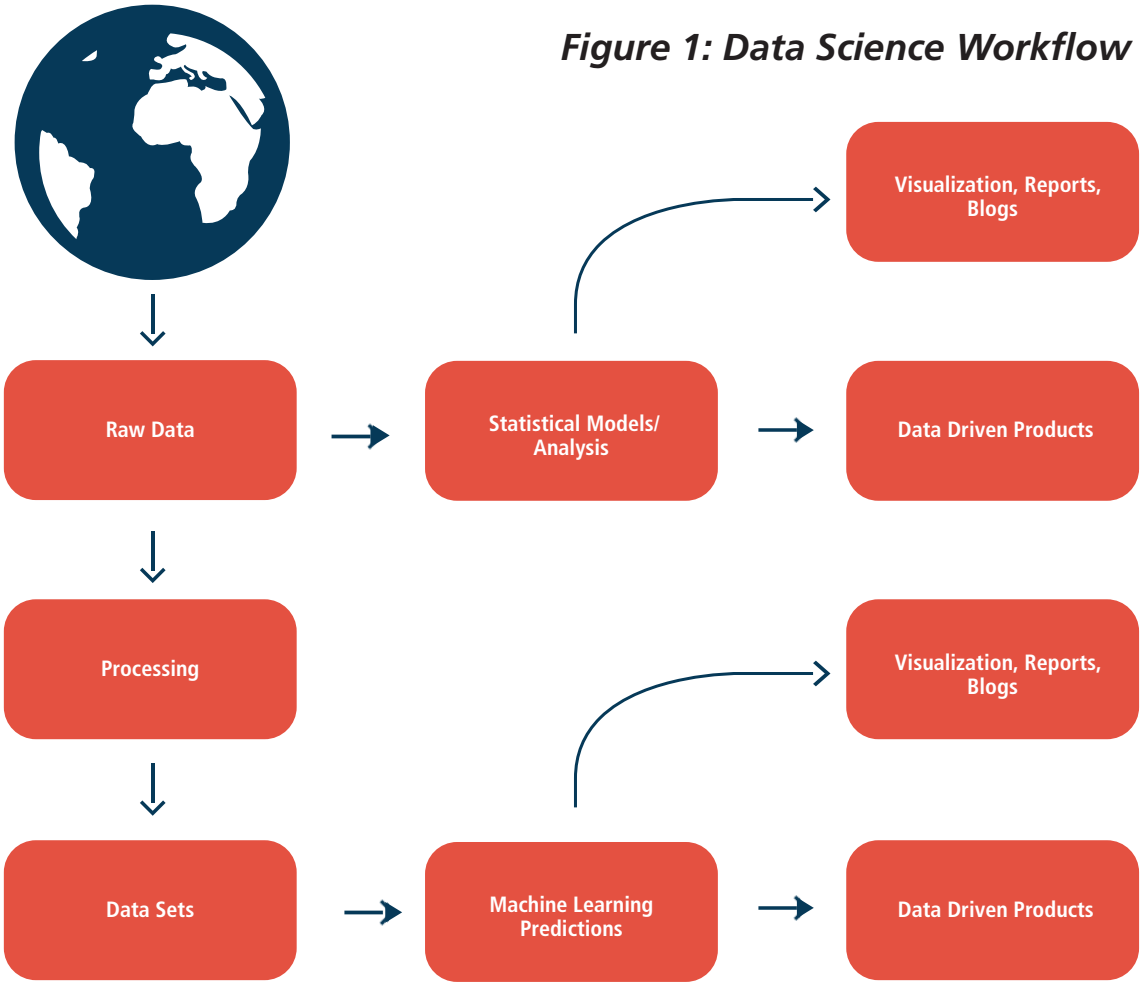
आई आई टी हैदराबाद
IIT Hyderabad

Executive M.Tech. in Data Science

Computer Science and Engineering Department

Indian Institute of Technology Hyderabad

Figure 1: Data Science Workflow



The department of Computer Science and Engineering at IIT Hyderabad has launched a new M.Tech. program, Executive M.Tech. in Data Science, exclusively for industry professionals since August 2015. The program is self-paced so that candidates can have flexibility in completing the program in 2-4 years. The program can be completed at the IIT-H campus or remotely through video-enabled online courses (with periodic visits to the IIT-H campus, typically scheduled on weekends).

