

BIG DATA ANALYTICS (CS138)

IV B.Tech: II Sem

L: 3 T: P: C: 3

Name of the Instructor(s): **Syed Nawaz Pasha**

No. of Hours/week: 3

Total number of hours planned: 48

Pre-requisite

- Data Warehousing and Data Mining

Learning Resources

Course notes, Text books, Course sites.

Required Resources

Text Books:

1. Seema Acharya and Subhashini Chellappan “Big data and Analytics” Wiley India Publishers, 2011.
2. Tom White “Hadoop the definitive guide” O’Reilly Publishers, 2015

Reference Books:

1. Eric Sammer, “Hadoop Operations”, O’Relley, 2012.
2. Vignesh Prajapati, Big data analytics with R and Hadoop, SPD 2013.

Web Links:

1. <http://hadoop.apache.org>
2. www.cloudera.com
3. <http://hortonworks.com>
4. <http://www.edureka.com>

Reading materials:

1. Course Material
2. Online Video links.

How to Contact Instructor:

Syed Nawaz Pasha:

- In-person office hours: 9:30 AM to 5:00 PM – Room no.: 1311, except class timings
- Online office hours: 9:30 AM to 5:00 PM - Except class timings, a mail or message
 - Mail: sd.nawazpasha@sru.edu.in
 - Phone numbers: 9959054349
- Other than office hours: A message to the above number from 6PM to 9PM from Monday to Saturday and 8AM to 10 AM on Sunday

Technology Requirements: (optional)

- Laptops for class work
- Software's – Hadoop, Pig, HIVE, HBase, Zookeeper
- Learning management system like Google classroom.

Overview of Course:

- What is the course about: its purpose?

Big data analytics is the often-complex process of examining large and varied data sets, or big data, to uncover information -- such as hidden patterns, unknown correlations, market trends and customer preferences -- that can help organizations make informed business decisions.
- What are the general topics or focus?
 1. Types of Digital Data.
 2. Big Data
 3. Hadoop
 4. HDFS Architecture
 5. Hive
 6. Pig Latin Programming
 7. HBase
- Why would students want to take this course and learn this material?
 1. To understand the 5V's paradigm of big data and categorizing digital data.
 2. To apply analytics to gain insights into the big data.
 3. To comprehend the Hadoop architecture for storage and processing the big data.
 4. To explore NoSQL databases and differentiating from SQL.
 5. To analyse the tools available for big data analytics.

Methods of instruction

- Lecture (chalk & talk / PPT)
- Collaborative Learning (Think pair share / Jigsaw / Brainstorming)

Workload

- Estimated amount of time student needs to spend on course readings: 2-3 hours per week
- Estimate amount of time to student needs to spend on course assignments and projects: 3-4 hours per week.

Assessment

S. No	Assessments	Assessment Methodology	No of assessments	Weightage in marks	Marks scaled to
	CIE	Quizzes	2	5	5
		Class test	2	10	5
		Assignment	--	--	--
		Course Activity	--	--	--
		Course Project	--	--	--
		Internal exams	2	20	20
	SEE	--	--	--	70

Note:

- Class test/ Quiz – 1. last week of February (Quize-1)
2. Third week of March (Class Test-1)
3. Second week of April (Quize-2)
4. Second week of May (Class Test-2)
- Grades will be shared within 3 days from the activity.
- Quizzes are having 10 Questions each carry equal marks for a total of 5 marks.
- Class Test is having 4 Questions each carry equal marks for a total of 10 marks to be scaled to 5 marks.
- Absentees for class assessments: In case the student is absent then an open-ended problem will be given as an assignment with a deadline, in case the assignment is not submitted in time then he/she will give zero marks.

