

Apache Doris 助力中科软打造保险行业 智能风控平台

刘钊毅

中科软科技 资深数据架构师

目录

01 智能风控平台

02 数据架构演进

03 实践场景

04 总结与展望

01

智能风控平台

「中科软简介」

27+ 年行业积累

- 公司成立时间：1996.5.30
- 注册资本：59,360万
- 国有控股：中科院软件研究所、海淀国有资产投资经营公司
- 2006年1月23日在深圳新三板上市
- 2019年9月9日上交所上市，股票代码603927
- 2023年实现营业收入约**65.03亿**，实现净利润**6.55亿**，同比增长**2.59%**

18000+ 行业应用专业化人才

- 7000+专注寿险、6000+专注财险、5000+专注公共卫生医疗及政府等行业；
- 涵盖众多行业的应用软件产品、开发服务、系统集成等；
- 公司拥有18000+人才，技术人员占比89%，82%+本科以上学历；
- 研发能力：与中科院软件所成立“行业信息化联合实验室”、“人工智能实验室”、“区块链实验室”；
- 在保险软件厂商中综合竞争力排名**第一**，并保持连续十年排名**第一**；
- 国家规划布局高新技术企业，国家鼓励的重点软件企业、软件百强企业。

「中科软简介 - 市场占有率」

自2011年起在IT解决方案市场竞争分析中连续数年占据领导者地位

| 表1 2023年中国保险行业IT解决方案市场主要厂商收入、排名及市场占有率 | | | |
|---------------------------------------|---------------------|---------|--------|
| 排名 | 厂商名称 | 收入（百万元） | 市场占有率 |
| 1 | 中科软科技股份有限公司 | 5371.0 | 41.07% |
| 2 | 新华保险信息技术中心 | 660.9 | 5.05% |
| 3 | 中国人寿保险股份有限公司信息技术中心 | 310.0 | 2.37% |
| 4 | 上海浦东发展银行股份有限公司信息技术部 | 302.0 | 2.31% |
| 5 | 上海农村商业银行股份有限公司信息技术部 | 185.4 | 1.42% |
| 6 | 中国工商银行股份有限公司信息技术部 | 180.0 | 1.38% |
| 7 | 中国建设银行股份有限公司信息技术部 | 158.0 | 1.21% |
| 8 | 交通银行股份有限公司信息技术部 | 154.0 | 1.18% |
| 9 | 中国银行股份有限公司信息技术部 | 145.0 | 1.11% |
| 10 | 中信银行股份有限公司信息技术部 | 95.0 | 0.73% |

数据来源：赛迪顾问 2024.08

「保险行业的知识体系框架 - 风险管理域」



「保险欺诈给各国保险行业造成损失估计」

赔付金额
10-30%

汽车保险



案件数量的
20%左右

赔付金额
10-40%

健康保险



案件数量的
10%左右

赔付金额
10-20%

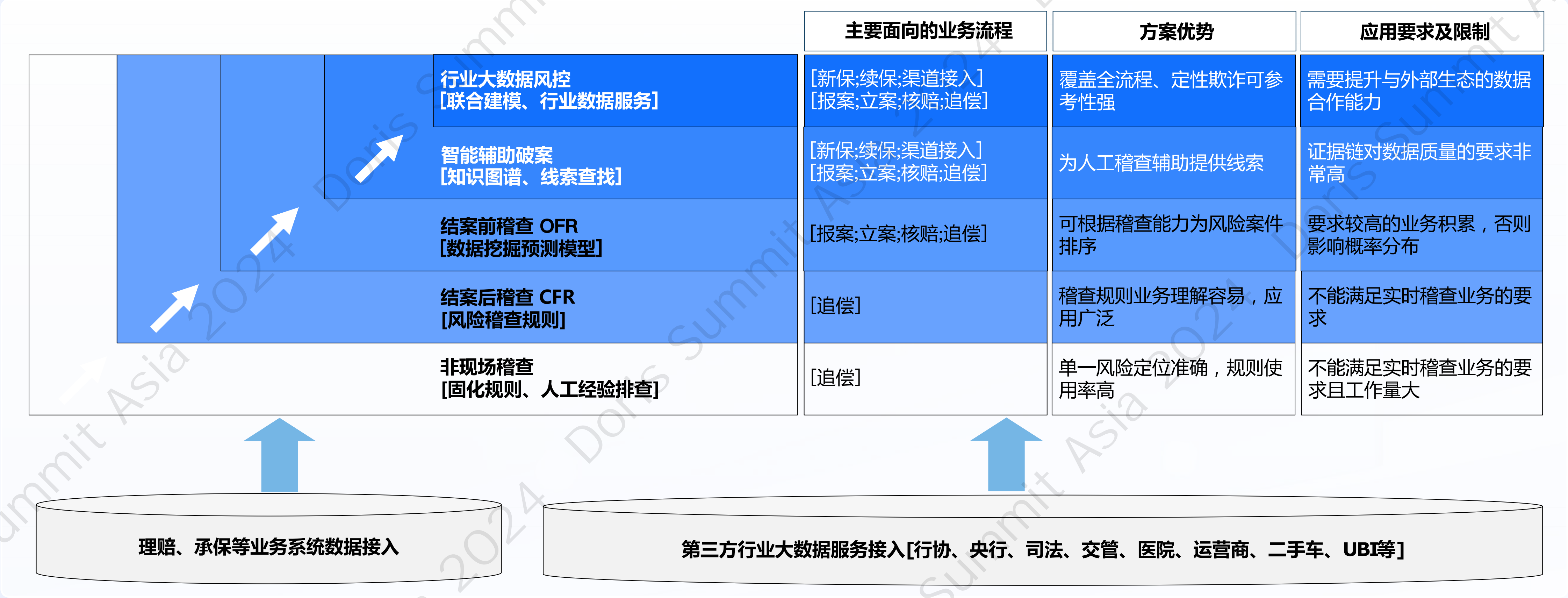
农业保险



案件数量的
10%左右

- 保险欺诈行为造成的损失是非常惊人的;
- 近年来通过强化组织并不断提升技术含量危害日益加剧 ;