The Need for Executive M.Tech. in Data Science Program

There are many applications, such as social media, healthcare, e-commerce, weather forecast, traffic monitoring, etc., that are producing massive amounts of data, the so-called "BIG DATA", with Volume, Velocity, Variety, Veracity and Value (the five "Vs" of Big Data challenges) at an unprecedented scale. This has led to a critical need for skilled professionals, popularly known as Data Scientists, who can mine and interpret the data. Making sense of this massive data is a very difficult challenge for scientific, technological and industrial disciplines.

Unfortunately, there is a gap between the demand and supply of data scientists and technologists. Following are the chief reasons behind this gap:

• Undergraduate courses are too generic for addressing issues in this area in a focused manner.

• There are not many postgraduate courses that focus explicitly on Data Science.

• Even if some generic postgraduate programs can be tailored to focus on Data Science through electives, professionals working in the industry or Research and Development establishments do not have the luxury of taking two years off for pursuing higher studies.

With these factors in mind, the CSE Department at IIT Hyderabad proposes a two-year course-work based M.Tech. program in Data Science area that is flexible and can be self-paced. This program is exclusively designed to cater to the needs of working individuals, wherein a candidate is expected to do eight 3-credit courses over a period of 2-4 years. Ideally, one can take 2 courses per semester. The classes will be held over the weekends (or other timings suitable for working professionals) with each class having a 3-hour duration.

Figure 1 shows the workflow of a data scientist, which can be broken down to following four essential steps:

1. **Data Collection:** Proliferation of smart devices, sensors, web, mobile and social media has led to explosive amount of complex data. To make use of this data, one needs expertise in Computer Networks and Databases to effectively collect and manage such huge volumes of data.

2. Data Processing: The next step is to convert the raw data into forms that can be scientifically analyzed, which includes data cleaning and transformation. For example, by transforming social network data into graph data, one can use concepts from Graph Theory to analyze social network data. To process huge volumes of data, one needs expertise in Databases and High Performance Computing. The data one needs to handle is a heterogeneous mix of different types of data, such as images, videos, text, social networks, etc. To handle these different types of data one needs expertise in areas such as Image and Video Analytics, Information Retrieval, Social Media Analytics, etc.

3. Data Analysis: The third step is to analyze the processed data using various Statistical, Data Mining and Machine Learning algorithms. Most of the existing data analysis algorithms do not scale to large datasets. As a result, one needs expertise in Statistics, Data Mining and High Performance Computing to design systems that can efficiently analyze large volumes of complex data.

4. Data Product: The final step is to make decisions from the data analysis and also deliver the analyzed information to the world in the form of various data products. This is often done using data visualization techniques, which are integrated with various smart devices. This step requires expertise in Information Visualization, Databases and Computer Networks.

From the above workflow, it can be understood that the task of a data scientist is quite complex and it requires expertise in multiple sub-disciplines of computer science and engineering. Through the Executive M.Tech. program in Data Science, our goal is to provide high-quality training in the aforementioned areas to meet the growing demands of the market for data scientists and technologists, and to serve our nation's economy from our capacity as an institution of national importance.

Curriculum Structure

Semester	Course Title	Course Type	Credits
Semester I	Mathematical Foundations of DS	Core	3
	Statistical Programming	Core	3
Semester II	Analytic Databases	Core	3
	Data Intelligence and Analytics	Core	3
Semester III	Applied Machine Learning	Core	3
	Elective 1	Elective	3
Semester IV	Elective 2	Elective	3
	Elective 3	Elective	3

Tentative List of Elective Courses

Image and Video Analytics	Parallel and Distributed Systems
Topics in Data Mining	Numerical Linear Algebra for Data Analytics
Scaling to Big Data	Programming Models for Multicore and GPU Architectures
Data Acquisition and Productization	Information Retrieval
Data Security	Web Databases and Information Systems
Data Privacy	Thick Data Analysis
	Social Media Analytics

