

WORK INSTRUCTION 工作指示

JABIL	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
		SHEET (页码) Page 1 of 25

Work cell (工作坊) :

Current Revision Details (当前版本修订信息)

Rev (版本)	Originator: (制作人)	Revision summary (文件更改概述)
A	Anny Wu	Create on 5/15/2025

Both the functional departments that are involved in the document content and the departments implementing the document must take approval (文件内容涉及的职能部门 和 执行文件内容的部门都必须审批)

Dept. (部门)	Approver (批核人)	Dept. (部门)	Approver (批核人)	Dept. (部门)	Approver (批核人)
OPS		MFG		IT	Dave Pan
Business		Inventory Control		IQC	
Engineering		Facility		HR	
IE		EHS		Finance	
QA		PMO		Logistic	
QMS		Purchasing		Automotive Ops	
SQE		Planning			
Lean					

Remark: Departments can be added as needed (上表未陈列出的部门可根据需要添加)

Environmental health and safety considerations (环境健康安全注意事项):

Jabil Confidential, For Internal Use Only. Photography Is Prohibited Unauthorized Printed Copies Are Reference Only.
公司机密，仅限内部使用，禁止拍照。非授权打印件仅供参考。

Form #:

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 2 of 25

1.0 PURPOSE (目的)

Guidelines for data asset inventory. 数据资产盘点指引。

2.0 SCOPE (范围)

This document applies to Jabil Huangpu plant.

本文档适用于捷普黄埔工厂。

This document is intended for data management related work.

本文档适用于数据管理相关的工作。

3.0 DEFINITION (定义)

3.1 数据资产盘点是什么

数据资产盘点是指对组织内所有数据资源进行系统化的识别、整理和评估的过程，是对组织的数据资产进行全面梳理，包括以物理或电子形式记录的数据库表、数据项、数据文件等结构化和非结构化数据资产，明确数据资产基本信息和相关方，形成数据资产清单。所以数据资产盘点的成果是数据资产目录或数据资产清单，它从全局层面直观地展现数据处理者拥有的数据资产情况。

通过数据资产盘点，可以全面了解数据的来源、类型、质量、用途等信息，为数据治理和数据利用奠定基础。

3.2 为什么要做数据资产盘点

目前，捷普广州经过多年的信息化建设，已积累了种类繁多、体量庞大的数据，并且随着业务的持续发展，业务范围、资产规模、客户规模均不断扩大，产生的数据规模也在快速增长。

而数据规模的快速增长为企业带来了数字化转变、智能化变革的机遇，同时也对数据资产的掌控能力提出更高要求。而数据资产盘点是解决企业数据资产管理的前提，通过对企业拥有的数据进行盘点，将帮助企业弄清楚以下问题：

- 企业有哪些数据？关注数据的分类
- 企业有多少数据？关注数据的存量、增量
- 企业的数据存储在什么地方？关注数据的存储和取用方式
- 企业的数据是由谁在管理？关注数据的归属部门和责任人
- 识别哪些是重要数据，哪些是敏感数据？关注数据的分级、共享条件和范围

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 3 of 25

数据安全和数据价值是数据资产梳理的两个重要应用方向。一方面盘点后的数据可运用于数据库审计、防水坝、防勒索软件中，提高数据安全，且根据分级分类的结果，可以明确数据保护目标，有针对性地采取适当、合理的管理措施和安全防护措施，协助组织完成数据安全保护体系的构建；另一方面数据资产目录可以用于数据脱敏、元数据管理和主数据建设方面，从而提升组织的内部数据质量，发挥数据价值，优化数据治理。

因此，全面推进数据资产盘点工作，是摸清数据资产家底、明确数据资产存量、识别数据资产范围、搭建数据资产地图的重要手段，也是准确识别出有价值的数据资产，并对数据资产开展统一规范管理，进而实现数据资产价值最大化和良性循环的重要基础性工作。

