Master of Science in Data Analytics
New Degree Proposal Approved by the Board of Governors
-Abridged Version-
I. PROGRAM DESCRIPTION

Data Analytics is an emerging discipline that seeks to infer insights from large amounts of data (“big data”) by using various statistical techniques and algorithms. The discipline is concerned with both statistical techniques that measure the validity of such insights and with computational techniques for managing data and resources efficiently. There is a great need for people with technical skills in these areas, prompted by the large amounts of information that governments and businesses are collecting. Thus, this degree program aims to train people to develop algorithms and computerized systems to facilitate the discovery of information from big data.

(a) This is an interdisciplinary Master of Science program in Data Analytics, offered jointly by the departments of Computer Science and Statistics.

(b) While there is no explicit tracks or specializations, the program emphasizes the technical aspects of big data analytics, including algorithm design, programming, acquisition, management, mining, analysis, and interpretation of data.

(c) It is a 30 credit hour program, with 24 credit hours of required courses and 6 elective credit hours. Students will be taught in cohorts by face-to-face instruction, and no transfer credits will be allowed. The program has a required semester-long project, but does not require a thesis.

(d) By graduation, students will be able to:
   1. Use state-of-the-art software tools to perform data mining and analysis on large structured and unstructured data sets, and transform such data into knowledge.
   2. Design and implement new algorithms for data mining and analysis, and study their time-, space-, and energy-efficiency.
   3. Perform data acquisition and management for extremely large and dynamic databases.
   4. Present and communicate knowledge derived from data in an unambiguous and convincing manner.

Thus, the overall goal is to provide technical skills to professionals in Florida. By 2020, zettabytes of data will be collected by governments and businesses. While governments want to use these data to improve the life of their citizens, businesses are keen on exploiting these data to better serve their clients. Consequently, there is an increasing demand for data analysts who can create, adapt, and use state-of-the-art tools to obtain insight from large structured and unstructured data sets, converting them into knowledge. Usually people with this training have the title of “data analyst” or “data scientist.” The US Bureau of Labor Statistics may classify people in these roles as statisticians, computer programmers, or other existing categories (such as “database administrator” or “software developer”). Graduates may go on to complete a Ph.D. in Computer Science, Statistics, or a related area and may also seek professional distinction.

This degree program will provide employable technical skills including the development of algorithms and computer systems to extract insight from big data. The curriculum includes several required courses that ensure that students have skills in algorithms and statistical techniques for extracting information, including:

- Parallel and Cloud Computation, which introduces students to parallel computing across the hardware-software stack with a special emphasis on emerging architectures and technologies,
- Network science, which helps students select and develop algorithms for finding relationships in large graphs, as occur, for example, in social networks,
- Machine learning, which helps students select and develop algorithms to classify data,
- Parallel and distributed databases, which teaches students how to work with data in the cloud, as is needed for processing big data, and
- Data mining methodology, which teaches students valid statistical techniques for finding insights from big data.
Students completing these required courses will take 2 electives from a set of 6 courses that provide more depth in visualization of data, extracting information from social networks, text mining, machine learning, and data preparation. They will also complete a required project in data analytics to demonstrate their mastery of the skills and techniques needed to extract knowledge from massive amounts of data.

The program will be offered on the main campus of UCF in face to face classes. The classes are planned for evenings and weekends. Students can use the STOKES cluster at UCF remotely, as well as cloud computing services from companies like Amazon, thus students will be able to access these computing resources from anywhere.

II. WHY AN MS IN DATA ANALYTICS?

Job demand for data analysts is already strong. As of March 23, 2016 on www.employflorida.com there were 60 active ads for positions that mention “data analyst” within a 50 mile radius of UCF, and 20 of these jobs were posted within the previous 14 days. In the preceding 14 days there were also 8 jobs posted for “data scientist” within a 50 mile radius of UCF, and 12 active positions including that job title. Since the title “data scientist” is more likely to be used for positions that require more technical background, as our degree, we found the following details out about the educational requirements for positions with the title “data scientist.”

