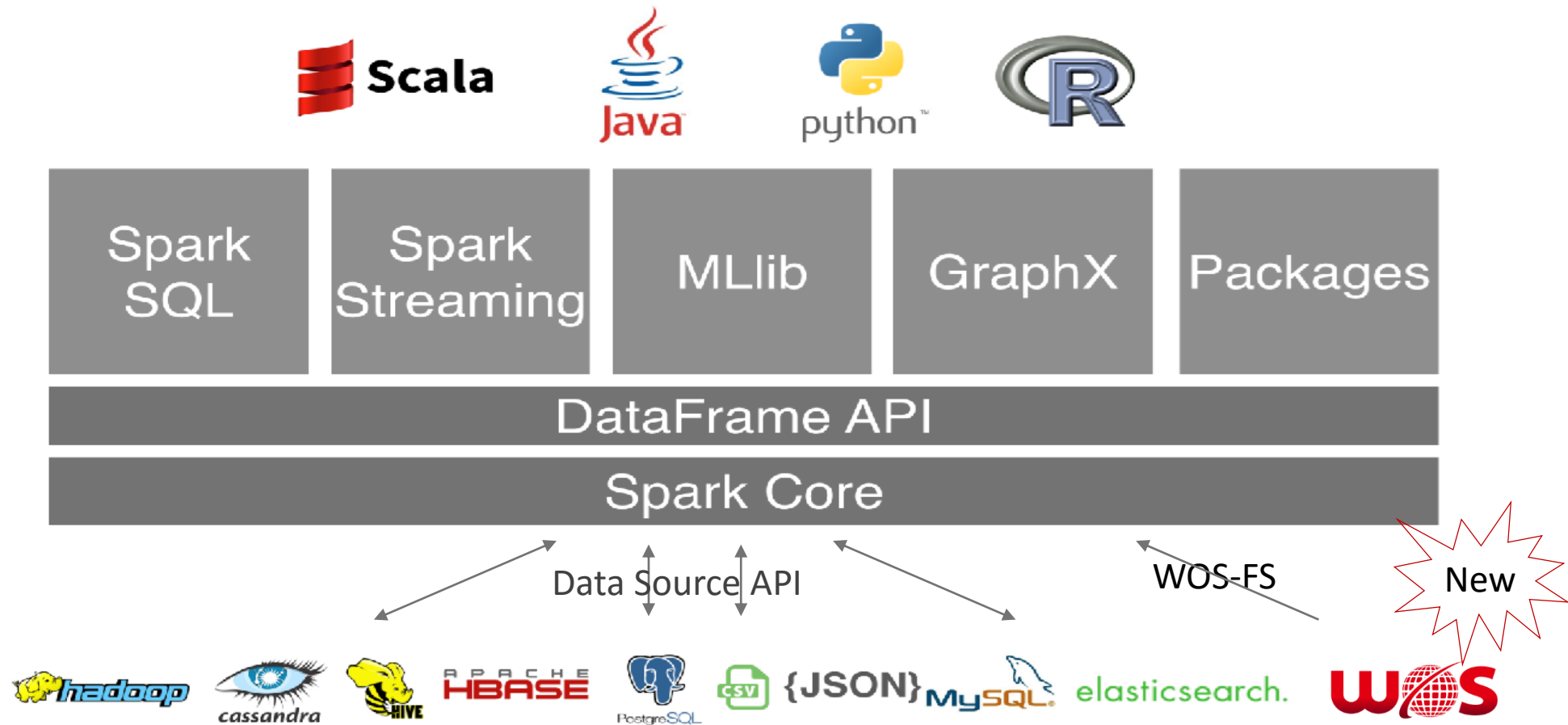


# WOS-FS for **Spark**<sup>TM</sup>

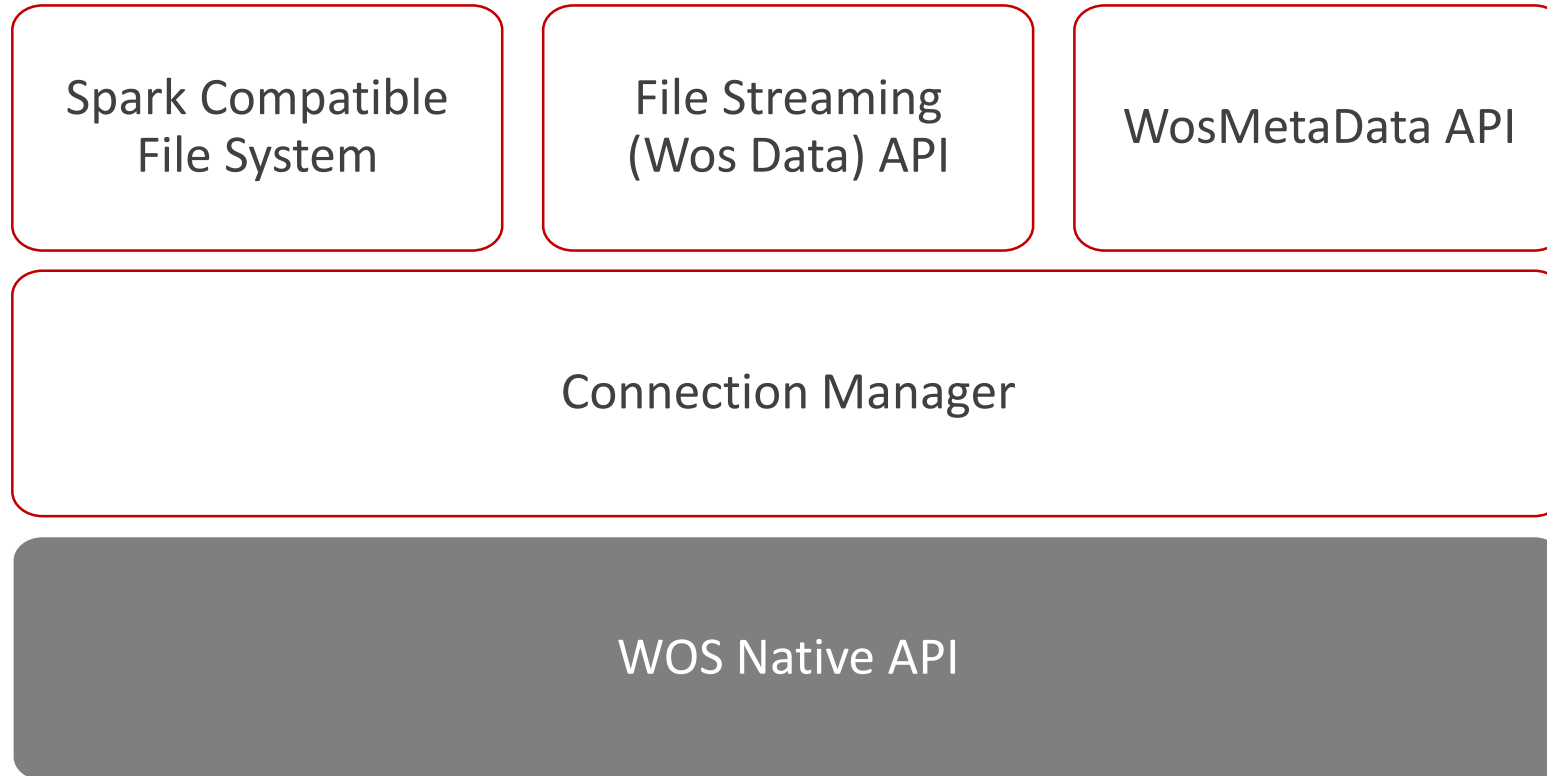
Hadoop File System Compatible

Developed by Ying LI  
([Mr.Ying.LEE@Gmail.com](mailto:Mr.Ying.LEE@Gmail.com))

# WOS-FS brings WOS to Spark cosystem



# WOS-FS for Spark architecture



# WOS-FS works like a read only HDFS

Sample: the famous WordCount using WOS as data source

```
initWos(sparkContext, String IP_of_WOS_Cluster);
```

```
.....
```

**\wos/**

```
JavaRDD<String> lines = sparkContext.textFile(("hdfs://...") );
```

```
JavaRDD<String> words =
```

```
    lines.flatMap(line -> Arrays.asList(line.split(" ")));
```

```
JavaPairRDD<String, Integer> counts =
```

```
    words.mapToPair(w -> new Tuple2<String, Integer>(w, 1))
```

```
    .reduceByKey((x, y) -> x + y);
```

# WOS-FS accelerate ETL performance

Filter WOS Objects by MetaData:

```
List<String> paths = Arrays.asList(wosFile1, ...,wosFileN);
```

```
String filteredWosFiles = paths.stream()  
    .filter(path -> (new String(WosConnection.getMeta(path, "postfix")).equals("txt")))  
    .collect(Collectors.joining(", "));
```

Thank you!

Developed by Ying LI  
(Mr.Ying.LEE@Gmail.com)