

Spark Streaming

2017.4 XenRon

CONTENTS

Spark Streaming

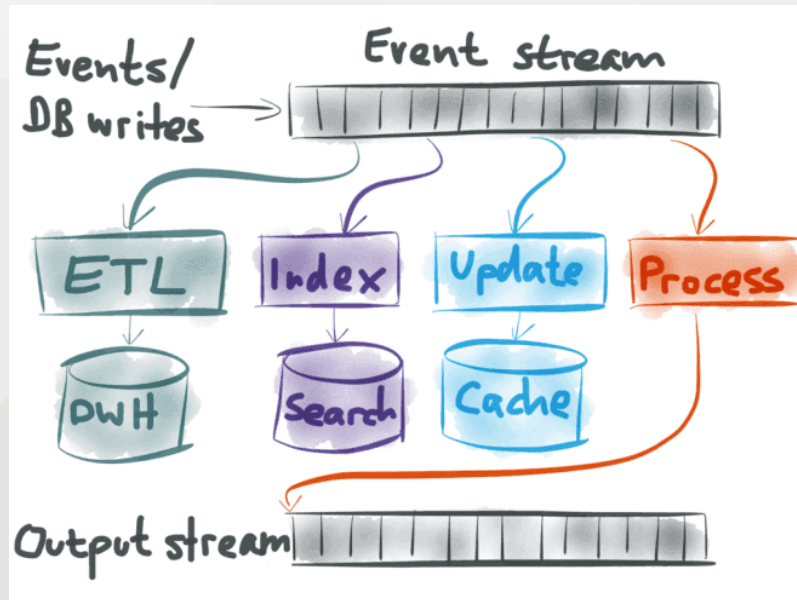
01

Spark ML

02

Spark Graph X

03



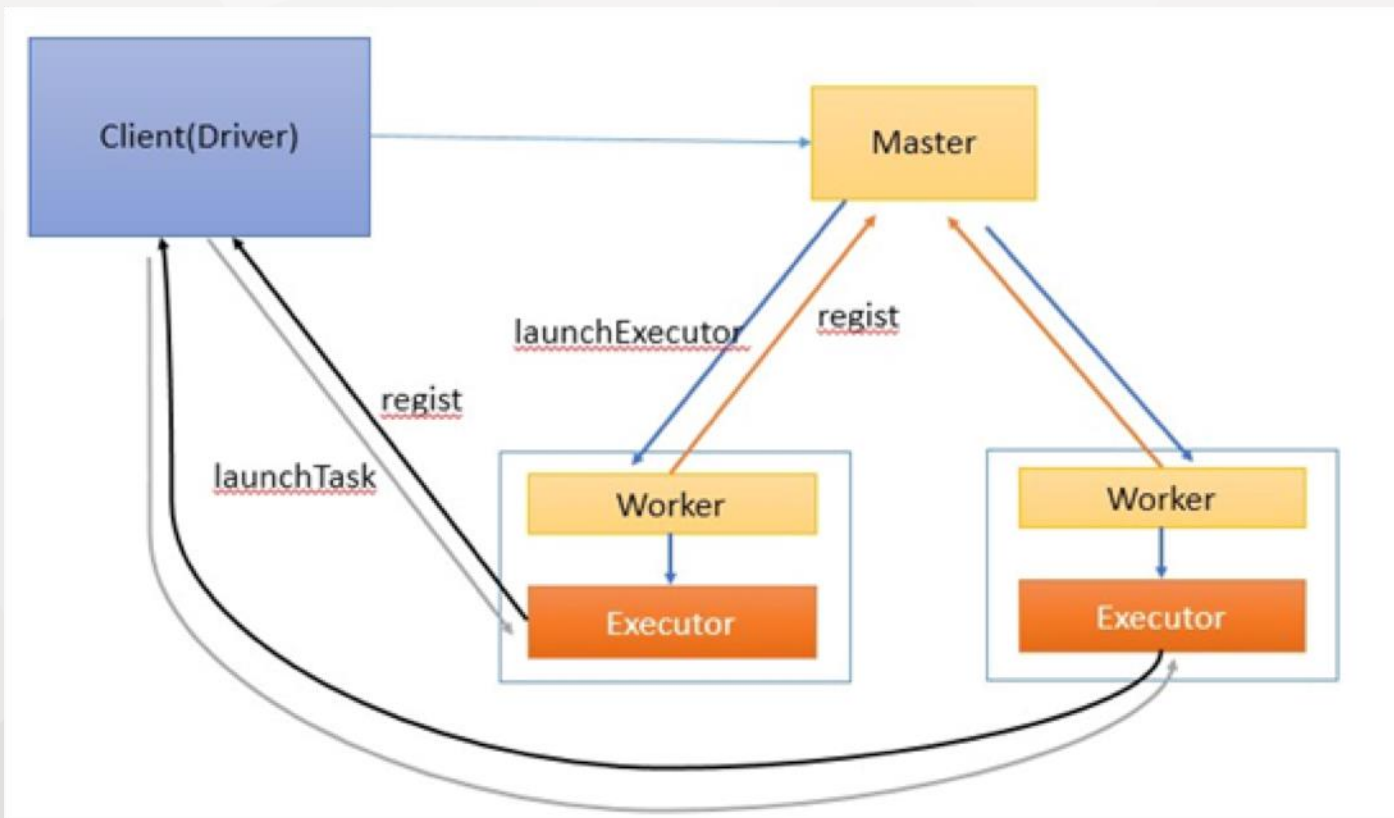


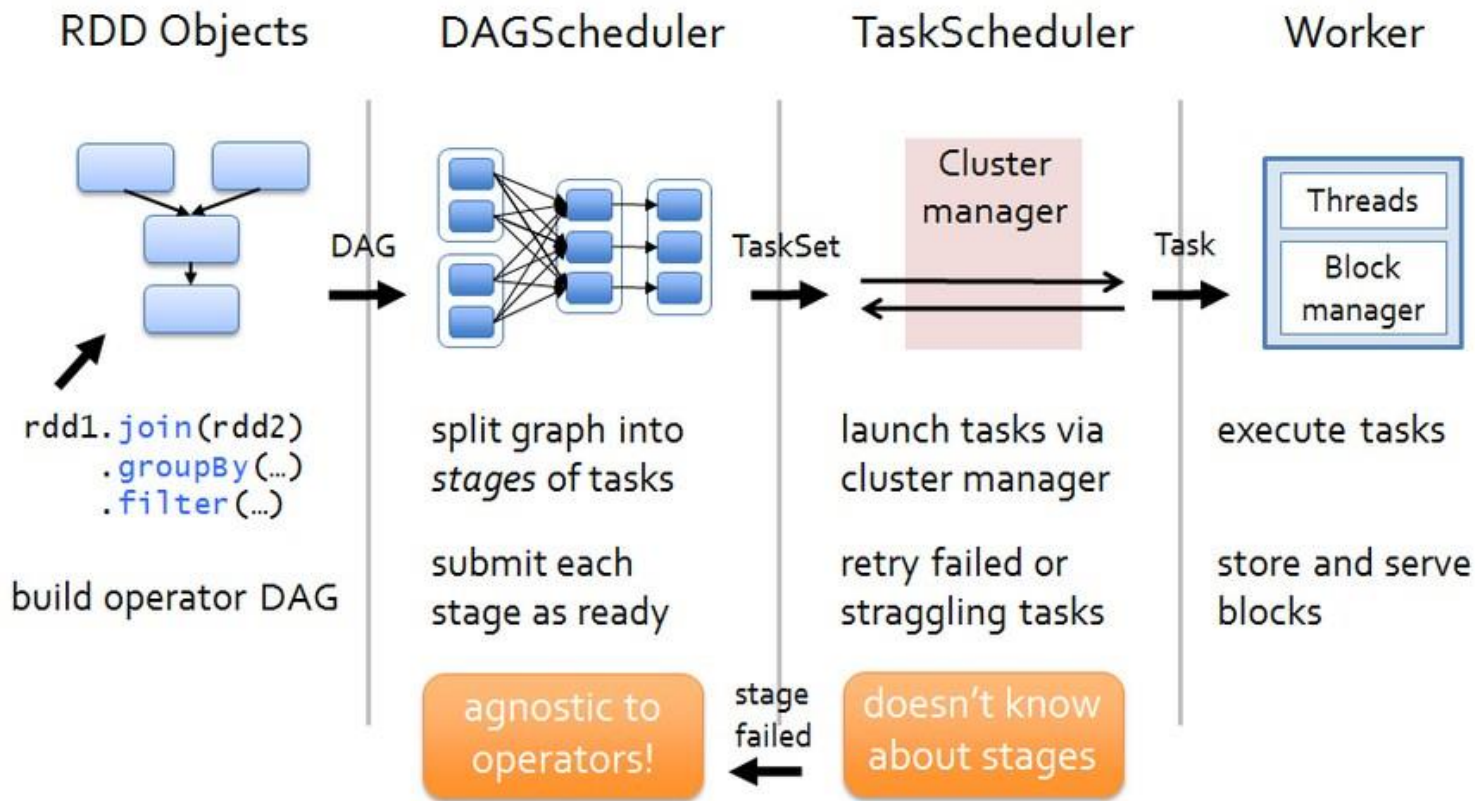
Review



Review

Spark Standalone



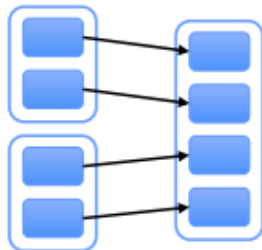


Dependency Types

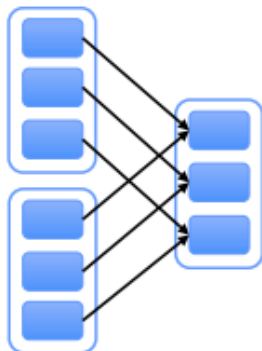
“Narrow” (pipeline-able)



map, filter

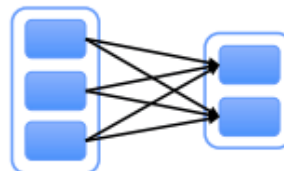


union

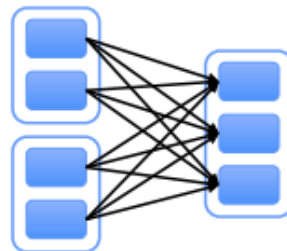


join with inputs
co-partitioned

“Wide” (shuffle)

