Apache Flink 与 Apache Hive 的集成

Integrating Apache Flink with Apache Hive

李锐
阿里巴巴技术专家

王刚
阿里巴巴高级开发工程师

FLINK FORWARD # ASIA
实时即未来 # Real-time Is The Future

FLINK FORWARD

目录
Contents

设计架构

Design & Architecture

01

背景

Background

提升Flink批处理能力，完善生态
Enable Flink to run batch jobs and enrich the Flink ecosystem

SQL是大数据领域中重要的分析工具
SQL is the most prevalent tool for big data analysis

FlinkSQL的不足:
FlinkSQL is incomplete and misses important features:

- 缺少完善的元数据管理系统
  Doesn't have a metadata management mechanism
- 缺少DDL的支持
  Lack of DDL support
- 无法方便的对接外部系统
  Unable to integrate with external systems conveniently

目标

Goals

统一的Catalog接口
Unified Catalog API

提供基于内存的和可持久和的Catalog实现
Provide both in-memory and persistent catalog implementations

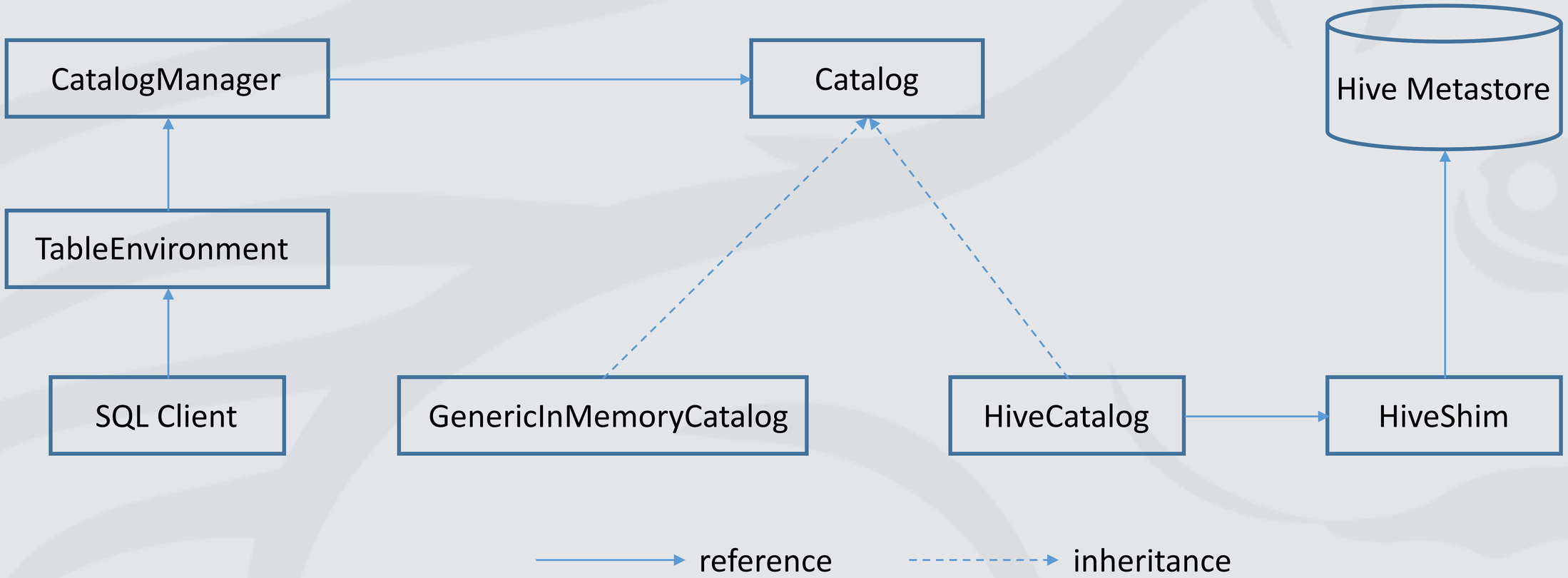
支持与Hive的互操作:
Interoperability with Hive:

- 访问Hive元数据
  Access to Hive metadata
- 读写Hive表
  Read/write Hive tables

支持Flink作为Hive的引擎（长期目标）
Add Flink as a Hive execution engine (long-term plan)

全新设计的 Catalog API （FLIP-30）

New Catalog API (FLIP-30)



读写 Hive 数据

Read/Write Hive Tables

兼容Hive数据格式
Compatibility with Hive data structure

Flink写入的数据Hive可以正常读取，且反之亦然
Data written by Flink can be consumed by Hive, and vice versa

复用Hive原有的Input/Output Format、SerDe等
Re-use Hive's native Input/Output Format and SerDe

- 减少代码冗余
  Reduce duplicated code
- 尽可能的保持兼容性
  Provide best compatibility

读Hive表：HiveTableSource、HiveTableInputFormat
Read Hive tables: HiveTableSource, HiveTableInputFormat

写Hive表：HiveTableSink、HiveTableOutputFormat
Write Hive tables: HiveTableSink, HiveTableOutputFormat

项目进展

Project Status

02

Flink 1.9.0 中的现状

Where're We in Flink 1.9.0

在Flink 1.9.0中首次发布
First released in Flink 1.9.0

在1.9.0中作为试用功能
Released as a preview feature in 1.9.0

1.9.0中的功能缺失较多:
Lots of limitations in 1.9.0:

- 较多不支持的数据类型
  Unsupported data types
- 不完善的分区表支持
  Incomplete support for partitioned table
- 不支持INSERT OVERWRITE
  INSERT OVERWRITE is not supported

Flink 1.10.0 中的新特性

What's New in Flink 1.10.0

支持静态与动态写分区表
Support both static and dynamic partitioning

支持表级别和分区级别的INSERT OVERWRITE
Support INSERT OVERWRITE at table and partition levels

支持更多数据类型，如CHAR、VARCHAR、DECIMAL、TIMESTAMP等
More data types are supported including CHAR, VARCHAR, DECIMAL, TIMESTAMP, etc.

支持更多的DDL
More DDLs are supported

支持在Flink中调用Hive的内置函数
Call Hive built-in functions via Flink

支持更多的Hive版本
More Hive versions are supported

性能优化，如Project/Predicate Pushdown、向量化的读取ORC数据等
Performance optimizations including project/predicate pushdown, vectorized reader for ORC tables, etc.

Module 接口

The Module Interface

Flink 1.10.0中引入了Module接口
Module interface is introduced in Flink 1.10.0

用户可以通过Module方便的接入外部系统的内置函数
With Module interface, users can conveniently plug in built-in functions in external systems

用户可以通过Table API和SQL的方式来动态地加载和卸载Module
Users can dynamically load or unload Module instances via Table API or SQL

可以同时加载多个Module，Flink在解析函数调用时，会根据加载顺序在多个Module中查找函数定义
Multiple modules can co-exist in a session. During function resolution, Flink looks for the function definition according to the order in which the modules are loaded

目前有两个Module的实现：
Currently there're two implementations of the Module interface:

- CoreModule提供了Flink原生的内置函数
  CoreModule provides Flink's native built-in functions
- HiveModule提供了Hive的内置函数
  HiveModule provides Hive's built-in functions

未来工作

Future Work

支持View
Support for views

改善SQL CLI易用性
Make SQL CLI easier to use

支持所有Hive常用的DDL
Support all DDLs that are commonly used in Hive

兼容Hive语法
Compatibility with HiveQL

支持SQL CLI远程模式
Gateway mode for SQL CLI

性能测试

Performance Benchmark

03

测试环境

Setup

21个节点的物理集群
A cluster of 21 nodes

节点硬件配置
Node specifications

- Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50GHz
  Intel(R) Xeon(R) CPU E5-2682 v4 @ 2.50GHz
- 256GB内存
  256GB RAM
- 网络端口聚合，速度20000Mb/s
  NIC bonding. Speed: 20000Mb/s
- 12块HDD硬盘，每块5.5TB容量
  12 HDD each with 5.5TB capacity

