Apache Spark
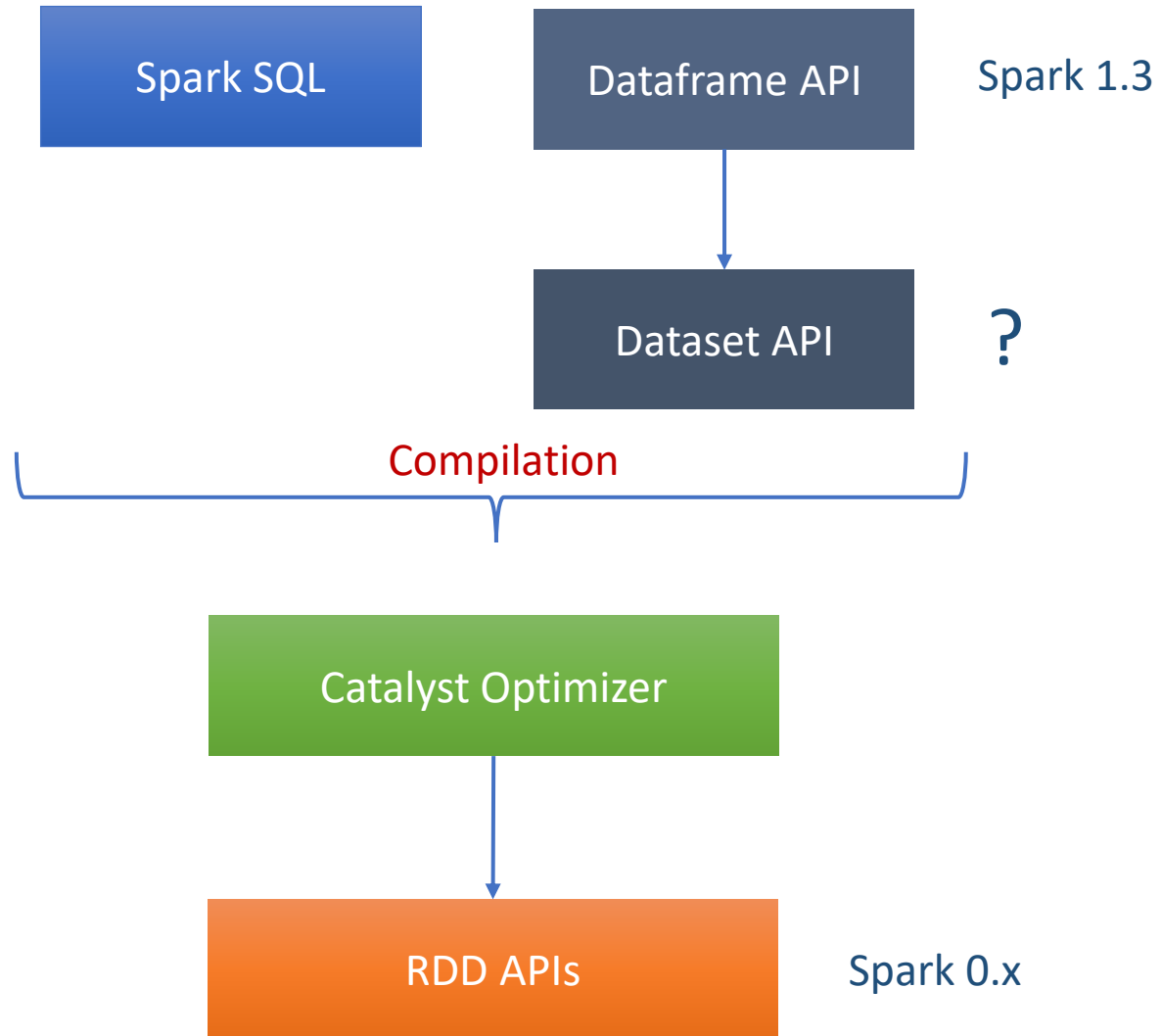
Spark Dataset Vs Dataframe

Spark APIs

Spark SQL

Dataframe API   Spark 1.3

Dataset API   ?

