数元灵湖仓数据智能 LakeInsight 产品白皮书

1. 产品介绍

LakeInsight 是北京数元灵科技有限公司研发推出的一站式湖仓数据智能产品,涵盖实时数据集成、实时数据建模、实时数据服务、指标报表和 AI 开发等功能,一站式服务企业数据智能业务,构建数据智能中台。

LakeInsight 基于数元灵自研开源湖仓一体框架 LakeSoul。LakeSoul 是数元灵自主研发的全新一代湖 仓一体框架,具备云原生、计算存储弹性、流批一体、Data+AI 一体化的设计。LakeSoul 的总体架构 如下:



2. 产品核心功能

2.1 实时数据同步

(1) 支持单库单表、单数据库多表同步,根据数据库 CDC 事件增量同步数据包含的信息,自行解析出数 据库,表名以及表的 schema 信息,然后实现建库建表的操作;

(2) 支持的数据类型: boolean, bit, binary, varbinary, blob, tinyblob, mediumblob, longblob, bigint, int, integer, mediumint, tinyint, smallint, float, double, date, datetime, timestamp, decimal, char, varchar, string, longtext, mediumtext, text, tinytext, json;

(3) 支持自动 Schema 变更同步。根据数据中的 Schema 信息,自动识别 DDL 变更,并同步到实时存 储平台表中。支持的 Schema 变更类型包含增、减列、列数据类型改变操作(int -> long,float -> double),并支持逻辑列删除(即数据本身删除某列后,平台可保留删除列的 schema,这样后续查询可 查到源表删除列字段信息);

(4) 保障数据准确性:实现全链路 Exactly-once,保证数据传输的不丢不重;实现数据延迟机制判断, 避免因上游数据延迟导致相关异常;

(5) 支持多种断点重传机制:在上游数据出现问题等情况下,支持多种重传方式,包括指定时间戳消 费,从最新数据消费等,方便快速恢复同步任务;

(6) 支持数据源数据安全机制。

2.2 实时计算存储

采用湖仓一体的实时数仓架构,兼容开源生态体系,支持多种开源计算引擎、存储引擎,能够进行增量计算、全量计算以及多种分层建模方式的 PB 级实时数仓框架,做到"流批一体、湖仓一体、AI 与BI一体"三位一体,满足数据可追踪、可管理、可查看以及集群可弹性伸缩的能力。具体功能包括: (1) 元数据管理功能,支持高可用、分布式部署方案,能够满足单节点千万级别数据管理要求,支持Domain(域)、Namespace、表、分区、数据文件多层级管理;支持高并发写入,支持 ACID,保证数据读写一致性。支持 TimeTravel,支持回滚、快照、增量读等;支持 Listen-Trigger-Notify 机制触发自动 Compaction 和数据清理支持高可用部署。

(2) 支持数据实时更新(append, upsert方式),数据实时读取方式(MOR,增量,快照):数据实时更 新情况下,可根据主键情况,自适配采用不同数据更新方式。无主键采用 Append 方式,有主键可采 用 Upsert 模式,数据在读取时会按照主键合并更新,保证同一个主键使用最新的数据;需提供 TimeTravel 能力,增量读和快照读。

(3) 支持计算引擎包括不限于批式、流式计算引擎:批式计算引擎包括 Spark等;流式计算引擎支持 Flink、Spark Streaming等;支持AI、MPP 等相关计算引擎: AI 计算引擎支持 PyTorch、Pandas、 Spark MLLib等; MPP 计算引擎支持 Presto、Doris 等;提供 API 接口以支持各类开源引擎的接入和整 合。

(4) 支持流批数仓建模:对接上游数据采集平台后,可使用实时、批量数据等方式写到存储计算平台; 支持实时增量建模,支持以流式的方式增量读取上游数据,读取时支持 Changelog 语义,增量数据, 支持 Flink 增量计算,包括双流 Join、LookupJoin、Aggregate 等,支持 CDC 输出,支持实时落盘到 计算存储平台,支持实时推送到下游数据服务平台;支持批量计算建模,支持以周期调度的方式,批 量执行建模计算任务。结果可以支持覆盖写(Overwrite)以及 Upsert 两种方式进行更新。批量计算 作业支持Spark SQL、Spark DataFrame API 进行相关开发。

2.3 实时数据建模分析

功能包括:基于 WEB-UI 的开发环境,在该平台可进行相关数据建模,任务提交,数据探查工作。

(1) 支持的开发语言:支持通过SQL方式进行相关数据研发工作;支持 Python、Java、Scala 等代码任务;支持数据任务的开发、测试、上线的一站式建模开发服务;

(2) 安全机制:支持企业内部单点登录对接;支持开发环境和生产环境隔离;支持数据域的划分,数据可读可写可执行等权限的隔离;支持各类租户的工作空间权限隔离与划分;支持并提供基于角色的权限管理功能,保证当前工作空间的各项业务及数据安全的要求;

(3)数据建模开发:支持开发人员使用基于 WEB 的在线编辑器,进行数据处理作业设计开发工作;支持开发人员使用交互协作的方式使用Flink、Spark等多种语言进行实时数据建模工作,包括数据清洗、分析、可视化显示,支持报告生成等;支持管理员角色用户,将开发人员在平台中完成的任务以可配置化的方式,发布在生产环境,支持任务实时状态监控、任务的启停以及日志查询等;

(4) 平台管理:支持工作空间管理,不同用户在不同的工作空间下,不同模块有不同的操作权限;基于 角色的数据开发和任务发布权限控制;基于各模块的自定义角色管理功能;支持配置开发集群和生产 集群,支持开发环境和生产环境隔离;支持平台任务管理和监控,支持资源、计算任务监控报警。