As of March 23, 2016 on indeed.com, there were 893 active ads for the “data scientist” position that are listed in the last 15 days. Out of these jobs, 395 (44%) of them mention a graduate degree (M.S. or Ph.D.) requirement or preference in their definition. Among the positions posted on dice.com within the last 30 days of March 23, 2016, there were 579 active ads for positions that mention “data scientist” and 221 (38%) of them list a graduate degree in their required or preferred qualifications. The same numbers for careerbuilder.com were 131 and 22 (17%) respectively.

The following give additional information about skills needed and demand for jobs.

1. Ferris Jumah, a data scientist at LinkedIn, mapped the most popular skills of data scientists by scraping LinkedIn profile data of data analysts who live in San Francisco. The listed skills are as follows: Data Mining, Machine Learning, R, Python, Data Analysis, Statistics, SQL, Java, MATLAB, Algorithms. Most of these skills are not included in the curriculum of a single undergraduate program and to become familiar with these skills, a professional either requires industry experience or a graduate level degree. The MS degree will teach these skills (or skills like them if the technology moves on).
2. RJMetrics released a benchmark report titled ‘The State of Data Science’, which includes the results of a study on 11400 data scientists currently employed by companies on LinkedIn. The study includes the analyses of 60200 records of professional experiences, 27700 records of education, and 254600 records of skills. The results of the study show that over 79% of data scientists analyzed in the study have earned a graduate degree, and 38% have earned a Ph.D. The ratio of scientists with graduate degrees is even higher for senior data scientists when compared to junior data scientists. The scientists with only a B.S. degree is 17.86% for senior positions.
3. In 2014, O’Reilly Media conducted an anonymous survey to examine factors affecting the salaries of data analysts and engineers. The study had over 800 respondents from 53 countries and 41 states, who work in and around the data space. The main findings of the study are as follows:
   - 44 percent of the respondents have a master’s degree and 20 percent have a Ph.D. degree.
   - Having a doctorate degree adds $11k to the annual salary.
   - The respondent sample had a median total salary of $98,000 (U.S.).
4. Burtch Works, an Evanston based executive recruiting firm, has conducted a study with 371 data scientists. In this study, phone interviews were held with each participant. The main findings of the study are as follows:
• 92 percent of the participants have at least a master’s degree and 48 percent have a Ph.D. degree. While 86 percent of other Big Data professionals also have at least a master’s degree, 20 percent have a Ph.D.
• More than one third of the participants are not US citizens. They either have F1, OPT, or H1B visas, or are permanent residents.
• Nearly one third (29 percent) of the participants hold a degree in mathematics or statistics and one fifth (18 percent) hold a degree in computer science.
• The median compensation of data scientists varies primarily with years of experience, depth of expertise, education level and management responsibility, but can range from $91,000 with one to three years of experience up to $250,000 for managers leading teams of ten or more.
• Data scientists earn more than other predictive analytics professionals that strictly work with structured data. When entering the job market, for instance, a data scientist with 13 years of experience can earn 24% more than a predictive analytics professional with the same level of experience.

5. The EMC Data Science Community Survey interviewed 497 data scientists and business intelligence professionals from around the world, including deliberate samples in the United States, India, China, the United Kingdom, Germany, and France in 2011. According to this study, 40 percent of the participants have a master’s degree, and an additional 17 percent have a Ph.D.

6. The ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) was established in 1998. In 2006, SIGKDD published a two course curriculum proposal in data mining. In 2011 at least 19 academic programs were available in the US and Canada with a connection to Knowledge Discovery and Data Mining either as a focus or as a supplement to a primary area of study. By August 2013, the number of programs listed had grown to 61 of which only four were clearly undergraduate programs. (See P. Anderson et al. “An undergraduate degree in data science: curriculum and a decade of implementation experience.” In Proceedings of the 45th ACM technical symposium on Computer Science Education. pp. 145-150. March 2014, ACM.)

