skillspeed

for the serious learner

# Apache Spark and Scala

## Module 1: Getting Started / Introduction to Scala

© 2015 BlueCamphor Technologies (P) Ltd.

# Course Topics

skillspeed
for the serious learner

**Module 1**
Getting Started / Introduction to Scala

**Module 2**
Scala – Essentials and Deep Dive

**Module 3**
Introducing Traits and OOPS in Scala

**Module 4**
Functional Programming in Scala

**Module 5**
Spark and Big Data

**Module 6**
Advanced Spark Concepts

**Module 7**
Understanding RDDs

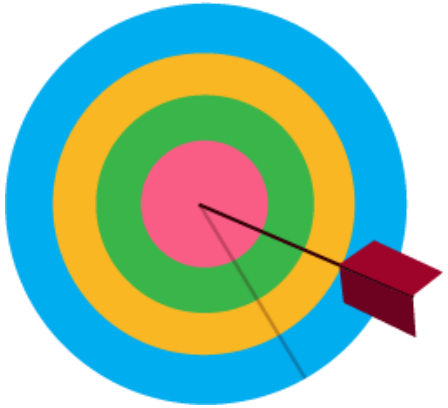
**Module 8**
Shark, SparkSQL and Project Discussion

© 2015 BlueCamphor Technologies (P) Ltd.

# Session Objectives

This session will help you to understand:

- Big Data
- IBM's Big Data Definition
- Some Big Data Examples
- Sparks Basics
- Why Spark ?
- Spark Components
- Scala Basics
- Why Scala ?
- Scala Job Trends
- Users of Scala
- Scala Frameworks
- Scala Usage
- Software Installation
- Scala Hands-on
- Scala community



© 2015 BlueCamphor Technologies (P) Ltd.

# Introduction to Big Data

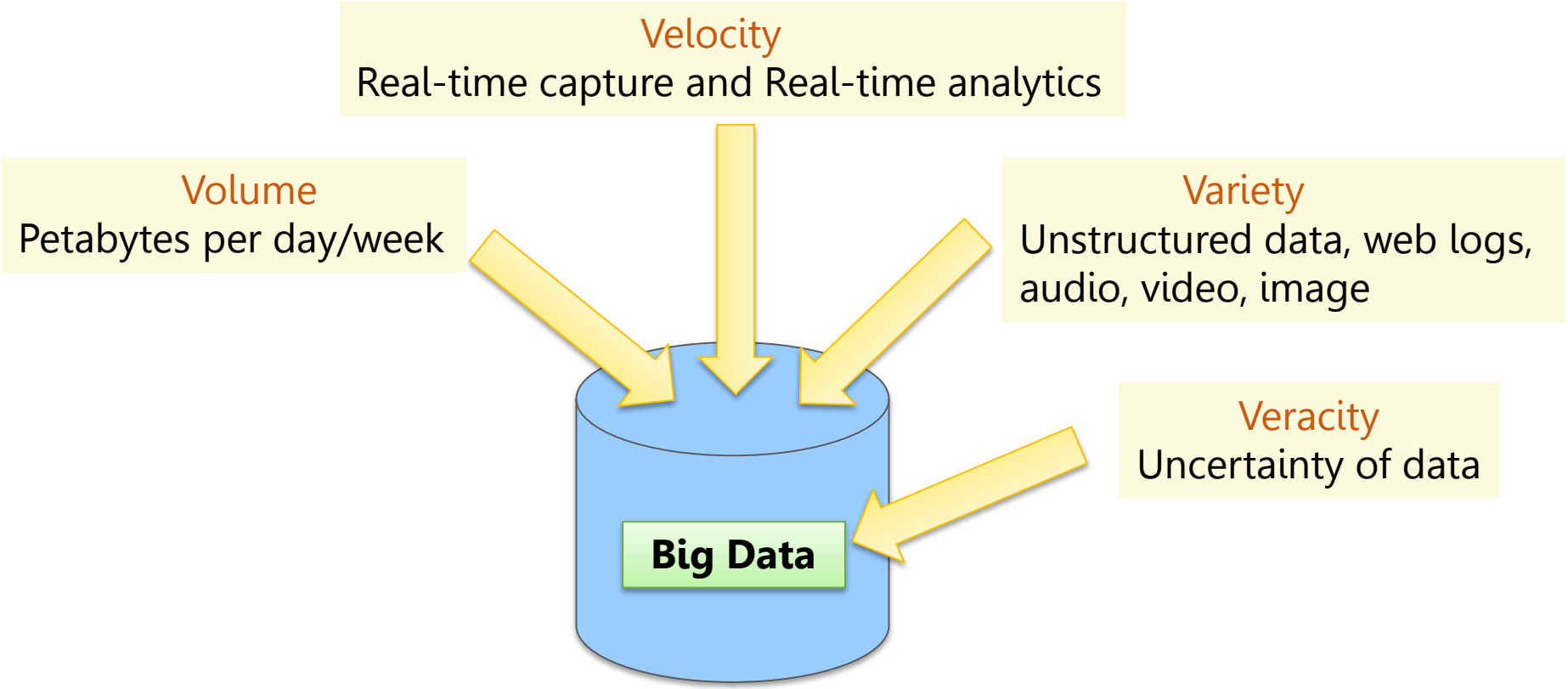

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate

The challenges of big data includes: analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy

© 2015 BlueCamphor Technologies (P) Ltd.

# IBM's Big Data Definition

- IBM's Definition – Big Data Characteristics
- http://www.ibmbigdatahub.com/infographic/four-vs-big-data/




© 2015 BlueCamphor Technologies (P) Ltd.

