

基于FLINK的实时计算平台

静行 邓小勇

阿里巴巴 高级技术专家

2018-09



Flink China



Flink China

个人介绍

xiaoyong.dengxy@alibaba-inc.com



Flink China

2010 — 2013

2013 — 2014

2014 — 2015

2015 — 至今

- 毕业加入阿里巴巴
- 搜索引擎测试
- Dump中心开发

- 实时数据同步
- 实时数据离线处理
- 实时数据质量

- 离线数据同步
- 分布式DataX
- 云上数据同步

- 流计算平台
- 基于Flink的开发运维
- 云上实时业务



Flink China

IT大咖说
知识共享平台



第一章

整体架构

第二章

任务开发

第三章

任务运维

第四章

发展规划

目录

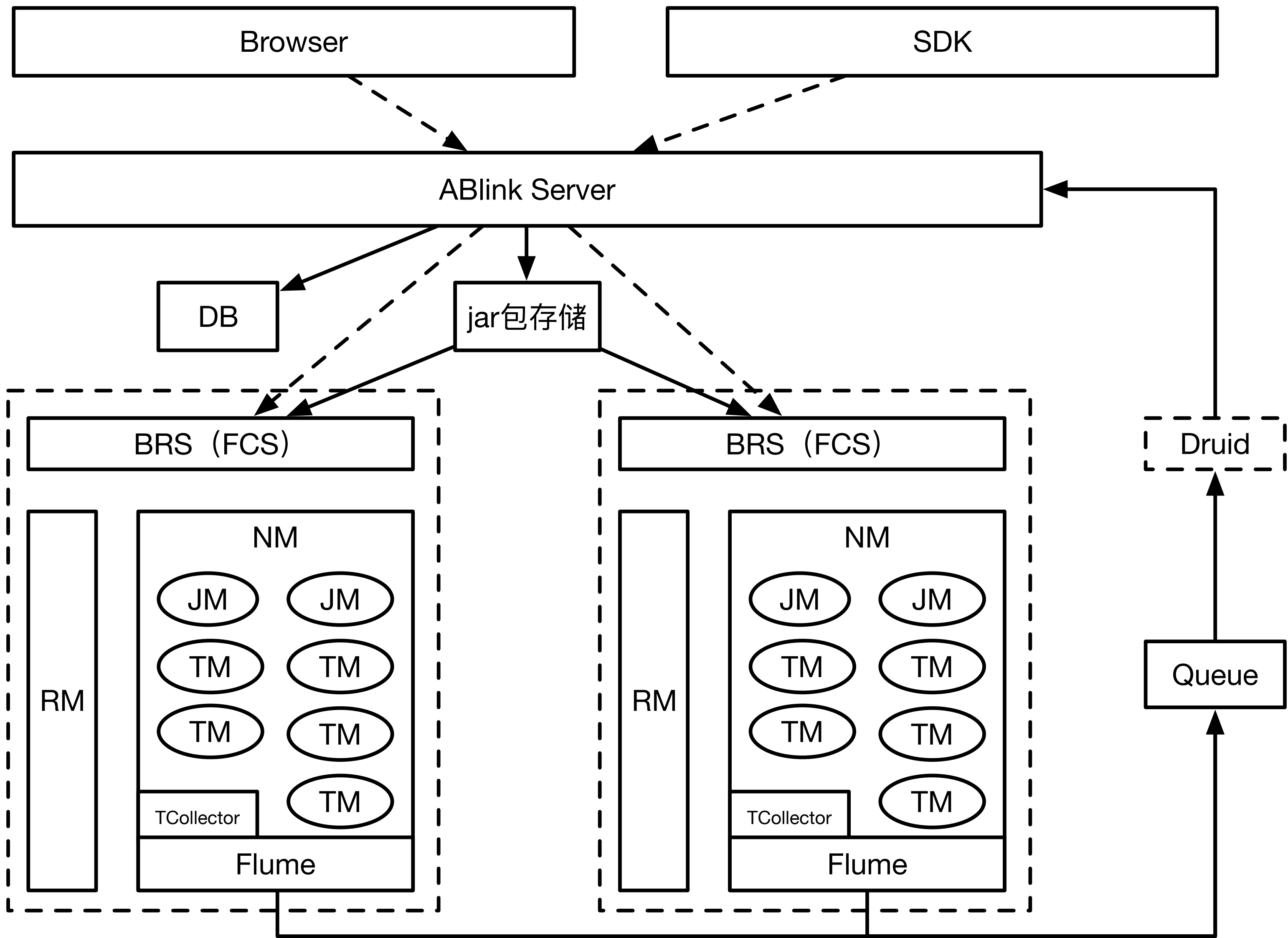




One

整体架构







整体架构



开发

版本管理	版本历史	版本diff	版本回滚
任务上线	资源配置	autoconfig	深度检查
任务调试	测试数据上传	测试数据下载	数据抽样
	调试日志	调试结果	
数据源管理	多种数据源接入	数据预览	生成DDL
SQL编写	语法高亮	语法检查	智能提示
	作业参数	代码结构	
资源管理	资源上传	资源下载	资源引用
引擎管理	引擎版本升降级	引擎关联任务	集群切换
基础操作	自动保存	另存为	撤销重做
	主题设置		
文件夹管理	任务新建	任务定位	任务搜索
	任务删除	文件夹刷新	