2.4 实时指标报表

功能包括:可视化低代码的指标定义,方便业务人员自助定义指标。支持原子、衍生、复合指标。支持可视化大屏面板自定义图表,支持在线编辑可视化图文数据报告。

(1) 指标定义:支持原子指标定义,即指定数据库表和维度列、聚合列、聚合函数,生成指标;支持衍 生指标定义,即任意指标添加过滤条件或聚合;支持复合指标定义,即任意多个指标四则运算;

(2) 图表定义:可以为每个指标,配置图表。支持折线图、柱状图、饼图、表格、数字等多种展示方式。支持将多个指标形成分组,每个分组形成图表面板,组成大屏展示;

(3) 图文报告:支持前端在线编辑图文报告,支持设置字体、段落等格式,支持在报告中选择指标和图 表形式插入,打开报告时自动刷新最新的指标值。

2.5 产品部署和集成

LakeInsight 产品支持部署在 Hadoop 环境、Kubernetes 容器化环境等。支持麒麟、统信、欧拉等国 产操作系统,支持海光、兆芯等国产 CPU 芯片。

LakeInsight 提供数据写入和读取接口。提供 JDBC 标准数据访问连接,可以与各类 BI 软件兼容集成。

3. 使用文档

3.1 安装

3.1.1 创建实例

• 用户可以从【云原生服务中心】进入【我的服务】,点击【创建实例】



• 勾选选择相应的版本,并点击右下角的【**下一步**】。

≡	👋 华为云 🗌 控制台				Q 搜索云服务、	快捷操作、资源、文档、API	备案 資源	「 费用 企」	k I具 If	单 D. Q ¹ (77) ⊕ 简体	hw045342058 hw_jnj_syl
<	创建 Lakeinsight										
		1 版本选择	(2) 基本信息 ————————————————————————————————————	(3) 依赖服务 ————————————————————————————————————	④ 实例参数	(5) 信息确认	- ⑥完成				
											Q
	服务版本	应用版本	创建时间		描述	架构		部署场展			
	O v1.0.0	-	2024/06/27 18:56:50 GMT+08:	00	A Helm chart for LakeInsight.	X86_64,ARM		CCE			
	• v1.0.1	-	2024/06/27 19:14:50 GMT+08:	00	A Helm chart for LakeInsight.	X86_64,ARM		CCE			

• 勾选相应的【**基本信息**】,并点击右下角的【**下一步**】。

< / 创建 Lakeinsight							
	○版本选择 ──	2 基本信息	(3) 依赖服务 ———	(4) 实例参数	(5) 信息确认	(6) 完成	
* 部署平台	云容翻傳 华为云UCS						Lakeinsight
* 区域	华东-上海—/default ~						A Helm chart for LakeInsight.
帮果器容★	cce-lest-cluster-v1 > 目前只支持v1.15及以上版本的集群 例范集群已经安装 v1.0.1 版本	0 彩薄熱機補詳					Service Version v1.0.1 Provider
★ 命名空间	default ~	Q 创建命名空间					 Maintainers
* 实例名称	lakeinsight-27/9ds						dmetasoul jiang@dmetasoul.com 🕐
*服务实体	HelmRelease A Helm chart for LakeInsight.						
描述信息	描述实例的用途						
	0/256	6					
✓ 我已知晓 1. 需要存储资源	酌实例在创建实例时会默认创建按需付费的云存储资源,计费洋情参考	https://www.huaweicloud.com/prici	ng.html?tab=detail 存储资源详情可以很	王云容器引擎->资源管理->存储管理	Éð		上一步 下一步: 灾州参数

• 配置访问方式(Ingress 和 NodePort 两种方式)

选择 Yaml 创建:

< 创建 Lakeinsight							
	○版本选择	── ◇ 基本信息 ──	── ⊘ 依赖服务 ───	④ 实例参数	(5) 信息确认	6) 完成	
表单创建 Yami创建							
1 揭示: 某些字段可能不在此"表单创建"中表示,可以改善。	至"YAML创建"中补充						×
✓ casdoor							
✓ dashboard							
✓ global							
✓ grafana							
∧ ingress							
enabled							
✓ hosts							
∨ tis							
V lakesoul_dashboard							
✓ mysql							

< 1 创建 Lakeinsight

	⊘版本选择 ───── ② 基本信息 ──── ○ 依赖服务 ·	() 实现参数 (5) 信息确认 (6) 完成
表单创建	Yamtitist	
16 17	internal: acceled true	lakeinsight
18	enabled: true	lakeinsight > spec > values > global > nginx
20	hist minin-syr	
21		• enapoint
22	volumesize: 20Gi	关型: string
23	postgresal:	描述:
24	host: postgresql-svc	
25	port: 5432	
26		• Ingress
27		
28		大型: object
29		
30	url: 'http://100.100.200/latest/meta-data/region-id'	· · · · · · · · · · · · · · · · · · ·
31		
32	endpoint: 'http://lakeinsight.mydomain.com'	
33	ingress:	nodePort
34	enabled: ture	業型・ pumber
35	nost: lakeinsight.mydomain.com	
30	niderstange	
38	arcestav: minioadmin	
39	bucket: lakesoul-bucket	
40	endpoint: 'http://minio-syc.CURRENT_NAMESPACE.syc:9000'	
41	secretkey: minioadmin	
42	postgresal:	Register-
43	casdoordb: casdoor	