Many of the advertised positions that would be suitable for our graduates have the title of “Database Administrator”; in Orange County (Florida) positions with this title have an annual salary in the range of $49K to $87K, with a median of $73K. Walgreen’s, in their support letter (see Appendix G), said that they would pay someone $73K annually who graduated with the proposed MS degree. AllPoints, in their support letter, said they would pay someone about $90K annually. Other positions might have a title of “Computer Programmer” and such positions in Orange County have an annual salary in the range of $47K to $92K, with a median of $69K. Positions with a title of “Computer Systems Engineers” or “Computer Systems Architect” in Orange County have an annual salary in the range of $49K to $89K, with a median of $78K. (The data in this paragraph are from www.employflorida.com for 2014.)

The Bureau of Labor Statistics does not collect information about Data Scientists, but the job categories under which such people fall (including computer programmers) are expected to experience strong job growth. A report from McKinsey & Company (May 2011) asserts that “By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills”. According to the November 2013 report of the Commission on Higher Education Access and Educational Attainment, Florida faces a “critical gap” in the areas of “computer and information technology.” Graduates would contribute to filling this gap with their skills in data analytics and computer programming.

In summary, all evidence points to strong demand for this degree program.

The National Science Foundation (NSF) and other grant agencies have been increasing their support for research in the area of big data since 2012. For example, the NSF had a solicitation (14-543) titled “Critical Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)”. There was an event at the White House (on March 29, 2012) that announced the federal big data initiative. Quoting http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607&org=NSF&from=news: “At an event led by the White House Office of Science and Technology Policy in Washington, D.C., [NSF director] Suresh joined other federal science agency leaders to discuss cross-agency big data plans and announce new
areas of research funding across disciplines in this field.” Faculty teaching in the program will have an opportunity to capitalize on this increased level of funding, starting collaborations with ideas from students and their employers that are derived from master’s projects.

III. CURRICULUM

A. Learning outcomes.

The following are the learning outcomes associated with the proposed degree program.

O1. Be able to program projects using a distributed database system (such as Hadoop) to analyze data.
    Measured by: grades on assignments (in the classes “Parallel and Distributed Database” and “Parallel and Cloud Computation” that related to using such a distributed database system), and by programming project grades (from the required project course and in other courses with programming projects).

O2. Be able to use a statistical software package (such as R or SAS) to analyze large data sets.
    Measured by: grades on assignments (in “Data Mining Methodology I” and “Data Mining Methodology II” that relate to using such statistical software packages), and by programming project grades (in the required project course and in other courses with programming projects that relate to using such statistical software packages).

O3. Be able to accurately estimate the order and real-time needed to run a given algorithm on a specific hardware configuration and data set.
    Measured by: grades on specific questions on exams (in the course “Parallel and Cloud Computation” and “Parallel and Distributed Database” that relate to computational order and time estimation), and by grades on assignments (in the same courses and in the required projects course).

O4. Be able to write programs that prepare unstructured data for analysis.
    Measured by grades on assignments (in the class “Data Preparation”), and by programming project grades (in the required project course and in other courses with programming projects) that require preparation of unstructured data.

O5. Be able to give a convincing report both orally and in writing that communicates knowledge derived from an analysis of a large set of data.
    Measured by grades (on assignments in the required projects course), and by the programming project grades (in the required project course and in other courses with programming projects) that require oral or written reports.

These learning outcomes will enable graduates to help companies use large amounts of data in innovative ways. The outcomes are tied to the goals of the program as described above.

B. Admission standards and graduation requirements.

For admission, an undergraduate degree in Computer Science, Statistics, Computer Engineering, or Information Technology is desirable but not required. Applicants without a strong undergraduate background in Computer Science or Statistics must demonstrate an understanding of the material covered in the following undergraduate courses, by either taking these courses or by convincing the program that the student’s work experience or courses at other institutions have covered this material:

- COP 3330 Object-Oriented Programming
- COP 3503C Computer Science II
- COP 4710 Database Systems
- STA 2023 Statistical Methods I
- Programming experience or STA 4164 (Statistical Methods III)

The program’s director, assisted by the program’s faculty, will evaluate student applications. At the discretion of the director, prospective students with sufficient industry experience in computer programming will be deemed to have programming experience and the director will decide which of the prerequisites the student may need to make up as a non-degree seeking student (at UCF or elsewhere).
In addition to the general UCF graduate application requirements, applicants to this program must provide:

- One official transcript (in a sealed envelope) from each college/university attended.
- Official, competitive GRE score taken within the last five years.
- Resume
- Letters of recommendation (encouraged but not required)

Faculty members may choose to conduct face-to-face or telephone interviews before accepting an applicant.