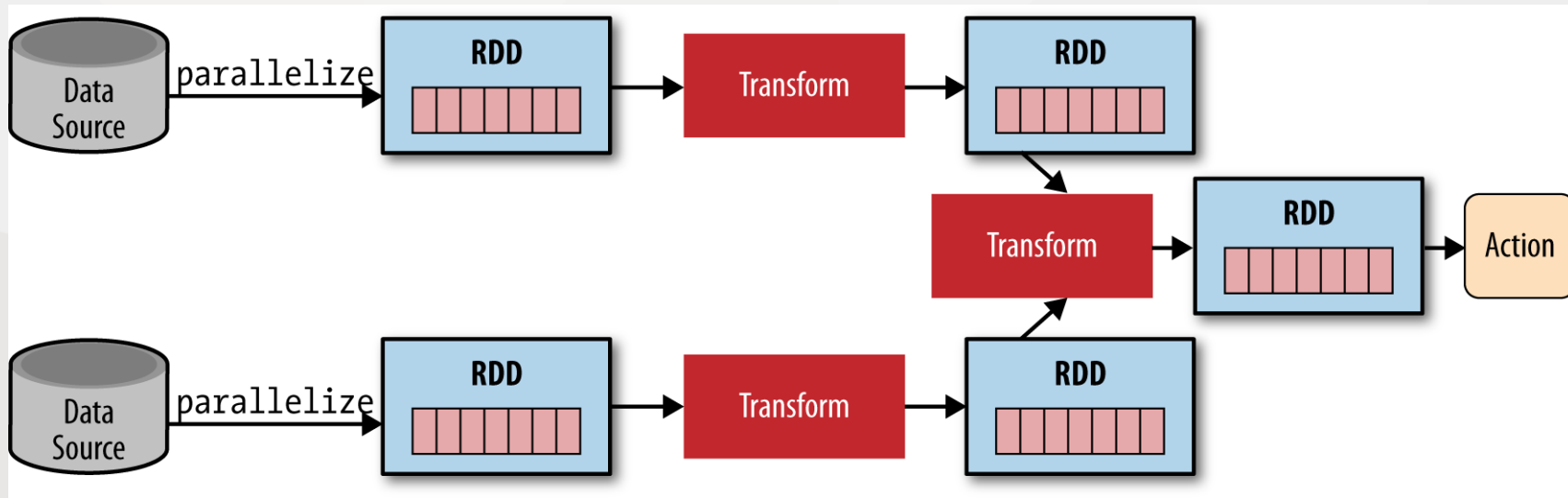


groupByKey on
non-partitioned data

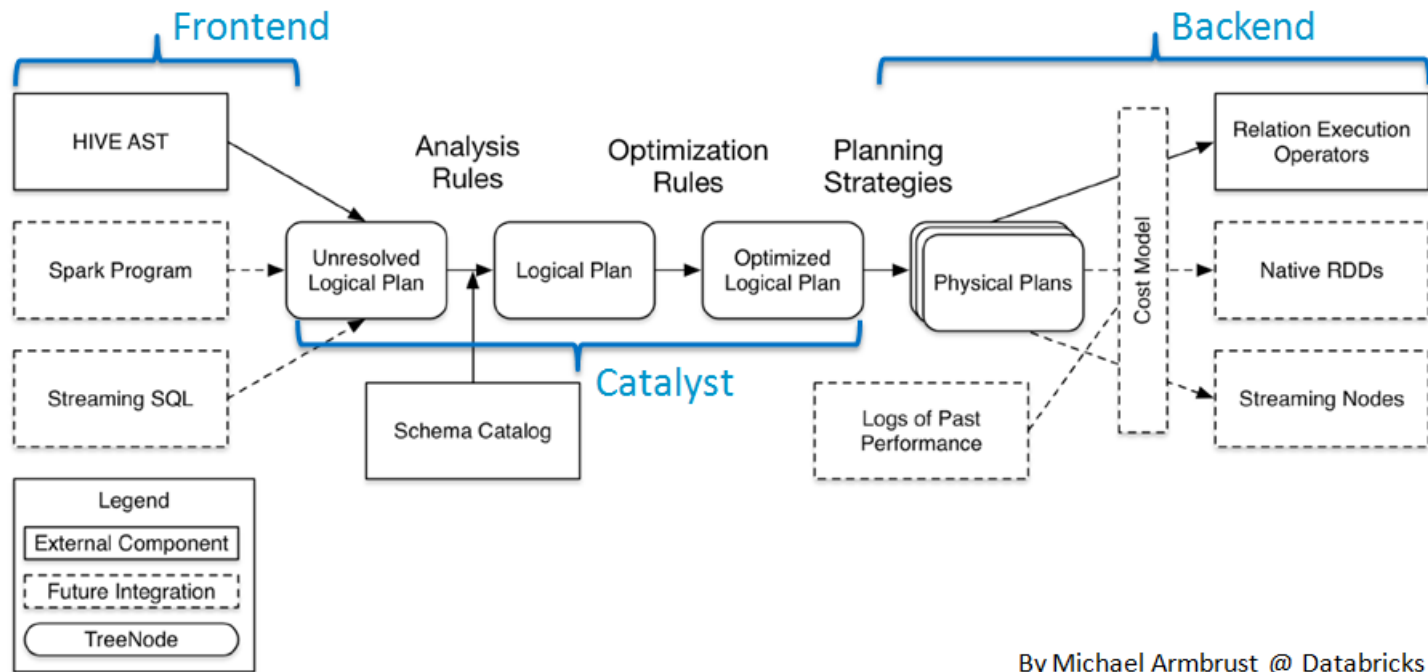


join with inputs not
co-partitioned

Dependency Type



Spark SQL Architecture



By Michael [Armbrust](#) @ [Databricks](#)