Course Content[1]

Core Course	Course Content
Mathematical Foundations of DS	– Matrices, Vectors and Properties – Vector Spaces, Norms, Basis, Orthogonality – Matrix Decompositions: Eigen decomposition, Singular Value Decomposition; – Applications – Differential Calculus: Derivatives and its significance, Partial derivatives – Optimization of single variable and multiple variable functions: Necessary and sufficient conditions – Real problems as optimization problems: Formulation and analytical solutions – Finding roots of an equation: Newton Raphson Method – Optimization via gradient methods – Probability basics, density function, counting, expectation, variance, independence, conditional probability, Poisson process, recurrences, Markov chains
Statistical Programming	– Probability and statistics – Statistical measures and tests – Introduction to statistical packages, such as SPSS, SAS, R, Stata, JMP. – Statistical analyses using R and Python
Applied Machine Learning	– Supervised learning – Unsupervised learning – Performance evaluation – Machine learning applications – Machine learning for large datasets
Data Intelligence and Analytics	– Online Analytical Processing (OLAP) – Decisive analytics – Descriptive analytics – Predictive analytics – Prescriptive analytics – Data analysis in specific domains
Analytic Databases	– SQL programming – Relational algebra – Database design – Creating database application – Traditional, NoSQL and NewSQL DBMS

[1] The course content will be adapted to the class requirements.

Elective Course	Course Content
Image and Video Analytics	– Image and video processing – Human visual recognition system – Detection/segmentation/recognition methods – Image and video classification models – Image and video analytics applications
Scaling to Big Data	– Distributed computing architecture – Parallel programming – Apache Hadoop framework – MapReduce programming
Social Media Analytics	– What is social network? – Network models – Network centrality measures – Finding network communities – Information diffusion – Contagion and opinion formation
Data Acquisition and Productization	– Data preprocessing: extraction, cleaning, annotation, integration – Information visualization – Dashboards, Android and iOS apps. – Internet of things
Data Security	– Cryptography – Web security – Hardware and software vulnerabilities
Data Privacy	– Managing personally identifiable or sensitive information – Hippocratic databases – Differential privacy – Privacy preserving data analysis
Thick Data Analysis	– Challenges with Big Data analysis – Limitations of data-driven Big Data analytics – Using human emotions in data analytics
Parallel and Distributed Databases	– Limitations of centralized and client-server DBMS model – Big Data analysis using parallel and distributed databases – Parallel vs. distributed databases – Parallel database architectures

Programming Models for Multicore and GPU Architectures	– Introduction to Flynn's taxonomy – Task vs. data-parallelism – Traditional models: OpenMP, MPI – Newer models: APIs: OpenCL, CUDA, etc.
Numerical Linear Algebra for Data Analytics	– Matrices and vector operations – SVD, PCA, LDA – QR decomposition, least squares problem, linear regression – Tensor decomposition – Sparse linear algebra
Web Databases and Information Systems	– Modern web-based information systems – Using n-tiered architectures to implement secure and scalable systems – Database-driven websites and applications – Utilize JavaScript to improve database-driven websites. – Critical components of the modern Web infrastructure: DNS, CDN, etc – Web programming technologies: HTTP, XML, SQL, JavaScript, AJAX, CSS, RSS, and others.
Information Retrieval	– Storing, indexing and querying document data – Scoring, term weighting and the vector space model – Text classification problem and Naive Bayes – Probabilistic information retrieval – Link analysis
Topics in Data Mining	– Data Preprocessing, – Data Warehousing & OLAP – Mining Frequent Patterns and Associations – Classification – Cluster Analysis – Mining Complex Types of Data (Sequence Data, Graphs, Social Networks, etc.) – Text Mining – Stream Data Mining

