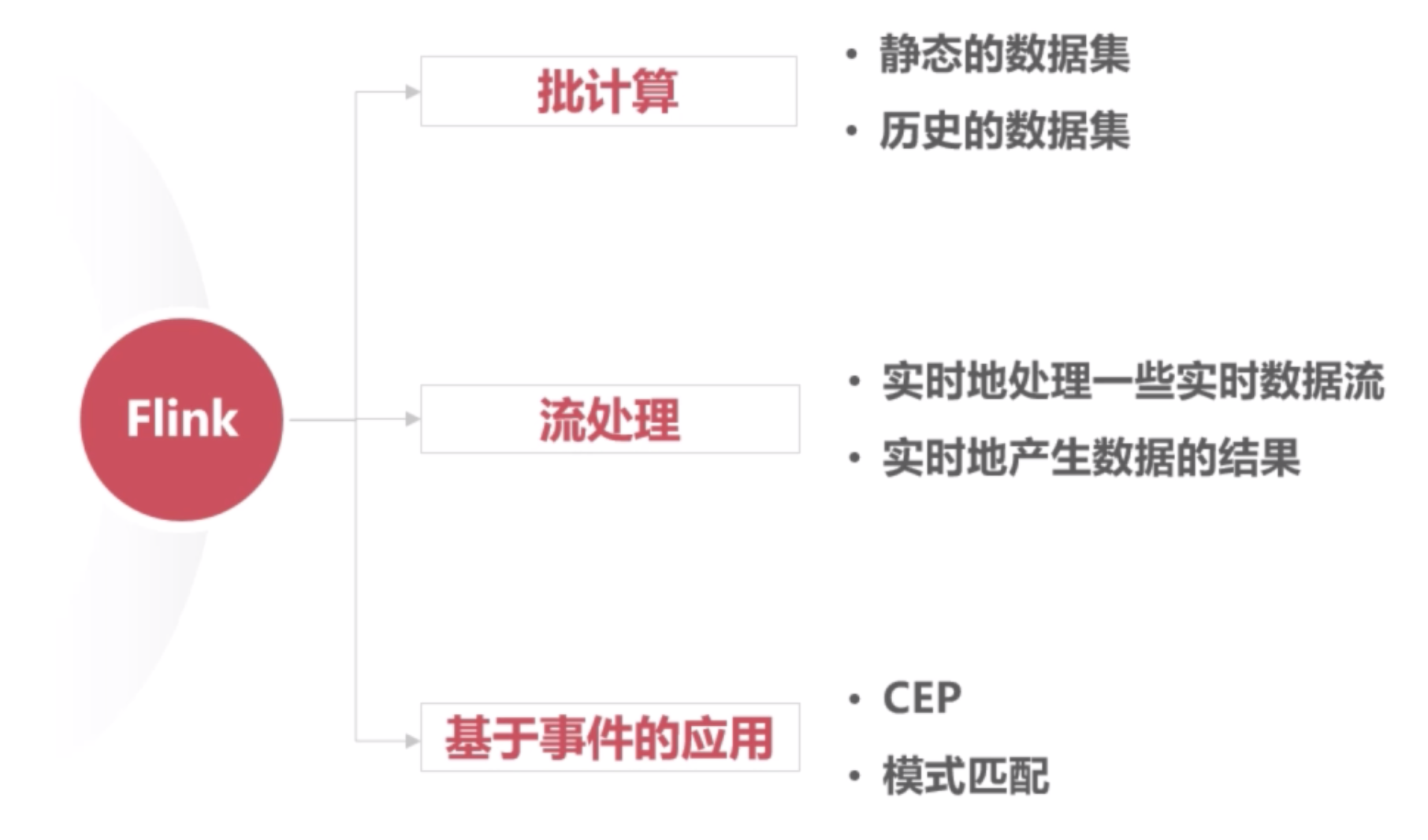
## 初识 Flink

### 课程导学

吹牛逼阶段

这里需要说明下一代的计算引擎-Flink

》》说一下 Flink 的可以用的地方：

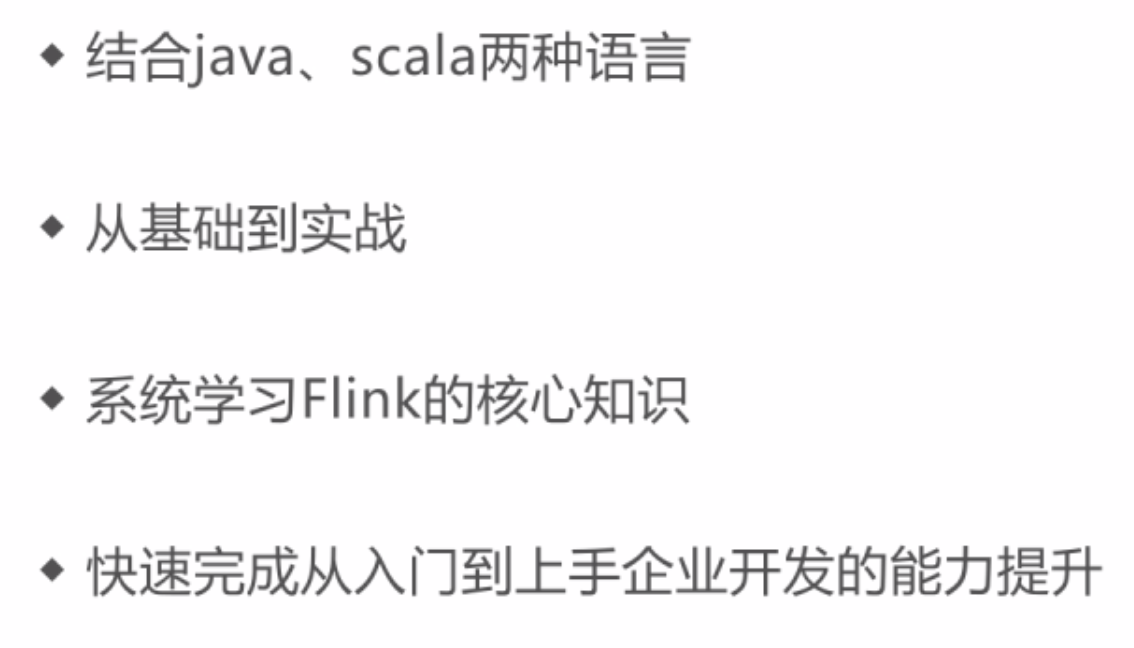


》》目前 Flink 的用户的公司：



并且阿里推出了 Blink 的一个基于 Flink 改造的

》》本课程的特点：



》》课程的安排：

1、初识 Flink

2、编程模型及核心概念

3、DataSetApi 编程：Flink 的批计算

4、DataStream API：流式操作的 API