PART1

Spark Streaming

Resource Management

Standalone

YARN

Mesos

Spark Ecosystems

Spark SQL

Spark Streaming

BlinkDB

Spark Machine Learning

GraphX

Tachyon

Spark Core

BACK TO BASICS

Spark DataFrame API



Java



Scala



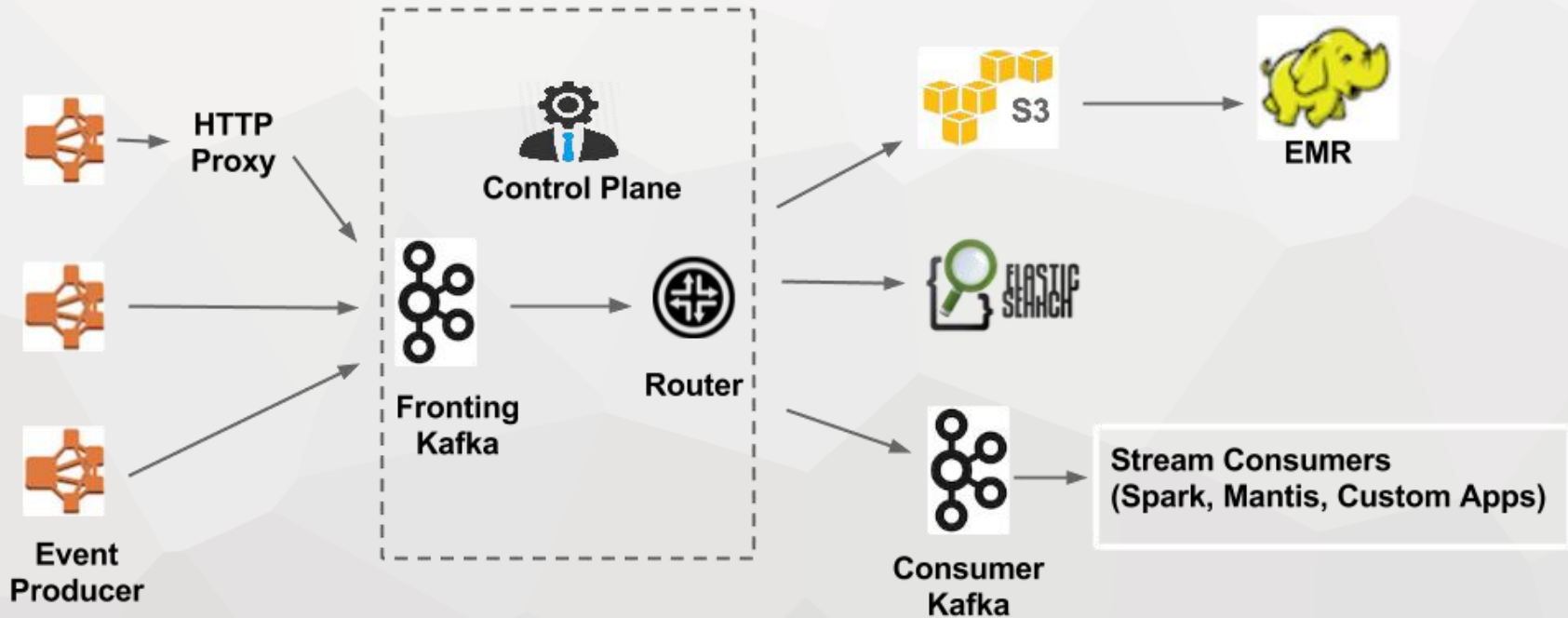
Python



R

Spark Core





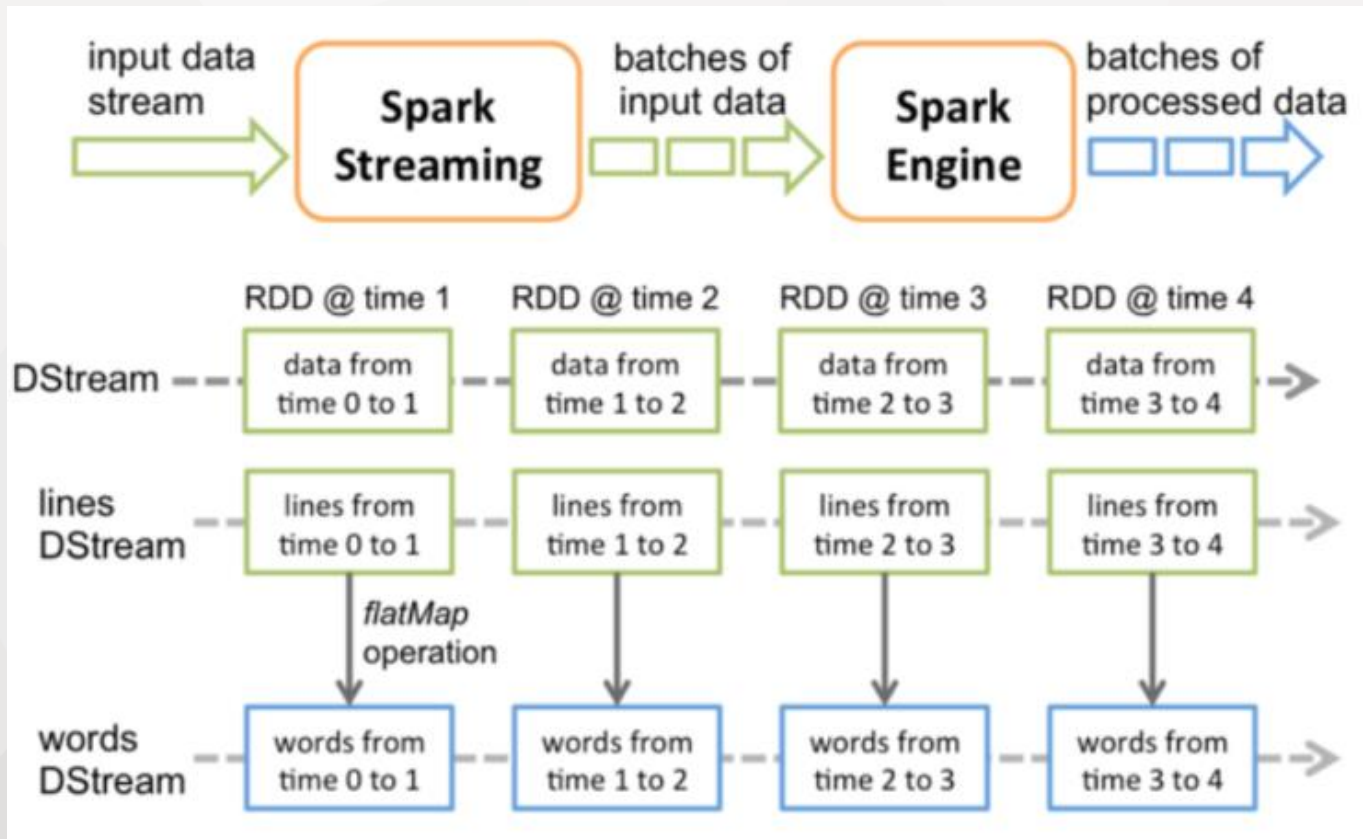
ELK (ElasticSearch LogStash Kibana)

