

阿里云开发者社区
ALIBABA CLOUD DEVELOPER COMMUNITY

阿里云数据中台解决方案 系列白皮书

水务行业数据 智能化解决方案

2020云栖大会 特别版

数智俱乐部



钉钉扫描二维码 加入

阿里云数据中台 出品



钉钉扫码加入
阿里云数据中台交流群



阿里云开发者“藏经阁”
海量免费电子书下载

目录

1.行业挑战	4
1.1 平台能力有待整合和提升	4
1.2 数据资产化广度和深度不足	4
1.3 数据运营管理过程中面临各种各样的管理困局	5
2.解决方案	6
2.1 基础数据智能平台	7
2.2 水务行业数据标准、数据模型架构	7
2.3 水务数据资产中心	7
2.4 水务专题分析	8
2.5 水务客户画像	8
2.6 水务管网档案	9
3.产品优势	10
3.1 数据规范统一	10
3.2 智能计算优化	10
3.3 一站式研发体验	13
3.4 系统化构建数据目录	13
3.5 可视化数据资产	15
3.6 数据使用简单可依赖	16
4.客户案例	19
4.1 某市水务公司	19
法律声明	22

1.行业挑战

放眼国内外水务的行业经验和发展趋势，智慧水务发展大体可分为三个阶段。第一阶段，以自动化控制为核心，着眼于制水工艺的自动化以及生产效率的提升；第二阶段，以信息化为核心，应用于企业业务管理、数据传输、移动应用、算法应用方面进行突破；第三阶段，以智慧化为核心，是大数据、人工智能、云计算、物联网的综合应用。目前，大部分水务集团处在第二阶段信息化阶段，头部企业开始向智慧水务方向拓展的阶段。而在向第三阶段的拓展过程中，又以数据智能化建设，为最重要的基础能力建设之一。

目前，数据智能化建设碰到的主要挑战包括：

1.1 平台能力有待整合和提升

在第一、第二阶段的建设过程中，积累了多样的软件产品或平台，并且部分产品能力已经比较老旧，从而导致：

- 1) 产品软件封闭，产品软件之间的可集成性差，在不同厂商的产品间难以进行集成。这就导致一些常见的管理过程需要依赖多个软件实现，一些彼此关联的信息被分散在不同系统中，对管理造成不便；
- 2) 产品软件的功能协同性低。产品软件尽管专注于某一方面的功能，但多个平行的功能软件之间难以协同，这在客观上限制了不同平台数据间的融合能力；
- 3) 产品软件的扩展缺乏弹性。产品软件采用传统技术架构模式，受限于节点数、CPU核数、采集点数等各种条件，一些产品基本无法做到线性扩展，严重制约数据的发展；

1.2 数据资产化广度和深度不足

随着水务业务从流程驱动到数据驱动，数据将成为企业的重要资产。水务行业在不同阶段的建设过程中，积累了大量的数据，但尚未提升到数据资产化运营的高度，数据的价值也尚未很好的体现出来。数据资产管理模式和意识，虽然初步成形，但在数据生命周期管理、标准

化存储管理、服务标准化等方面仍有待提高。并且受限于工具和机制，数据资产管理仍滞后于数据产生的过程，即数据资产管理与信息化项目的管理之间缺乏整合，人员、组织、项目、系统等要素与数据资产管理仍然脱节，因此在深化数据服务时，难以确保数据标准的遵从，也难以保障数据服务的安全和可控。

1.3 数据运营管理过程中面临各种各样的管理困局

- 1) 难以宏观掌控数据平台全网的整体运行状况
- 2) 数据服务故障通常是事后响应，难以提前发现问题
- 3) 异常问题定位十分困难，涉及面太广
- 4) 数据质量、服务效率状况难以衡量和评价

2.解决方案

数据智能化解决方案，可实现对“原、制、供、排、污”全业务链的生产数据和业务数据，进行集成和整合，构建“人”，“水”，“厂”全方位的水务行业数据资产，并提供多维度数据资产的感知、分析、监控的能力；可构建面向水务基础设施数字化管理、管养精细化、调度智能化、服务扁平化等应用场景的水务数据服务总线，实现企业级的数据服务发布和共享，支撑技术中心管理人员，长期高效的运营数据服务能力，提高企业整体数据运营能力和工作效率。