运维

任务大盘	任务状态	业务延时	资源消耗
	启动位点	操作人	优先级
任务操作	启动	暂停	恢复
	停止	下线	批量操作
Vertex信息	瞬时指标	任务DAG	使用资源状态
	反压点	Vertex DAG	Vertex 指标曲线
Container信息	Container List	Container日志	Container指标
数据曲线	延迟曲线	输入/输出TPS曲线	checkpoint曲线
	state曲线	GC曲线	watermark曲线
	JVM曲线	CPU消耗曲线	MEM消耗曲线
其他展示	FailOver日志	血缘关系	告警配置
系统运维	Blink版本滚动升级	大促封网	优先级管理
	成本计量	作业压测	

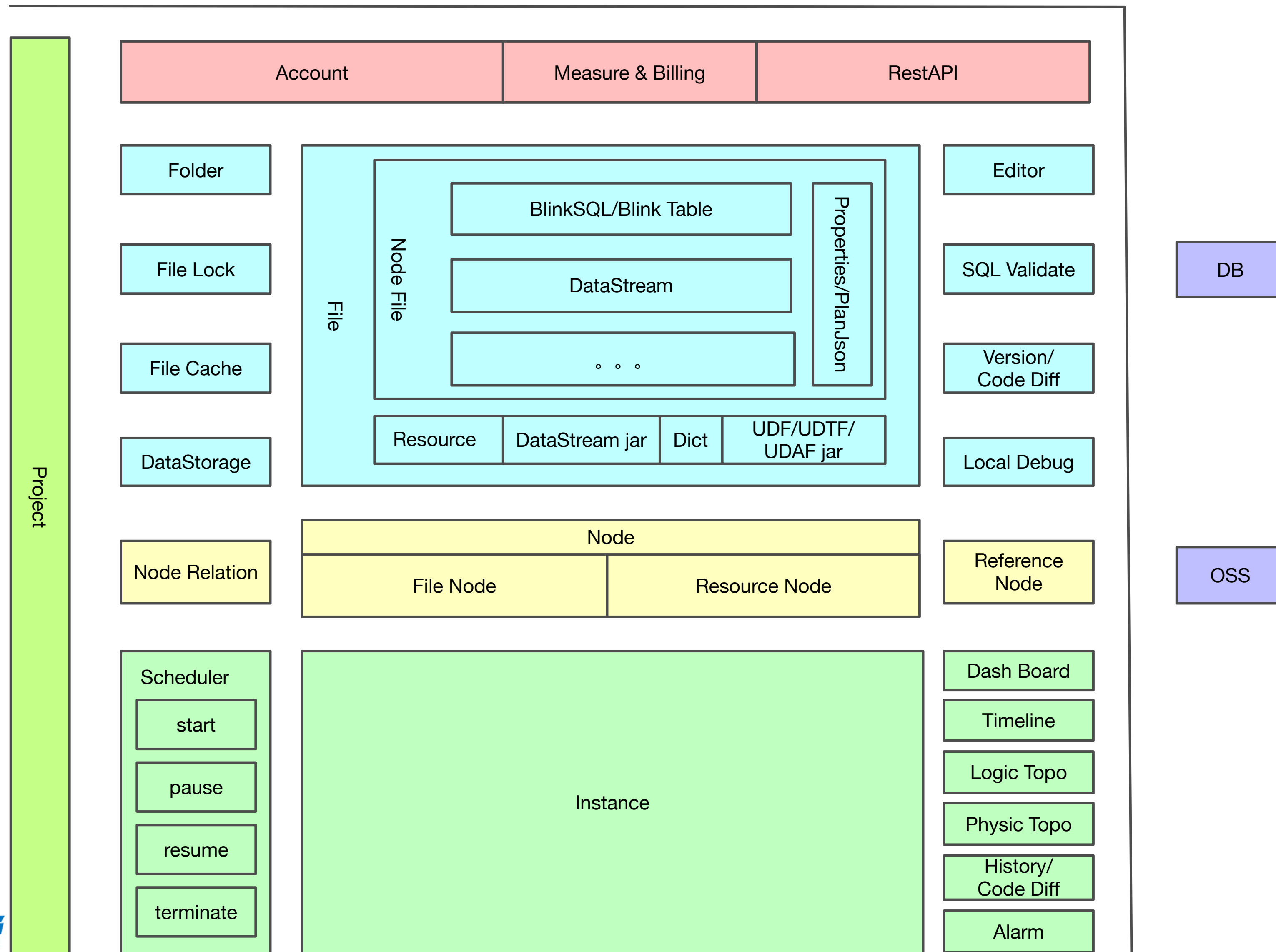




Two

任务开发







Flink China

阿里流计算开发平台

bayes_team

总览 开发 运维

作业开发 数据仓储 资源引用 运行引擎

新建作业 另存为 保存 撤销 重做 查找 调试 语法检查 上线 运维 格式化 设置 全屏

brs2_repro test_dat_stream test_data_stream test_data_stream... debug_fortest

```
25 uid7 VARCHAR,
26 time8 VARCHAR,
27 uid9 VARCHAR,
28 time10 VARCHAR,
29 uid11 VARCHAR,
30 time12 VARCHAR,
31 uid13 VARCHAR
32 ) WITH (
33   type = 'jdbc',
34   endPoint = 'jdbc:mysql://10.10.10.10:3306',
35   accessId = 'jdbc:mysql://10.10.10.10:3306',
36   accessKey = 'jdbc:mysql://10.10.10.10:3306',
37   project = 'jdbc:mysql://10.10.10.10:3306',
38   logStore = 'jdbc:mysql://10.10.10.10:3306'
39 );
40 CREATE TABLE print_test ('name' TIMESTAMP) WITH (TYPE = 'jdbc');
41 INSERT INTO
42   print_test
43 SELECT
44   CURRENT_TIMESTAMP
```

SOURCE

sls_output

SELECT

CURRENT_TIMESTAMP

SINK

print_test

运行结束 print_test

1 [2018-08-22 16:33:25] [INFO] Blink-SQL local debug task created for job debug_fortest_48731 with id 540a256665d30943ad0dd517a655ebac

2 [2018-08-22 16:33:25] [INFO] Start to compile debug_fortest_48731

3 [2018-08-22 16:33:26] [INFO] Created file print_test.csv to store result data of print_test

4 [2018-08-22 16:33:26] [INFO] Compile finished, start to run 540a256665d30943ad0dd517a655ebac

5 [2018-08-22 16:33:27] [INFO] class org.apache.flink.streaming.api.functions.source.TimestampedFileInputSplit is not a valid POJO type

6 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

7 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

8 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

9 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

10 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

11 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

12 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

13 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

14 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

15 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

16 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

17 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

18 [2018-08-22 16:33:27] [WARN] No state backend has been specified, using default state backend (Memory / JobManager)

19 [2018-08-22 16:33:27] [INFO] State backend is set to heap memory (checkpoint to jobmanager)