# Some Big Data Examples

- NYSE broadcasts several levels of data, including trade prices, sizes
- NYSE Technologies receives four to five terabytes of a data in a day and which is used for complex analytics, market surveillance, capacity planning and monitoring


NYSE generates about one terabyte of new trade data per day to perform stock trading analytics to determine trends for optimal trades

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – 1

Which of the following are the Big Data Solutions Candidates?

a) Processing 1.5 TB data everyday

b) Processing 30 minutes Flight sensor data

c) Interconnecting 50K data points (approx. 1 MB input file)

d) Processing User clicks on a website

© 2015 BlueCamphor Technologies (P) Ltd.

# Introduction to Spark



mongoDB

BIG DATA

hadoop

STORM

Spark

cassandra

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – Solution

Which of the following are the Big Data Solutions Candidates?

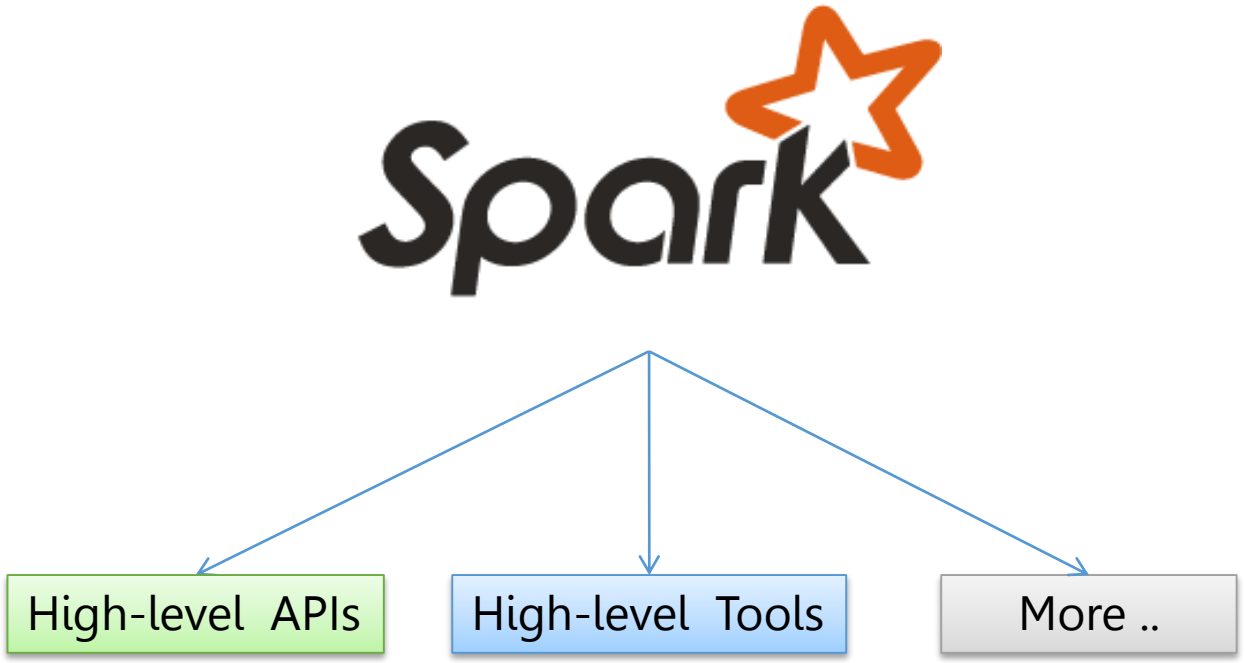
- a) Processing 1.5 TB data everyday
- b) Processing 30 minutes Flight sensor data
- c) Interconnecting 50K data points (approx. 1 MB input file)
- d) Processing User clicks on a website

**ALL** of the options are Big Data solutions Scenario. Even if the input size of the problem is small, the processing might make the scenario as Big Data Problem

© 2015 BlueCamphor Technologies (P) Ltd.

# Spark Basics

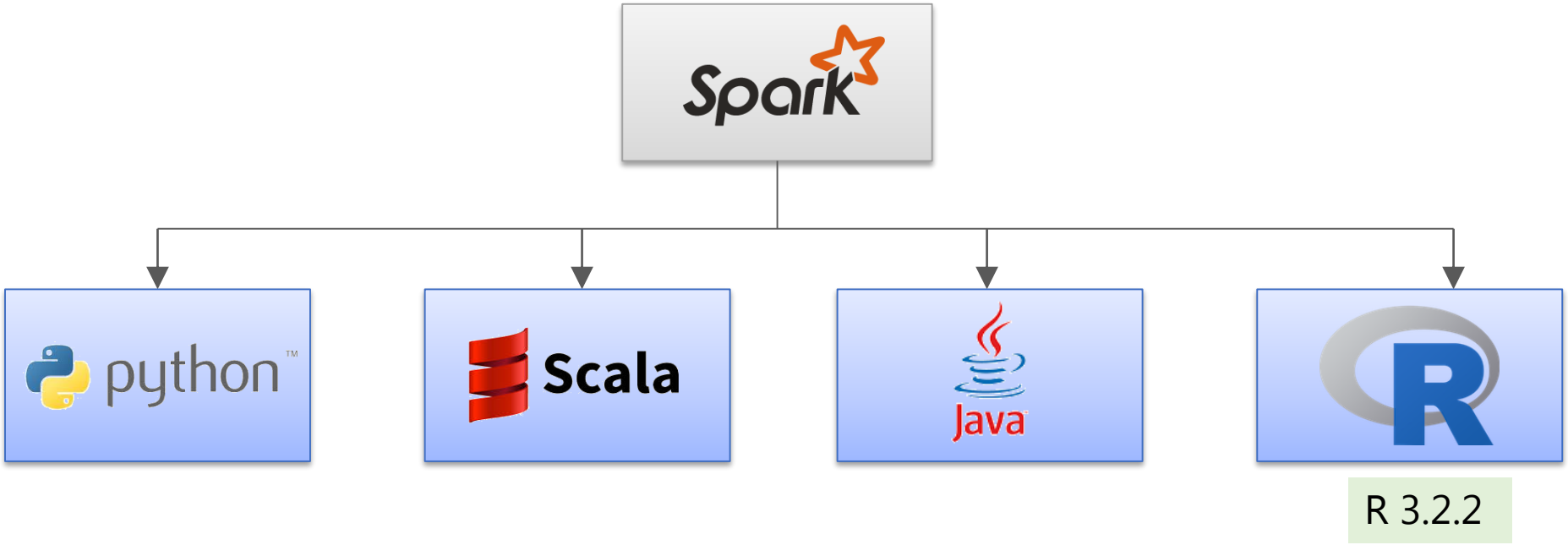
- Apache Spark is a general-purpose cluster in-memory computing system which is used for data analytics
- It provides high-level APIs in Java, Scala and Python and an optimized engine that supports general execution graphs
- Apache Spark Provides various high level tools like Spark SQL for structured data processing, R programming Language for analyzing large datasets and MLlib for Machine Learning etc.