通过该方案的统一数据汇聚和标准管理的能力、统一维护并对外提供服务的模式，盘活融通数据资产、激活数据服务能力，促进“产业数据化、数据产业化”的正循环效应，从而实现“统一标准规范、统一数据存储、统一数据资产、统一技术平台、统一数据服务、百花齐放应用”的效果。



数据智能化解决方案，主要包括：

2.1 基础数据智能平台

提供一站式集数据引入、规范定义、数据建模、数据研发、数据萃取、数据资产管理和数据服务的全链路智能数据构建及服务平台。包含存储组件、计算组件、设计组件、开发组件、调度组件、运维组件、服务组件、监控组件、管理组件、分析组件的底座，可以满足集团、子公司、厂区的数据全域融合存储、复合计算、实时响应等需求；

2.2 水务行业数据标准、数据模型架构

提供水务大数据的标准体系，含数据标准规范、数据集成标准规范、数据分析指标规范、数据交换与共享规范，确定数据资产目录和数据模型清单；

2.3 水务数据资产中心

基于数据智能平台，采集并整合公司各个部门及所属企业的相关数据，建立水务综合数据模型和共享交换能力，并且建立对数据资产的管理、维护、更新和使用的长效管理机制，实现企业各部门之间的互联互通、资源共享，各个部门可以及时、方便的获取与当前业务密切相关的其他部门共享的业务信息资源，优化信息共享流程，促进跨部门的业务协同，实现数据资源的全面汇聚和高效利用。

1) 数据资产类别：

- a. 设施数据：水务所有相关基础设施数据；
- b. 量测数据：水务全业务链条中各个环节的生产数据(如：流量、压力、水质等信息)；
- c. 业务数据：来自核心业务系统，主要为供水业务的数据，包括：物联网远传表计、营收、工单等数据；
- d. 专题数据：主要为重大活动保障、数据对接与互联互通等专题数据；

2) 数据资产应用:

- a. 可视化数据资产：系统化构建水务数据资产大图，数据视角还原业务系统、提取业务数据知识，并可快速感知业务关键环节及数据。
- b. 数据标签化管理：根据规则自动化建模，生成基于业务含义、工艺、工况的多维体系的数据标签化管理。
- c. 数据资产目录：基于规范化建模、高效自动化的元数据抽取，以标准的技术框架系统地自动化构建规范可读的业务化数据目录，形成数据资产地图，方便业务查找及应用。

2.4 水务专题分析

依赖数据智能平台，利用水务数据资产，构建水务专题分析，实现水务全过程、全要素分析，形成对水务数据和行业数据的深度挖掘与融合分析能力，主要包括：

- 客户分析主题
- 收益分析主题
- 供水水量分析主题
- 计划维护类分析主题
- 客户服务分析主题
- 水质水压分析主题
- 管网数据分析主题
- 大客户及商业客户分析主题

2.5 水务客户画像

依赖数据智能平台，利用水务数据资产，全面打通客户数据，全面整合客户信息，实现全渠道客户的唯一识别，构建客户画像及标签体系。为 CIS 及客户服务系统提供统一的客户数据服务支撑：

1) 客户身份识别（OneID 模型）

通过手机、身份证、姓名、户号、地址、城市、mac 地址、open_id、uid、union_id、wx_id 等参数将分散各个系统的顾客信息连接为一个整体，生成每个顾客的个人识别号 OneID。基于 OneID 可以更加全面的为顾客打标签，然后将标签用于会员群体画像输出和人群筛选及触达等服务场景下。

2) 客户洞察

实现对客户画像的分析，客户画像主要包含单个客户画像及客群画像，主要对用户关心的客户数据进行统一的展示，一般包括基本属性、社交属性、消费标签、行为标签、互动标签等分类，最终根据实际的数据定义和分类划分，对客户各类型数据和标签进行定义。

