SFU Dean of Graduate Studies

Maggie Benston Student Services Centre 1100
8888 University Drive
Burnaby, BC
Canada V5A 1S6

TEL 778.782.3042
FAX 778.782.3080

report-dgs@sfu.ca
www.sfu.ca/Dean-GradStudies

MEMORANDUM

ATTENTION	Senate	DATE	8 January 2014
FROM	Wade Parkhouse, Dean of Graduate Studies	No.	GS2014.01
RE:	Proposal for a Professional Master's Program in Big Data under Special Arrangements		

For information:

At its meeting of 6 January 2014, SGSC approved the Proposal for a Professional Master's Program in Big Data under Special Arrangements.

Effective Date: September 2014

Faculty of Applied Science **[GS2014.01]**
School of Computing Science
Proposal for a Professional Master's Program in Big Data under Special Arrangements

New courses:
CMPT 731-6 Programming for Big Data 1
CMPT 732-6 Programming for Big Data II

Program proposal for a Professional Masters Program in Big Data

(Cohort Special Arrangements Program)

School of Computing Science
Faculty of Applied Science
Simon Fraser University

Executive Summary

The School of Computing Science wishes to propose the creation of a professional master's degree program in Big Data. This program will initially be offered as a cohort special arrangements master's program. Our plan is to prototype this program with the objective of converting it to a regular program after it has been run for 3 years under special arrangements.

Big Data is a fast growing area of computer science that has transformed the information technology industry. Highly qualified personnel who are trained in computational methods dealing with Big Data are in great demand, and we feel that we can offer a program that will train students to fill this need. Additional information about the field of Big Data and the timeliness of this proposal are included later in this document.

Our goal is to train computational specialists who can construct models, develop algorithms and write software that can extract actionable knowledge from Big Data. The graduates from this program will know how to handle, analyze, and visualize massive data sets; they will be able to ask feasible questions and answer these questions efficiently using computational methods. Our intended audience will be our own undergraduates, international students and professionals who wish to advance their knowledge in this upcoming and lucrative area of information technology.

The program tuition will be $22,437 and students will normally complete the program in 4 semesters. The per-term fee for the program will be $5610. There will be a continuing fee per term, for those students who take longer than 4 terms, of $2805, which is one half of the regular fee. Students will complete 30 units of graduate work: 15 credits of graduate course work that is relevant to Big Data analysis; 12 credits of specialized lab work; and 3 credits of co-op. The 15 credits of coursework will come from graduate courses already offered in our graduate program. The 12 credits of specialized lab work will provide hands on experience with Big Data. The high quality lab coursework, the exclusive lab facilities and the small cohort of students justify our premium fees for these credits. There will also be a significant experiential component, with a co-op term in industry as an integral part of this program.

Curriculum

Program novelty

Big Data is a term that has been recently coined to cover the computational treatment of the massive datasets that are being collected in all areas of our society. The world's technological per-capita capacity to store information has approximately doubled every 40 months since the 1980s[1] and as of 2012, each day 2.5 exabytes (2.5×10^{18}) of data are being created. Our society is increasingly dependent on efficient analysis of this massive volume of data to support decision making. The proposed program will focus on the capture, curation, storage, search, sharing, transfer, analysis, and visualization of Big Data.

Big Data may be the result of scientific experimentation (e.g. the Large Hadron Collider at CERN), the Internet (e.g. links between web pages and the content on the web), social networking (e.g. Tweets, Facebook posts), sensor networks (e.g. elderly patients wearing heart-rate monitors), and there are a multitude of other examples. New sources of Big Data are appearing at an increasing pace due to our increasingly connected technological society. Today one can spot business trends, detect environmental changes, predict forthcoming social change, and combat crime, etc. by analyzing Big Data.

Big Data comes with novel computational challenges and requires a novel set of skills in computer scientists. The scale of Big Data is currently measured in exabytes or 10^{18} bytes. When data reaches such a scale both storage and computation requires a massively distributed architecture (e.g. Google is the most famous Big Data company; Facebook has a 100 petabyte storage cluster used to store and search for user information). Because of the scale of Big Data and the challenges of Big Data analysis, computer scientists require special technical skills. They need expertise in formulating algorithms that can scale to massive data and the ability to write software for the massively distributed systems that are currently used to store and analyze Big Data.

Big Data is a brand-new, emerging discipline. We are aware of only six related master's programs (all in the US). We hope to be the first comprehensive master's program in Big Data in Canada. We also aim to be one of the most competitive in terms of experiential learning among similar programs in North America. These programs are:

- Northwestern University: Master of Analytics Program
- North Carolina State University: Institute for Advanced Analytics
- Carnegie Mellon University: Very Large Information Systems
- New York University: master's in Data Science
- DePaul University: master's in Predictive Analytics
- Worcester Polytechnic Institute: master's in Data Science

Some additional details and a comparison with these programs are provided in Appendix VI.

Curriculum Details

---

[1] Hilbert, Martin; López, Priscila (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information". Science 332 (6025): 60–65.

The program curriculum will cover the following areas:

- Analysis of scalability of algorithms to big data.
- Data warehouses and online analytical processing.
- Efficient storage of big data including data streams.
- Scalable querying and reporting on massive data sets.
- Scalable and distributed hardware and software architectures.
- Software as a service. Cloud Computing (e.g. Amazon EC2, Google Compute Engine)
- Big data programming models: map-reduce, distributed databases, software for implementing streaming and sketching algorithms.
- Dealing with unstructured data such as images, text or biological sequences.
- Scalable machine learning methods such as online learning.
- Data mining: methods for learning descriptive and predictive models from data.
- Distributed algorithms over very large graphs and matrices.
- Social media analysis.
- Visualization methods and interactive data exploration.

Students will complete 30 units of graduate work. These units are divided into three main sections: 15 credits of graduate course work; 12 credits of specialized lab work; 3 credits for co-op.

Course work

Five 3 credit courses must be taken from the choices below:

- One of the following two courses:
    - CMPT 881 Algorithms for Massive Data. (3 cr) Spring 2013 Instructor: Funda Ergun
    - CMPT 705 Design and Analysis of Algorithms (3cr; offered every semester by several faculty in our School)
- CMPT 886 Systems for Big Data. (3 cr) Spring 2012 Instructor: Alexandra Fedorova
    - also can be taught by Arrvindh Shriraman, Mohamed Hefeeda, J.C. Liu.
- CMPT 741 Data Mining. (3 cr) Fall 2013 Instructor: Jian Pei
    - also can be taught by Martin Ester, Ke Wang.
- CMPT 726 Machine Learning. (3 cr) Fall 2013 Instructor: Greg Mori
    - also can be taught by Oliver Schulte, Anoop Sarkar.
- 1 Elective course (3 credits) from the following list
    - CMPT 825 Natural Language Processing. Spring 2013 Instructor: Anoop Sarkar
        - also can be taught by Fred Popowich.
    - CMPT 711 Bioinformatics Algorithms. Fall 2012 Instructor: Cenk Sahinalp
    - CMPT 767 Visualization. Fall 2010 Instructor: Torsten Moeller
    - IAT 814 Knowledge, Visualization and Communication
    - Any special topics course in Computing Science: CMPT 829, CMPT 880, CMPT 881, CMPT 882, CMPT 884, CMPT 885, CMPT 886, CMPT 887, CMPT 888, CMPT 889.
    - CMPT 894 Directed Reading.

Lab work

Students will take the following two lab courses with 6 credits each. Only students enrolled in the professional master's in Big Data will be permitted to enroll in these courses:

- CMPT 731 Programming for Big Data 1 (6 cr). *new course.* see Appendix I.
- CMPT 732 Programming for Big Data 2 (6 cr). *new course.* see Appendix I.

The lab courses will focus on hands-on learning of various models, algorithms, and software related to Big Data. The detailed curriculum and learning outcomes are given in Appendix I. The course development will be done by research faculty (primarily those teaching the five required courses). They will develop a set of assignments that will provide the learning experience for these labs. The labs themselves will be run using a limited term faculty member who is knowledgeable in this area. The students will have considerable help from teaching assistants.

The lab courses will be run in the Vancouver Visual Analytics Institute (VIVA) lab space in IRMACS.

Since the lab courses provide the premium experience in this program, our plan is to use the VIVA lab as the location for the lab courses for at least the program's three year term as a Cohort Special Arrangements Program. The VIVA lab has recently acquired sophisticated, state of the art computer hardware and software. The VIVA lab which is located in IRMACS also includes breakout discussion areas and useful rooms where the instructor or TAs can conduct impromptu discussions with groups of students. These rooms have built-in laptop projectors and other amenities. In order to accommodate the entire incoming cohort of students, we intend to limit enrollment in the first year of the program to 28 spots. This will ensure that we can offer a truly premium lab experience by ensuring we have a high quality lab for the students admitted into this proposed program. A letter of support from VIVA is attached to this proposal.

For the software and hardware requirements for the labs we intend to use the following resources:

- Premium workstations and displays available in the VIVA lab.
- Amazon compute cloud EC2 / Google Compute Engine.
- Local Hadoop cluster for programming assignments.
- Other cloud computing experimental testbeds already in use in our School for research.

Co-op

A term of co-operative education is an integral part of this program. Students will register for one co-op term. The student will be expected to find a suitable industry partner for the co-op term with the assistance of the co-op office or the student may opt to conduct research into Big Data at one of the various Computing Science research labs as a paid research assistant. Alternatively, the student may appeal to the program director to substitute an elective course for the co-op term. Students will be required to register for at least one of the required courses in the semester after the co-op term.

Academic Requirements within the Graduate General Regulations

All graduate students must satisfy the academic requirements that are specified in the graduate general regulations (residence, course work, academic progress, supervision, research competence requirement, completion time, and degree completion), as well as the specific requirements for the program in which they are enrolled, as described previously.

Timeline

Semester	Courses	Credits
Fall	3 graduate courses	9
Spring	1 graduate course + 1 lab course	9
Summer	Co-op	3
Fall	1 graduate course + 1 lab course	9

The students will take three 3 credit graduate courses in their first semester (Fall); one 3 credit graduate course and one 6 credit lab course in their second semester (Spring); one 3 credit co-op in their third semester (Summer); one 3 credit graduate course and one 6 credit lab course in their fourth semester (Fall). This will enable most students to finish 30 credits in four semesters.