Key concepts

1. Types of Digital Data.
2. Big Data
3. Hadoop
4. HDFS Architecture
5. Hive
6. Pig Latin Programming
7. HBase

LESSON PLAN

Course Outcomes (COs):

At the end of the course the student should be able to:

1. Classify various data analytics.
2. Distinguish hdfs operations for different applications.
3. Implement map reduce application for various data.
4. Relate various programming constructs with respect to big data.
5. Demonstrate Hadoop ecosystem for data analytics

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs)

Course Outcomes (COs) / Program Outcomes (POs)	1	2	3	4	5	6	7	8	9	10	11	12	PSO1	PSO2
CO1	2	2											2	
CO2	2	2	3	2	3	1						2		2
CO3	2	2	3	3	3	1								3
CO4		2	3	2										2
CO5			2	3										2

Course Content (Syllabus)

UNIT – I

Types of digital data: Structured, Semi-structured and Un-structured. Big data definition: volume, velocity, variety, veracity and value. Big data analytics skills: analytics classification, In-memory analytics, parallel processing and distributed systems. CAP theorem. The big data technology landscape: Sql and NoSql. Introduction to Hadoop.

UNIT – II

Exploring Hadoop architecture: HDFS: The hadoop distributed file system. Basic concepts, namenode, datanode, The design of HDFS, Interfaces Basic commands for file system operations, Hadoop I/O.

UNIT –III

Introduction to MapReduce programming: Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression. Developing a MapReduce application (programming language is not binding), MapReduce workflows, MapReduce Types & Formats, Features of MapReduce.

UNIT – IV

Introduction to HIVE: Architecture, File Format, Hive Query Language (HQL).

Programming with Hive.

Introduction to Pig: Pig Latin Overview, data types, Running Pig, Execution models, Pig Commands, Application development using Pig. Pig vs Hive.

UNIT – V

Introduction to HBase: NoSQL a column-oriented database. Example schemas & comparisons with RDBMS. Introduction to Zookeeper: Installing & running zookeeper. Zookeeper service. Introduction to Sqoop: A sample import using Sqoop. Importing large objects. Working with imported data. A deeper look at Sqoop export.

Lecture No.	Topic	Delivery Method/ Activity
UNIT-I		
L1	Introduction	Brainstorm
L2	Types of digital data: Structured, Semi-structured and Un-structured.	Chalk and Talk
L3	Big data definition: volume, velocity, variety, veracity and value	Chalk and Talk
L4	Big data analytics skills: analytics classification	Chalk and Talk
L5	In-memory analytics	Chalk and Talk
L6	parallel processing and distributed systems	PPT
L7	CAP theorem	Discussion
L8	The big data technology landscape: Sql and NoSql	Chalk and Talk
L9	Introduction to Hadoop	Brainstorm
UNIT-II		
L10	Exploring Hadoop architecture	Chalk and Talk
L11	Quiz-1 from Lecture 1 to Lecture 10	
L12	HDFS: The hadoop distributed file system	Chalk and Talk
L13	Basic concepts	Chalk and Talk
L14	Name node	Chalk and Talk
L15	Data node	Chalk and Talk
L16	The design of HDFS	PPT
L17	Interfaces	Chalk and Talk
L18	Basic commands for file system operations	Discussion
L19	Hadoop I/O	Chalk and Talk

UNIT-III		
L20	Introduction to MapReduce programming	Brainstorm
L21	Mapper, Reducer, Combiner	Chalk and Talk
L22	Partitioner, Searching, Sortitng, Compression	Chalk and Talk
L23	Class Test-1 from Lecture 12 to Lecture 22	
I – Mid Examinations		
L24	Developing a MapReduce application	Chalk and Talk
L25	MapReduce workflows	PPT
L26	MapReduce Types & Formats	Discussion
L27	Features of MapReduce	Chalk and Talk
UNIT-IV		
L28	Introduction to HIVE, Architecture	Brainstorm
L29	File Format	Chalk and Talk
L30	Hive Query Language (HQL)	Chalk and Talk
L31	Programming with Hive	Chalk and Talk
L32	Quize-2 from Lecture 24 to Lecture 31	
L33	Introduction to Pig, Pig Latin Overview	Chalk and Talk
L34	data types	Discussion
L35	Running Pig	Chalk and Talk
L36	Execution models	Brainstorm
L37	Pig Commands	Chalk and Talk
L38	Application development using Pig	Chalk and Talk
L39	Pig vs Hive	Chalk and Talk
UNIT -V		
L40	Introduction to HBase	Chalk and Talk
L41	NoSQL a column-oriented database	Brainstorm
L42	Example schemas & comparisons with RDBMS	Chalk and Talk
L43	Introduction to Zookeeper, Installing & running zookeeper	Chalk and Talk
L44	Zookeeper service	Chalk and Talk