5、Flink Table&SQL：类似于 SparkSQL 的一种 sql 操作

6、Window 和 Time 操作：

7、FlinkConnectors：跟外部应用的的读写，这里介绍 HDFS，以及 Kafka的读写操作

8、Flink 部署及作业提交：单机模式，standlone 模式等

9、Flink 监控及调优：监控一些参数，或者调优参数提高性能

学和练结合，多动手，多思考，多交流

最后还需要实战：**互联网直播平台的日志分析**，主要是 scala 语言进行开发



这里涉及到的 4 大核心点：

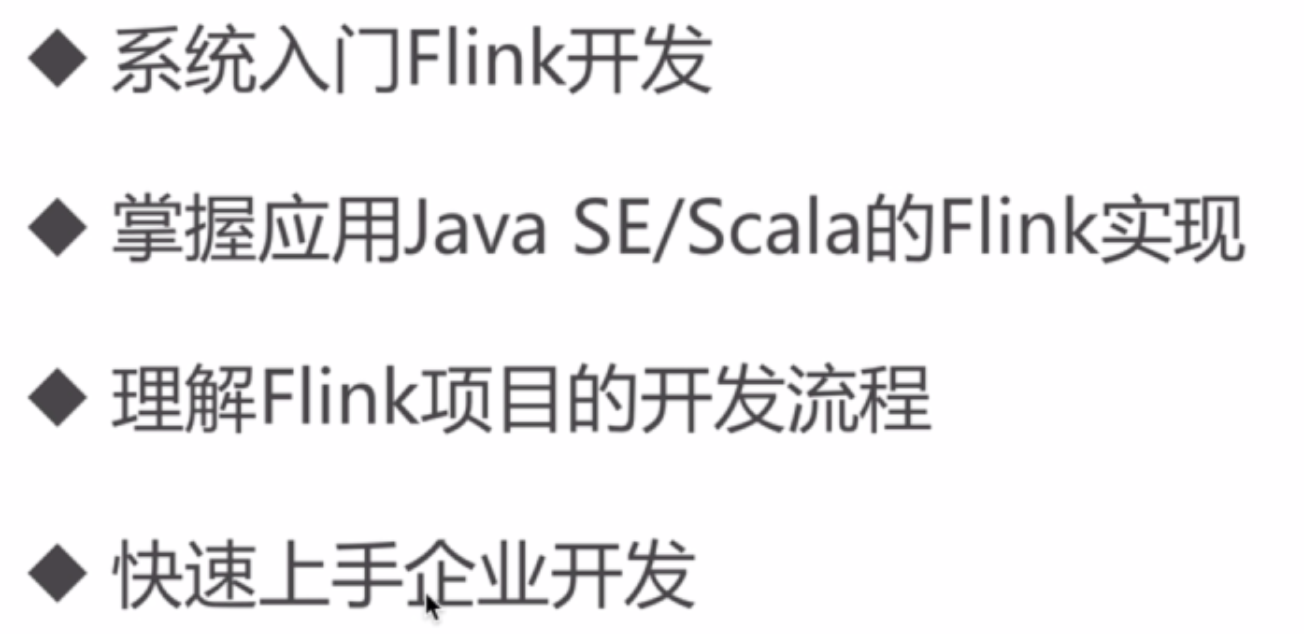
1、使用 Flink 自定数据源读取配置数据：项目里很多自定义的配置数据，需要进行读取

