, and Transport Systems (CBET)
  - Christina Payne
- Division of Civil, Mechanical and Manufacturing Innovation (CMMI)
  - Alexis Lewis
- Division of Electrical, Communications and Cyber Systems (ECCS)
  - Anthony Kuh

Directorate for Mathematical & Physical Sciences (MPS)

- Office of the Assistant Director
  - Bogdan Mihaila
- Division of Mathematical Sciences
  - Nandini Kannan

Directorate for Social, Behavioral & Economic Sciences (SBE)

- Division of Social and Economic Sciences
  - Sara Kiesler

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Directorate Specific Priorities - EHR

The **Directorate for Education and Human Resources (EHR)** is interested in fostering novel, transformative, multidisciplinary approaches that address the use of large data sets and/or learning analytics software to create actionable knowledge for improving STEM teaching and learning environments (formal and informal) in the medium term, and to revolutionize learning in the longer term.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Directorate Specific Priorities - ENG

The **Directorate of Engineering (ENG)** The Directorate for Engineering (ENG) is interested in topics specified by the programs belonging to the Divisions of Civil, Mechanical, and Manufacturing Innovation (CMMI); Chemical, Bioengineering, Environmental, and Transport Systems (CBET); and Electrical, Communications, and Cyber Systems (ECCS). Engineering-specific applications, however, must incorporate aspects of data science, data analytics, or innovation in data-driven discovery at a level significantly beyond the scope of existing ENG programs.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Directorate Specific Priorities - MPS

The **Directorate for Mathematical and Physical Sciences (MPS) Division of Mathematical Sciences (DMS)** is interested in foundational mathematical and statistical approaches to big data challenges generated by complex dependence structures, missing information, sparsity, and heterogeneous data. Proposals must describe the specific big data challenges being addressed and justify why the proposed research cannot be considered in one of the "core" DMS programs.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Directorate Specific Priorities - SBE

The **Directorate for Social, Behavioral, and Economic Sciences (SBE)** is interested in research that advances computational social science and analytic methods using social network, sensor, text, video, administrative, and other big data. Example topics include algorithms for social policy choices; cognitive assistance; promoting American competitiveness and economic opportunity, health, and well-being in different regions and populations; new forms of work; financial markets and critical institutions of society; and the intersection of law, governance, and data science.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option

- Continues the cloud option, introduced in FY 2017, in partnership with:
  - Amazon Web Services (AWS);
  - Google Cloud; and
  - Microsoft Azure.

  The objective is to encourage projects that focus on large-scale experimentation and scalability studies. If additional cloud providers join the program, resources/credits from those providers will be available under the same terms and conditions as described in the BIGDATA solicitation.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Related Programs

- Before preparing a proposal in response to this BIGDATA solicitation, applicants are strongly urged to review other related programs and solicitations and contact the respective NSF program officers to identify whether those solicitations are more appropriate.
  - Advances in Biological Informatics, managed by the BIO Division of Biological Infrastructure (DBI)
  - EarthCube program for *Developing a Community-Driven Data and Knowledge Environment for the Geosciences*
  - *Cyberinfrastructure for Sustained Scientific Innovation - Data and Software* program (CSSI)
  - *Computational and Data Enabled Science and Engineering* program (CDS&E)
  - Directorate for Mathematical & Physical Sciences (MPS) Division of Mathematical Sciences (DMS)
  - Core programs in CISE:
    - Computer and Network Systems (CNS);
    - Computing and Communication Foundations (CCF); and
    - Information and Intelligent Systems (IIS).

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

The NSF 18-539 Solicitation

- Schedule
- 2018 Updates
- Additional Documents
- Review criteria

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Schedule

- All proposals to NSF 18-539 have the same deadline window: **May 7, 2018 – May 14, 2018, due by 5:00 PM submitter's local time.**

- **Schedule:**

Solicitation Issued:	February 2018
Proposals Due:	May 7-14, 2018
Review:	June-July, 2018
Announcement of Awards:	Fall, 2018

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Awards

- Projects will typically receive NSF funding in the range of $200,000 to a maximum of $500,000 per year, for 3 to 4 years of support.
- The minimum award size will be $600,000 of total NSF funding, reflecting the minimum expected level of effort for BIGDATA projects.
- The maximum award size will be $2,000,000, or $500,000 per year over four years, of total NSF funding.
- Any allocation of cloud credits/resources from Amazon, Google or Microsoft will be in addition to the NSF funding.