Research Expectations

Since this is a professional master's program there will be no thesis or research project requirement. Students may participate in research into Big Data as part of their co-op term if they obtain a suitable research assistant position in the various Computing Science research labs.

Learning Methodologies

Learning Environment and Methodologies

Proficiency in large-scale data analysis requires expertise across a range of computer science subfields. Analysts must be adept at systems for storing, computing and managing large-scale data. They must be well-versed in highly efficient algorithms for processing massive data. Expertise in data mining to extract patterns from large-scale data and machine learning to improve performance of analysis algorithms given large amounts data are essential aspects of their toolkit. Finally, they must be skilled at visualization so that they can present the information obtained from the data to users who can act on this information. These five pillars of large-scale data analysis – systems, algorithms, data mining, machine learning, and visualization form the core teaching areas of the Big Data program.

Experiential Learning

A co-op term that will provide direct experience in industry is integral to this program. In addition, the two new lab courses proposed for this program will enable students to develop the

skills required in industry and provide them with hands-on experience with the kinds of software programming that are relevant for Big Data, thus making them more qualified for their co-op terms.

Faculty

The main need is for a limited term faculty member to take charge of and teach each of the two lab courses. The need for faculty members can be split up into two parts:

1. Five faculty members are required to teach the classroom courses. No new faculty will be required as all of these courses already exist in the calendar and there are existing faculty members who teach these courses on a fairly regular schedule.
2. A limited term faculty member is needed to teach the lab courses. These are innovative lab courses and the course development necessitates the creation of a series of programming assignments that will impart a hands-on experience in programming with the tools commonly used in Big Data analytics. The steering committee along with selected faculty members will be responsible for developing these assignments along with suitable video lectures that will provide instruction on techniques and methodologies for solving the problems presented in the assignments. Each of the 6 credit lab courses will require 12 hours per week for students (using the typical 0.5 credit per weekly lab hour). A limited term faculty member will be assigned to each 6 credit lab course. There will also be substantial teaching assistant support in the labs to assist students with their assignments. Thus, for an ongoing commitment to these lab courses one limited term faculty member will be required.

We will be able to teach the proposed program with existing faculty and a limited term faculty member hired to teach the lab courses.

Need for this Program

There is a rapidly growing demand for data scientists in industry, science, medicine, and government. According to a McKinsey Global report, "*By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of Big Data to make effective decisions.*" IDC estimates that the worldwide spending on business analytics alone was $90 billion in 2011. EMC's Chuck Hollis says of big data analytics: "*The race is now on to acquire -- and maximize the productivity of -- the key talent behind this wave: data scientists and their supporting data science teams".*

We have had discussions with several industry partners including IBM, SAP, Simba, Demonware, BC Genome Centre, Hootsuite among others, and we have observed that all of them are keen to hire highly qualified personnel who are proficient in Big Data analytics. Companies like Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics. In 2010, this industry on its own was worth more than $100 billion and was growing at almost 10 percent a year, approximately twice as fast as the software industry as a whole.[2]

---

[2] Data, data everywhere. The Economist. 25 February 2010.

In addition, we commissioned a survey to reach potential applicants to the proposed program. We received close to 500 responses to this survey and there was overwhelming interest in this program. The summary of the survey results is attached as an appendix to this proposal (see Appendix VI).

The main group of people we were able to reach in our survey who were enthusiastic about the proposed program were SFU students. The majority of these students, from the Faculty of Applied Sciences and from other units across campus, stated that they would be interested in obtaining a professional master's in Big Data. The survey clearly shows that there is a perceived need for such a fulltime professional master's program among our own students.

We have also received more than 50 emails expressing interest in this program from international students who have joined a mailing list that will provide details about the proposed program when it is approved.

Admission Requirements

To qualify for admission to the M.Sc. program in Big Data, a student must satisfy the university admission requirements for a master's program as stated in Section 1.3.3 of the Graduation Admission section of the SFU calendar and the student must hold a bachelor's degree or equivalent in Computing Science or a related field with a cumulative grade point average (GPA) of 3.0 (on a scale of 0.0 - 4.0) or the equivalent.

The School's Graduate Admissions Committee may, at its discretion, offer M.Sc. admissions to exceptional students who do not have an undergraduate degree in Computing Science or a related field. Minimally we require demonstrated competence in computing science at the third year level equivalent to CMPT 300 (Operating Systems 1), CMPT 307 (Data Structures and Algorithms) and CMPT 354 (Database Systems and Structures).

Students who do not have the proper background in Computing Science may take the three courses listed above in the Summer semester before the Fall cohort begins and then join the M.Sc. program in Big Data.

We wish to be inclusive of a variety of incoming students, including SFU students from related disciplines who wish to obtain a master's degree; international students; and professionals who wish to advance their knowledge in the area of Big Data analysis.

Appendices

Appendix I: New Course Outlines

Please see the attached new graduate course proposal forms for CMPT 731 and CMPT 732, the two new lab courses that will be part of the proposed program. The course description for the SFU Calendar and the learning outcomes are provided below for the two new courses. The course content will be updated each year as Big Data is a fast growing and therefore rapidly

changing area with new ideas from research and industry combining to create new software environments that are better able to scale to larger and faster changing sources of data.

CMPT 731 Programming for Big Data 1 (Big Data Lab 1)

Course Description

The curriculum for this lab course will cover the following topic. We have also provided an estimate of the number of assignments that will be needed for each topic and how long in terms of lab hours each assignment will take. This information provides the justification for setting the number of credits for this course at 6. In these labs, students will have the opportunity to acquire the hands-on experience necessary for a successful career in Big Data in the information technology sector. Many of the assignments will be completed on massive publically available data sets giving them appropriate experience with cloud computing and the algorithms and software tools needed to master programming for Big Data.

- Analysis of scalability of algorithms to big data.
    - Two assignments (12 hours of work; 1 week)
- Data warehouses and online analytical processing.
    - Two assignments (24 hours of work; 2 weeks)
- Efficient storage of big data including data streams.
    - Three assignments (24 hours of work; 2 weeks)
- Scalable querying and reporting on massive data sets.
    - Two assignments (12 hours of work; 1 week)
- Scalable and distributed hardware and software architectures.
    - Three assignments (24 hours of work; 2 weeks)
- Software as a service. Cloud Computing (e.g. Amazon EC2, Google Compute Engine).
    - One assignment (12 hours of work; 1 week)
- Big data programming models: simple map-reduce algorithms
    - Two assignments (24 hours of work; 2 weeks)
- Distributed databases (nosql)
    - Two assignments (24 hours of work; 2 weeks)

Thus, over 13 weeks of lab work and 12 hours per week of lab time, the students will obtain a solid background in programming for Big Data.

Learning Outcomes

By the end of this course, the student will be able to ...

- use a distributed file system such as (or similar to) HDFS (Hadoop Distributed File System).
- write software that can interact with a distributed file system using programming tools that are part of Apache Hadoop.
- write simple distributed software using the map-reduce programming model.
- be able to formulate queries over multidimensional data in an OLAP query language.
- write software that can interact with at least one so-called "nosql" database.
- summarize information compactly using scripting tools similar to Pig or Sawzall

- be able to query data in a distributed streaming model such as Storm.

CMPT 732 Programming for Big Data 2 (Big Data Lab 2)

Course Description
The curriculum of this lab course will cover the following topics.

- Big data programming models: complex map-reduce algorithms, software for implementing streaming and sketching algorithms.
    - Four assignments (48 hours of work; 4 weeks)
- Dealing with unstructured data such as images or text and visualization.
    - Two assignments (24 hours of work; 2 weeks)
- Scalable machine learning methods such as online learning and visualization.
    - Two assignments (24 hours of work; 2 weeks)
- Data mining: methods for learning descriptive and predictive models from data.
    - Two assignments (24 hours of work; 2 weeks)
- Social media analysis and visualization.
    - One assignment (12 hours of work; 1 week)
- Distributed algorithms over very large graphs and matrices.
    - Two assignments (24 hours of work; 2 weeks)

We also provide an estimate of the number of assignments needed for each topic and how long in terms of lab hours each assignment will take. This provides the justification for assigning 6 credits to this lab course. The students will get the majority of the hands-on experience needed for a successful career in Big Data in the information technology industry. Many of the assignments will be completed on massive publically available data sets giving the students appropriate experience with cloud computing and the algorithms and software tools needed to master programming for Big Data.
Thus, over 13 weeks of lab work and 12 hours per week of lab time, and building on the previous lab course CMPT 731, the students will obtain a solid background in programming for Big Data.

Learning Outcomes
By the end of this course, the student will be able to ...

- write more complex map-reduce programs that iterate through several map and reduce stages
- write programs that can read, tokenize and predict useful classes over text data
- write programs that can search through large amounts of image data
- use scalable online machine learning algorithms such as stochastic gradient descent over large data sets
- compute matrix operations on matrix data stored as shards on a distributed file system
- visualize large amounts of data using visualization tools such as Tableau, Spotfire or d3.js
- implement streaming and sketching algorithms over large data sets

Appendix II: Proposed Calendar Entry

Computing Science

Professional Master of Science in Big Data

The school offers a professional Master of Science that involves taking courses and a co-op term.

Admission Requirements

To qualify for admission to the M.Sc. program in Big Data, a student must satisfy the university admission requirements for master's program as stated in Section 1.3.3 of the Graduation Admission section of the SFU calendar and the student must hold a bachelor's degree or equivalent in Computing Science or a related field with a cumulative grade point average (GPA) of 3.0 (on a scale of 0.0 - 4.0) or the equivalent.

The School's Graduate Admissions Committee may offer, at its discretion, M.Sc. admissions to exceptional students without an undergraduate degree in Computing Science or a related field. Minimally we require demonstrated competence in computing science at the third year level equivalent to CMPT 300 (Operating Systems 1), CMPT 307 (Data Structures and Algorithms) and CMPT 354 (Database Systems and Structures).

Students who do not have the proper background in Computing Science may take the three courses listed above in the Summer semester before the Fall cohort begins and then join the M.Sc. program in Big Data.

Program Requirements

Students will complete 30 units of graduate work. These units are divided into three sections: 15 credits of graduate course work; 12 credits of specialized lab work; 3 credits for co-op.