2、使用 Flink 完成实时数据清洗：进行数据符合我们规则的数据

3、使用 Flink 完成实时结果统计：这里进行 API 的逻辑统计

4、统计结果可视化展示（Kibana）

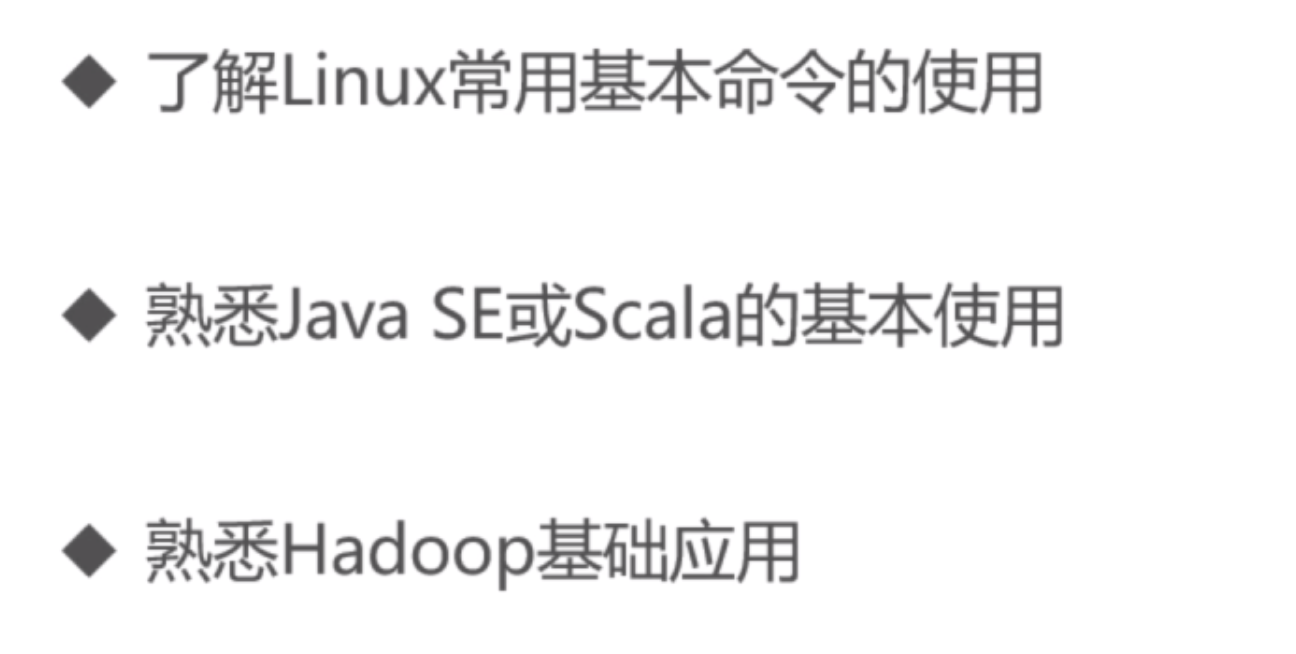
》》学习收获：



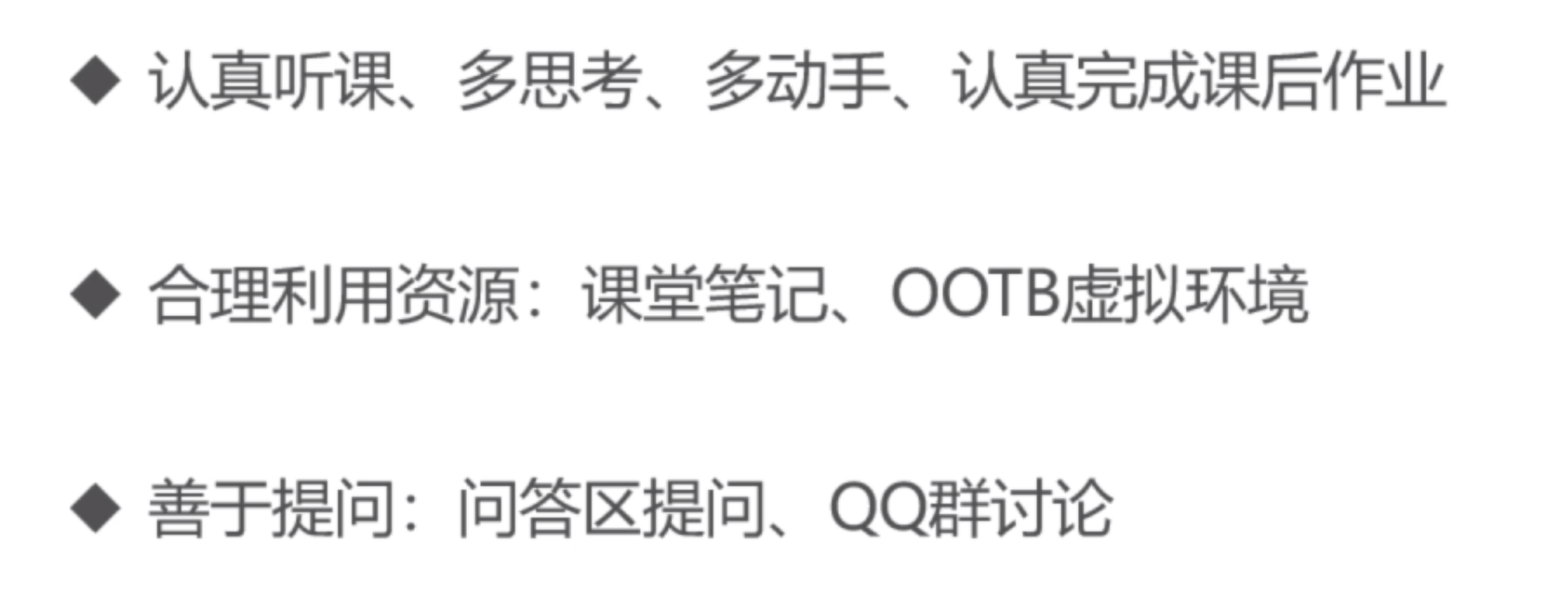
》》环境参数



》》前置知识：



》》学习方法：



### 学前必读

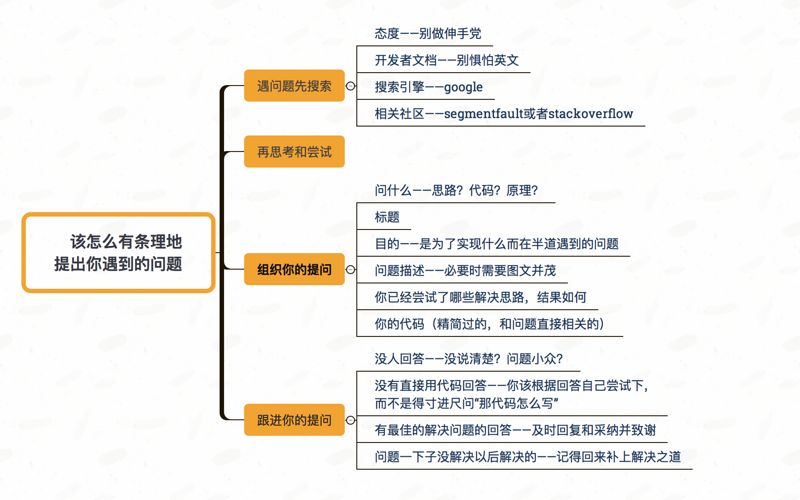
1、课程源码，教辅材料怎么下载

2、PK 老师在慕课网上制作了那么多门课程，我该如何学习更加顺畅了？

Hadoop -> scala -> flink -> spark sql -> spark streaming -> pyspark -> storm

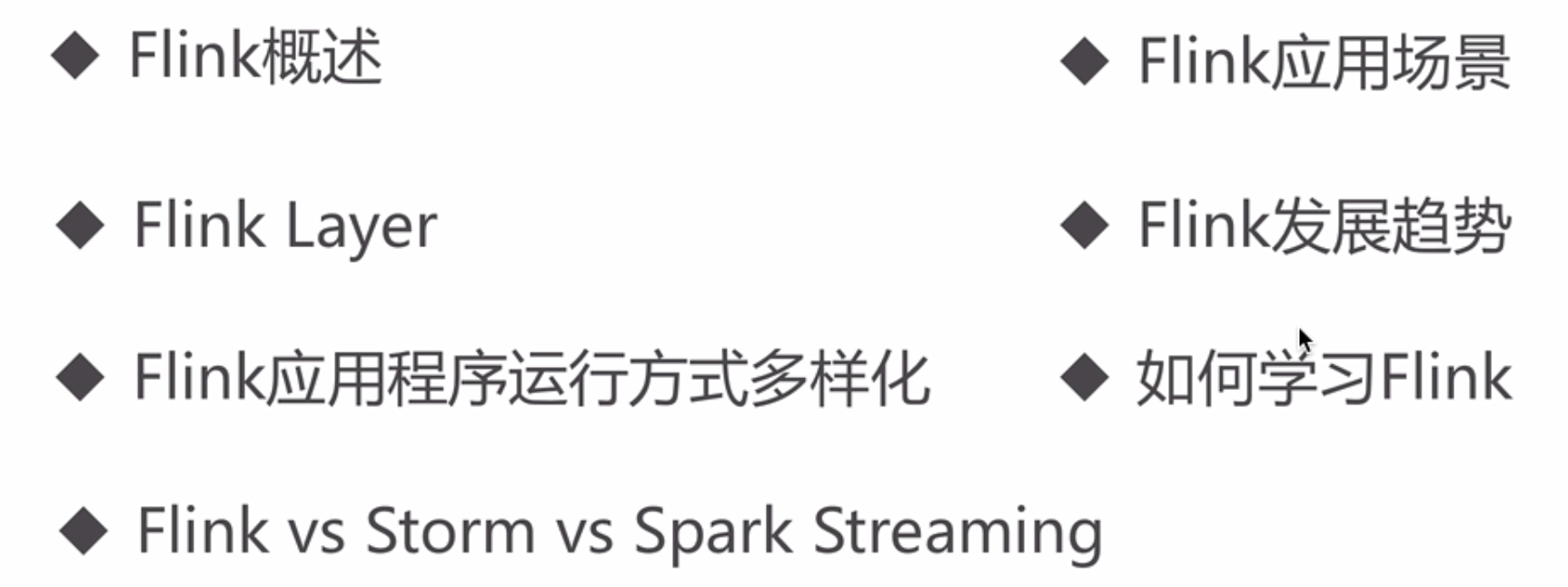
3、课程 qq 交流群怎么加入

