University of Toronto
Major Modification Proposal:
New Field or Concentration within an Existing Graduate Program

Program: Master of Science in Applied Computing (MScAC) (offered through the Department of Computing Science)

Proposed New Field / Concentration: Data Science. This proposal also closes the existing fields of study in MScAC.

Units: Computer Science & Statistical Science

Faculty: Arts & Science

1 Summary

We propose a new concentration, Data Science, as part of the current Master of Science in Applied Computing [MScAC]. The MScAC is offered through the Department of Computer Science and the new Data Science concentration is offered jointly between Computer Science and the Department of Statistical Sciences.

This proposed concentration is part of ongoing efforts in faculty recruitment and curricular development between the Departments of Computer Science and Statistical Sciences. These efforts are critical to the research and training missions of both departments in recognition of the increasing importance of the interface between our disciplines, particularly in machine learning and data analytics. These efforts are also aligned with the University’s ambitions in Computational and Data Science.

As data becomes ubiquitous and easier to acquire, particularly on a massive scale, the demand for trainees with expertise in both computer and statistical science has become acute. The proposed concentration in Data Science builds on the current MScAC infrastructure by bringing in expertise from Statistical Sciences, to meet:
1. Current student needs for master’s level education in Data Science,
2. Industrial demand for Data Scientists as interns
3. Student demand for an internship experience as a Data Scientist
4. UofT’s burgeoning research emphasis in Data Science.

Both needs and demand are acutely evident within the current MScAC program, with several students in previous years’ cohorts cobbling together courses to try to create an ad-hoc data science concentration.

When the MSc in Applied Computing was approved to commence in January 2010, the proposal identified the following 10 fields of study:

- Computational Complexity
- Applied Discrete Mathematics
- Scientific and Numerical Computation
- Artificial Intelligence
- Database and Information Systems
- Computer Graphics and Human-Computer Interaction
- Computer Systems: Hardware and Software
- Programming Languages and Methodology
- Software Engineering
- Bioinformatics

As part of this proposal, the Department of Computer Science is closing its existing fields of study within the MScAC listed above. Prior to the change to quality assurance processes in Ontario in 2011, new graduate programs approved by the Ontario Council on Graduate Studies (OCGS) were required to have at least one field, which led to the creation of these fields. They capture broadly research areas embedded in the Department, and therefore in MScAC. Students do not apply to these fields, they are not admitted or enrolled in a field, and they do not graduate with a field designation on their academic record. As such, these fields have never been created in any student systems (i.e. ROSI) and there are no students enrolled in the fields. This proposal seeks to clean up formal records and systems by removing the field offerings, which have been inactive/dormant since creation.

2 Effective Date

Effective September 1, 2017. In-progress students who have taken the required coursework will be allowed to switch to the concentration as of September 1, 2017.
3 Academic Rationale

As a discipline, statistical science emerged from various scientific domains, established foundations and developed structure through mathematics, and has evolved to become inseparable from scientific computation, machine learning and computer science. Training in Data Science, as well as the field itself, captures all aspects of this evolution and makes it experiential for the trainee. The field of Data Science has emerged as a response to our increasing, and relentless, ability to generate data, particularly on a massive scale, and our reliance upon it in nearly every facet of human endeavour. The importance of the field has led to entirely new type of professional – the Data Scientist.

A good Data Scientist requires expertise in statistical reasoning and inference; training in data management, manipulation, computation and analysis; and experience in scientific or industrial collaboration. The proposed concentration in Data Science is unique at the University of Toronto and would have no peer in Canada due to the strength of expertise in the Departments of Computer Science and Statistical Science, as well as the unrivalled wealth of potential collaborators across campus and within the Greater Toronto Area. As a new concentration in the current MScAC it is characterised by the involvement of another unit within the FAS that is responsible for a significant portion of the curriculum, as well as a shift in the current emphasis of the MScAC. The concentration bears the name of the related discipline Data Science.

The proposed concentration evolved from efforts to renew the current M.Sc. program within the DoSS to offer a sequel to the highly successful undergraduate specialist program in Applied Statistics that would be modeled on our new professional Master’s program in Financial Insurance. The Applied Statistics program requires focused study in other disciplines and a significant collaborative research project in 4th year. The current M.Sc. is an 8-month course-based Ph.D. streaming program with no significant collaborative project or internship.

Currently the Applied Statistics program does not offer a disciplinary focus in computer science and dialogue began to do just that. The result of that dialogue was for both units to begin planning an undergraduate specialist program in Data Science and a Data Science concentration within the current MScAC. The latter because the structure and outcomes of the MScAC can readily accommodate a concentration in data science, and the former because the curricular demand for training in both computer and statistical sciences cannot be met through a computer science concentration in the Applied Statistics program.