• 方式一: Ingress配置方式(默认)

如果用户希望通过Ingress独立域名的去访问,需要将ingress开启,为endpoint配置 http://ingress_host,可参考如下配置方式:

2 •	global:
3 *	admin:
4	password: 123456
5 •	nginx:
6	endpoint: http://lakeinsight.mydomain.com
7	#if use nodePort, set endpoint like: http://192.168.18.167:30443
8	nodePort: 30447
9	#if use ingress, set endpoint like: http://lakeinsight.mydomain.com
10 -	ingress:
11	enabled: true
12	host: lakeinsight.mydomain.com
13	
14 -	internal:
15	enabled: true
16 •	minio:
17	host: minio-svc
18	port: 9000
19	volumesize: 8Gi
20 -	postgresql:
21	host: postgresql-svc
22	port: 5432
23	volumesize: 8Gi

• 方式二: NodePort配置方式

如果用户希望通过NodePort方式访问,需要预定义集群中未被占用的nodePort,为endpoint配置 http://NodeIP:NodePoint,可参考如下配置方式:

1	######################################						
	<pre>global:</pre>						
	admin:						
	password: 123456						
	nginx:						
	endpoint: http://192.168.18.167:30447						
	#if use nodePort, set endpoint like: http://192.168.18.167:30443						
	nodePort: 30447						
	#if use ingress, set endpoint like: http://lakeinsight.mydomain.com						
	ingress:						
11	enabled: false						
	host: lakeinsight.mydomain.com						
	internal:						
	enabled: true						
	minio:						
	host: minio-svc						
	port: 9000						
	volumesize: 8Gi						
	postgresql:						
	host: postgresql-svc						
	port: 5432						
	volumesize: 86i						

• 信息确认并完成创建

< 创建 Lakeinsight								
		⊘版本选择	─── ◇ 基本信息 ───	─────────────────────────────────────	─────────────────────────────────────	5 信息确认	6)完成	
基本信息								
部署平台	云容器引擎		区域	华东-上海一				
容器集群	cce-test-cluster-v1		命名空间	lakeinsight				
实例名称	lakeinsight-27i9ds		描述信息					

3.2 注册 Licence

登录首页,如果没有申请过 licence 或 licence 已过期,则会弹出提示窗口,点击【更新license】

license过期提示	×
huawei license invalid	
	更新license

鼠标移动到右下角,会提示向邮箱【hw-support@dmetasoul.com】发送邮件申请 license



将获取到的 licence 粘贴至输入框里,并点击【**验证**】

FD7 18415E61B350E093F55CB299063E314C7EFF76088B5BF08B29DAF103 6932 963008.3421252A216FD5140FF75CD629D40983A38B98B6C714F6D 72049E2836731 エレットを2317&ユーロ 01847 B98A JUCC90C1EDEDB68D60F75D5E26DCFA17764A142D66264 30A1B2A2CA32DE86E648CC4BA1BF376D90EC07B4B2FBR07063334B15 9A6C12058E69EFB8EA66F749CD68F313F71CD7E7F03F19143D433427 4CC27E6 D48C69E775 6A989D33C812B8744DDE808888A7DC 10059575861E068 9B6DB24ACCP071 20105214755A62 01E3		LICENSE	
	*	FD7 18415E61B350E093F55CB299063E314C7EFF76088B5BF08B829DAF03 6932; 38B763 96308A3421252A216FD514DEFF26D080RAD940983A38B98B8CF14F6D 72049E63627; JULE CONTENDED680D6DF75D6E26DCFA17764A1E42D86264 30A1B2A2CA32DE86E648CC4BA1BF376D90EC07R4R2FR0D066334B15 9A6C7FA 98D455B69EF8BEA666F749CD68F313E71CD7E7F093F19143D433427 4CC27E6 4C27E7E6 4C27E7E7 4C27E7E7 4C27E7E7 4C27E7E7 4C27E7E7E7E7E7E7E7E7	
近回 ····································		反 登证	

右上角【关于】可以查看当前 licence 版本。

elnsight		lake-public > 승고야함이 옷 admin ^
下面是您目前使用的版本信息。不同版本	对应特益的权限集,如果忽希望获取更大权限,确升级Licensel	進出登委
^{出机版本} 大企业版	可伯羅工作空闲歌: 天殿制 可伯服歌思聞今雲: 天殿制 可配墨巴中歌: 天殿制 短周服券: 支持 JDBC, HTTP Rest 動詞数据 JDBC信振想地:: lakesoul-dashboard-thrift-binary.lakeinsight.svc:10009 技术武神振券: 7°24 在线、电话技术支持	

3.3 登录

1. 进入 LakeInsight 前,未登录的情况下,先跳转到登录页进行登录认证。使用上述安装时配置的账 号和密码进行登录(我们当前的用户管理,引入了第三方的用户管理系统casdoor)。

D LakeInsight	•
A 用户名、Email或手机号 盘 密码	ø
✓ 下次自动登录 登录	

2. 登录认证通过后,会进入到LakeInsight的首页,根据需求通过切换左侧导航选择内容。(可以通 过右下角logo处 单独打开LakeInsight)

D Lakel	nsight	lake-public ~ 신正作空间 온 admin ~
Ŧ	作业统计 资源监控	
🚽 作业监控	任务总数	开始时间 ③ 选择时间
🔒 数据源管理		结束时间 🕞 2023-12-06 18:34:28
📓 数据入湖		
✔ 发布任务		查看146
数据开发	模块任务状态信息	
🔊 数据出湖	🔲 运行中 🔛 运行完成 🔛 未运行 🛑 运行失败	
▲ 用户管理		