测试环境

Setup

测试工具：hive-testbench
Benchmark tool: hive-testbench

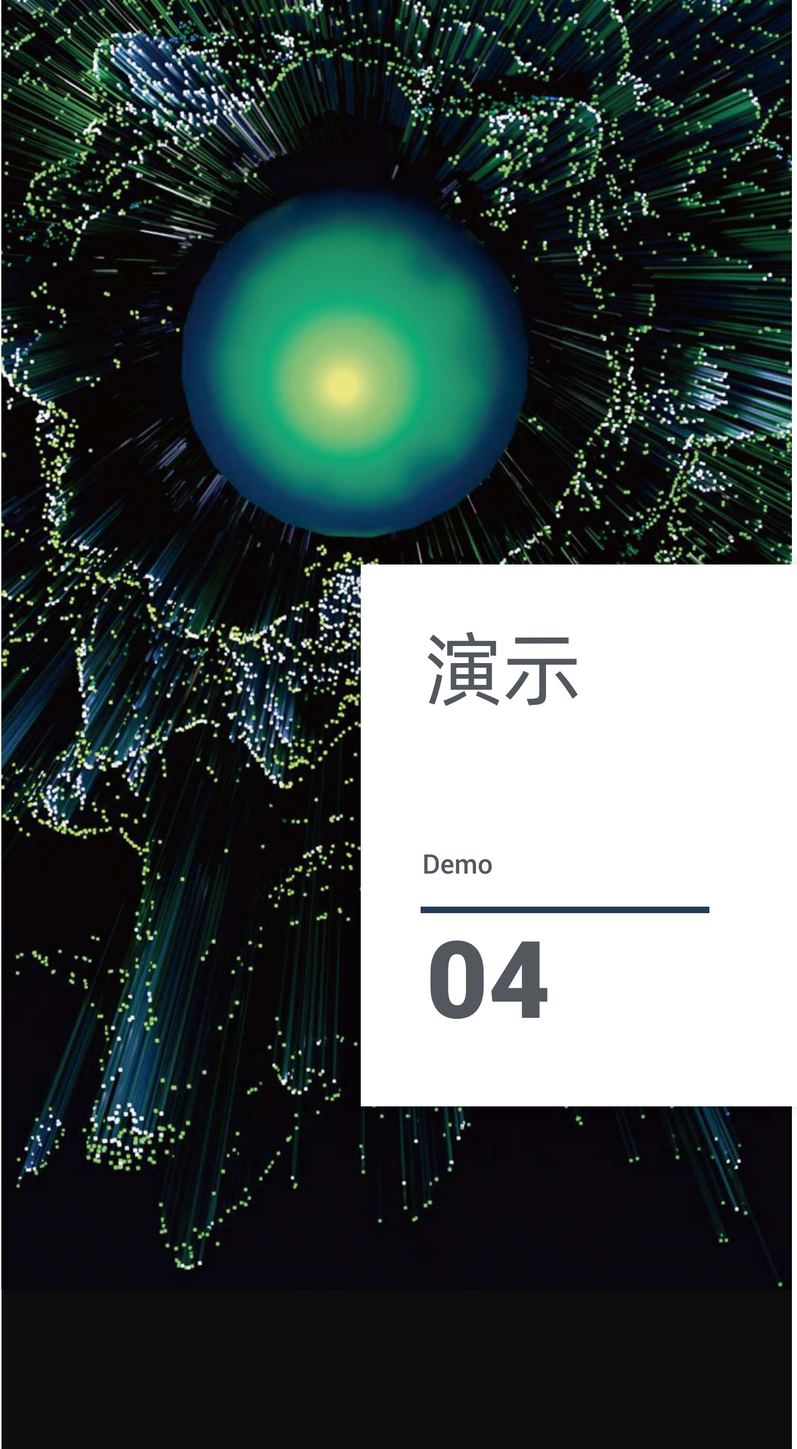
测试数据集：10TB TPC-DS
10TB TPC-DS dataset

测试组件版本
Software component versions

- Hive 3.1.1
  Hive 3.1.1
- Flink master分支
  Flink master

测试结果