4、学习遇到问题后怎么，怎么办？



### 课程目录

》》初识 Flink本章节的内容



### Flink 概述

》》Flink 是什么

打开挂网的链接：http://flink.apache.org

Stateful Computation over Data Streams

基于数据流的一个有状态的计算，简单来说就是这样

点开 What is Apache Flink？（<https://flink.apache.org/flink-architecture.html>）

Flink 是一个框架，并且是一个分布式的计算引擎，这个框架用来有状态的计算，这个计算是作用在有界和无界的数据上。

Apache Flink is a framework and distributed processing engine for stateful computations over ***unbounded and bounded* data streams**. Flink has been designed to run in *all common cluster environments*, perform computations at ***in-memory speed*** and at ***any scale*.**

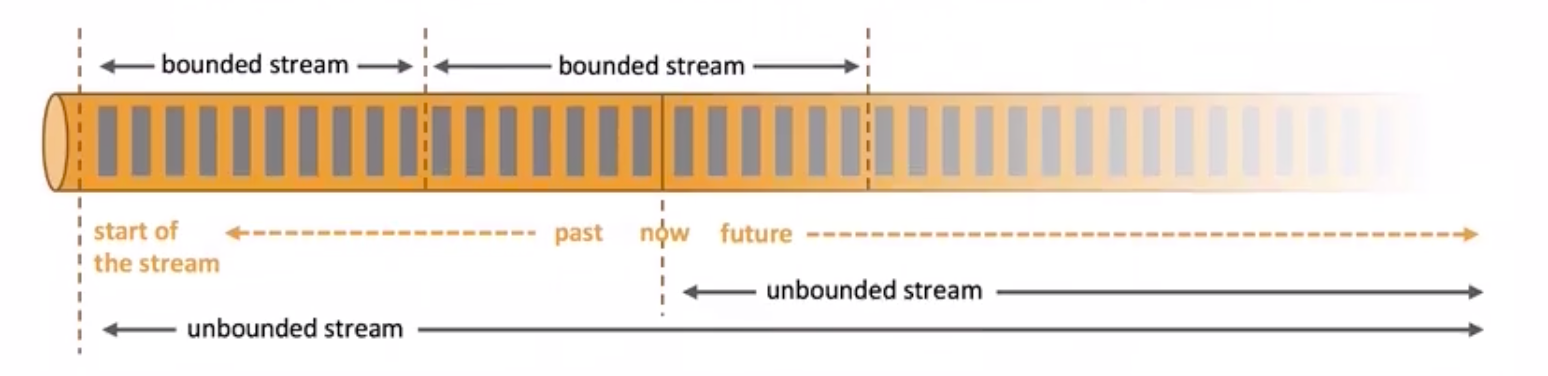
》》Unbounded VS Bounded Data

下面都是具体的解释了：

无界和有界的数据，如信用卡交易数据、传感数据、机器日志，这些数据都是流式的形式产生出来的。

Unbounded streams：有开始的但是没有结束的，像水流一样的，源源不断的产生，并且是不间断的处理的

Bounded streams: 处理有界的数据就是批处理(batch processing)



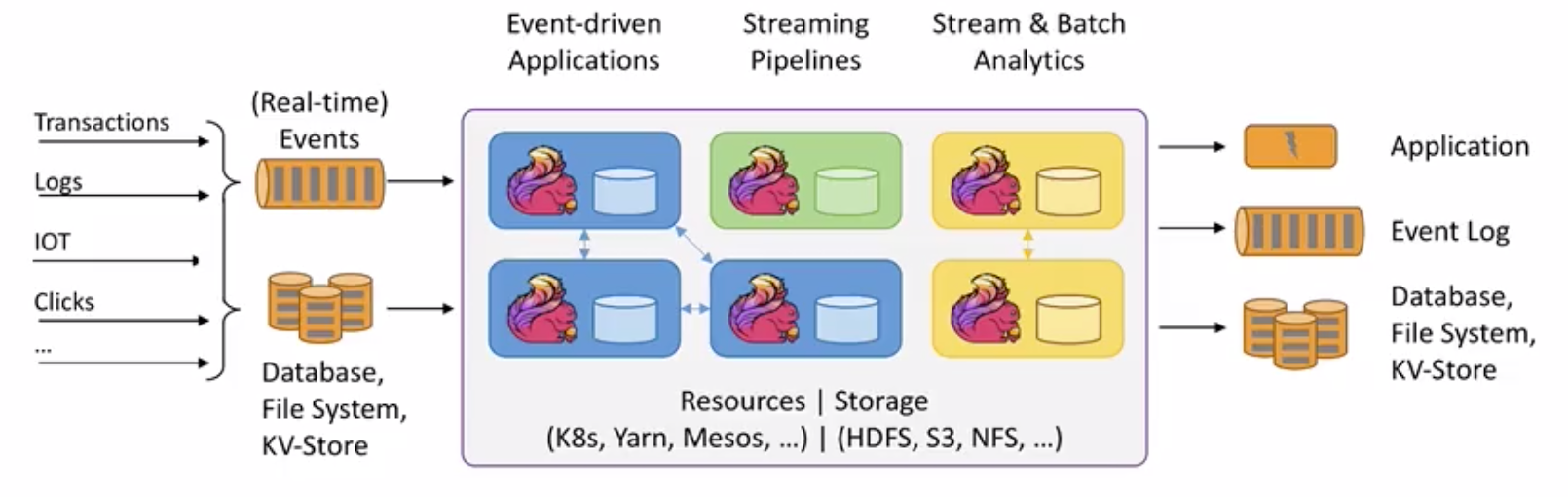
**Apache Flink excels at processing unbounded and bounded data sets.**

