

解决方案实践

# 翼海云峰大数据服务解决方案实践

文档版本 1.0  
发布日期 2024-07-15



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 安全声明

## 漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

---

## 目录

---

1 方案概述.....	1
2 资源和成本规划.....	4
3 实施步骤.....	8
4 修订记录.....	10

# 1 方案概述

## 应用场景

以互联网、零售、制造这三个行业为例，介绍客户的典型方案场景和痛点

### 场景一：互联网行业电商大数据迁移

#### 客户的痛点:

在不同云厂商间迁移，需要迁移脚本和作业，通常在不同厂商云之间，脚本的语法上存在兼容性和性能、配置等问题。

#### 解决方案:

- 提供专业的大数据迁移技术服务。
- 提供不同厂商云之间的脚本语法映射表。
- 提供不同厂商云的平台配置、和性能调优手册。

### 场景二：零售行业，数字化营销，大数据运营。

#### 客户的痛点:

- 如何高效选址
- 如何高效准确定位客户群体
- 如果高效为客户提供线上线下一体化的服务

#### 解决方案:

- 提供专业的大数据运营服务。
- 提供先进的大数据治理工具、平台。
- 提供高性能的用户画像、市场画像等大数据运营工具。

### 场景三：制造行业，电子元件，大数据运维

#### 客户的痛点:

在制造行业，企业通常需要时刻监控设备的运行情况。

#### 解决方案:

- 提供专业的大数据运维服务。

- 提供先进的大数据运维监控平台工具。
- 提供全部的大数据运维案例和预案。

本章节介绍如何基于华为云基座、华为云服务平台、华为云工具平台设计，提供从大数据咨询、调研到大数据运营分析的全技术栈和端到端的服务和交付能力。

## 方案架构

图 1-1 业务架构图



本架构基于华为云基座、华为云服务平台、华为云工具平台设计，提供从大数据咨询、调研到大数据运营分析的全技术栈和端到端的服务和交付能力。

## 方案优势

此解决方案本身具备如下优势：

- **方案灵活：**可拆分配置的8大服务模块，满足客户定制服务需求，节约客户开支，提高付费意愿。
- **一专多能：**以大数据迁移、接入、开发为主的技能优势为入口，拓展全场景的大数据服务能力。
- **通用模式：**7年+的大数据技术专业经验，打磨出面向底层方法论的通用大数据服务模式、模型。

此人力服务型解决方案依托深厚的公司实力、团队能力和技术能力：

- **公司实力：**

9年的行业经验，20+领域的覆盖，服务过1000+家的企业。发明专利10+项，各类ISO/政府/企业认证资质10+项，软件著作权50+项。公司投入巨大，大数据相关行业的项目管理经验丰富。

- **团队能力：**

专业人才充足：数据管理与分析服务团队200+人，共考取100+张大数据开发/数据仓库相关的能力认证证书。

服务分布广泛：分布北京、深圳、江苏、湖北等地，能及时响应全国项目的服务需求。

项目经验丰富：2022-2023年度服务10+家大数据客户。

- **技术能力：**

具备从大数据咨询、调研到大数据运营分析的全技术栈和端到端的服务和交付能力。具备7+年的大数据技术使用和服务经验。

## 约束与限制

此方案以且仅以华为云公有云、华为混合云为云基座，以且仅以ECS、OBS、VPC等华为云服务为服务基础，以且仅以MRS、DLI、CDM、DIS、DataArts Studio等华为云大数据套件为工具基础。

# 2 资源和成本规划

表 2-1 云服务成本规划

序号	云服务名称	云服务类型	规格	数量	计费模式	计费周期	参考价格
1	VPC	VPC名称: iotstage- vpc-1872911	/	1	无需 选择 计费 模式	无需选 择售 卖 周 期	--
2	SG	SG	安全组名称: iot_sg	1	无需 选择 计费 模式	无需选 择售 卖 周 期	--
3	ModelAr ts	MODELARTS	计算型 GPU(Vnt1 NVLINK 32G) 实例 modelarts.bm.g pu.v100NV32 购买时长: 1000h	1	包周 期	1年	2450 0
4	DLV	GeneralCloudS ervice	基础班	1	包周 期	1年	60
5	ServerGr oup	ECS.ServerGro up	云服务器组名 称: ServerGroup- d39fa3   策略: 反亲和	1	无需 选择 计费 模式	无需选 择售 卖 周 期	--



