Request for Certificate in Big Data

Institution Submitting Request: *School of Computing*
Proposed Title: *Certificate in Big Data*
Currently Approved Title: *N/A*
School or Division or Location: *School of Computing*
Department(s) or Area(s) Location: *Computer Science*
Recommended Classification of Instructional Programs (CIP) Code[1] (for new programs): *00.0000*
Current Classification of Instructional Programs (CIP) Code (for existing programs): *11.0701*
Proposed Beginning Date (for new programs): *08/23/2014*
Institutional Board of Trustees' Approval Date: *02/11/2014*

Proposal Type (check all that apply):

Regents' General Consent Calendar Items		
R401-5 OCHE Review and Recommendation; Approval on General Consent Calendar		
SECTION NO.		**ITEM**
5.1.1	☐	Minor*
5.1.2	☐	Emphasis*
5.2.1	☐	(CER P) Certificate of Proficiency*
5.2.3	☒	(GCR) Graduate Certificate*
5.4.1	☐	New Administrative Unit
	☐	Administrative Unit Transfer
	☐	Administrative Unit Restructure
	☐	Administrative Unit Consolidation
5.4.2	☐	Conditional Three-Year Approval for New Centers, Institutes, or Bureaus
5.4.3	☐	New Center
	☐	New Institute
	☐	New Bureau
5.5.1	☐	Out-of-Service Area Delivery of Programs
5.5.2	☐	Program Transfer
	☐	Program Restructure
	☐	Program Consolidation
5.5.3	☐	Name Change of Existing Programs
5.5.4	☐	Program Discontinuation
	☐	Program Suspension
5.5.5	☐	Reinstatement of Previously Suspended Program
	☐	Reinstatement of Previously Suspended Administrative Unit

**Requires "Section V: Program Curriculum" of Abbreviated Template*

Chief Academic Officer (or Designee) Signature:
I certify that all required institutional approvals have been obtained prior to submitting this request to the Office of the Commissioner.

R Watkins
Signature

Date: 02/12/2014

Printed Name: Ruth Watkins
Ruth Watkins

[1] CIP codes must be recommended by the submitting institution. For CIP code classifications: http://nces.ed.gov/ipeds/cipcode/Default.aspx?y=55.

Program Request - Abbreviated Template
College of Engineering, School of Computing
Certificate in Big Data
11/25/2013

Section I: Request

The School of Computing in the College of Engineering is seeking to provide a graduate certificate in big data for several types of students outside the school. The proposed certificate would consist of 15 credits made up from 5 classes: Data Mining, Machine Learning, Database Systems, Visualization, and Advanced Algorithms, all offered in the School of Computing. The program would serve both graduate students from related areas such as Biomedical Informatics, Electrical and Computing Engineering, and Math, as well as professional computer scientists working in local industry who hope to expand their skills to deal with the many related emerging big data analysis and management problems.

This proposal was formed as a joint effort from a committee of 12 faculty in the School of Computing: Mary Hall, Charles Hansen, Chris R. Johnson, Mike Kirby, Feifei Li, Miriah Meyer, Valerio Pascucci, Jeff Phillips (chair), Paul Rosen, Jur Van Den Berg, Kobus Van Der Merve, and Suresh Venkatasubramanian.

The proposed certificate was discussed in a School of Computing faculty meeting on October 25, 2013, and was endorsed by the faculty.

Section II: Need

There is a quickly growing need for skills in Big Data. The McKinsey report Big Data: The Next Frontier for Innovation, Competition, and Productivity, estimates that the United States will face a shortage of 140 – 190 thousand deep analytical talent positions and 1.5 million data savvy manager positions in the coming years. While many existing degrees aim to specifically train the data savvy managers, including the MSIS degree in the Utah Business School, this program is more focused on the more technical skills required for the deep analytical talent positions. Companies will highly value individuals with these technical skills since, for instance, data in the health care sector alone is estimated to be worth more than $300 billion every year. To harness this data, skilled big data scientists are essential to *understand*, *process,* and *present* this data. These three interlocking skills require more than just how to run a piece of rapidly changing analytical software, but also how to adapt to new techniques, to know which options to attempt based on the data properties, how to scale these approaches to every growing data sets in size and complexity.

