为什么要学习 Apache Flink?

巴真 （陈守元）

Apache Flink Community China Training

CONTENT
目录 >>

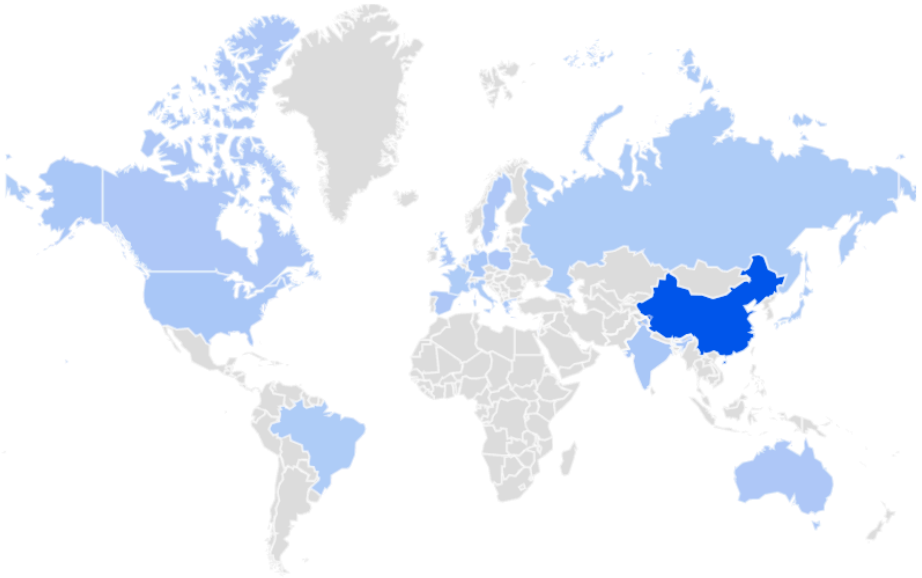
01

课程介绍

为什么会开设系列课程?

Flink在中国采用情况

Flink在全球热度情况



1	中国	100
2	新加坡	7
3	以色列	5
4	荷兰	5
5	德国	5

Flink社区希望解决的问题

丰富需求端

扩大整个IT行业对于Flink技术栈的诉求，让大量公司基于Flink Stack构建完整大数据体系

赋能供应端

扩大整个IT从业人员对于Flink熟悉掌控程度，让大量从业人员成为Flink深度开发者

搭建供需桥梁

搭建企业与员工的桥梁，让更多使用Flink Stack公司接触到更多Flink专业人员

我们系列课程目标

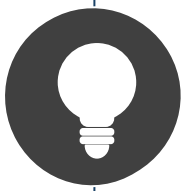
面向人群

对于Flink、或者Bigdata感兴趣的IT初学者、在校大学生

前置知识

熟悉基本的编程语言(Java、Python)

大数据、数据处理有基本的了解



课程目标

(第一季)课程能够初步了解Flink技术栈，初步掌握Flink API，完成简单生产业务开发

后续规划

前期面向Flink开发者

后期面向架构师

我们系列课程计划

扫码访问课程地址

02

Apache Flink

定义/原理/应用

Flink Application
Flink应用开发相关知识

Flink Architecture
Flink基本架构原理以及核心逻辑

Flink Operation
Flink运维管理相关内容

Apache Flink Definition

Apache Flink is a **framework** and **distributed** processing engine for **stateful** computations over ***unbounded and bounded*** **data streams**.

