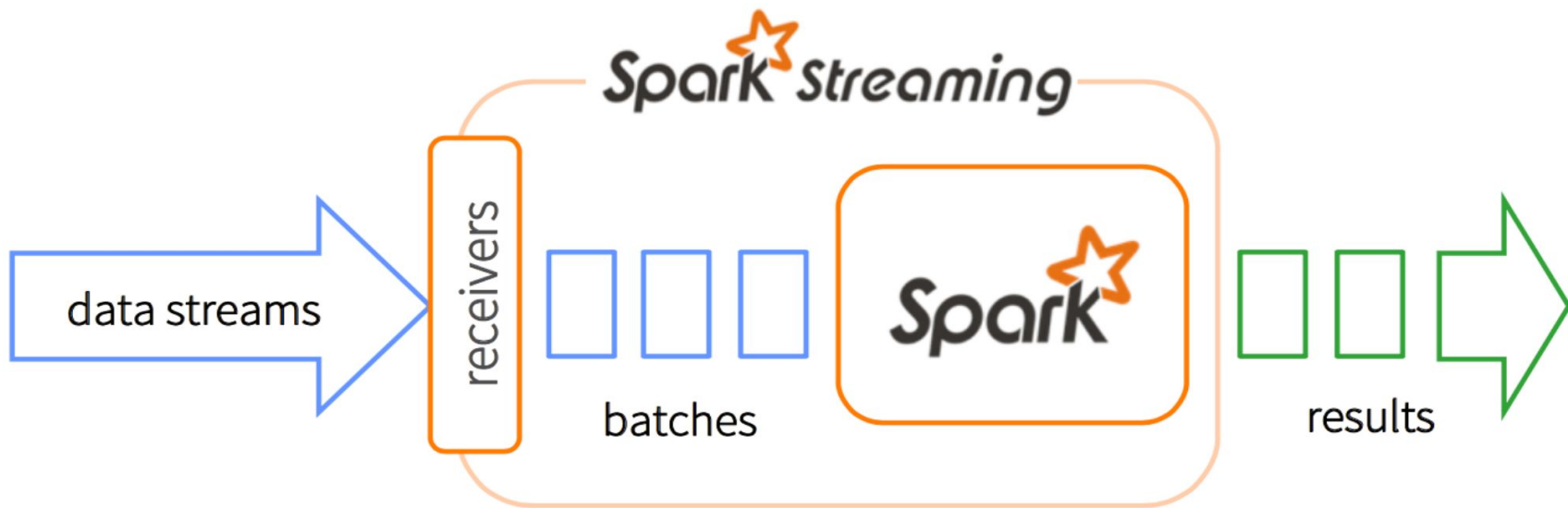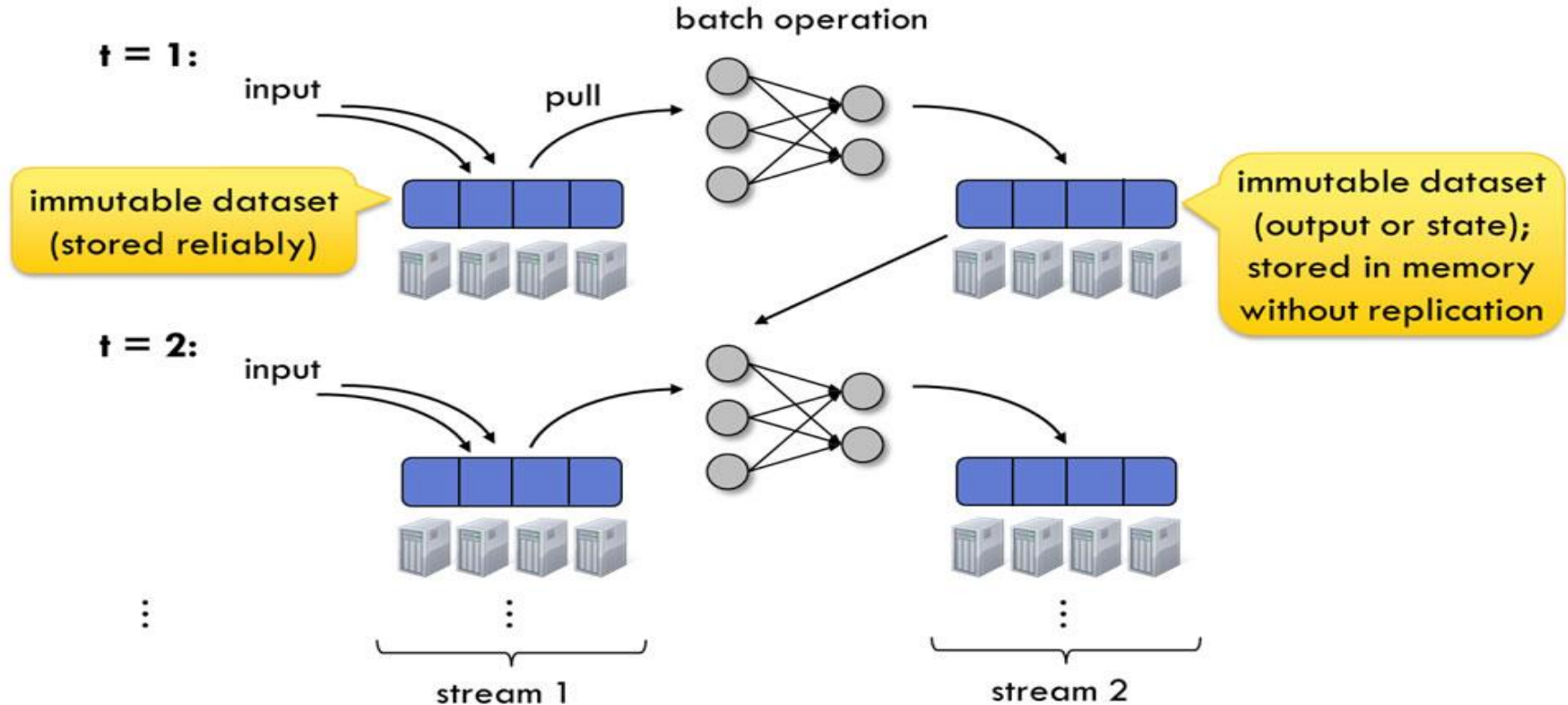# Spark Streaming