「保险风控业务方案的发展」



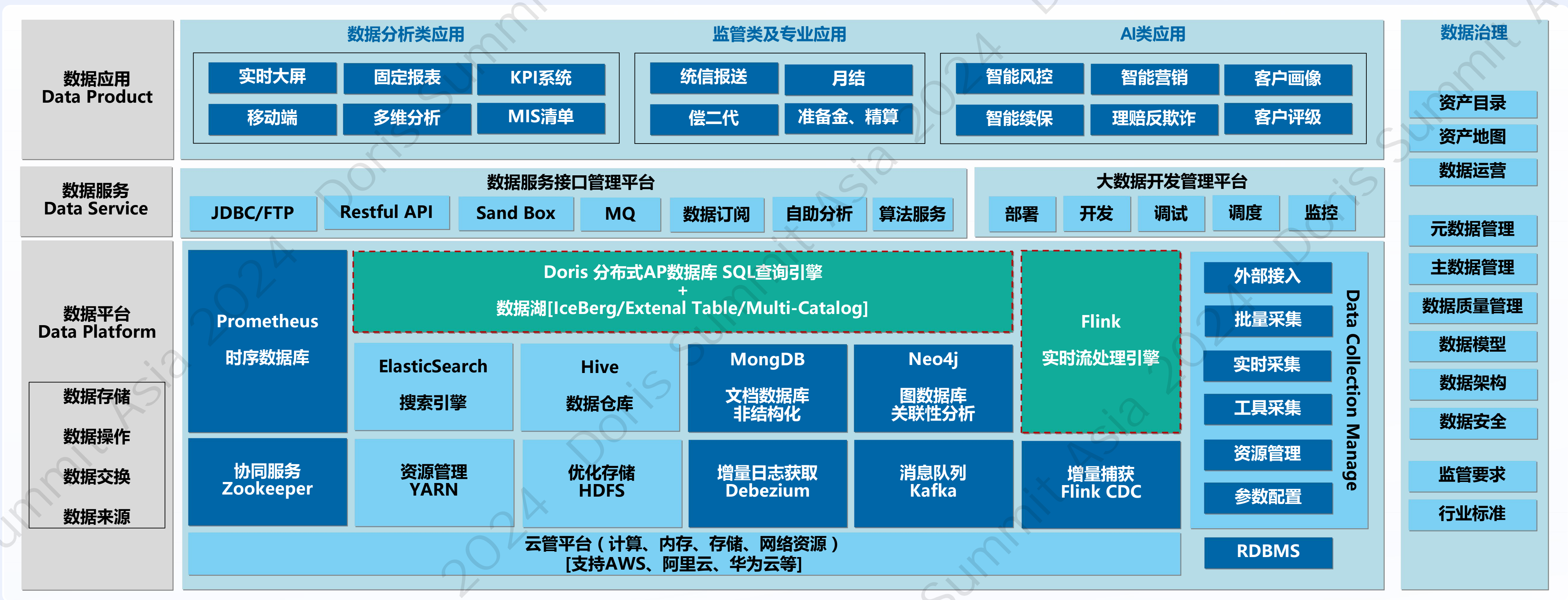
01

技术架构演进

「原数据架构-Hadoop生态」



「升级后的数据架构-Doris生态」



「大数据技术开发管理平台的建立」

用户需求

| | | | | | | | |
|---|---|---|---|---|---|---|--|
|  <ul style="list-style-type: none">● 集群部署管理● 组件服务管理 |  <ul style="list-style-type: none">● 数据API构建● ETL job构建 |  <ul style="list-style-type: none">● 代码调试● 任务管理 |  <ul style="list-style-type: none">● 调度流程监控● 可视化自定义 |  <ul style="list-style-type: none">● 系统运行监控● 消息队列监控 |  <ul style="list-style-type: none">● 增量数据捕获● 多数据源管理 |  <ul style="list-style-type: none">● 流批一体处理● 流式入湖入仓 |  <ul style="list-style-type: none">● 高性能查询● 强兼容mysql |
|---|---|---|---|---|---|---|--|