13

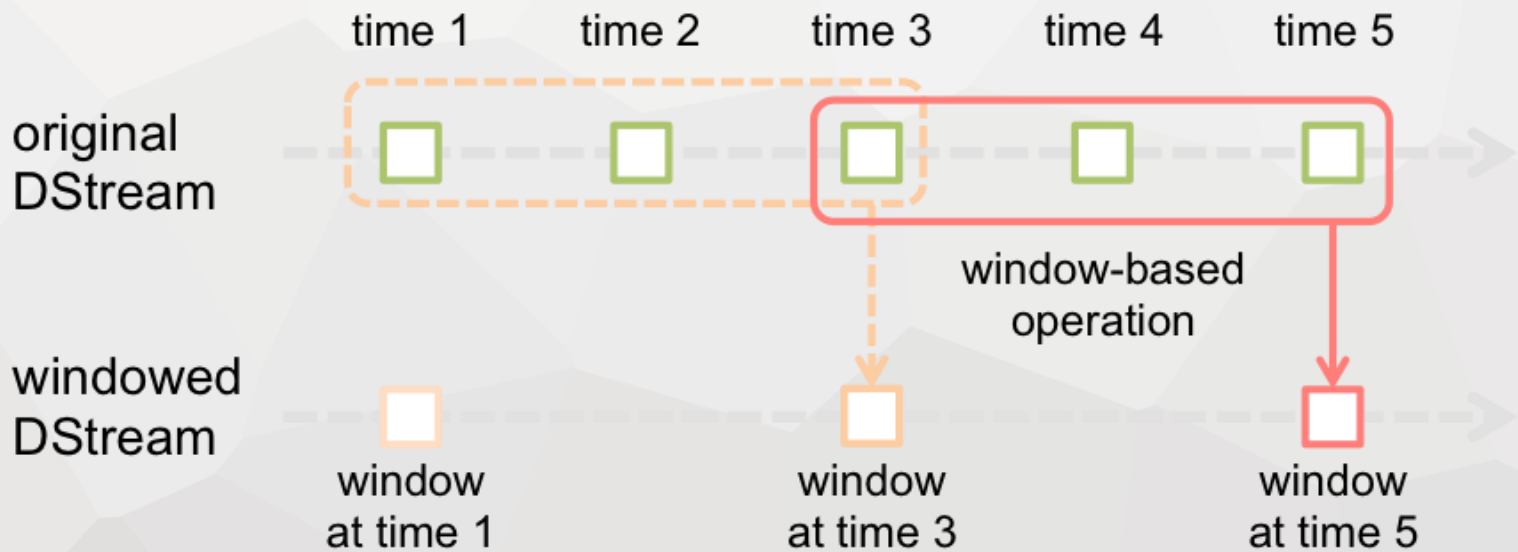


华为		搜索
华为p9	约39个商品	金融 61
华为手机	约431个商品	
华为v8	约1273个商品	
华为p8	约38个商品	
华为p7	约41个商品	
华为5X	约60个商品	
华为G9	约810个商品	
华为荣耀	约187个商品	
华为5c	约778个商品	
华为5s	约19个商品	
关闭		