M Solaimani
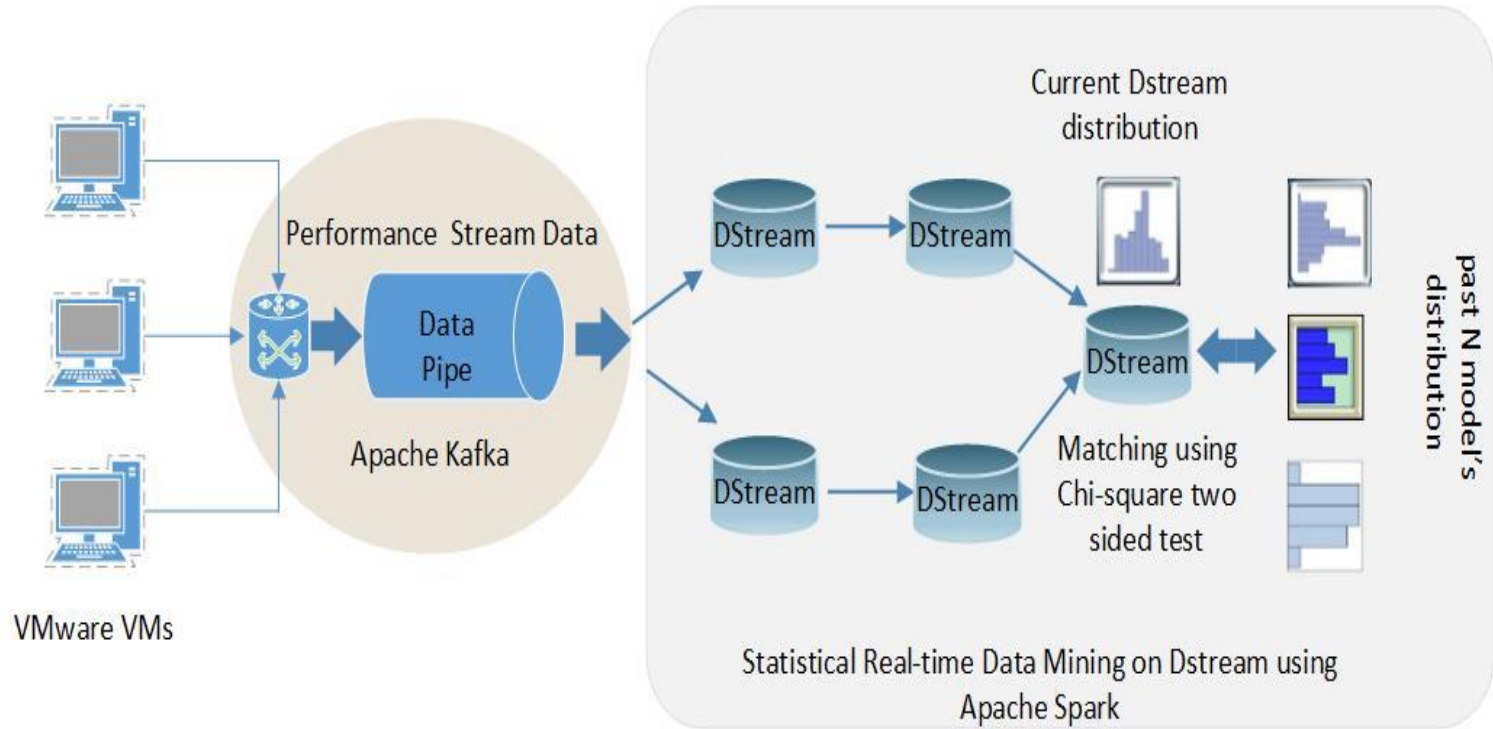
What is Spark Streaming

What is Spark Streaming

Application

Real-time anomaly detection using Apache Spark

Setup

Message Broker

Spark

Message Broker

Apache Kafka

- Transport message from producer to receiver
- Communication with IP
- Multiple producer and receiver with same topic
- Guaranteed message delivery