Graduation with an MS in Data Analytics requires 30 credit hours beyond the bachelor’s degree, including 24 credit hours of required courses and 6 credit hours of electives. The degree involves a 3 credit hour required project course, which is a culminating experience. Students must receive a B or better grade in all courses in the program. However, if a student gets a B- or worse grade in a course, they may repeat that course in a future semester. Such a student will be permitted to join another cohort to make up such a class. The required and elective courses are listed in section E below.

The independent learning requirement is met by successful completion of a capstone project in the required course (CAP 6XXX) Project in Data Analytics.

C. Sequenced course of study.

Students will enroll in cohorts that start each fall semester. Each semester they take two courses; thus all students are part-time students and it will not be possible to take courses full-time in the program. The rationale for offering only enough courses for part-time students are: (1) we expect part-time students to be our largest student population and (2) we need to run the program in cohorts for administrative and economic reasons. Thus the students would all take the first 6 required courses in the first year of the program. The second year would be devoted to taking two more required courses, including the required projects course, and two electives. The required projects course will be offered in the summer as well as in spring, so that students who want to take that separately or who want to take more electives can be accommodated.

The sample schedule follows.

<table>
<thead>
<tr>
<th>Yr</th>
<th>Semester</th>
<th>Course Number</th>
<th>Course Name</th>
<th>Credits</th>
<th>Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>fall</td>
<td>CNT 5805</td>
<td>Network Science</td>
<td>3</td>
<td>(undergrad degree)</td>
</tr>
<tr>
<td>1</td>
<td>fall</td>
<td>STA 5206</td>
<td>Statistical Analysis</td>
<td>3</td>
<td>STA 2023</td>
</tr>
<tr>
<td>1</td>
<td>spring</td>
<td>CAP 5610</td>
<td>Machine Learning</td>
<td>3</td>
<td>(CAP 4630)</td>
</tr>
<tr>
<td>1</td>
<td>spring</td>
<td>STA 5703</td>
<td>Data Mining Methodology I</td>
<td>3</td>
<td>STA 5206</td>
</tr>
<tr>
<td>1</td>
<td>sum.</td>
<td>COP 5711</td>
<td>Parallel and Distrib. Database</td>
<td>3</td>
<td>COP 4710</td>
</tr>
<tr>
<td>1</td>
<td>sum.</td>
<td>STA 6704</td>
<td>Data Mining Methodology II</td>
<td>3</td>
<td>STA 5703</td>
</tr>
<tr>
<td>2</td>
<td>fall</td>
<td>COP 6XXX</td>
<td>Parallel and Cloud Computation</td>
<td>3</td>
<td>COP 5711</td>
</tr>
<tr>
<td>2</td>
<td>fall</td>
<td>6000-level</td>
<td>Elective 1 (or Project course)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>spring</td>
<td>6000-level</td>
<td>Elective 2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>spring</td>
<td>CAP 6XXX</td>
<td>Project in Data Analytics</td>
<td>3</td>
<td>(required courses)</td>
</tr>
</tbody>
</table>