Course work

Five 3 credit courses must be taken from the choices below:

- Pick one of the following two courses:
  - CMPT 881 Algorithms for Massive Data. (3 cr) Spring 2013 Instructor: Funda Ergun
  - CMPT 705 Design and Analysis of Algorithms (3cr; offered every semester by several faculty in our School)
- CMPT 886 Systems for Big Data. (3 cr) Spring 2012 Instructor: Alexandra Fedorova
  - also can be taught by Arrvindh Shriraman, Mohamed Hefeeda, J.C. Liu.
- CMPT 741 Data Mining. (3 cr) Fall 2013 Instructor: Jian Pei
  - also can be taught by Martin Ester, Ke Wang.
- CMPT 726 Machine Learning. (3 cr) Fall 2013 Instructor: Greg Mori
  - also can be taught by Oliver Schulte, Anoop Sarkar.
- 1 Elective course (3 cr)
  - CMPT 825 Natural Language Processing. Spring 2013 Instructor: Anoop Sarkar
    - also can be taught by Fred Popowich.
  - CMPT 711 Bioinformatics Algorithms. Fall 2012 Instructor: Cenk Sahinalp
  - CMPT 767 Visualization. Fall 2010 Instructor: Torsten Moeller

- IAT 814 Knowledge, Visualization and Communication.
- Any special topics course in Computing Science: CMPT 829, CMPT 880, CMPT 881, CMPT 882, CMPT 884, CMPT 885, CMPT 886, CMPT 887, CMPT 888, CMPT 889.
- CMPT 894 Directed Reading.

Lab work

Students will take the following two lab courses worith 6 credits each. Only students enrolled in the professional master's in Big Data will be able permitted to enroll in these courses:

- CMPT 731 Programming for Big Data 1 (6 cr). *new course.*
- CMPT 732 Programming for Big Data 2 (6 cr). *new course.*

Co-op

A co-op internship is an integral part of this program. Students will register for one co-op term. With assistance from the co-op coordinator for this program, students will be expected to find a suitable industry partner for the co-op term. The student may instead choose to conduct research into Big Data at one of the various Computing Science research labs as a paid research assistant to satisfy their co-op requirement. In extenuating circumstances, a student may appeal to the program director to take an elective course from the list of electives for this program instead of a co-op. Students are required to enroll in at least one of the required courses in the semester following the co-op term.

Academic Requirements within the Graduate General Regulations

All graduate students must satisfy the academic requirements that are specified in the graduate general regulations (residence, course work, academic progress, supervision, completion time, and degree completion), as well as the specific requirements for the program in which they are enrolled, as shown above.

Appendix IIIa: Graduate Scholarships

If we admit more than 20 students (the current plan is to cap enrollment in the first year to 28 students) the plan is to offer additional graduate fellowships to Computing Science graduate students. For instance, if we are able to enroll 28 students in the first year (as we plan) then once we account for increased expenses for hardware/software as well as increased co-op expenses, we will be able to fund 8-10 entrance scholarships for Computing Science graduate students to make us more competitive in admitting the best students to enhance our graduate research program. This will in turn benefit the professional master's students in the proposed program as they will receive high quality mentoring from some of these students who will be their teaching assistants.

Appendix IIIb: Start and End Dates

It is planned that this proposed program will commence in the Fall semester 2014. Admissions decisions will be made by April 2014. The last date of admission to this program under this Cohort Special Arrangements Program will be September 2016.

Appendix IV: Steering Committee

The following faculty members from Computing Science are the members of the steering committee for this program:

- Alexandra Fedorova, Associate Professor, fedorova@sfu.ca
- Anoop Sarkar, Associate Professor, anoop@sfu.ca
- Martin Ester, Professor and Director, ester@sfu.ca
- Greg Mori, Associate Professor, mori@sfu.ca

For now, please direct all correspondence about this proposal to Anoop Sarkar.

Appendix V: Faculty CVs

Please refer to the attached documents for the CVs of the following faculty involved in teaching this program.

- Funda Ergun
- Martin Ester
- Alexandra Fedorova
- Mohamed Hefeeda
- J.C. Liu
- Greg Mori
- Jian Pei
- Cenk Sahinalp
- Anoop Sarkar
- Oliver Schulte
- Arrvindh Shriraman
- Ke Wang

Appendix VI: Supplementary Information

Related Master of Science programs in Big Data

Big Data is a brand-new, emerging discipline. We are aware of only six related master's programs and one online course (all in the US), which offer some elements of our vision, but none of which are strictly comparable to the program we envision. These programs are:

- Northwestern University: Master of Analytics Program
- North Carolina State University: Institute for Advanced Analytics
- Carnegie Mellon University: Very Large Information Systems
- New York University: master's in Data Science
- DePaul University: master's in Predictive Analytics
- Worcester Polytechnic Institute: master's in Data Science
- Harvard University: online course on Data Analytics

Existing programs either specialize in analytics (Northwestern, North Carolina, DePaul) or on systems aspects (CMU). None of them offers a "complete package" that combines the various aspects of Big Data analysis in a meaningful way.

Some of these programs offer job placement services and emphasize employment prospects as one of their key advantages. Many of these programs include an industrial co-op as an integral part of the program. Most or all programs have a capstone project, where students solve a real problem from a real industry partner. This is consistent with our own co-op term. To provide avenues to future employment, we can possibly leverage the co-op as a way to improve chances of employment for our graduates, for example, through internship opportunities provided by MITACS.

Consultations with a number of companies have led us to create a more complete program which includes a background in systems, programming, and human computer interaction specifically for Big Data analysis. Our program combines all of the aspects of Big Data analysis in one program while at the same time being pragmatic about the scope of a professional master's program. Many of the universities listed above do not exclusively focus on computational aspects, but rather on using tools aimed at specific use cases such as healthcare, hospitality and marketing.

In terms of geographical location, if we can launch this program in 2014 we should be among the first, if not the first, such master's program in Canada. The programs listed above are all in the United States, and tuition is often twice or even three times the tuition we are planning to charge for our proposed program.

FACULTY OF COMMUNICATION, ART AND TECHNOLOGY
School of Interactive Arts and Technology

13450 102 Avenue, Surrey BC
Canada V3T 0A3

TEL 778.782.7532
FAX 778.782.9422

mhatala@sfu.ca
www.sfu.ca/siat

MEMORANDUM

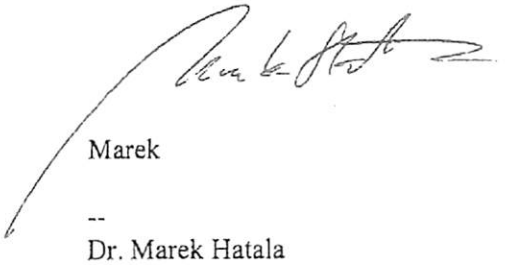
ATTENTION	Dr. Martin Ester, Director, Computing Science	DATE	December 11, 2013
FROM	Dr. Marek Hatala, Director	PAGES	1
RE:	**Professional Masters Program in Big Data (Cohort Special Arrangements Program)**		

Dear Martin,

Our Graduate Program Committee reviewed your School's proposal for *Professional Masters Program in Big Data* (Cohort Special Arrangements Program) and discussed the suitability of including the course IAT 814 Knowledge Visualization and Communication into the list of electives for this program. Based on our GPC's recommendation, I am happy to confirm that SIAT will make seats available in IAT 814 for students of the proposed program.

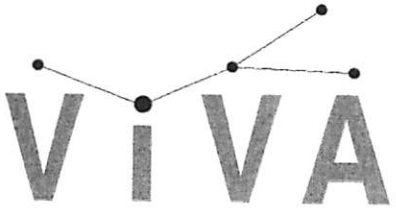
I would also like to communicate to you SIAT's strong support for the proposed program. Our core strength and interest in Visual Analytics is complementary to Big Data and we are looking forward to further collaboration between our programs.

Sincerely,



Marek

--
Dr. Marek Hatala
Professor and Director



Vancouver Institute for Visual Analytics

Simon Fraser University
VIVA (ASB 10905)
8888 University Drive
Burnaby, British Columbia
Canada V5A 1S6
Contact: info@viva-viva.ca
Web: www.viva-viva.ca

6 December 2013

Dr. Martin Ester, Director
School of Computing Science
Simon Fraser University
8888 University Drive
Burnaby, BC
V5A 1S6

Re: Support for Professional Masters Program in Big Data

Dear Dr. Ester,

As Director of the Vancouver Institute for Visual Analytics (VIVA), I am writing to express our enthusiastic support for the School of Computing Science's proposed Professional Masters Program in Big Data.

VIVA is a joint UBC-SFU institute dedicated to developing Visual Analytics (VA) expertise in industry, government and institutions across Canada, as well as fostering collaborative VA research. VA is the science of analytical reasoning facilitated by interactive visual interfaces. VA tools and techniques enable analysts to synthesize and explore information and derive insight from massive, dynamic and possibly conflicting data collections. VA constantly deals with issues in big data, so graduates from the proposed program would be well suited for careers with the industrial partners of VIVA, and in careers dealing with visual analytics research, development and applications.

VIVA is committed to assisting the School of Computing Science on the proposed program in a number of ways. First, we will provide students in new program with full access to our Visual Analytics Research and Instructional Labs (VARI Labs). The laboratory contains a wide selection of state of the art visual analytics software, along with the capacity to develop and deploy new software for use on a wide range of platforms. Note that Western Economic Diversification Canada is providing $513,141 in funding for the labs, with another $1.5 million in-kind contribution is coming from IBM, along with other big data software provided by VIVA partners. Second, we will assist in finding internships for the graduate students through our partner organizations and through our Andrew Wade Visual Analytics Challenge Program. Third, faculty members affiliated with VIVA will be willing to act in advisory roles in curriculum development activities. Finally, VIVA will also work at promoting the new program through its partnerships, including the newly established national not-for-profit Canadian Network for Visual Analytics (CANVAC), involving researchers from Dalhousie University, OCAD University, Simon Fraser University, University of British Columbia, the University of Alberta, the University of Calgary, and York University.

We sincerely look forward to the opportunity of working with the School of Computing Science on this initiative.