Flink Application

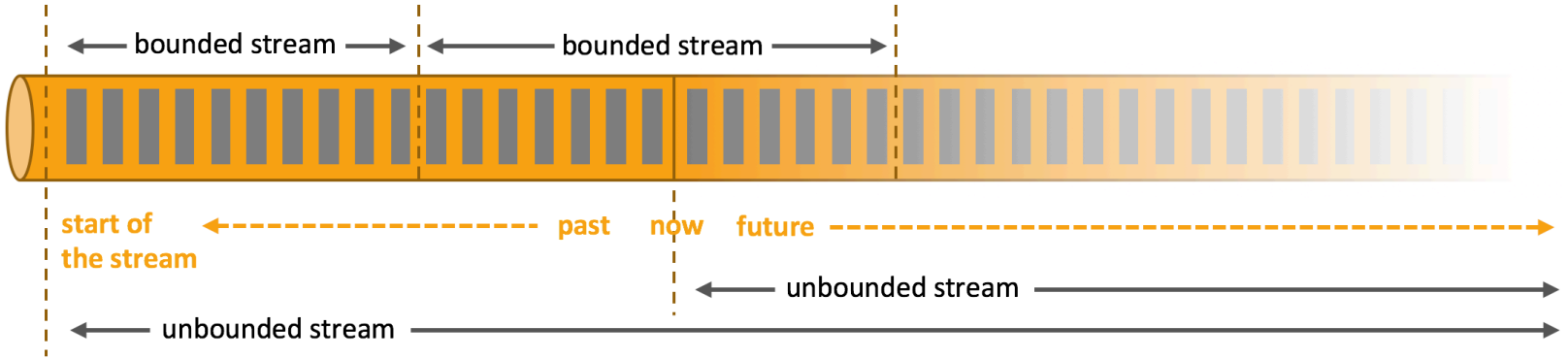
基础处理语义

Streams、State、Time

多层次API

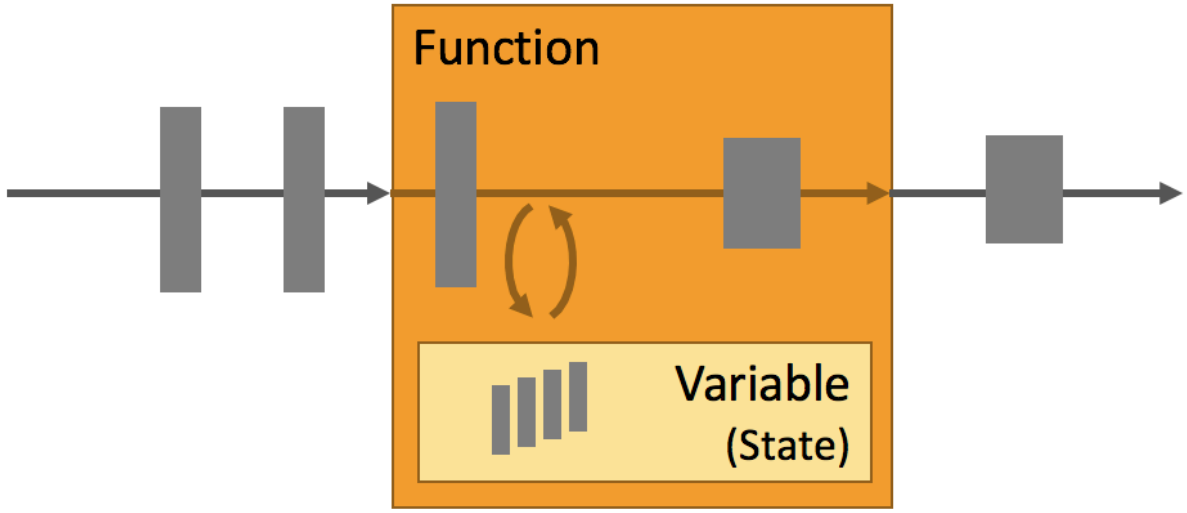
灵活性和方便性的兼顾

Flink Application – Streams



Unbounded streams have a start but no defined end.
Bounded streams have a defined start and end.

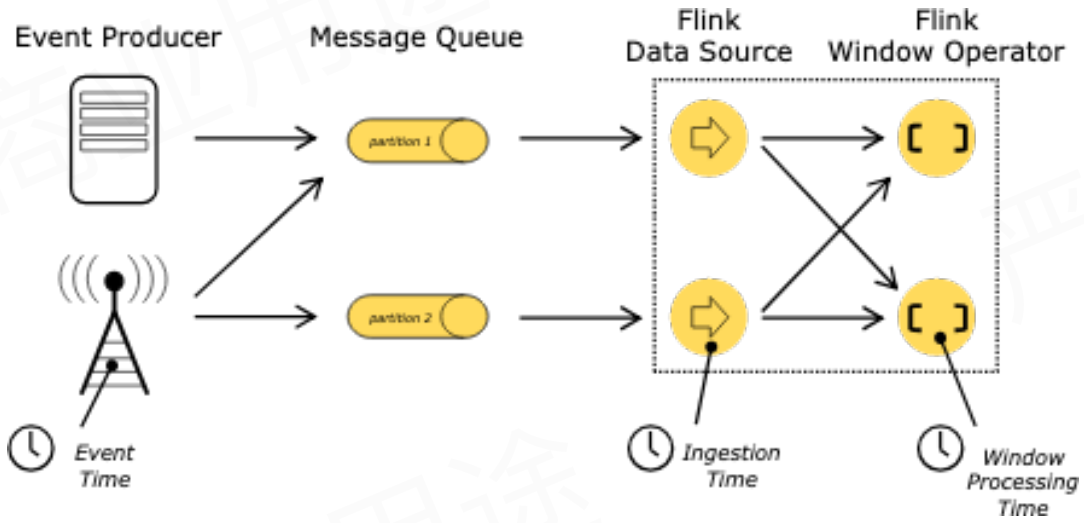
Flink Application – State



Apache Flink is xxxx processing engine for **stateful** computations. Application state is a first-class citizen in Flink.