Spark Streaming



Windows based operations



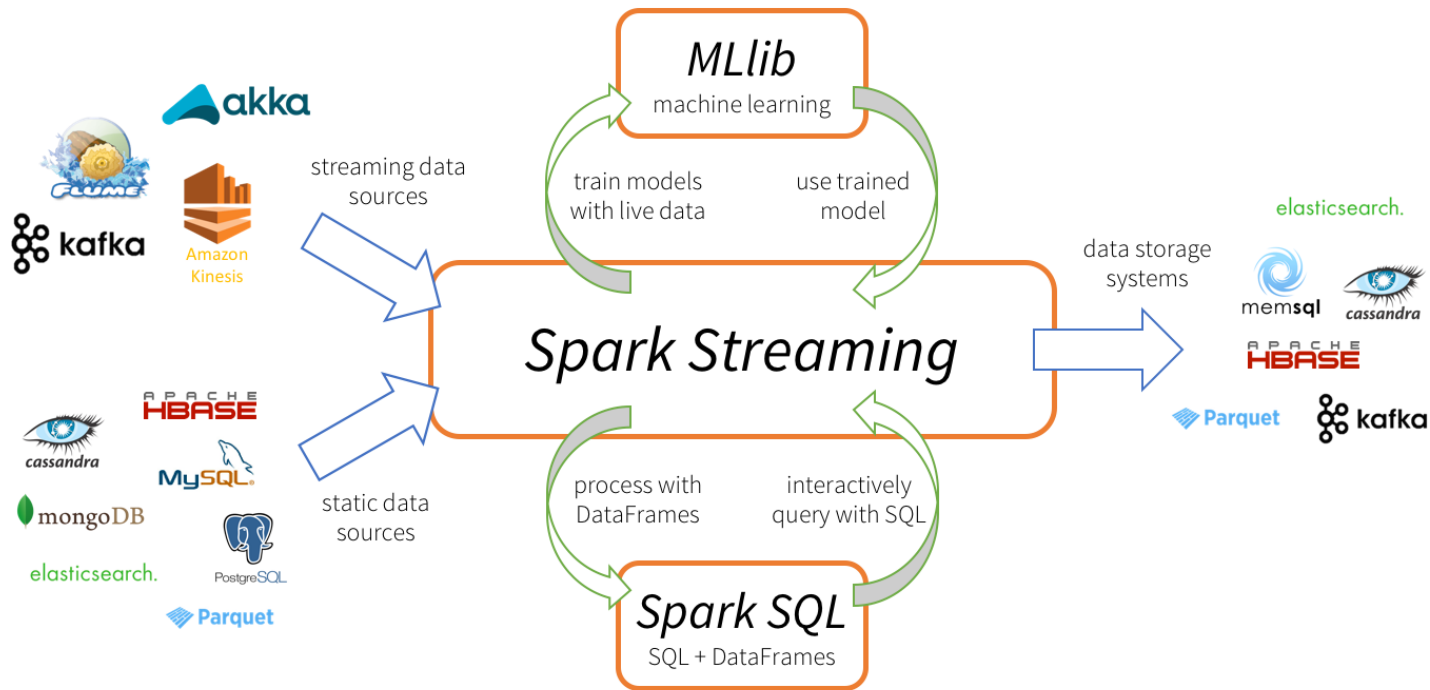
Real-Time Traffic





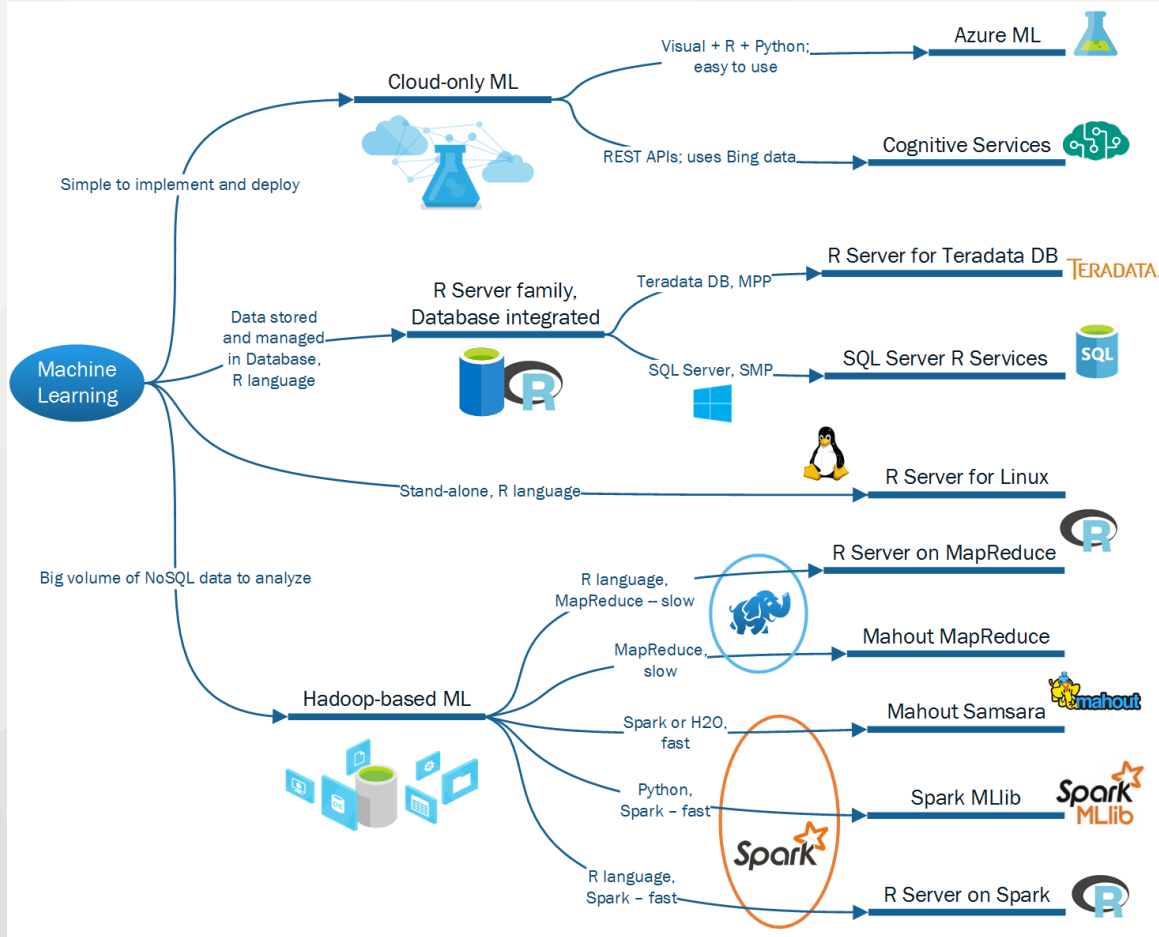
PART2

Spark Machine Learning



Ecosystems

20



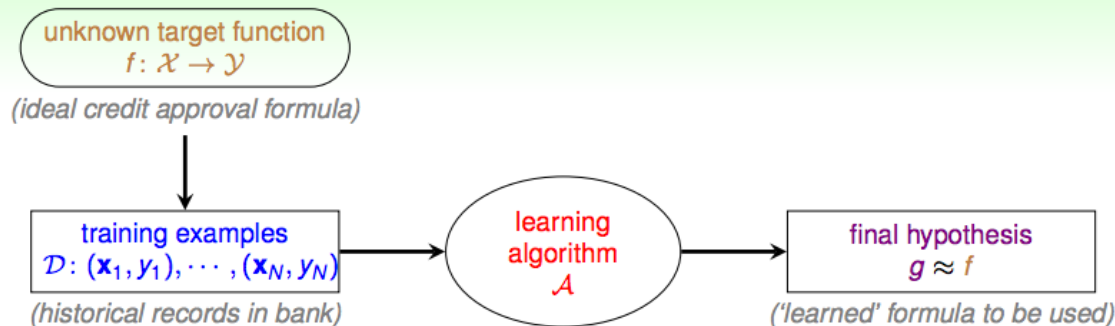
Use Case : Credit Approval

21

The Learning Problem

Components of Machine Learning

Learning Flow for Credit Approval



- target f **unknown**
(i.e. no programmable definition)
- hypothesis g hopefully $\approx f$
but possibly **different** from f
(perfection 'impossible' when f unknown)

What does g look like?