3.3 数据资产盘点的目标是什么

提升数据质量：通过系统化的盘点，识别并清理低质量、重复或冗余的数据，提升整体数据质量。

提高数据利用效率：明确数据的分布和使用情况，帮助组织更高效地访问和利用数据资源。

支持数据决策：为数据分析和业务决策提供可靠的数据基础，提升决策的科学性和准确性。

促进数据合规：确保数据的管理和使用符合相关法律法规和行业标准，降低数据合规风险。

3.4 基本方法

数据资产盘点有两个方法可以帮助我们梳理出企业的数据资产清单或数据资产目录，分别是“自上而下梳理”和“自下而上盘点”。

1) 自上而下梳理

自上而下梳理是一种以业务视角进行数据梳理的方法，通过对企业的相关制度文件、智能体系、业务流程、业务单据等进行全面分析，逐层分解，梳理数据资产的三级目录、业务属性和相关管理属性。

- 三级目录：即数据资产的分类，是按照业务视角对企业数据资产的梳理和分解，例如：数据域-数据主题-数据子主题-数据对象（注：三级目录不限于三级，但一般建议控制在五级之内为宜）。
- 业务属性：即用来描述数据资产的业务元数据。常见业务属性包括：所属数据域、数据主题等分类属性，数据对象、业务定义、业务规则、敏感等级等。
- 管理属性：即用来描述数据资产的管理、维护、使用相关元数据。常见管理属性包括：管理部门、管理人员、联系方式、更新频率、最后更新时间、数据共享条件等（注：业务视角下，数据资产的管理属性可能无法全部梳理出来，这就需要在技术盘点环节对其进行补充完善）。

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 4 of 25

2) 自下而上盘点

数据资产盘点的另一个方面是以技术的视角，从 IT 系统—数据库表—数据结构出发，进行自下而上归纳，逐步明确数据资产相关的系统信息项（技术属性）。

技术属性：即用来描述数据资产的技术元数据。常见技术属性包括：来源系统、数据库表、字段类型、字段格式、取值范围、存储方式、血缘关系等。

最后，对业务视角的梳理出的目录中数据项与技术视角的盘点出的系统信息项进行关联，建立起两者的映射关系，这样一个完整的数据资源目录就成型了。通过数据资产目录可以从多个视角（业务或 IT）进行数据的查找，并确保目录中的每一个数据项都可以在真实的 IT 系统中找到。

3.5 基本原则

众所周知，数据作为一种特殊的资产，确权难、虚拟性、可复制是其主要特征，这也为数据资产盘点造成了一定的困难。数据资产盘点的原则是“谁生产，谁负责”，“谁使用，谁负责”，“谁管理，谁负责”。

因此，企业数据资产的盘点需要找到一个具有全局思维的人来进行统筹，规划出数据盘点的相关原则、框架和蓝图，定义出数据盘点的内容，制定出数据盘点的模板，再由生产或使用数据的业务部门执行梳理，完成数据盘点工作。这个统筹的人可以是 IT 部门、数据管理部门，或是外部聘请的数据专家。

除此以外，为了更好地开展企业数据资产盘点工作，保证盘点过程安全稳定、保证盘点结构准确可靠，还应遵循如下原则开展具体工作：

- 前瞻性：数据资产盘点还应该站在整个盘点旅程的角度进行规划和实践，充分考虑数据资产规范、搜索获取、分析应用、绩效评估、可视化展示等需求；
- 全面性：数据资产盘点范围要全面覆盖企业的所有数据资产，但在落地实践过程中，企业也可以分阶段开展盘点工作；
- 基础性：选择数据资产最稳定的本质属性或特征作为盘点内容，确保盘点内容不因环境因素而发生变化；
- 系统性：将需要盘点的数据资产的属性或特征按一定排列顺序以系统化，并形成一个合理的分类体系；
- 确定性：对于盘点范围内的任何一项数据资产，在分类体系中应该有唯一确定的基本单元与之相应；

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 5 of 25

- 可拓展性：目录框架应满足数据资产不断发展和变化的需求，允许在目录框架中增加新的盘点内容而不影响原有内容。为使用者进行延拓细化创造条件；
- 安全性：减少数据盘点工作对于业务活动和系统运行的影响，避免出现数据丢失或泄露等失误，给企业带来损失；
- 保密性：数据是企业重要资产，数据资产盘点过程中，盘点人员要严格遵守保密要求，避免触及敏感信息。

