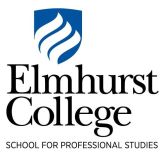
****

**MDS535: Programming Languages and Environments**

***Elmhurst College School for Professional Studies (SPS)***

**Facilitator and Contact Information**

David A. Ostrowski, Ph.d. , Adjunct Professor

Office Hours: E-Mail at any time, On Campus by Appointment

Phone Number: (248) 880-6632

Email: davidaostrowski@gmail.com

Website: [my.elmhurst.edu](http://www.elmhurst.edu/sps)

**MDS523 Course Description**

This course covers the application of appropriate high-level programming languages, for expressing software design patterns used for extracting and processing big data within the Hadoop environment. These high-level languages include imperative, object-oriented languages, including PIG, HIVE and Scala. Examples will also be presented in Java and Python. The languages will be presented in support of Big Data Processing relying on the map-reduce paradigm. Additional libraries will be explored in order to support activities of Data Mining as well as Machine Learning.

**Prerequisite Courses:**   *Course equivalent to CS 200 or working knowledge of a high-level programming language and/or query language; also, completion of the Data Science Curriculum prerequisites for fundamental programming and quantitative skills, if needed.*

**Course Outcomes**

At the completion of this course, students should:

* *Demonstrate understanding and competency of the Apache Hadoop Environment applied using Amazon Web Services.*
* *Express algorithms relying on the Map-Reduce paradigm to apply languages in the context of a parallel architecture using widely-used software engineering practices and respective high-level language pragmatics ;*
* *Write high-level programs that express algorithms used in accessing, extracting, and presenting of large data sets.*
* *Write appropriate queries in query languages (e.g. SQL) for purpose of data filtering as well as intermediate storage;*
* *Develop applications that will apply Statistical and Machine Learning techniques for purpose of Analytics.*

**Required Textbook for MDS535**

* **Hadoop Beginner’s Guide by Gary Turkington, Packt Publishing** ISBN-10: 1849517304ISBN-13: 978-1849517300, **© 2013**
* **Hadoop: the Definitive Guide by Tom White, OReilly pub.,** ISBN: 978-1-449-31152-0,**© 2012**

**(free online: http://ce.sysu.edu.cn/hope/UploadFiles/Education/2011/10/201110221516245419.pdf)**

**Additional Recommended Textbook for MDS535**

* Programming PIG by Alan Gates, Yahoo! Inc. ISBN-13: 978-1449302641

ISBN-10: 1449302645 **© 2011**

* Learning Spark: Lightning Fast Data Analysis by Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, OReilly, ISBN-13: 978-1449358624 , ISBN-10: 1449358624 **© 2015**

**Course Deliverables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Unit** | **All Students** | **Quiz** | **Activities** |
| 1 | Hadoop Fundamentals  Discussion Forum  Lab One,  Quiz 1, Section 1 | Quiz 1 | Lab Assignment One: installation of Hadoop, “Hello Hadoop” java demos  Discussion one |
| 2 | Map-Reduce Programming  Discussion Forum  Lab Two, Section 2, Quiz 2  Signup for Quiz with Proctor U | Quiz 2 | Lab Two: Streaming I/O using Python programming language  Discussion two |
| 3 | Introduction To PIG  Discussion Forum  Lab Three, Section 3  Quiz 3 | Quiz 3 | Lab Three : PIG programming  Discussion three |
| 4 | Advanced PIG  Discussion Forum  , Section 4  Midterm Exam | Quiz 4 | Lab Four :Advanced PIG programming  Discussion four |
| 5 | Introduction to HIVE  Discussion Forum  Data Project, Section 5 | Quiz 5 | Lab Four: HIVE Programming  Discussion five |
| 6 | Advanced HIVE  Discussion Forum  Data Project, Section 6 | Quiz 6 | Lab Six: Scala Programming  Discussion six |
| 7 | Introduction to Scala / Spark  Discussion Forum  Data Project, Section 6 Updated | Quiz 7 | Lab Seven : Scala/Spark  Discussion Seven |
| 8 | Advanced Scala, Spark ML  Discussion Forum  Data Project, All Sections  Final Exam | Quiz 8 | Lab Eight: Spark ML (Scala/Spark)  Discussion Eight |