3. 登出

点击**退出登陆**按钮,可以退出当前系统。

4. 关于

点击**关于**按钮,可以查看当前平台的版本信息。

O Lakelı	nsight		lake-public · 신工作空间 _ 스 고作空间
	下面是您目前使用的版本信息。不同版本对应	转定的权限集。如果整希望获取更大权限,请升级License!	退出登录 关于
🖢 作业监控		可创建工作25问数: 1	
局数据源管理	当前版本	可创建宽旗旗小型: 3 可配置用户数: 3	
2≧ 数据入湖	基础版	数据服务: 支持 JDBC、HTTP Rest 查询数据 JDBC连接地址: lakesoul-dashboard-thrift-binary.extension-lakeinsight.svc:10009	
☆ 友布任务 ⑤ 数据开发		技术支持服务: 5*8 在线技术支持	
🔓 数据出湖			
🛆 用户管理			
		済前往 <u>KubeSphere Marketplace</u> 秩J 更新Licons	\{License

3.4 数据源管理

数据源管理界面,可以配置各种类型的数据源,目前包括Mysql、Kafka以及MaxCompute三类数据 源。针对不同的License,创建的数据源实例个数不同。

1. 点击左侧数据源管理,查看当前已有的数据源列表。

D Lakel	Insight		lake-public	∨ し工作空间	$\stackrel{\scriptsize o}{\sim}$ admin $$
	数据源管理 创建数据源			根据名称查询	
🖵 作业监控	名称	类型	更新日期	操作	
🗟 数据源管理	maxcompute-ds	MaxCompute	2023-12-06 09:39:04		
🔓 数据入湖	mysql-ds	MySQL	2023-12-06 09:38:01		
✔ 发布任务		共2条 20条页 >	〈 1 〉 前往 1 页		
3 数据开发					
🗟 数据出湖					
🛆 用户管理					
					0
					Ö

点击创建数据源按钮,新增数据源。红色星号标识着必填项。需要输入数据源名称、选择数据源类型(目前支持 MySQL、Kafka和MaxCompute 三种数据类型),输入不同数据源类型所需要的服务地址以及配置相关安全信息。最后点击创建按钮即可新增一条数据源。

D Lakelı	nsight	lake-public	~ し 工作空间	은 admin ~	
<u> </u>	数据源管理				
_	创建数据源			根据名称查询	
┗┙ 作业监控	名称	类型	更新日期	操作	
🔓 数据源管理	maxcompute-ds	MaxCompute	2023-12-06 09:39:04		
🗟 数据入湖	mysql-ds	MySQL	2023-12-06 09:38:01		
✔ 发布任务		共2条 20条页 🗸	1 > 前往 1 页		
💣 数据出湖					
🛆 用户管理					

0

D Lakel	nsigh	t		lake-public		¦admin ∨
Œ	数据派	* 名称			民名称查询	
🖵 作业监控	名称	* 类型	MySQL			
日 数据源管理	maxcom	* 地址				
遏 数据入湖	mysql-d	* 3世口				
✓ 发布任务		* 数据库名				
国 数据开发		* 数据库用户名				
		* 数据库密码				
圖 数据田湖		扩展配置				
🛆 用户管理		描述				
			创建取消	<i>L</i>		
						0

3. 对创建完成的数据源可以点击**编辑**按钮来进行修改,更新数据源相关信息。

Ö Lakel	nsight			lake-public	~	も 工作空间	은 admin ~
<u>—</u>	数据源管理						
	创建数据源					根据名称查询	
닐 作业监控	名称	类型	更新日期		操作		
🔓 数据源管理	maxcompute-ds	MaxCompute	2023-12-06 09:39	9:04	2 🗇		
🖥 数据入湖	mysql-ds	MySQL	2023-12-06 09:38	3:01	2		
✔ 发布任务		共2条 20条页 >	1 > 前	1 页			
♂数据出湖							
🛆 用户管理							
							Ċ

4. 点击**删除**按钮,可以删除一条数据源。

D Lake	Insight	lake-public	✓ ● 工作空间	
<	数据源管理 创建数据源			根据名称查询
🖵 作业监控	名称	类型	更新日期	操作
🔒 数据源管理	maxcompute-ds	MaxCompute	2023-12-06 09:39:04	
🔓 数据入湖	mysql-ds	MySQL	2023-12-06 09:38:01	
✔ 发布任务		共 2 条 20 条/页 V	< 1 → 前往 1 页	
🔊 数据出湖				
🛆 用户管理				
				0

3.5 数据入湖

1. 点击左侧数据入湖,查看当前已有的数据入湖任务列表。

D Lake	Insight			lake-public	✓ も工作空间 ♀ admin ∨
<	数据入湖				
	添加入湖任务				根据名称查询
」 作业监控	名称	类型	任务启动状态	更新日期	操作
🗟 数据源管理	mysql2lakesoul	MySQL	•运行中	2023-12-06 09:44:15	2 🔟 🕕 🖹
🖥 数据入湖	maxcompute2lakesoul	MaxCompute	• 未启动	2023-12-06 09:39:40	2 🗓 💿 🖹
✔ 发布任务		共:	2条 20条/页 > < 1	> 前往 1 页	
🗳 数据出湖					
🛆 用户管理					
					6

点击添加入湖任务按钮,可以创建一个入湖任务。请从上到下依次填写资源信息,分别输入入湖任务名称、选择数据源(不同数据源会显示不同的配置项),以及其他所需的资源信息。点击创建按钮即可新增一个入湖任务。