Sincerely,

Dr. Fred Popowich
Director
Vancouver Institute for Visual Analytics
fred@viva-viva.ca +1.778.782.4193

SFU SIMON FRASER UNIVERSITY
DEAN OF GRADUATE STUDIES

New Graduate Course Proposal Form

PROPOSED COURSE

Subject (eg. MAPH) CMPT	Number (eg. 810) 731	Units (eg. 4) 6

Course Title (max 80 characters)
Programming for Big Data 1

Short Title (appears on transcripts, max 25 characters)
Big Data Lab 1

Course Description for SFU Calendar ☑ see attached document ☑ Learning outcomes identified

See Appendix I: New Course Outlines in the Program Proposal (also attached)

Available Course Components: ☐ Lecture ☐ Seminar ☑ Laboratory ☐ Practicum ☐ Online ☐ _____

Grading Basis ☑ Letter grades ☐ Satisfactory/Unsatisfactory ☐ In Progress/Complete | This is a capstone course ☐ Yes ☐ No

Prerequisites (if any) ☑ see attached document (if more space is required)

☐ This proposed course is combined with an undergrad course: Course number and units: _____

Additional course requirements for graduate students ☐ See attached document (if this space is insufficient)

Campus at which course will be offered (check all that apply) ☑ Burnaby ☐ Vancouver ☐ Surrey ☐ GNW ☐ _____

Estimated enrolment	Date of initial offering	Course delivery (eg. 3 hrs/week for 13 weeks)
20	Fall 2014	12 hrs/week for 13 weeks

☐ Yes ☑ No Practicum work done in this class will involve children or vulnerable adults
(If the "Yes" box is checked, all students will require criminal record checks)

Justification ☐ See attached document (if more space is required)
This is one of two lab courses that are part of the Professional Master of Science in Big Data. These courses offer hands-on software programming experience in the area of Big Data.

RESOURCES

If additional resources are required to offer this course, the department proposing the course should be prepared to provide information on the source(s) of those additional resources.

Faculty member(s) who will normally teach this course ☐ information about their competency to teach the course is appended
Limited term faculty member

Number of additional faculty members required in order to offer this course

Additional space required in order to offer this course ☐ see attached document

Additional specialized equipment required in order to offer this course ☐ see attached document

Additional Library resources required (append details) ☐ Annually $_____ ☐ One-time $_____

PROPOSED COURSE *from first page*

Program (eg. MAPH) CMPT	Number (eg. 810) 731	Units (eg. 4) 6
Course title (max 80 characters) Programming for Big Data 1		

APPROVAL SIGNATURES

When a department proposes a new course it must first be sent to the chairs of each faculty graduate program committee where there might be an overlap in course content. The chairs will indicate that overlap concerns have been dealt with by signing the appropriate space or via a separate memo or e-mail (attached to this form).

The new course proposal must also be sent to the Library for a report on library resources.

Once overlap concerns have been dealt with, signatures indicate approval by the department, home faculty and Senate Graduate Studies Committee.

Other Faculties
The signature(s) below indicate that the Dean(s) or designate of other Faculties affected by the proposed new course support(s) the approval of the new course.

Name of Faculty	Signature of Dean or Designate	Date

Departmental Approval (non-departmentalized faculties need not sign)

Department Graduate Program Committee	Signature	Date
Department Chair Martin Ester	Signature M Este	Date Dec 12, 2013

Faculty Approval
Faculty approval indicates that all the necessary course content and overlap concerns have been resolved, and that the Faculty/Department commits to providing the required Library funds and any other necessary resources.

Faculty Graduate Program Committee	Signature	Date
R. CAMERON	R. Cameron	Dec 12, 2013

Senate Graduate Studies Committee Approval
SGSC approval indicates that the Library report has been seen, and all resource issues dealt with. Once approved, new course proposals are sent to Senate for information.

Senate Graduate Studies Committee	Signature	Date
		Jan 8/14

CONTACT

Upon approval of the course, the Office of the Dean of Graduate Studies will consult with the department or school regarding other course attributes that may be required to enable the proper entry of the new course in the student record system.

Department / School / Program	Contact name	Contact email
Computing Science	Martin Ester	ester@sfu.ca

Course Outline - CMPT 731 - Programming for Big Data 1

Information

Subject	CMPT
Catalog number	731
Section	D100
Semester	2014 Fall (1147)
Title	Programming for Big Data 1
Instructor(s)	Limited Term Faculty Member
Campus	Burnaby Mountain Campus

Calendar Objective / Description

This course is one of two lab courses that are part of the Professional Masters Program in Big Data in the School of Computing Science. This lab course aims to provide students with the hands-on experience needed for a successful career in Big Data in the information technology industry. Many of the assignments will be completed on massive publically available data sets giving them appropriate experience with cloud computing and the algorithms and software tools needed to master programming for Big Data. Over 13 weeks of lab work and 12 hours per week of lab time, the students will obtain a solid background in programming for Big Data.

Instructor's Objectives

By the end of this course, the student will be able to ...

- use a distributed file system such as (or similar to) HDFS (Hadoop Distributed File System).
- write software that can interact with a distributed file system using programming tools that are part of Apache Hadoop.
- write simple distributed software using the map-reduce programming model.
- be able to formulate queries over multidimensional data in an OLAP query language.
- write software that can interact with at least one so-called "nosql" database.
- summarize information compactly using scripting tools similar to Pig or Sawzall
- be able to query data in a distributed streaming model such as Storm.

Prerequisites

None

Topics

- Analysis of scalability of algorithms to big data.
    - Two assignments (12 hours of work; 1 week)
- Data warehouses and online analytical processing.
    - Two assignments (24 hours of work; 2 weeks)

- Efficient storage of big data including data streams.
    - Three assignments (24 hours of work; 2 weeks)
- Scalable querying and reporting on massive data sets.
    - Two assignments (12 hours of work; 1 week)
- Scalable and distributed hardware and software architectures.
    - Three assignments (24 hours of work; 2 weeks)
- Software as a service. Cloud Computing (e.g. Amazon EC2, Google Compute Engine).
    - One assignment (12 hours of work; 1 week)
- Big data programming models: simple map-reduce algorithms
    - Two assignments (24 hours of work; 2 weeks)
- Distributed databases (nosql)
    - Two assignments (24 hours of work; 2 weeks)

Grading

17 assignments (100% of the grade).

Recommended Books

- Hadoop: The Definitive Guide, 3rd Edition. Storage and Analysis at Internet Scale. Tom White. O'Reilly Media. 2012. Ebook ISBN:978-1-4493-1151-3
- Mahout in Action. Sean Owen, Robin Anil, Ted Dunning, Ellen Friedman. Manning Publications. 2011. ISBN-13: 978-1935182689
- Data-Intensive Text Processing with MapReduce. Jimmy Lin and Chris Dyer. Morgan and Claypool. 2010. (doi:10.2200/S00274ED1V01Y201006HLT007)

Academic Honesty Statement

Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct Policies (http://www.sfu.ca/policies/gazette/student.html).

SFU SIMON FRASER UNIVERSITY
DEAN OF GRADUATE STUDIES

New Graduate Course Proposal Form

PROPOSED COURSE

Subject (eg. MAPH) CMPT	Number (eg. 810) 732	Units (eg. 4) 6

Course Title (max 80 characters)
Programming for Big Data 2

Short Title (appears on transcripts, max 25 characters)
Big Data Lab 2

Course Description for SFU Calendar ☑ see attached document ☑ Learning outcomes identified

See Appendix I: New Course Outlines in the Program Proposal

Available Course Components: ☐ Lecture ☐ Seminar ☑ Laboratory ☐ Practicum ☐ Online ☐ _____

Grading Basis ☑ Letter grades ☐ Satisfactory/Unsatisfactory ☐ In Progress/Complete | This is a capstone course ☐ Yes ☐ No

Prerequisites (if any) ☐ see attached document (if more space is required)

☐ This proposed course is combined with an undergrad course: Course number and units: _____

Additional course requirements for graduate students ☐ See attached document (if this space is insufficient)

Campus at which course will be offered (check all that apply) ☑ Burnaby ☐ Vancouver ☐ Surrey ☐ GNW ☐ _____

Estimated enrolment	Date of initial offering	Course delivery (eg. 3 hrs/week for 13 weeks)
20	Fall 2014	12 hrs/week for 13 weeks

☐ Yes ☑ No Practicum work done in this class will involve children or vulnerable adults
(If the "Yes" box is checked, all students will require criminal record checks)

Justification ☐ See attached document (if more space is required)
This is one of two lab courses that are part of the Professional Master of Science in Big Data. These courses offer hands-on software programming experience in the area of Big Data.

RESOURCES

If additional resources are required to offer this course, the department proposing the course should be prepared to provide information on the source(s) of those additional resources.

Faculty member(s) who will normally teach this course ☐ information about their competency to teach the course is appended
Limited term faculty member

Number of additional faculty members required in order to offer this course

Additional space required in order to offer this course ☐ see attached document

Additional specialized equipment required in order to offer this course ☐ see attached document

Additional Library resources required (append details) ☐ Annually $_____ ☐ One-time $_____

PROPOSED COURSE *from first page*

Program (eg. MAPH) CMPT	Number (eg. 810) 732	Units (eg. 4) 6
Course title (max 80 characters) Programming for Big Data 2		

APPROVAL SIGNATURES

When a department proposes a new course it must first be sent to the chairs of each faculty graduate program committee where there might be an overlap in course content. The chairs will indicate that overlap concerns have been dealt with by signing the appropriate space or via a separate memo or e-mail (attached to this form).

The new course proposal must also be sent to the Library for a report on library resources.

Once overlap concerns have been dealt with, signatures indicate approval by the department, home faculty and Senate Graduate Studies Committee.

Other Faculties

The signature(s) below indicate that the Dean(s) or designate of other Faculties affected by the proposed new course support(s) the approval of the new course.

Name of Faculty	Signature of Dean or Designate	Date

Departmental Approval (non-departmentalized faculties need not sign)

Department Graduate Program Committee	Signature	Date
Department Chair Martin Ester	Signature M. Ester	Date Dec 12, 2013

Faculty Approval

Faculty approval indicates that all the necessary course content and overlap concerns have been resolved, and that the Faculty/Department commits to providing the required Library funds and any other necessary resources.

Faculty Graduate Program Committee	Signature	Date
R. CAMERON	R. Cameron	Dec. 12, 2013

Senate Graduate Studies Committee Approval

SGSC approval indicates that the Library report has been seen, and all resource issues dealt with. Once approved, new course proposals are sent to Senate for information.

Senate Graduate Studies Committee	Signature	Date
	W Paulhus	Jan 8/14

CONTACT

Upon approval of the course, the Office of the Dean of Graduate Studies will consult with the department or school regarding other course attributes that may be required to enable the proper entry of the new course in the student record system.