© 2015 BlueCamphor Technologies (P) Ltd.

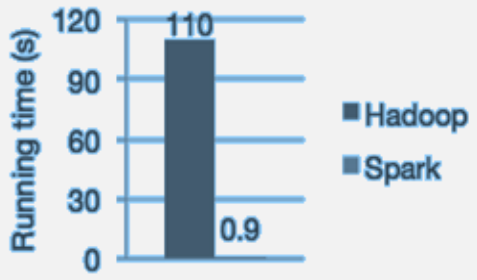
# Spark Basics (Cont'd)

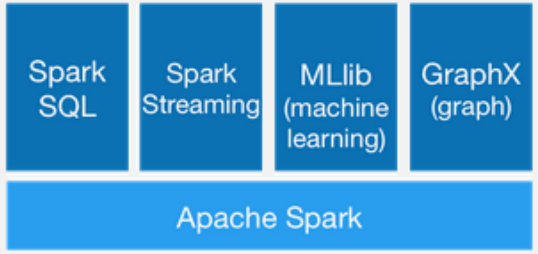
Spark framework is polyglot – It can be programmed in several programming languages (Java, Scala ,R 3.2.2 and Python supported)




© 2015 BlueCamphor Technologies (P) Ltd.

# Why Spark?

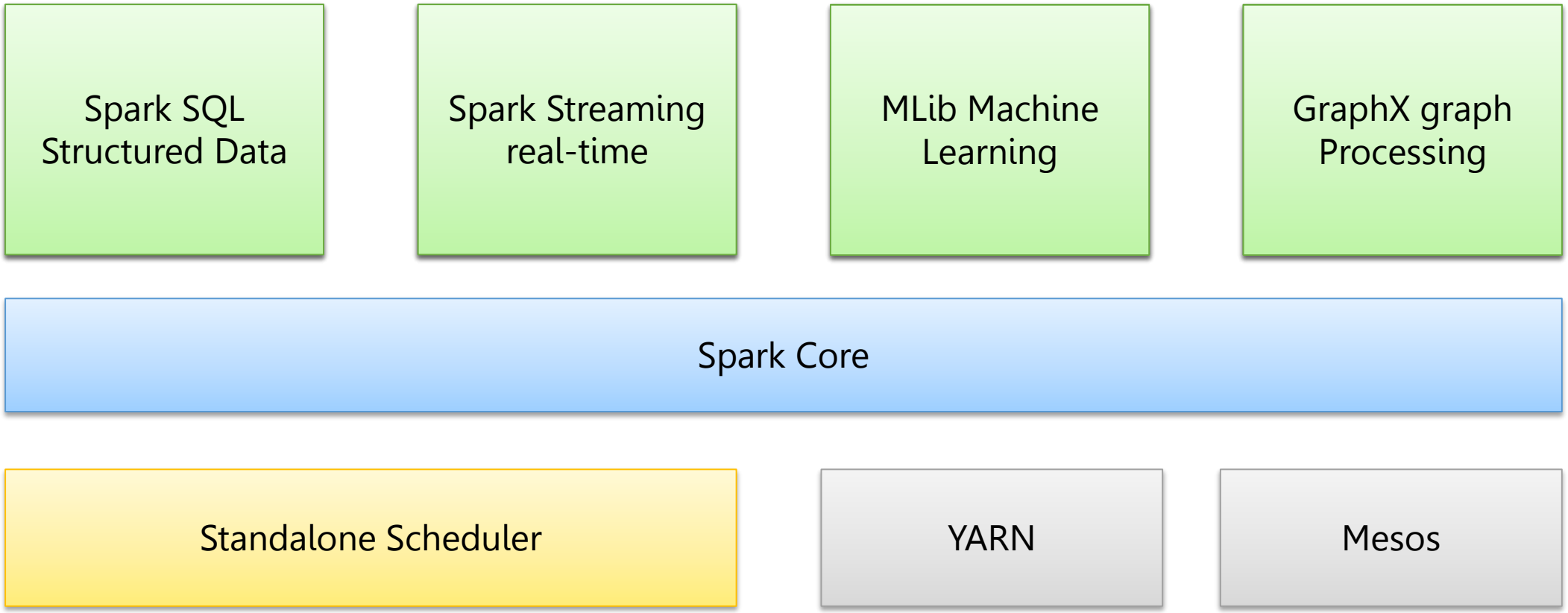



| | | |
|---|---|---|
| Speed<br>Run programs up to 100x faster than Hadoop Map Reduce in memory | Generality<br>Combine SQL ,streaming and complex analytics into one platform | Runs Everywhere<br>Spark runs on Hadoop, Mesos.standalone or in cloud |