2.6 水务管网档案

依赖数据智能平台，利用水务数据资产，构建水务管网档案信息：

1) 管网档案

建立完整的管网资产电子档案，包括：管道规格、管道位置、阀门表计位置、管道材质、品牌、竣工年限、管道维修次数等信息。

2) 管网分析

可通过管网画像分析评价不同管道使用寿命，对存在漏损风险的管段发出预警提示。并可基于管网资产档案，构建管道资产状态的定期报表、管道常规使用年限等关键指标。

3.产品优势

3.1 数据规范统一

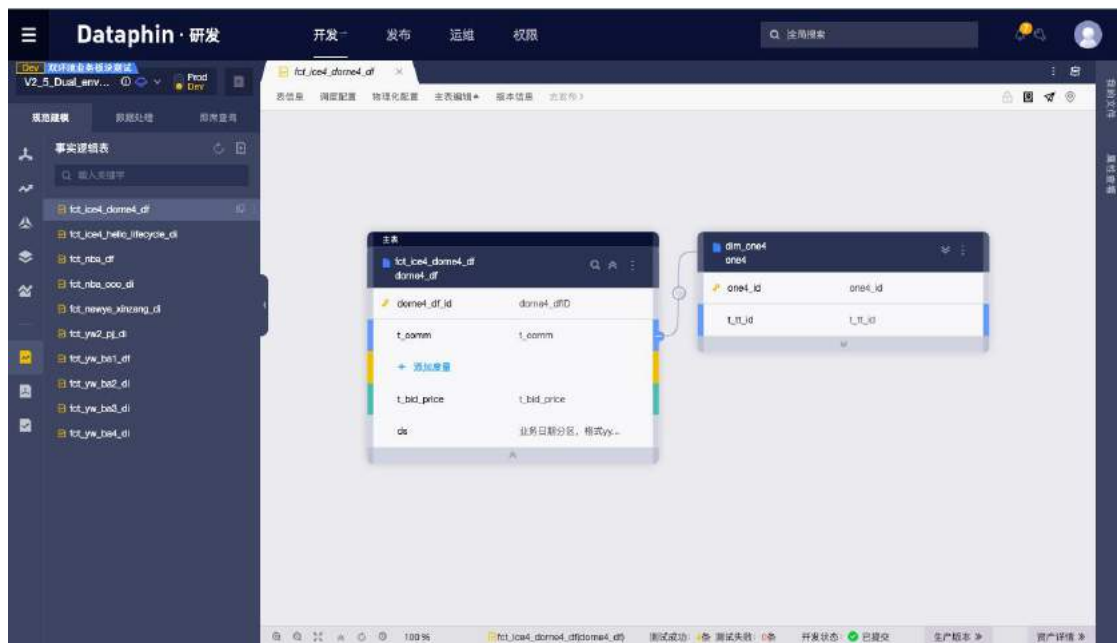
采用维度事实建模理论，对维度、维度属性、业务过程、指标字段等进行严格的 标准化规范化定义，保障数据质量，避免数据指标定义的二义性。



3.2 智能计算优化

支持从业务视角进行逻辑建模，逻辑模型发布后，系统自动化进行物理建模、编码，从而降低对开发人员的技术能力依赖。

第一步，定义业务过程及分析维度，设计逻辑模型：



第二步，系统自动化物理建模和编码，并可一步投入生产调度：

特别地，实时研发支持本地采样数据调试，保障开发质量、保障数据质量

The screenshot displays the Dataphin interface for configuring and debugging a Flink SQL task. The task is named 'dwd_ae_log_clk_pc_us'. The configuration includes a Flink SQL query that inserts data into the 'dwd_ae_log_clk_pc_us' table. The query selects 'date_time', 'stat_date', 'stat_hour', and 'log_time' from a source table. The execution results are shown in a table with columns 'log_id', 'logkey', 'ip', and 'user'.

任务参数 任务配置 版本信息

1 --Flink SQL
2 -----
3 --Author: 闻超(36226)
4 --CreateTime: 2019-12-24 15:18:22
5 --Comment: 写入美国点击明细
6 -----
7 set dwd_ae_log_clk_pc_us.accessKey='f8a83d15-e96d-44d5-b16d-becc001e4900';
8
9 INSERT INTO dwd_ae_log_clk_pc_us
10 select
11 date_time
12 ,stat_date
13 ,stat_hour
14 ,log_time
15 in

Console Result 1

调试数据 中间结果 调试结果

aecdm_dev.dwd_log_hjij_guoji_rj_us_new

	log_id	logkey	ip	user
1	356947a0BdqD	/ae_pc_ctr.statweb_ae_ctr	198.16.66.44	Mozilla/5.0 NT ...
2	66a668a2K3Xa	/x.p.d	185.29.243.101	Mozilla/5.0 NT ...
3	356947a0BdqD	/ae_pc_ctr.statweb_ae_ctr	198.16.66.44	Mozilla/5.0 NT ...
4	7d0e2f3pDSGII	/sc.migrate.event	71.105.117.112	Mozilla/5.0 CP...