Finally, the emergence of large-scale complex data in every facet of academic and daily life has been accompanied by an increasing demand for expertise at the interface of the computational and statistical sciences – particularly machine learning. The importance of this interface has only grown in time – a trend that argues for the continuing integration of the two disciplines.
No professional statistician can possibly hope to make a meaningful contribution in collaboration or research without a serious computational skill-set. Conversely, computer scientists need far more in-depth statistical training to fully understand the behaviour and impact of the tools and algorithms they invent. Our expectation is that the proposed concentration will serve to meet the above demand, further integrate the activity of the two partnering Departments, elevate the quality of training within both units particularly the M.Sc. within the DoSS and the MScAC within CS, and align with the University’s ambitions in Data Science.

4 Need and Demand

Our media constantly reminds us of the emergence of large-scale complex data in nearly every facet of daily life. These reports speak to the increasing demand for expertise at the interface of computer and statistical science as well as the increasing opportunities for employment.

Student interest in Data Science naturally flows from this and both this, and industrial demand for Data Scientists, are currently evident in the MScAC program. The following is a partial list of employers that have engaged our students in data science internships: Addictive Mobility, SmartFinance, Rakuten Kobo, NeuRecall, Geotab, Google, Ipsos, Wattpad, VerticalScope, Amazon, Scotiabank, PwC, and Facebook.

This is not a fleeting trend. There has been a massive increase in the amount of data available from new technologies that seem to be emerging on a daily basis. New data sources, such as network data, image data, streaming data, are all a part of a trend that is set to intensify. So too will the need and demand for Data Scientists.

### Table 1: Graduate Enrolment Projections

<table>
<thead>
<tr>
<th>Year in program</th>
<th>Academic Year 2016-17</th>
<th>Academic year 2017-18</th>
<th>Academic year 2018-19</th>
<th>Academic year 2019-20</th>
<th>Academic year 2020-21*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total #</td>
<td># in concentration</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>5</td>
<td>50</td>
<td>10</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>0</td>
<td>40</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>5</td>
<td>90</td>
<td>15</td>
<td>110</td>
</tr>
</tbody>
</table>
These projections are commensurate with known demand. For example, the financial technology (FinTech) sector in Toronto has seen a rapid increase in demand for Data Scientists in recent years. In 2016, a total of 25 Data Science FinTech internship positions were posted to the MScAC cohort, and only 14 were filled due to lack of available students. FinTech is not the only source of demand for Data Science industry positions, it is just the area of most significant growth. Over the last five years, more than half of MScAC internships involved Data Science.

Student demand for Data Science, Machine Learning, and Data Analytics courses and internships also continues to increase.

5 Admission Requirements

Students entering the Data Science Concentration of the MScAC program at the University of Toronto will register in the Department of Computer Science. The minimum admission requirements listed below are consistent with those criteria in Department of Statistical Sciences MSc program, and are identical to the Department of Computer Sciences MScAC program.

Minimum Admission Requirements

Applicants are admitted under the General Regulations of the School of Graduate Studies. Applicants must also satisfy the graduate units’ additional admission requirements stated below.

Admission to the MScAC program is on a competitive basis. For an applicant to be considered in the Data Science concentration, they must have:

- A Bachelor’s degree from a recognized university in a related field such as Statistics, Computer Science, Mathematics, or any discipline where there is a significant quantitative component. Studies must include significant exposure to Statistics, Computer Science and Mathematics, including coursework in advanced calculus, linear algebra, probability and statistics, programming languages and computational methods.
- An average grade equivalent of at least a University of Toronto B+ in the final two years, as defined in the SGS Admissions Manual.
• Three letters of reference.
• A letter of intent (limited to 500 words) explaining their interest in data science and objectives for the program.
• Applicants whose primary language is not English and who have graduated from a university where the language of instruction and examination is not English must demonstrate facility in English using one of the official methods as outlined in the Calendar of the School of Graduate Studies.

Admission to the Data Science Concentration is competitive. Achievement of the minimum standards does not guarantee admission into the program. Those accepted will normally have achieved a standing considerably higher than the minimum B+ standing or have demonstrated exceptional ability through appropriate workplace experience.

6 Program Requirements

The program requirements for the proposed Data Science concentration will be:

• 1.0 FCE chosen from the STA 2000-level or higher. This may include a maximum of 0.5 FCE chosen from the STA 4500-level of six week (quarter) courses.
• 1.0 FCE chosen from the Computer Science (CSC course designator) graduate course listings.
• 1.0 FCEs required courses in technical communications (CSC 2701H) and technical entrepreneurship (CSC 2702H)
• Students must also complete an eight-month industrial internship. The internship CSC 2703H (3.5 FCEs) is coordinated by the department, and evaluated on a pass/fail basis.
• Students in the DS concentration may choose a supervisor from either the Department of Statistical Sciences or the Department of Computer Science. It is also possible for a joint supervision of faculty members from both departments.