Eligibility Criteria and Admission Process

Duration	Two to four years
Intake	Maximum 35
ADMISSION	
Eligibility	The candidate must have a minimum work experience of 2 years in industry and be employed at the time of applying. The candidate must have a BTech/BE/AMIE degree in CS/EE/IT/ECE, or a MCA, a MSc/MS degree in CS/IT/Maths/Stat, and have an excellent academic record. For candidates from others branches, they must have passed GATE exam in Computer Science with more than 80 percentile rank.
Selection Procedure	Candidates must fill an online application. Shortlisted candidates will go through test and/or Interview. Prior research exposure and/or industry experience in areas related to data science will be considered a plus. The final selection of the candidate will be based on performance in the written test and/or interview, qualifying examination, work experience, and any other criteria deemed suitable by the admission committee.
Program Fee	The fee for applying to the program is ₹500/- (₹250 for Women/SC/ST/PWD candidates). There is a registration fee of ₹10,000 per semester the student enrolls for one or more courses. The course fee is ₹17,500 per credit. For all the 24 credits the total is ₹4.2 Lakhs.

Important Dates:

Date of opening of site for submitting online applications	April 1, 2017
Last date for submitting online applications	May 25, 2017
Written/Objective test	June 9, 2017
Interview over Phone/Skype	June 15 – June 25, 2017
Announcement of selected candidates	June 30, 2017
Registration	To be announced
Commencement of class work	1st week of August, 2017

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

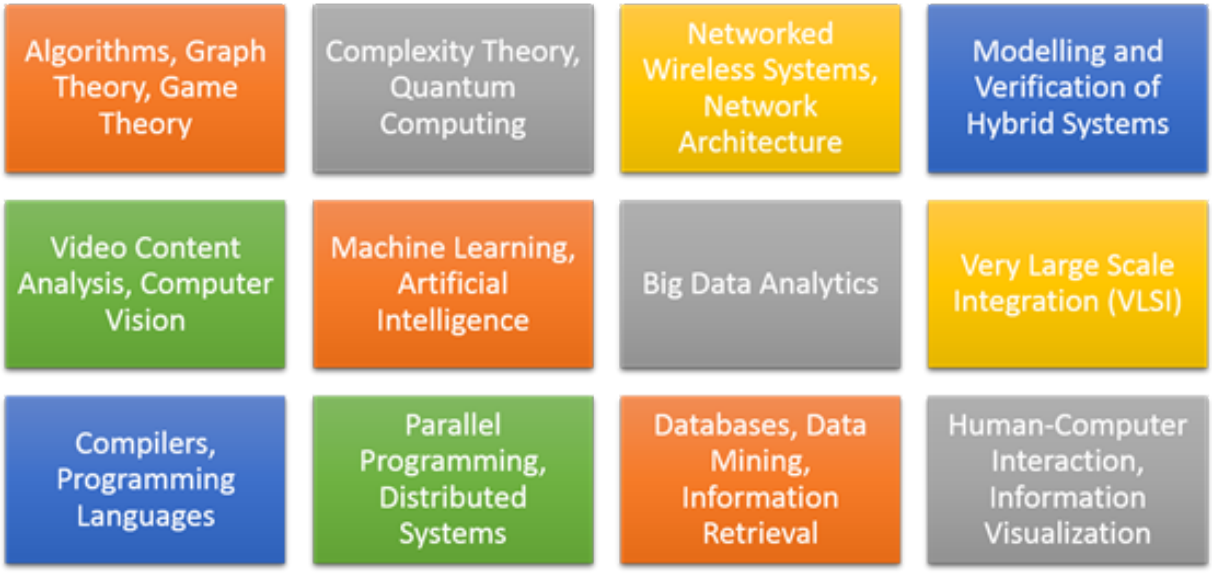
The Indian Institute of Technology, Hyderabad (IITH) was established in the year 2008 and is currently operating out of its permanent campus since 2015. The Institute has a sprawling permanent campus of about 570 acres at Kandi near Sangareddy in Medak; about 50 minutes' drive from the Rajiv Gandhi International Airport, Shamshabad.

