

GIGASPACE INSIGHTEDGE TRAINING

This training is designed to provide you with the knowledge required to build high throughput, low latency applications for scaling with GigaSpaces including analytical capabilities.

You will not only learn how to code such applications, but you will also gain a better understanding of how GigaSpaces is a unique enabler of highly transactional, high volume, low latency applications, benefiting the analytical capabilities of InsightEdge platform as well as exactly what types of architecture GigaSpaces is best suited for.

AUDIENCE Developers | Architects | Data Scientist

KNOWLEDGE REQUIREMENTS Some previous coding experience – recommended scala, java or python

LENGTH 5 Days

COMPUTER REQUIREMENTS

- RAM: minimum 8 GB of RAM required for exercises and platform to operate.
- Disk Space: At least 50 GB of free disk space
- Internet connection
- User with sufficient privileges for creating environment variables and execute processes
- Trainees should have a user with Administrator privileges (to edit system files for environment variables)

SYSTEMS OPERATING SUPPORTED

- Linux (64bit)
- Mac OS (64 bit)

REQUIREMENTS HW CLASSROOM

- Projector 768*1024 minimum resolution
- White Board
- Erasable markers
- Internet connectivity for all participants

WHAT WILL BE LEARNING

- Spark for ML and leading DL frameworks
- Push-down predicate for ultra-low latency filter
- RDDs/DataFrames
- Streaming
- Big Data
- Scala
- Spark & Hadoop
- Graph processing, text mining, geospatial
- AnalyticsXterme
- Tableau

GIGASPACE'S INSIGHTEDGE TRAINING

AGENDA DAY 1

Lesson 1

Course Introduction

🕒 30 minutes

Lesson 2

Introduction to Big Data

🕒 30 minutes

Lesson 3

Introduction to Scala

🕒 150 minutes

Lesson 4

Intro to Spark & Hadoop

🕒 90 minutes

Lesson 5

InsightEdge Under the Hood

🕒 60 minutes

AGENDA DAY 2

Lesson 6

RDD API

🕒 150 minutes

- Lab-5.1 - RDD Transformations and Actions
- Lab-5.2 - RDD Word Frequency
- Lab-5.3 - RDD Cities
- Lab-5.4 - RDD Pair RDDs

Lesson 7

Structured API - DataFrame

🕒 50 minutes

- Lab-6.1 - DataFrame Loading and Saving Data
- Lab-6.2 - DataFrame Operations
- Lab-6.3 - DataFrame Using SQL
- Lab-6.4 - DataFrame Operations and SQL Usage
- Lab-6.5 - DataFrame Save and Load DataFrame from Grid

Lesson 8

Architectural Approaches

🕒 60 minutes

- InsightEdge
- Lab-8.1 - SQL Query Benchmark Spark vs InsightEdge
- Administration and Deployment

Lesson 9

InsightEdge

🕒 1 Hour

- Lab-8.1 - SQL Query Benchmark Spark vs InsightEdge

Lesson 10

Administration and Deployment

🕒 60 minutes

GIGASPACE INSIGHTEDGE TRAINING

AGENDA DAY 3

Lesson 11

Structured Streaming

🕒 60 minutes

- Lab-10.1 - Structured Streaming

Lesson 12

Microbatch Streaming

🕒 180 minutes

- Lab-11.1 - Save Stream to Grid
- Lab-11.2 - Twitter Notebook
- Lab-11.3 - Kafka and Geospatial
- Lab-11.4 - Spark Streaming Examples
 - Twitter Access
 - Using State

Lesson 13

Machine Learning

🕒 120 minutes

- Lab-12.1 - Machine Learning

AGENDA DAY 4

Lesson 14

GraphX

🕒 180 minutes

- Lab-13.1 - GraphX Creation Structural Operators
- Lab-13.2 - GraphX Connected Components
- Lab-13.3 - GraphX Neighbourhood Aggregation
- Lab-13.4 - GraphX Airline Demo

Lesson 15

Event Containers

🕒 30 minutes

Lesson 16

MemoryXtend

🕒 60 minutes

Lesson 17

Kubernetes

🕒 120 minutes

- Lab-16.1 - Kubernetes

GIGASPACE INSIGHTEDGE TRAINING

AGENDA DAY 3

Lesson 18

Flight Delay Demo

🕒 30 minutes

- Lab-17.1 - Flight Delay Demo

Lesson 19

End 2 End Exercise

🕒 150 minutes

Lesson 20

Grafana

🕒 45 minutes

Lesson 21

Tableau

🕒 45 minutes

Lesson 22

Jupiter

🕒 30 minutes

Lesson 23

AnalyticsXterme

🕒 60 minutes