L45	Introduction to Sqoop, A sample import using Sqoop	PPT
L46	Importing large objects, Working with imported data	Chalk and Talk
L47	A deeper look at Sqoop export	Chalk and Talk
L48	Class Test-2 from Lecture 33 to Lecture 47	
II – Mid Examinations		

INFORMATION RETRIEVAL SYSTEMS (CS141)

IV B.Tech: II Sem

L: 3 T: P: C: 3

Name of the Instructor(s): **Dadi Ramesh**

No. of Hours/week: 3

Total number of hours planned: 48

Learning Resources

Ex: Course Notes, PPT, Web Links

Required Resources

Text Books:

1. Kowalski, Gerald and Mark T Maybury, "Information Retrieval Systems: Theory and Implementation", Kluwer Academic Press, 2006.
2. Robert Korfhage, "Modern Information Retrieval by Yates Pearson Education Information Storage and Retrieval", John Wiley & Sons, 1997.

Reference Books:

1. W.B. Frakes and Ricardo Baeza-Yates, "Information Retrieval Data Structures and Algorithms", Prentice Hall, 1992.
2. Gobinda B Chowdhury, "Information to Modern Information Retrieval", Library Accusation Publication.
3. Kowalski, Gerald and Mark T Maybury, "Information Storage and Retrieval Systems: Theory and Implementation", Kluwer Academic Press.

How to Contact Instructor:

Dadi Ramesh:

- In-person office hours: 9:30 AM to 5:00 PM – Room no.: 1311, except class timings
- Online office hours: 9:30 AM to 5:00 PM - Except class timings, a mail or message
 - Mail: d.ramesh@sru.edu.in
 - Phone numbers: 9848142720
- Other than office hours: A message to the above number from 6PM to 9PM from Monday to Saturday and 8AM to 10 AM on Sunday

Technology Requirements: (optional)

- Laptops for class work
- Google Classroom

Methods of instruction

- Lecture (chalk & talk / ICT)

- PPT
- Video Lecture presentation

Workload

- Estimated amount of time to spend on course readings
Students are informed to spend half an hour per day (any four days of a week) or maximum of two hours per week on course readings
- Estimate amount of time to student needs to spend on course assignments and projects (per week)
One assignments is given during the delivery of this course. Students will need to spend couple of hours per day for a maximum of three days or one hour per day for a maximum of six days to finish the assignment

Assessment

S. No	Assessments	Assessment Methodology	No of assessments	Weightage in marks	Marks scaled to
1	CIE	Quizzes	--	--	--
2		Class test	4	10	2.5
3		Assignment	2	10	7.5
4		Course Activity	--	--	--
5		Course Project	--	--	--
6		Internal exams	2	20	20
7	SEE	--	--	--	70

Note:

- Class test: The class test will be scheduled after completion of each unit with prior intimation to students and the grades will be announced 3 days after the completion of test.
- Two assignments will be given each of 10 marks and scaled to total of 10. The students need to submit the assignment in time.
- Absentees for class assessments.
The absentees for class assessment will be awarded zero marks unless the reason for absent is prior and a genuine one. In case of genuine reason and have permission from concerned authority, a retest will be conducted with a new set of assessment questions. In case of assignment full marks will be awarded for those who submit proper assignment in time, late assignments will be marked as zero.

