



Apache SeaTunnel API 重构

SeaTunnel与计算引擎解耦

Apache SeaTunnel Contributor 李宗文

2022-05-14



CONTENT

01

背景与动机

02

重构目标

03

整体设计

04

Source API

05

Sink API



01

背景与动机



背景与动机



SeaTunnel与引擎耦合

- Connector实现次数多
- 参数不统一
- 难以支持多个版本的引擎
- 引擎升级难



02



重构目标



重构目标



SeaTunnel与引擎解耦

- Connector只实现一次
- 支持多个版本的Spark & Flink引擎
- 明确Source的分片并行逻辑，Sink的提交逻辑
- 支持实时场景下的数据库整库同步
- 支持元信息的自动发现与存储

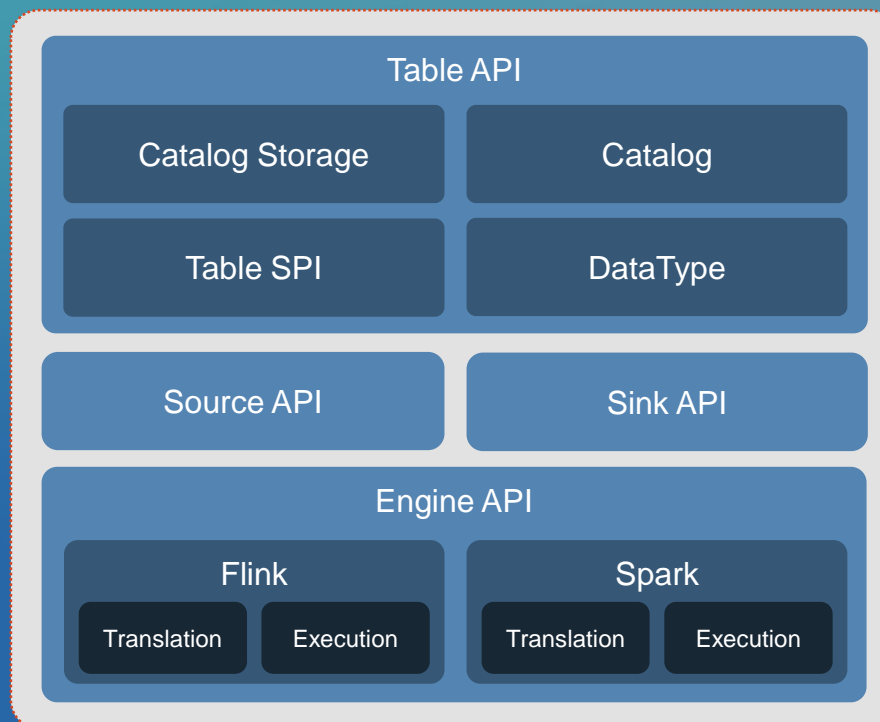


03

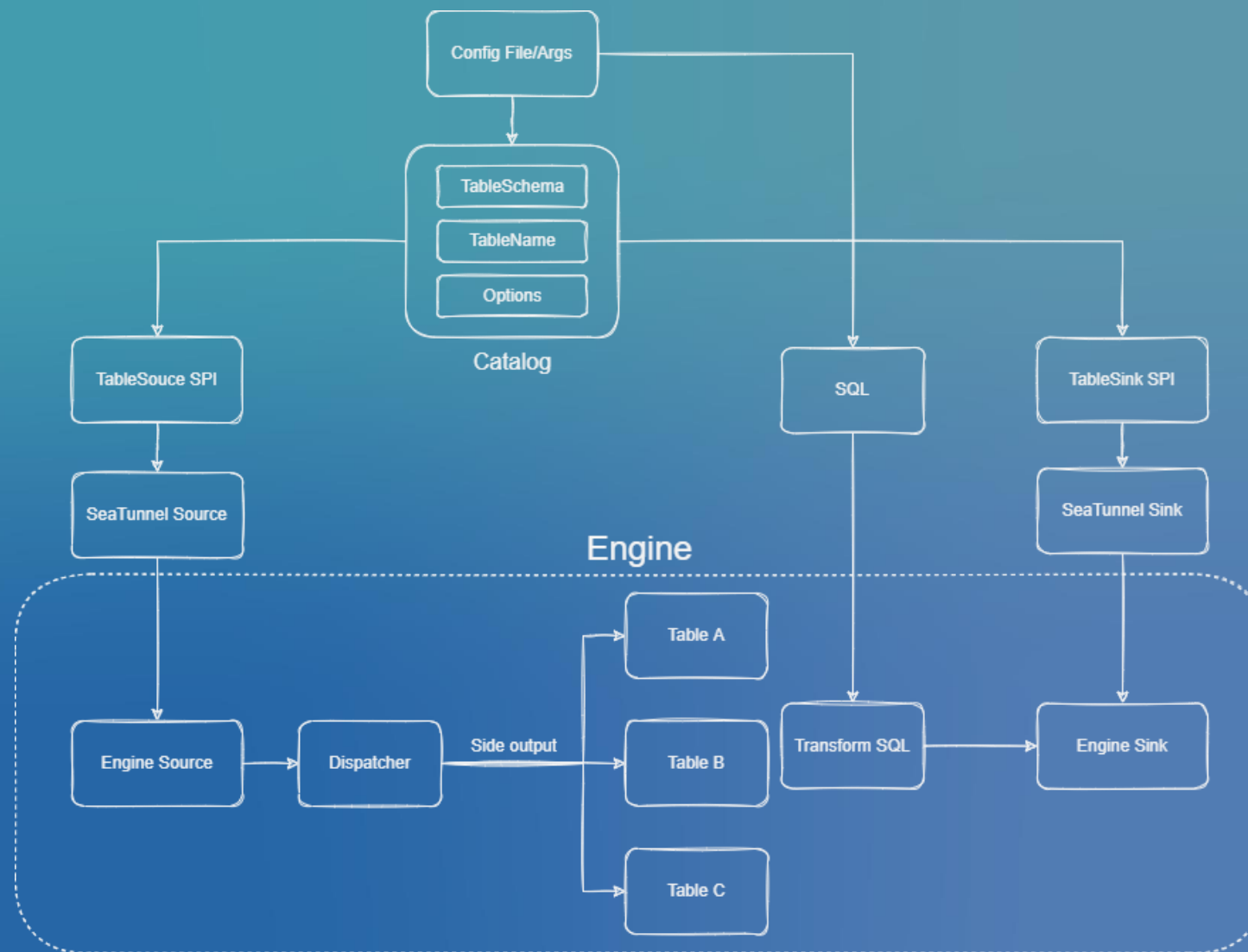


整体设计

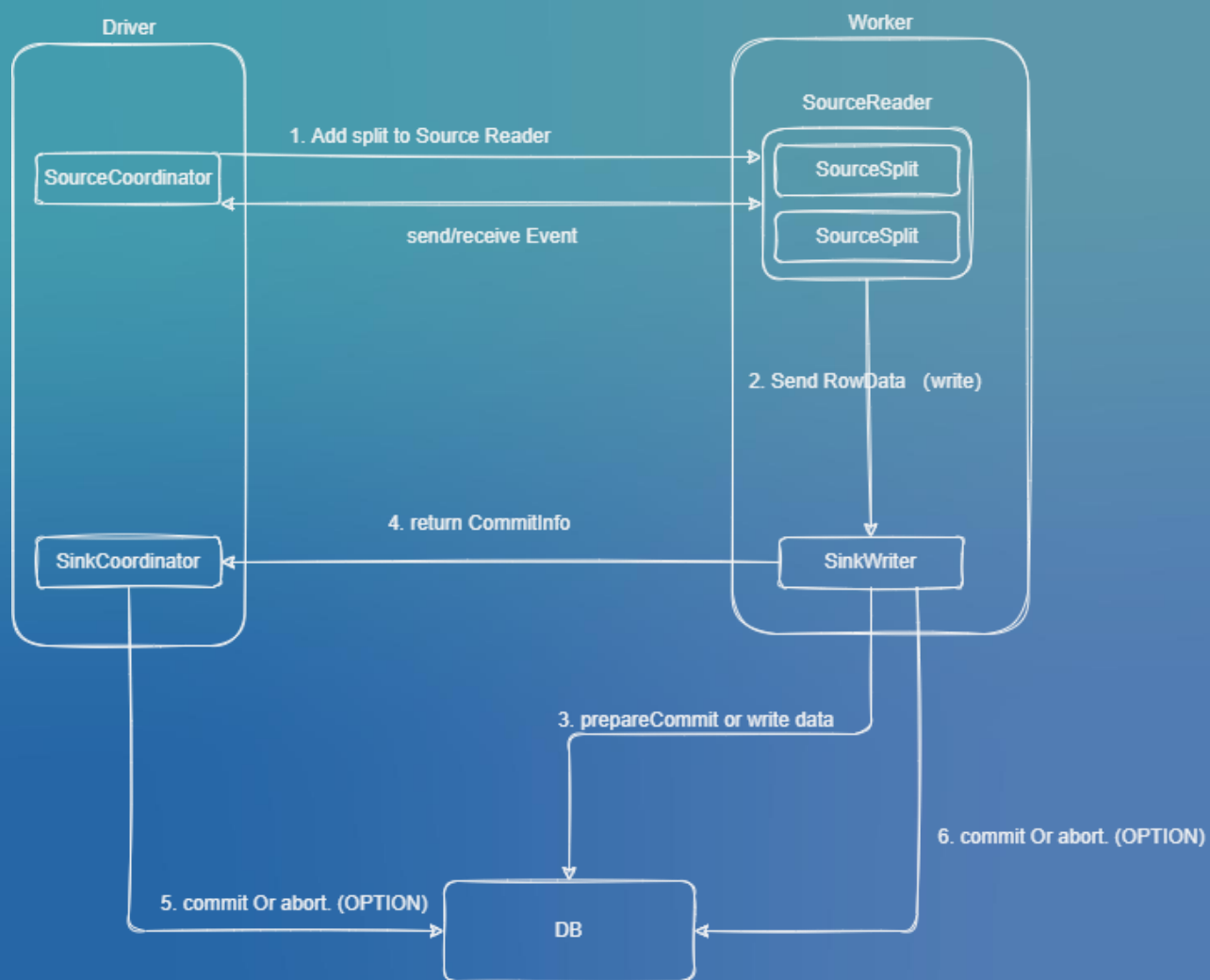
SeaTunnel API



Execution Flow



Connector Flow





04

Source API



Source API



Source 所需的特性

- 统一离线与实时API
- 支持并行读取
- 支持动态添加分片
- 支持协调读取器工作
- 支持单个读取器处理多张表

Source API



Source 所需的特性

- 统一离线与实时API
- 支持并行读取
- 支持动态添加分片
- 支持协调读取器工作
- 支持单个读取器处理多张表

API

- Boundedness
- SourceReader
- SourceSplit & Enumerator
- SupportCoordinate & SourceEvent
- SupportMultipleTable