In the steady state (starting in year two), courses would be offered on the following schedule. One can see that every semester three required courses are offered. In addition, three electives will be offered in the Fall and Spring semesters.
<table>
<thead>
<tr>
<th>Semester</th>
<th>Course Number</th>
<th>Course Name</th>
<th>Cr.</th>
<th>Instructor</th>
<th>Required or Elective?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall (1)</td>
<td>CNT 5805</td>
<td>Network Science</td>
<td>3</td>
<td>Mainak Chatterjee</td>
<td>Req</td>
</tr>
<tr>
<td>Fall (1)</td>
<td>STA 5206</td>
<td>Statistical Analysis</td>
<td>3</td>
<td>Daoji Li</td>
<td>Req</td>
</tr>
<tr>
<td>Fall (2)</td>
<td>COP 5XXX</td>
<td>Parallel and Cloud Computation</td>
<td>3</td>
<td>Sumit Jha</td>
<td>Req</td>
</tr>
<tr>
<td>Fall (2)</td>
<td>CAP 6307</td>
<td>Text Mining I</td>
<td>3</td>
<td>Fei Liu</td>
<td>Elec</td>
</tr>
<tr>
<td>Fall (2)</td>
<td>STA 6714</td>
<td>Data Preparation</td>
<td>3</td>
<td>Hsin-Hsiung Huang</td>
<td>Elec</td>
</tr>
<tr>
<td>Fall (2)</td>
<td>CAP 6318</td>
<td>Computational Analysis of Social Com</td>
<td>3</td>
<td>Guo-Jun Qi</td>
<td>Elec</td>
</tr>
<tr>
<td>Spring (1)</td>
<td>CAP 5610</td>
<td>Machine Learning</td>
<td>3</td>
<td>Guo-Jun Qi</td>
<td>Req</td>
</tr>
<tr>
<td>Spring (1)</td>
<td>STA 5703</td>
<td>Data Mining Methodology I</td>
<td>3</td>
<td>Daoji Li</td>
<td>Req</td>
</tr>
<tr>
<td>Spring (2)</td>
<td>CAP 6315</td>
<td>Social Media and Network Analysis</td>
<td>3</td>
<td>Gita Sukthankar</td>
<td>Elec</td>
</tr>
<tr>
<td>Spring (2)</td>
<td>COP 6XXX</td>
<td>Interactive Data Visualization</td>
<td>3</td>
<td>Sumanta Pattanaik</td>
<td>Elec</td>
</tr>
<tr>
<td>Spring (2)</td>
<td>CAP 6545</td>
<td>Machine Learning Meth. for Bio. Data</td>
<td>3</td>
<td>Haiyan (Nancy) Hu</td>
<td>Elec</td>
</tr>
<tr>
<td>Spring (2)</td>
<td>CAP 6XXX</td>
<td>Project in Data Analytics</td>
<td>3</td>
<td>Kien Hua</td>
<td>Req</td>
</tr>
<tr>
<td>Sum (1)</td>
<td>COP 5711</td>
<td>Parallel and Distrib. Database</td>
<td>3</td>
<td>Kien Hua</td>
<td>Req</td>
</tr>
<tr>
<td>Sum (1)</td>
<td>STA 6704</td>
<td>Data Mining Methodology II</td>
<td>3</td>
<td>Daoji Li</td>
<td>Req</td>
</tr>
<tr>
<td>Sum (2)</td>
<td>CAP 6XXX</td>
<td>Project in Data Analytics</td>
<td>3</td>
<td>Ivan Garibay</td>
<td>Req</td>
</tr>
</tbody>
</table>

D. Brief descriptions of courses.

**REQUIRED COURSES**

**COP 5711 Parallel and Distributed Databases.** 3(3,0). PR: COP 4710 or C.I. Storage management, implementation techniques for parallel DBMSs, distributed DBMS architectures, distributed database design, query processing, multi-database systems.

Note: CAP 5610 will admit all students in the program regardless of whether they have had CAP 4630 previously; that is, for this degree program CAP 5610 will not be taught in a way the requires knowledge of CAP 4630’s material.


**COP 6XXX Parallel and Cloud Computation.** 3(3,0) PR: COP 5711 or C.I. The course introduces students to parallel computing across the hardware-software stack. Special emphasis is placed on parallel programming using emerging architectures and technologies.

**CNT 5805 Network Science.** 3(3,0) PR: Undergraduate degree in CS, EE, or CpE. This course deals with the emerging science of complex networks and their applications. Focus will be on algorithms, mathematical theories, and computational methods that analyze complex networks and predict their behavior.

Note: if a student has taken STA 4164 (Statistical Methods III) as an undergraduate, they will be allowed to take an additional elective instead of taking STA 5206.

