

Big Data Analytics

Training and Certification Program

13th June -23rd July 2022

Organized by Department of CSE, The NorthCap university, Gurugram



Top UG
Data Science
Programmes
in India

By Analytics India Magazine

2021

PIM



About The NorthCap University

The NorthCap University (NCU), founded in 1996 as Institute of Technology and Management (ITM) is known for delivering high quality of education that transforms lives and forges alliances to handle the new world order locally and globally. Over the last 25 years, the institute has transformed into a multi-disciplinary university and a destination of choice, offering industry and socially relevant undergraduate, postgraduate, and doctoral degree programs in Engineering, Applied Sciences, Management, Liberal Arts and Law. NCU has earned itself a place among the top universities in the world for its academic offerings with a perfect score of 5 stars in Teaching, Employability, Academic Development, Online Learning, and Inclusiveness and an overall rating of 4 stars by the prestigious Quacquarelli Symonds (QS) Stars Ratings, the world's largest rating agency. Further, The School of Engineering and Technology at The NorthCap University has been ranked amongst the Top 100 Engineering Institutions across the country in All India NIRF Ranking 2021. NCU has been rated among the **BEST PERFORMING** Universities in the category of 'Private-Self-Financed University' in the Atal Rankings of Institutions on Innovation & Achievements (ARIIA). The NorthCap University has been accredited by National Assessment and Accreditation Council (NAAC) and the Assessment Services to International Colleges (ASIC), UK in the category of 'Premier' Universities with 'Commendable' Grades. NCU has consistently achieved top rankings in different surveys for a good placement record of students and imparts value-based education.

Centre of Excellence Data Science

The Centre of Excellence Data Science, under the Department of Computer Science & Engineering, addresses specific development challenges by offering value-added programs & training for enhancing employability skills of our students as per the latest requirements of the industry in the domain of Data Science & Machine Learning. It aims to provide an ecosystem for innovation to thrive and embrace entrepreneurship and create quality human capital. The Centre is well equipped with **iMac machines with 27-inch iMac retina 5K display, 3.8GHz 8-core 10th-generation Intel Core i7 processor, turbo Boost up to 5.0GHz, 8 GB RAM and 1TB SSD storage, Radeon Pro 5500XT graphics card with 8GB of GDDR6 memory** perfectly suitable for undertaking research in emerging areas.

Data Science Program at NCU has been ranked among top UG & PG Data Science programs in India, according to the popular Analytics India Magazine survey . The Centre is offering Six Weeks Training and Certification Program in Big Data Analytics. The Big Data Analytics Training and Certification program has been created with a complete focus on the practical aspects of Big Data Analytics. It is a comprehensive program designed by experts from Industry and Academia and the best starting point for anyone interested in either moving into or up in the Big Data industry.

Big Data Analytics: Training and Certification program

Prerequisites

General computer skills and access to the Internet for research and data gathering.

Who should pursue this certificate?

- Programming Developers and System Administrators
- Experienced working Professionals and Project Managers
- Mainframe Professionals, Architects, and Testing Professionals
- Business Intelligence, Data Warehousing, and Analytics Professionals
- Academic Professionals and Research Scholars
- Graduates and undergraduates eager to learn Big Data Analytics

Resource Persons (Experts from Industry and Academia)

- IIT Roorkee certified
- BITS Pilani certified
- IBM Certified

Duration & Time

- Total of six weeks
- Total 90 Hours

Course Fee

- **INR 30,000/-**
- **Register before 30th April 2022** and avail **Early bird discount of 30%**

Registration Details

Registration Link : <https://tinyurl.com/yckxcyk7>

Contact Details :

Dr. Meghna Sharma , Associate Professor,

Email: meghnasharma@ncuindia.edu.

Dr. Poonam Chaudhary, Assistant Professor,

Email: poonamchaudhary@ncuindia.edu.

Phone: + 91 124 2365811 / 12 / 13

Program Overview

The Big Data Analytics training and certification program prepares you to apply the science of Data Analytics to the modern phenomenon of “Big Data Analytics,” that is, data whose scale and complexity offer vast challenges and opportunities. The Big Data Analytics Training and Certification program is designed to make you a certified Big Data practitioner by providing you rich hands-on training on Hadoop ecosystem and best practices about HDFS, MapReduce, Hive, Pig and Spark.

In this 100 hours long training and certification program with 7 tools and 2 hands-on projects, the aspirants will evaluate trends and understand the process of Big Data. It will take them through platforms, processing, management, and analytics. Leverage this program to grow within the organization, switch to new profiles, explore tremendously through interactive sessions and industry projects.