最近保存时间: 2018-07-27 16:58:56

BLINK_SQL 0:0 BLINK版本: blink-1.4.5

使用推荐版本 使用测试版本 降级版本

IT大咖说
知识共享平台



Flink China



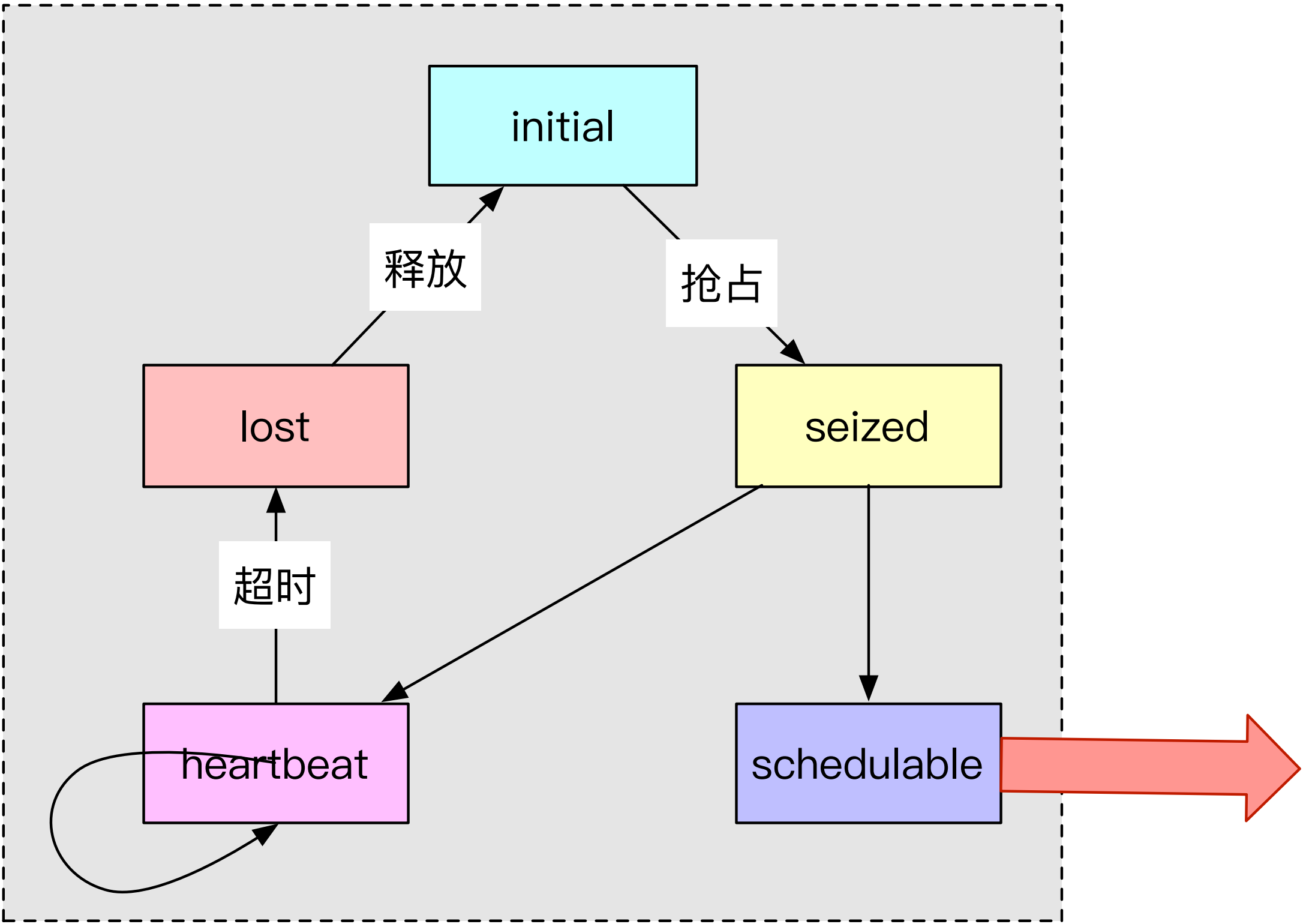
知识共享平台



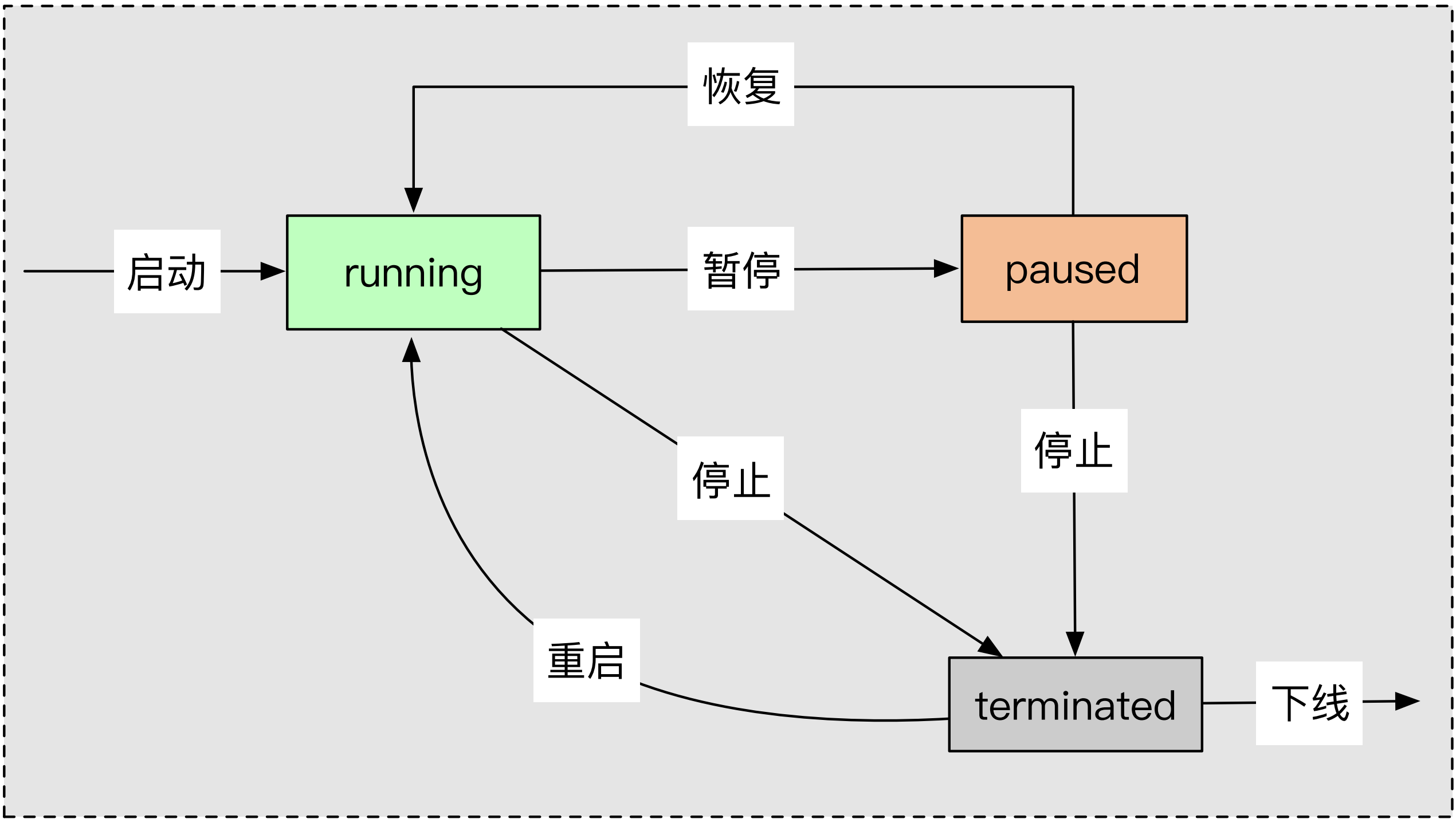