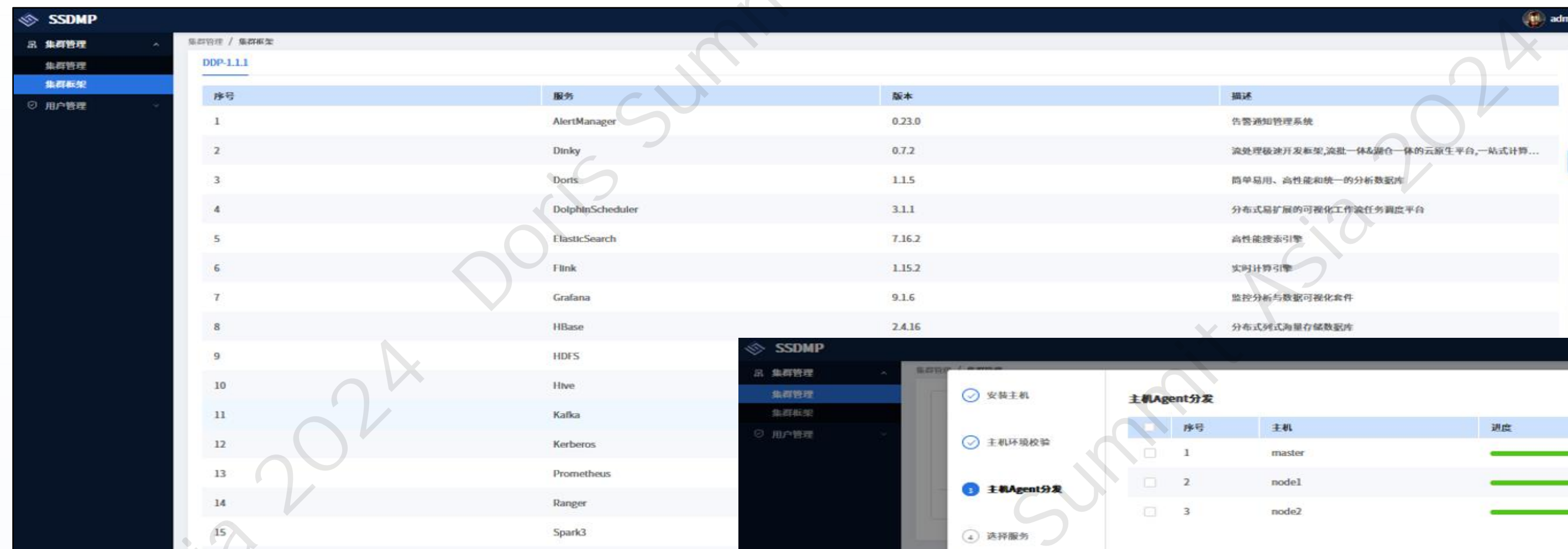
产品功能

| | | | | | | | |
|----------------------------|---------------------|------------------------|---------------------|-------------------|--------------------------------------|-------------------|-------------------|
| 中科软大数据技术开发管理平台) | | | | | | | |
| 组件托管平台 Platform Manager | 快速开发平台 Developer | 任务管理平台 Task Manager | 智能调度平台 Scheduler | 运维监控平台 Monitor | 实时数据获取 Flink-CDC【独占】 Kafka【多播】 | 流批一体处理引擎 Flink | 分布式AP数据库 Doris |

03

实践场景

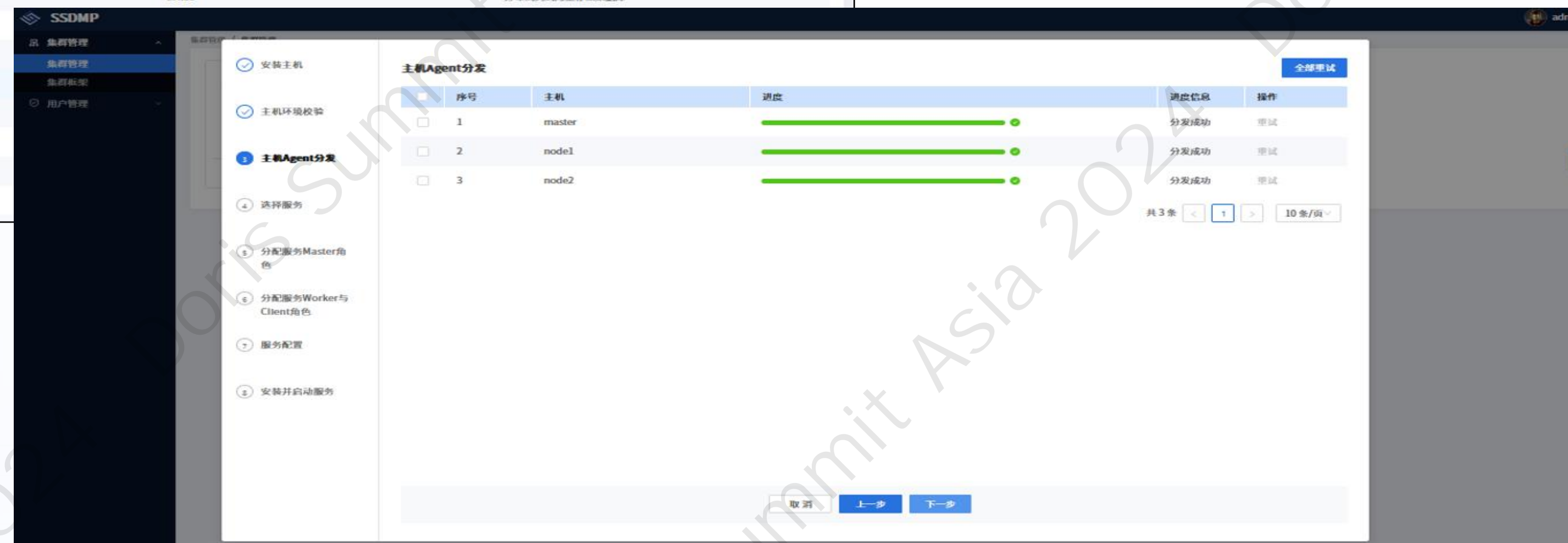
「技术组件托管能力」



The screenshot shows the SSDMP cluster management interface. On the left is a sidebar with navigation options: 集群管理 (Cluster Management), 集群框架 (Cluster Framework), and 用户管理 (User Management). The main area displays a table of services for a cluster named 'DDP-1.1.1'.