These needs are not just a national trend; they are especially relevant in Utah. The School has been approached by various technology companies with strong local ties, such as Adobe, EMC, Doma, Goldman Sachs, Zions Bank, NSA, as well as many smaller start-up operations with foresight to request these skills in future employees.

Conveniently, the School of Computing offers graduate level classes in just these areas, taught by experts in these areas. Moreover, the relevant faculty of these courses are cognizant of the role of each in the big data phenomenon, as demonstrated by a technical white paper written in the last year: Rethinking Abstractions for Big Data: Why, Where, How, and What? (www.arXiv.org/abs/1306.3295).

By formalizing this certificate, it will guide students towards the particular skills and methodologies to succeed not just in the next few years, but for the next decades. It will allow the University of Utah to put a uniform and consistent stamp of accomplishment on a focused set of five classes that will help employers identify these quality future employees.

Related Programs

At the University of Utah, the two most related programs are larger and at the Masters level. The School of Computing offers a MS in Computing with a specialization track in Data Management & Analysis (soon to be renamed, and henceforth referred to as the Data track) and the David Eccles School of Business offers an MS in Information Science (MSIS). Both degrees require twice the coursework as the proposed certificate, which may be more than desired for students outside of those particular Schools. In particular, the proposed certificate will be managed by the faculty chairperson of the School of Computing Data track, and the certificate courses will strongly overlap with required courses within that track. In contrast to the MSIS, which is intended to cater to students who do not necessarily have a strong CS or programming background, the Big Data Certificate is designed to build on existing CS and programming skills. The MSIS program also offers a certificate associated with that program; it requires 18 elective credits within the Operations and Information Systems department at the DESB.

There already exist several intermountain programs in the related areas of Big Data, Data Science, and Business Analytics. The main distinguishing factor in the proposed program is that these classes are intended to build on existing computer science and programming skills, and teach the fundamentals of advanced computing and analytical techniques for big data.

- Business Intelligence @ University of Denver. Masters Degree from the Business School. http://daniels.du.edu/specialized-masters-degrees/business-intelligence/
- Business Analytics @ Arizona State University. Masters Degree from the Business School. http://wpcarey.asu.edu/master-business-analytics/
- Data Science Certificate @ University of Washington. 3 course tools-based certificate. Taught by adjunct professors from local industry. http://www.pce.uw.edu/certificates/data-science.html

On a national level there are a handful of programs that may be comparable to the one proposed, and that number is quickly growing. Many are like the tool-based programs mentioned above, and are often taught out of business schools or by adjunct professors. However, there are a few programs aiming to train the deep analytical talent positions. For instance, in New York City there are two programs catering to professionals living in the city:

- Data Science @ NYU. Masters Degree. http://cds.nyu.edu/academics/ms-in-data-science/
- Data Science and Engineering @ Columbia. Interdisciplinary Masters Degree. http://idse.columbia.edu/foundations-data-science

And there is another related nationally prominent program in North Carolina. It is an intensive 1 year program that dates back to 2007.

- Advanced Analytics @ NC State. Masters of Science in Analytics. http://analytics.ncsu.edu

We believe that the proposed program has the potential to not only be a compliment to these programs in the West, but also to compete with them nationally given the expertise of the faculty involved.

This certificate program will make gaining a post-secondary degree or certificate more accessible for many Utahns who are well qualified to make this next step. It targets students who are have a bachelor's degree and may serve as a feeder program into a more comprehensive MS degree in Computing under the related track. Altogether, this will have a positive and tangible effect on the Governor's call that 66% of Utahns will have a post secondary degree or certificate by 2020.