Three

任务运维





单机对Instance的维护

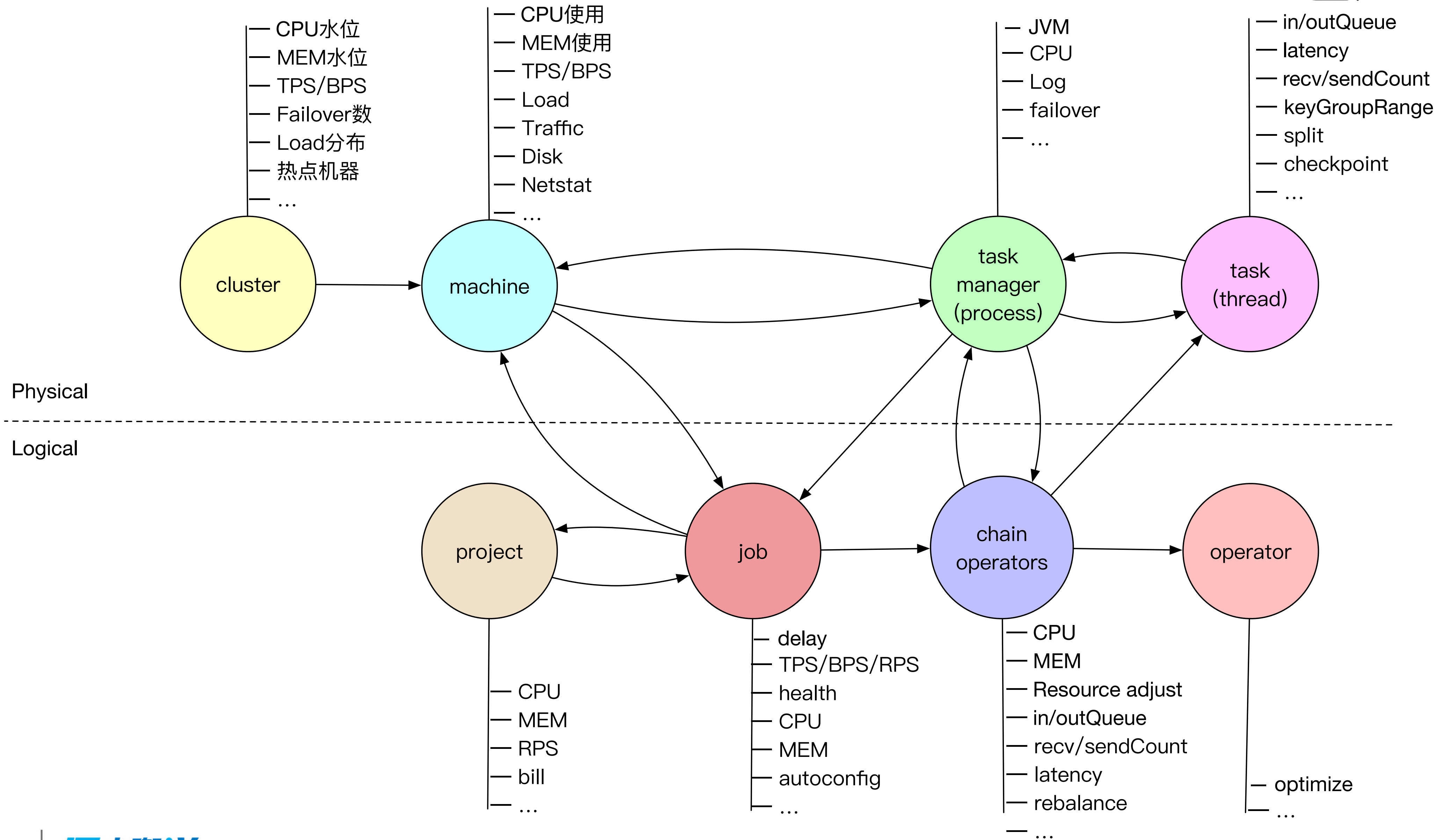


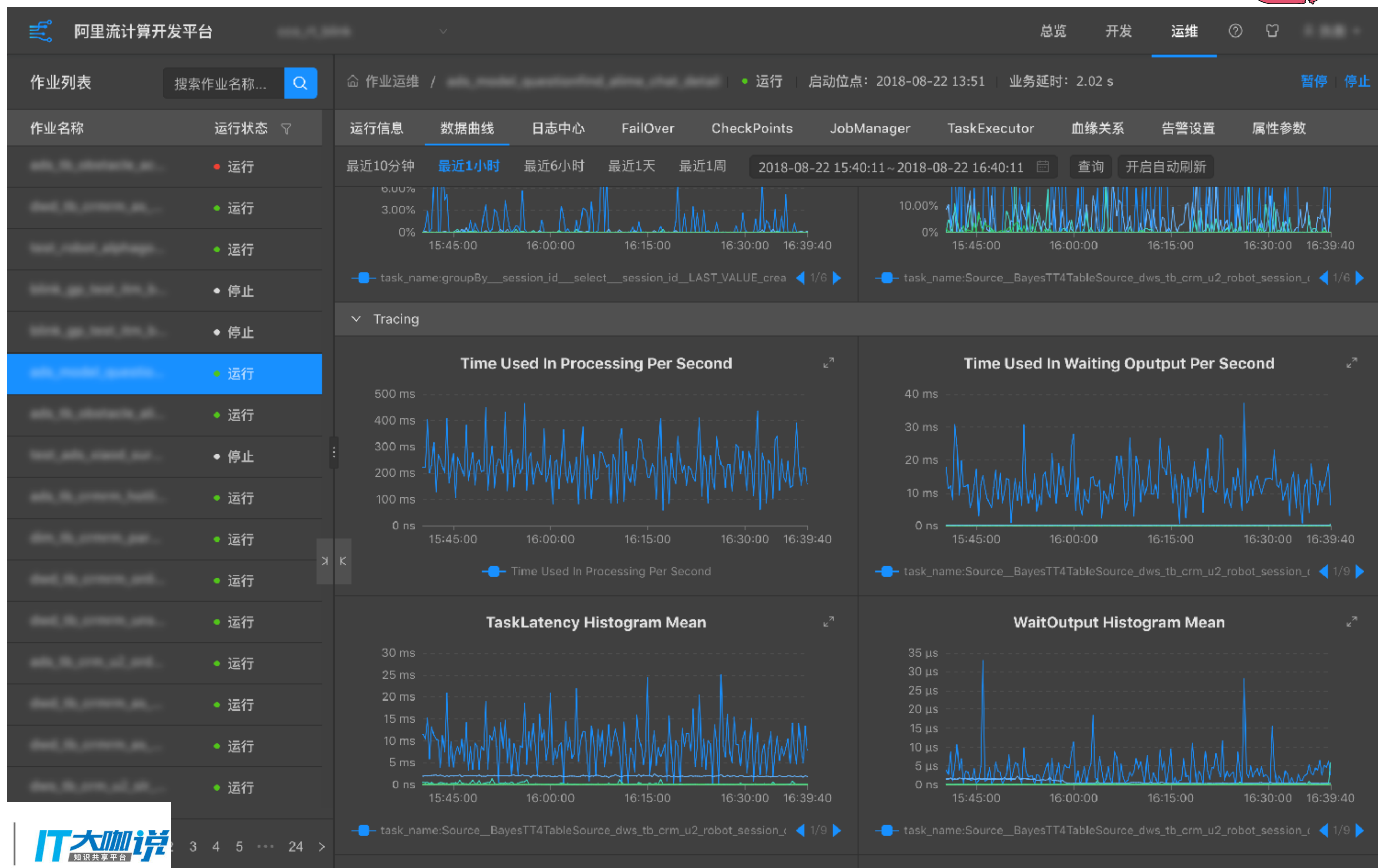