Message Broker

Apache Kafka Setup

- Version compatibility issue with Spark and Scala
- One of the compatible version is Kafka 0.9.0.1, Spark 1.6.1, Scala 2.10.5
- Download from https://www.apache.org/dyn/closer.cgi?path=/kafka/0.9.0.1/kafka_2.10-0.9.0.1.tgz

Kafka Setup

Follow the step to set up Kafka

https://kafka.apache.org/090/documentation.html#quickstart

Kafka Setup

- Topic
- Producer
- Consumer

Creating a Spark Streaming Project

- Use any IDE like Eclipse/IntelliJ IDEA
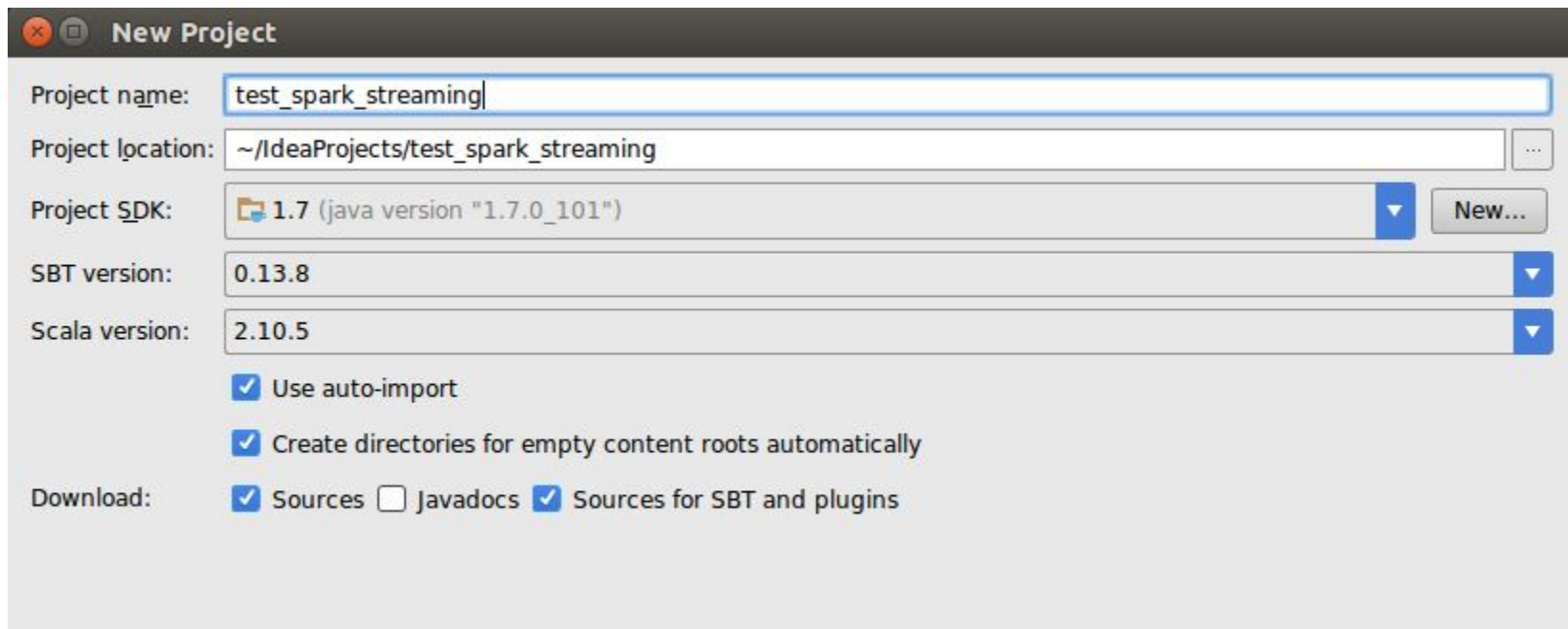- Install maven/sbt

Creating a Spark Streaming Project

- Install IntelliJ IDEA Community edition from
  https://www.jetbrains.com/idea/download/#section=linux
- Documentation for SBT setup in IntelliJ
  https://www.jetbrains.com/help/idea/2016.2/getting-started-with-sbt.html

Spark Install:

- Download Spark 1.6.1
  http://d3kbcqa49mib13.cloudfront.net/spark-1.6.1-bin-hadoop2.6.tgz
- Extract it and put it to /usr/local
- Name it to spark. You can create symbolic link if you want

Creating a Spark Streaming Project

New Project

Creating a Spark Streaming Project

SBT:

```
name := "test_spark_streaming"

version := "1.0"

scalaVersion := "2.10.5"

libraryDependencies ++= Seq("org.apache.spark" %% "spark-core" % "1.6.1",

 "org.apache.spark" % "spark-streaming_2.10" % "1.6.1",

 "org.apache.spark" % "spark-streaming-kafka_2.10" % "1.6.1"

)
```

Creating a Spark Streaming Project

Spark Streaming project is uploaded to elearning