3.6 流程步骤

数据资产盘点工作，整体包括准备阶段、盘点阶段、结果输出三部分流程，其中盘点阶段根据所盘点系统的不同类型和特征，包含4个子流程、分别针对套装软件、自开发系统、分析系统、数据仓库开展盘点工作。

3.6.1 准确阶段

不同系统对于盘点文档的需求存在差异，盘点开始前，需对获取或生成的业务及技术文档准备情况进行核查，确认文档可供盘点使用。同时，向各盘点系统运维支持人员申请各系统查询权限并确认。准备阶段主要工作可分为：

1) 组建团队

成立跨部门小组（业务、IT、数据治理团队）。

2) 确认操作权限

由各系统运维支持人员提供各系统前台显示和后台数据库查询权限，盘点人员对所需用户权限在目标系统进行验证，为后续进行数据资产盘点打下基础。

3) 形成数据字典

系统数据字典是系统盘点过程中的重要依据，如技术支持人员能提供盘点系统的完整数据字典，由技术支持人员提供；如技术支持人员不能提供完整数据字典，需由技术支持人员通过数据库工具导出该系统完整数据字典。

4) 形成功能操作清单

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 6 of 25

针对系统功能文档不全的历史遗留系统，由盘点业务人员创建系统功能操作清单，支撑系统功能与业务分类的关系对照。

5) 进行业务分类

根据盘点系统的业务定位，参考组织已有的业务分类体系，补充完善业务分类体系。

6) 确认盘点文档完整

根据收集与自主生成的文档，确认盘点文档是否完整。

3.6.2 盘点阶段

3.6.2.1 定位数据资产

企业开展数据资产盘点的过程中，需要结合所盘点的系统情况，划定系统中需要进行盘点的数据表范围。原则上仅针对各系统在业务环节源端产生的基础表，以及终端产生直接应用结果的数据表进行盘点。

具体来讲，可以结合数据资产的概念圈定盘点范围，即对“企业在运营活动中形成的，由企业拥有、全过程可控，并能给企业带来价值的数据”开展盘点，当拥有、可控、具有价值三个条件全部满足时，识别为数据资产盘点的对象范畴。企业数据资产可分为如下几个类型：

(1) 基础表：系统直接产生、直接反应业务运营情况的基础数据表，同时具有由该系统拥有、可控、具备应用价值的特征，是数据资产盘点的主要内容。需要注意的是，根据数据资产概念，仅对该系统拥有的数据表进行盘点，调用其他系统的数据不进行盘点，例如分析系统中调用的业务数据不进行盘点。

(2) 代码表：系统中的代码相关数据，属于参考数据，需要进行盘点。

(3) 中间过程表：中间过程表由于具有临时性、变动性大的特点，不具备数据资产的可控性特征，且大数据分析、应用价值不高，因此不进行盘点。

(4) 报表、指标：属于分析数据，是系统在运营分析后产生的相关结果，具有分析、共享价值。盘点时需要根据情况进行识别，仅对业务系统和分析系统中的固定报表进行盘点，对于变化频率高、自定义的报表不进行盘点。

此外，针对系统中不同数据表的特殊情况，进行实际盘点工作时，可根据数据特征、数据应用场景、数据分析价值进行进一步识别。

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 7 of 25

因此，圈定需要盘点的数据资产范围，明确盘点对象清单，可以为后续进行数据资产盘点打下基础。具体执行即为数据源的梳理。

数据源梳理是数据资产识别的第一步，也是最关键的步骤之一。它就像是绘制一张企业数据地图，帮助我们找到所有数据的来源和存放位置。这个过程可能会让人感到繁琐，但却是不可或缺的。