D Lakel	nsight					lake-public		~	し工作空间	⊖ admin ∨
Ξ	3	* 名称								
🖵 作业监控	添 名:	* 数据源	maxcompute-ds	~	还没有? 去添加				查询	
🗟 数据源管理	my	* 表名	请选择	~					•	
2 数据入湖	ma	Range分区	请选择	~					0 [
☞ 发布任务		土曜	请选择	~					. 8	
数据开发			执行器cpu数 驱动器cpu数	2	▲ 执行器内存大小 ▲ 驱动器内存大小	4	*	G ×		
ਡ 数据出湖			执行器数	4		-	~	0		
🛆 用户管理		描述								
			创建取消			ĥ			- 11	
										•
										0

- 点击编辑按钮,可以修改并更新一条数据入湖任务的相关信息(操作同上述数据源管理的编辑操作)。
- 4. 点击**删除**按钮,可以删除数据入湖任务(操作同上述数据源管理的删除操作)。
- 5. 点击**启动任务**按钮,即可以启动一个数据入湖任务,可以实时查看任务的运行状态,启动后可以随 时停止运行中的任务。

Ö Lakel	nsight					lake-public	~	し 工作空间	$\stackrel{\ }{\sim}$ admin $$
☲	数据入湖			1				根、各称查询	
🖵 作业监控	名称	类型		任务启动状态		更新日期		操作启动任务	
🔓 数据源管理	mysql2lakesoul	MySQL		• 停止		2023-12-06 09:44:15		2 🖬 🎯 🖻	
🔓 数据入湖	maxcompute2lakesoul	MaxCompute		• 未启动		2023-12-06 09:39:40		2 🗓 💿 🖻	
♥ 发布任务			共2条 20	条页 ~ 〈	1 >	前往 1 页			
3 数据开发									
🔓 数据出湖									
🛆 用户管理									
									0
10 Labor	to bet								0
	nsignt					lake-public		8 工作空间	i≍ admin ∨
=	数据入湖								
	添加入湖任务							根族文称查询	
🚽 作业监控	名称	类型		任务启动状态		更新日期		操作停止任务	
🔒 数据源管理	mysql2lakesoul	MySQL		• 启动中		2023-12-06 09:44:15		2 🗰 🔍 🖻	
🖥 数据入湖	maxcompute2lakesoul	MaxCompute		• 未启动		2023-12-06 09:39:40		2 🗓 💿 🖻	
♥ 发布任务			共2条 20	条/页 ~ 〈	1 >	前往 1 页			

70

6. 点击**日志**按钮在任务启动成功后,可以查看日志信息。

♂数据出湖
○ 用户管理

Ö Lakel	Insight			lake-public	~ む工作空间	ho admin $$
☲	数据入湖 添加入湖任务				根据名称之为	
🖵 作业监控	名称	类型	任务启动状态	更新日期	操作	
🔓 数据源管理	mysql2lakesoul	MySQL	• 启动中	2023-12-06 09:44:15		J
🛃 数据入湖	maxcompute2lakesoul	MaxCompute	• 未启动	2023-12-06 09:39:40	2 🖬 💿 🖻	
♥ 发布任务			共2条 20条页 ∨ < 1 >	前往 1 页		
ਡ 数据出湖						
▲ 用户管理						
						U
Ö Lakel	Insight		日志			×
	数据入湖		++ id -u			Î
	添加入湖任务		+ myuid=185 ++ id -g			
し 作业监控	名称	类型	+ mygld=0 + set +e ++ getent passwd 185			
🔒 数据源管理	maxcompute2lakesoul	MaxCompute	+ uidentry= + set -e			
2 数据入湖	mysql2lakesoul	MySQL	+ '[' -z '' ']' + '[' -w /etc/passwd ']'	0		
			+ echo '185:x:185:0:anonymous uid:/opt/spark + '[' -z /usr/local/openjdk-11 ']' + SPARK_CLASSPATH=':/opt/spark/jars/*'	::/bin/false		
■ 数据开发			+ env + grep SPARK_JAVA_OPT_ + sed 's([^]=)*=)(*))//1/a'			
🗳 数据出湖			+ sort -tk4 -n + readarray -t SPARK_EXECUTOR_JAVA_OPTS			
🛆 用户管理			+ ' [' -n '' ']' + ' [' -z ']' + ' [' -z ']'			
			+ '[' -n '' ']' + '[' -z ']'			
			+ '[' -z x ']' + SPARK_CLASSPATH='/opt/spark/conf::/opt/spa	rk/jars/*'		
			+ case \$1 in + shift 1 + CMD=(\${TAVA HOME}/hin/java "\${SPARK EVECUT	OR TAVA OPTS[@]}" -Yms\$SPARK FY	ECUTOR MEMORY -Xmx\$SPARK F	XECUTOR MEMORY -
			+ exec /usr/bin/tini -s /usr/local/openjd Using Spark's default log4j profile: org/apa	k-11/bin/java -XX:+IgnoreUnreco che/spark/log4j2-defaults.prop	ognizedVMOptionsadd-open erties	s=java. base/java
			23/12/05 09:22:10 INFO KubernetesExecutorBac	kend: Started daemon with proce	ess name: 20@maxcompute21ak	esoul-11-73df9f8c

3.6 数据开发

1. 点击左侧数据开发按钮,进入Zeppelin界面,进行相关Flink SQL任务开发。



详细的 Flink SQL 语法请参考 LakeSoul Flink Connector, Flink SQL 以及 Flink on Zeppelin 文档:



https://nightlies.apache.org/flink/flink-docs-release-1.17/docs/dev/table/sql/overview/