用户操作Instance状态机





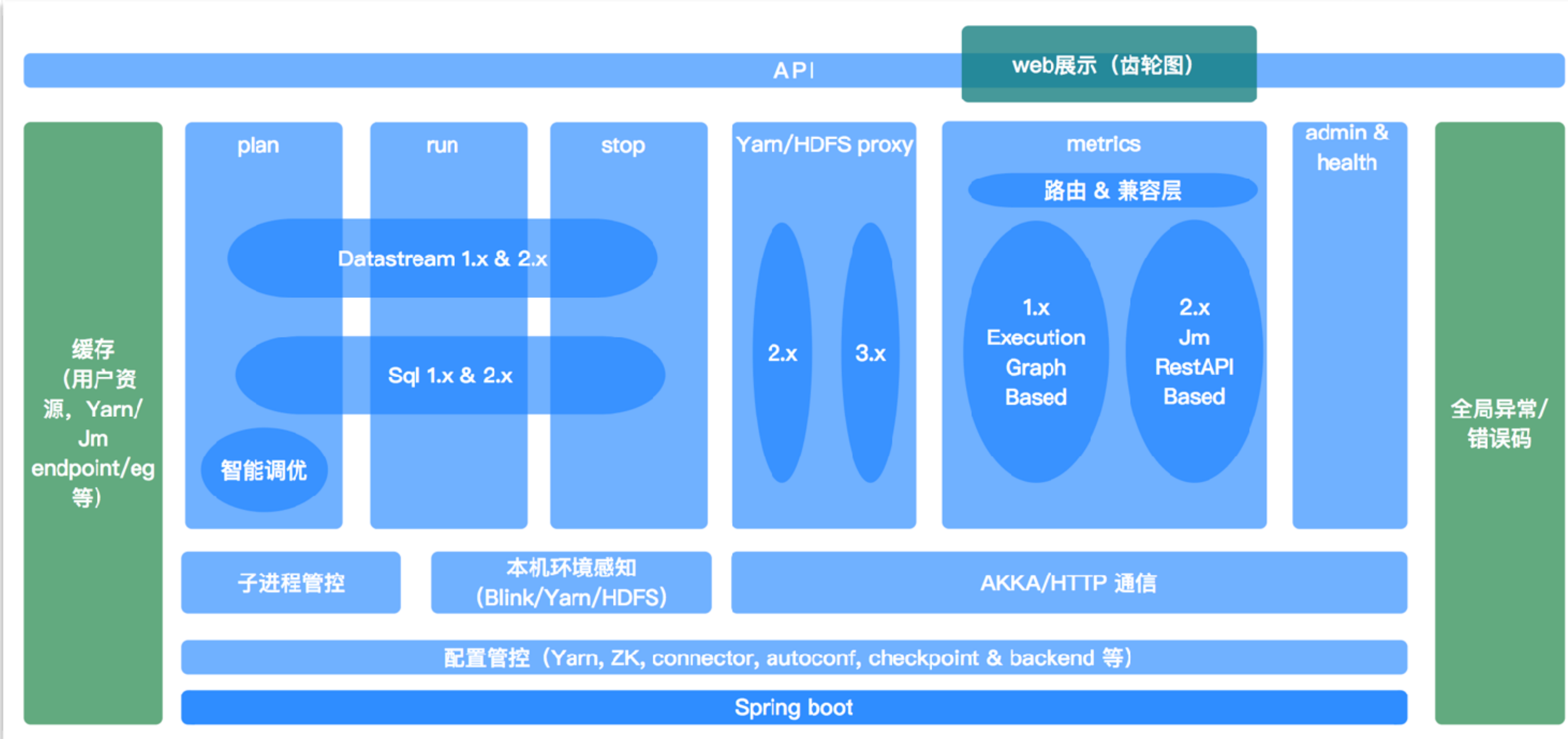
任务DEBUG