3.6.2.1.1 常见数据源类型

在开始梳理之前，我们需要了解企业中常见的数据源类型。这些类型包括但不限于：

a) 业务系统:这是企业数据的主要来源。例如，ERP系统存储了大量的交易数据、财务数据和库存数据;CRM系统包含了客户信息、销售线索和商机数据;人力资源系统则涵盖了员工信息、薪酬数据等人力资源相关数据。

b) 数据仓库和数据湖:这些是企业为了更好地管理和分析数据而建立的集中存储和处理平台。比如，一家大型零售企业可能使用Amazon Redshift作为其数据仓库，存储来自各个销售渠道的汇总数据;而一家制造企业可能使用Hadoop数据湖来存储和分析来自物联网设备的海量传感器数据。

c) 文件服务器和共享文件夹:不要小看这些看似普通的存储位置。很多重要的数据可能以Excel表格、Word文档或PDF文件的形式存储在这里。例如，市场部门的调研报告、财务部门的月度报表，甚至是高管的战略规划文档，都可能存在与公司的SharePoint或Google Drive中。

d) 云存储服务:随着云计算的普及，越来越多的企业数据存储在云端。常见的服务如AWS S3、Google Cloud Storage、Microsoft Azure Blob Storage等，可能存储了从日志文件到备份数据等各种类型的信息。

e) 个人电脑和移动设备:虽然从数据治理的角度来看并不推荐，但现实是很多重要数据可能存储在员工的个人设备上。例如，销售人员的客户联系表，研发人员的创意草图，这些都可能存在与个人电脑或平板设备中。

f) 外部数据源:在数据驱动决策的时代，企业越来越多地利用外部数据来补充自身的数据资产。这可能包括从数据供应商购买的行业报告、通过API接入的社交媒体数据，或者是合作伙伴共享的数据集。

3.6.2.1.2 数据源梳理的实施步骤

了解了常见的数据源类型后，我们来看看如何进行全面的数据源梳理。这个过程需要IT部门、业务部门和数据管理团队的紧密合作。以下是一个详细的实施计划：

步骤 1:组建跨部门梳理小组

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 8 of 25

首先，我们需要组建一个跨部门的梳理小组。这个小组包括：

- IT 部门代表:了解公司的技术架构和系统布局
- 业务部门代表:了解各自部门的数据使用情况
- 数据治理团队:负责协调整个梳理过程
- 信息安全部队:确保梳理过程符合安全规范

步骤 2:制定数据源调查表

接下来，我们需要设计一份详细的数据源调查表。这份调查表应该包含以下关键信息：

- 数据源名称和类型
- 所属部门/业务线
- 数据描述(存储了什么类型的数据)
- 数据量估算
- 更新频率
- 访问方式和权限
- 负责人/联系人

举个例子，一份针对 ERP 系统的调查表可能如下：

● ● ●
数据源名称: SAP ERP系统
类型: 业务系统
所属部门: 财务部、供应链部
数据描述: 包含财务交易、采购订单、库存、生产计划等数据
数据量估算: 约500GB, 增长速度约50GB/年
更新频率: 实时更新
访问方式: SAP GUI客户端, 部分数据通过BI工具访问
负责人: 张三(IT部SAP管理员)

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 9 of 25

步骤 3:展开全面调查

有了调查表，下一步是在全公司范围内展开调查。这个过程可以分为几个阶段：

- a) IT 系统清单确认:**IT 部门首先提供一份公司所有 IT 系统的清单，包括内部部署的系统和云服务。
- b) 部门问卷调查:**向各个部门发放调查问卷，了解他们使用的系统和数据。这里要特别注意那些"影子 IT"，即未经 IT 部门批准私自使用的系统或工具。
- c) 关键人员访谈:**对于一些复杂的系统或数据密集型部门，可以安排面对面的访谈，深入了解情况。
- d) 技术扫描:**使用网络扫描工具，发现网络中的数据存储位置，这可以帮助发现一些被遗漏的数据源。

步骤 4:验证和补充

收集到初步信息后，我们需要进行验证和补充：

- a) 技术验证:**IT 团队需要验证每个报告的数据源的技术细节，如实际的数据量、数据结构等。
- b) 采样检查:**对于重要的数据源，可以进行数据采样，了解实际的数据内容和质量。
- c) 补充调查:**针对调查过程中发现的问题或疑点，进行补充调查。