配置调试数据

数据表列表 所有数据表都需要已采样，才可进行调试

数据表	元数据采样
to_cdm_gmv_biz_order	未采样
to_cdm_biz_pay_order	未采样
to_cdm_biz_pay_order	未采样

元数据采样

	auction_price	biz_type	adjust_fee
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

取消 确定

Dataphin · 研发

开发 发布 运维 权限

全局运维 运维去运维

两期任务 输入节点在运维中

我的节点 今天发布

常规任务 逻辑表物理化任务

Node ID	节点名称	更新时间
n_7738298...	test_nba_pos_dk_0000_v1	2019-08-18 15:55:18
n_7108522...	test_newys_sirang_d_0000_v1	2019-08-28 10:48:43
n_7091874...	dws_yimei_00000_v1	2019-08-27 21:28:03
n_7091884...	dim_yimei_core_00001_v1	2019-08-27 21:21:41
n_7091856...	dim_yimei_core_00001_v1	2019-08-27 21:20:15
n_8956804...	test_test_dome1_cf_0000_v2	2019-08-23 09:48:28
n_8901378...	test_yw_bat_0_0000_v2	2019-08-21 14:03:45
n_8876898...	test_yw_bat_0f_0000_v1	2019-08-20 16:59:50
n_8876830...	dws_shengwei_00000_v1	2019-08-20 15:38:12
n_8876713...	dim_shengwei_core_00001_v1	2019-08-20 15:24:30
n_8876549...	dim_shengwei_core_00001_v1	2019-08-20 15:18:33
n_8876490...	dim_gw1_z_pos_00001_v1	2019-08-20 15:21:01

非生产数据运维，仅供非生产数据操作

展开节点 1

请搜索节点ID或名称

virtual_root_node_6...

dim_yimei_core_000...

test_spark

test_MssCompute_S...

dws_yimei_00000_v1

节点ID: a_833518860515430401 新增

节点: virtual_root_node_83351886051 5430400 新增

优先级: 高优先级

所在项目: Basic

负责人: SuperAdmin(SuperAdmin@dataphin)

更新时间: 2019-08-02 09:37:09

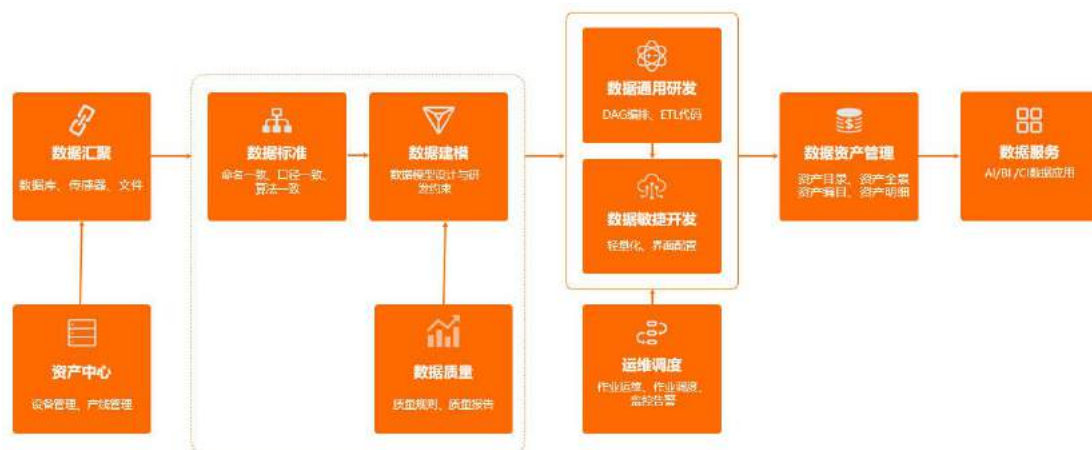
节点描述: virtual root node of Dag[a_833518860515430400]