**STA 5206. Statistical Analysis.** 3(3,0). PR: STA 2023; not open to students who have completed STA 4164. Graduate status or senior standing or C.I. Data analysis; statistical models; estimation; tests or hypotheses; analysis of variance, covariance, and multiple comparisons; regression and nonparametric methods.

Note: for the next class, STA 5703, the undergraduate admissions requirements will meet the prerequisites.
of the next class, because programming experience will be accepted in place of STA 5104.

STA 5703. Data Mining Methodology I. 3(3,0). PR: STA 5104 and STA 5206, graduate status or senior standing, or C.I. Data mining to uncover valuable information through SEMMA (Sample, Explore, Model, Modify, and Access). Data processing with neural networks and decision trees provides the basics of data mining for uncovering knowledge.

STA 6704. Data Mining Methodology II. 3(3,0). PR: STA 5703 or C.I. Statistical techniques for data mining that include discriminant analysis, logistic regression, and factor analysis.

CAP 6XXX. Project in Data Analytics. 3(3,0). PR: COP 5711, CAP 5610, CNT 5805, STA 6704. A project-focused course that demonstrates mastery of data analytics through development of novel algorithms or innovative application of existing techniques for data mining applications.

ELECTIVE COURSES

CAP 6307 Advanced Text Mining. 3(3,0) PR: CAP 5610; or C. I. Extracting knowledge from unstructured text collections. Classification, clustering, named entity recognition, information extraction, topic modeling, summarization, programming assignments.

CAP 6315 Social Media and Network Analysis. 3(3,0) PR: CNT 5805. The course will cover:
1) Computational approaches for social network analysis;
2) Data processing and machine learning techniques for extracting information from social media datasets (e.g., Twitter).

CAP 6318 Computational Analysis of Social Complexity. 3(3,0) PR: CNT 5805. The course will cover computational concepts, principles, modeling and simulation approaches used to analyze complex social and economic phenomena, leveraging the availability of large amounts of data, and elements of complexity theory.

CAP 6XXX Interactive Data Visualization. 3(3,0) PR: COP 5711. Principles and techniques for interactive data visualization that are useful for analyzing, presenting, and exploring quantitative information. The emphasis will be on algorithmic aspects of developing interactive visualization. The students will receive practical experience in building interactive visualization systems.

Note: the next course, STA 6714 will be changed to accept programming experience in other courses in the degree program instead of STA 5104. We plan to change the catalog copy to allow consent of instructor (C.I.) as an alternative prerequisite.

E. Identification of industry-driven competencies and incorporation into the curriculum.

The department of Electrical Engineering and Computer Science at UCF had an Industrial Advisory Board, the EECS IAB. The June 2014 meeting of the EECS IAB meeting was attended by representatives from a data analytics company (Splyt), Harris Corp., Leidos, “.decimal”, Lockheed Martin, Siemens, and HD Supply. At this meeting and the board approved of the general goals and outline of the program. They agreed with the programs goals, which are stated above in section I.A, are specific competencies
that students should have after graduation. (Unfortunately, the minutes of this meeting are not adequate in recording this discussion, and so are not shown in an appendix.)

To get more focused advice from industry, Computer Science also convenes an Advisory Council. The CS Advisory Council has members from Splyt, CNL Financial Group, Lockheed Martin, Disney, Boeing, and Symantec. The CS Advisory discussed the program in March 2015 (attended by representatives from Splyt, CNL Financial Group, and Symantec), and affirmed that there was indeed a shortage of people (“talent”) in data analytics, especially of people who can obtain data, store it sensibly, and write code to extract information from data, particularly in a distributed cloud-based environment. These competencies are part of the curriculum, as reflected in the Data Preparation course, the Parallel and Distributed Databases course, and the Parallel and Cloud Computation course.

The Computer Science advisory council will continue to give advice on the program and will provide input for curriculum development.

The program itself will form its own industrial advisory council to provide regular input on the curriculum and to help with student projects and assessment.

