

D.Sc. in IT Qualifying Examination Guidelines

1. The exam will cover the following *Nine* areas:

- A. Operating Systems
Study guide-COSC519.
- B. Data Structures and Algorithms
Study guide-COSC600.
- C. Database Management Systems
Study guide – COSC 578.
- D. Data Communications and Networking
Study guide – COSC 650.
- E. Software Engineering
Study guide – COSC 612.
- F. Project Management
Study guide --- COSC 609.
- G. Human Computer Interaction
Study guide – COSC 605.
- H. Computer Security
Study Guide—COSC734.
- I. Big Data Analytics
Study Guide---COSC760 Big Data Analytics.

2. Date of the exam and the nature of the exam.

The D.Sc. in IT qualifying exams will be written exams and will be offered twice a year, once in the First week of January, and once in the Last week of May. You can attempt partial or all parts in the exam.

3. Evaluation Procedure

- Four areas (out of 9; the number of sections may change in the future) must be passed to pass the qualifying examination. Each subject has a Pass or Fail status.
- A student can take the Qualifying Exam at most two times. The student must pass at least 4 sections (as indicated in the previous paragraph) in the two attempts combined; Otherwise, the committee will recommend the student for dismissal from the program.
- A student must pass the qualifying exam within the first 4 years in the program. If a student does not pass the qualifying exam within the time period, the committee will recommend the student for dismissal from the program.

4. The students will be notified of results in about two weeks after the exam.

Detail Topics for Each Subject Area

You must pass at least 4 out of 9 areas to pass the qualifying exam.

(1) Data Structures and Algorithms

General Information:

- The topics below are covered at different level of details in the following courses offered by our department: COSC 600.
- These topics are standard and are covered in many textbook on algorithms and data structures. We recommend the following textbooks:
 - (W) *Data Structures and Algorithm Analysis (C++ or Java editions)*, by Mark A. Weiss, Addison-Wesley Publishing Co.
 - (CLRS) *Introduction to Algorithms* by Cormen, Leiserson, Rivest, Stein, published by MIT Press and McGraw-Hill.

Topics:

- 1 Basics of Algorithm Analysis:
 - Asymptotic notation (Big-O, little-O, Big-Omega, basic manipulation rules) – W (Chapt. 2), CLRS (Chapt 3)
 - Analysis of recursive algorithms, Master Theorem – CLRS (Chapt 4)
- 2 Linear Data Structures: Lists, Stacks, Queues - W (Chapter 3), CLRS (Chapter 10)
 - Abstract data type characterization
 - Implementations (array, linked lists, doubly linked lists, circularly linked lists)
- 3 Trees
 - Binary trees, Binary search trees (W – Chapter 4)
 - AVL trees (W – Chapter 4)
 - Red-black trees (CLRS – Chapter 13)
 - Splay Trees (W – Chapter 4, Section 4.5)
 - B-trees (W – Chapter 4, Section 4.7)
- 4 Hashing (W – Chapter 5)
 - hash functions, open addressing hashing
 - extendible hashing
- 5 Heaps and Priority Queues
 - binary heaps (W – Chapter 6)
 - leftist heaps (W – Chapter 6, Section 6.6)
 - binomial heaps (W Section 11.2, CLRS – Chapter 19)
- 6 Sorting Algorithms:
 - insertion sort, select sort, bubble sort (W – Chapter 7)
 - heapsort (W – Section 7.5, CLRS – Chapt. 6)

- mergesort (W – Section 7.6, CLRS – Section 2.3.1)
- quicksort (W – Section 7.7, CLRS – Chapt. 7)
- bucket sort, radix sort (CLRS – Chapter 8)
- external sorting (W – Section 7.11)