Within a very short time since its inception, IIT Hyderabad (IITH) has made significant strides in research as well as in pedagogy. IITH became operational on August 20, 2008 with three departments: CSE, EE & ME, with the first batch comprising 111 B.Tech. students. Today, IITH has 142 full-time world class faculty: 6 Professors, 19 Associate Professors, 101 Assistant Professors, 13 On Contract Assistant Professors and 3 Visiting Assistant Professors. Moreover, IITH has hosted several international faculty for research collaborations as well as teaching engagements. IITH currently has 14 departments, which span across all the major departments found in any of the older IITs. The current student strength is 2126: 862 B.Tech., 108 M.Sc., 448 M.Tech., 32 M.Des., 8 M.Phil., and 668 Ph.D. IITH has an equal number of students in its postgraduate and undergraduate programs, which is a testimony to its emphasis on becoming a leading world-class research institution. IITH has so far graduated 330 B.Tech., 246 M.Tech., 5 M.Phil., 61 M.Sc., and 25 Ph.D. students.

Many state-of-the-art laboratories have been established over the last few years at IITH. In all, there are 102 labs that are operational, of which 41 are research labs, 30 are research/teaching labs, and 31 are teaching labs. IITH has nearly ₹120+ crore of sponsored research funding, and about ₹10+ crore of consultancy funding. On average, IITH faculty have published about 175 journals and referred conference papers annually over the last few years (which will only increase in years to come).

IITH has a very strong collaboration with Japan at an institutional level in research and development, as well as in the architectural design of the permanent campus.

This collaboration gives IITH a great impetus for quickly being among the world leaders for cutting-edge research. IITH already has some transdisciplinary centers, such as Center for Nano-X, X-Materials Center, Center for Sustainable Development, Center for IoT and Cyber Physical Systems, Center for High Performance Computing, Center for Smart Cities. IITH also has strong industry collaborations, both at national and international levels. Besides, IITH has MoUs and active collaborations with 8 leading US universities and two leading Japanese Universities. IITH has had several visiting faculty from USA, France, and Canada who have taught short duration (Fractional Credit) courses. IITH aims to create an environment that fosters innovation and invention, and seeks to realize the dreams of aspiring top-class students: dreams for higher knowledge, dreams for scientific inquiry, dreams for technology creation, dreams for co-curricular activities, and dreams to change the world.