F. Enrollment projections.

We limit each incoming cohort to 34 students enrolling each year. The number 34 was chosen as a limit on enrollment to maintain quality. Since the program is designed to serve Florida industry, we anticipate that 22 of these students will be from local industry, as well as six recent UCF bachelor’s graduates and 6 recent bachelor’s graduates from other universities. Historically, for the MS in Computer Science, the average dropout rate is 26% over the last 3 years, but only 10% in the last 2 years. For the MS in Statistical Computing the average dropout rate is about 15% over the past 3 years, but only 13% in the past 2 years. We assume that 15% will drop out of the program after the first year, leaving 30 of the original students in the program’s second year, and that this continues to be the steady state. Thus in the steady state there are 64 students in the program every year (30 second year students and 34 first year students).

The program’s curriculum is set up so that students will form a cohort in their first year of the program that all take the same required classes, and then take electives and the required project class in their second year. The required projects class will be offered in both spring and summer of the second year, in case students want to take electives that are only offered in the other semester.

Students cannot take more or less than 6 credits per semester and they are not allowed to transfer credits into the program.

IV. FACULTY PARTICIPATION

A. Existing and anticipated full-time (not visiting or adjunct) faculty.

We have 13 faculty involved in the Master of Data Analytics program. Of these one is a lecturer in Computer Science, two are assistant professors in Computer Science, five are (or will be by the time the program gets started) associate professors in Computer Science, and one is a full professor in Computer Science. In addition, two other full professors in Computer Science have contributed to the design of the degree program, but are not planned to teach in the program. From Statistics there are two assistant professors and one full professor. The director of the program is an assistant professor in the UCF Office of Research and Commercialization. All of these faculty members contribute to teaching and are members of the UCF graduate faculty.

The faculty from Computer Science are: Dr. Fei Liu, Dr. Kien A. Hua, Dr. Guo-Jun Qi, Dr. Demetrios Glinos, Dr. Mainak Chatterjee, Dr. Sumit Jha, Dr. Gita Sukthankar, Dr. Sumanta Pattanaik, and Dr. Haiyan (Nancy) Hu. The faculty from Statistics are: Dr. Daoji Li, Dr. Hsin-Shiung Huang, and Dr. Chung
Ching (Morgan) Wang. There is also one faculty member from the Institute for Simulation and Training: Dr. Ivan Garibay.

Dr. Ivan Garibay will be the interim director of the program. There will also be a program committee composed of tenured and tenure-track faculty from Computer Science and Statistics who are drawn from the above faculty teaching in the program.

B. Academic unit(s) associated with the degree.

The 32 tenured and tenure track faculty in the department of Computer Science at UCF teach over 2,400 undergraduate majors and over 300 graduate students (according to the preliminary Fall 2015 enrollment data) and in the academic year 2013-2014 they graduated over 280 undergraduate majors, 75 MS students, and 6 PhD students. The course load for research-active tenured and tenure-track faculty is 3 courses per year (with more for tenured faculty who are not research active), and the course load for the 11 non-tenure-track faculty is 8 courses per year.

In the past 3 years, average expenditures on grants and contracts were $149,206 per faculty member. The tenured faculty include five fellows of the Institute of Electrical and Electronics Engineers (IEEE), one fellow of the Association for Computing Machinery (ACM), and two fellows of the American Association for the Advancement of Science (AAAS). The faculty in Computer Science have been awarded five NSF CAREER awards in the last six years, and average six journal and conference papers per faculty per year over the last three years. The 2010 NRC rankings gave a range instead of a precise ranking; in it UCF Computer Science was about 70th among the 126 doctoral programs in Computer Science in the US.

