

George Mason University | School of Computing
Department of Information Sciences and Technology

AIT 614: *Big Data Essentials*
Course Syllabus

For each section, a customized syllabus with information specific to that section will be made available to registered students via the [Blackboard Learning System](#).

Course Description

AIT Applied Information Technology
614 Big Data Essentials

Pre-requisites:

Graduate standing
AIT 524 (Recommended)

Catalog Description:

Students participate actively through class dialogues and the crafting of IT solutions to specific problem areas. Course cannot be used to satisfy course requirements for PhD IT students.

Special Topics Description:

This is a **hands-on** course that discusses emerging technologies for big data analytics and their applications in the real-world environment. It helps students gain *essential **industry-ready skills** effectively*. Students apply learned concepts and best practices using a large variety of emerging technology tools by following detailed **step-by-step** instructions to implement a series of **hands-on labs** on the *modern **Databricks** cloud platform* and then employ these techniques to develop a **final project**, a functional big data analytical system. The topics include, but are not limited to:

- **Big Data Analytics Concepts and Architectures**
 - Big Data Analytics, RDBMS, NoSQL Databases
 - Hadoop Ecosystem, Stream Data Analytics
 - Advanced In-DB Data Analytics & Machine Learning
 - Data Warehouse, Data Lake, Statistics for Data Analysis
 - Big Data Security, Big Data in Cloud, Internet of Things (IoT)
- **Big Data Essential and Modern Tools**
 - SQL, MongoDB, PyMongo – MongoDB with Python
 - Databricks (*Microsoft Azure, AWS, Google Cloud*)
 - Databricks SQL Query and Visualization
 - Hadoop, HDFS, HIVE, MapReduce Programming Paradigm
 - Spark, PySpark - Spark with Python
 - Spark MLlib for Machine Learning
 - Microsoft Azure, Raspberry Pi Simulator to Azure IoT Hub
- Other technologies in the Big Data ecosystem for unstructured data analytics.

Rationale

This course is designed to provide an understanding of the impetus, value proposition, and trends of emerging technologies in the big data analytics field. The course will help students fully understand the usage, opportunities, and challenges of emerging big data technologies and their applications in decision making. Emphasis for the course will be placed on a real-world orientation through the analysis of case studies and hands-on activities that emphasize the importance of a comprehensive approach in big data analytics. The course will be in demand by students working in or interested in working with big data, which is one of the fastest growing specializations within IT. As part of this specialization, students need to learn how emerging big data analytics technologies are continuing to transform the IT industry, including their use in government programs.

Objectives

On successful completion of this course, students will be able to:

- Understand the importance of Data Science and Big Data Analytics in decision-making
- Perform analytics on unstructured data using Big Data technologies, including Hadoop, HDFS, HIVE, Spark, and others
- Use NoSQL for Big data database management and basic analysis
- Use *modern Databricks* as a Big Data cloud development platform
- Use Databricks SQL to query data and visualize queries in a dashboard
- Use Spark and Spark MLlib for data queries and data analytics
- Use R or Python for statistical analysis
- Use Microsoft Azure Virtual Machine (VM) for stream data analytics and Internet of Things (IoT)
- Understand Big Data security, and how Big Data is used in combination with Cloud and IoT

References

Textbooks and Reference Materials

There are no required textbooks for this course. Students are recommended to use the following books for reference:

- EMC open-source materials on Data Science and Big Data Analytics
- A lot of tutorials, code examples, and other learning resources in Blackboard

Faculty

AIT 614 Instructor: as shown in the Course Registration Site

AIT 614 Course Coordinator: Lindi Liao, Ph.D. dliao2@gmu.edu

Course Duration

Dates: GMU Academic Calendar: <https://registrar.gmu.edu/calendars/>

Total Duration: **16 weeks** or **8 weeks**

Grading Policy

Student grades will be determined based on the following components¹:

| Grade Component | Weight |
|---|---------------|
| Discussion Boards including self-introduction | 10% |
| Written Assignments | 25% |
| Labs | 20% |
| Midterm (1.5 hours, online, open book/notes/Internet) | 15% |
| Big Data Project ² | 30% |
| <i>Extra Credit</i> | TBD |

¹Subject to revision before and throughout the course

²Teamwork (4-5 students)

Class Participation

This course is designed for an *in-person*, "*synchronous*", or "*asynchronous*" delivery (<https://masononline.gmu.edu/course-delivery-methods>). Please go to **Blackboard/Welcome to AIT 614** for more details.