7 Graph Algorithms:

- data structures for representing graphs (adjacency matrix, adjacency list) (W – Section 9.1.1, CLRS – Section 22.1)
- Breadth first search (W- Chapter. 9, CLRS – Section 22.2)
- Depth-first search (W – Section 9.6, CLRS – Section 22.3)
- Topological sorting (W – Section 9.2, CLRS - Section 22.4)
- Shortest-path algorithms (Dijkstra, Floyd-Warshall) (W – Section 9.4, CLRS – Chapt 24)
- Minimum spanning tree algorithms (Kruskal, Prim) (W – Section 9.5, CLRS – Chapt. 23)
- Network flows ; maximum-flow algorithm (W – Section 9.4.1)

8 General techniques in the design of algorithms (W – Chapter 10, CLRS – Chapters 15 and 16)

- Divide-and-conquer
- Dynamic programming
- Greedy strategies
- Backtracking
- Branch-and-bound

9. Basic notions of computational complexity (W – Section 9.7, CLRS – Chapter 34)

- lower bounds for sorting
- reductions among problems
- NP and NP-complete problems

(2) Operating Systems

General Information:

- Most of the topics below are covered in an introductory graduate course on operating systems, COSC519 offered at Towson University.
 - The following text book or equivalent should be studied for the given topics.
Operating System Concepts, Eight Edition, Silberschatz, Galvin, and Gagne, Wiley Publications.
- Operating Systems Essentials 2nd Edition Silberschatz, Galvin and Gagne, Wiley Publications

Topics:

1. Fundamentals
 - o Types of Systems
 - o Computing Environments
 - o Storage Hierarchy
 - o System Calls
 - o System Structures
 - o Virtual Machines (Java)
 - o Interrupts
2. Process Management
 - o Process Scheduling
 - o IPC
 - o Client/Server Systems
3. Threads
 - o Multithreading
 - o Difference between processes and threads
 - o Threading Issues
4. CPU Scheduling
 - o Scheduling Criteria
 - o Scheduling Algorithms
5. Process Synchronization
 - o Critical Section
 - o Synchronization using flags
 - o Hardware Methods (Test and Set)
 - o Semaphores and examples
6. Memory Management
 - o Swapping
 - o Contiguous Memory Allocation
 - o Paging, TLB
 - o Segmentation
 - o Segmentation with paging
7. Virtual Memory
 - o Demand Paging
 - o Page Replacement
 - o Allocation Frames

- o Thrashing

8. File-system Implementation

- o File System Structure
- o File System Implementation
- o Directory Implementation
- o Allocation Methods

9. Mass Storage Structure

- o Disk Structure
- o Disk Scheduling

10. I/O Systems

- o I/O Hardware
- o Application I/O Interface
- o Kernel I/O subsystem
- o Streams

11. The Linux System

- o Kernel Modules
- o Process Management
- o Scheduling
- o Memory Management
- o File Systems

12. Basic Concepts (Can be studied from any source)

- o Script files
- o Batch files
- o Compilation and Linkage
- o Boot programs
- o Loader programs
- o Command Line Interpreters
- o Shell Commands
- o General OS Concepts
- o Current trends in OS

(3) Computer Networks

General Information:

- Most of the topics below are covered in introductory graduate courses on computer networks or data communications, COSC 650.
- While many books and websites cover these topics, we recommend the following books that are used in the above course (COSC 650).
 - (PD) Larry L. Peterson and Bruce S. Davie, *Computer Networks (4th edition)*, Morgan Kaufmann, 2007.

Topics:

- 9 Network Architecture (PD 1.3.1-1.3.3; S 5.1-5.3, 5.5; C 10.1-10.7.1, 10.9-10.11)
 - Layers and Protocols
 - OSI Architecture and TCP/IP Architecture
- 10 Network Performance Calculations (PD 1.5.1-1.5.4; S 2.1-2.5)
 - Transmit Time
 - Data Rates Required to Support Applications
- 11 Data Communications (PD 2.3.3, 2.4.3; S 17.5, 16.4)
 - SONET
 - Error Detection and CRC
- 12 Flow Control and Reliability (PD 2.5.1, 2.5.2; S 17.1; C 12.4, 12.5)
 - Stop-and-Wait (simple positive acknowledgement with retransmission)
 - Sliding Window
- 13 LAN Technologies (PD 2.6.2, 2.8.2, 3.2.1; S 10.2, 10.3, 11.2; C 2.4.5, 2.4.6, 2.4.8, 2.4.9, 2.4.10, 2.5)
 - Ethernet Frame Format and Address
 - 802.11 Architecture and Distribution System
 - 802.11 Collision Avoidance and Reliable Data Delivery
 - CSMA and Backoff
 - Learning Bridges and LAN Switches
- 14 IP (PD 4.1.2-4.1.5, 4.2.3, 4.3.1, 4.3.2; C 4.3-4.5, 4.11, 4.16, 5.5-5.12, 6.2-6.7.1, 6.7.3-6.7.8, 7.2-7.10, 9.2, 9.3, 9.5-9.8, 9.10-9.12, 9.16-9.20.1, 13.9, 15.15-15.15.5)
 - IPv4 Header Format and IPv4 Addresses
 - IP Forwarding
 - IP Fragmentation
 - ARP
 - Link State Routing and OSPF Protocol
 - Subnetting and CIDR
- 15 UDP and TCP (PD 5.1, 5.2.1-5.2.7, 6.2.1, 6.3.1, 6.3.2, 6.4.2; C 11.3-11.9, 12.2-12.19, 12.22-12.33.3)
 - UDP Header Format and Checksum
 - TCP Header Format
 - Connection Establishment/Closing and State Transitions
 - Reliable Delivery and Flow Control
 - Timeout and Adaptive Retransmission Algorithms
 - Congestion Control and Avoidance

(4) Fundamentals of Software Engineering

Suggested Textbooks

- Ian Sommerville's "Software Engineering" by Addison Wesley

Suggested topics

- Software Engineering history
- Software Life-cycle Models
 - Major software processes
 - Software Process Models
- Software Quality Factors:
 - Attributes of good quality software
- Software Requirement Acquisition and Analysis
 - Requirement types
 - Requirement Elicitation and Analysis process
- Software Requirements Specification (SRS) Tools
 - Data Dictionary,
 - Data Flow Diagram,
 - Process Specification
 - Entity-Relationship Diagram,
 - UML: Use Cases, Sequence Diagram, State Chart
- Software Design
 - Architectural Design
 - Types of software architecture
 - Detailed Design
- Object Oriented Analysis and Design
 - Object Characteristics
 - Object Associations and Hierarchy
 - Services and Attributes of objects
- Software Metrics
 - Size related metrics: source lines of code
 - Complexity related metrics: Function point, Object Point, Cyclomatic Complexity
- Software Configuration Management
 - What to maintain in software
 - Change management
- Software Testing
 - Dynamic / Static
 - Black box/ White Box
 - Test Plan
 - Test Case

- Software Cost Estimation
 - Software Attributes
 - Software Cost Estimation Models

- Software Maintenance
 - Maintenance process, activities
 - Maintenance effort
- Software Reuse
- Software Re-engineering

(5) Database Management Systems

General Information:

- Most of the topics below are covered in graduate courses on database management systems I, and II. Computer science courses are COSC578.
- While many books and websites cover these topics, we recommend the following books that are used in the respective courses above.
 - (EN) Elmasri and Navathe, Fundamentals of Database Systems, 7th Edition, Addison-Wesley.
 - (CB) Thomas Connally, and Carolyn Begg, Database Systems, A Practical Approach to Design, Implementation, and Management, 6th Edition, Pearson.

The following topics are covered in both the books.