90 %

n_709188470257221632 dim_yimei_core_00001_v1

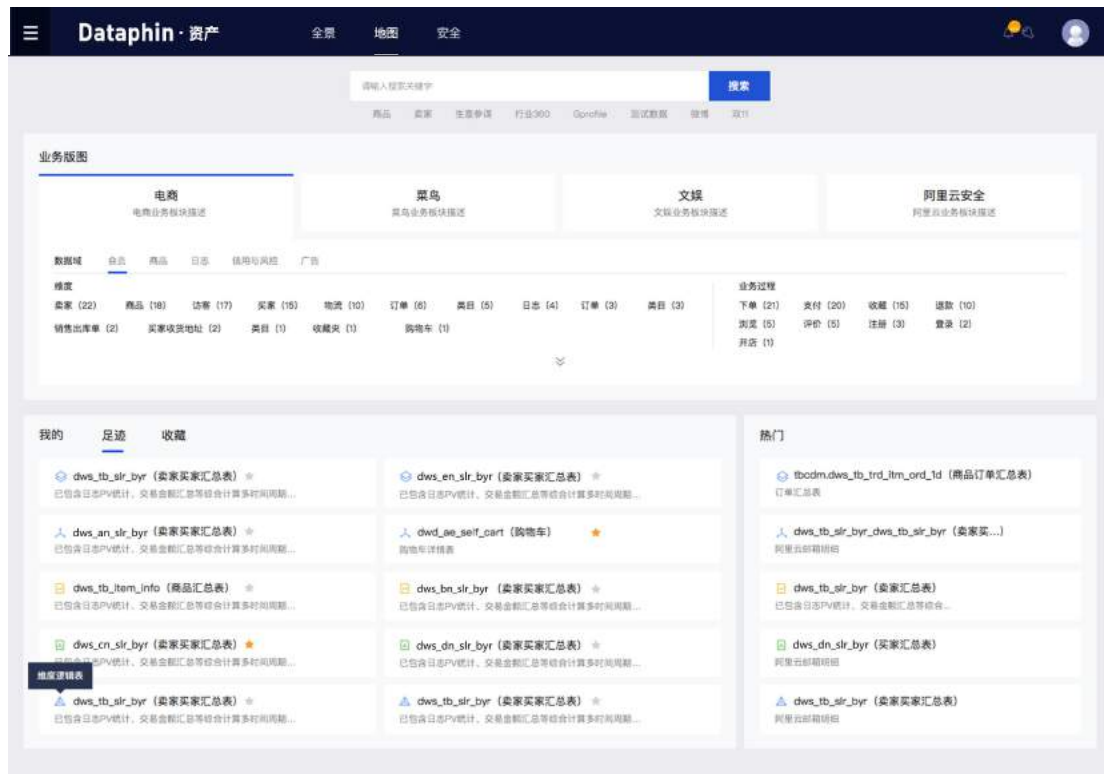
3.3 一站式研发体验

数据引入、建模、研发、运维、数据查找及探查等过程一气呵成，研发链路统一高效。



3.4 系统化构建数据目录

基于规范化建模、高效自动化的元数据抽取，以标准的技术框架系统地构建规范可读的业务化数据目录，形成数据资产地图，方便业务查找及应用。



3.5 可视化数据资产

系统化构建业务数据资产大图，数据视角还原业务系统、提取业务数据知识，并可快速提炼业务关键环节及数据。



3.6 数据使用简单可依赖

定义即服务，研发构建的业务主题式数据逻辑表可被直接、快速地查询和访问，简化查询代码。

1) 数据使用方式 1: API

通过 API 模板化或自定义 SQL 的方式，提供数据 API 查询能力。



个人后台 > 开发中心 > API > 新建API

新建API

API创建方式选择 ----- API基础信息配置 ----- 参数配置

参数配置

服务单元分组配置

测试

单位服务单元

交易数据服务单元

服务单元字段:

☒ 字段名称

- ☒ ord_id
- ☐ pay_time
- ☒ pay_type
- ☐ pay_cardno
- ☐ pay_amt
- ☐ itm_qty
- ☐ cust_id
- ☐ item_id
- ☐ shop_id
- ☐ refund_id

