NPN Training

Training is the essence of success and we are committed to it.

Apache Spark Online Training

Apache Spark In-depth Training Program

About Us

NPN Training is a software training institute which believes that technology has to be learnt under experienced practitioners and that technology training has to be industry relevant and tailor made to the real world. At NPN Training importance is given to experiential learning where you learn from **Real case studies** and **Live projects**. Our aim is not just to land our students in better jobs but to ensure that they are prepared to become the best software engineers

---

NPN Training
#35 31st main BTM 2nd Stage,
Bengaluru, Karnataka 560068
www.npntraining.com

Contact : +91 9535584691 / +91 8095918383

www.youtube.com/NPNTraining

www.facebook.com/NPNTraining

Courses we offer :

Big Data Engineering Online Training | Big Data Live Project Training | Big Data Architect Masters Program | AI with Machine Learning & Deep Learning | Full Stack Test Automation | Selenium Live Project Training

About the Program

The **Apache Spark and Scala Developer Training Program** is designed to empower professionals to develop relevant competencies and accelerate their career progression in Big Data Hadoop Spark technologies through complete Hands-on training.

We are ranked one among the **Best Apache spark training institutes in Bangalore**, all our courses and coaching classes are conducted in compliance with industry standards that help beginners, experts and professionals in sync with present industry trends guiding them to face and solve real industry challenges confidently. Our team of skilled and certified professionals are specialists in delivering and imparting their excellence and experience to the students.

We believe in the philosophy **"Learn by doing"** hence we provide complete **Hands-on training** with a **real-time project development**.

Course Duration

60 hours extensive class room training.
30 sessions of 2-3 hours each.
Course Duration : 3 Months

Assignment

For each of the module multiple Hands-on exercises, assignments and quiz are provided in **Google Classroom**

Project work

We follow **Agile Methodology** for the project development. Each project will have a Feature **Study** followed by **User stories.**

Mock Interview

There will be a dedicated **1 to 1 interview call** between you and a Big Data Architect. Experience a real Mock Interview.

Forum

We have a community forum for all our students wherein you can enrich their learning through peer interaction and knowledge sharing.

Certification

From the beginning of the course, you will be working on a project. On completion of the project NPN Training certifies you as a "**Big Data Engineer**" based on the project.

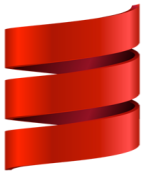
Trainer Profile

Naveen

12+ years of experience in IT with vast experience in executing complex projects using Java, Micro Services , Big Data and Cloud Platforms. I found **NPN Training** Pvt Ltd a India based startup to provide high quality training for IT professionals. I have trained more than 3000+ IT professionals and helped them to succeed in their career in different technologies. I am very passionate about Technology and Training. I have spent 12 years at **Siemens, Yahoo, Amazon and Cisco**, developing and managing technology.

Program Structure

Being a Big Data Developer requires you to be a master of multiple technologies, and this program will ensure you to become an industry-ready Big Data Developer / Architect who can provide solutions to Big Data projects.



Comprehensive Scala Programming : Functional + Object Oriented

4 Sessions

Learning Objectives:

- ✓ Use the REPL (the Scala Interactive Shell) for experimentation and fun
- ✓ Define and use function literals and higher order functions
- ✓ Create factory methods in companion objects
- ✓ Use case classes to get lots of free, best practice functionality
- ✓ Obtain an in-depth grounding in the Scala collections and examine the enormous potential and power in the collections API

- ✓ Know how to mix Java and Scala on a project, if you need to.