- Flink任务的启停
- Yarn的管理
- 智能调优
- Jar包等资源缓存
- 集群状态监控





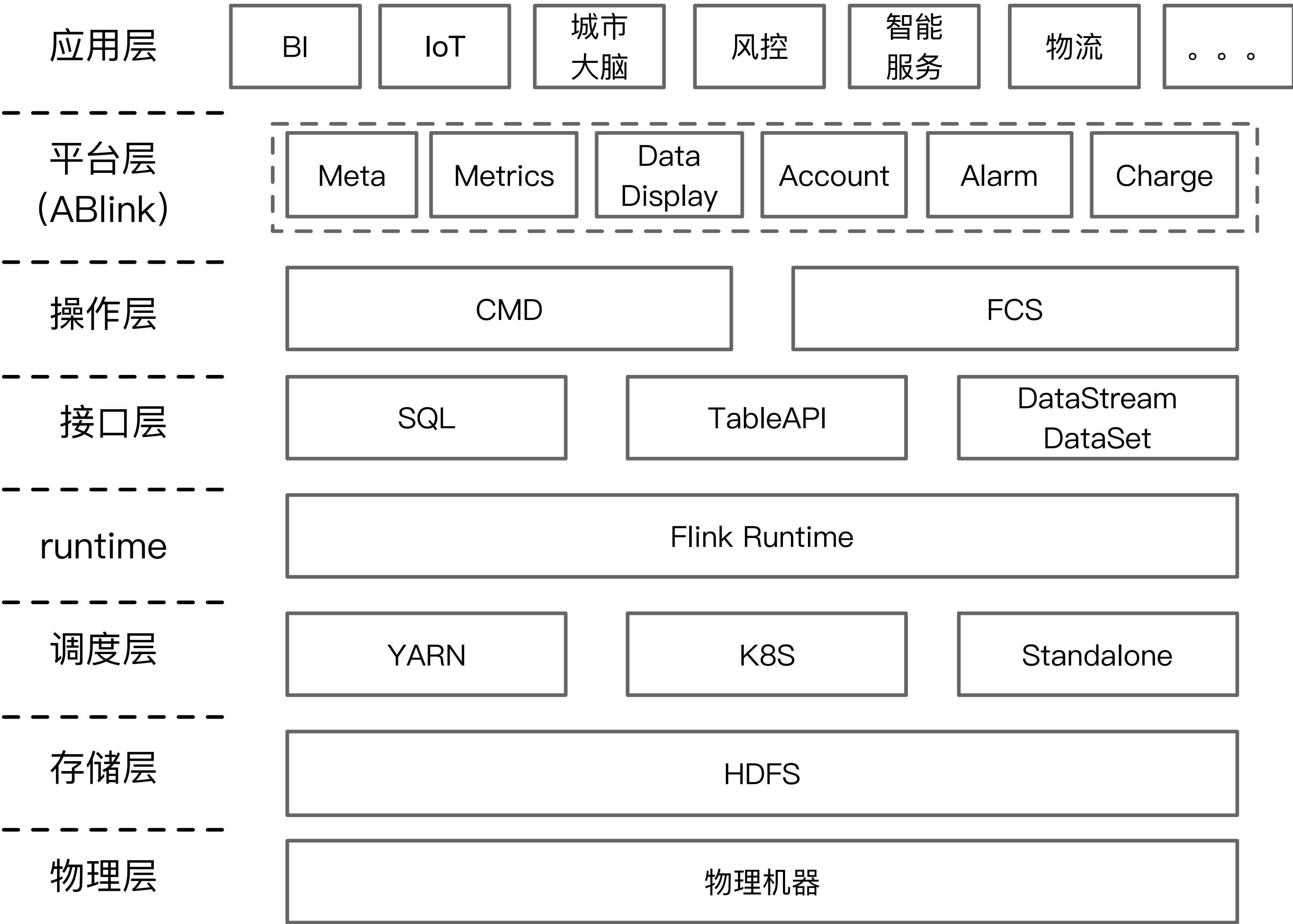
Four