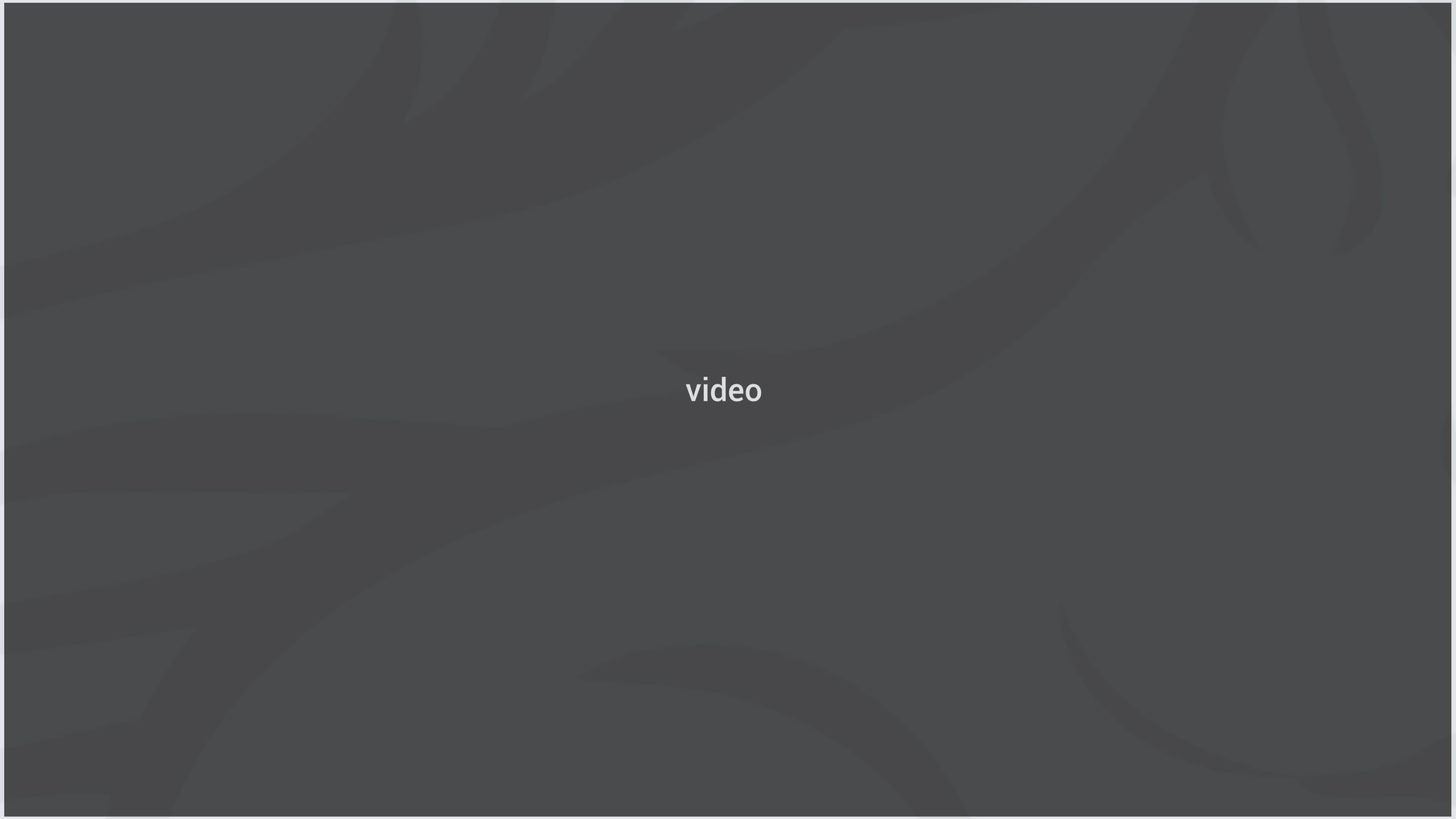

Benchmark Results
Flink vs Hive
TPC-DS 10 TB
~ 7
X
8000
7000
6000
5000
4000
3000
2000
1000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99 101
Flink
Hive on MR



演示

Demo

04


video


THANKS