SQL

SQL # This page describes the SQL language supported in Flink, including Data Definition Language (DDL), Data Manipulation Language (DML) and Query Language. Flink's SQL support is based on Apache Cal

https://zeppelin.apache.org/docs/latest/interpreter/flink.html

zeppelin.apache.org

2. 在 LakeInsight 上使用 Flink SQL 进行实时任务开发的流程:

进入Zeppelin界面,点击**Create New Note**按钮,创建一个新的Zeppelin Note进行开发,亦可导入已有的Zeppelin的文档。下面是一个简单例子,展示创建Lakesoul Catalog、数据写入以及数据查询的过程。

• 创建catalog

```
1 create catalog lakesoul with('type'='lakesoul');
```

- 2 use catalog lakesoul;
- 3 use `default`;

D LakeInsight

lake-public · 人工作空间 久 lakeinsight-user-01 ·

	<- 返回上一级 <i>Seppelin</i> Notebook -				Q Search e lak	teinsight-user-01 🗸				
🖵 作业监控	flink-sql-dev-02	21 ▲ ▲ 3 答 @ ⊕ == Head Q Û				🖙 🏟 🔒 default 🗸				
局 数据源管理	%flink.bsql				FINE	ISHED D 💢 🗐 🐵				
昌 数据入湖	<pre>create catalog lakesoul with('type'='lake use catalog lakesoul; use `default`;</pre>	esoul');								
√ 发布任务	Catalog has been created.									
🗐 数据开发										
🗳 数据出制	SFLINE.bagl PENSHED D X 🗐 🔍 rest default: rest.table ('default': r									
🛆 用户管理	'path'='\$3://lakeSoul-bucket/\$ink/te: Table has been created. Took 3 sec. Last updated by lakeInsight user 01 at December 06 20	'path'='33://lakesoul-bucket/sink/test', 'hashBucketNum'='4'); Table has been created. Took 3 sec. Last updated by lakensight user 01 at December 05 2023, 652-43 PM.								
	#fink.bsql mEnux.JOB FNUSHED use 'default'; Str 'table.local-time-zone' = 'Asia/Shangheit'; Str 'table.local-time-zone' = 'Asia/Shangheit'; Strine'); Insert into 'lokesoul'.'default'.test.table values (1,'AAA', 90, '2823-85-11', 'China'); Insert into 'lokesoul'.'default'.test.table.values (3,'AAA', 98, '2823-85-11', 'China'); Insert into 'lokesoul'.'default'.test.table.values (3,'AAA', 98, '2823-85-18', 'China'); Insertion successfully. Insertion successfully. Insertion successfully.									
	%flink.bsgl ■FLINK.JOB FINISHED ▷ X 등 use 'default': :setect 'from 'iskesul'.'default'.test_table; ■ ● ▲ > settings ~									
	id	v name v	score ×	date ~	region	~ =				
	1	ААА	100	2023-05-11	China					
	2	BBB	100	2023-05-11	China					
	3	AAA	98	2023-05-10	China					

• 创建table

1 use `default`;

```
2 create table if not exists test_table (`id` INT, name STRING, score
INT,`date` STRING,region STRING, PRIMARY KEY (`id`,`name`) NOT ENFORCED )
PARTITIONED BY (`region`,`date`) WITH ( 'connector'='lakeSoul',
'use_cdc'='true','format'='lakesoul', 'path'='s3://lakesoul-
bucket/sink/test', 'hashBucketNum'='4');
```

🐌 Lakelı	nsight			lake-public	 人工作室间 名 lakeinsight-user-01 ~ 			
	<- 返回上一级 Took 54 sec. Last updated by lakeinsight-user-01 at December 06 2023, 6:3	52.29 PM.						
_	(flink heal				PARTIES N X PR A			
└── 作业监控	use `default`;				HINISHED D X B @			
局数据源管理	create table if not exists test_table ('id' INT, name STRING, score INT, 'date' STRING, region STRING, PRIMARY KEY ('id', 'name') NOT ENFORCED) PARTITIONED BY ('region', 'date') WITH ('connector'='lakeSoul', 'use_cdc'='true', 'format'='lakeSoul', 'path'='s3://lakeSoul-bucket/sink/test', 'hashBuckethum'='4');							
2 数据入湖	Date: Has been treated.							
☞ 发布任务	別flink.bsgl							
🗐 数据开发	use 'default'; SET 'table.local-time-zone' = 'Asia/Shanghai' insert into `lakesoul`.'default'.test_table v	; ////////////////////////////////////						
▲ 数据出湖	insert into lakesoul , derault .test_table v insert into `lakesoul`.`default`.test_table v Insertion successfully.	alues (2, 888', 100, 2023-05-11', 'Unina'); values (3,'AAA', 98, '2023-05-10', 'China');						
요 用户管理	Insertion successfully. Insertion successfully.							
	Took 10 min 27 sec. Last updated by lakeinsight-user-01 at December 06 2	023, 7:03:24 PM.						
	<pre>%flink.bsql use 'default'; select * from 'lakesoul', 'default'.test_table</pre>	; इत रू			■FLINKJOB FINISHED ▷ X 団 ⊕			
	id	v name	score ~	date ~	region ~ =			
	1	AAA	100	2023-05-11	China			
	2	BBB	100	2023-05-11	China			
	3	AAA	98	2023-05-10	China			
	Took 2 sec. Last updated by lakeinsight-user-01 at December 06 2023, 7:0	3:26 PM. (outdated)						
	%flink.bsql				READY D 💥 💷 🐵			

```
1 use `default`;
```