序号	云服务名称	云服务类型	规格	数量	计费模式	计费周期	参考价格
6	OBS	OBS	OBS桶名: iotstage- obs-8t8cie 数据 冗余存储策略: 多AZ存储 存储 类别: 标准存储 桶策略: 私有	1	按需	无需选择 售卖周期	0.139
7	RDS	RDS	数据库引擎: MySQL 数据库 版本: 8.0.28 实 例类型: 主备 性 能规格: 独享型 云盘SSD   16vCPUs   128GB 存储类 型: SSD云盘 磁 盘大小: 40GB	1	包周 期	1个月	6944
8	DataArtsStudio	DGC	版本 初级版 企 业项目: 0	1	包周 期	1个月	2000
9	DIS	GeneralCloudService	["分区:普通   分 区数量: 1","消 息数量:普通   平 均消息大小: 1KB   平均写入 消息数: 1"]	1	按需	无需选 择售卖 周期	0.1
10	CDM	CDM	实例规格: cdm.large   8核 16G   0.8/3 Gbps   20	1	按需	无需选 择售卖 周期	¥2.25 /小时
11	GesPackage	GESPackage	鲲鹏百万边图	1	包周 期	1个月	3000
12	CSS	CSS	集群版本:7.10.2 节点规 格:ess.spec-4u3 2g   4 vCPUs   32 GB 节点存储 类型:超高I/O   100 GB	1	包周 期	1个月	4566. 96

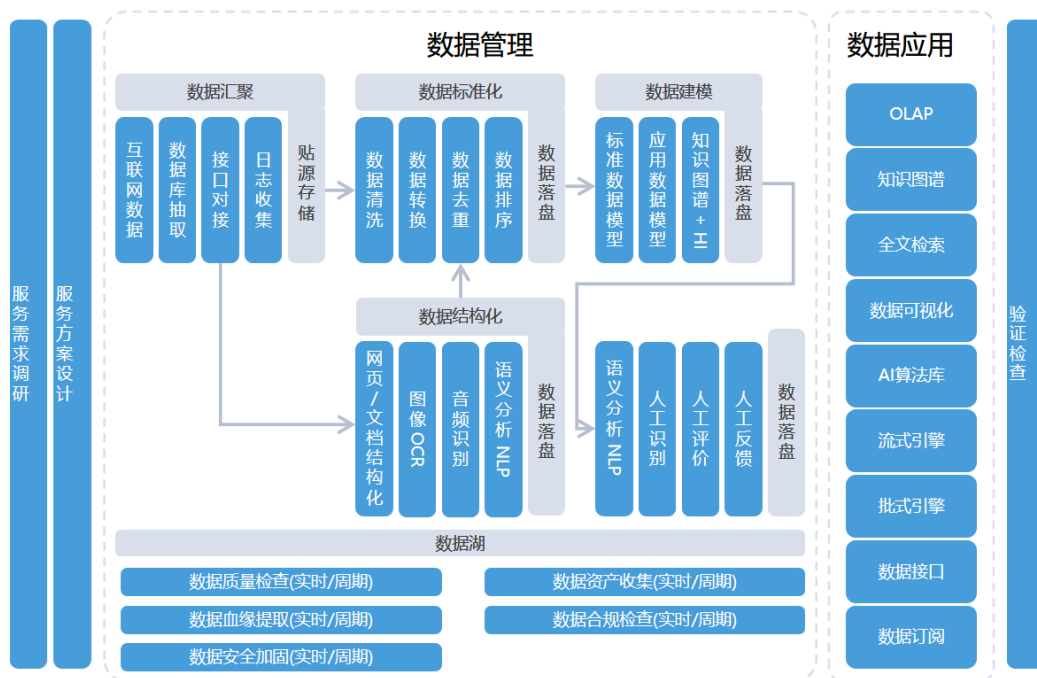
序号	云服务名称	云服务类型	规格	数量	计费模式	计费周期	参考价格
13	CloudTable	GeneralCloudService	["Front End计算规格:4U16G * 3节点","Front End存储规格:高IO   300GB * 3节点","BE计算规格:4U16G * 3节点","BE存储规格:高IO   400GB * 3节点"]	1	按需	无需选择售卖周期	8.829
14	DWS	DWS	产品类型: 标准数仓 计算类型: 弹性云服务器 存储类型: SSD 云盘 部署类型: 集群 CPU 架构: X86 节点规格: dwsx3.16U64G.16DPU   16 vCPUs   64 GB 热数据存储: 100 GB 企业项目: 0	1	包周期	1个月	8664
15	General CloudService	GeneralCloudService	["CUs:弹性资源池   64"]	1	包周期	1个月	10880