Topics:

- ❑ **Module 01 :** Scala Language Fundamentals
- ❑ **Module 02 :** Functional Programming using Scala
- ❑ **Module 03 :** Object Oriented Programming
- ❑ **Module 04 :** Collections

Apache Spark 2.x : In-Memory Cluster Computing Framework

12 Sessions

Description :

- ✓ Understand Apache Spark's primary data abstraction (RDDs) to process and analyze large data sets
- ✓ Optimize Spark jobs through partitioning, caching, and other techniques.
- ✓ Learn how RDDs caching works and use it for advanced performance optimization
- ✓ Build, deploy, and run Spark scripts on Hadoop clusters
- ✓ Understand how Spark SQL lets you work with structured data
- ✓ Understand how Structured Streaming lets you process continuous streams of data in real time
- ✓ Understand development life-cycle of Apache Spark Applications in Scala
- ✓ You will learn to Provision your own Databricks workspace

Topics

Apache Spark Core

- ❑ **Module 01 :** Getting Started with Apache Spark
- ❑ **Module 02 :** Learning Spark Core : RDD's

Batch Processing - Spark SQL

- ❏ **Module 01 :** Spark SQL : DataFrame + DataSet API
- ❏ **Module 02 :** DataFrame + DataSet API Operations
- ❏ **Module 03 :** Exploring Data Source API
- ❏ **Module 04 :** Deep Dive Spark SQL

Stream Processing

- ❏ **Module 01 :** Structured Streaming
- ❏ **Module 02 :** Advance Structured Streaming

Application Deployment

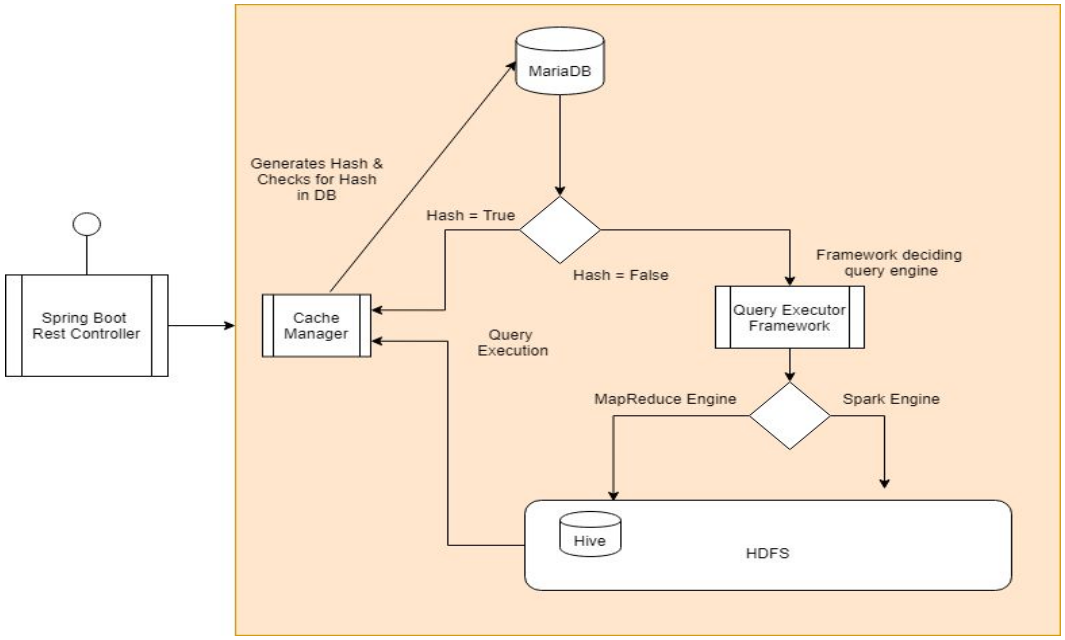
- ❏ **Module 01 :** Deploying Spark on Production
- ❏ **Module 02 :** Performance Tuning of Spark Applications

[Capstone Project] - Query Cache Scheduler

⚠ Sample projects of previous batches :

Project description

The purpose of this project is reder the query response faster than actually executing. The project aims to provide flexibility to cache the previous executed query in MariaDB and render for any future request. The project provides a convenient way to schedule or trigger based on the configuration.



⚠ Sample projects of previous batches :

Project description

The purpose of this project is to dynamically allocate resources for a Hive job at run time. The job details are present in a XML file read during execution . Based on the job name present in the XML file, the business logic lookup a XML file with the matching job name and assign dynamically the hive job a Queue in capacity scheduler , set multiple resource values and start running the job and the job execution status is visible in Hadoop resource manager Web UI.

