

OUTLINE FOR AN INTERDISCIPLINARY CERTIFICATE PROGRAM

I. Basic Information

1. Institution: University of Georgia Date: September 30, 2015
2. School/College: Franklin College of Arts and Sciences
3. Department/Division: Department of Computer Science
4. Certificate Title (as it will appear in the *Bulletin*): Applied Data Science
5. Level (undergraduate or graduate): Undergraduate
6. Proposed starting date for program: Fall 2016
7. Abstract of the Certificate Program (for the University Council's agenda):

Purpose: The Computer Science Department is proposing the following new certificate program in Applied Data Science. This certificate would be useful for students in a variety of mathematical, scientific and engineering fields. The certificate would develop expertise in the collection, storage, analysis, visualization and interpretation of data. Efficient processing of large quantities of data via frameworks like Hadoop and Spark would also be explored. The need for expertise in data analytics and data science has grown tremendously in recent years. The proposed certificate program is intended to help meet these needs.

Eligibility: Admission is open across the university, but is targeted to the mathematical, scientific (including biological, physical and social sciences) and engineering fields. The minimal prerequisite for courses in the certificate program is MATH 1113 Precalculus.

Curriculum: the Undergraduate Certificate Program in Applied Data Science requires 18-20 hours of coursework with 12-hours of core courses and 6-8 hours of electives. Bold indicates new courses.

Core Courses (12 hours):

CSCI 1301-1301L Introduction to Computing and Programming or
CSCI 1360 Foundations for Informatics and Data Analytics

CSCI 2150-2150L Introduction to Computational Science or
CSCI (MATH) 2610 Discrete Mathematics for Computer Science

CSCI 2360 Data Science I

Electives (6-8 hours):

CSCI 4130 CUDA C Programming on GPUs for High Performance Computing
CSCI 4150 Numerical Simulations in Science and Engineering
CSCI 4210 Simulation and Modeling

Electives continued:
CSCI 4250 Computer Security
CSCI 4360 Data Science II
CSCI 4370 Database Management
CSCI 4380 Data Mining
CSCI 4780 Distributed Computing Systems
CSCI 4850 Biomedical Image Analysis
STAT 4210 Statistical Methods
STAT 4230 Applied Regression Analysis
STAT 4280 Applied Time Series Analysis
STAT 4630 Statistical Methods in Bioinformatics I

8. Letters of support with signatures: see attached letter.

II. Response to the Criteria for All Programs

1. Purpose and educational objectives:

- A) Purpose and educational objectives: The Computer Science Department is proposing a new certificate program in Applied Data Science. Data Science has emerged as an important new discipline that is most closely aligned with Computer Science and Statistics. Contributing areas from Computer Science include Machine Learning, Database Systems, Data Mining, Cybersecurity, Computational Science, Distributed Systems and Data Intensive Computing. In Applied Statistics, contributing areas include Regression, Time Series Analysis and Bayesian Models. This certificate would be useful for students in a variety of mathematical, scientific and engineering fields. The certificate would develop expertise in the collection, storage, analysis, visualization and interpretation of data. Efficient processing of large quantities of data via frameworks like Hadoop and Spark would also be explored. The need for expertise in data analytics and data science has grown tremendously in recent years. Forbes reports an 89.9% increase in the last twelve months in the demand for employees with big data expertise (<http://www.forbes.com/sites/louiscolombus/2014/12/29/where-big-data-jobs-will-be-in-2015/>). The article also lists the median salary as \$103,000 per year. The proposed certificate program is intended to help meet these needs.
- B) Interdisciplinary nature: The program involves elective courses from Computer Science and Statistics departments and is targeted to students in mathematical, scientific and engineering fields.

2. There must be a demonstrated and well-documented need for the program:

- A) Why this certificate program is necessary: Students majoring in several disciplines would benefit from the proposed certificate program as new classes will be provided and existing courses will become more structured to support the certificate program. As mentioned in the Forbes article, Data Science offers an excellent career path with many well-paying jobs being available. As the volume of data continues to increase, the expectation is that the need for Data Scientists will increase as well.