| 序号 | 服务 | 版本 | 描述 |
|----|------------------|--------|------------------------------------|
| 1 | AlertManager | 0.23.0 | 告警通知管理系统 |
| 2 | Dinky | 0.7.2 | 流处理敏捷开发框架,流批一体&融合一体的云原生平台,一站式计算... |
| 3 | Doris | 1.1.5 | 简单易用、高性能和统一的分析数配库 |
| 4 | DolphinScheduler | 3.1.1 | 分布式易扩展的可视化工作流任务调度平台 |
| 5 | ElasticSearch | 7.16.2 | 高性能搜索引擎 |
| 6 | Flink | 1.15.2 | 实时计算引擎 |
| 7 | Grafana | 9.1.6 | 监控分析与数据可视化套件 |
| 8 | HBase | 2.4.16 | 分布式列式海量存储数据库 |
| 9 | HDFS | | |
| 10 | Hive | | |
| 11 | Kafka | | |
| 12 | Kerberos | | |
| 13 | Prometheus | | |
| 14 | Ranger | | |
| 15 | Spark3 | | |

1、安装包集成了常用的30个大数据组件，安装时可以自行选择需要安装服务的组件。



The screenshot shows the SSDMP installation wizard. The left sidebar lists steps: 安装主机 (Install Host), 主机环境校验 (Host Environment Check), 主机Agent分发 (Host Agent Distribution), 选择服务 (Select Services), 分配服务Master角色 (Assign Service Master Role), 分配服务Worker与Client角色 (Assign Service Worker and Client Role), 服务配置 (Service Configuration), and 安装并启动服务 (Install and Start Services). The main area is titled '主机Agent分发' and shows a table of hosts with progress bars and status.

| 序号 | 主机 | 进度 | 进度信息 | 操作 |
|----|--------|-------------|------|----|
| 1 | master | <div></div> | 分发成功 | 重试 |
| 2 | node1 | <div></div> | 分发成功 | 重试 |
| 3 | node2 | <div></div> | 分发成功 | 重试 |

共 3 条 < 1 > 10 条/页

取消 上一步 下一步

「平台快速开发能力-数据库」

1、支持常用的**各类数据库、hadoop系统**，可以通过配置数据源新增对于目标数据库系统的支持。

2、**提供友好的代码调试界面**，提供关键字高亮，代码提醒，代码片段保存等功能。

The screenshot displays the Doris SQL development platform interface. On the left, a sidebar lists various data sources including 'doris', 'mysql', 'dlink', 'dolphinscheduler', 'know_streaming', 'metabase', 'performance_schema', 'sys', and 'test'. The main area features a query editor with a SQL query that includes a union of three SELECT statements. Below the editor, a results table shows the output of the query. The table has two columns: 'ods' and 'count(*)'. The results are as follows:

| | 'ods' ↑ | count(*) ↑ |
|---|---------|------------|
| 1 | ods | 201 |
| 2 | dwd | 100 |
| 3 | dm | 100 |

At the bottom of the interface, there is a status bar indicating '3行-60ms(24ms执行)' and a pagination control showing '50条/页' and '1' page.