Key concepts

1. Information retrieval
2. Data structures
3. Stemming
4. Clustering
5. Text mining

LESSON PLAN

Course Outcomes (COs):

At the end of the course, the students will develop ability to

1. Apply the catalogue, indexing, the student should design effective information.
2. Accomplish the data structure and automatic indexing for the hypertext.
3. Synthesize programs to implement search engines.
4. Build skills in problem solving using systematic approaches.
5. Analyze the limitations of different information retrieval techniques.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs)

Course Outcomes (COs) / Program Outcomes (POs)	1	2	3	4	5	6	7	8	9	10	11	12	PSO1	PSO2
Apply the catalogue, indexing, the student should design effective information	3	2				2							3	
Accomplish the data structure and automatic indexing for the hypertext	3	2				2							3	
Synthesize programs to implement search engines.	2	3		2		3							2	
Build skills in problem solving using systematic approaches.				3		3							3	2
Analyze the limitations of different information retrieval techniques	2	1		3		1							2	

Course Syllabus:

UNIT – I

Introduction: Definition, Objectives, Functional Overview, Relationship to DBMS, Digital libraries and Data Warehouses.

UNIT – II

Information Retrieval System Capabilities: Search, Browse, Miscellaneous

Cataloging and Indexing: Objectives, Indexing Process, Automatic Indexing, Information Extraction.

UNIT –III

Data Structures: Introduction, Stemming Algorithms, Inverted file structures, N-gram data structure, PAT data structure, Signature file structure, Hypertext data structure.

Automatic Indexing: Classes of automatic indexing, Statistical indexing, Natural language, Concept indexing, Hypertext linkages

UNIT – IV

Document and Term Clustering: Introduction, Thesaurus generation, Item clustering, Hierarchy of clusters.

User Search Techniques: Search statements and binding, Similarity measures and ranking, Relevance feedback, Selective dissemination of information search, weighted searches of Boolean systems, Searching the Internet and hypertext. **Information Visualization:** Introduction, Cognition and perception, Information visualization technologies.

UNIT – V

Text Search Algorithms: Introduction, Software text search algorithms, Hardware text search systems. **Information System Evaluation:** Introduction, Measures used in system evaluation, Measurement example – TREC results.

Lecture No.	Topic	Delivery Method/ Activity
UNIT-I		
L1	Definition, Objectives	Brainstorm
L2	Functional Overview	Chalk and Talk
L3	Relationship to DBMS, .	Chalk and Talk
L4	Digital libraries and Data Warehouses	Chalk and Talk
UNIT-II		
L5,6	Information Retrieval System Capabilities	Chalk and Talk
L7,8	Search Capabilities	Chalk and Talk
L9	Browse Capabilities	Chalk and Talk
L10, 11	Miscellaneous Capabilities	Chalk and Talk
L15,16	Cataloging and Indexing	Chalk and Talk
L17,18	Indexing Process	PPT
L19,20	Automatic Indexing	Chalk and Talk
L21,22	Information Extraction.	Discussion
UNIT-III		
L23	Data Structures: Introduction	Brainstorm
L24,25	Stemming Algorithms	Chalk and Talk
L26	Inverted file structures	Chalk and Talk
L27,28	N-gram data structure	Chalk and Talk
L29,30	PAT data structure	Chalk and Talk
L31,32	Signature file structure, Hypertext data structure	Chalk and Talk
L33,34	Automatic Indexing: Classes of automatic indexing,	Chalk and Talk
L35	Statistical indexing	Chalk and Talk
L36	Natural language Indexing	Chalk and Talk

UNIT-IV		
L37	Document and Term Clustering: Introduction, .	Brainstorm
L38, 39	Thesaurus generation Item clustering, Hierarchy of clusters	Chalk and Talk
L40	User Search Techniques: Search statements and binding,	Chalk and Talk
L41	Similarity measures and ranking,	Chalk and Talk
L42,43	Relevance feedback, Selective dissemination of information search,	Chalk and Talk
L43	weighted searches of Boolean systems, Searching the Internet and hypertext	Discussion
L44	Information Visualization: Introduction,	Chalk and Talk
L44	Cognition and perception, Information visualization technologies	Brainstorm
UNIT -V		
L45	Text Search Algorithms: Introduction, .	Chalk and Talk
L45	Software text search algorithms,	Brainstorm
L46	Hardware text search systems.	Chalk and Talk
L46	Information System Evaluation: Introduction,	Chalk and Talk
L47	Measures used in system evaluation,	Chalk and Talk
L48	Measurement example – TREC results	PPT

CYBER LAWS (OE115)

IV B.Tech: II Sem

L: 3 T: P: C: 3

Name of the Instructor(s): Dr. D. Kothandaraman.