The Data Science concentration program requirements follow the structure of the existing MScAC program. The existing program and new concentration are designed as a 16-month (4 session, F/W/S/F) full-time program comprised of 4 half courses (2.0 FCEs) that will be completed in 8-months (2 sessions), 2 required courses in technical communications and entrepreneurship (1.0 FCE) and an 8-month (2 session) industrial internship (3.5 FCE).

Students in the new Data Science concentration, like those in the existing MScAC program, follow a course of study that is fully integrated; course projects and assignments will be designed to integrate the material learned from a variety of the courses and to utilize it in a practical context. Excellent communication and presentation skills will be emphasized in both the oral and written components of the projects and assignments. The program contains an 8-month internship component from May – Dec. The students will enter the internship immediately after coursework is completed at the end of the Winter term.
The required course work can be completed during the regular academic year. The course load of two half courses per session is identical to that of the current MScAC program in the Department of Computer Science.

In addition to the 4 half courses taken during the first 2 sessions, an additional two specialized half courses (in Technical Communication, and in Technical Entrepreneurship) are required. One of these courses is taken during the first eight months of the student’s term, while the other is taken during the eight-month internship period.

Whereas the Province’s Quality Assurance Framework requires that students complete a minimum of 2/3 courses at the graduate level, the University of Toronto requires graduate students to complete all of their course requirements from amongst graduate level courses. This proposed Data Science Concentration complies with this requirement.

The creation of the Data Science concentration does not require the creation of any new courses at the graduate level. Students enrolled in the Data Science concentration will select existing courses from the two departments, DoSS and DCS, and receive approval of their choices from each graduate unit and/or the director of the program.

Among the total 2.0 FCEs, one FCE will come from the STA 2000-level or higher, with a maximum of 1 HCE (0.5 FCE) coming from the STA4500-level six week (quarter) courses. The other 1.0 FCE will be selected from the CSC graduate course listings.

Within the MScAC program support for establishing and maintaining industrial partners for internships is key to attracting students. The internship provides a critical experiential learning component, and helps students improve their communication skills. Students will not only gain practical experience in knowledge and technology transfer, but will also have access to well-trained professional support staff in their host company to realize their vision and make further connections in industry. An internship will be required for all students in Data Science Concentration.

Our current relationship with Mathematics and Information Technology and Complex Systems (MITACS) indicate that these internships should qualify for partial funding under the MITACS ACCELERATE CANADA program, which would offset $15,000 of the employer’s costs with a current average total compensation of $51,500 per student over eight months. This would make interns even more attractive (making this opportunity consistent with the current MScAC program).

For academic supervision, students in the DS concentration may choose a supervisor from either DoSS or DCS faculty. It is also possible for a joint supervision of faculty members from both departments. The selection of an appropriate academic supervisor is facilitated by the program director once the student has accepted a qualifying internship placement.

All students in the MScAC, including students in the new concentration, receive individualized advising to ensure that they select courses that a) meet the program requirements, including any requirements specific to the concentration; b) have sufficient academic preparation for
each course; and c) support their professional goals.

Students currently completing the general requirements for the MScAC as of September 2017 will be allowed to opt-in to the new concentration if they are eligible.

Please see Appendix A for the proposed calendar entry.

Please see Appendix B for a list of suggested course numbers and titles. Note that these are existing courses offered by the participating departments, and specifically that none of these courses were created as a result of this proposed concentration.

### 7 Degree Level Expectations, Program Learning Outcomes and Program Structure

<table>
<thead>
<tr>
<th>MASTER’S DEGREE LEVEL EXPECTATIONS (based on the Ontario Council of Academic Vice Presidents (OCAV) DLEs)</th>
<th>MASTER’S PROGRAM LEARNING OBJECTIVES AND OUTCOMES</th>
<th>HOW THE PROGRAM DESIGN AND REQUIREMENT ELEMENTS SUPPORT THE ATTAINMENT OF STUDENT LEARNING OUTCOMES</th>
</tr>
</thead>
</table>

### EXPECTATIONS:

This Data Science (DS) concentration in the Master of Science in Applied Computing is awarded to students who have demonstrated:

1. **Depth and Breadth of Knowledge**

   A systematic understanding of knowledge, and a critical awareness of current problems and/or new insights, much of which is at, or informed by, the forefront of the academic discipline, field of study, or area of professional practice.