Topics:

1. Files versus Databases (Comparison)
2. Three Schema Architecture or Three-Level ANSI-SPARC architecture
3. Data Independence
4. Database Languages, DDL, DML, DCL.
5. Data Modeling Using E-R Models (Entities, Relationships, Cardinalities), Chen's Model or UML model representation which ever is familiar to you.
6. The Relational Algebra, Selection, Projection, and Join operations. Representation of queries using relational algebra. Database Constraints.
7. SQL: Simple queries, and nested queries (Data Manipulation Only)
8. Normalization, Normal forms 1, 2, 3.
9. Disk storage, file structures, hashing, and indexing (CH13, CH14 from Elmasri & Navathe)
10. Transaction Management or Transaction Processing and Theory
11. Query Processing
12. Object Database Standards, Languages, and Design
or
Object-oriented DBMSs – Concepts and Design
13. Object-relational DBMSs Concepts only.
14. XML data models and concepts related to database.

(6) Project Management

AIT 630 (COSC609): References

Jack Gido and James P. Clements, Successful Project Management, 5th ed., South-Western, Ohio, 2012.

A Guide to the Project Management Body of Knowledge, 5th edition, Project Management Institute, Inc., Pennsylvania, 2013.

General Study Topics	Gido & Clements	PMBOK
I. General Concepts: <ul style="list-style-type: none"> • Project management • Constraints • Life cycle • Project management process 	Ch. 1.	Ch. 1, 2
II. Initiating a Project: <ul style="list-style-type: none"> • Identification and selection • Charter • RFP & proposal development and processes • Types of contracts 	Ch. 2, 3	Ch. 3, 4
III. Planning, Performing, and Controlling a Project: <ul style="list-style-type: none"> • Defining the project objective and scope • WBS, network diagram, and other schedule aspects • Controlling and updating the schedule • Resource planning, leveling, and scheduling • Determining costs and cost performance measures • Risk management • Project closing activities 	Ch. 4, 5, 6, 7, 8, 9	Ch. 3, 4, 5, 6, 7, 8, 11
IV. Human Factors: <ul style="list-style-type: none"> • Project manager roles, skills, and development • Delegation • Change management and tracking changes • Stages of team development and effective project teams • Ethics, conflict and problem solving • Communication; project communication plan • Meetings, presentations, and reports • Project management organizational structures 	Ch. 10, 11, 12, 13	Ch. 1, 2, 9, 10

(7) Human Computer Interaction

General information:

- The topics below are covered at different level of details in the following courses offered by our department: COSC605
- These topics are covered in many textbooks on Human-Computer Interaction and Research methodology. We recommend the following text books:
 - Shneiderman, B. and Plaisant, C. (2009). Designing the user interface: Strategies for effective human-computer interaction. Addison Wesley. 5th edition. ISBN: 0321537351
 - Lazar, J., Feng, J., and Hochheiser, H. (2010). Research methods in HCI. Wiley. ISBN: 0470723378
 - Sharp, H., Rogers, Y., and Preece, J. (2007) Interaction Design, Beyond Human-Computer Interaction. Wiley. ISBN: 0470018666.

Topics:

1. Human cognition in HCI
 - a. Human memory
 - b. Human visual perception
 - c. Human cognition relates to output
 - d. Implications for design (7+-2 rules, recognition vs. recall)
2. Human errors
 - a. Mistakes vs. slips
 - b. Types of slips
 - c. Design strategies to prevent, detect, and recover from errors
3. Input and Fitts' Law
 - a. Input technologies
 - b. Fitts' law
4. Usability
 - a. Usability concepts
 - b. Usability measures
 - c. General design guidelines and principles
5. Accessibility
 - a. Accessibility concepts
 - b. Why is accessibility important
 - c. Accessibility guidelines and regulations
6. User-centered design
7. How to know your user and evaluate your design
 - a. Requirement analysis
 - b. Ethnography
 - c. Observation
 - d. Survey
 - e. Diary
 - f. Interview
 - g. Focus group