Source API



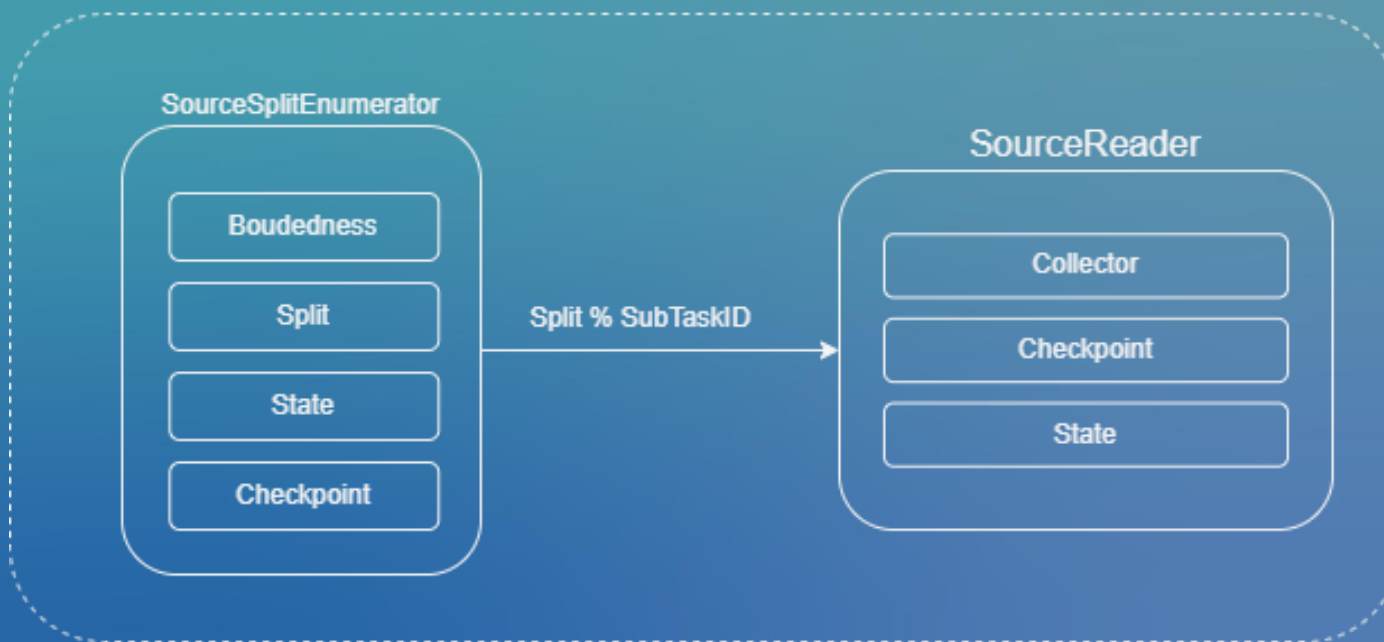
如何适配Spark与Flink引擎?

- 如何支持动态添加分片?
- 如何支持协调读取器?
- 如何支持单个读取器处理多张表?