Section III: Institutional Impact

No substantial institutional changes are needed. All classes already exist, and a similar track (which results in a MS or PhD in Computing) is offered in Data. This certificate program may serve as a feeder program into the related MS program. The Big Data Certificate and its students will be administered by the Data track faculty.

The program is also planning serveral features designed for making the classes more convenient for proferssional students (online video, internet-video office hours, scheduling considerations). These will be deployed in such a schedule so that more distance education features are offered proportional to the increased enrollment, and so that they will not incur any additional net costs. Early experimentation with these features has yeilded extremely positive feedback. One such feature for professional students would be to offer the classes in consecutive blocks in the late afternoons and evenings, and the relevant faculty slated to teach the required classes have agreed to reschedule their classes if they are not already at these times.

In fact, it is hoped that if this certificate program and the related masters degree are successful, then many of the distance education techniques can be easily extended to other programs and tracks offered by the School of Computing and other departments at the University.

Section IV: Finances

No outside financing is required to offer this certificate. The department hopes it will attract a significant number of new students to existing classes already taught on a regular basis.

The program hopes to offer videotaped lectures to the enrolled students. These students may be regular matriculated students, professional masters students, in addition to the students primarily in the classes for the certificate. Utah Video Services Coordinators offer to videotape lectures for roughly $4000 per 3-credit class. This cost can be offset by a few additional students per class. Alternatively, the School has been experimenting with TAs or a rotating set of students recording videos for class, and then uploading them with minimal processing to a free hosting site (such as a private channel on YouTube). This incurs no additional cost, using existing off-the-shelf camera equipment. It does require some time, but is not a major burden on a TA, especially since making videos of lectures available often reduces the questions asked of the TA from students who can watch a missed lecture, or re-watch the relevant part of a lecture.

It is hoped that experimentation with video taping of lectures can serve as a model for many more classes in the future. With the gained experience, the cost of such expenditures should only decrease as the program grows.

Section V: Program Curriculum

Course Prefix and Number	Title	Credit Hours
Required Courses	CS 6140: Data Mining CS 6350: Machine Learning CS 6630: Visualization CS 6150: Advanced Algorithms	3 3 3 3
	Sub-Total	12
Elective Courses	CS 6530: Database Systems CS 5530: Database Systems	3 3
	Sub-Total	3
Track/Options (if applicable)		
	Sub-Total	0
	Total Number of Credits	15

Program Schedule

All certificate classes are offered every year in either the Spring or Fall semester. In the past CS 6350: Machine Learning was offered in the Fall semester, but starting with the 2014-2015 academic year, it will be offered in the Spring semester to accommodate a more balanced schedule for, among other things, this certificate. That is, starting from Fall 2014, courses CS 6530, CS 6630, and CS 6150 are offered every Fall, and course CS 6140, CS 6350, and CS 5530 are offered every Spring.

A typical schedule for a student aiming to complete the certificate in a single year would look as follows:

Fall (Semester 1):
CS 6630: Visualization – 3 credits
CS 6530: Database Systems – 3 credits
CS 6150: Advanced Algorithms – 3 credits

Spring (Semester 2):
CS 6140: Data Mining – 3 credits
CS 6350: Machine Learning - 3 credits

Alternatively a 4 semeseter variant, might be popular among very busy professional students or a student completing a 2 year degree in another department. A suggested more spread out schedule is as follows:

Fall (Semester 1):
CS 6530: Database Systems – 3 credits

Spring (Semester 2):
CS 6360: Data Mining – 3 credits

Fall (Semester 3):
CS 6630: Visualization – 3 credits
CS 6150: Advanced Algorithms – 3 credits

Spring (Semester 4):
CS 6350: Machine Learning - 3 credits

Training in Big Data Ethics

Big Data raises many ethical issues, and the School feels it is important that the students obtaining this certificate are educated in these issues. All of the core courses include lectures which discuss ethical issues associated with data management and analysis. The School of Computing will ensure that these topics remain a point of emphasis in the required classes.