- h. Automated usability testing
 - i. Lab-based usability testing
 - j. Empirical studies
8. Experiment design
- a. Dependent variables vs. independent variables
 - b. Between group design
 - c. With group design
 - d. Split-plot design
 - e. When should a specific design be used
 - f. Internal, external, and theoretical validity in experiment design
9. Hypothesis testing
- a. Null hypothesis vs. alternative hypothesis
 - b. Type I error vs. type II error
 - c. Statistical significance
10. Statistical data analysis methods
- a. Why do we need to run statistical analysis on our data
 - b. T test
 - c. Analysis of Variance
 - d. Correlation
 - e. Regression
 - f. Repeated measures Analysis of Variance
 - g. Non-parametric tests
 - h. Factor analysis
11. Analyzing qualitative data (primarily text-based data)
- a. Grounded theory
 - b. How to code text-based data
 - c. How to evaluate the quality of the coding
 - d. How to interpret the data

(8) Computer Security

General Information:

- The topics below are covered at different level of details in the following course offered by our department: COSC734
- These topics are standard and are covered in many textbooks on computer and network security. We recommend the following textbooks written by the same author:

(WS-N) William Stallings, Network Security Essentials, Fourth Edition, Prentice Hall, 2011. ISBN-10: 0-13-610805-9

(WS-C) William Stallings, Cryptography and Network Security – Principle and Practice, Prentice Hall, ISBN-10: 0-13-609704-9

Topics:

Notice that following listed topics are mainly based on book WS-N. However, similar materials related to Basic Compute Security Concept, Cryptography, and Network Security Applications can be found WS-C book as well.

Basic Computer Security Concept:

- Computer Security Concept (WS-N: Chapter 1.1), OSI Security Architecture (WS-N: Chapter 1.2) Security Attacks (WS-N: Chapter 1.3), Security Service (WS-N: Chapter 1.4), Security Mechanisms (WS-N: Chapter 1.5), A Model for Network Security (WS-N: Chapter 1.6)

Cryptography

- Symmetric Encryption and Message Confidentiality: Symmetric Encryption Principle (WS-N: Chapter 2.1), Symmetric Block Encryption Algorithm (WS-N: Chapter 2.2), Steam Ciphers (WS-N: Chapter 2.4)
- Public-key Cryptography and Message Authentication: Approaches to Message Authentication (WS-N: Chapter 3.1), Message Authentication Code - HMAC (WS-N: Chapter 3.2), Public-Key Cryptography Principles (WS-N: Chapter 3.4). Public-Key Cryptography Algorithm (WS-N: Chapter 3.5)

Network Security Applications

- Key Distribution and User Authentication: Symmetric Key Distribution Using Symmetric Encryption (WS-N: Chapter 4.1), Kerberos (WS-N: Chapter 4.2)
- Transport-Level Security (WS-N: Chapter 5): Security Socket Layer and Transport Layer Security (WS-N: Chapter 5.2)
- IP Security: IP Security Overview (WS-N: Chapter 8.1), Encapsulating Security Payload (WS-N” Chapter 8.3)

System Security

- Intruders: Intruders (WS-N: Chapter 9.1), Intrusion Detection (WS-N: Chapter 9.2)
- Malicious Software: Types of Malicious Software (WS-N: Chapter 10.1), Viruses (WS-N: Chapter 10.2), Virus Countermeasures (WS-N: Chapter 10.3), Worms (WS-N: Chapter 10.4), Distributed Denial of Service Attacks (WS-N: Chapter 10.5)

- Firewalls: Firewall Characteristics (WS-N: Chapter 11.2), Types of Firewalls (WS-N: Chapter 11.3), Firewall Basing (WSN: Chapter 11.4)

(9) COSC 760 –Big Data Analytics

Course Description: This course provides students with an understanding of big data. Students will gain knowledge and hands-on experience on various big data framework and NoSQL database tools and real-time stream processing tools such as Spark. The course also includes machine learning based big data analytics. Students are expected to obtain insights on producing and consuming big data collection, ingestion, storage, analytics, and application. In addition, this course explores cutting-edge research in big data analytics published in academic journals and conferences.