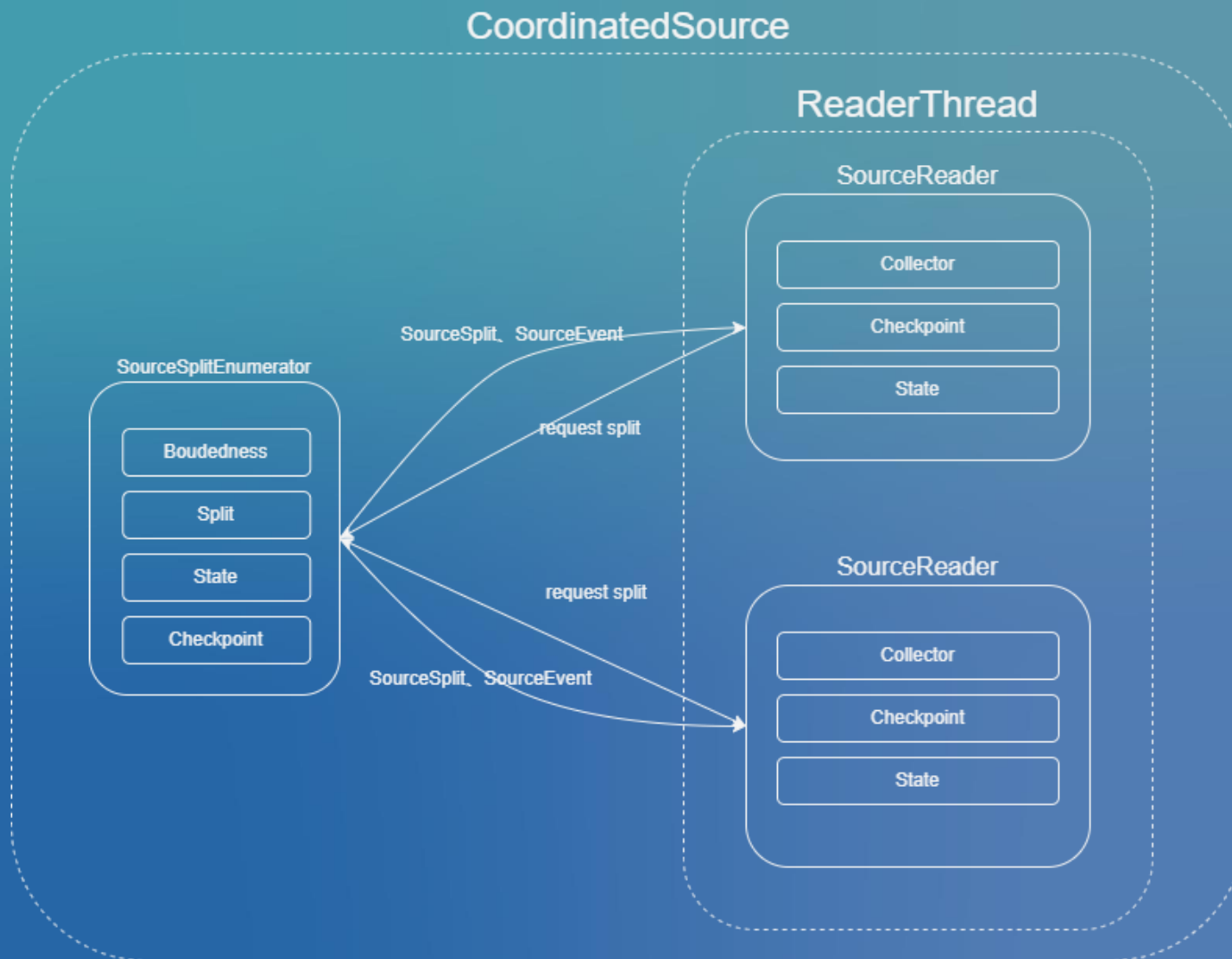
Source API



ParallelSource



Source API





05

Sink API



Sink API



Sink 所需的特性

- 分布式事务
- 聚合提交



Sink API

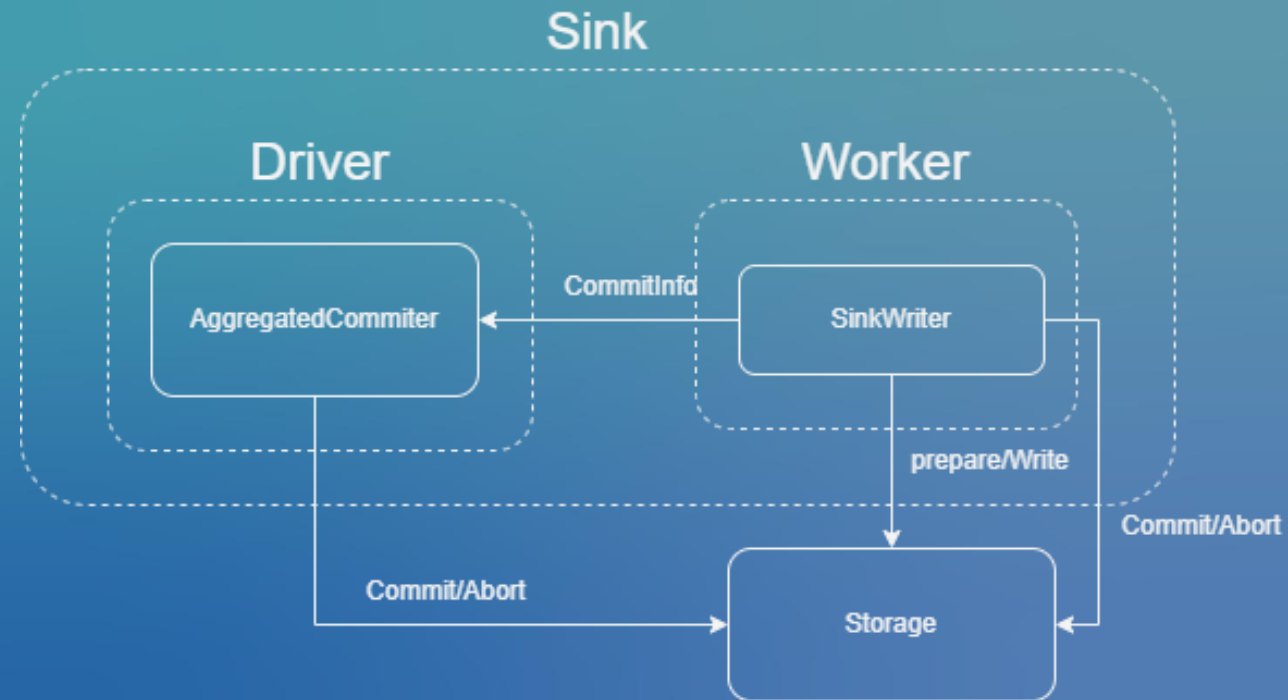