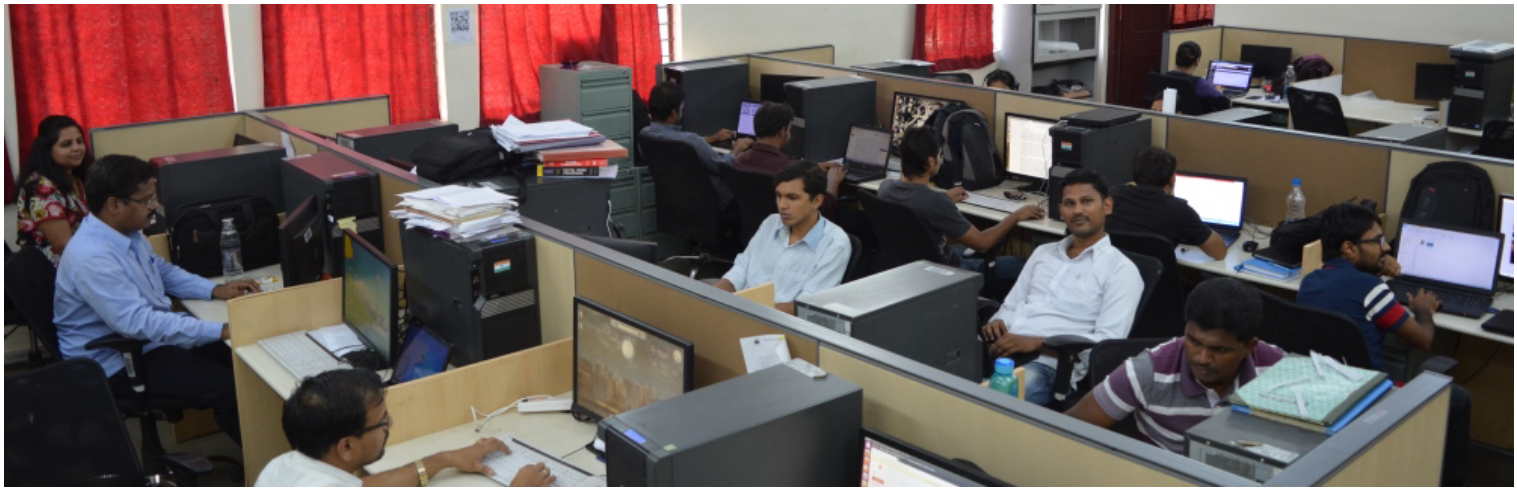
Research Areas

About CSE Department

The Department of Computer Science and Engineering (CSE) at IIT Hyderabad is poised for a giant leap through research in cutting-edge computing and technology, while imparting top-class education through innovative pedagogy. The department offers undergraduate B.Tech. program with current annual intake of around 45 students, and postgraduate programs such as 2-Year M.Tech., 3-Year M.Tech., EMDS and Ph.D. with current annual intake of around 65 students. The current student strength is 379. It had already graduated 5 batches of B.Tech. and M.Tech., and 8 Ph.D. student. Over the last few years, the department has also offered short courses, such as Foundations of Predictive Analytics, Big Data Analytics, etc., as a part of Continuing Education Program (CEP), customized to the needs of academia and industry. The department comprises of seventeen full-time faculty members and one adjunct faculty from reputed academic and industry backgrounds.

The faculty are actively engaged in research areas such as data analytics, machine learning, databases, image and video processing, information visualization, formal methods, computational complexity, algorithms, graph theory, networking, IoT, distributed systems, compilers, programming languages, and parallel processing. The department has several state-of-the-art computing facilities for advanced research in the areas of data analytics, networking, databases, cloud computing, etc. Following are some of the research/teaching labs in the CSE department:

1) Core Labs
   - Networked Wireless Systems (NeWS) Lab
   - Visual Intelligence and Learning (VIGIL) Lab
   - Data Analytics Lab
   - Theoretical Computer Science Lab
   - Compilers and Programming Languages Lab
   - Parallel Programming and Distributed Systems Lab
   - Databases and Data Mining Lab
2) DISANET (Disaster Management) Lab
3) CPS (Cyber-Physical Systems) Lab



Over just a few years, CSE faculty has received large sponsored research project grants, some of which are interdisciplinary and hence jointly with faculty of other departments at IITH, from various national and international agencies

Industry	Academic
• KDDI R&D Labs, Japan • Microsoft Research • IBM Research • AMD • Hitachi • NetApp • LG Soft (R&D Division of LG Electronics) • Uurmi Systems, Hyderabad	• Indian Institute of Science (IISc), Bangalore • Tata Institute of Fundamental Research (TIFR) • Purdue Univ., USA • Univ. of Illinois Urbana-Champaign, USA • Georgia Inst. of Tech., USA • University of Texas – Dallas, USA • UC - San Diego, USA • UC - Santa Cruz, USA • Univ. of Buffalo, USA • Univ. of Utah, USA • Osaka Univ., Japan • Univ. of Tokyo, Japan • Keio University, Japan • INRIA, Paris, France • Tel Aviv University • Paris Tech, France • Chennai Mathematical Institute, India • Royal Holloway University of London • Nanyang Technological University • University of Tartu

Research Collaborations

The upcoming CSE-EE department building in the permanent campus is expected to be ready by 2018. It is designed to be a smart building. The building structure is designed to be inspiring to precipitate interaction, creativity, and collaboration. To know more about the department and research interests of the faculty, please visit ***http://cse.iith.ac.in/.***

Join us.
Let's make a smarter world together.

Contact Us

Admin Office
Indian Institute of Technology Hyderabad
E-mail: emds@cse.iith.ac.in
Phone : (040) 2301 7006
Website: *http://cse.iith.ac.in/?q=emds*

Executive M.Tech. in Data Science

Computer Science and Engineering Department

Indian Institute of Technology Hyderabad