Course Objectives: Upon completion of the course, students will understand and be familiar with the general concepts associated with Big Data processing and analytics using Big Data framework and NoSQL databases. Specifically, students will be able to:

- Gain in-depth understanding about data-driven decision-making environment with big data;
- Figure out and classify big data datasets, storages, data platforms, analytics mechanisms;
- Formulate big data analytics model and choose a proper framework and tool(s) to discover hidden/implicit values;
- Perform big data analytics using big data cluster framework and real-time stream processing tools;
- Understand machine learning based big data analytical methods with machine learning libraries available on Big Data frameworks;
- Develop big data analytics applications.

Textbook:

Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale, Ofer Mendeleevitch, Casey Stella, Douglas Eaoline, Addison-Wesley, 2017

Recommended Readings:

Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. David Loshin. Elsevier, 2013.

Hadoop: The Definitive Guide (Fourth Edition), Storage and Analysis at Internet Scale. Tom White O'Reilly Publishers, 2015.

MangoDB: The Definitive Guide (Second or Third Edition), Powerful and Scalable Data Storage. Shannon Bradshaw, Kristina Chodorow O'Reilly Publishers, 2013/June 2017 (scheduled).

Elasticsearch in Action.

Radu Gheorghe, Matthew Lee Hinman, Roy Russo. Manning Publications Co., 2016.

Learning Spark, Lightning-fast Data Analysis.

Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia. O'Reilly Publishers, 2015.

Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Foster Provost and Tom Fawcett O'Reilly Publishers, 2015.

References:

Apache Hadoop, <http://hadoop.apache.org/>

Apache HBase, <https://hbase.apache.org/>

Apache Spark, <http://spark.apache.org/>

Apache Mahout, <http://mahout.apache.org/>

Apache Cassandra, <http://cassandra.apache.org/>

Apache CouchDB, <http://couchdb.apache.org/>

MongoDB, <http://www.mongodb.com>

Elasticsearch, <https://www.elastic.co/>

Hortonworks Data Platform, <http://hortonworks.com/>
Cloudera Data Platform, <https://www.cloudera.com>
MapR Data Platform, <https://www.mapr.com/>
TensorFlow, <https://www.tensorflow.org/>

Course Context/ Background: Over the last decade, technological advancements have enabled the collection of a vast amount of data (data-at-rest and data-in-motion) from various data sources including Cyber-Physical Systems (CPS), Internet of Things (IoTs), social media, and public/private sectors. The discovery of values in Big Data can enable a better understanding of business, society, the environment, health care, and all aspects of life and a data-driven decision making based on machine learning algorithm. This course provides students with an opportunity to build their Big Data analytical capabilities across all Big Data processing stages including its accumulation the cloud framework, data processing for analysis, and value discovery.

Course Topics: This course provides students with a background in the concepts and applications of Big Data analytics. Potential course topics include the following:

- Introduction to data science and Big Data analytics
- Big Data cluster framework
- Introduction to Hadoop Ecosystem/Hadoop Stack
- Big Data classification
- Hadoop Distributed File System (HDFS)
- Map/Reduce Big Data processing approach
- Hadoop Application Data Frameworks
- NoSQL databases for non-structured datasets
- Real-time stream data processing
- Introduction to machine learning algorithms
- Cutting-edge Big Data analytics trends and research directions

Course Format: Active learning techniques, lectures, discussion sessions (in class AND online), presentations, lab exercises, hands-on programming, and projects may be used. Students are expected to read and use current contents on the course subjects using the library, internet and provided resources. Some assignments will necessitate collaborative learning while others will require individual research and presentation.