No. of Hours/week: 3 Hours per week

Total number of hours planned: 30

Pre-requisite

- Basic knowledge about internet crimes
- Cyber issues related to the internet, communication technology, and electronic elements including hardware, software, computer and information systems.
- Legal and illegal issues related to the internet crimes.

Learning Resources

1. Lecture Notes

Required Resources:

Text Books:

1. Harish Chander, “Cyber Laws and IT Protection”, PHI, 2012.

Reference Books:

1. George Kostopoulos, “Cyberspace and Cyber Security”, Auerbach Publications, 2012.
2. Albert Marcella, Jr., Doug Menendez, “Cyber Forensics: A Field Manual for Collecting, Examining, and Preserving Evidence of Computer Crimes”, Auerbach Publications, 2nd Edition, 2007.

Reading materials:

1. Lecture notes soft copy will be provided to the students.
2. Online video links will be provided.

Additional Resources:

<https://www.itlaw.in/>

<http://ili.ac.in/e-learnCL.htm>

How to Contact Instructor:

- In-person office hours:
 1. Students can able to meet, whenever we have free schedule during the college hours.
 2. Students can able to meet 3:00 pm to 4:00 pm in college working hours.
- Online office hours: time and how to access
 - Email address: kothanda_raman_d@srecwarangal.ac.in.
 - Phone numbers: 9944380724
 - LMS: 7 pm to 10pm

Technology Requirements:

- Google classroom

Overview of Course:

- **What is the course about: its purpose?**

A cyber law plays a vital role in all electronic contents in internet. Cyber-law is created to help protect people and organizations on the internet from malicious people on the internet and help to maintain order. If someone breaks a cyber-law or rule, it allows another person or organization to take action against that person or have them sentenced to a punishment.

- **What are the general topics or focus?**

1. Scope of cyber laws
2. Online resources related to cyber laws
3. Security information's
4. Regulation of cyber crimes
5. Types of E-commerce
6. IPR Issues

- **How does it fit with other courses in the department or on campus?**

Cyber laws course common to all branches.

- **Why would students want to take this course and learn this material?**

In future all electronic contents are processing in internet. So that everyone should aware about the cyber-crime related to the personal and office data's.

Methods of instruction

- Lecture
- Group Discussion
- Team based learning

Workload

- Estimated amount of time student needs to spend on course readings: **1 hour per week**
- Estimate amount of time to student needs to spend on course assignments and projects: **1 hour per week**

Assessment

S. No	Assessments	Assessment Methodology	No of assessments	Weightage in marks	Marks scaled to
1	CIE	Quizzes	--	--	--
2		Class test	--	--	--
3		Assignment	2	2.5	5
4		Course Activity	2	2.5	5
5		Course Project	--	--	--
6		Internal exams	2	10	20
7	SEE	--	--	--	70

Assessments:

Two types of assessments: 1. Assignments 2. Activities

1. Assignment:

1.1. Assignment:

Schedule: Before the I-Internal Examination

Syllabus: I-Unit, II-Unit and III-Unit.

1. 2. Assignment:

Schedule: Before the II-Internal Examination

Syllabus: IV-Unit, V-Unit.

Note: If the students submit the assignment in time then, will be given with 2.5 marks, otherwise 0 marks

2. Activities:

2. 1. Activity:

Topic: Analyze in details about different role of electronic signatures

Activity: Think Pair Share

Description of the Activity: Each team has to analyses the role of electronic signature.

Rubrics:

H.T.No	Identify the electronic signature(1m)	Identify different electronic evidence (1.5m)	Total Marks (2.5m)

2. 2. Activity:

Topic: Offences under the Cyberspace law and the Internet in India

Activity: Think Pair Share

Description of the Activity: Each team has to write important offences under the cyberspace laws in internet in india.

Rubrics:

H.T.No	Identify the important offences (1m)	What are the offences under cyber laws and the internet in india (1.5m)	Total Marks (2.5m)

Note: If done the activity 2.5 marks, otherwise 0 marks

Key concepts

- Introduction to Computer Technology and Programming.
- Legal Dimensions of IPR in Cyber World.
- Fundamentals of Information Security.
- E-Business Security.
- Communication and Soft Skills.
- IT Acts and Cyber Crime.
- OS: Linux, Server (Apache)
- IT Act, Information assurance & e-Governance/ Data Information & Cryptography.