请求参数

参数名	绑定字段	参数类型	操作符	示例
ord_id	ord_id	STRING		
pay_type	pay_type	STRING		

返回参数

参数名	绑定字段	参数类型	示例
ord_id	ord_id	STRING	
pay_type	pay_type	STRING	

个人后台 > 开发中心 > API > 新建API

新建API

API创建方式选择 ----- API基础信息配置 ----- 参数配置

* API脚本编辑 ①

参考示例

```
1
```

解析参数

SQL试运行

服务单元

测试

单位服务单元

集成平台数据服务单元

服务单元对应字段

ord_id

pay_time

pay_type

pay_cardno

pay_amt

itm_qty

cust_id

请求参数

参数名	绑定字段	参数类型	示例	描述	是否必填
-----	------	------	----	----	------

2) 数据使用方式 2：SDK

1 SDK简介

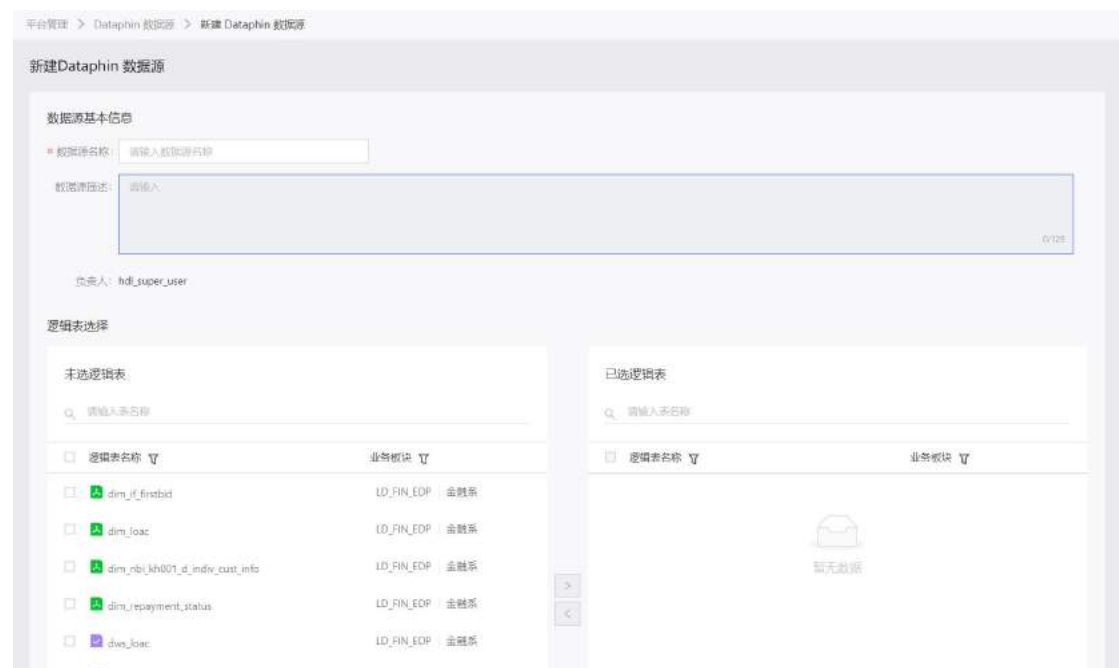
Dataphin-服务SDK是根据您自定义的所有API接口，自动生成的Java调用代码，让您无需复杂编程即可访问Dataphin-服务。这里向您介绍如何使用Dataphin-服务SDK

代码文件的层级结构如下：

- SDK文件夹
 - sdk/
 - ClientDemo.java 这里提供了统一的API调用示例和方法
 - lib
 - sdk-core-java-1.1.0.jar sdk的核心包，为本sdk的核心依赖包
 - sdk-core-java-1.1.0-sources.jar 上述依赖包的源码
 - dataphin-sdk-core-java-v1.0.0.jar sdk的param包，为本sdk的构造请求参数的依赖包
 - dataphin-sdk-core-java-v1.0.0-sources.jar 上述依赖包的源码
 - fastjson-1.2.55.jar 格式化json包
 - ApiDocument.md API接口文档
 - Readme.md 本SDK使用指南
 - LICENSE 版权许可

3) 数据使用方式 3：数据模型数据源

将数据模型作为数据源，以完整的数据模型能力，提供给需要应用系统。



4.客户案例

4.1 某市水务公司

某市水务公司的数据智能化建设，从顶层设计出发，推进企业内部各运营生产数据的采集融合，同时结合城市地理信息数据、用户画像等数据，实现数据资源的深度加工、知识挖掘，提升企业整体数据分析能力，为更多智能应用场景落地和经营分析探索提供坚实数据基础。

该公司通过数据智能平台的建设，深度整合和升级平台能力，强化数据资产管理和运营能力，建立数据智能服务体系，包括指标中心和报表中心、数据一张图等。对内，为业务部门的数据分析、数学建模、业务创新等提供智能数据支撑，实现基于数据的智能应用，将数据转化成核心资产。对外，通过数据协同生态，进行信息共享和数据能力开放。建立了“用数据说话、用数据决策、用数据管理、用数据创新”的管理机制。