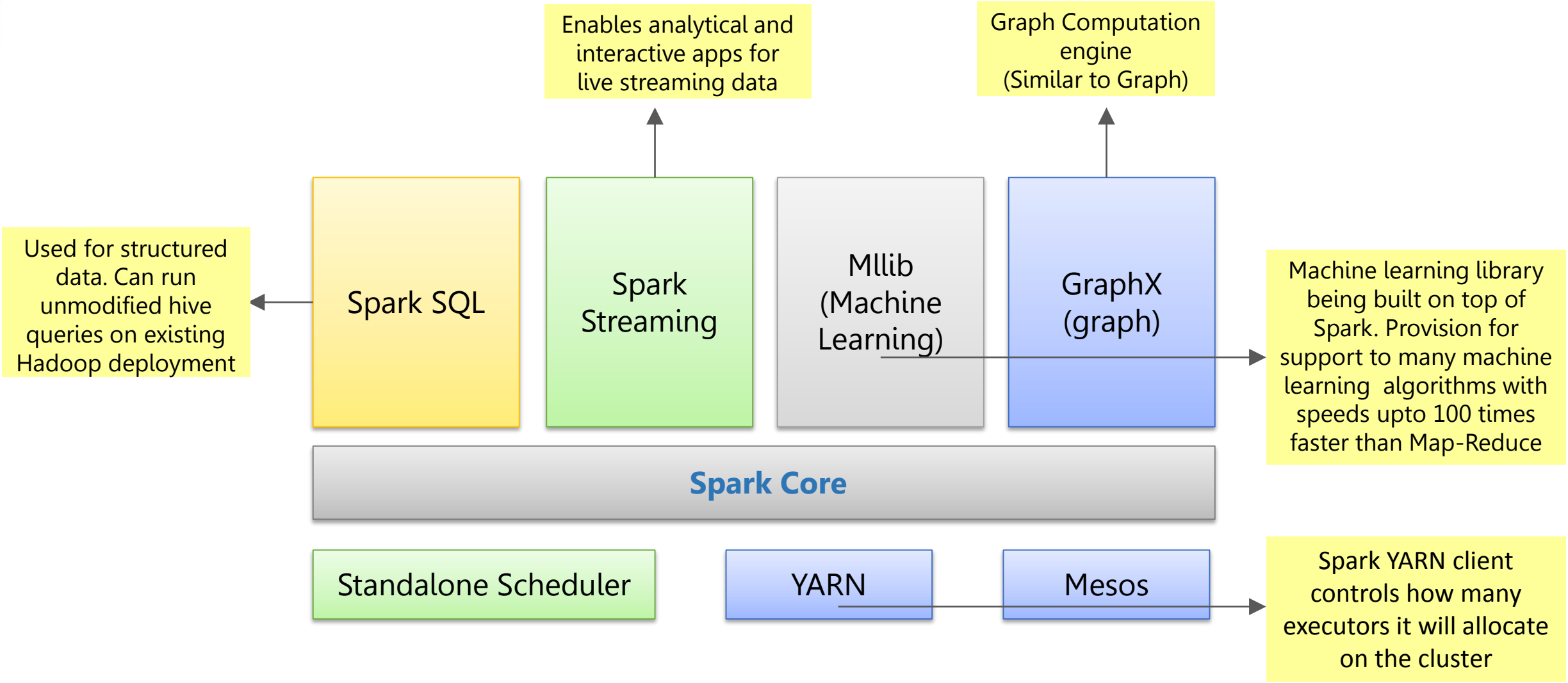
© 2015 BlueCamphor Technologies (P) Ltd.

# Spark Components




© 2015 BlueCamphor Technologies (P) Ltd.

# Spark Components (Cont'd)




© 2015 BlueCamphor Technologies (P) Ltd.

Spark + Scala = CREATES MAGIC

© 2015 BlueCamphor Technologies (P) Ltd.

# Introduction to Scala

Martin Odersky and his team started developing Scala in 2001

© 2015 BlueCamphor Technologies (P) Ltd.

# Introduction to Scala (Cont'd)

- Scala is a  general purpose programming language, multiparadigm object oriented, functional, scalable
- Aimed to implement common programming patterns in a concise, elegant, and type-safe way
- Supports both object-oriented and functional programming styles, thus helping programmers to be more productive
- Publicly released in January 2004 on the JVM platform and a few months later on the .NET platform

© 2015 BlueCamphor Technologies (P) Ltd.

# Introduction to Scala (Cont'd)

Scala is Statically Typed

- Statically typed language binds the type to a variable for its entire scope

```
scala> val x = "skillspeed"
x: String = skillspeed

scala> x = 10
<console>:8: error: reassignment to val
       x = 10
         ^
```

- Dynamically typed languages bind the type to the actual value referenced by a variable .Example : python

```
>>> x="skillspeed"
>>> x=20
>>> x=34.5
```

- Fully supports Object Oriented Programming
- Everything is an object in Scala
- Unlike Java, Scala does not have primitives
- Supports "static" class members through Singleton Object Concept
- Improved support for OOP through Traits

© 2015 BlueCamphor Technologies (P) Ltd.

