

## **BigDataX: From theory to practice in Big Data computing at eXtreme scales**

<http://datasys.cs.iit.edu/grants/BigDataX/index.html>

### **Program Highlights**

The BigDataX program will host 10 students spending 10 weeks each summer doing research in one of two interdisciplinary laboratories: 1) the DataSys Lab in CS at IIT (5 students) or 2) the Systems Group in the CI/CS at UChicago (5 students). The two labs are physically close (4 miles apart), and students will have weekly activities that place all students in a single lab to strengthen the cohort experience. The PIs will aim to recruit most of the students from outside of the host institutions, focusing on recruiting students from institutions without research opportunities as well as women and minorities.

Each 10-week program will begin in the 3rd or 4th week of May aiming to end the program by the end of July or early August. Students will be encouraged to work in teams of two to help reinforce the positive aspects of teamwork, and each team will be assigned both a graduate student and faculty mentor. The BigDataX program will emphasize collaborative research teams in which graduate students are formally involved in the mentoring activities of the REU students. Each REU student will be involved in conducting research as part of a larger group, presenting research results to the entire BigDataX group, and as well as other audiences such as the DataSys lab and Globus Labs. The students will learn to both read research papers, and to write up their research work and results in the format of a conference/journal article. The significant collaboration between the mentors over the past eight years coupled with advanced graduate students working with the mentors at both institutions, will enable REU students to become productive faster on more complex research problems than is typically possible in such a short timeframe. The mentors will work with the students during the summer, and continue working on refining the research output towards publishing the students' summer work. The mentors will also help students prepare for presentations, and help them with the application process to graduate schools. Students will also be encouraged to continue their work at their respective institutions as part of an undergraduate thesis.

### **Orientation**

Orientation will take place during the first week and will introduce students to the research environments at IIT and UChicago, including labs, libraries, and both campuses. Further, each faculty will present an overview of their particular research projects and how they fit into big data computing. The mentors will also discuss how the proposed research projects differ from regular class or lab work. As part of this presentation, each mentor will present their approach to mentoring and expectations of students involved in research. Students will be asked to list their ranked top three projects; this process is meant to simulate the early stages of PhD programs where students choose a research topic and associated advisor.

Since most of the REU students will not have a background and research experience in big data, several hands-on tutorials provided by mentors and graduate students will be offered during the orientation week. There will be a general discussion with the students concerning presentations and peer review, both critical skills in research. Students will be given pointers on what constitutes a good presentation and then given suggestions on how to critique others' work constructively. Students will also be introduced to the top conferences and journals in the field related to their projects. Lunches will be organized in the first week to promote cohort building and to create closer ties between mentors and students.

Between the end of orientation and the beginning of student research, the research mentors will discuss the allocation of students to projects, based on student preferences, student backgrounds, and mentor

perceptions of the students gained during the application process and orientation week. We are proud to say that in the prior three summers, we managed to place 24 of the 25 students with their top preferences.

### **Mentoring Activities**

The PIs plan to enhance undergraduate education by introducing research early into the undergraduate career of students. Through research, young and diverse students will learn important skills early in their undergraduate studies: teamwork, written and oral skills, computational thinking, interdisciplinary skills, and experimental skills. Furthermore, participation in research groups can promote retention by increasing personal attachment to the research group, research objectives and research advisor, and captivate more students to pursue graduate degrees. Research at the undergraduate level is arguably one of the most important skills to attain in order to increase one's chances of being accepted in a top computer science graduate program. The students will be exposed to research and development in big data computing.

They will also be exposed to other research environments in the Chicago area, through research meetings and presentations at Argonne National Laboratory (ANL), University of Chicago (UChicago), and Illinois Institute of Technology (IIT). At ANL, they will be exposed to a large government laboratory, and be able to see one of the largest supercomputers in the world (the IBM BlueGene/Q Mira and Cray XC40 Theta). At the Computational Institute at UChicago, they will be exposed to world class academic research environment that caters to the general sciences. At IIT, they will be exposed in cutting edge research projects in data-intensive computing at extreme scales. All these institutions are easily accessible from each other, with IIT and UChicago being 4 miles apart, and ANL being 22 miles away.