Sink 所需的特性

- 分布式事务
- 聚合提交

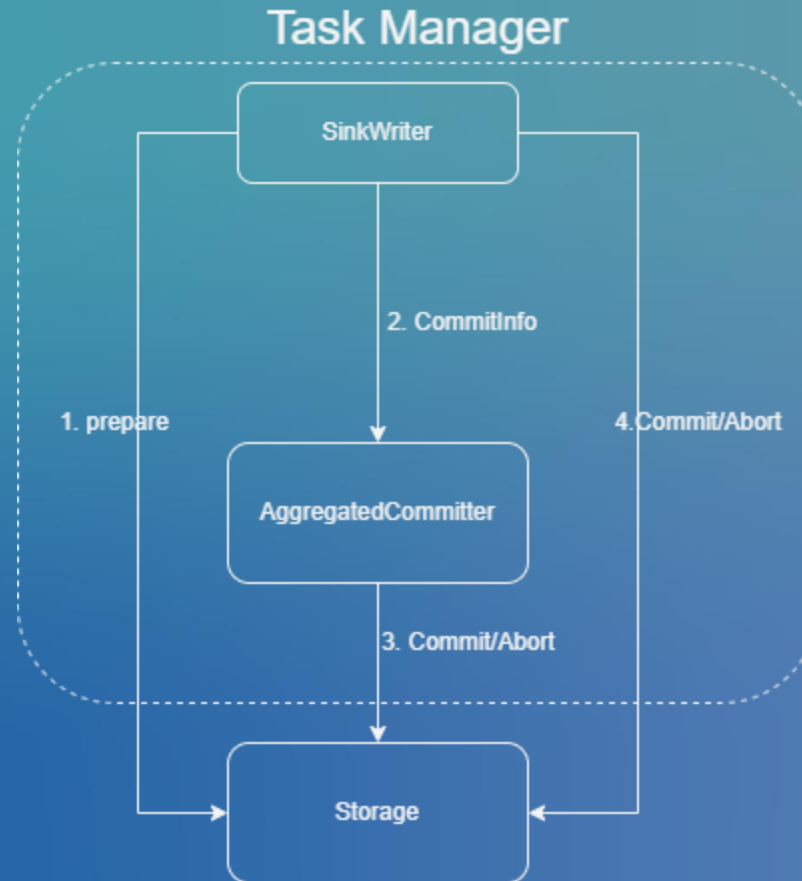
API

- SinkWriter
- SinkCommitter
- SinkAggregatedCommitter

Sink API



Sink API





Thanks