1) 指标中心

指标中心提供统一的数据指标服务，相关业务场景可以调用指标中心完成数据指标运算、加工和展示。



指标中心主要能力使用场景包括：

- 指标定义服务：主要定义指标的名称、指标分级、指标含义以及指标对象，在指标运营过程中不断积累和完善，形成完备的指标体系；

- b. 指标计算服务：根据指标的定义完成指标接入和指标运算的功能；指标计算服务在数据智能平台体系支撑上构建，充分借助数据治理的能力构建规范的数据指标，利用大数据平台强大的算力完成指标的计算；
- c. 指标运维服务：基于数据智能平台对数据指标生产过程进行运维监控，及时发现作业和任务的异常，降低数据运维的复杂度，简化运维过程；
- d. 指标发布服务：以数据 API 服务接口方式，对产生的指标进行发布；

2) 报表中心

报表中心提供统一的数据报表服务。通过统一的报表中心，能够提供完备的报表生成、监控和发布服务，自动化完成报表制作，减少人工作业的负荷，提升报表的数据质量，也可不断适应新的报表需求。



报表中心主要能力使用场景包括：

- a. 报表管理服务：用于定义报表的样式、报表生成规则以及管理报表对象等；
- b. 报表生成服务：提供基于数据的报表生成服务，包括报表数据的接入、报表数据计算以及按照样式自动生成报表；
- c. 报表监控服务：提供报表的运维监控服务，包括对报表生成的过程以及生成的结果进行持续性监控；
- d. 报表发布服务：提供报表发布服务，支持发布成 web 页面、图片、PDF、PPT、周报、月报、邮件格式等；

3) 数据一张图

数据一张图是建立全程数据追溯和展示,实现从龙头到源头全流程数据打通,形成围绕水的数据全过程追溯能力。通过建立全程数据追溯和展示,为客户在向食品级饮用水服务转型过程中提供数据支撑。主要实现了两个方面的目标:

- a. 对内目标:建立供水全链条生产运营考评体系,对供水全链条生产运行情况进行监督和考评;
- b. 对外目标:建立供水服务品质的评价标准,通过数据化的品控体系,全面掌握从源水、制水、供水管网到小区二次供水的水质指标,尤其是终端客户明显感知的指标,如水质、水量状况等。

以数据协同为基础,通过指标体系建立和数据可视化手段将对内和对外的指标评价结果显性化,推动基于数据的科学决策体系建立。

4) 生态大协同

通过数据智能平台及相关数据应用的建设,有效加强了与政府、产业上下游以及客户之间的协作互动,加速融入政府一网通办和智慧城市建设进程,与市民形成更加紧密的联结,提升在产业中的优势地位。协同能力,主要包括:

a. 信息共享管理

实现与相关部门共享交换与水务运营和应急指挥相关的数据,包括但不限于:

- 管线数据:向政府备案供排水管网的地理信息数据,用户统筹协调城市规划、建设与运行维护工作;
- 市政工程施工数据:如工程施工的时间、地理位置、施工单位、责任人等,用于应急事件的判断和处置;
- 供水服务运行数据:向政府开放供水服务的运行情况,如供水量、水质、水压、客户满意等相关指标数据;
- 应急事件信息:当出现应急事件时通过平台向政府传达事件的信息和处置情况;

b. 数据服务开放

通过数据服务开放,向政府开放基于供水大数据的决策支撑能力。如人口迁徙、用水规律、疫情后企业复工复产辅助决策等。数据服务开放根据政府侧对水务数据的需求,调用“指标中心”和“报表中心”的能力,对来自于数据层的原始数据进行加工,形成数据指标或数据报表,以 API 接口的方式提供给政府部门使用。

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档的，您的阅读或使用行为将被视为对本声明全部内容的认可。

- 1) 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本文档内容或提供给任何第三方使用。
- 2) 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
- 3) 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
- 4) 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以大数据集成服务平台的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
- 5) 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“AliCloud”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
- 6) 如若发现本文档存在任何错误，请与阿里云取得直接联系。



钉钉扫码加入
阿里云数据中台交流群



阿里云开发者“藏经阁”
海量免费电子书下载