# Why Scala?

- Scala is pure object-oriented language. Conceptually, every value is an object and every operation is a method-call
- Scala is also a functional language and supports immutable data structures
- Many big data technologies use Scala like Spark, Kakfka, Storm, Akka, Scalding and web frameworks like Play

© 2015 BlueCamphor Technologies (P) Ltd.

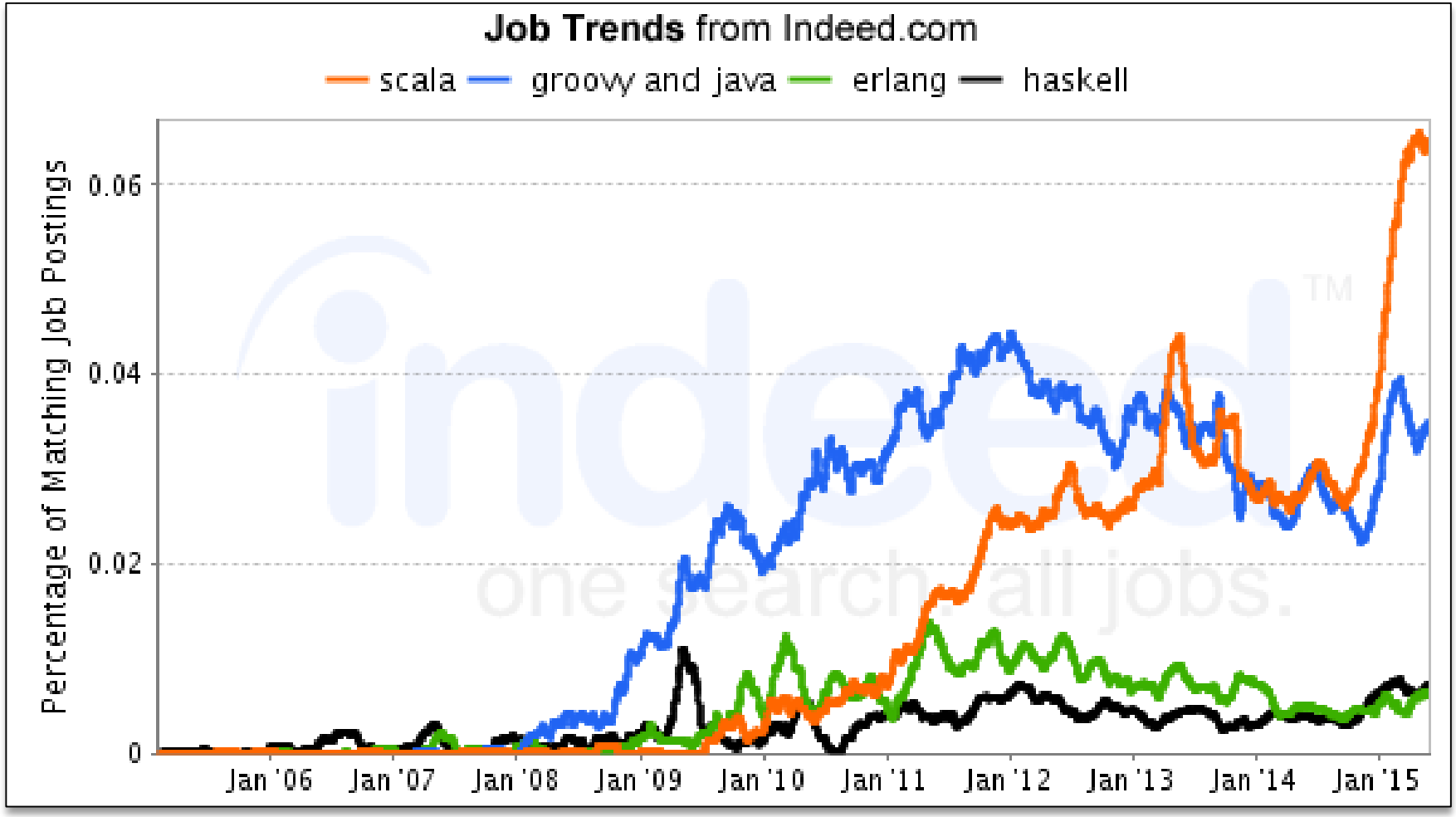
# Why Scala? (Cont'd)

Scala code compared to Java code

| Java | Scala |
|---|---|
| Java Code<br><br>List<String> list = new ArrayList<String>();<br>list.add("1");<br>list.add("2");<br>list.add("3"); | Scala Code<br><br>val list = List("1", "2", "3") |

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala – Job Trends




© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Users

# Scala Frameworks

### Play – For Web Development

Play is a high-productivity Java and Scala web application framework that integrates the components and APIs you need for modern web application development

### Scalding – For Map/Reduce

Scalding is a Scala library that makes it easy to specify Hadoop MapReduce jobs. Scalding is built on top of Cascading, a Java library that abstracts away low-level Hadoop details

### Akka – Actors Based Framework

Akka is a toolkit and runtime for building highly concurrent, distributed, and fault tolerant applications on the JVM. Akka is written in Scala

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Frameworks (Cont'd)

## Spark – In – memory Processing

Apache Spark is a general-purpose cluster in-memory computing system. It is used for fast data analytics and it abstracts APIs in Java, Scala and Python, and provides an optimized engine that supports general execution graphs

## Apache Kafka

Apache Kafka is publish-subscribe messaging rethought as a distributed commit log

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Usage

Scripting


Web Application

Messaging

Mobile Android Apps


Digital Subscriber Line

GUI (Graphical User Interface)

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – 2

Which Features are supported by Scala?

a) Less error prone functional style

b) High maintainability and productivity

c) High scalability

d) High testability

e) Provides features of concurrent programming

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – Solution

Which Features are supported by Scala?