### **Activities aimed at Improving Verbal and Written Communication Skills**

Perhaps the most important ability of a researcher is the ability to communicate results – an essential activity for expanding scientific and technical knowledge. To this end, students' research activities will involve many opportunities to increase their communication skills – from sharing their work at the individual level to sharing it with the entire scientific community.

At the individual level, each mentor will meet with students several times per week in a small team meeting, including the mentor, the graduate student, and the undergraduate student team. This will give the students practice in summarizing their work and presenting it to others. Further, it will give the students practice in formalizing their work and being able to explain it to an audience, while still allowing feedback from interested and knowledgeable parties.

At the team level, students will meet as part of the larger DataSys/Systems/Labs research groups. These weekly meetings will generally be held on Friday afternoons, and will generally be informal enough to facilitate an exchange of ideas, while still providing students an opportunity to become proficient at presenting their work. Furthermore, students will be asked to present formally to all REU participants, as well as interested outsiders. There will be two such presentations during each summer, a mid-term presentation and a final presentation. During the mid-term presentation, each student team will be asked to present to the entire group a summary of their project as well as describing motivation, goals, and progress. At the second meeting, each student team will be asked to present their final results. These presentations will be prepared with mentor input and formally critiqued by distributing a critique sheet to each member of the audience. Each presentation will be recorded digitally using IIT's distance-learning facilities both to provide feedback to the presenter and to offer a memento of the REU experience.

Students will be encouraged to present their work in formal settings. One such opportunity will be a presentation at the student's home institution, which will not only provide recognition among that student's peers, but will also widen the recognition of the REU BigDataX site and enhance the following

year’s recruitment. Other venues include numerous symposia for undergraduate research, such as the Chicago Area Undergraduate Research Symposium (for Chicago land-area students) as well as conferences like the Grace Hopper Conference. Where appropriate, students will be encouraged to prepare their work for presentation at a refereed conference or for submission to a journal. One likely venue is the IEEE/ACM Supercomputing conference series, which has yearly deadlines on July 31<sup>st</sup> for posters, and has workshop papers due in September every year. Both posters at the main conference and/or workshop papers are a good and realistic goal for the BigDataX students. For exceptional research results, there are various conferences with deadlines between September and January, such as USENIX NSDI, IEEE IPDPS, IEEE/ACM CCGrid, and ACM HPDC, all top venues which could make a significant positive difference for these undergraduate students if they decide to pursue doctorate studies.

### Typical Summer Schedule & Special Events

A recent book on “Undergraduate Research in the Sciences: Engaging Students in Real Science” by Laursen outlines six major types of benefits from undergraduate research experience: personal/professional gains, intellectual gains, gains in professional socialization, gains in skills, enhanced preparation for graduate school, and career clarification and confirmation. A typical summer schedule of the BigDataX site is organized around these six areas as shown in Figure 1.

<b>Week 1</b>	Orientation, Tutorials (Amazon AWS, Swift, Hadoop, Spark, Chameleon, Mystic, and applications), how to search for and review research papers tutorial, entrance survey
<b>Week 2</b>	Progress reports at UChicago, “Career Paths with a PhD in Computer Science” panel
<b>Week 3</b>	Progress reports at IIT, “How to Identify the best conferences/journals” presentation, Reading Group
<b>Week 4</b>	Progress reports at UChicago, “How to give a good presentation” presentation, Reading Group
<b>Week 5</b>	Students’ midterm presentation; Social Activity Summer Picnic, Survey
<b>Week 6</b>	Progress reports at UChicago, "A Day in the Life of a PhD student" panel, Reading Group
<b>Week 7</b>	Progress reports at IIT, Argonne National Laboratory field trip, Reading Group
<b>Week 8</b>	Progress reports at UChicago, “Research at IIT/UChicago” event, Reading Group
<b>Week 9</b>	Progress reports at IIT, “Women in Science” Panel, Reading Group
<b>Week 10</b>	Students’ final presentations, deliverables, exit survey

Figure 1: BigDataX typical summer schedule

**“A Day in the Life of a PhD Student” Panel:** This event will be a round-table discussion where several PhD students will be invited to talk about the application process, their experiences as PhD students, and their expectations after graduation (see Figure 3).