**Discussion Forums**

Each week, a question or topic will be posted in the Discussion Forum. This key interactive learning activity is designed to promote class discussions on a key topic each week. Students are asked to create their initial post by Wednesday of each week and respond to at least one other student’s post by Saturday

**Course Assignments**

The deliverables and points breakdown is as follows:

1000 pts total

* Weekly Assignments – due at conclusion of each module (44% of grade 55 pts \* 8 )
* Weekly Quizzes  – due at the conclusion of each module (32% of grade 50 pts \* 8 )
* Weekly Discussions  - due at the conclusion of each module (24% of grade 30 pts \* 8)

**Tentative Course Schedule, Schedule of Assignments, and Assessments:**

|  |  |  |
| --- | --- | --- |
| **UNIT** | **Learning Outcomes** | **Learning Activities, Assignments, and Assessments** |
| **01** | • Develop an understanding of Apache Hadoop architecture and setup  • Comprehend the application of the Amazon Web Services  • Describe basic Hadoop based introductory examples (hello hadoop)  • Demonstrate an understanding of Hadoop File System (HDFS) in the Apache Hadoop environment. | **Readings (1):**  • Chapter 01, 02 (Turkington)  **Weekly Adobe Connect Session.**  • Topic: Signing up for Amazon AWS, Installation of Apache Hadoop  **Powerpoint.**   * **Introduction to AWS,** * **Installation of Apache Hadoop**   **Assignments and Assessments**   * **Discussion Forum** * **Lab 1** * **Quiz 1** |
| **02** | • Justify the application of Map-Reduce Paradigm.  • Solve concurrent problems through the application of Hadoop Streaming I/O.  • Assess Map-Reduce Programming applications..  • Describe basic Map-Reduce examples provided in the Apache Hadoop installation.  • Assemble Map-Reduce - based applications within a language of your own choice.  • Critique the utility of the Map Reduce Paradigm in the context of the Hadoop architecture | **Readings (1):**  • Chapter 03: Map Reduce  • Chapter 04: Developing Map-Reduce Programs, (Turkington)    **Weekly Adobe Connect Session.**  • Topic: Map-Reduce  **Assignments and Assessments:**   * Discussion Forum * Unit Quiz 2 * Lab Two |

|  |  |  |
| --- | --- | --- |
| **UNIT** | **Learning Outcomes** | **Learning Activities, Assignments, and Assessments** |
| 03 | • Demonstrate competency of writing programs using apache PIG  • Apply the LOAD command in PIG programming languauge  • Construct filter jobs leveraging the foreach command in PIG programming language  • Arrange data retrieval applications through the application of the OrderBy, Distinct and Join commands in PIG programming language | **Readings:**  **Chapter 1,2,3,4,5 ( Programming PIG by Alan Gates, Yahoo! Inc. ISBN-13: 978-1449302641**  **ISBN-10: 1449302645 © 2011)**  **Week Three Notes (powerpoints)**  **Weekly Adobe Connect Session.**  • Topic: PIG programming  **Activity three**  **Quiz three**  **Discussion Three** |
| 04 | • Compute solutions in the Apache Hadoop environment relying on advanced features of foreach in Apache Pig.  • Analyze the process of Integrating Pig with legacy Code and MapReduce.  • Demonstrate an understanding of how to develop and test Pig Scripts. | **Readings:**  **• Chapter 04, 05, 06, 07, ( Programming PIG by Alan Gates, Yahoo! Inc. ISBN-13: 978-1449302641**  **ISBN-10: 1449302645 © 2011)**  **Week Four Notes (powerpoints)**  **Weekly Adobe Connect Session.**  • Topic:Advanced PIG  **Activity Four**  **Quiz Four**  **Discussion Four** |