Anticipated Number of Awards: 25 to 33

Anticipated Amount of Funding: $24,000,000

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option

BIGDATA proposals may be submitted with or without any request for cloud resources. You should request cloud resources/credits, only if your project requires such resources.

The section, Use of Cloud Resources, in the solicitation specifies how to request cloud resources.

The request for cloud resources will be reviewed along with the rest of the proposal. Credits will be allocated either for the cloud provider requested in the proposal, or for equivalent resources from an alternative cloud provider, if necessary.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option Request Instructions

How to request cloud credits/resources:

- Within the project summary's list of keywords, include the term "Cloud Resource" as the first keyword (i.e., Keywords: Cloud Resource,…).
- As a supplemental document, submit a Cloud Costing and Annual Usage Plan. The two essential items appearing in this **two-page document** are 1) the total cost of cloud resources and 2) an annual usage plan.
- **Separately, you must fill out all the information requested in the Excel workbook that is linked to from the solicitation, and available at this location:**

  https://www.nsf.gov/cise/iis/cloudrequestform.xlsx.
- **After filling completing this workbook, you must email the file to bigdatacloud@nsf.gov, by Monday, May 21st, 2018.**
- **If cloud resources are requested to provide longer-term storage, this should also be detailed in the Data Management Plan.**

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option Workbook

The completed workbook should include the FastLane proposal ID (which is assigned after the proposal is submitted to NSF).

NSF Directorate for Computer & Information Science & Engineering
BIGDATA Cloud Request
(Version Date: January 2018)

Instructions:
Before starting, read all instructions. Please note: This template is locked to allow automated capture of submitted information. You can select any cell but will only be able to enter or change the content of white cells.
1. Complete this template, listing all cloud resources needed.
2. In order to retain original format, direct entry (typing) into cells is best.
3. Click on any column heading for additional help.
4. When complete, save this file with a filename in the format: [(lead) proposal ID#]_[(lead) PI's last name]. E.g., 1800000_Smith.
5. Send a copy of the properly named file as an attachment to bigdatacloud@nsf.gov, with the message subject line the same as the filename. Send only one completed file per project using only the lead proposal ID number (i.e., one file for all members of a multi-institutional collaborative tied to the lead proposal ID number) after submission of the proposal per the solicitation instructions. You will not receive a confirmation email; we suggest you cc yourself as proof of emailing.

Type, do not paste, the 7-digit FastLane Proposal ID number into the adjacent white box. (If a multi-institutional collaborative, enter only the lead proposal's ID number) **This number is provided after you have submitted the proposal.** Do not use the temporary proposal number provided prior to submission.

Question?
Email: bigdatacloud@nsf.gov

REMINDER: Prior to submitting the completed file,
please ensure the **filename and email subject line** both follow the format:
[(lead) proposal ID#]_[(lead) PI's last name]
e.g., 1800000_Smith.

Send completed file to:
bigdatacloud@nsf.gov

Resource Description	Unit Rate	Year 1		Year 2		Year 3		Year 4 (optional)		Total Cost ($)
		Total Units	Total Cost ($)	Total Units	Total Cost ($)	Total Units	Total Cost ($)	Total Units	Total Cost ($)	
Totals			$0.00		$0.00		$0.00		$0.00	$0.00

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option Reminders

The **minimum** cloud request is $100,000. Requests for less than $100,000 in cloud credits/resources will not be considered.

The **maximum** cloud request is based on a 70-30 split. For example, a project that requests, $700,000 on the NSF Budget page may request a maximum of $300,000 of cloud credits/resources. Or a project that requests $1,000,000 from NSF can request a maximum of $430,000 of cloud credits/resources.

Proposers are strongly encouraged to visit resource web pages and contact cloud providers to obtain assistance with estimating cloud resources needed for the project.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Cloud Option Contacts

The cloud providers and points of contact are listed below.