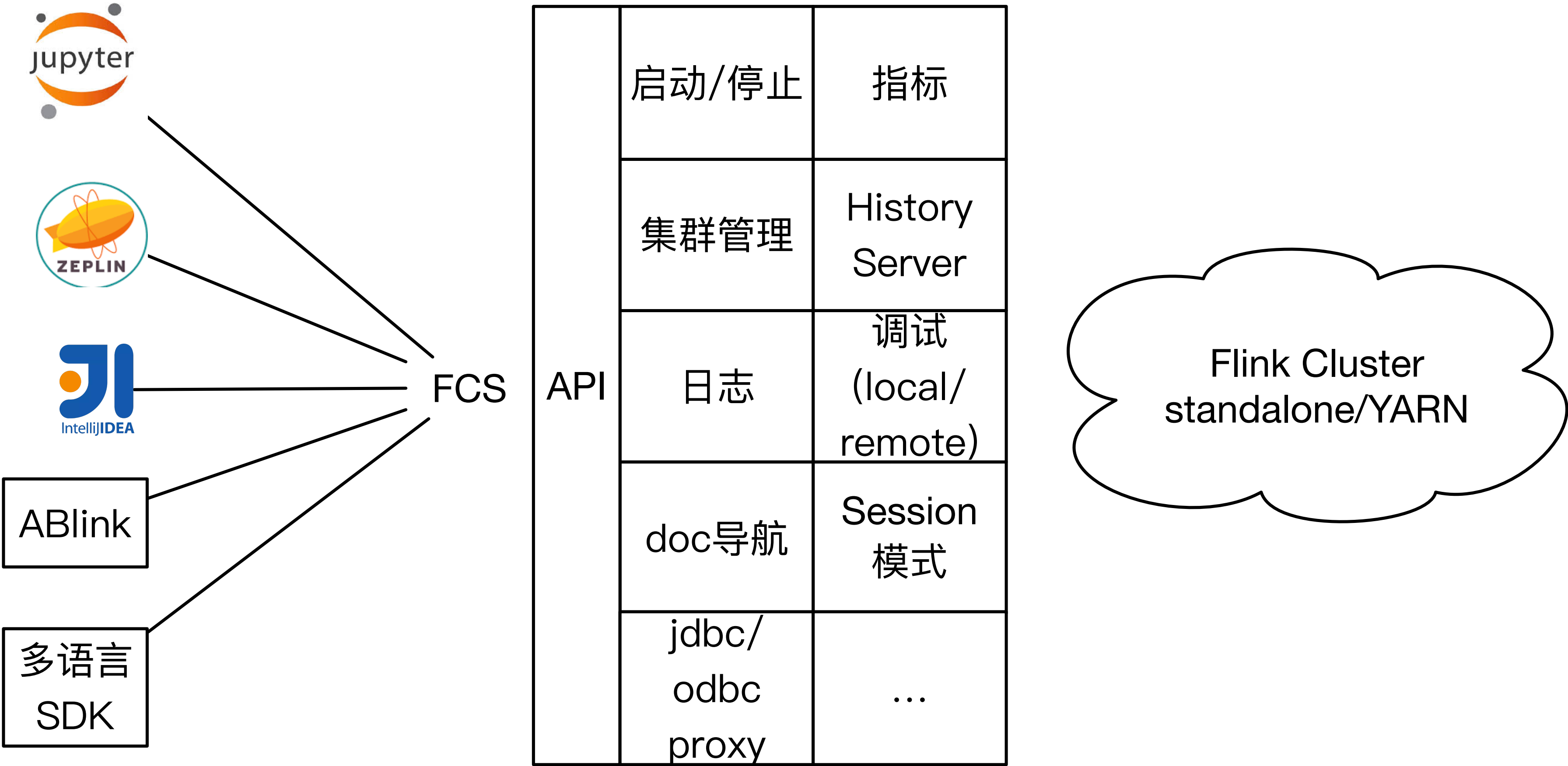
发展规划





ABLINK平台化（内部）





Thanks



Flink China

IT大咖说
知识共享平台