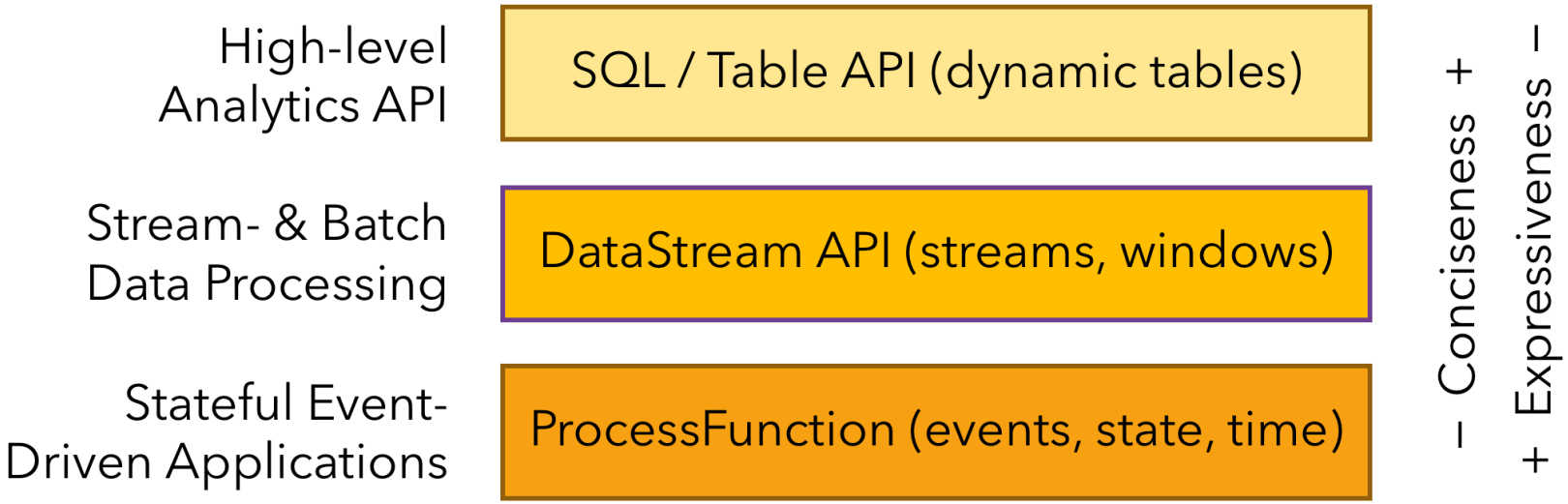
Keys: **Incremental Processing**, **Exactly-once Semantics**

Flink Application – Time



- **Event Time** is the time when an event was created. It is usually described by a timestamp in the events
- **Ingestion time** is the time when an event enters the Flink dataflow at the source operator.
- **Processing Time** is the local time at each operator that performs a time-based operation.

Flink Application – API



- **ProcessFunctions**： the most expressive function interfaces that Flink offers. Flink provides ProcessFunctions to process individual events from one or two input streams or events that were grouped in a window.
- **DataStreamAPI**： provides primitives for many common stream processing operations, such as windowing, record-at-a-time transformations, and enriching events by querying an external data store.
- **SQL/TableAPI:** relational APIs.