3、**基于Web的开发管理平台**，提供高效、所见即所见的SQL调试界面。

「平台快速开发能力-ETL」

4、支持对指定Catalog元数据的查询、SQL生成、可用Function查询等功能。

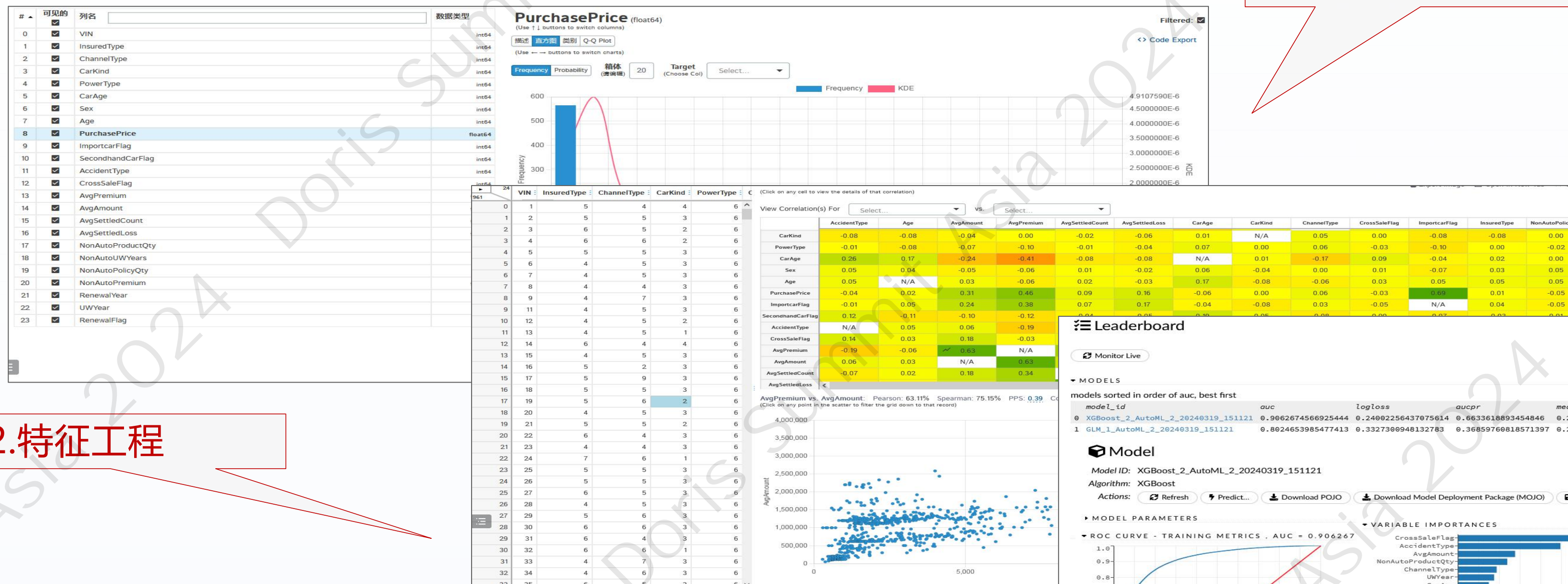
5、支持对执行模式、集群选择、默认Catalog、任务并发度等参数的调整。

The screenshot displays the Doris ETL development interface. On the left, a catalog tree shows the structure of the 'my_catalog.default_database', including tables like 'dm_lclaim_agri' and 'source_lclaim_agri', and functions like 'AGG_DECIMAL_PLUS'. The central panel is a SQL editor with a dark theme, showing a SQL script for creating a table 'source_lclaim_agri' with various columns and a primary key. The right panel contains configuration options for the task, such as '执行模式' (Execution Mode) set to 'Standalone', 'Flink集群' (Flink Cluster) set to 'standalone', and 'FlinkSQL 环境' (FlinkSQL Environment) set to 'DefaultCatalog'. The bottom panel shows the task execution history, listing tasks with their IDs, names, execution modes, and status (e.g., 'SUCCESS').