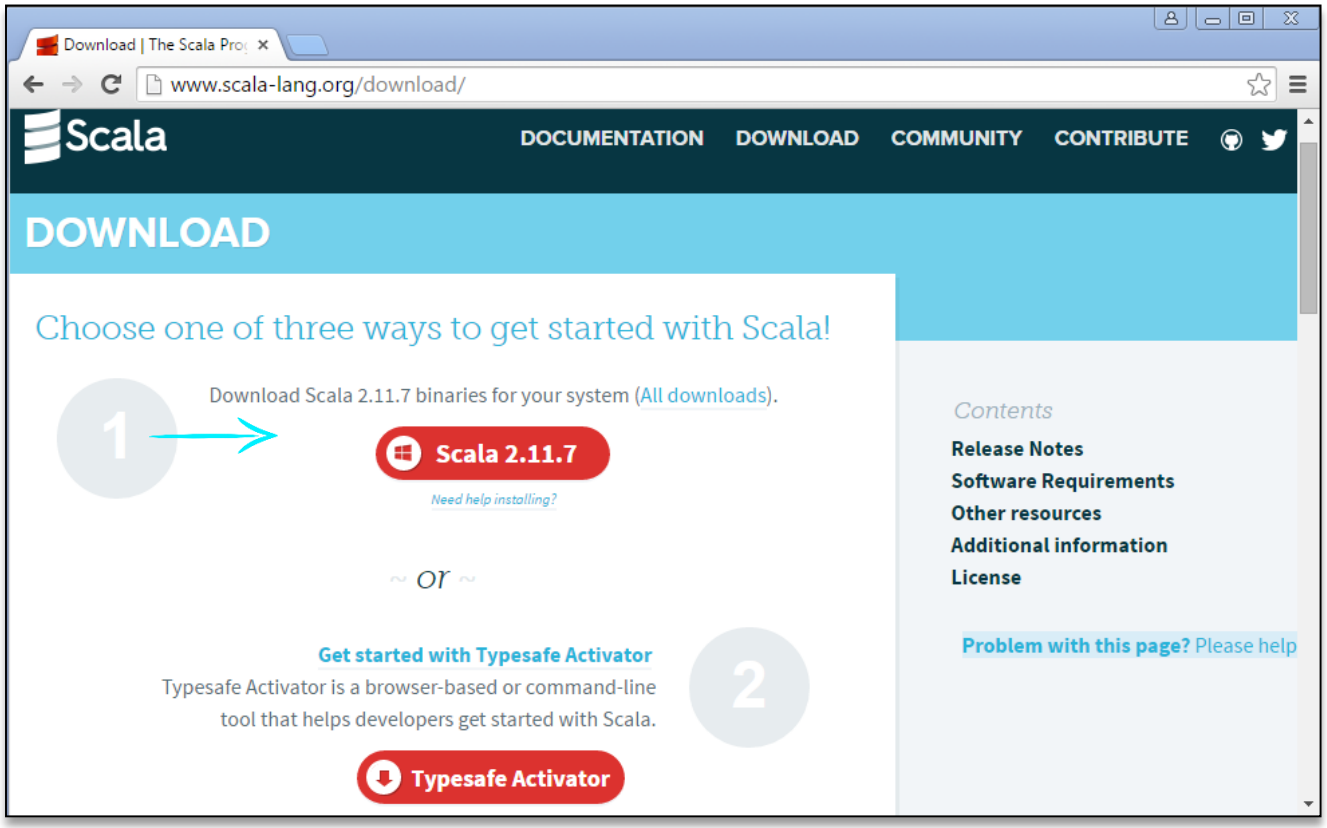
- a) Less error prone functional style
- b) High maintainability and productivity
- c) High scalability
- d) High testability
- e) Provides features of concurrent programming

All of these are the features of Scala

© 2015 BlueCamphor Technologies (P) Ltd.

# Software Installation

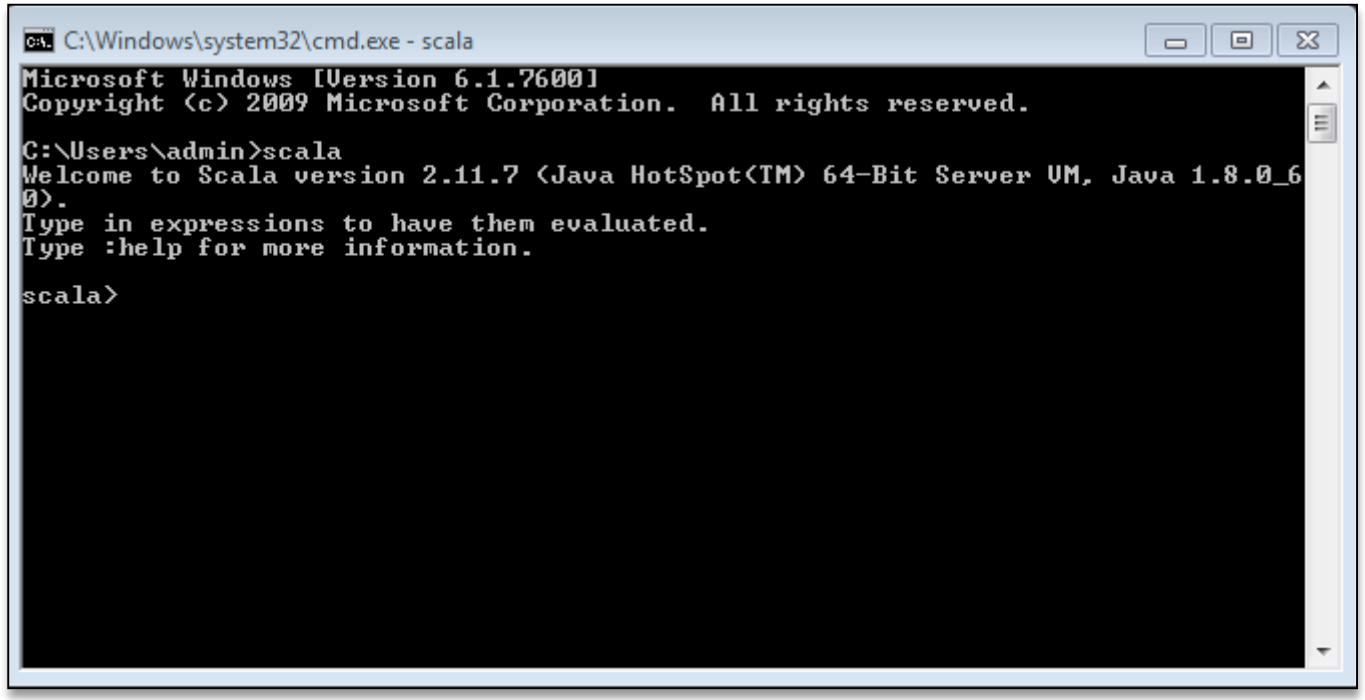
- Latest version can be downloaded from: http://www.scala-lang.org/download/
- Install the Scala and Set the Scala Path in Machine