Flink Architecture

有界和无界数据流

Flink 具备一套框架处理两种数据集合

部署灵活

Flink支持多种部署方式，包括Yarn、K8S

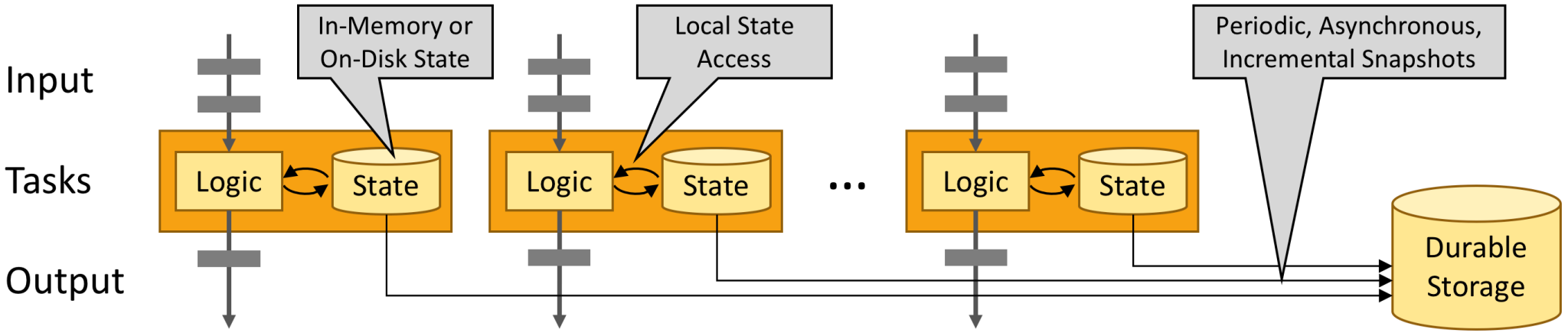
极高可伸缩性

峰值达17亿条/s，无需任何业务语义调整

极致流式处理性能

本地状态存取，极致性能优化

Flink Architecture – Stateful



Stateful Flink applications are optimized for **local state access.** Flink guarantees exactly-once state consistency in case of failures by periodically and asynchronously checkpointing the local state to durable storage.