Department / School / Program	Contact name	Contact email
Computing Science	Martin Ester	ester@sfu.ca

Course Outline - CMPT 732 - Programming for Big Data 2

Information

Subject	CMPT
Catalog number	732
Section	D100
Semester	2014 Fall (1147)
Title	Programming for Big Data 2
Instructor(s)	Limited Term Faculty Member
Campus	Burnaby Mountain Campus

Calendar Objective / Description

This course is one of two lab courses that are part of the Professional Masters Program in Big Data in the School of Computing Science. This lab course aims to provide students with the hands-on experience needed for a successful career in Big Data in the information technology industry. Many of the assignments will be completed on massive publically available data sets giving them appropriate experience with cloud computing and the algorithms and software tools needed to master programming for Big Data. Over 13 weeks of lab work and 12 hours per week of lab time, , and building on the previous lab course CMPT 731, the students will obtain a solid background in programming for Big Data.

Instructor's Objectives

By the end of this course, the student will be able to ...

- write more complex map-reduce programs that iterate through several map and reduce stages
- write programs that can read, tokenize and predict useful classes over text data
- write programs that can search through large amounts of image data
- use scalable online machine learning algorithms such as stochastic gradient descent over large data sets
- compute matrix operations on matrix data stored as shards on a distributed file system
- visualize large amounts of data using visualization tools such as Tableau, Spotfire or d3.js
- implement streaming and sketching algorithms over large data sets

Prerequisites

CMPT 731: Programming for Big Data 1

Topics

- Big data programming models: complex map-reduce algorithms, software for implementing streaming and sketching algorithms.
    - Four assignments (48 hours of work; 4 weeks)
- Dealing with unstructured data such as images or text and visualization.

    - Two assignments (24 hours of work; 2 weeks)
- Scalable machine learning methods such as online learning and visualization.
    - Two assignments (24 hours of work; 2 weeks)
- Data mining: methods for learning descriptive and predictive models from data.
    - Two assignments (24 hours of work; 2 weeks)
- Social media analysis and visualization.
    - One assignment (12 hours of work; 1 week)
- Distributed algorithms over very large graphs and matrices.
    - Two assignments (24 hours of work; 2 weeks)

Grading

13 assignments (100% of the grade).

Recommended Books

- Hadoop: The Definitive Guide, 3rd Edition. Storage and Analysis at Internet Scale. Tom White. O'Reilly Media. 2012. Ebook ISBN:978-1-4493-1151-3
- Getting Started with Storm. Continuous streaming computation with Twitter's cluster technology. Jonathan Leibiusky, Gabriel Eisbruch, Dario Simonassi. O'Reilly Media. 2012. Ebook ISBN:978-1-4493-2400-1.
- Mahout in Action. Sean Owen, Robin Anil, Ted Dunning, Ellen Friedman. Manning Publications. 2011. ISBN-13: 978-1935182689
- Data-Intensive Text Processing with MapReduce. Jimmy Lin and Chris Dyer. Morgan and Claypool. 2010. (doi:10.2200/S00274ED1V01Y201006HLT007)
- Beautiful Visualization: Looking at Data through the Eyes of Experts. Julie Steele, Noah Iliinsky. O'Reilly Media. 2010. Ebook ISBN:978-1-4493-7987-2

Academic Honesty Statement

Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct Policies (http://www.sfu.ca/policies/gazette/student.html).

*

CMPT 731: Programming for Big Data 1, and
*

CMPT 732: Programming for Big Data 2.

No additional library resources are required.

The courses have therefore be added to the appropriate list at
http://www.lib.sfu.ca/collections/course-assessments

This will be enough to indicate library sign-off as they move through the approval process.

Best,

Megan

Megan L. Crouch
Health Sciences Librarian
Collections Librarian
Simon Fraser University / W.A.C. Bennett Library
8888 University Drive, Burnaby, B.C. V5A 1S6
mcrouch@sfu.ca / Tel: 778.782.4962 / Fax: 778.782.3023
I am on campus Monday, Tuesday, Wednesday, and alternate Fridays

--
Dr. Martin Ester
Professor and Director
School of Computing Science
Simon Fraser University
ester@cs.sfu.ca

Analysis of Survey Results Regarding Potential Interest in SFU's Big Data Professional Masters Program

Research Assistant: TJ Gill
August 1, 2013

Simon Fraser University
Supervisor: Dr. Alexandra Fedorova
Special thanks to Stephen Price for his large contribution in the creation of the survey

Link to the survey: http://websurvey.sfu.ca/survey/137905546

Contents

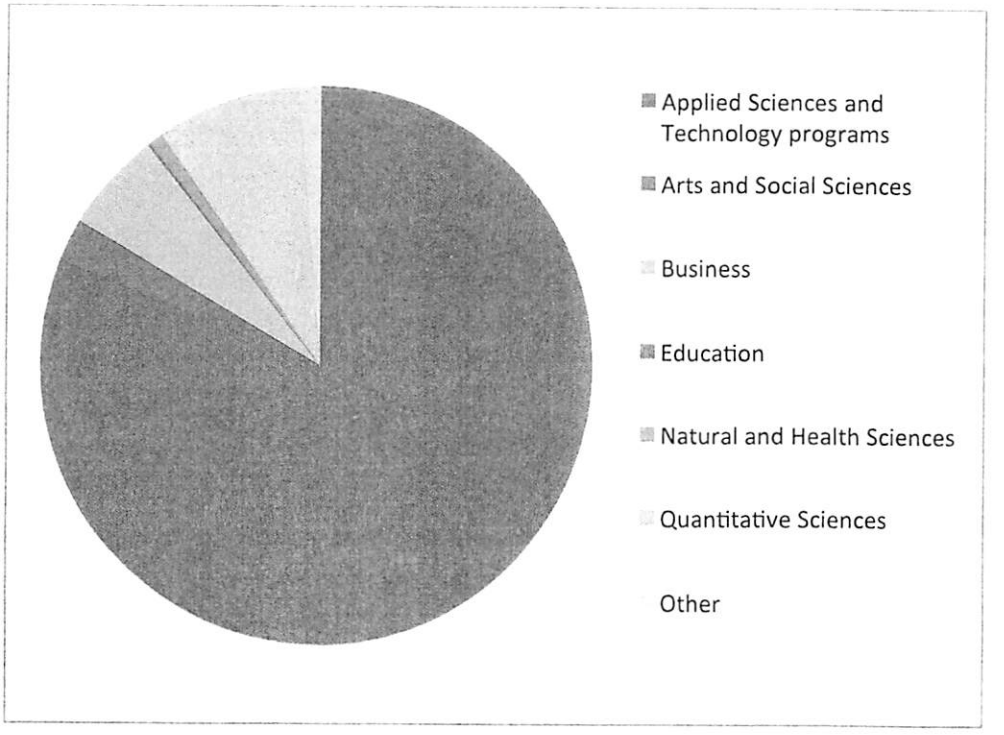
In-depth Analysis

For questions one and three (see Appendix), the survey allows the respondent to select all applicable options. The problem that this causes is that it is unclear how many respondents have picked more than one option. Moreover, it is not known how many options respondents have chosen if they have indeed chosen more than one option. The following section shows a comparison between the original data, and the data that has been modified to remedy the above problems.

Area of Concentration Breakdown

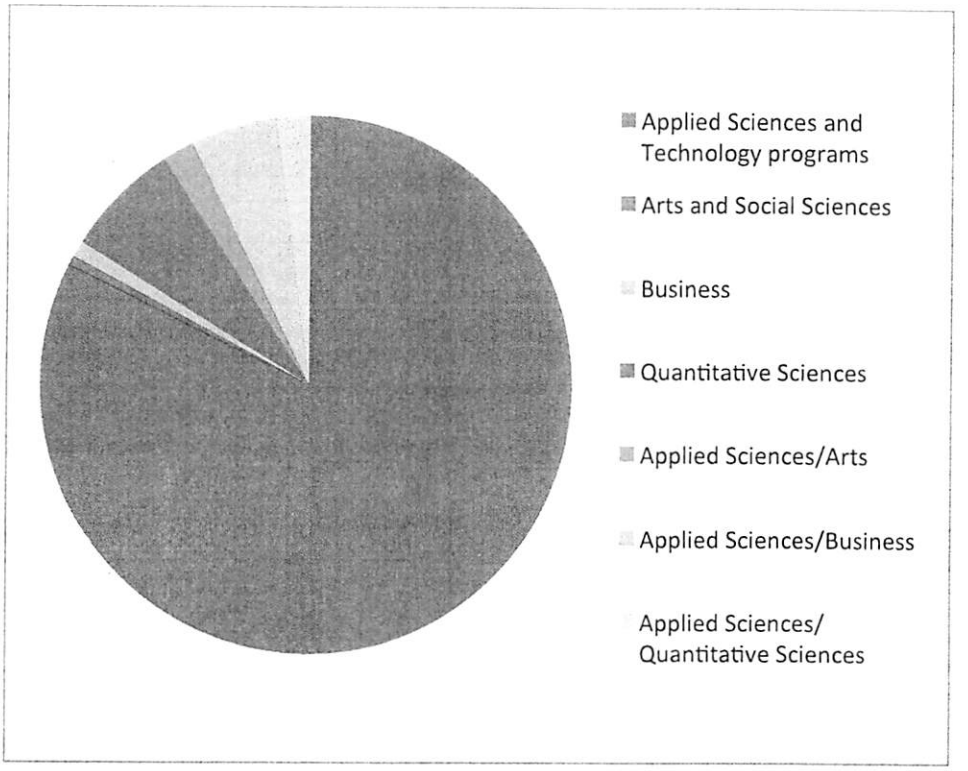
This section shows a comparison between the original and modified data for question one of the survey, which inquires about respondent's area of concentration.

Original Data



The original data shows that a large majority of respondents to the survey concentrate in 'Applied Sciences and Technology Programs.' This pie chart does not show how many of the respondents concentrate in more than one discipline. The following pie chart solves this problem.