6、支持对任务执行历史的概要日志的查询，其中包括开始时间、结束时间、是否成功等信息。

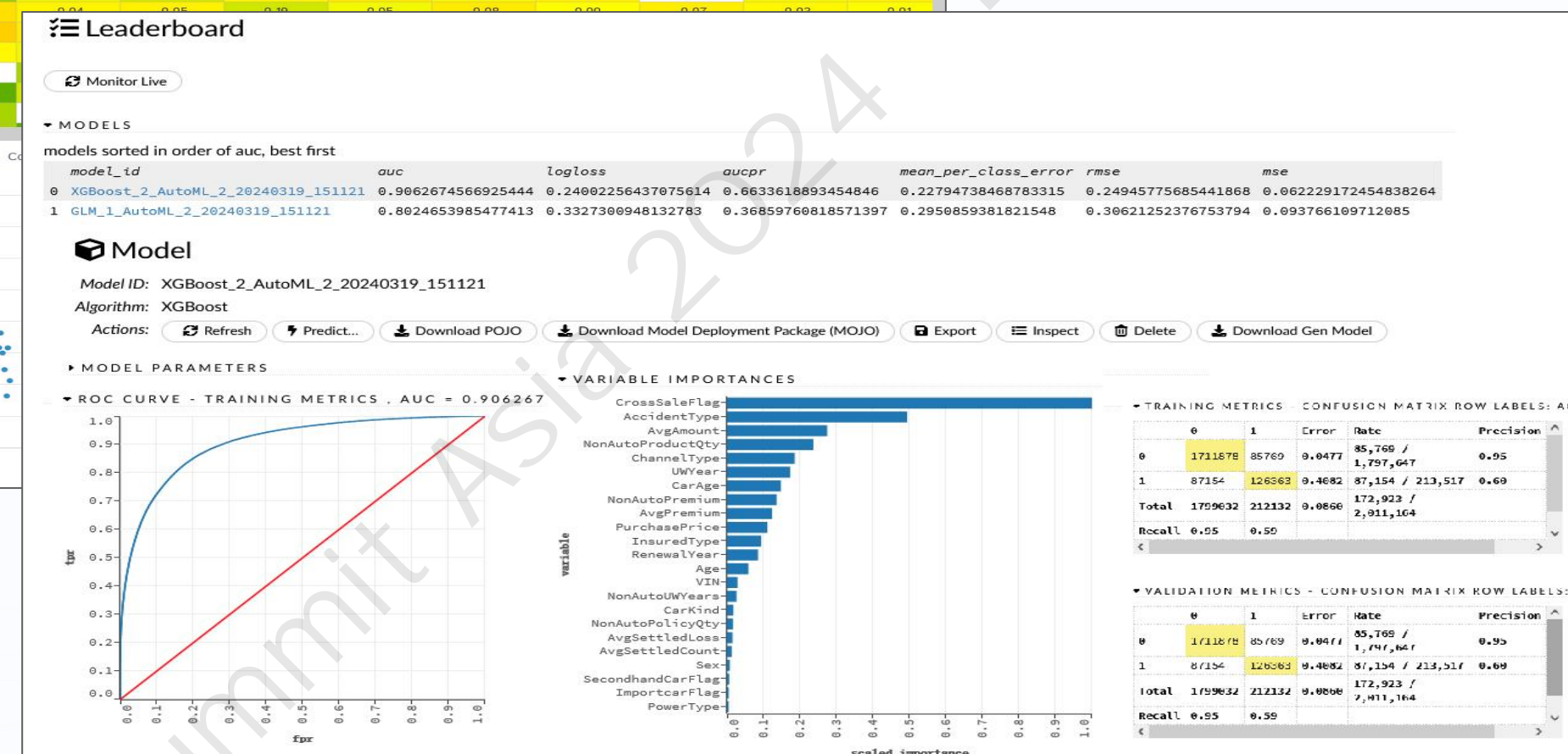
「平台快速开发能力-机器学习」

1.EDA(数据探索性分析)



2.特征工程

3.模型自动化调优



「平台智能调度能力」

1、 workflows 定义，支持各种执行对象，工作流串并行逻辑编排等能力。

2、 workflows 调度，支持各种任务执行方式，定义灵活的定时任务等能力。

The screenshot displays the DolphinScheduler web interface, divided into two main sections: workflow definition and workflow scheduling.

Workflow Definition (Left Panel): The left sidebar shows the navigation menu with options like '项目概览', '工作流', '任务', and '任务实例'. The main area shows the 'wf_zeppelin_test' workflow definition. It includes a list of supported execution engines (SHELL, SUB_PROCESS, PROCEDURE, SQL, SPARK, FLINK, MapReduce, PYTHON, DEPENDENT, HTTP, DataX, PIGEON, SQOOP, CONDITIONS) and a visual graph showing the workflow structure with nodes like 'src-ods', 'ods-dwd', and 'dwd-dm'.