   Depth and breadth of knowledge is defined in DS Concentration of MScAC program as an ability to assimilate and distill complex data and models into informed decisions.

   This is reflected in students who are able to:
   - Explore, manipulate and visualize large-scale complex and massive data.
   - Develop statistical and learning models and methods using current software infrastructure to analyze data arising from science, business and other activities.
   - Understand latent

   The program design and requirement elements that ensure these student outcomes for depth and breadth of knowledge are from the 1.0 FCE chosen from the STA2000-level or higher (max. 0.5 FCE at STA4500-level) and 1.0 FCE chosen from the Computer Science (CSC course designator) graduate course listings. Relevant courses could include:

   - STA2080 - Fundamentals of Statistical Genetics
   - STA2101 - Methods of Applied Statistics I
   - STA2102 - Computational Techniques in Statistics
   - STA2104 - Statistical Methods for Machine Learning and DataMining
   - STA2201 - Methods of Applied
### MASTER’S DEGREE LEVEL EXPECTATIONS (based on the Ontario Council of Academic Vice Presidents (OCAV) DLEs)

**MASTER’S PROGRAM LEARNING OBJECTIVES AND OUTCOMES**

<table>
<thead>
<tr>
<th>How the Program Design and Requirement Elements Support the Attainment of Student Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>relationships and patterns obtained from applying statistical methods or machine learning algorithms to complex data.</td>
</tr>
</tbody>
</table>
| Statistics II  
STA2453 - Statistical Consulting  
STA4273 - Large Scale Machine Learning  
STA4501 - Functional Data Analysis and Related Topics  
STA4507 - Extreme Value Theory and Applications  
STA4515 - Modelling and Analysis of Spatially Correlated Data  
STA4516 - Topics in Probabilistic Programming  
CSC2508 - Advance Data Management Systems  
CSC2525 - Evaluating Data Curation  
CSC2541 - Topics in Machine Learning  
CSC2542 - Topics in Knowledge Representation & Reasoning  
CSC2545 - Kernel Methods & Support Vector Machines  
CSC2515 - Machine Learning  
CSC2501 - Computational Linguistics  
CSC2511 - Natural Language Computing  
CSC2506 - Uncertainty & Learning  
CSC2502 - Knowledge Representation & Reasoning |

### 2. Research and Scholarship

**A conceptual understanding and methodological competence that i) Enables a working comprehension of how established techniques of research and inquiry are used to create and interpret knowledge in the discipline; ii) Enables a critical understanding of the challenges and opportunities within research and scholarship in the discipline.**

Research and Scholarship is defined in in Concentration in DS Concentration of MScAC program as case study based on real-world industry specific experience.

This is reflected in students who are able to:

- Define and describe a variety

The program design and requirement elements that ensure these student outcomes for research and scholarship are from the 1.0 FCE chosen from the STA 2000-level or higher (max. 0.5 FCE at STA4500-level) and 1.0 FCE chosen from the Computer Science (CSC course designator) graduate course listings. Relevant courses include:

<table>
<thead>
<tr>
<th>Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
STA2453 - Statistical Consulting  
STA4273 - Large Scale Machine Learning  
STA4501 - Functional Data Analysis and Related Topics  
STA4507 - Extreme Value Theory and Applications  
STA4515 - Modelling and Analysis of Spatially Correlated Data  
STA4516 - Topics in Probabilistic Programming  
CSC2508 - Advance Data Management Systems  
CSC2525 - Evaluating Data Curation  
CSC2541 - Topics in Machine Learning  
CSC2542 - Topics in Knowledge Representation & Reasoning  
CSC2545 - Kernel Methods & Support Vector Machines  
CSC2515 - Machine Learning  
CSC2501 - Computational Linguistics  
CSC2511 - Natural Language Computing  
CSC2506 - Uncertainty & Learning  
CSC2502 - Knowledge Representation & Reasoning |
### MASTER’S DEGREE LEVEL EXPECTATIONS (based on the Ontario Council of Academic Vice Presidents (OCAV) DLEs)

- evaluation of current research and advanced research and scholarship in the discipline or area of professional competence; and
- Enables a treatment of complex issues and judgments based on established principles and techniques; and, on the basis of that competence, has shown at least one of the following:
  - The development and support of a sustained argument in written form; or
  - Originality in the application of knowledge.

### MASTER’S PROGRAM LEARNING OBJECTIVES AND OUTCOMES

- Be able to place the analysis of results into the broader context of enterprise wide data analysis and algorithm development.
- Formulate statistical models and learning methods, spanning existing techniques and algorithms, which are tailored to new problems and applications.
- Communicate models and their analysis to non-experts verbally and in written form.