Data Mining:
The course includes a segment about the privacy of data. It describes successful and un-successful de-anonymization techniques to prevent the leakage of private information while allowing global structure to be mined. Case studies are presented where these techniques have failed; the consequences for all parties are discussed. The related ethical issues are also discussed in various forms throughout the class, including in a hands-on project with real data.

Machine Learning:
This course will be interspersed with discussions about the societal impact of machine learning and the choices that practitioners may face in the design and choice of learning algorithms. It will cover the interpretability and auditabilty of decisions made by learned programs and how this impacts the choice of learning algorithms. It will also include discussions about ethical questions involved when machine learning is applied to sensitive information (such as health records, personal emails or law enforcement data).

Database Systems:
This course discusses access control and security issues in database systems, and the related ethical issues hinged on such topics. Numerous sensitive data are stored in a database system. Without proper mechanism in place to ensure and enforce a set of desirable access policies, such databases are not deployable in practice. The design and implementation of modules inside the modern database systems that safeguard access to the data will be discussed. Nevertheless, any database systems cannot prevent inside-attacks. Hence, as part of the ethical training, the ethical behavior of database administrators and application developers will be discussed.

Algorithms:
The course will include a unit on the larger social context of algorithm design. Topics that might be covered include work on algorithm design to ensure fairness in allocating resources, the quantifying of privacy as a resource and how to preserve it, techniques to obfuscate (or conversely, make transparent) code execution, as well as theoretical concepts that form the basis of explanation and interpretation of data analysis. The consequences of these choices will be discussed.

Visualization:
This course covers the ethics behind creating visual representations of data. We discuss principles such as visualizing data in context, the importance of meaningful baselines, the Lie Factor, and proper labeling. Using these principles students not only critique existing visualizations, but also design their visuals to tell stories with data in an open, and transparent, way.

Council Approval

Note: This form is intended to track the progress of a proposal (whether from Academic Affairs or Health Sciences) through the Undergraduate and Graduate Councils.

Proposal: Graduate Certificate in Big Data

This proposal needs to go through:

Undergraduate Council	☐
Graduate Council	X
Both Approvals	☐
Grad Approval/Undergrad Notification	☐

This proposal has been approved by:

Chair of Undergraduate Council _____ Date: _____

Chair of Graduate Council _____ Date: 12/4/13

Once the appropriate signature(s) have been obtained, please forward this completed form to the Office of the Senior Vice President for Academic Affairs. *(NOTE: The SVP-AA is the Chief Academic Office for the University of Utah and reports to the Board of Regents in this capacity. When necessary, the CAO will get a signature from the SVP-HSC.)*

Chief Academic Officer _____ Date: _____

Once the Chief Academic Officer's signature has been obtained, this approval document will be forwarded to the **Office of the Academic Senate**.

David Eccles
School of Business
THE UNIVERSITY OF UTAH

November 11, 2013

Jeff Phillips
School of Computing
University of Utah
50 South Central Campus Drive, 3190
Salt Lake City, UT 84112

Dr. Phillips:

It is our understanding that the School of Computing is proposing a graduate certificate in big data. I am writing this letter in response to your request for a letter of support from the David Eccles School of Business for your proposal.

As you noted in your proposal, the need for big data professionals has increased dramatically in recent years. Social media, cloud computing, and mobile technologies are contributing to demand for data savvy graduates from both the information systems and computing science departments at the University of Utah.

The David Eccles School of Business currently offers a Master of Science in Information Systems (MSIS) degree with a specialization in business intelligence and analytics and a graduate certificate in information systems. Included in both of these programs, are courses in business intelligence and analytics, data mining, database theory and design, data visualization, networking and servers, cloud computing, system analysis and design, etc. We note that your proposal includes courses that may appear to overlap with some of these areas (e.g. data mining, database technology, and visualization).