Use Case : WeChat

22



Use Case : Hatsune Miku Upscale

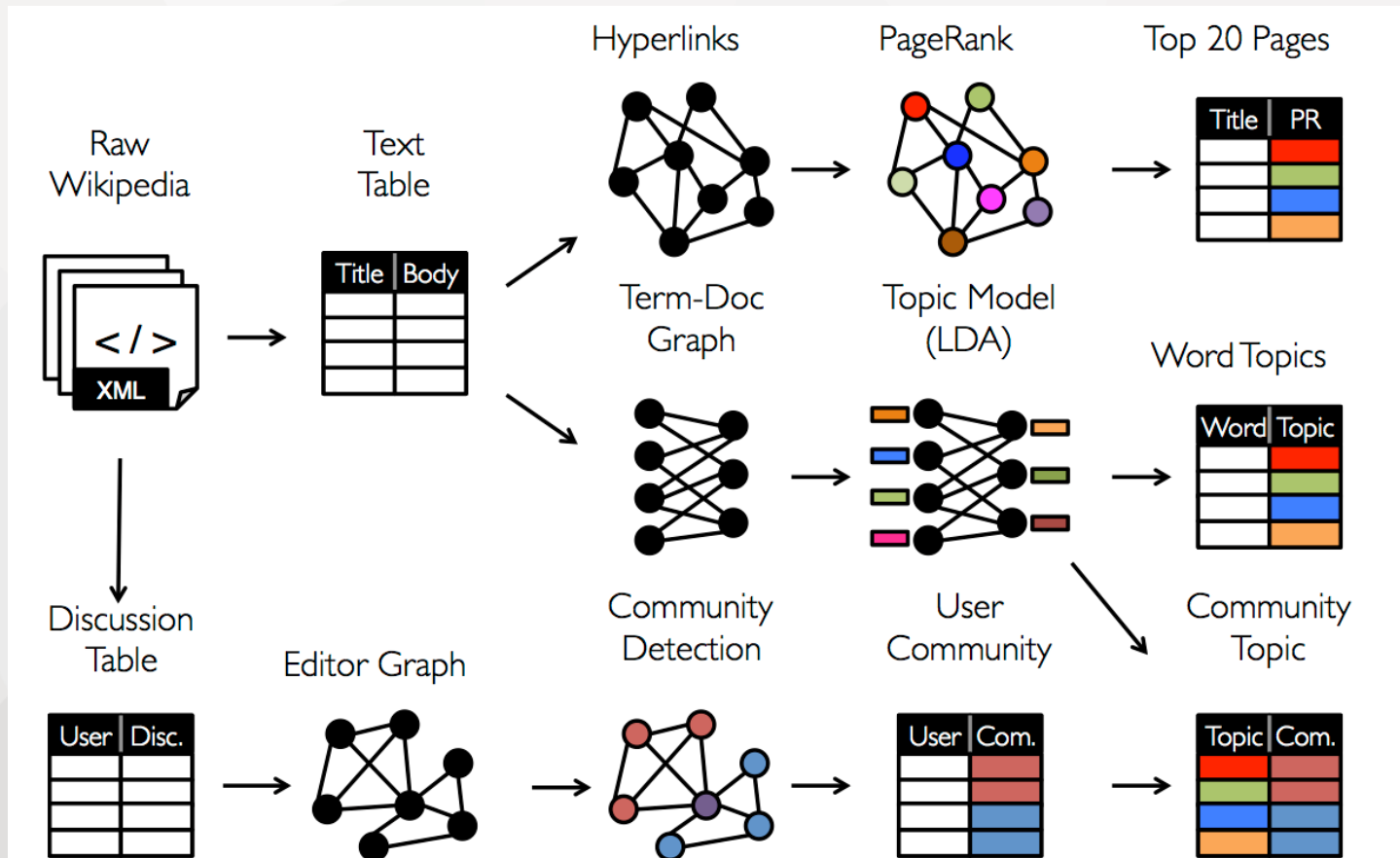


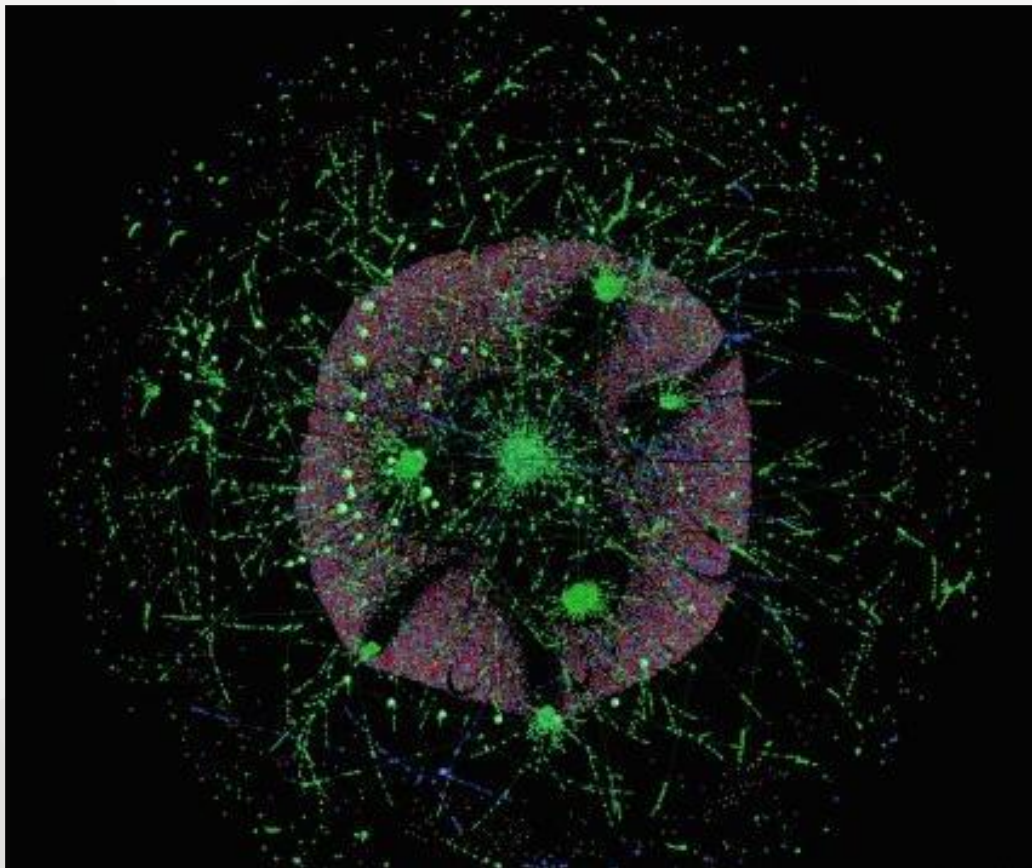