### HOW THE PROGRAM DESIGN AND REQUIREMENT ELEMENTS SUPPORT THE ATTAINMENT OF STUDENT LEARNING OUTCOMES

- could include:
  - STA2080 - Fundamentals of Statistical Genetics
  - STA2101 - Methods of Applied Statistics I
  - STA2102 - Computational Techniques in Statistics
  - STA2104 - Statistical Methods for Machine Learning and Data Mining
  - STA2201 - Methods of Applied Statistics II
  - STA2453 - Statistical Consulting
  - STA4273 - Large Scale Machine Learning
  - STA4501 - Functional Data Analysis and Related Topics
  - STA4507 - Extreme Value Theory and Applications
  - STA4515 - Modelling and Analysis of Spatially Correlated Data
  - STA4516 - Topics in Probabilistic Programming
  - CSC2508 - Advance Data Management Systems
  - CSC2525 - Evaluating Data Curation
  - CSC2541 - Topics in Machine Learning
  - CSC2542 - Topics in Knowledge Representation & Reasoning
  - CSC2545 - Kernel Methods & Support Vector Machines
  - CSC2515 - Machine Learning
  - CSC2501 - Computational Linguistics
  - CSC2511 - Natural Language Computing
  - CSC2506 - Uncertainty & Learning
  - CSC2502 - Knowledge
## 3. Level of Application of Knowledge

*Competence in the research process by applying an existing body of knowledge in the critical analysis of a new question or of a specific problem or issue in a new setting.*

**Application of Knowledge** is defined in DS Concentration of MScAC program as competence in applying an existing body of knowledge in the critical analysis of a new question or of a specific problem or issue in a new setting.

This is reflected in students who are able to:
- Demonstrate expertise in data analysis involving large-scale, complex structures and massive sizes.
- Critically and comprehensively assess problems arising from a variety of application contexts.
- Provide quantitative solutions to a new question or a specific problem in a new setting, with statistical thinking and reasoning.

### Industrial Internship:
The main goal of their internship experience, students will be required to apply their knowledge of statistics and machine learning to the real-world problems in an industrial setting.

## 4. Professional Capacity/Autonomy

**Professional Capacity/Autonomy** is defined in DS Concentration of MScAC program as the qualities and transferable skills necessary for employment requiring:

- The exercise of initiative and of personal responsibility and accountability; and
- Decision-making in complex situations;
- The intellectual independence required for continuing professional development;
- The ethical behaviour consistent with academic integrity and the use of appropriate guidelines and procedures for responsible conduct of research; and
- The ability to appreciate the broader implications of applying knowledge to particular

### Industrial Internship:
The exploration of new or specific problems is coupled with the students’ quantitative skills and statistical rationale, with the aim of solving practical issues in an environment of complex and/or massive datasets.

In addition, as outlined in section 6, there is a special course on technical entrepreneurship and business, CSC2702 – Technical.
<table>
<thead>
<tr>
<th>MASTER’S DEGREE LEVEL EXPECTATIONS (based on the Ontario Council of Academic Vice Presidents (OCAV) DLEs)</th>
<th>MASTER’S PROGRAM LEARNING OBJECTIVES AND OUTCOMES</th>
<th>HOW THE PROGRAM DESIGN AND REQUIREMENT ELEMENTS SUPPORT THE ATTAINMENT OF STUDENT LEARNING OUTCOMES</th>
</tr>
</thead>
</table>
| contexts. | implications of applying knowledge to particular contexts. This is reflected in students who are able to:  
- Prepare written reports and deliver oral presentations to expert (quantitative teams) and non-expert audiences (upper management)  
- Provide a holistic perspective on data analysis, statistical or learning method development for industry problems in specific settings. | Entrepreneurship. This is a required course. |

5. Level of Communications Skills

The ability to communicate ideas, issues and conclusions clearly.

Communications Skills is defined in DS Concentration of MScAC program as the ability to communicate ideas, issues and conclusions clearly.

This is reflected in students who are able to:

- Construct a credible argument and present it in appropriate formats
- Construct detailed research reports and executive summaries
- Deliver professional presentations to expert (quantitative teams) and non-expert audiences (upper management)

The program design and requirement elements that ensure these student outcomes for level of communication skills are:

The required written report on the internship experience is designed for students to connect their course work with their industrial experience. The oral presentation in front of faculty, industry experts and students will require students to discuss and critically assess their success at applying their academic knowledge to specific problems they encountered in their internship.