Apache Kafka : Message Queuing System

4 Sessions

Learning Objectives:

- ✓ This section of the training will help you gain proficiency to work with
- ✓ Kafka Cluster Architecture in depth
- ✓ Master Concepts such as Topics, Partitions, Brokers, Producers, Consumers
- ✓ Create fault-tolerant clusters with topics replication across multiple brokers
- ✓ How to create Kafka cluster with multiple Brokers
- ✓ Kafka in-depth Producer API Programming in Java
- ✓ Kafka in-depth Consumer API Programming in Java

Topics

- ❑ **Module 01 :** Getting Started with Kafka
- ❑ **Module 02 :** Exploring Kafka Core API
- ❑ **Module 03 :** Deep Dive Kafka Producer API
- ❑ **Module 04 :** Deep Dive Kafka Consumer API

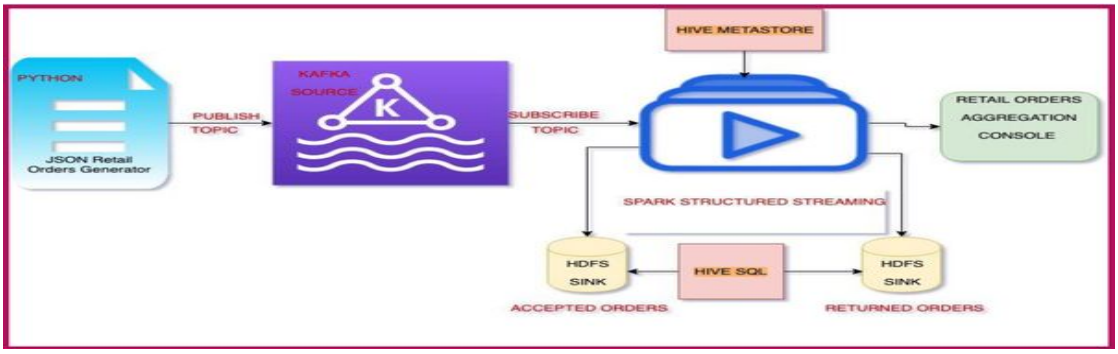
[Capstone Project] - Building Real-time data pipeline

⚠ Sample projects of previous batches :

Project description

The purpose of the project is to subscribe to KAFKA topic from spark structured streaming read stream API and the JSON records are generated automatically using the python retail data generator script. We use the HDFS sink of CSV format to write the accepted orders and rejected orders to different locations in HDFS and the aggregation of orders like average amount and the count of order Quantity is written to the console. The hive tables orders and orders reject is used to query the accepted and rejected retails orders

Project flow



Candidate Evaluation

We follow **Assessment and Project based** approach to make your learning maximized. For each of the modules there will be multiple Assessment/Problem Statements.

Each of the Assessments in the **Google Classroom** helps students to grasp the concepts taught in class and apply in business problem scenarios.

- **Module Test:** You will have a test for each of the topics covered in the previous class/week. These tests are usually for 15-20 minute duration.
- **Hands-on Test** : Each candidate will be given an exercise for evaluation and the candidate has to solve.
- **Coding Hackathon** : Coding hackathon will be conducted during the middle of the course. This is conducted to test application of concepts to the given problem statement with tools and techniques that have been covered and to solve a problem quickly, accurately.

- **Capstone Projects** : At the end of each course there will be a Real-world Capstone Project that enables you to build an end-to-end solution to real world problems. You will be required to write a project report and present to the audience.

Process We Follow For Project Work

We follow **Agile Methodology** for the project development. Each batch will be divided into **Scrum Teams** of size 4-5 members and before the project implementation we will start with **Feature Study** where we will discuss various aspects of a feature like High level design of the component, Framework to use, challenges, test strategies and many more. Once the Feature Study is finalized we will start implementing the feature by splitting into **User Stories** and **Tasks** and also help candidates to define **Definition of Done** for each User Stories and then finally we will be discussing Test **Strategies** for testing the project or a feature.