Labs and Assignments: Big data analysis tasks will be assigned throughout the semester to give students hands-on experience using big data cluster framework and real-time stream data processing software tools such as Hadoop, MongoDB, CouchDB, Spark, Mahout, Elasticsearch, etc. Students are expected to select a tool as their big data clustering framework and NoSQL database, present the summary of the tools, and share the installation procedures during the correspondent lab sessions. We will also complete a number of big data analytics labs using the Python/Scala/Java to experiment with a number of datasets: data-at-rest and data-in-motion. There will be three big data analytics assignments that require students to setup a big data cluster framework and implement big data analytical approaches on real datasets. All assignments are due at the assigned time on Blackboard. All work should be thoroughly backed up before turning it in and all submissions should be well documented with proper citations. Homework may also be assigned in the lecture at the instructor's discretion and as the need arises.

Article Summary and Presentation: Each student will be required to select a research article related to a course topic in the beginning of the semester. The student will write a summary of the article to be submitted to the professor and prepare a 20-minute presentation on the article. The presentation should give details about the approach presented in the article and summarize the results. During the first class, students will select a topic and sign up for a presentation slot.

Project: The semester project will require students to apply a big data analytics technique covered during the semester to discover values in a real-world dataset. Project examples include classification of Baltimore City crime/income data, world-wide/nation-wide climate data, clustering of stock market data, social network data, transaction data, CPS/IoT data, and many other public/private datasets. Students will be required to submit a project proposal to the instructor before carrying out any execution of the project. Projects should generally follow the steps of the big data analytics process and should give details on the data pre-processing, clustering and mapping, reducing and transforming, modeling and analyzing, discovering knowledge and value of the data, and

learning from the data. This will include selecting an appropriate algorithm and a software tool based on the knowledge and value discovery problem at hand. Projects should also report on the evaluation of the approach and use visualization to present the discovered knowledge and value. Students are encouraged to use available big data cluster framework and real-time stream data processing software tools such as Hadoop, MongoDB, CouchDB, Spark, Elasticsearch, etc.

Tentative Schedule: The following is a tentative schedule. Note that these topics and chapters are subject to change based on time and discretion of the instructor.

Week	Lectures	Labs/Assignments: Reading Article, Discussions, and Development with Tools
Week 1	Course Overview/Introduction to Data Science and Big Data Analytics	Reading/Overview: Introduction to Python and Scala
Week 2	Big Data Cluster Framework	Reading/Overview: Hadoop Ecosystem
Week 3	Introduction to Hadoop Ecosystem	LAB #1: Hadoop Stack and Hadoop Application Data Framework
Week 4	Big Data Classification (data-at-rest & data-in-motion)	ASSIGNMENT #1: Supplemental Readings Article Presentations
Week 5	HDFS, Big Data Store	LAB #2: HDFS as a Big Data Store/MapReduce
Week 6	MapReduce	ASSIGNMENT #2: MapReduce Applications
Week 7	Hadoop Application Data Framework	LAB #2: HDFS as a Big Data Store Hortonworks, Cloudera, MapR, Amazon EMR, etc.
Week 8	Project Discussion	Project Discussion
Week 9	NoSQL Data Management for unstructured data processing	LAB #3: NoSQL Databases MongoDB, CouchDB, Elasticsearch, Cassandra, etc. *Students choose and install their preferred NoSQL database. Lab 4
Week 10	Building NoSQL Database	ASSIGNMENT #3: NoSQL Application Unstructured data (ex, JSON documents) into the selected NoSQL Database. Lab 5
Week 11	Cyber-Physical Systems (CPS)/Internet of Things (IoTs)/Smart Grid Environment	LAB #4: Real-time Stream Data Processing for CPS/IoTs Apache Spark, etc.
Week 12	Real-time Stream Data Processing: Spark	ASSIGNMENT #4: Real-time Stream Data Processing for CPS/IoTs Application
Week 13	Introduction to Machine Learning	Optional work: Machine Learning Tools Apache Mahout, Spark MLlib, TensorFlow, etc. *Students choose their preferred machine learning library or tool for their term project
Week 14	Machine Learning Algorithms	Supplemental Readings: Machine Learning Algorithms
Week 15	Project Presentation/Final exam	Project Presentations