|  |  |  |
| --- | --- | --- |
| **UNIT** | **Learning Outcomes** | **Learning Activities, Assignments, and Assessments** |
| 05 | • Develop the schema for a HIVE database.  • Compare means of access to HIVE tables.  • Demonstrate the import of data into HIVE tables. | **Readings:**  **• Chapter 8 ( Hadoop Beginner’s Guide by Gary Turkington, Packt Publishing ISBN-10: 1849517304ISBN-13: 978-1849517300© 2013)**  **Weekly Adobe Connect Session.**  • Topic: Introduction to HIVE  **Activity Five**  **Quiz Five**  **Discussion Five** |
| 6 | • Solve advanced data retrieval issues using advanced queries in HIVE.  • Explain the basic syntax of Scala.  • Paraphrase an understanding of functional programming.  • Demonstrate an understanding of database administration in the context of the Apache Hadoop environment. | **Readings:**  **• Chapter 12, Hadoop: the Definitive Guide by Tom White, OReilly pub., ISBN: 978-1-449-31152-0,© 2012**  **(free online:** [**http://ce.sysu.edu.cn/hope/UploadFiles/Education/2011/10/201110221516245419.pdf**](http://ce.sysu.edu.cn/hope/UploadFiles/Education/2011/10/201110221516245419.pdf)**)**  **• Chapter 1 (Learning Spark: Lightning Fast Data Analysis by Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, OReilly, ISBN-13: 978-1449358624 , ISBN-10: 1449358624 © 2015 )**  **Weekly Adobe Connect Session.**  • Topic: Scala Programming  **Activity Six**  **Quiz Six**  **Discussion Six** |

|  |  |  |
| --- | --- | --- |
| **UNIT** | **Learning Outcomes** | **Learning Activities, Assignments, and Assessments** |
| 7 | • Construct and apply solutions to Big Data applications on the spark platform.  • Develop applications in Scala programming language on the spark platform.  • Produce applications in the Python programming language using pySpark on the Apache Spark platform.  • Construct programming examples in Java language on the Apache Spark platform. | **Readings:**  **• chapter 1,2,3,4 : Learning Spark: Lightning Fast Data Analysis by Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, OReilly, ISBN-13: 978-1449358624 , ISBN-10: 1449358624 © 2015**  **Weekly Adobe Connect Session.**  • Topic: Apache Spark platform  **Activity Seven**  **Quiz Seven**  **Discussion Seven** |
| 8 | • Demonstrate the application of statistical and machine learning algorithms including nearest neighbor, clustering and logistic regression using Spark ML. | Readings:  • chapter 9,10,11: Learning Spark: Lightning Fast Data Analysis by Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, OReilly, ISBN-13: 978-1449358624 , ISBN-10: 1449358624 © 2015  **Weekly Adobe Connect Session.**  • Topic: Libraries on the Apache Spark Platform.  **Activity Eight**  **Quiz Eight**  **Discussion Eight** |

**Webinar Schedule**

Each Wednesday there will be a webinar using the Adobe Connect platform. The webinars will start at 8:00 PM CST and will be recorded for those unable to attend the sessions live. You must use the appropriate URL to enter into the virtual meeting room. You can find the URL within each unit under the Weekly Webinar heading.

To join the webinar:

1. Click the **URL** contained inside the online classroom.
2. Select **Guest** and enter your full name.
3. Check the audio on your PC and be sure your speaker volume is at an adequate level.
4. Use the interactive tools to actively participate in class discussions.

The primary topic for all webinars will be:

1. The Database and Data Warehouse Project
2. Q&A from e-mails received during the week
3. Open Q&A on anything related to the course, Elmhurst College, SPS, or virtually any other topic of interest

**Grading Scale**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Grading Component** | **Points** | | Activities | 440 | | Quizzes | 320 | | Discussion Forums | 240 | |  |  | |  |  | |  |  | |  |  | | Total Points | 1000 | | |  |  |  | | --- | --- | --- | | **Points** | **Grade** | **GPA** | | 940-1000 | A | 4.0 | | 900-939 | A- | 3.7 | | 870-899 | B+ | 3.3 | | 830-869 | B | 3.0 | | 800-829 | B- | 2.7 | | 770-799 | C+ | 2.3 | | 730-769 | C | 2.0 | | 700-729 | C- | 1.7 | | 600-699 | D | 1.0 | | 000-599 | F | 0.0 | |

**Format of Exams**

The Midterm and Final Exam will be a combination of objective questions, i.e. Multiple Choice, True False, and Matching, and Short Answer. Weekly quizzes provide sample questions and should be used to study for the exams.