Flink Operation

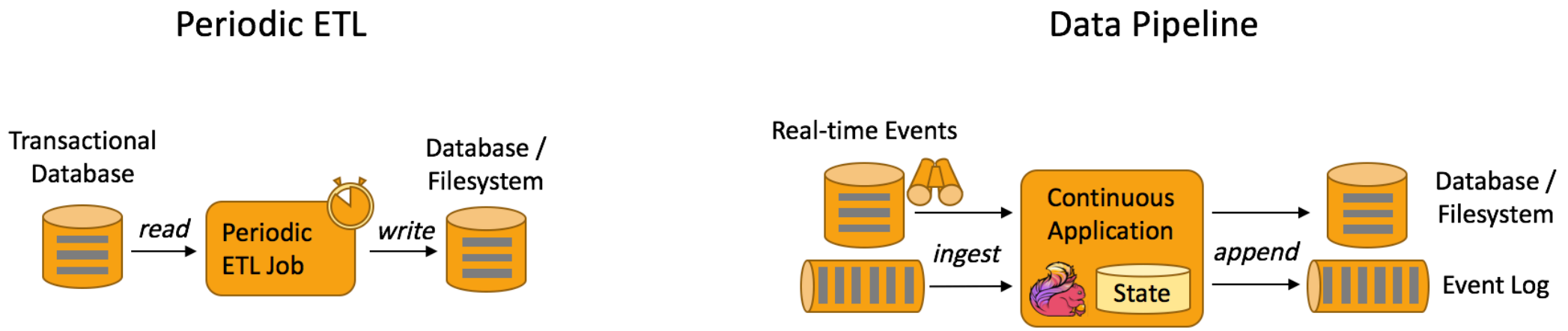
7x24小时高可用

一致性Checkpoint、高效的Checkpoint

业务应用监控运维

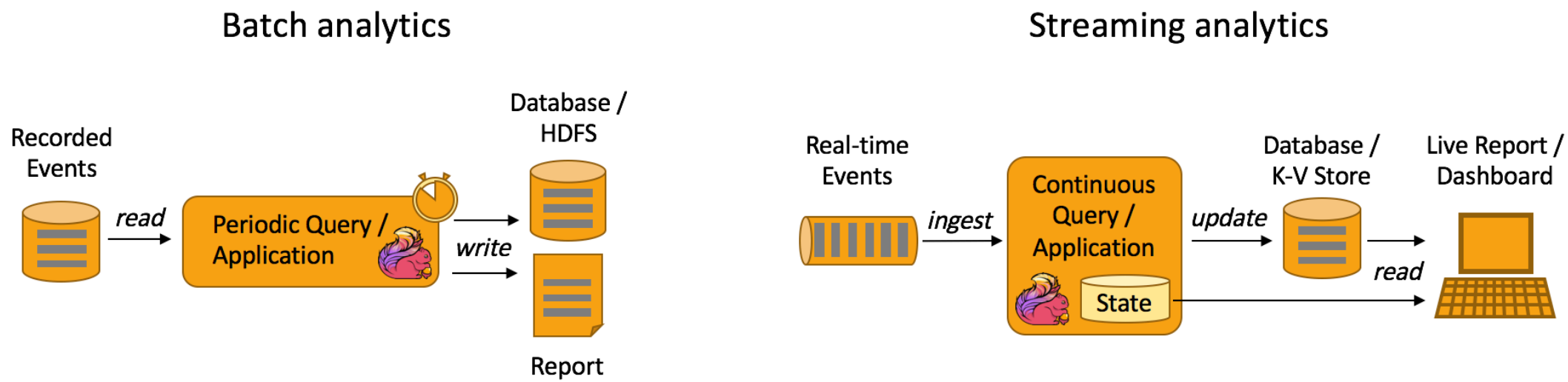
WebUI、Metric

Flink Scenario: Data Pipeline



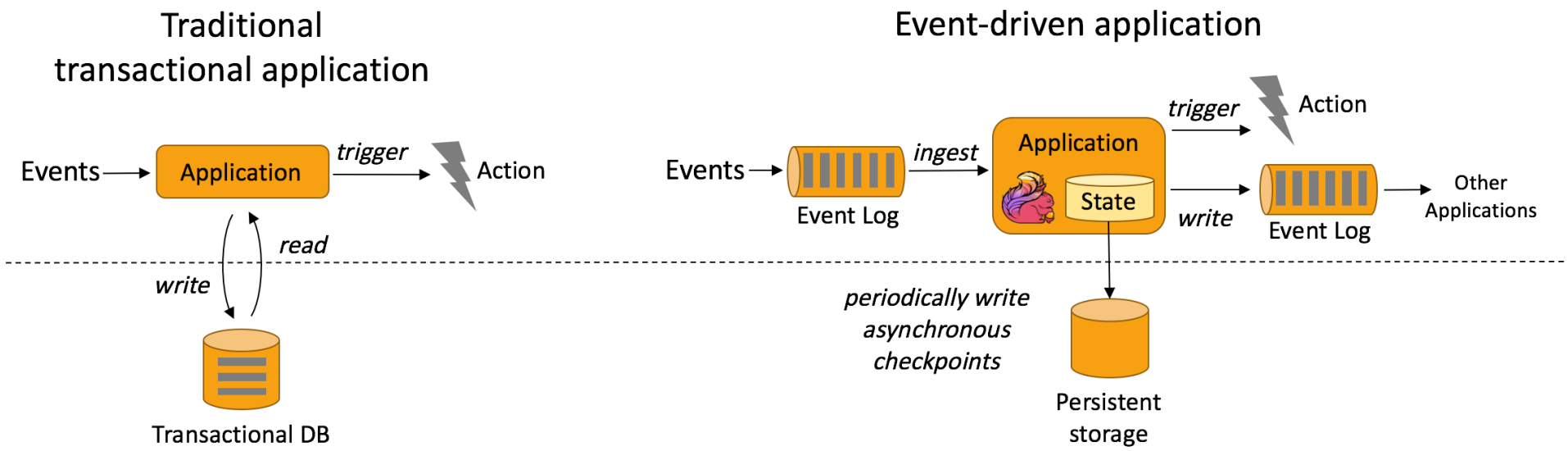
Extract-transform-load (ETL) is a common approach to convert and move data between storage systems. Flink data pipelines operate in a continuous streaming mode instead of being periodically triggered

Flink Scenario: Data Analytics



Analytical jobs extract information and insight from raw data.

Flink Scenario: Data Driven



An event-driven application is a stateful application that ingest events from one or more event streams and reacts to incoming events by triggering computations, state updates, or external actions.

03

高效学习

学前准备以及学习方法

学习准备

环境条件

操作系统: **Linux**、**MacOS**、Windows

JDK版本: 8.x以上

Flink本地环境搭建

下载地址: https://flink.apache.org/downloads.html

搭建方法: https://ci.apache.org/projects/flink/flink-docs-release-1.7/tutorials/local_setup.html

DEMO演示

DEMO

学习建议

1. **先实践再理论。**先学习应用，尝试构建复杂的Flink Application
2. **横向扩展。**在构建复杂Flink生产业务后，横向使用学习Storm、Spark、DataFlow等系统，知识是演化过来的，必有前置和铺垫。多横向看看，打开视野。
3. 关注下Apache Flink以及Flink China社区，多交流、多提问、**多输出**。

课后作业

画一个Flink的思维导图
https://www.xmind.net/m/6fk4/

THANKS