

# Module 11 – Apache Spark Streaming

---

## Case Study I

edureka!

**edureka!**

© Brain4ce Education Solutions Pvt. Ltd.

## Case Study: Spam Detection

### Domain: Telecom

A telecom software provider is building an application to monitor different telecom components in the production environment. For monitoring purpose, the application relies on log files by parsing the log files and looking for potential warning or exceptions in the logs and reporting them. The POC we had been working on, for SPAM Detection on the data of telecom operator forum, has been accepted and the stakeholders has asked us to work on the real-time example for predicting SPAM messages.

#### Tasks:

This POC will focus on saved machine learning model for spam prediction with streaming data to do real-time prediction. Now with model and data pipeline ready, you are required to predict the spam message on the steaming data.

1. Modify the model application to train the model and persist it.
2. Create a new spark streaming application to predict the spam messages
3. Application will connect to the flume to retrieve the data
4. Load the model
5. Predict the SPAM messages and print the SPAM in the logs
6. Test the application by sending dummy data rows from the consumer

#### Hint:

Below is the example for Spam Messages:

- HAM: What you doing? How are you?
- SPAM: Sunshine Quiz! Win a super Sony DVD recorder if you can name the capital of Australia? Text MQUIZ to 82277