Compilation

Catalyst Optimizer

RDD APIs   Spark 0.x

LEARNING JOURNAL
www.learningjournal.guru

Spark APIs

Spark SQL   Dataframe API   Spark 1.3

1. Generic Typed APIs
2. Expression based transformation
   select() where() join() orderBy()

Compilation

Tungsten/Catalyst Optimizer

rdd()   Without Optimization

RDD APIs   Spark 0.x

1. Strongly Typed APIs
2. Lambda based transformation
   map() filter() reduce() foreach()

LEARNING JOURNAL
www.learningjournal.guru

Spark APIs

Spark Structured API

Spark SQL

Dataset[Row]

Dataset

Spark 2.0

Compilation

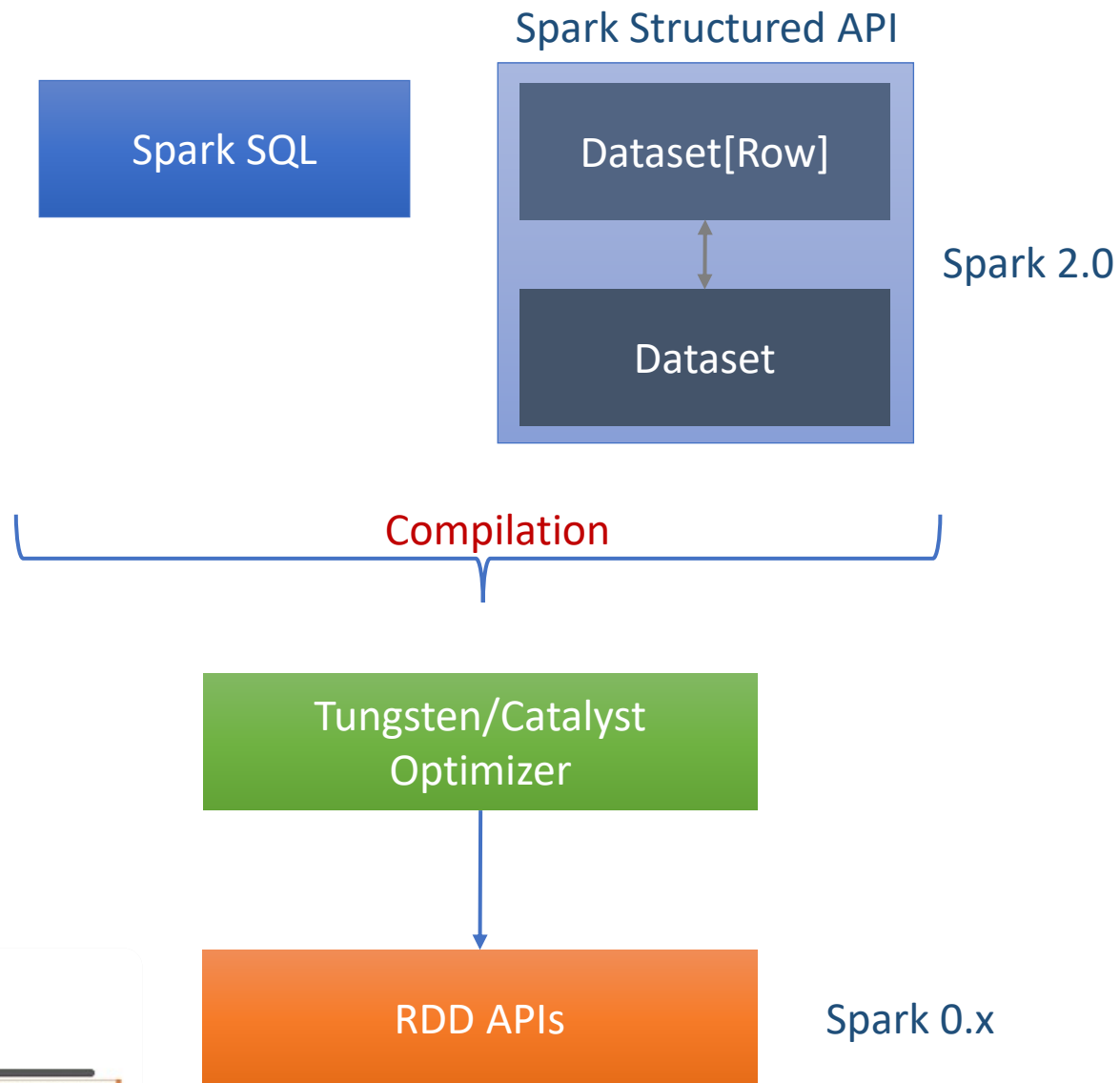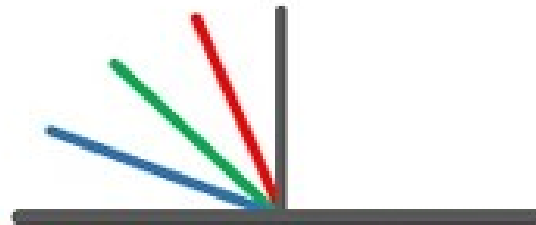Tungsten/Catalyst Optimizer

RDD APIs

Spark 0.x

1. Generic Typed APIs
2. Expression based transformation
   select() where() join() orderBy()

Spark implicit Encoders

1. Strongly Typed APIs
2. Lambda based transformation
   map() filter() reduce() foreach()

LEARNING JOURNAL
www.learningjournal.guru

www.learningjournal.guru