Modified Data

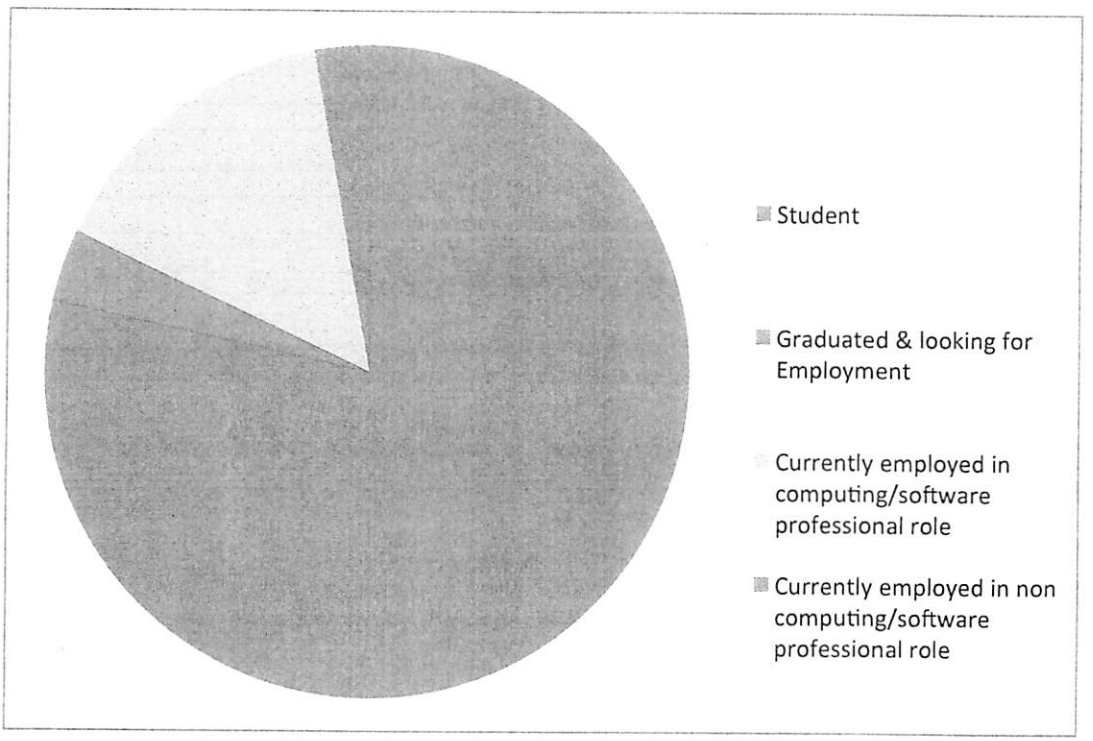


The modified data, which separates the original data so that we see how many respondents chose multiple options, shows fewer students focusing solely in 'Arts,' 'Business,' and 'Quantitative Sciences' and a greater number of students focusing in 'Applied Sciences' along with another discipline.

Current State of Career Breakdown

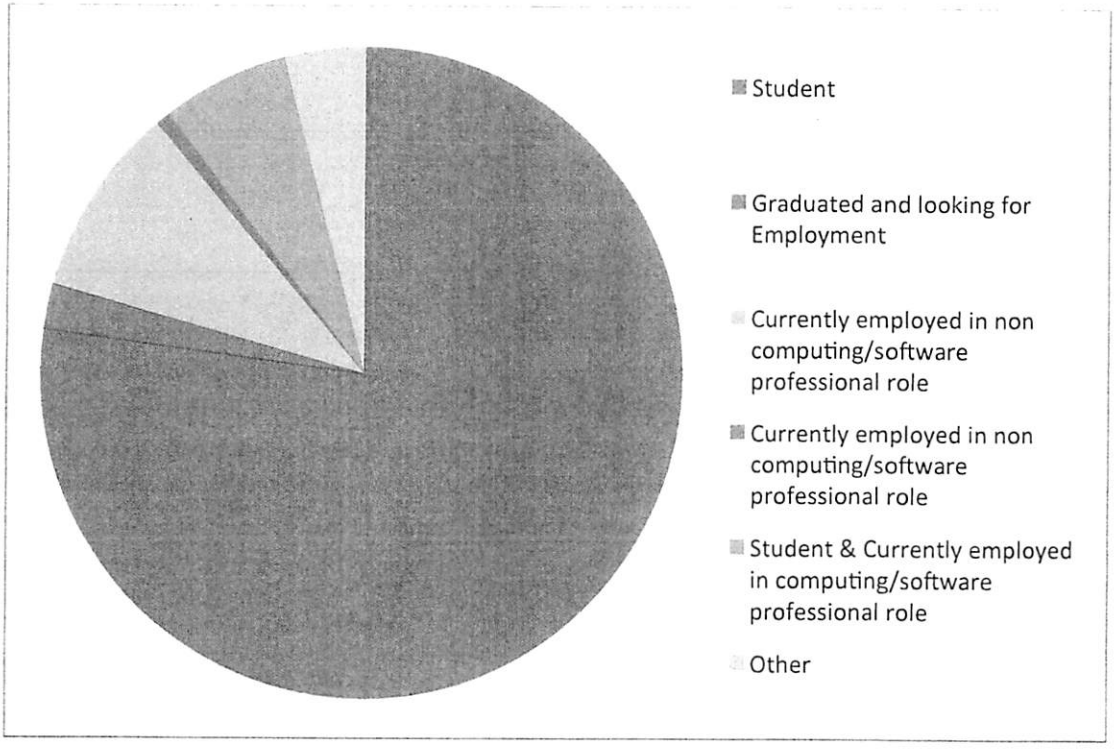
This section shows a comparison between the original and modified data for question three of the survey, which inquires about respondent's current state of career.

Original Data



The original data shows that a large majority of respondents to the survey are at least students. This pie chart does not show how many of them are students as well as another option in the list. The following pie chart solves this problem.

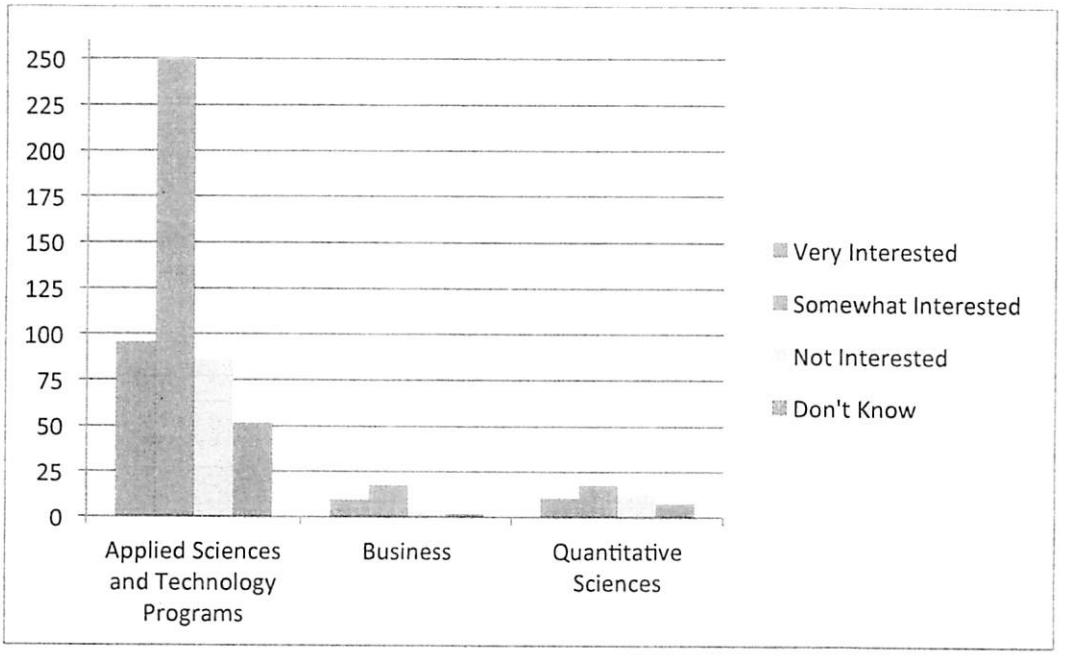
Modified Data



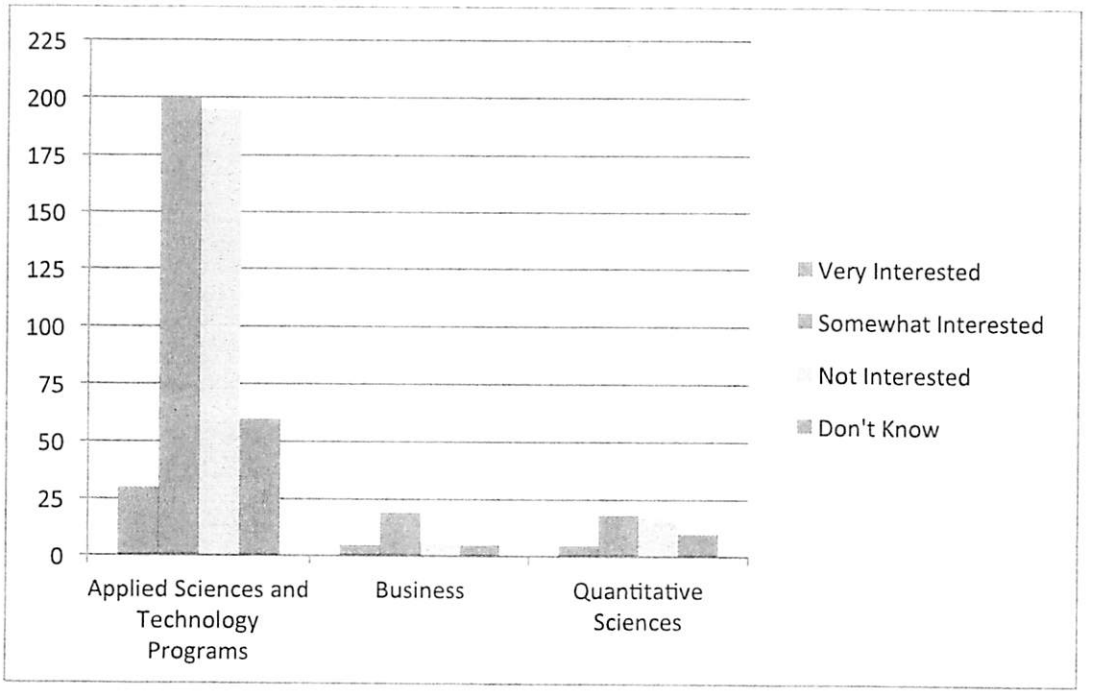
The modified data, which separates the original data so that we see how many respondents chose one or more option, shows that a noticeable chunk of students are also currently employed in a computing software/professional role. The 'other' category simply represents respondents who chose other combinations of options, none of which were sizeable enough to display in the pie chart.