Workflow Scheduling (Right Panel): The right sidebar shows the '智能调度' (Intelligent Scheduling) section. It includes a search bar and a table of workflow instances. A modal window is open, showing the scheduling configuration for a specific workflow instance. The modal includes options for frequency (e.g., '每天', '每隔 2 天'), start time, and end time.

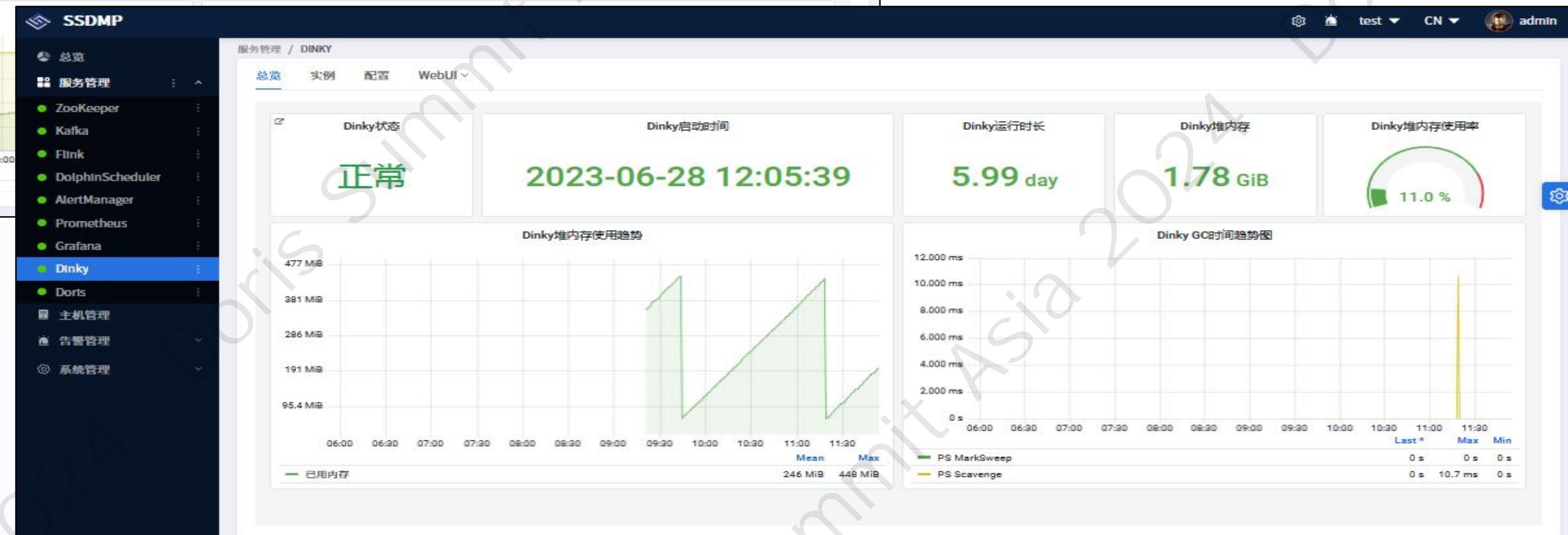
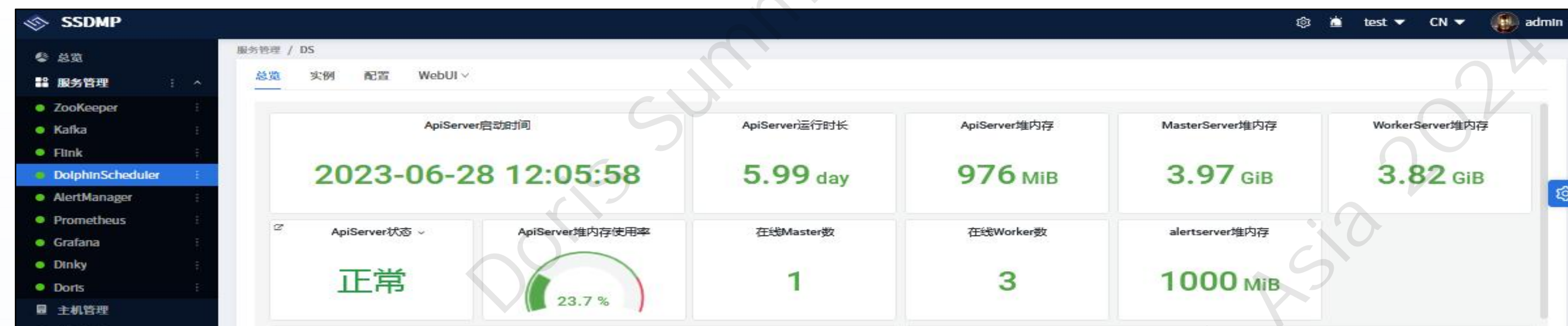
Task Instance Table (Bottom): A table at the bottom lists task instances. The table has columns for task name, workflow instance, execution user, node type, status, submission time, start time, and end time. The first row shows a task instance 'wf_mysql_2_sr' with status '失败' (Failed).