》》 Batch VS Streaming

streams 是有自己的一些特征的，Flink 是一个多功能的框架，可以处理各种各种的 streams。

Bounded and unbounded streams：如无界的和有界的数据集，

Real-time and recorded streams：

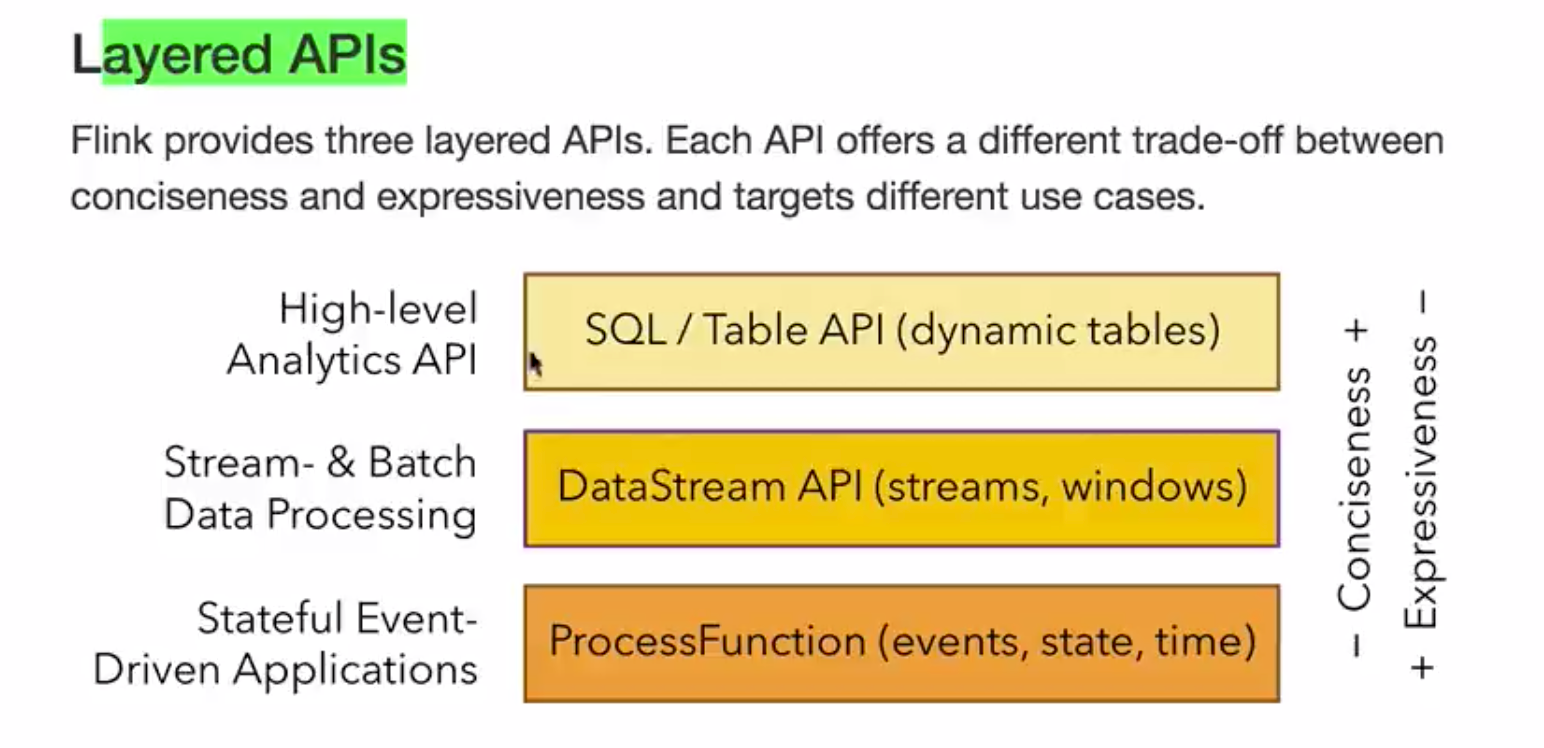


官网的图，必备，必须理解

### Flink Layered API

提供接口的层级：

Applications:



三个层级的 API

第一个：事件驱动的 API 级别，

第二个：流处理和批处理的

第三个：高级别的 API ，就是 SQL

这三个级别就是在简介性和表达性做一个权衡。目前工作中用的最对的就是中间的一层，SQL 的 api 还不是很成熟

### Flink 运行多样化

Deploy Applications Anywhere

Flink 可以运行在 yarn，mesos，k8s，也可以在 stand-alon 上面的。

如果有问题，就会替换 container，去重新申请新的资源(resource)，并且都是通过 REST 的调用方式，这样就会很容易的集成到别的环境。

Run Application at any Scale

exactly-once

### 业界流处理框架对比

Spark：对于流处理的是：SparkStreaming，Structured Streaming

Spark Streaming 是用 minibatch 方式来进行处理的

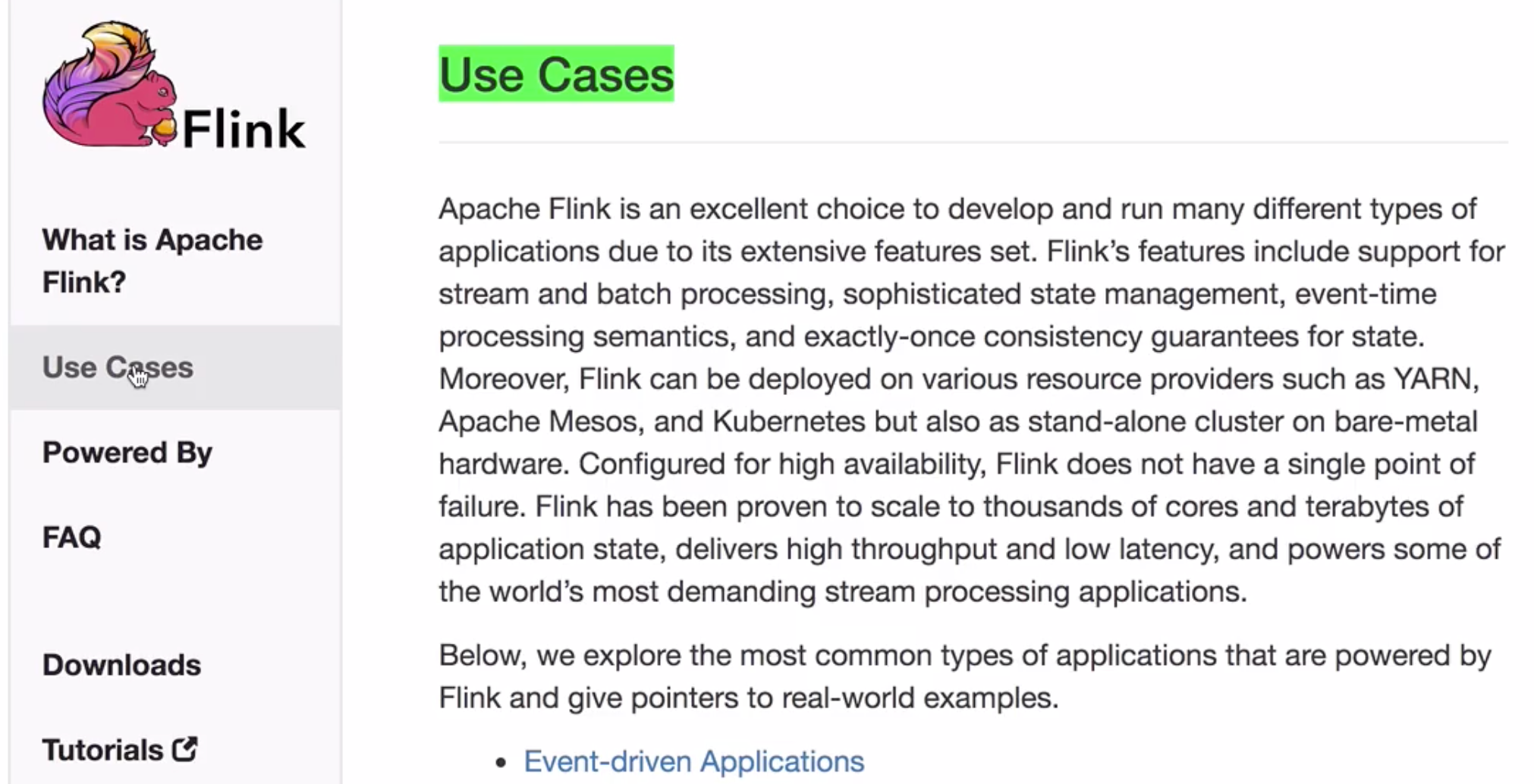
生态圈比较强大

Flink：相反的，流式为主的，批处理是流式处理的一个特例

Storm：Tuple，真正的流式处理框架，但是只能进行流式处理

### Flink Use Cases

Use Cases:



这是官网的一个介绍，使用案例。

可以给我们一些很好的应用提示和思考。

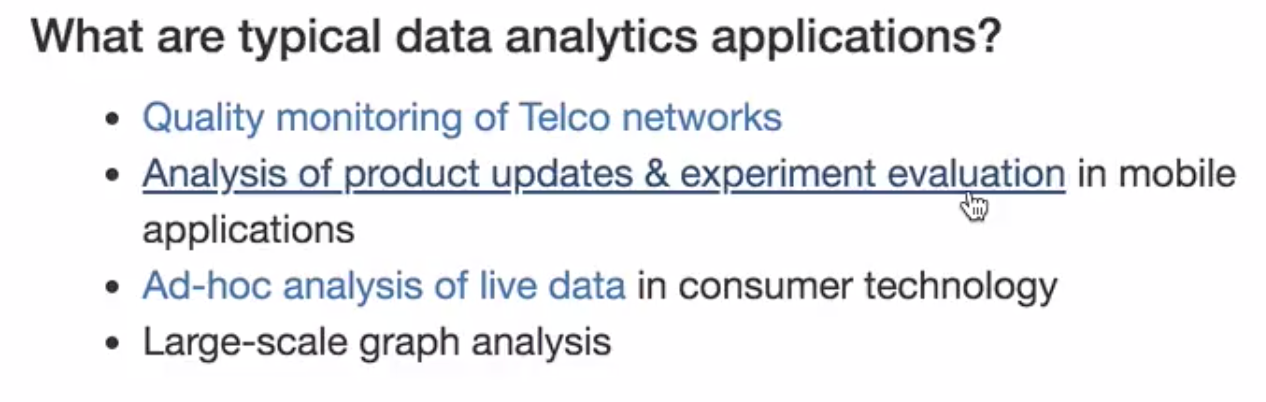
下面是一些的事件的应用：



规则的告警

欺诈监控

下面还有数据分析的一些应用：



数据管道的应用：

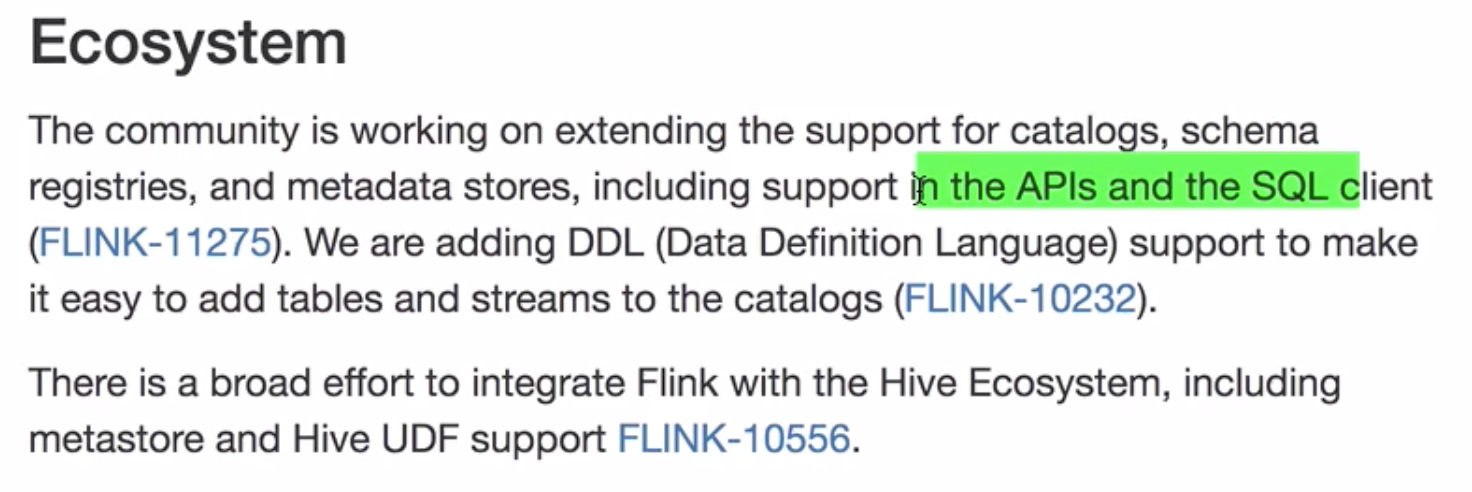


### Fink 发展趋势

Apache Fink Roadmap



下面是生态：



阿里收购了 Flink 的母公司，从这个来看，内部是用 Blink。

### 如何正确的姿势来学习 Flink

1、官方文档：一定是一首好资料，要看官方文档

Documents

Application Development

Deployment & Operations

Debugging & Monitoring

2、源码，maven 源码关联上，并且 github 里面的 examples 代码学习。

## 快速上手开发第一个 Flink 应用程序

### 课程目录

### 开发环境准备-JDK 安装

### 开发环境准备-Maven 安装

### 开发环境准备-IDEA 安装

### Flink 批处理应用- Java 开发环境准备

### Flink 批处理应用-八股文编程

### Flink 批处理应用-Java 开发功能实现

### Flink 批处理应用-scala 开发环境准备

### Flink 批处理应用-scala 开发功能实现

### 使用 Java和 scala 开发 Flink 应用程序对比

### Flink 实时处理应用 Java开发功能实现

### Flink 实时处理应用 Java 开发代码重构

### Flink 实时处理应用scala 开发代码重构

### 开发过程中依赖的注意事项