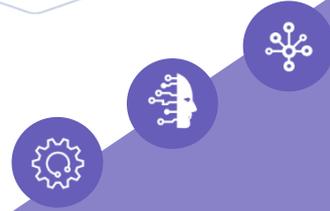


# Introduction to Big Data Analytics



NASSCOM FutureSkills Platform Partner

<https://academy.iZen.ai>  
Email: [info@iZen.ai](mailto:info@iZen.ai)

**future skills**  
A NASSCOM initiative



## CERTIFICATE PROGRAM



## OVERVIEW: BIG DATA LANDSCAPE

Today, it is very important for all of us to realize how data generation has exploded, and learn about Big Data technology evolution and important milestones of Big Data technologies. Knowingly or unknowingly, many companies have accumulated and invested in volumes of data which is nothing but “Big Data”. At a very high-level, Big Data Analytics technologies provide a means to analyze data sets and draw conclusions about them. This helps organizations make informed business decisions. Traditionally, Big Data systems used to be deployed on premises. However, cloud platform vendors, such as Amazon Web Services (AWS) and Microsoft (Azure), Google Cloud (GC) have made it easier to set up and manage Hadoop clusters in the cloud.

This course is a first big step towards mastering Big Data Analytics.



### HIGH DEMAND

Corporations are realizing the value of Big Data Analytics and are aggressively pursuing it to gain competitive advantage



### HUGE SUPPLY GAP

Demand for Big Data Engineers far exceeds supply. That means less competition and more pay for trained-experts.



### BRIGHT FUTURE

The global Big Data and business analytics market is forecasted to grow up to \$274.3 billion by 2022 and these jobs are not going away.

This program covers the basic knowledge and skills that an individual must possess before being trained on the core concepts of Big Data Analytics in addition to Big Data systems and platforms, Data Modelling and Management. This is an “Outcome-based training” with a focus on enabling students to explore the fundamentals of Big Data Analytics, to provide them with a base from where they can upskill themselves for specific Big Data Analytics job roles.

## BENEFITS OF THIS PROGRAM

- ✓ Kick-start or re-ignite your career by completing this course. The salaries of Big Data Engineers will only keep rising as the world can't keep up with the enormous demand
- ✓ Learn Big Data Analytics from a world-class professor with decades of experience teaching at the University of California, added with industry experience of building successful start-ups
- ✓ Learn by doing:
  - Get hands-on experience by doing several industry-specific projects
  - Access our “Big-Data-Lab” on the cloud to learn by doing
- ✓ Real-life case-studies and use-cases to give you practical insights
- ✓ Get a certificate of completion in “Big Data Analytics”



## WHO IS THIS PROGRAM FOR?



- ✓ If you are a student seeking employment to kickstart your career, this is a great opportunity for you to learn about concepts of Big Data and build a rewarding, future-proof, and meaningful career
- ✓ If you are already employed but are trying to rekindle your career in the exciting world of Big Data, this program is perfect for you
- ✓ Regardless of your background, this program can give you introductory knowledge in the area of Big Data and Analytics

### PREREQUISITES:

- ✓ Knowledge of the fundamentals of programming including data sequences such as stacks, queues, strings, arrays, linked lists, trees, maps and the concepts of Object-Oriented Programming

## WHAT WILL YOU LEARN?

Foundational Curriculum for Big Data Analytics is aimed at upskilling those who have a basic understanding of programming and data sequences, to help them expand their knowledge and learn the fundamentals of Big Data Analytics technologies at a beginner level.

- ✓ This course will train students in Big Data and tools and will expose students to a wide variety of concepts surrounding Big Data Analytics
- ✓ This is a very hands-on course and emphasis will be on how the students can use Big Data to derive value out of stored data by applying analytics techniques. Upon completing this course, you will learn to:
  - Evaluate trends in Big Data and discuss how Big Data is transforming businesses
  - Evaluate the different platforms used for processing Big Data
  - Develop the fundamentals of data structures and algorithms that form the basis of Big Data systems
  - Evaluate the features of databases e) Write Map and Reduce codes for distributed processing of data
  - Understand key concepts behind Big Data modelling and management and gain practical skills needed for modelling Big Data projects
  - Select appropriate data models that suit the requirements of data
  - Differentiate between a traditional Database Management System and a Big Data Management System
  - Retrieve data from Big Data management systems Execute simple Big Data integration and processing operations



This certificate program consists of 3 major modules:

## Module-1: Introduction to Big Data Analytics

1. Evolution of Big Data technologies
2. Characteristics of Big Data and the six Vs: Volume, Variety, Velocity, Veracity, Valence and Volume
3. Introduction to Data Science and how it helps in getting value out of Big Data
4. Big Data use cases such as fraud prevention, security intelligence, price optimization, recommendation engines, preventive maintenance and operational efficiency
5. Foundations for Big Data Systems and Programming including Distributed File Systems, scalable computing, cloud computing and cloud service models