All assignments, assessments, class announcements, schedules, files, and presentations will use Blackboard. Additionally, students need to contribute actively and participate in **online discussions** on Blackboard for grading.

Schedule

A detailed schedule will be published on Blackboard. As many factors may affect the development and progress of a class, the instructor reserves the right to alter the schedule as may be required to assure attainment of course objectives. The schedule is subject to revision before and throughout the course.

Registered students should see the Blackboard Learning System for the latest class schedule.

Grading

Grades will be awarded in accordance with the Mason Grading System for graduate students. See the university catalog for policies: <http://catalog.gmu.edu> for more information.

Grading Scale

The grading scale for this course is:

| | | |
|------------|----|---------|
| 97 – above | A+ | Passing |
| 93 – 96% | A | Passing |
| 90 – 92% | A- | Passing |
| 87 – 89% | B+ | Passing |
| 83 – 86% | B | Passing |
| 77 – 82% | B- | Passing |
| 70 – 76% | C | Passing |
| 0 – 69% | F | Failing |

NOTE: Study success takes constant effort!

Students are responsible for checking the currency of their grade books. Grade discrepancies must be brought to instructor's attention within one week of assignment submission, 48 hours of exam submission and 72 hours of final course score grade posted on the Grade Center of the current course in Blackboard.

No make-up for any activity including exams, unless arranged in advance. Only in special cases, such as medical problems and family emergency, make-ups and late assignments may be allowed with verifiable proof. Arrive promptly to exams. Late students may not be admitted.

Final letter grades will be posted to [PatriotWeb](#), which is the only vehicle for students to obtain those grades. A student with a "hold" on his/her PatriotWeb account will be unable to access final grades until the hold has been removed by the Registrar.

Important Dates

Dates for dropping, adding the course etc. are available via <http://registrar.gmu.edu/calendars/>

Religious Holidays

A list of religious holidays is available on the [University Life Calendar](#) page. Any student whose religious observance conflicts with a scheduled course activity must contact the Instructor **at least 2 weeks in advance** of the conflict date in order to make alternative arrangements.

Attendance Policy

Students in in-class sections are expected to attend each class, to complete any required preparatory work and to participate actively in lectures, discussions and exercises. As members of the academic community, all students are expected to contribute regardless of their proficiency with the subject matter.

Students are expected to make prior arrangements with Instructor if they know in advance that they will miss any class and to consult with the Instructor as soon as possible if they miss any class without prior notice. Any student who expects to miss more than one class session is strongly advised to drop the course and take it in a later semester when he/she can attend every class.

However, the policy for some courses will vary, so please refer to the class attendance requirements on Blackboard.

Departmental policy requires students to take exams at the scheduled time and place, unless there are truly compelling circumstances supported by appropriate documentation. Except in such circumstances, failure to attend a scheduled exam will result in a score of zero (0) for that exam, in accordance with [Mason policy on final exams](#). Students should not make travel plans or other discretionary arrangements that conflict with scheduled classes and/or exams. If the University is closed due to weather or other unforeseen conditions, final exams may be rescheduled – students are strongly advised not to make plans that would prevent them from attending exams that may be rescheduled during the entire exam period.

NET Version attendance: During each week the students must perform all the requirements published for that week. A detailed week-by-week schedule of classes will be published on the net version of the course.

Classroom Conduct

Students are expected to conduct themselves in a manner that is conducive to learning, as directed by the Instructor. Any student who negatively impacts the opportunity for other students to learn will be warned – if disruptive behavior continues, the student will be asked to leave the classroom.

Electronic devices are potential distractions in the classroom environment. Cell phones, pagers and other handheld devices must be turned off or set to "silent" mode and not used while class is in session. Laptop computers and similar devices may be used only if such use is directly related to the classroom activity in progress – for some activities the Instructor may require that such devices not be used in order to maximize student engagement.

Communications

Registered students will be given access to a section of the [Blackboard Learning System](#) for this course. Blackboard will be used as the primary mechanism (outside of lectures) to disseminate course information, including announcements, lecture slides, homework and other assignments, and scores for homework and exams.

Communication with the Instructor on issues relating to the individual student should be conducted using Mason email, via telephone, or in person - **not** in the public forums on Blackboard. Mason Mail is the preferred method – for urgent messages, you should also attempt to contact the Instructor via telephone. Federal privacy law and Mason policy require that any communication with a student related in any way to a student's status be conducted using secure Mason systems – if you use email to communicate with the Instructor you **MUST** send messages from your Mason email account.