PART3



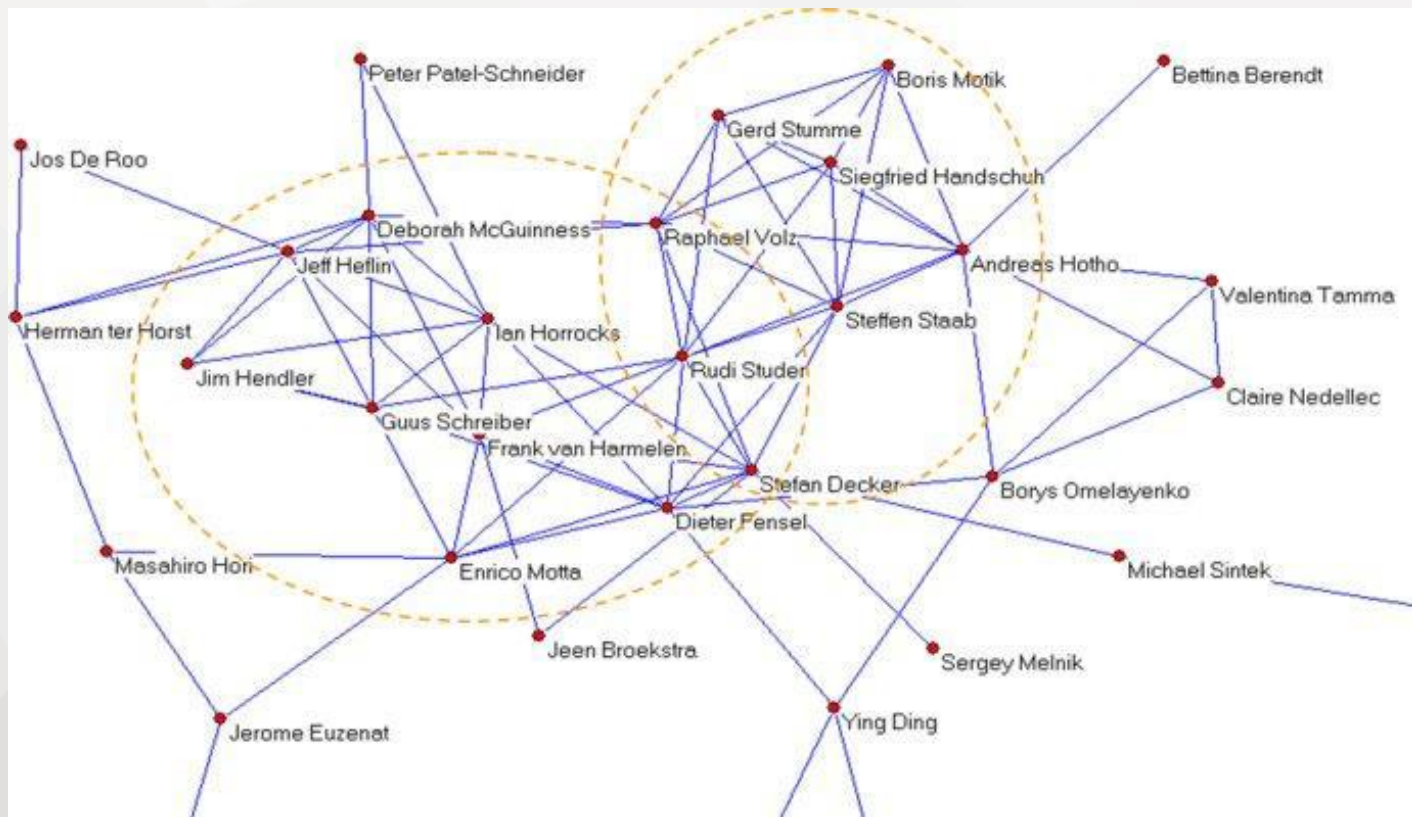
Spark GraphX





Social Network Site community

27





The End