**Note: Extensive installation Guide is available in LMS**

# Scala Hands-on (Cont'd)

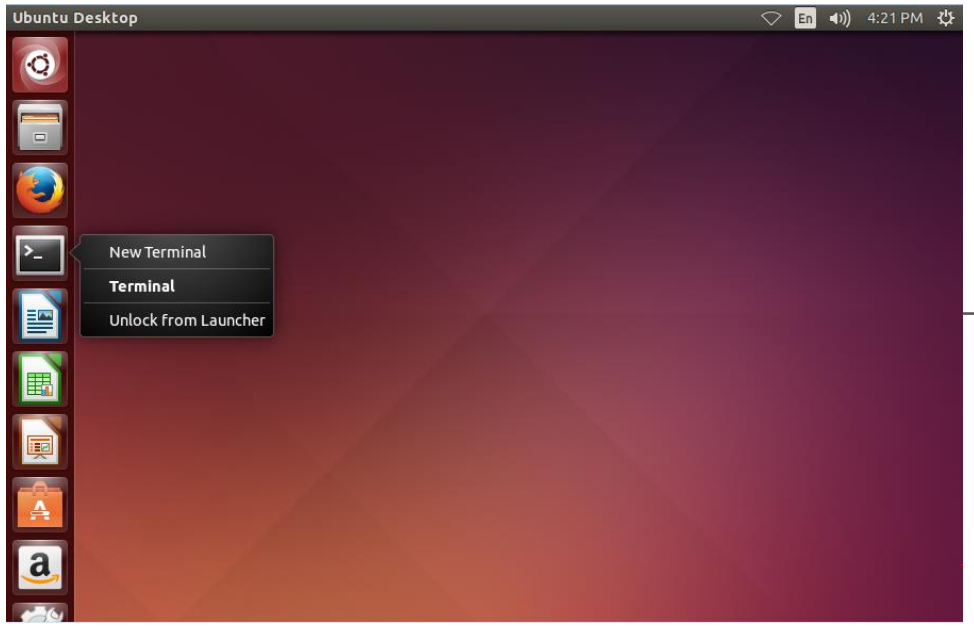
- Start → Type Run → Type cmd
- Type scala

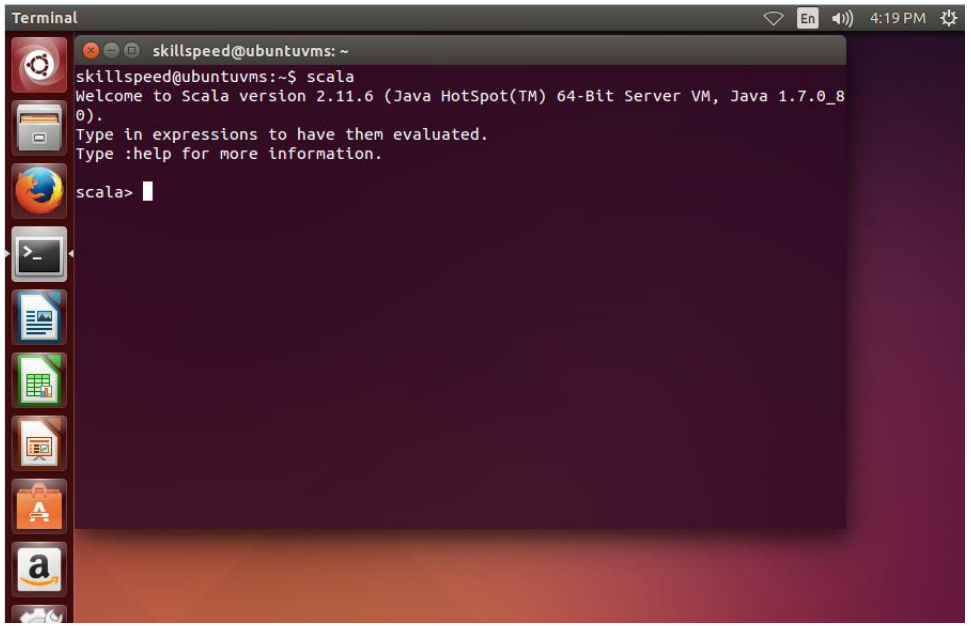



© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Hands-on (Cont'd)

- Download Vmware Player
- Double-click on Vmware Player Workstation → Open the Virtual Machine (it will open the Ubuntu desktop)
- Install scala
- Then select New Terminal → It will open the Terminal and install Scala then type scala




Note: Installation Guide for Linux is Available in LMS

© 2015 BlueCamphor Technologies (P) Ltd.