When sending an e-mail to the instructor, please include the following:

- **Course number**
- **Section number**
- **Your full name**

Privacy

Instructors respect and protect the privacy of information related to individual students.

As described above, issues relating to an individual student will discussed via email, telephone or in person. Instructors will not discuss issues relating to an individual student with other students (or anyone without a need to know) without prior permission of the student.

Assessable work other than final exams will be returned to individual students directly by the Instructor (or by a faculty or staff member or a Teaching Assistant designated by the Instructor, or via another secure method). Under no circumstances will a student's graded work be returned to another student.

Faculty and staff will take care to protect the privacy of each student's scores and grades.

Disability Accommodations

[The Office of Disability Services \(ODS\)](#) works with disabled students to arrange for appropriate accommodations to ensure equal access to university services. Any student with a disability of any kind is strongly encouraged to register with ODS as soon as possible and take advantage of the services offered.

Accommodations for disabled students **must** be made in advance – ODS cannot assist students retroactively, and at least one week's notice is required for special accommodations related to exams. Any student who needs accommodation should contact the Instructor during the first week of the semester so the sufficient time is allowed to make arrangements.

Using Generative-AI Models*

ChatGPT or other Generative-AI models **may be used** in this course ONLY as an **assistant** in projects and homework assignments. Any use must follow the fundamental principles of the [Honor Code](#) and include the following statement with assignment submission: **The ideas in this submission are original and were generated by (my name). ChatGPT (or name other Generative-AI model) was used as an editorial/coding assistant, however, I take full responsibility for the originality and accuracy of the content.**

Risk accompanies use of any powerful tool. Students are cautioned that sharing their own original ideas with Generative-AI models can lead to loss of control and ownership of those ideas and coding. Furthermore, in terms of learning in this class, students who replace their own learning and project work with materials prepared by Generative-AI models:

- Surrender control over the material's truthfulness and accuracy, and violate the university's [Honor Code](#).
- Sacrifice the opportunity to acquire the knowledge, skills, and critical thinking taught in the course.
- Risk being unable to perform to expectations when Generative-AI models are unavailable, such as in exams.
- Ultimately endanger their employability if they are unable to produce work other than that produced by Generative-AI models.

** D. J. Goodings, J. Nelson, and I. Rytikova. "Stating guidelines for using ChatGPT and other Generative-AI models in F2023 CEC course syllabi".*

Honor Code

All members of the Mason community are expected to uphold the principles of scholarly ethics. Similarly, graduating students are bound by the ethical requirements of the professional communities they join. The ethics requirements for some of the communities relevant to IST graduates are available via the following links:

- [ACM Code of Ethics and Professional Conduct](#)
- [IEEE Code of Ethics](#)
- [EC-Council Code of Ethics](#)

On admission to George Mason University, students agree to comply with the requirements of the [Mason Honor System and Code](#). The Honor Code will be strictly enforced in this course. Honor Code cases are heard by a panel consisting of students – students who meet the requirements are encouraged to nominate themselves to serve on the Honor Committee.

Any use of the words or ideas of another person(s), without explicit attribution that clearly identifies the material used and its source in an appropriate manner, is **plagiarism** and will not be tolerated. The Instructor reserves the right to use manual and/or automated means (including such services as [SafeAssign.com](#)) to detect plagiarism in any work submitted by students for this course, and to direct Teaching Assistants and/or other faculty and/or staff members to do likewise in support of this course.

For this course, the following requirements are specified:

- All assessable work is to be prepared by the individual student, unless the Instructor explicitly directs otherwise.
- All work must be newly created by the individual student for this course for this semester.
- Any usage of work developed for another course, or for this course in a prior semester, is strictly prohibited without prior approval from the instructor.

Students may seek assistance with assigned work (and are encouraged to do so if they feel the need), **provided**:

- The directions for the assigned work do not prohibit such assistance.
- Such assistance is acknowledged in the submitted work, clearly identifying the person(s) giving assistance and the nature of the assistance given.
- Any work to be submitted is prepared entirely and exclusively by the student submitting it. Students are expressly prohibited from sharing any assessable work for this course in any manner with other students (except students assigned as Teaching Assistants to this course and the student's section), unless all students involved have had their work graded and returned by the Instructor, or the Instructor has explicitly approved such sharing.

NET Sections only:

For the net section you will need to have the following additional resources:

- Computer with fast internet connection
- Microphone and web camera (optional)