B) Describe the expected stage of development:

1. Semester/Year of Program Initiation: Fall 2016
2. Semester/Year Full Implementation of Program: Fall 2016
3. Semester/Year First Certificates will be awarded: Spring 2017
4. Annual Number of Graduates expected (once the program is established): 50
5. Projected Future Trends for number of students enrolled in the program: expect continued growth

3. Evidence of student demand for program sufficient to sustain reasonable enrollments:

- A) Student interest in the program: In Computer Science (with over 745 majors and 135 minors), the current courses related to the certificate program have expanding enrollments. Nationally, a large number of universities have started programs in Data Science or Data Analytics. The trend is continued expansion.
- B) Diversity: The program is expected to have diversity similar to that in degree programs it draws from. Advertising and outreach to minorities will be included.

4. Design and curriculum of the program:

A) Detailed Curriculum:

Eligibility: Admission is open across the university, but is targeted to the mathematical, scientific (including biological, physical and social sciences) and engineering fields. The minimal prerequisite for courses in the certificate program is MATH 1113 Precalculus.

Curriculum: the Undergraduate Certificate Program in Applied Data Science requires 18-20 hours of coursework with 12-hours of core courses and 6-8 hours of electives. Bold indicates new courses.

Core Courses (12 hours):

CSCI 1301-1301L Introduction to Computing and Programming **or**
CSCI 1360 Foundations for Informatics and Data Analytics

CSCI 2150-2150L Introduction to Computational Science **or**
CSCI (MATH) 2610 Discrete Mathematics for Computer Science

CSCI 2360 Data Science I

Electives (6-8 hours):

CSCI 4130 CUDA C Programming on GPUs for High Performance Computing

CSCI 4150 Numerical Simulations in Science and Engineering

CSCI 4210 Simulation and Modeling

CSCI 4250 Computer Security

CSCI 4360 Data Science II

CSCI 4370 Database Management

CSCI 4380 Data Mining

CSCI 4780 Distributed Computing Systems

CSCI 4850 Biomedical Image Analysis
STAT 4210 Statistical Methods
STAT 4230 Applied Regression Analysis
STAT 4280 Applied Time Series Analysis
STAT 4630 Statistical Methods in Bioinformatics I

Descriptions of New Courses

CSCI 1360 Foundations for Informatics and Data Analytics

Prerequisite: MATH 1113.

Provides an introduction to concepts in scientific programming and data science using the Python language. Students are given hands-on opportunities to learn techniques applicable to quantitative analyses across a broad range of fields. These core techniques involve formulating solutions in terms of their inputs and outputs (functional programming), repeated operations (loops), branching operations (conditionals), different methods of organizing data (data structures), how to implement an optimal problem-solving strategy (algorithm design), and methods for visualizing and interpreting results.

Sample Textbook: *Python for Informatics: Exploring Information*

CSCI 2360 Data Science I

Prerequisites: (CSCI 1301 or CSCI 1360) and (MATH 2250 or CSCI 2150).

For students with a basic knowledge of programming. The course will cover techniques for collecting, storing, and analyzing data in varying formats. Scientific programming, supervised and unsupervised analytics, and data visualization techniques will be covered.

Sample Textbook: *Data Science from Scratch: First Principles with Python*

CSCI 4360/6360 Data Science II

Prerequisite: CSCI 2360 or equivalent.

A basic course on sophisticated data science techniques for making sense of data. Topics involve decision trees, convex optimization, randomized algorithms, out-of-core data processing, and just-in-time compilation. Students will also be introduced to distributed programming frameworks (e.g. Spark, Hadoop).

Sample Textbook: *An Introduction to Statistical Learning: with Applications in R*

Implementation: Typically, students may complete the certificate program in two years as a supplement to their degree program. A student seeking a B.S. in Computer Science may obtain a certificate in Applied Data Science by choice of electives without any delay in graduation. Many of the existing courses are already offered frequently. For the new courses, CSCI 1360 and CSCI 2360 would be offered every semester, while CSCI 4360/6360 would be offered once per year.

- B) Model Programs and Curricula:** The University of California, San Diego has a five course undergraduate certificate program in Data Mining (<http://extension.ucsd.edu/programs/index.cfm?vAction=certDetail&vCertificateID=128&vStudyAreaID=14>) that is similar to the proposed certificate. The proposed certificate program is broader in scope though. The University of Montana offers a certificate in Big Data Analytics (<http://www.umt.edu/datascience/big-data-courses.php>) with courses similar or the proposed certificate, but has more of an emphasis in Business. The Department of Computer Science and Engineering at the University of Washington has a certificate

program (<http://www.pce.uw.edu/certificates/data-science.html>) that is very similar and is described as follows,

“Acquire the computer science, mathematics and analytical skills needed to enter the field of data science. Use data science techniques to analyze and extract meaning from extremely large data sets, or big data. Become familiar with relational and non-relational databases. Apply statistics, machine learning, text retrieval and natural language processing to analyze data and interpret results. Learn to apply data science in fields such as marketing, business intelligence, scientific research and more.”