**Research at IIT/UChicago:** This event will be an afternoon session where IIT and UChicago faculty will be invited to give short research presentations to students on topics outside the big data area. The goal of the event is to introduce the students to other research areas in the computing discipline.



Figure 2: “A Day in the Life of a PhD Student” Panel in summer 2017

**“Career Paths with a PhD in Computer Science” Panel:** This event will be a round-table discussion where several REU mentors will describe possible career paths of someone with a PhD degree in computer science. The goal of the event is to introduce the students to the possible paths they can take if they

choose to complete a PhD degree: 1) academia (teaching faculty and tenure-track faculty), 2) national laboratories (researcher and developer), and 3) industry (researcher, developer, and entrepreneur). The panel tries to discuss not just what the different tracks are, but also what the pros and cons are of each career path.

**“Women in Science” Panel:** This event will be a round-table discussion where several successful women in computing discuss the challenges and opportunities of being a woman in a male-dominated field (see Figure 3). The aim is to show that success in computing can be achieved even for women and minorities.



Figure 3: “Women in Science” Panel in summer 2017

**Argonne National Laboratory Field Trip:** This event will be a whole day event for the REU students to visit Argonne National Laboratory (see group photos below). We plan to have many activities, such as hearing from scientists describe their positions, careers, and research; we will aim to have both computer scientists and domain scientists from physics, chemistry, and biology present in order to give students a well-rounded view of research careers in STEM disciplines. We plan to participate in a tour of the ALCF machine room where the #9 (Mira) and the #16 (Theta) supercomputers from the Top500 list can be found. This field trip has been a highlight for most of the REU students in the past, and has proven to be quite inspirational with where they could be in 10 years in the future.

**Reading Group:** This will provide a setting in which REU researchers can review literature in an informal setting with graduate students, mentors, and UChicago and IIT faculty. The Reading Group will meet weekly and a paper will be presented by one REU team. Under the supervision of their mentors, the teams will select papers relevant to their projects and summarize the content to be presented. All REU students will be encouraged to review each article by answering the following: 1) what are the contributions?; 2) what are the strengths?; and 3) what are the weaknesses?.

**Summer Picnic:** The picnic aims to forge friendships between the REU students, and to get them to open up with their mentors in a non-academic setting (see Figure 5). Mentors would cook food on a coal grill, while students would mingle and enjoy outdoor activities.



Figure 4: REU Picnic in summer 2017, joint between BigDataX REU, MEDIX REU, and SCC programs

### Group Photos from 2015 to 2017

Photos from the prior summers can be found at:

- <http://datasys.cs.iit.edu/grants/BigDataX/2015/pictures/index.html>
- <http://datasys.cs.iit.edu/grants/BigDataX/2016/pictures/index.html>
- <http://datasys.cs.iit.edu/grants/BigDataX/2017/pictures/index.html>

Below is BigDataX REU 2017 Site along with the MEDIX REU 2017 Site at the summer picnic in July 2017.