## Module-2: Big Data Fundamentals and Platforms

### Big Data Fundamentals

1. Fundamentals of databases including topics such as relational databases, tables, data types and SQL and NoSQL databases
2. Linear data structures including Hashtables, Hashmaps and Bloom Filters
3. Non-linear data structures including Binary Search Trees and KD Trees

### Big Data Platforms

1. Features of popular Big Data systems and platforms (such as Hadoop)
2. Algorithm Design using MapReduce f) Data Storage and batch processing operations

## Module-3: Big Data Processing, Management and Analytics

### Big Data Management

1. Concepts including data ingestion, data storage, data quality, data operations, scalability and security
2. Concepts including relational and semi structured data models, data formats, data streams and data lakes
3. Structures, operations and constraints of Data Modelling
4. Differences between DBMS and BDMS

### Big Data Processing

1. Fundamentals of Big Data integration and processing
2. Fundamentals of Big Data pipelines and workflows
3. Fundamentals of Big Data Pipelines and analytical operations in Big Data Pipelines

### Big Data Analytics

1. Fundamentals of Big Data Analytics
2. Building visualizations using Big Data
3. Introduction to key tools in the Spark toolkit including Spark MLlib and GraphX



## Module-4: Hands-on Projects

Learn practical applications in Big Data Analytics

## WHO WILL YOU LEARN FROM?



### PROGRAM FACULTY

**Dr. Raju Pandey** is a Professor Emeritus in the Computer Science department at the University of California at Davis, where he developed and taught graduate and undergraduate courses in programming languages, operating systems, distributed systems, Internet of Things, Wireless sensor networks, Web-based systems, and compilers. He is also the CEO and founder of Thinking Books, a software Infrastructure and Tools company.

Dr. Pandey has a deep interest in math and computer science education and has developed novel interactive methods and tools for teaching both algorithmic and system aspects of Computer Science courses.

- Dr. Pandey's first startup, SynapSense, was a pioneering IoT company, later acquired by Panduit.
- His research and entrepreneurial interests lie in AI, Programming Languages, Blockchain, Internet of Things, Cloud, Security, and Privacy. Specifically, his interests are driven by the need to build software systems that are easier to build, analyze and deploy.
- In this regard, he has developed a novel software platform for building multi-platform AI, Blockchain, Mobile, and IoT applications. The platform includes a next-generation programming language, Ankur, that Dr. Pandey has designed and implemented. The platform will enable development of AI applications in which both algorithm-driven (deterministic) and data-driven (non-deterministic) components of AI applications can be integrated seamlessly.
- In addition, he consults extensively with companies on AI, Blockchain, IoT, Cloud, Mobile Computing, and Distributed Systems.
- He has published 40+ papers in conferences and journals and holds 16+ patents in software, visualization, wireless networks, data analytics, security, and control systems.
- Dr. Pandey holds a B.Tech. degree in Computer Science from IIT (Indian Institute of Technology), Kharagpur, and Ph.D. in Computer Science from the University of Texas at Austin.



## HOW WILL YOU LEARN?

- ✓ Online using desktop, laptop or mobile devices
- ✓ Learn at your own convenient time, and pace
- ✓ Video lectures delivered from a cloud LMS platform
- ✓ Quizzes are given remotely
- ✓ Hands-on projects, and industry case studies for the reinforcement of the learning



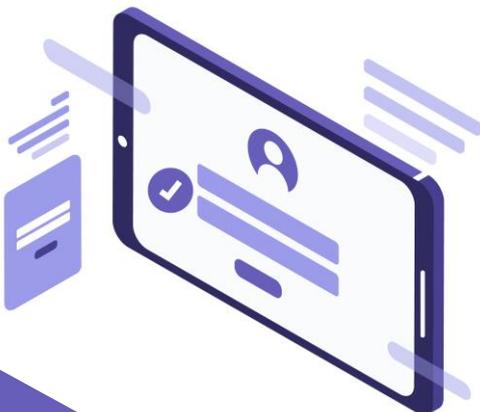
## WHAT IS THE DURATION OF THE PROGRAM?

- ✓ 6 weeks, around 6-8 hours per week, or a total of 36-48 hours
- ✓ Rolling enrolment allows you to start any time. The duration can be aligned to your requirements

## ABOUT iZen



iZen is a Talent Empowerment company, offering end-to-end solutions for skill development and employability, leveraging the power of AI and other digital technologies. The company was founded in Silicon Valley, California with a global vision to incubate innovation and to provide a platform that gives access to knowledge, skills, and advisory to empower the next generation workforce and students. iZen brings you internationally recognized standard programs, to set you apart and to future-proof your career.



## HOW DO I ENROLL IN THE COURSE?

<https://academy.iZen.ai>  
Email: [info@iZen.ai](mailto:info@iZen.ai)