序号	云服务名称	云服务类型	规格	数量	计费模式	计费周期	参考价格
16	MRS	MRS	集群版本 MRS 3.1.0 集群类型 自定义集群（仅 MRS 3.x版本支持） 组件 [Hadoop, ZooKeeper, Ranger] 集群节点组 master_node_default_group 实例规格: ac7.4xlarge.4.linux.bigdata(16 U 64G) 系统盘: 超高IO 480GB X1 数据盘: 超高IO 600GB X1 实例数量: 3 node_group_1 实例规格: ac7.4xlarge.4.linux.bigdata(16 U 64G) 系统盘: 超高IO 480GB X1 数据盘: 超高IO 600GB X1 实例数量: 3	1	包周期	1个月	18983.04
17	ECS	ECS.CloudServer	ECS 名称 iotstage-ecs-fb1c38 CPU架构 x86计算规格 通用计算增强型   ac7.4xlarge.4   16vCPUs   64GiB 镜像来源 公共镜像 镜像 CentOS 7.6 64bit 系统盘 极速型SSD, 100GiB 数据盘 1块   极速型SSD, 100GiB	1	包周期	1个月	2184.32

# 3 实施步骤

本方案包括八大模块，每大模块的每个功能模块都有其独立的实施步骤

图 3-1 数据管理&数据应用



以大数据管理服务为例子，其实施步骤大体包含以下内容：

表 3-1 大数据管理服务

项目阶段	服务内容
服务需求调研	/
服务方案设计	/

项目阶段	服务内容
数据汇聚	包括 互联网数据采集 数据库数据抽取 接口对接 日志文件收集
数据标准化	数据清洗 数据转换 数据去重 数据排序
数据建模	标准数据模型 应用数据模型 知识图谱+HI专家智能
数据结构化	网页/文档的数据结构化 图像OCR 语音识别 语义分析NLP
数据质量检查与管理	实时/周期的数据质量检查 实时/周期数据质量报告 检查内容包括:准确性、一致性、重复性、实体完整性、参照完整性、用户自定义完整性等。
数据资产收集与管理	实时/周期数据资产收集 实时/周期数据资产报告
数据血缘提取与管理	实时/周期数据血缘提取 数据血缘地图维护 其提取的粒度为数据列
数据合规检查与管理	实时/周期数据合规检查 实时/周期数据合规报告 其内容包括法律遵从性、隐私保护、数据最小化、数据保留政策、数据主体权利、数据泄露管理等。
数据安全加固与管理	实时/周期数据安全检查 及时的数据安全加固 其内容包括访问控制、数据加密、数据备份、漏洞管理、安全审计、风险评估、数据脱敏等

# 4 修订记录

表 4-1 修订记录

发布日期	修订记录
2024-07-15	第一次正式发布。