For the second part of the quantitative analysis, questions six and seven, the two questions that gauge the respondent's interest in the program were compared to questions one to four, four questions that inquire about certain characteristics of the respondent (see Appendix). The reason why these four questions were compared to question six was so that the type of respondent most interested in the program could be determined. This allows SFU to identify its target market for its Big Data Professional Masters Program. The reason why these four questions were compared to questions six and seven was to see the difference that tuition cost had on respondent interest in the program. The charts are shown first, with their explanations shown later.

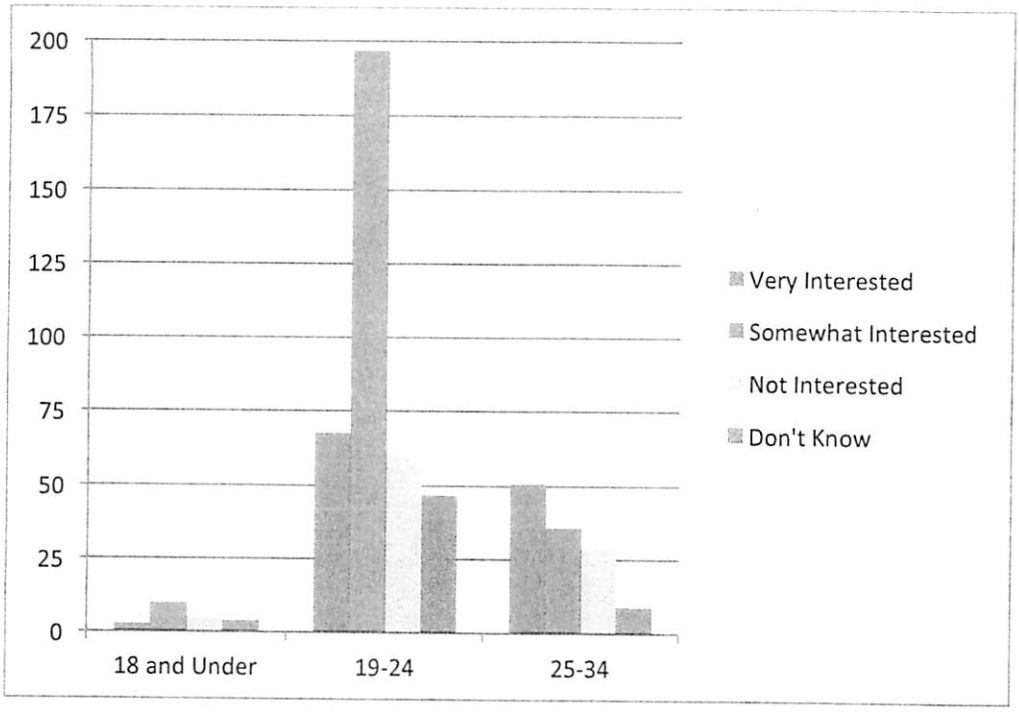
Respondent Area of Concentration vs. Interest Setting Aside Tuition



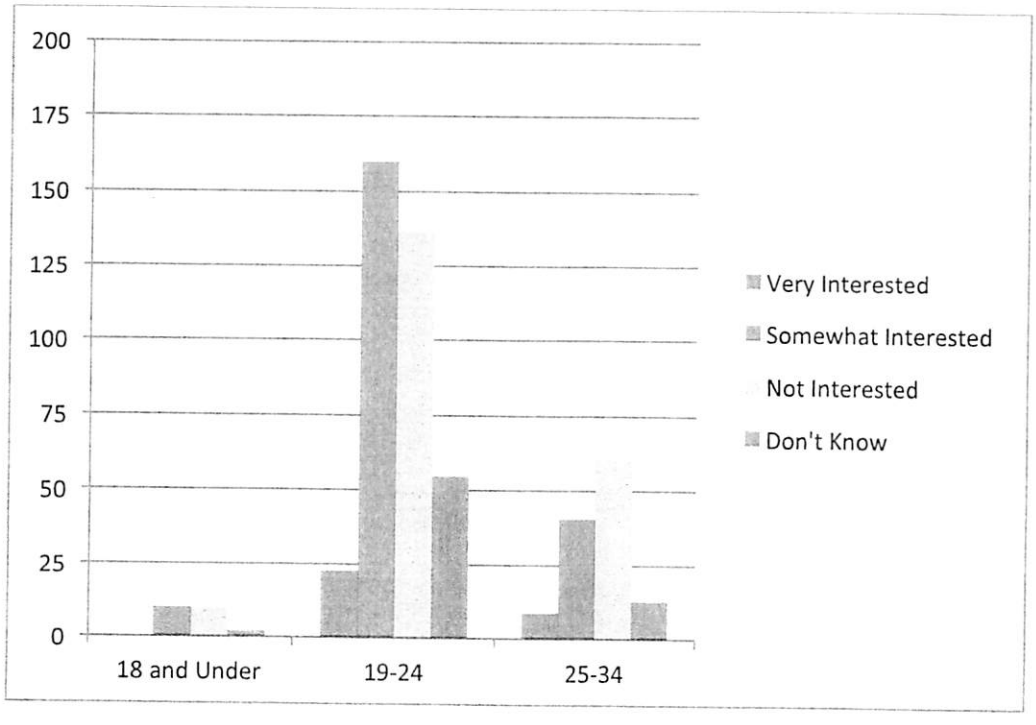
Respondent Area of Concentration vs. Interest Including the Cost of Tuition



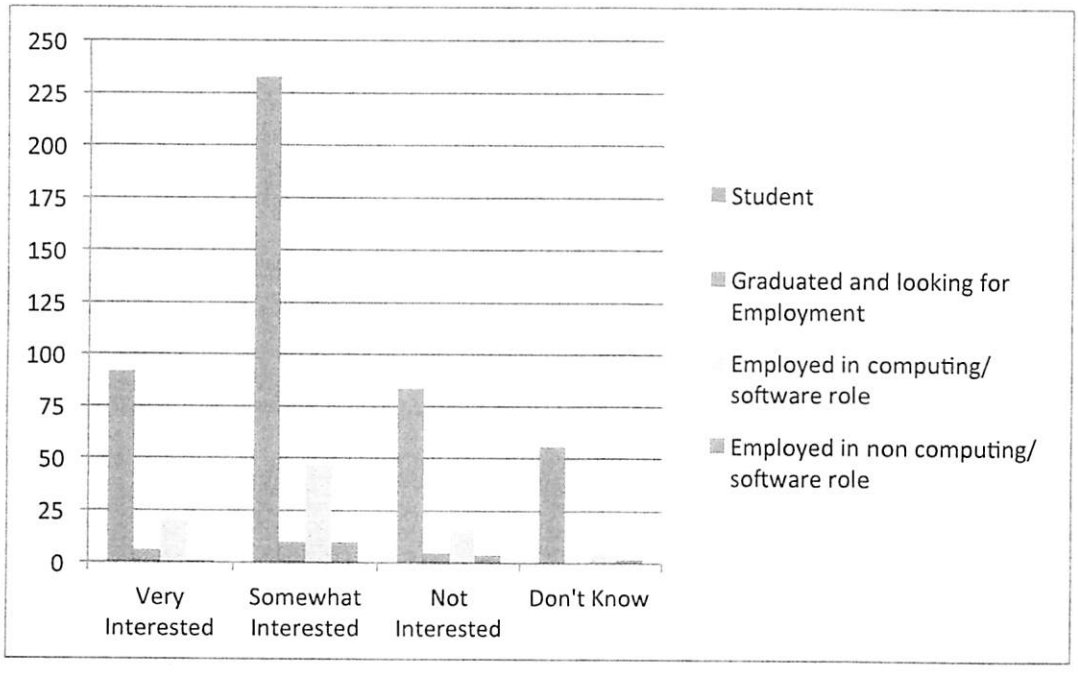
Respondent Age vs. Interest Setting Aside Tuition



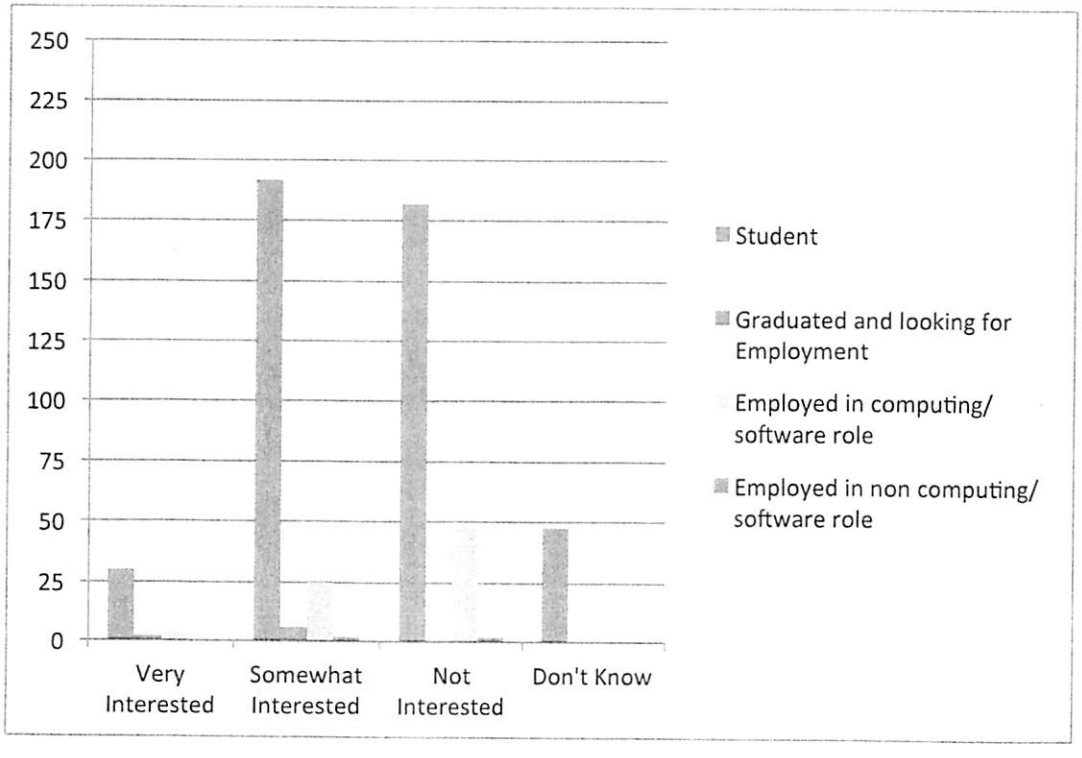
Respondent Age vs. Interest Including the Cost of Tuition