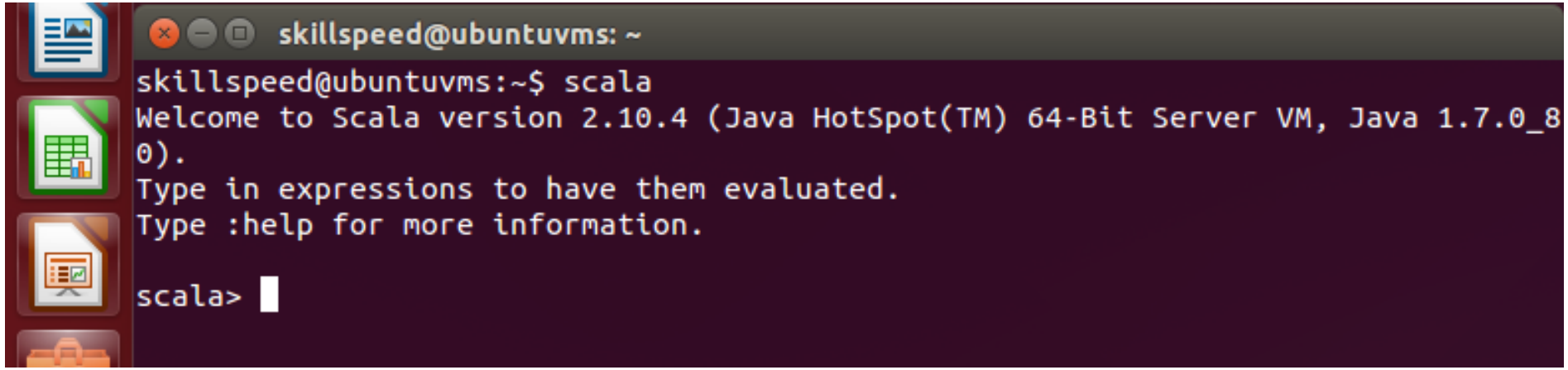
# Scala Hands-on (Cont'd)

- Scala IDE provides excellent and enhanced editing and debugging support for the development of pure Scala (mixed Scala-Java also) applications
- The best choices for Scala IDEs are IntelliJ IDEA and Eclipse because they are excellent in terms of stability and features like type interference, code inspection and memory consumption

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Hands-on (Cont'd)

- REPL: Read - Evaluate - Print - Loop
- Easiest way to get started with Scala, acts as an interactive shell interpreter
- Even though it appears as interpreter, all typed code is converted to Bytecode and executed
- Invoked by typing Scala as shown below




© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Hands-on (Cont'd)

After you type an expression, such as 10 + 30, and hit enter:

```
scala> 10 + 30
```

The interpreter will print:

```
res0: Int = 40
```

```
scala> 10+30
res0: Int = 40
```

This line includes:

- An automatically generated or user-defined name to refer to the computed value (res0, which means result 0),
- A colon (:), followed by the type of the expression (Int),
- An equals sign (=),
- The value resulting from evaluating the expression (30)

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Hands-on (Cont'd)

In the beginning, you started with the REPL

```
scala> println("Hello, Welcome to Skillspeed")
Hello, Welcome to Skillspeed
```

- Scala scripts can be written in text files and saves the script with a .scala extension
- It indicates to the operating system and programmer that the file is actually a scala program

Hello.scala

```
println("Hello,welcome to skillspeed")
```

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Hands-on (Cont'd)

The scripts can be read into the interpreter in several ways:

scala Hello.scala # here Hello is Script name

```
skillspeed@ubuntuvms:~$ scala Hello.scala
Hello,welcome to skillspeed
skillspeed@ubuntuvms:~$
```

The script is executed and the REPL is immediately closed

- scala –i Hello.scala (Output prints and opens the scala REPL )

```
skillspeed@ubuntuvms:~$ scala -i Hello.scala
Loading Hello.scala...
Hello,welcome to skillspeed

Welcome to Scala version 2.11.6 (Java HotSpot(TM) 64-Bit Server VM, Java 1.7.0_80).
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

- scala –nc Hello.scala

```
skillspeed@ubuntuvms:~$ scala -nc Hello.scala
Hello,welcome to skillspeed
skillspeed@ubuntuvms:~$
```

© 2015 BlueCamphor Technologies (P) Ltd.

# Scala Community

Developers in countries all over the world are using Scala for a large variety of applications across a broad range of industries

Popular ways to connect with the Scala community are via mailing lists or IRC channels

Though there are plenty of opportunities to connect face-to-face with others in the community– for example, via local Scala Meetups, or local Scala user groups

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – 3

Scala REPL acts as scala Interpreter

a) True

b) False

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – Solution

Scala REPL acts as scala Interpreter

a) True

b) False ✔

False

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – 4

Scala supports primitive and wrapper classes ?

a) True

b) False

© 2015 BlueCamphor Technologies (P) Ltd.

# Check your Understanding – Solution

Scala supports primitive and wrapper classes ?

a) True

b) False ✓

False

© 2015 BlueCamphor Technologies (P) Ltd.

# Questions

© 2015 BlueCamphor Technologies (P) Ltd.

skillspeed

for the serious learner

# thank you!