- Amazon Web Services

  Contact: Sanjay Padhi, sanpadhi@amazon.com

- Google Cloud

  Contact: Karan Bhatia, karanbhatia@google.com

- Microsoft Azure

  Contact: Vani Mandava, vanim@microsoft.com

Supplementary Documents

Proposals should include all standard supplementary documents. We highlight certain supplementary documents here:

- **Project Coordination Plan** [required, page limit: 2]
- **Cloud Costing and Annual Usage Plan** [required of proposals requesting the cloud option, page limit: 2]
- **Data Management Plan, including Software Sharing** [required of all proposals, including those requesting the cloud option, from the lead institution, page limit: 2]

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

NSF Review Criteria

In their reviews, panel discussions, and panel summaries, reviewers and review panels will address:

- Intellectual Merit,
- Broader Impacts, and

Standard NSF Review Criteria

- BIGDATA Specific Review Criteria

BIGDATA-Specific Review Criteria

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

BIGDATA-Specific Review Criteria (1)

- The extent to which the BIGDATA proposal pursues a big data research effort, as opposed to a traditional data research effort.
- For BIGDATA Foundations proposals, the extent to which the research is motivated by application use cases and/or characteristics of real big data datasets and addresses the innovation with respect to computer science, mathematical, statistical, and/or computational theories proposed.
- For BIGDATA Innovative Applications proposals, the extent to which genuine partnership between researchers in the technical community (i.e., computer scientists, statisticians, mathematicians) and researchers in science and engineering domains (e.g., science, engineering, social sciences, finance, education) is demonstrated, along with corresponding novel contributions.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

BIGDATA-Specific Review Criteria (2)

- In addition to the Broader Impacts detailed in the PAPPG, the extent to which additional broader impacts that are particularly relevant to BIGDATA activities are demonstrated (e.g., for software sharing and dissemination).
- The extent to which the proposal includes clear plans for *evaluation* and *validation* that demonstrate the effectiveness and efficacy of proposed solutions, in the context of specific big data characteristics.
- For proposals that request cloud resources, the extent to which the use of these resources and the levels of the requests are justified. As a reminder, it is expected that cloud resources will be requested and used for large-scale experimentation and scalability studies.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Questions?

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

Questions and Answers (1)

- **How do I submit a proposal to this program?**
  Please carefully read and follow the instructions provided in the solicitation itself (https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm) and the NSF *Proposal & Award Policies & Procedures Guide (PAPPG)* available at (https://www.nsf.gov/pubs/policydocs/pappg18_1/index.jsp). If you need additional help preparing and submitting your proposal, we recommend that you contact your institution's Sponsored Projects Office.

- **Do I need to use Grants.gov or Fastlane to apply?**
  You may use either Grants.gov or Fastlane.

- **Do I need to include keywords in my Project Summary?**
  All project summaries should include a list of three to six keywords at the bottom of the Broader Impacts text box, beginning with the phrase "Keywords:". Proposals using cloud resources should include the term "Cloud Resource" as the first keyword (i.e., **Keywords: Cloud Resource,…**).

Questions and Answers (2)

- **Who may submit BIGDATA proposals?**
  The BIGDATA accepts proposals from all categories of proposers listed in the NSF 18-1 Proposal & Award Policies & Procedures Guide (PAPPG), Chapter I, Part E.

- **BIGDATA incorporates a partnership with cloud providers. Is it essential to request cloud resources/credits in every project?**
  No. BIGDATA proposals may be submitted without any request for cloud resources. You should request cloud resources/credits, only if your project requires such resources. In general, projects may obtain access to the necessary compute, storage, and network resources by utilizing existing facilities, or purchasing new equipment, or requesting cloud credits.

https://www.nsf.gov/pubs/2018/nsf18539/nsf18539.htm

On behalf of the National Science Foundation and the BIGDATA team

THANK YOU!

Questions?

- Chaitanya Baru: cbaru@nsf.gov, or 703-292-4541
- Sylvia Spengler: sspengle@nsf.gov, or 703-292-8930

These slides, an audio recording, and a script of this webinar are available at http://www.nsf.gov/events/ and the BIGDATA program website.