

Libraries

SQL

Table

CEP

Gelly / Graphs

ML

Setup Runner Core

DataStream

DataSet

Distributed
Dataflow Runtime

Cluster

Backend

Config

(a) The Flink Software Stack

`stream.map(..).`
`..window(..)...`

Client

*Optimised Logical
Dataflow Graph*

Zookeeper

Job
Manager

- *Passive Failover*
- *Metadata for Execution and Snapshots*

- **Managed Memory Allocation**
- **Resource management**
- **Local Snapshot Execution**

Task
ManagerTask
Manager*Physical Tasks*
snapshot
storeLocal
Snapshots*managed state
(heap / out-of-core)*

(b) The Flink System Architecture