Compared to undergraduate certificates, there are more universities offering graduate certificates as well as full degree programs at the Bachelors, Masters and Doctoral levels. See (101.datascience.community/2012/04/09/colleges-with-data-science-degrees/) for an extensive list of such programs.

- C) Program Accreditation: The certificate program is administered by the Computer Science Department which is ABET accredited.

5. Faculty resources

- A) Full-time faculty: Full-time faculty within the Department of Computer Science are sufficient to initiate the proposed certificate program. A good fraction of the department's faculty conducts research in Data Science and Big Data Analytics. Most of the courses listed in the curriculum are already being taught and the three new courses are concurrently being formally proposed.

B) List of Involved faculty:

- Hamid R. Arabnia, Professor, Ph.D., Kent at Canterbury
- Ismailcem Budak Arpinar, Associate Professor, Ph.D., Middle East Technical University
- Brad Barnes, Lecturer, Ph.D., University of Georgia
- Suchendra M. Bhandarkar, Professor, Ph.D., Syracuse University
- Liming Cai, Professor, Ph.D., Texas A&M University
- Prashant Doshi, Associate Professor, Ph.D., University of Illinois at Chicago
- Daniel M. Everett, Assistant Professor, Ph.D., University of Wisconsin
- Funk, Shelby H., Associate Professor, Ph.D., University of North Carolina
- Bill Hollingsworth, Lecturer, Ph.D., University of Cambridge
- Maria Hybinette, Associate Professor, Ph.D., Georgia Institute of Technology,
- Krzysztof J. Kochut, Professor, Ph.D., Louisiana State University
- Kyu Hyung Lee, Assistant Professor, Ph.D., Purdue University
- Kang Li, Professor, Ph.D., Oregon Graduate Institute
- Tianming Liu, Professor, Ph.D., Shanghai Jiaotong University
- John A. Miller, Professor, Ph.D., Georgia Institute of Technology
- Roberto Perdisci, Associate Professor, Ph.D. University of Cagliari
- Walter D. Potter, Professor, Ph.D., South Carolina
- Shannon Quinn, Assistant Professor, Ph.D., Carnegie-Mellon-University of Pittsburgh
- Lakshmish Ramaswamy, Associate Professor, Ph.D.: Georgia Institute of Technology
- Khaled, Rasheed, Associate Professor, Ph.D.: Rutgers University
- Thiab Taha, Professor and Head, Ph.D., Clarkson University

Note: More detailed information about the listed Faculty above can be found at:
<http://www.cs.uga.edu/directory/front>

- C) Additional Faculty: Big data analytics and data science is already a hiring priority for the Department of Computer Science, but we anticipate that CSCI 1360 will become a popular class with high enrollments as we hope to list it as a course supporting the university's Informatics Initiative. To help with this course and others in the certificate program, we anticipate a future need to hire a Lecturer to support it and the university's Informatics Initiative.

6. Resources needed to support the program:

- A) Library resources: no need for additional library resources
- B) Equipment: no need for additional equipment

7. Physical facilities: no need for additional physical facilities

8. Expense to the institution:

- A) Funding to initiate the program (first three years): No amount of funding is needed for Years 1-3.
- B) Support for students: As an undergraduate certificate program, several students will likely have Hope Scholarship. Also, students in the certificate program would be well positioned for internships.

9. Commitments of financial support:

- A) Sources of additional funds: Current funding through the Department of Computer Science will be sufficient to initiate and maintain the certificate program. Additional funding will be needed for expansion.
- B) Long-range plans: Additional funding in the form of a Lecturer position will be needed for expansion of the certificate program.

10. Administration of the program:

The proposed undergraduate certificate in Applied Data Science will be administered by the undergraduate coordinator of the Computer Science Department. The administrator in conjunction with the Department Head will be responsible coordinating course offerings, maintaining student records, promoting activities, securing additional funding, and consulting with the department's curriculum committee regarding courses in the certificate program. Students having taken the prerequisite course, MATH 1113 Pre-calculus, are eligible to take courses in the certificate program. All courses in the certificate program must be passed with a grade of C or better. The semester before completing the certificate, students will be required to fill out a certificate completion form. The undergraduate certificate will be awarded to the student upon the completion of the undergraduate degree.