Outcomes

- Fundamentals of Hadoop and YARN and write applications using them
- HDFS, MapReduce, Hive, Pig tools expertise.
- Spark, Spark SQL, Data Frame, RDD, and MLLib writing Spark applications
- Practicing real-life projects using Hadoop and Apache Spark
- Be equipped to clear Big Data Hadoop Certification

Career Opportunities



Data Analyst
Database Administrator
Data Scientist
Big Data Architect
Big Data Researcher



Database Manager
Big Data Engineer
ML Engineer
Business Analyst
Big Data Visualizer



Data warehouse Manager
Data Modeler
Portal Administrator
Data Mining Analyst
Chief Data Analyst

Glimpse of Program

Course Contents

Week 1 : 3 Hours / Day

Day	Theory	Practical
1	Introduction to Big Data- Big Data Ecosystem, Applications, Key Components in Big Data Analytics Environment.	Familiarization of Hadoop Components
2	Introduction to Apache Hadoop, Hadoop Ecosystem	Installation of Hadoop in Windows
3	The Apache Hadoop Framework: Basic Modules, HDFS – Architecture, Configuration, Design & Role in Hadoop.	Hadoop services on GCP
4	Installation of Hadoop, HDFS Commands Hadoop MapReduce Framework- Introduction to MapReduce, Map/Reduce Framework, Data flow in MapReduce	Installation of Hadoop in Windows
5	HDFS commands	Implementation of HDFS commands

Week 2 : 3 Hours / Day

Day	Theory	Practical
1	MapReduce real-world Applications, Mapper and Reducer	Familiarization of MapReduce Components
2	Various terminologies in MR like Input Format, Output Format, Shuffle, and Sort	Basic structural understanding of MapReduce classes
3	Partitioners and Combiners	Mapper, Reducer, Partitioner and Combiner flow process
4	MapReduce Programming with MRJob package	Implementation of MapReduce Program
5	MapReduce Programming with MRJob package	Implementation of MapReduce Program

Week 3 : 3 Hours / Day

Day	Theory	Practical
1	Introduction, Apache Pig, Pig Architecture	Familiarization of PIG scripting modes
2	Program structure & execution process	Basic structural understanding of a PIG script
3	The available functions in Pig, Bags, Tuples, and Fields	PIG scripts
4	Working with Pig in MapReduce and local mode	IPIG scripts
5	Loading of data, Limiting data to 4 rows, storing the data into files and working with Group By, Filter By, Distinct, Cross	IPIG scripts

Week 4 : 3 Hours / Day

Day	Theory	Practical
1	Apache Hive-Hive Architecture, PIG Vs hive, Installation, Hive Data Types, HIVEQL	Familiarization of HIVE environment
2	Create, alter & drop table and database	HIVE scripts
3	Loading data with different filters	HIVE scripts
4	Hive Partitions & Buckets, Hive Queries	HIVE scripts
5	Hive Indexes and view, Hive join and Subquery, Hive Built-in function	HIVE scripts

Week 5 : 3 Hours / Day

Day	Theory	Practical
1	Introduction of Spark, Components of Spark Unified stack, Spark Architecture, Resilient Distributed Dataset (RDD): Create Parallelized collection & external Datasets. Various operations of RDD	Familiarization of Spark ecosystem
2	The execution flow in Spark, Overview of RDD persistence, Spark terminology, distribution shared memory vs. RDD, RDD limitations, RDD Lineage.	Basic structural understanding of Spark program
3	Key-value pair for sorting implicit conversions like CountByKey, ReduceByKey, SortByKey, and AggregateByKey	RDDs with function
4	Creating RDDs, RDD partitioning, Types of RDDs, Operations and transformation in RDD, RDD action for the collect, count, collect map, save-as-text-files, and pair RDD functions	Spark Program implementation
5	Spark SQL and Spark DataFrames	Spark SQL and Spark data frame implementation

Week 6 : 3 Hours / Day

Day	Theory	Practical
1	Introduction to Spark MLlib	Capstone Project
2	Machine Learning algorithms supported by MLlib	
3	Classification techniques using Spark MLlib	
4	Clustering techniques using Spark MLlib	
5	Case study	

Why NCU?

Ranked amongst
Top100 Institutions in
Engineering Category, NIRF 2021.

Legacy of 25 years,
11000+ Strong
Alumni.

Well-qualified faculty
and learner-centered
collaborative pedagogy.

Rated Overall
4 Stars - QS World University
Ratings, 2021.

Academic collaboration with
Top Ranked International
Universities.

Analytics India Magazine
Top UG & PG Data Science
Program in India

Why Big Data Analytics?

As per NASSCOM, by 2025,
the Indian analytics sector to
grow to USD 16 Billion.

Demand supply gap of
1,40,000 AI & Big Data
Analyst Professionals.

India to become the
Third Largest Market
in the world for Data Analytics.

Demand Across Industries
ranging from Auto, Healthcare, Consulting,
FMCG, E-Commerce, Retail, Banking, etc.

Program Highlights



24X7 Content Repository &
Best-In-Class Support



Eminent and Certified Experts



Face-to-face &
Project Based Learning



Hands-on exposure &
Industry Connect



Placement ready outcomes
& Career advancement



Curriculum in sync with
the latest industry needs

Patrons



Prof. Prem Vrat,
Pro-Chancellor ,
The NorthCap University, India



Prof. Nupur Prakash,
Vice-Chancellor ,
The NorthCap University, India

Program Chair



Prof. Rita Chhikara ,
HOD CSE
The NorthCap University, India



Dr. Meghna Sharma,
Associate Professor,
The NorthCap University, India



Dr. Poonam Chaudhary
Assistant Professor,
The NorthCap University, India

Program Conveners

Our Partners



Association for
Computing Machinery
Student Chapter



Brain Mentors Pvt. Ltd.
We Develop IT Brains



POWERED BY
Arizona State University

HUDA Sector 23-A, Gurugram, Haryana, India - 122017

www.ncuindia.edu