- 2 SET 'table.local-time-zone' = 'Asia/Shanghai';
- 3 insert into `lakesoul`.`default`.test_table values (1,'AAA', 100, '2023-05-11', 'China');
- 4 insert into `lakesoul`.`default`.test_table values (2,'BBB', 100, '2023-05-11', 'China');
- 5 insert into `lakesoul`.`default`.test_table values (3,'AAA', 98, '2023-05-10', 'China');

lake-public ~ 소工作空间 온 lakeinsight-user-01 ~

D LakeInsight

_ !	<- 返回上一级						
<u></u>	%flink.bsql				FINISHED D	XBO	
🖵 作业监控	use 'default'; create table if not exists test_table ('id' INT 'path''31/(]akesail_bucket/cink/test' 'b	T, name STRING, score INT,`date` STRING,region ST	'RING, PRIMARY KEY (`id`,`name`) NOT ENFORCED) P/	ARTITIONED BY (`region`,`date`) WITH ('connector	'='lakeSoul', 'use_cdc'='true','format'='la	kesoul',	
局数据源管理	Table has been created.						
8L 数据入湖	Took 3 sec. Last updated by lakeinsight-user-01 at December 06 2023, 6:52:42	Tooli 3 see. Last updated by latelinsight-user-01 at December 66 2023, 652-43 PM.					
ቑ 发布任务	%flink.bsql				FLINK JOB FINISHED	XØØ	
■ 数据开发	<pre>use aerault; SET 'toble.local-time-zone' = 'Asia/Shanghai'; insert into 'lakesoul`.'default'.test_table val insert into 'lakesoul`.'default'.test_table val insert into 'lakesoul`.'default'.test_table val</pre>	ues (1,'AAA', 100, '2023-05-11', 'China'); ues (2,'888', 100, '2023-05-11', 'China'); ues (3,'AAA', 98, '2023-05-10', 'China');					
▲ 数据出湖	Insertion successfully. Insertion successfully. Insertion successfully.						
🐣 用户管理	Took 10 min £7 ace. Last updated by lake insight user 01 at Occember 00 £055	5, 7.08.04 PM.					
	%flink.bsql				SFLINK JOB FINISHED D	XØØ	
	<pre>use `default`; select * from `lakesoul`.`default`.test_table;</pre>						
	III 🕼 🗳 🕍 🖄 🕹 🕶 settings	•					
	id ~	name ~	score ~	date ~	region	~ =	
	1	ААА	100	2023-05-11	China		
	2	BBB	100	2023-05-11	China		
	3	AAA	98	2023-05-10	China		
	Took 2 sec. Last updated by lakeinsight-user-01 at December 06 2023, 7:03-26	5 PM. (outdated)					
	%flink.bsql				READY D	XØØ	

• 数据查询

- 1 use `default`;
- 2 select * from `lakesoul`.`default`.test_table;

D Lake	Insight			lake	-public 〜 し工作空(□ A lakeinsight-user-01 ~
	<- 返回上一级					
- 作业监控	<pre>%flink.bsql use `default`; create table if not exists 'anth'='s1/d akesul-b</pre>	test_table (`id` INT, name STRING, score INT,`date`	STRING, region STRING, PRIMARY KEY (`id`,`name`) NOT ENFORCED) PARTITIONED BY ('region','date')	WITH ('connector'='lakeSoul', 'use_cdc'=	FINISHED ▷ X 🗐 ۞
动 数据源管理	Table has been created.	accessitio test ; manuaccettum = + ;;				
🏜 数据入湖	Took 3 sec. Last updated by lakeinsight-user-0	11 at December 06 2023, 6:52:43 PM.				
♥ 发布任务 ■ 数据开发	<pre>%flink.bsql use 'default'; SET 'table.local-time-zone' insert into 'lakesoul'.'def insert into 'lakesoul'.'def</pre>	= 'Asia/Shanghai'; ault'.test_table values (1,'AAA', 100, '2023-05-11' ault'.test_table values (2,'888', 100, '2023-05-11' ault'.test_table values (3,'AAA', 95, '2023-05-11'	, 'China'); , 'China'); 'China');		20 F	LINK JOB FINISHED D 🔀 🗟
출 数据出湖	Insertion successfully. Insertion successfully. Insertion successfully.					
🕰 用户管理	Took 10 min 27 sec. Last updated by lakeinsig	ht-user-01 at December 06 2023, 7:03:24 PM.				
	use 'default'; select * from 'lakesoul'.'default'.test_table; I dat de las le2 LE2 de settings ~					
	id	∽ name	~ score	~ date	✓ region	~ =
	1	AAA	100	2023-05-11	China	1
	2	BBB	100	2023-05-11	China	
	3	AAA	98	2023-05-10	China	
	Took 2 sec. Last updated by lakeinsight-user-0	11 at December 06 2023, 7:03:26 PM. (outdated)				
	%flink.bsql					READY D X 🗐 🖗

3.7 任务发布

1. 点击左侧发布任务,查看当前已有的发布任务。

D Lakel	Insight	lake-public	 く 工作空间 名 admin マ 		
-	发布任务				
	创建任务				根据名称查询
''」作业监控	任务名	处理引擎	运行状态	更新时间	操作
🔓 数据源管理	task-publish-01	Flink	• 未启动	2023-12-06 09:52:16	2 🗓 💿 🗐
2 数据入湖			共1条 20条页 🗸 1	〉 前往 1 页	
✓ 发布任务					
💣 数据出湖					
🛆 用户管理					
					7
					0