**Due Dates and Policy Regarding the Submission of Late Assignments**

Assignments (Projects) are due by 11:59 PM on Saturday of the week they are assigned. There is a 10% penalty per day (up to 11:59 PM of the day after the due date) for late assignments, up to a maximum of 50%. Late assignments after 50% will be accepted up to 11:59 PM of the last day of class; however, since most assignments build on the previous assignment, it is critical to keep up with the assigned work each week.

**Extra Credit**

There are no extra assignments available in this course.

**Elmhurst College Policies and Procedures**

**Statement Regarding Netiquette**

Netiquette refers to Online Etiquette, and it is the expectation of the College that all students will use appropriate online etiquette for interaction in online courses and using online communication tools, such as email and instant messaging. Common guidelines include:

* Avoid using strong or offensive language;
* Be forgiving. If something offends you, work with the instructor regarding the incident, rather than engaging in further enflamed discussion;
* Remember that your online messages and course postings live forever, so proof-read your work before you press Send.
* Avoid typing ALL IN CAPS. This is considered “shouting” online.
* Be concise; your points may be lost in a lengthy text;
* Give credit where credit is due: Use citations as appropriate;
* Avoid using Reply All in email messages.

**Statement of Disability**

The College will make reasonable accommodations for persons with documented disabilities.  If you have a disability that may have some impact on your work in this course, please contact the Disability Services Provider at 630-617-3753.

**Code of Academic Integrity**

Elmhurst College is a community of scholars.  Such a community places the highest value on academic honesty and integrity.  Therefore, Elmhurst College’s Code of Academic Integrity is designed to ensure:

1. A fair academic environment where students are evaluated on the scholarly merits of their work.
2. An academic environment where person integrity is fostered.
3. An understanding of standard academic practices when formulating ideas, conducting research, and documenting sources.

It is the responsibility of each participant to know and understand this academic integrity policy.  There are general campus-wide definitions of academic honesty as well as department guidelines.  Participants are responsible for educating themselves on this matter since sanctions for academic dishonesty may be severe.  Thus, Elmhurst College participants in this seminar should act honestly in all academic work.  Please refer to the[***Elmhurst College E-Book***](http://media.elmhurst.edu/documents/EBook_2011_2012.pdf) for additional information.

**Statement Regarding the Use of Plagiarism Detection Services**

Faculty members may at any time submit a participant’s work for review by a plagiarism detection service.

**Proctor-U**

To ensure the academic integrity of online courses, all Elmhurst College Online Center students are required to complete one proctored test per course using Proctor-U. Online proctors will check photo IDS, ask students to complete a short authentication quiz, and will start the quiz for the students by using a password provided by the faculty member. In order to take the quiz, students must pass the identity authentication portion of the process. In the event that a student fails to schedule and complete the required proctored quiz, fails to pass the authentication quiz, fails to provide photo identification, or demonstrates an attempt to subvert the integrity of the quizzing process, he/she will be required to meet with the Dean, School for Professional Studies after which a determination will be made as to whether or not a violation of the Code of Academic Integrity has occurred. Refer to the Elmhurst College E-Book for more information.

**Non-Discrimination/Non-Harassment Policy**

The College believes that all employees and students should be able to work and learn in an educational environment free from discrimination and harassment.  Harassment includes unwelcome conduct, whether verbal, physical, written or graphic, that is based on protected group status such as race, color, national origin, religion, sex, sexual orientation, gender identity or expression, age, disability, genetic information, veteran’s status, or other protected status.  Please refer to the[***Elmhurst College E-Book***](http://media.elmhurst.edu/documents/EBook_2011_2012.pdf) for additional information.

**Code of Conduct**

The College attempts to establish a climate which encourages the assumption of responsibility by participants for their individual and collective actions with a minimum of rules and regulations.  It shall be the responsibility of each seminar participant student to comply with appropriate standards of conduct and decorum befitting a member of an educational community.  Please refer to the[***Elmhurst College E-Book***](http://media.elmhurst.edu/documents/EBook_2011_2012.pdf) for additional information.