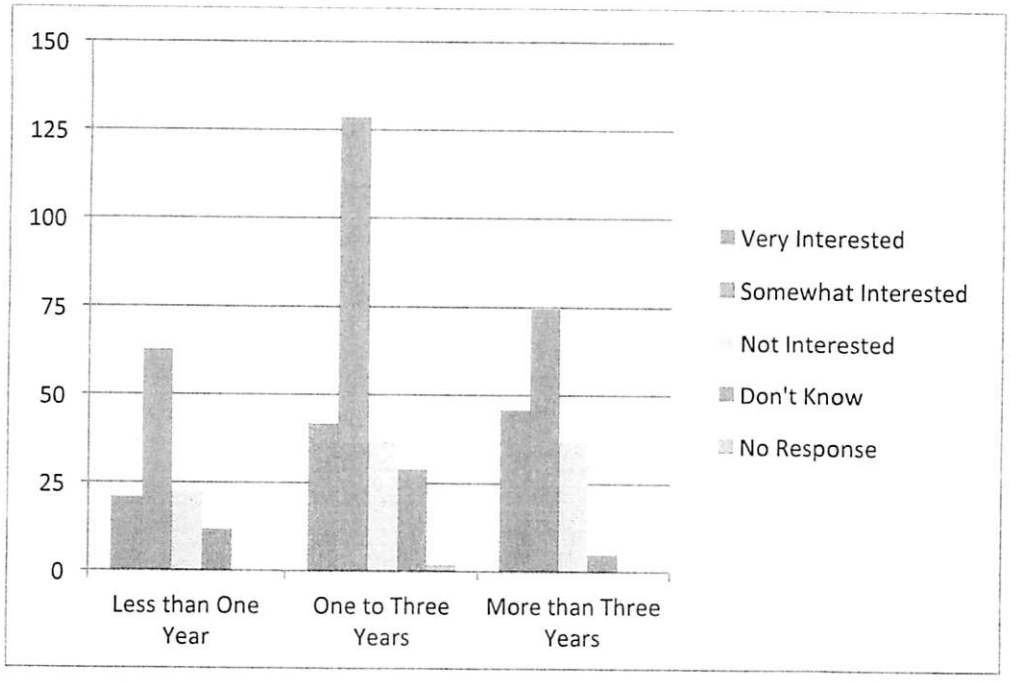
Where in Career vs. Interest Setting Aside Tuition



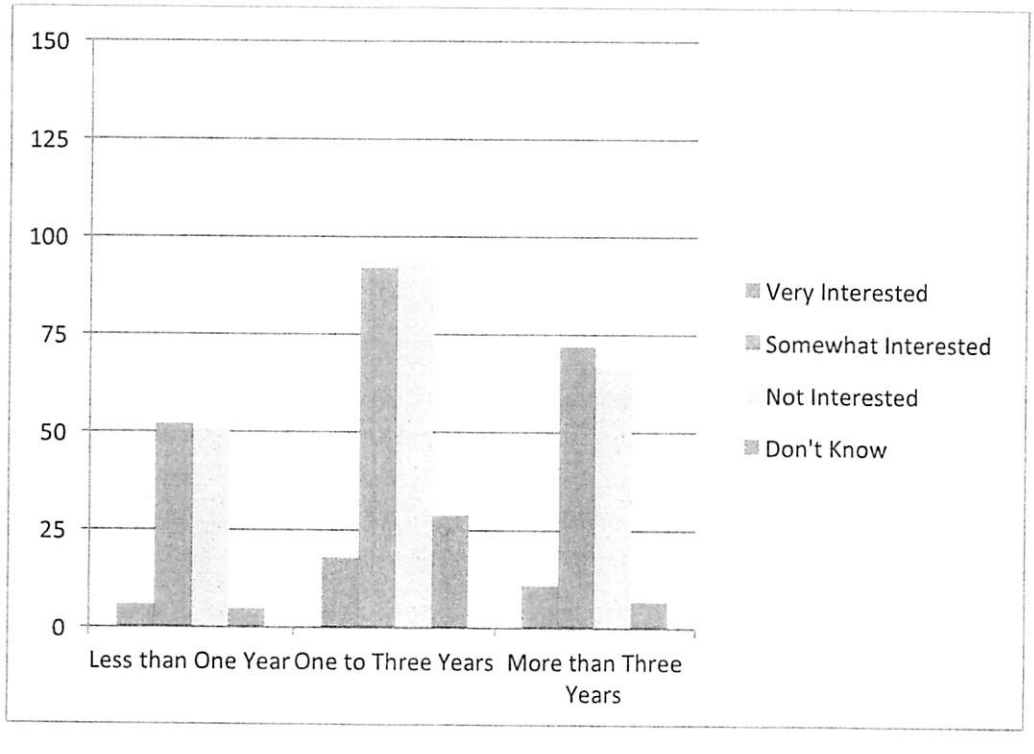
Where in Career vs. Interest Setting Aside Tuition



Years of Programming Experience vs. Interest Setting Aside Tuition



Years of Programming Experience vs. Interest Including the Cost of Tuition



The results of all the column charts are pretty clear. The level of interest in the Big Data Professional Masters Program definitely decreases after the cost of tuition is taken into account. This seems to be the case regardless of the respondent's chosen area of concentration, age, current position in career, and years of programming experience.

General Analysis

This section looks at the results of the survey and summarizes general findings. The data shows that a substantial majority of the respondents are concentrating in at least 'Applied Science and Technology' (80%). With such a large number, the results of the survey are heavily skewed to only evaluate this group of people. This is not too problematic, as SFU wishes to target this group of students for their program. Furthermore, most of the respondents are young, with 69.7% being between the ages of 19-24 and 23.2% being between the ages of 25-34. Overall, 97% of respondents are under the age of 35. Also, a high majority of the respondents are students (79.5%). A sizeable chunk (15.1%) is currently employed in a professional role in computing/software development. The years of programming experience seems to be fairly varied amongst the respondents; 23.1% of respondents have less than one year of programming experience, 45.4% of respondents have one to three years of programming experience, and 31.2% of respondents have more than three years of programming experience.

Furthermore, a big chunk of respondents (28.3%) were already somewhat familiar with the Big Data Program. The majority of respondents (51.7%) were instead not very knowledgeable about the program, but were interested in learning more about it. Most respondents felt that the cost of tuition of the program was too high. Before the cost of tuition was taken into account, 20.4% of respondents were very interested in the program, 50.3% of respondents were somewhat interested, and only 18% of respondents were not interested. After the cost of tuition was taken into account, only 6.7% of respondents were very interested in the program, 40.6% of respondents were somewhat interested,

and 39.3% of respondents were now not interested in the program. 47.2% of respondents instead felt that the cost of tuition should be under $18,000 and 36.2% believed that the cost of tuition should instead be in between $18,000 to $21,999. Regardless of fact that most people feel the tuition was too high, many respondents believed that it was important for a university in British Columbia to develop such a program; 32.8% of respondents felt that this type of program would be very important and 52.9% of respondents felt that the program would be somewhat important.

In regards to program format, respondents were basically split between wanting a full-time and part-time format, with slightly more people wanting a full-time format. However, in regards to the times and days respondents wanted the program to be implemented, opinions varied. 44.1% of respondents would prefer to have the program take place during regular business hours. 22.4% of respondents would prefer the program to take place on the weekend. Finally, 30.6% of respondents would prefer the program to take place on a combination of evenings and weekends.

Lastly, a strong majority of respondents would be much more likely to participate in the program if it included a paid internship (71.2%).

Limitations

The biggest limitation of this survey is that the majority of respondents were young students who concentrated in 'Applied Sciences and Technology.' As a result, most of the results of the survey were skewed to reflect the opinions of this group of people. This is not too problematic as SFU just needs to see if there is enough interest in the program amongst any group of people.

A limitation of this analysis is that for the data depicted by the clustered columns, students who have selected more than one discipline or have selected more than one current state of their career, have their opinions essentially expressed twice and a couple times, they have their views expressed three times. This could be remedied by splitting up the groups of people so that the combined

categories are also shown, however the problem here is that most of the combined categories were too small to have substantial finding on their own. However, leaving these categories of respondents isn't a big concern, as only a small group of respondents selected more than one option for these questions.

Implications

The results of the survey are mostly representing the views of young students who are concentrating in 'Applied Sciences and Technology.' These students generally seem to have at least some sort of interest in the program. The main factor that seems to be deterring these students is the cost of tuition which they feel is too high. This makes sense, as most of these students are not currently employed.

Appendix

Below are all the questions that were asked in the survey:

(Q1) What is/was your chosen field of study at SFU? (Please select all that apply)

(Q2) Please indicate the age group that you fall under?

(Q3) Where are you in your career? (Please select all that apply)

(Q4) How many years of programming experience (in school and/or work) do you have?

(Q5) How much did you know about Big Data before participating in this survey?

(Q6) Setting aside the cost of tuition, how interested are you in taking a Professional Master's degree program at SFU that is focused on Big Data?

(Q7) The proposed SFU Big Data program is a cost-recovery program. The current estimate is that the cost of delivering the program is approximately $30k per student. Taking into account the cost of the program, how interested are you in taking such a program?

(Q8) If the cost of the program was too high for you to be interested, what tuition rate would be acceptable to you?

(Q9) Based on your own experience, how important is it that a university in British Columbia develop a Professional Master's degree program focused on Big Data?

(Q10) If you are interested in the SFU Big Data program, which program format would you prefer?

(Q11) If you are interested in the SFU Big Data program, when would you prefer classes to be scheduled?

(Q12) Would you be more likely to enroll in the SFU Big Data program if it included a paid internship in the Big Data industry?

(Q13) Do you have any other comments or questions regarding the SFU Big Data program?

(Q14) Please leave your preferred contact email if you would like to participate in a draw to win one of the two iPad minis or a $20 gift card to Starbucks.