LESSON PLAN**Course Outcomes (COs):**

At the end of the course the student should be able to:

1. Analyze and evaluate the current trends and technologies such as e-commerce and e-governance.
2. Sketch the importance of digital signature in electronic records.
3. Formulate the importance and role of cyberspace laws and cyber-crimes.
4. Design and motivate law relating to electronic records and intellectual property rights in India.
5. Summarize about the IT act in India and generate the new IT acts for current cyber space.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs)

Course Outcomes (COs) / Program Outcomes (POs)	1	2	3	4	5	6	7	8	9	10	11	12	PSO1	PSO2
CO1	3	3								1		1	3	3
CO2	3	3	3	2									3	3
CO3	3	2	2	2					1				3	2
CO4	3	3	2	2									3	2
CO5	3	2	2	2					1				3	2

Course Syllabus

UNIT I

Internet, E-Commerce and E-Governance with reference to Free Market Economy:

Understanding Computers, Internet and Cyber laws, Conceptual Framework of E-commerce: E-Governance, the role of Electronic Signatures in E-commerce with Reference to Free Market Economy in India.

UNIT II

Law Relating to Electronic Records and Intellectual Property Rights in India: Legal aspects of Electronic records / Digital signatures, The roles and regulations of Certifying Authorities in India, Protection of Intellectual Property Rights in Cyberspace in India.

UNIT III

International Efforts Relating to Cyberspace Laws and Cyber Crimes: International efforts related to Cyber laws, Council of Europe (COE) convention on Cyber Crimes

UNIT IV

Penalties, Compensation and Offences Under the Cyberspace and Internet in India:

Penalties, Compensation and Adjunction of violations of provisions of IT Act and Judicial review, some important offences under the Cyberspace law and the Internet in India, Other offences under the Information Technology Act in India.

UNIT V

Miscellaneous Provisions of IT Act and Conclusions: The role of Electronic Evidence and miscellaneous provisions of the IT Act.

Lecture No.	Topic	Delivery Method/ Activity
UNIT – I Internet, E-Commerce and E-Governance with reference to Free Market Economy:		
L1	Introduction	Chalk and Talk
L2	Chalk and Talk	Chalk and Talk
L3,L4	Internet and Cyber laws	Chalk and Talk
L5,L6	Conceptual Framework of E-commerce	PPTs
L7,L8	E-Governance	PPTs
L9	Role of Electronic Signatures (2.1 Activity)	Think Pair Share
L10	E-commerce with Reference to Free Indian Market	PPTs

UNIT – II Law Relating to Electronic Records and Intellectual Property Rights in India		
L11	Introduction	PPTs
L12	Legal aspects of Electronic records	PPTs
L13	Digital signatures	Think Pair Share
L14	The roles and regulations of Certifying Authorities in India	PPTs
L15,L16	Protection of Intellectual Property Rights in Cyberspace in India	PPTs
	(1. 1. Assignment)	
UNIT – III International Efforts Relating to Cyberspace Laws and Cyber Crimes:		
L17	Introduction	Chalk and Talk
L18,L19	International efforts related to Cyber laws	PPTs
L20	Council of Europe (COE) convention on Cyber Crimes	PPTs
UNIT – IV Penalties, Compensation and Offences Under the Cyberspace and Internet		
L21	Penalties	Chalk and Talk
	(1.2. Assignment)	
L22,L23	Compensation and Adjunction of violations of provisions of IT Act and Judicial Review	PPTs
L24,L25	Some important offences under the Cyberspace law and the Internet in India (2.2. Activity)	Think Pair Share
L25,L26	Offences under the Information Technology Act in India	PPTs
UNIT-V - Miscellaneous Provisions of IT Act and Conclusions		
L26,L27,L28	The role of Electronic Evidence	Team-based learning
L29,L30	Miscellaneous provisions of the IT Act	PPTs