In addition, as outlined in section 6, there is a special course on technical communications skills; CSC2701 – Communication for Computer Scientists. This is a required course.
8 Assessment of Teaching and Learning

Student performance in the program will be assessed through a variety of methods including reports, presentations, assignments, case studies, and exams. Students will receive letter grades for their performance in all courses except that CR/NCR is given for their written and oral reports on industrial internship.

<table>
<thead>
<tr>
<th>Teaching and Learning Outcomes</th>
<th>Assignments</th>
<th>Projects</th>
<th>Exam</th>
<th>Presentations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Depth and Breadth of Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Display expertise in statistical methods and machine learning algorithms</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Critically assess a problem that is complex and has alternative design approaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjust communications to address different audiences</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Identify key debates that result from conflicting practitioner/scientists/business views</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>2. Scholarship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptualize, design, implement a statistics/machine learning project</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Make informed judgments on complex issues in the context of complex data analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articulate those strategies and judgments</td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>3. Application of Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assess a complex problem from the viewpoints of practitioners/scientists/business</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>4. Professional Capacity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete the degree requirements in a timely manner</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Demonstrate project management skills</td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>5. Communication Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate complex ideas effectively</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Prepare reports and presentations that outline the problem, option and solutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
9 Consultation

The proposed data science concentration is the result of a lengthy collaboration between the Departments of Computer Science and Statistical Sciences. In addition, there has been, to varying degrees, some consultation with the Office of the Vice President Research concerning the University’s plans to establish an institute for computational and data sciences, the Dean’s Offices and science Chairs within the FAS. The focus of the latter is to identify data intensive courses and activities within these units with the expectation that these consultations will be extended to each unit within the FAS [Social Sciences and Humanities].

10 Resources

As the program grows, there will be resource implications on various fronts. In terms of faculty complement, we anticipate growing faculty complement by several FTEs in the Data Science area (see below). Space needs will be minimal in the initial years. Administrative staffing will be handled by the MScAC program, and will be funded through revenues from the concentration itself, so no new central funds will be required as this is intended to be a self-sustaining concentration, within the existing MScAC program, and with a formal agreement between Computer Science and Statistical Sciences.

10.1 Faculty Complement

The concentration will require faculty expertise in various areas of data science. This includes statistical machine learning, data analytics, data visualization, usability, knowledge representation. Both DoSS and DCS are currently in the process of hiring several FTEs in the area of machine learning. In the coming years, given the departments’ and university’s significant focus on data science, we expect to be hiring additional faculty in these data science related areas, which will sync with the expected growth rate of the proposed concentration.
### Table 3: Detailed Listing of Committed Faculty

<table>
<thead>
<tr>
<th>Faculty name and rank</th>
<th>Home unit</th>
<th>Area(s) of Specialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radu Craiu, Professor</td>
<td>Statistical Sciences</td>
<td>MCMC, Statistical Computing, Genetics, Copula Modelling</td>
</tr>
<tr>
<td>Alison Gibbs, Associate Professor, Teaching Stream</td>
<td>Statistical Sciences</td>
<td>Statistical Education, Consulting</td>
</tr>
<tr>
<td>Keith Knight, Professor</td>
<td>Statistical Sciences</td>
<td>Robustness, Time Series</td>
</tr>
<tr>
<td>Radford Neal, Professor</td>
<td>Computer Science &amp; Statistical Sciences</td>
<td>Machine Learning, MCMC, Neural Networks</td>
</tr>
<tr>
<td>Nancy Reid, University Professor</td>
<td>Statistical Sciences</td>
<td>Likelihood, Experimental Design</td>
</tr>
<tr>
<td>Jeffrey Rosenthal, Professor</td>
<td>Statistical Sciences</td>
<td>MCMC, Stochastic Processes</td>
</tr>
<tr>
<td>Daniel Roy, Assistant Professor</td>
<td>Statistical Sciences</td>
<td>Machine Learning, Probabilistic Programming</td>
</tr>
<tr>
<td>Ruslan Salakhutdinov, Assistant Professor</td>
<td>Computer Science &amp; Statistical Sciences</td>
<td>Statistical Machine Learning, Deep Learning</td>
</tr>
<tr>
<td>Jamie Stafford, Professor &amp; Chair</td>
<td>Statistical Sciences</td>
<td>Disease Mapping, Spatial Temporal Modelling</td>
</tr>
<tr>
<td>Lei Sun, Professor &amp; Associate Chair Research</td>
<td>Statistical Sciences</td>
<td>Statistical Genetics, False Discovery, Multiple Testing</td>
</tr>
<tr>
<td>Nathan Taback, Assistant Professor, Teaching Stream</td>
<td>Statistical Sciences</td>
<td>Statistical Education, Consulting</td>
</tr>
<tr>
<td>Fang Yao, Professor &amp; Associate Chair Graduate</td>
<td>Statistical Sciences</td>
<td>Functional Data Analysis, Nonparametric Regression</td>
</tr>
<tr>
<td>Zhou Zhou, Associate Professor</td>
<td>Statistical Sciences</td>
<td>Functional Data Analysis, Time Series</td>
</tr>
<tr>
<td>Allan Borodin, University Professor</td>
<td>Computer Science</td>
<td>Mathematical foundations of Computer Science</td>
</tr>
<tr>
<td>Faculty name and rank</td>
<td>Home unit</td>
<td>Area(s) of Specialization</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Eugene Fiume, Professor</td>
<td>Computer Science</td>
<td>Software systems and parallel algorithms, with a special focus on modelling natural phenomena</td>
</tr>
<tr>
<td>Matt Medland, Assistant Professor, Teaching Stream &amp; Director, Professional Programs &amp; External Relations</td>
<td>Computer Science</td>
<td>Computer Science Education, Software Engineering</td>
</tr>
<tr>
<td>Khai Truong, Professor &amp; Associate Chair Research</td>
<td>Computer Science</td>
<td>Ubiquitous computing (ubicomp) and human-computer interaction (HCI)</td>
</tr>
<tr>
<td>Richard Zemel, Professor</td>
<td>Computer Science</td>
<td>Machine learning methods, with a specific focus on unsupervised learning, and probabilistic models of neural representations</td>
</tr>
<tr>
<td>Ravin Balakrishnan, Professor &amp; Chair</td>
<td>Computer Science</td>
<td>Human-computer interaction (HCI) with special focus on methods for interacting with and visualizing large amounts of data</td>
</tr>
<tr>
<td>Anna Goldenberg, Assistant Professor</td>
<td>Computer Science</td>
<td>Machine learning methods to decipher human disease heterogeneity</td>
</tr>
</tbody>
</table>