While there are some similarities between courses in the information systems and computer science departments, our courses have generally included different content. Our departments have also traditionally served different markets. Although the business school offers coursework in computer programming, our graduates do not receive the same level of training in programming or pursue the same positions as computer science graduates. Our MSIS and graduate certificate programs are intended to prepare students for business careers where an understanding of technology is of financial and strategic value to a firm. On the other hand, computing science graduates typically pursue careers in computer engineering or software development. Companies such as Adobe, EMC, Zions Bank, Goldman Sachs, and others recognize the differences between our programs and how we uniquely approach the big data domain. This has been readily apparent in discussions with technology leaders from these companies who serve on the MSIS advisory board.

In your graduate certificate proposal, you include a brief review of big data, data science, and business analytics programs in the intermountain west. We note that your review includes several business schools. You propose to differentiate your certificate program from a business-oriented program by focusing on computer science and programming skills. Our support for your proposal comes with the explicit understanding that the proposed certificate will not compete directly with programs or courses in the David Eccles School of Business or restrict our ability to continue to offer programs or courses in the area of big data. For example, we currently offer several courses in the area of business intelligence

and analytics. We view this as a uniquely business focused area and would expect students interested in such training to come to the business school. We recognize that courses such as data mining, can be taught from a business school or computer science perspective to meet the needs of different student audiences. For these classes, our expectation is that each school will continue to focus class content and presentation in a manner consistent with their core strengths. Further, our support of this certificate should not preclude the addition of a business focused big data certificate or degree program similar to that offered by many business schools throughout the country and as referenced in your proposal.

We also note the reference in your proposal to several multi-disciplinary programs in Data Science. Given the differences in expertise between our two departments, we remain optimistic that opportunities exist in the future for us to collaborate on additional courses or programs in the area of big data. We desire to continue discussions with the School of Computing regarding opportunities for cross-over between our many programs.

Thank you for the opportunity to review and respond to your proposed certificate program. We wish you the best as you take the next steps in the approval process.

Sincerely,

Taylor Randall
Dean
David Eccles School of Business

Richard B. Brown
Dean of Engineering
1692 Warnock Engineering Building
72 S. Central Campus Drive
Salt Lake City, Utah 84112
PH: (801) 585-7498 FAX: (801) 581-8692
brown@utah.edu
http://www.coe.utah.edu/~brown

November 5, 2013

Al Davis
Professor and Director
School of Computing
University of Utah
Salt Lake City, Utah

Dear Prof. Davis,

I am pleased to write a letter of support for the masters certificate in Big Data Engineering as proposed by Jeff Phillips and 11 other faculty from the School of Computing. This is a new area of strength within the College of Engineering, as seen by several recent faculty hires in this area and the creation of the Center for Extreme Data Management and Visualization (CEDMAV). It is also an area that has received a lot of interest from students. Within the School of Computing the number of students specializing in this area has been quickly increasing.

Masters students in the School of Computing can focus in this area as a track in the computing degree. But other students who may want to gain a background in Big Data have had no way to have that expertise recognized. The graduate certificate in Big Data will meet that need. The area has already been attracting a diverse set of students from outside the School of Computing. The core classes draw students from departments such as Civil Engineering, Biomedical Informatics, and Mathematics. Several students from local industry have also been taking these classes, often through continuing education, and I anticipate that this cohort will grow with a certificate being available.

The certificate, as designed and administered by Jeff and the other School of Computing faculty, will provide guidance on which classes are most important for this emerging topic. It will put a stamp of achievement on a transcript for those students. Furthermore, we believe that formalizing this certificate will attract new students, and encourage some of these students to continue on to obtain a masters degree in related areas.

In summary, Big Data is a priority technical area for the College of Engineering, and a certificate in this area fits well with the University's academic goals. We have considerable strength in this important emerging area, and establishing this graduate certificate in Big Data will help the U to better serve the needs of our students and local industry while building the reputation of of our School of Computing.

Sincerely,

Richard B. Brown
Dean of Engineering