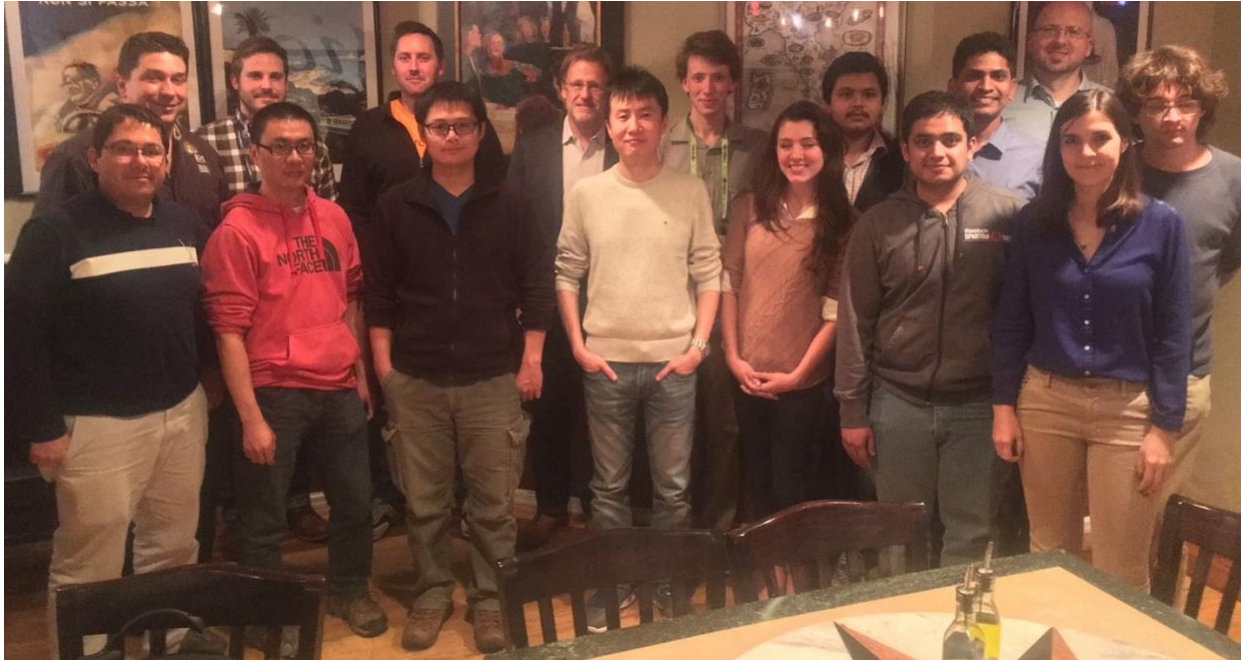
Below is Blue Keleher and Emily Herron (REU 2017 students) won 2nd and 3rd place respectively in the ACM Undergraduate Student Research Competition at IEEE/ACM SC 2017 conference.



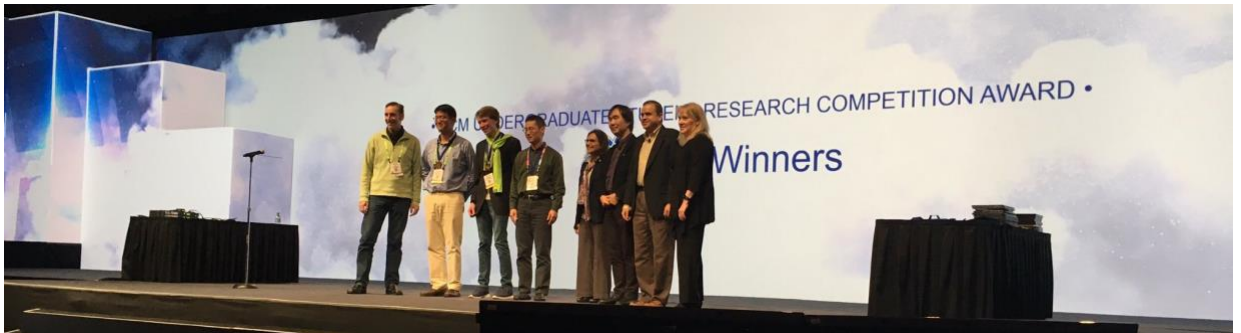
Below is BigDataX REU 2016 Site along with the MEDIX REU 2016 Site visiting Argonne National Laboratory in June 2016.



Below are REU and DataSys students at the IEEE/ACM Supercomputing/SC 2016 conference in Salt Lake City Utah in November 2016.



Below is William Agnew (REU 2016 student) receiving the ACM Undergraduate Student Research Award at the IEEE/ACM SC 2016 conference.



Below is BigDataX REU 2015 Site along with the MEDIX REU 2015 Site visiting Argonne National Laboratory in July 2015.