### 10.2 Space/Infrastructure

Students in the concentration will be provided office space in the Bahen Centre, alongside the current MScAC students. IT support is provided by a 0.25FTE IT staff member in the Department.
of Computer Science. There is no lab space or specialized equipment requirement for the program.

## 11 UTQAP Process

<table>
<thead>
<tr>
<th>Steps</th>
<th>Approval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development/consultation within Unit</td>
<td>Fall 2015/Winter 2016</td>
</tr>
<tr>
<td>Consultation with Dean’s Office (&amp; VP,AP)</td>
<td>Winter/Spring/Fall 2016</td>
</tr>
<tr>
<td>September 16, 2015</td>
<td>Graduate unit approval as appropriate</td>
</tr>
<tr>
<td>October 6 and October 12, 2016</td>
<td>Faculty/Divisional Council</td>
</tr>
<tr>
<td>Report to AP&amp;P</td>
<td>Anticipated spring 2017</td>
</tr>
<tr>
<td>Report to Ontario Quality Council</td>
<td>Anticipated spring 2017</td>
</tr>
</tbody>
</table>
Appendix A: Calendar Entry

2015-16 SGS Calendar
Computer Science
Faculty Affiliation
Arts and Science

Degree Programs

Applied Computing

<table>
<thead>
<tr>
<th>MScAC</th>
<th>Concentrations:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data Science</td>
</tr>
</tbody>
</table>

Computer Science

<table>
<thead>
<tr>
<th>MSc</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhD</td>
</tr>
</tbody>
</table>

Collaborative Programs

The following collaborative programs are available to students in participating degree programs as listed below:

1. Genome Biology and Bioinformatics
   - Computer Science, PhD
2. Knowledge Media Design
   - Computer Science, MSc, PhD
3. Neuroscience
   - Computer Science, MSc, PhD

Overview

The Department of Computer Science offers a graduate program leading to three degrees: Master of Science, Master of Science in Applied Computing, and Doctor of Philosophy. The program consists of courses and either research (MSc and PhD) or practicum (MScAC), both of which are conducted under the supervision of a faculty member.

Graduate faculty in the Department of Computer Science are interested in a wide range of subjects related to computing, including programming languages and methodology, software engineering, operating systems, compilers, distributed computation, networks, numerical analysis and scientific computing, financial computation, data structures, algorithm design and analysis, computational complexity, cryptography, combinatorics, graph theory, artificial intelligence, neural networks, knowledge representation, computational linguistics, computer vision, robotics, database systems, graphics, animation, interactive computing, and human-computer interaction.