There are currently 2 tenure-track, nine tenured faculty, and three instructors in the Statistics department. (The two tenure-track faculty were hired for 2015-16, both in Data Analytics.) These faculty teach 137 undergraduate majors and 46 graduate students, as well as thousands of non-majors in general education courses every semester. The tenured faculty teach four courses per year, the instructors teach eight and the tenure-track faculty teach three the first two years and four after that. The Statistics department has been teaching data mining classes for the past 16 years and has the distinction of being the first academic program in the US to offer a data mining MS degree. The Data Mining Track in the Department of Statistics has established strong connections with local business community. Many students have been hired by local companies such as Disney, Universal, Siemens, Florida Blue, and Everbank, etc. There are also many students who work as interns for these companies. Due to the strong position of statistics in big data analytics, the department is in the process of developing a Ph.D. program in big data analytics. Statistics has one fellow of the American Statistical Association (ASA). In the past 6 years, average grants and contracts were $21,203 per faculty member per year, and average two peer-reviewed journal publications per faculty per year over the last two years. (The UCF Statistics department was not eligible to participate in the 2010 NRC rankings, as at the time it had no doctoral program.)

V. NON-FACULTY RESOURCES

A. Library resources.

Databases: UCF Libraries’ database list is strong compared to the other institutions with the advantage of having Business Source Premier, Web of Science, and IEEE. The databases in computer sciences also offer solid resources for students in the program that include journal articles, conference proceedings, review papers, and technical reports.

Journals: The University of Central Florida Libraries’ journal listing compares favorably with the other institutions and there are some open access titles (i.e. Big Data & Society from Sage and Journal of Big Data from Springer). Since data analytics covers a broad territory in computer science and statistics, many of the publications in data analytics/big data are already occurring in established journals in those fields to which UCF already has access.

Books: Because the M.S. in Data Analytics draws from a variety of subject sources for which we are
already well established, UCF compares quite favorable with it peers. The numbers of titles within the
areas such as “big data” and “data mining” are small but that is in part due to the fact that titles are
finally beginning to emerge devoted solely to the subject area alone. Since these topics are also quite
relevant to students not only in data analytics but to students in existing programs, more titles should be
purchased from monies that already exist for those areas. UCF also has access to Springer E-Books as
well as IEEE/Wiley E-Books.

Costs: The UCF Libraries has a collection strength that is ample to support the initial start-up of the M.S.
in Data Analytics. A request for $3,000.00 per year for two years is supported by the fact that the UCF
collection is somewhat weak in Statistical Methods [this cost is built into the program’s budget]. There
are also new areas including Big Data and Data Mining in which materials are needed. Publication in
these areas is expected to rise and the library will need new funding to keep current with these and other
new areas of research. The new funding will enable the library to purchase new books to support the
program.

According to the library report (above) the library needs $3000 per year for 2 years, in order to strengthen
the library’s collection in Statistical Methods, and to increase its collection in Big Data and Data Mining.

B. Specialized equipment.

The UCF Stokes cluster, which is a free service of UCF for all research and teaching, could be used for
some work in the program, in particular the Parallel and Cloud Computing class. The plan is to use free
and open source software for instruction and research, as that is the state of the art at present. The
program plans to purchase cloud computing time for student use, at a cost of $10K per year.

The only additional specialized equipment needed is computer support, which would be provided by
cloud computing vendors, such as Amazon Web Services or Google, and rented on a demand basis from
such vendors.

VI. BENEFIT OF THE PROGRAM TO THE LOCAL COMMUNITY AND STATE

The degree program will help Orlando’s burgeoning computing industry, which already has some
startup companies in the area of data analytics. Other industries in the area, particularly health care
(Florida Hospital, etc.), the hospitality industry (Disney World, Universal Studios, etc.), and the electronic
video games industry (e.g., Electronic Arts), already make use of data analytics to increase revenues. We
anticipate that these companies will partner with UCF to provide opportunities for both research (in the
form of data or problems to solve) to faculty members, and also that they will provide contracts or grants
to better understand how to solve problems. Some existing relationships with companies are highlighted
in the support letters (see Appendix G). We will strengthen these relationships before the program gets
underway by forming an advisory board and having some meetings to discuss the program and how
companies can participate.

The benefit to the state will be in increasing the pool of qualified employees in the area of data analytics
by approximately 30 people per year. Note that this is a STEM area that is projected to quickly grow and
that can help businesses be more competitive. This will help expand the state’s high technology business
profile and will help encourage more businesses re-locate to Florida.