一 Basic API Concepts

Flink程序是常规的应用程序(实现了分布式集合的转换(如filtering, mapping, updating state, joining, grouping, defining windows, aggregating)). 集合最初是由数据源(如 reading from files, kafka topics, local, in-memory 集合)读出来的. 结果通过sink(槽)返回(比如可以将数据写到分布式文件中或者标准输出—控制台). Flink应用程序可以运行早local本地模式 或者 多个机器构成的集群模式.

根据数据源的类型(有边界有限的数据源和无边界无线的数据源),你可以编写批处理应用程序和流应用程序, 其中批处理应用程序使用的是DataSet API, 流应用程序使用的是DataStream API.

注意官网的案例本节展示的API 例子都是使用的StreamingExecutionEnvironment 和 DataStream API, 这些概念和DataSet API一模一样, 仅仅是使用ExecutionEnvironment 和 DataSet API 进行替换而已.

* 1. DataSet and DataStream
  2. Anatomy of a Flink Program
  3. Lazy Evaluation
  4. Specifing Keys
  5. Specifying Transformation Functions
  6. Supported Data Types
  7. Accumulators & Counters