For further details, consult the graduate student handbook prepared by the department and available online.
Contact and Address

Web: www.cs.toronto.edu
Email: gradadmissions@cs.toronto.edu
Telephone: (416) 978-8762
Fax: (416) 946-1932

Department of Computer Science Graduate Office
University of Toronto
Room 4242, Bahen Centre for Information Technology
40 St. George Street
Toronto, Ontario M5S 2E4
Canada

Degree Programs

Applied Computing

Master of Science in Applied Computing

Minimum Admission Requirements

- Applicants are admitted under the General Regulations of the School of Graduate Studies. Applicants must also satisfy the Department of Computer Science’s additional admission requirements stated below.
- An appropriate bachelor’s degree in computer science for applicants to the general program option.
- A minimum average grade of B+ over the final two years of undergraduate studies.
- Applicants whose primary language is not English and who have graduated from a university where the primary language of instruction is not English must achieve a Test of English as a Foreign Language (TOEFL) score of at least 580 on the paper-based test and 4 on the Test of Written English (TWE); 93/120 on the Internet-based test and 22/30 on the writing and speaking sections.
- Three letters of support from faculty and/or employers.
- A statement of purpose.

Applicants to the Data Science concentration (offered jointly by the Department of Computer Science and the Department of Statistical Sciences), must fulfil the following additional admission requirements:

- Applicants must indicate a preference for a concentration in Data Science in their application. Admission to the Data Science concentration is on a competitive basis. Students admitted to the MScAC program are not automatically admitted to the Data Science concentration upon request.
- An appropriate bachelor’s degree from a recognized university in a related field such as Statistics, Computer Science, Mathematics, or any discipline where there is a significant quantitative component. The bachelor’s degree completed must include significant exposure to Statistics, Computer Science and Mathematics, including coursework in advanced calculus, linear algebra, probability and statistics, programming languages and computational methods.

Program Requirements for the General Program (No Concentration)

- This is a 16-month professional master’s program comprising of:
  - 3.0 full-course equivalents (FCEs) of coursework including 1.0 FCEs required courses in technical communications (CSC 2701H, 0.5 FCE) and technical entrepreneurship (CSC 2702H, 0.5 FCE)
  - and an eight-month industrial internship. The internship CSC 2703H (3.5 FCE) is coordinated by the department, and evaluation on a pass/fail basis.
- There is no thesis requirement.

Program Requirements for the Data Science Concentration

- Students admitted to the Data Science concentration are required to complete the 3.0 (FCEs) in coursework as follows:
1.0 FCE chosen from the STA 2000-level or higher. This may include a maximum of 0.5 FCE chosen from the STA 4500-level of six week (quarter) courses.

1.0 FCE chosen from the Computer Science (CSC course designator) graduate course listings.

1.0 FCEs required courses in Technical Communications (CSC 2701H, 0.5 FCE) and Technical Entrepreneurship (CSC 2702H, 0.5 FCE)

Course selections should be made in consultation with the Program Director.

Students enrolled in the Data Science Concentration must also complete an eight-month industrial internship. The internship CSC 2703H (3.5 FCEs) is coordinated by the department, and evaluated on a pass/fail basis.

There is no thesis requirement.

Program Length
4 sessions full-time (typical registration sequence: F/W/S/F)

Time Limit
3 years full-time
Appendix B: List of Courses associated with the new concentration

All students in the MScAC, including students in the new concentration, receive individualized advising to ensure that they select courses that a) meet the program requirements, including any requirements specific to the concentration; b) have sufficient academic preparation for each course; and c) support their professional goals.

Students pursuing the new concentration in data science may select graduate-level courses from the participating departments, in compliance with the requirements listed in the SGS calendar, and subject to approval of the program director. Of eligible courses, the following are examples that are particularly relevant to the data science concentration:

- STA2080 - Fundamentals of Statistical Genetics
- STA2101 - Methods of Applied Statistics I
- STA2102 - Computational Techniques in Statistics
- STA2104 - Statistical Methods for Machine Learning and DataMining
- STA2201 - Methods of Applied Statistics II
- STA2453 - Statistical Consulting
- STA4273 - Large Scale Machine Learning
- STA4501 - Functional Data Analysis and Related Topics
- STA4507 - Extreme Value Theory and Applications
- STA4515 - Modelling and Analysis of Spatially Correlated Data
- STA4516 - Topics in Probabilistic Programming
- CSC2508 - Advance Data Management Systems
- CSC2525 - Evaluating Data Curation
- CSC2541 - Topics in Machine Learning
- CSC2542 - Topics in Knowledge Representation & Reasoning
- CSC2545 - Kernel Methods & Support Vector Machines
- CSC2515 - Machine Learning
- CSC2501 - Computational Linguistics
- CSC2511 - Natural Language Computing
- CSC2506 - Uncertainty & Learning
- CSC2502 - Knowledge Representation & Reasoning