| # | 任务名称 | 工作流实例 | 执行用户 | 节点类型 | 状态 | 提交时间 | 开始时间 | 结束时间 | 操作 |
|---|---------------|-----------------------------|-------|---------|----|---------------------|---------------------|---------------------|------|
| 1 | wf_mysql_2_sr | wf_mysql_2_sr-1-20220915... | admin | ZEPELIN | 失败 | 2022-09-16 02:31:14 | 2022-09-16 02:31:14 | 2022-09-16 02:33:25 | 查看日志 |

3、 job 运行详细日志查看，支持问题快速定位等能力。

「平台运维监控能力」

4、平台集成了各功能组件的性能指标监控能力，例如：Doris、Flink、DS 等。

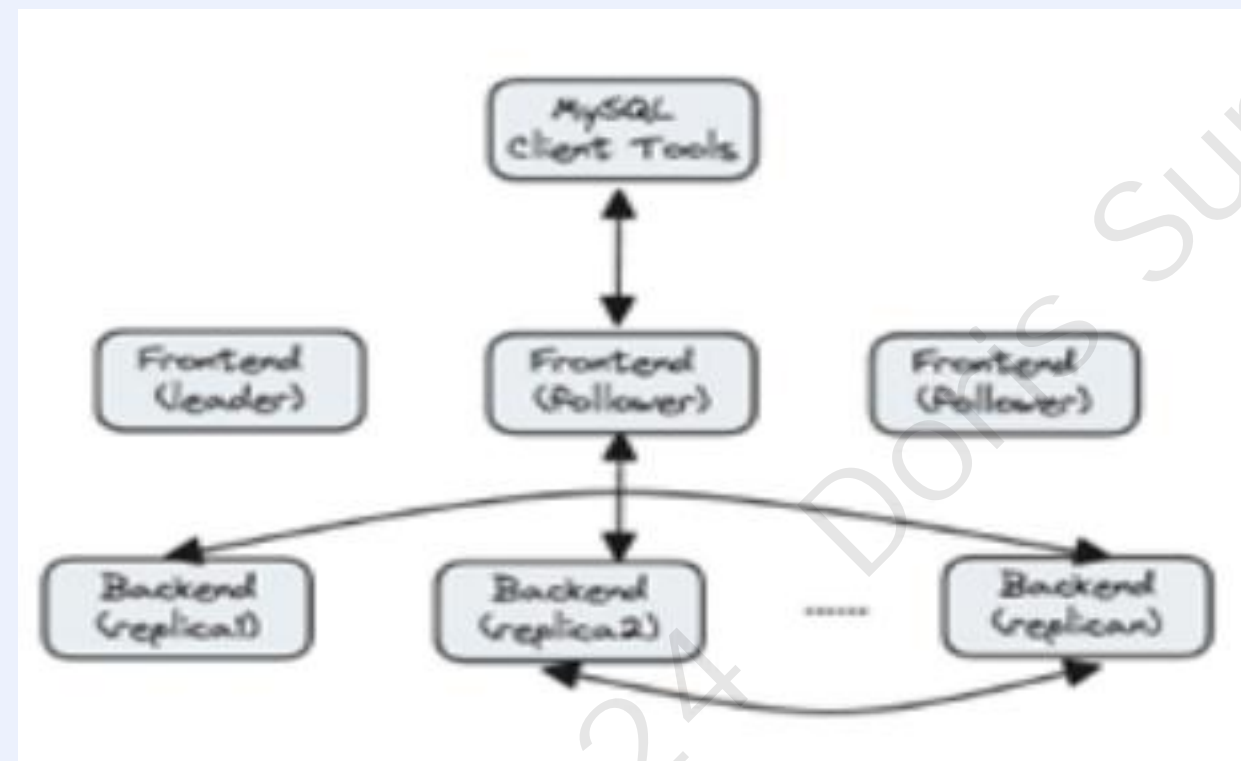


04

总结展望

「对服务高可用、数据高可靠的保障能力提升」

服务高可用



FE和BE的多节点部署，宕机容错

- FE 分为 Leader, Follower 和 Observer 三种角色
- 默认一个集群，只能有一个 Leader，可以有多个 Follower 和 Observer。
- Leader 和 Follower 组成一个 Paxos 选举组，如果 Leader 宕机，则剩下的 Follower 会自动选出新的 Leader，保证写入高可用。
- Observer 同步 Leader 的数据，但是不参加选举。如果只部署一个 FE，则 FE 默认就是 Leader。
- 第一个启动的 FE 自动成为 Leader，上面输出可以看到 IsMaster: true。在此基础上，可以添加若干 Follower 和 Observer。

数据高可靠

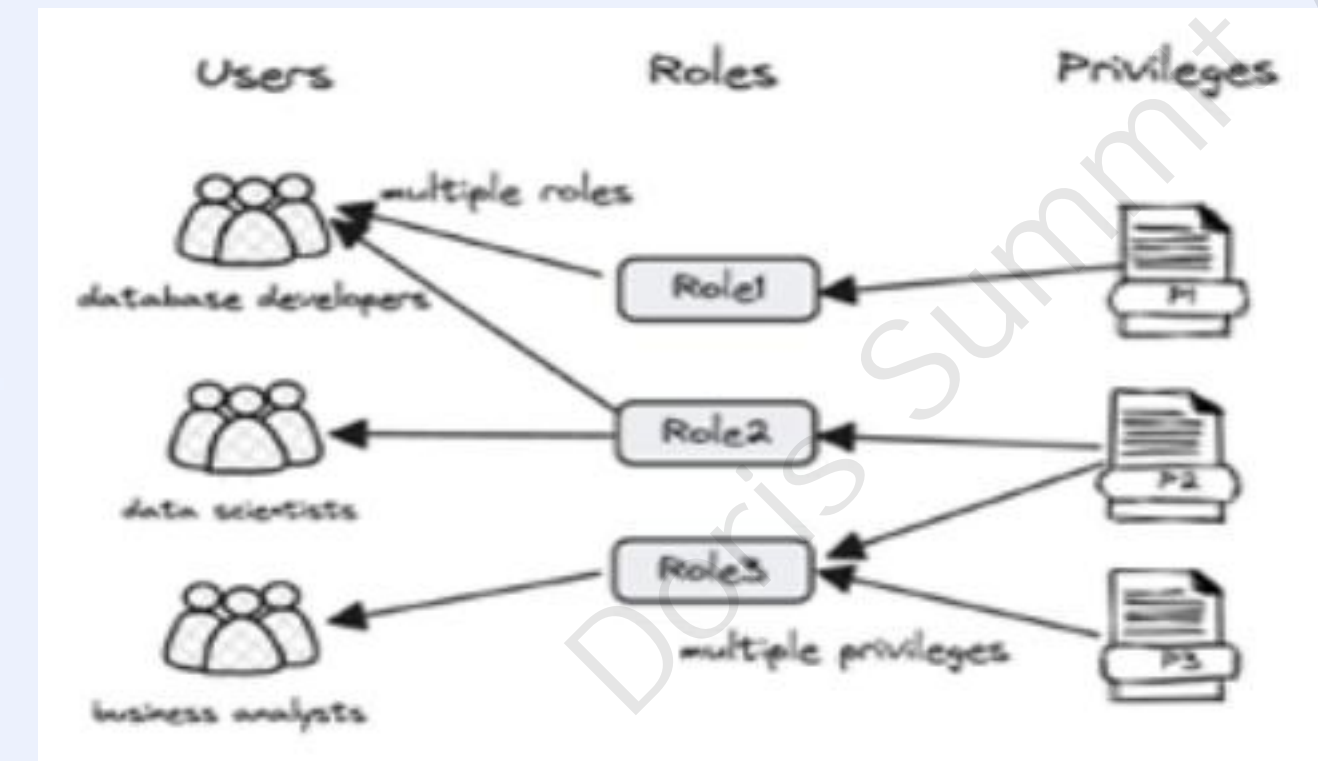
```
BACKUP SNAPSHOT example_db.snapshot_label2
TO example_repo
ON (
  example_tbl PARTITION (p1,p2),
  example_tbl2);

RESTORE SNAPSHOT example_db.snapshot_1
FROM "example_repo"
ON ( "backup_tbl" )
PROPERTIES (
  "backup_timestamp"="2022-04-08-15-52-29");
```

元数据、数据多副本存储，宕机数据自动修复

- 备份/恢复**：支持全表或分区表的备份；
- 备份方式**：mysqldump备份、Hadoop备份等
- 数据恢复**：Doris 支持将当前数据以文件的形式，通过 broker 备份到远端存储系统中。之后可以通过 恢复 命令，从远端存储系统中将数据恢复到任意 Doris 集群。通过这个功能，Doris 可以支持将数据定期的快照备份，也可以在不同集群间进行数据迁移。

系统强安全



- 完善的 RBAC 访问控制**
- 数据细粒度权限**：库、表、行、列
- 支持 SSL、TLS 安全连接和数据传输**

Thanks for Watching !