- 点击创建任务按钮,弹出创建发布任务页面,依次输入任务名称、选择任务处理引擎、...选择sql语 句来源(自助SQL或者来自zeppelin)、任务运行资源以及描述信息,输入完成后,点击发布按钮, 即可完成一个发布任务。
 - a. SQL语句来源包括: 自助SQL以及zeppelin

如果选择编辑SQL,可以在面板中直接写入Flink SQL语句

如果选择**Zeppelin**,可以从已有的zeppelin中选择已开发的Flink SQL任务(即上述数据开发所 创建的数据)。选择后,数据信息会并自动导入下面面板中

🔁 Lakelı	nsight			lake-public	∨ ら工作空间	⊖ admin ∨	
-							Î
	* 任务名						
🖵 作业监控	处理引擎	Flink					
🔓 数据源管理	任务类型	流式任务					
2 数据入湖	checkpoint时间间隔(/ms)	300000					
✔ 发布任务	SQL来源	编辑SQL Zeppelin					
		1			sql	~	
📓 数据出湖							
🛆 用户管理							
		重置				9	•
			=			10	
	任务运行资源					_	ſ

- 3. 点击编辑按钮,可以修改并更新任务发布相关信息(操作同上述数据源管理的编辑操作)。
- 4. 点击删除按钮,可以删除一条已发布的任务(操作同上述数据源管理的删除操作)。
- 5. 点击**启动任务**按钮,即可以启动一个发布任务,可以实时查看任务的运行状态(操作同上述数据入 湖的启动/停止操作)。
- 6. 任务启动成功后,可点击**日志**按钮在,查看日志信息(操作同上述数据入湖的查看日志操作)。

3.8 数据出湖

1. 点击左侧数据出湖,查看当前已有的数据出湖任务列表。

D Lakel	nsight	lake-public	✓ し工作空间	\red{A} admin \lor		
Ţ	数据出湖					
_	添加出湖任务				根据名称查询	
▶ 作业监控	名称	类型	任务启动状态	更新日期	操作	
🔓 数据源管理	lakesoul2mysql	MySQL	• 未启动	2023-12-06 09:54:23	2 🖬 🛛 🖡	
圖 数据入湖	lakesoul2maxcompute	MaxCompute	• 未启动	2023-12-06 09:53:54	Z 🗓 🛛 🖻	
✔ 发布任务		共2条	20条/页 < 1 >	前往 1 页		
圖 数据开发						
🔊 数据出湖						
🛆 用户管理						
						7

 点击添加数据出湖按钮,弹出数据出湖任务页面。依次输入出湖任务名称、选择出湖数据包括数据 库名和表名;选中导出到数据库信息包括选择数据源(不同类型的数据源对应的配置信息不同), 设置表名库名等信息,设置任务运行资源以及描述信息。输入完成后,点击创建按钮,即可创建一 个数据出湖任务。

D Lake			A admin ∨
	* 出湖任务名		
—	出湖数据		
□ 作业监控		* 数据库名	3 清选择 >
		* 表名	る 「清选择 ~
🔓 数据源管理			
2 数据入湖	导出到		
		* 数据源	● 请选择 ∨ 还没有? 去添加
✓ 发布任务			创建新表 选择已存在的表
■ 数据开发		* 库名	
		* = 4	7
		花石	
🛆 用户管理			
	运行资源配置	source并行度	き 4 へ sink并行度 4 へ
			*/#7450-157
		cneckpoint仔陌时间响(/ms)	300000 数据库时区 UTC
	描述		

- 3. 点击编辑按钮,可以修改并更新数据出湖相关信息(操作同上述数据源管理的编辑操作)。
- 4. 点击删除按钮,可以删除一条数据出湖任务(操作同上述数据源管理的删除操作)。
- 5. 点击**启动任务**按钮,即可以启动一个数据入湖任务,可以实时查看任务的运行状态(操作同上述数 据入湖的启动/停止操作)。
- 6. 在任务启动成功后,点击**日志**按钮,可以查看日志信息(操作同上述数据入湖的查看日志操作)。

3.9 作业监控

- 1. 点击左侧作业监控按钮。
 - a. 选择**作业统计,**查看当前各模块任务数及运行状态信息,可通过右侧筛选出某段时间内的相关 信息。

D Lakel	nsight	lake-public	 ・ ・ ・
☲	作业统计 资源监控		
🚽 作业监控	任务总数	开始时间	③ 选择时间
🗟 数据源管理	###2/6 ###2.3/8 ###2.3/8	结束时间	© 2023-12-06 18:51:27
🖥 数据入湖			
✔ 发布任务		查看详情	
数据开发	模块任务状态信息		
💣 数据出湖	■ 运行中 ■ 运行完成 ■ 未运行 ■ 运行失数		
▲ 用户管理			

b. 选择**资源监控,**查看当前任务运行的状态。

D Lakel	nsight		lake-public > 人工作空间 名 admin >
Œ	作业统计资源监控		
□ 作业监控 ◎ 数据源管理	맘 Dashboards Playlists	Dashboards Create and manage dashboards to visualize your data	New ~
a 数据入湖 ✓ 发布任务		Search for dashboards and forders Filter by tag Image: Starred	1≡ Sort ✓
🗐 数据开发		Name BB Flink Running App Monitor	ags fink job manager task manager
🔊 数据出湖		B Spark Running App Monitor	apache flink job manager task manager
▲ 用户管理			

D LakeInsight