在验证过程中，我们可能会发现之前报告的 ERP 系统实际数据量达到了 1TB，远超初始估计。这可能意味着我们需要重新评估数据存储和备份策略。

步骤 5:汇总形成数据源清单

最后，我们将所有收集到的信息汇总，形成一份全面的企业数据源清单。这份清单应该是动态的，需要定期更新。清单可以采用电子表格形式，也可以录入专门的数据目录工具中。

一个简化版的数据源清单可能如下所示：

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 10 of 25

数据源名称	类型	所属部门	数据描述	数据量	更新频率	负责人
SAP ERP	业务系统	财务、供应链	财务、采购、库存数据	1TB	实时	张三
Salesforce	CRM 系统	销售、客服	客户、销售、服务数据	500GB	实时	李四
共享文件服务器	文件存储	全公司	文档、报表、图片等	5TB	每日	王五
AWS S3 存储桶	云存储	IT 部门	日志文件、备份数据	10TB	每小时	赵六

3.6.2.2 数据分类标准制定

完成数据源梳理后，我们已经对企业拥有的数据资产有了一个全局的认识。下一步是建立一个科学、合理的数据分类标准。这就像是为图书馆里的书籍制定分类系统，目的是让每一份数据都能被正确识别、轻松找到，并得到恰当的管理。

3.6.2.2.1 数据分类的重要性

在深入讨论如何制定分类标准之前，我们先来看看为什么数据分类如此重要：

- 1) 提高数据查找和使用效率：清晰的分类让用户能够快速定位所需的数据，大大提高工作效率。
- 2) 支持数据治理：不同类别的数据可能需要不同的管理策略，如安全级别、保留期限等。
- 3) 优化存储和处理：通过分类，我们可以为不同类型的数据选择最合适的存储和处理方案。
- 4) 合规要求：很多行业规范和法律（如 GDPR）要求对某些类型的数据进行特殊管理。
- 5) 价值评估：分类有助于识别高价值数据，为数据价值评估提供基础。

3.6.2.2.2 数据分类维度的选择

在制定数据分类标准时，我们需要从多个维度来考虑。每个维度都像是观察数据的一个视角，综合这些视角，我们才能对数据有全面的认识。以下是一些常用的分类维度，我们会详细讨论每个维度，并给出具体的例子：

a) 业务领域维度

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 11 of 25

这个维度根据数据所属的业务功能或部门来分类。它帮助我们了解数据在业务中的用途和重要性。

示例分类:

- 销售数据
- 客户数据
- 财务数据
- 人力资源数据
- 生产数据
- 供应链数据

一家大型零售企业可能会这样细分其客户数据:

- 会员信息
- 购物历史
- 客户反馈
- 忠诚度计划数据

这样的分类有助于不同部门快速定位他们需要的客户相关信息。

b) 数据类型维度

这个维度基于数据的结构和形式来分类，有助于选择合适的存储和处理方法。

分类:

- 结构化数据:如数据库表格中的数据
- 半结构化数据:如 JSON、XML 文件
- 非结构化数据:如文本文档、图片、视频

一家医疗器械公司的数据可能包括:

- 结构化数据:患者基本信息、设备参数记录
- 半结构化数据:设备生成的诊断报告(JSON 格式)
- 非结构化数据:医生手写笔记的扫描件、X 光图像

了解这些数据类型有助于公司选择合适的数据存储和分析工具。

c) 敏感度维度

这个维度关注数据的机密性和重要性，直接影响数据的访问控制和保护措施。

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 12 of 25

分类:

- 公开数据:可以自由传播的信息
- 内部数据:仅供公司内部使用
- 机密数据:只有经过授权的人员才能访问
- 高度机密数据:对公司至关重要, 需要最高级别保护的数据

一家科技公司可能这样分类其数据:

- 公开数据:产品说明书、公司新闻稿
- 内部数据:员工通讯录、内部培训材料
- 机密数据:在研产品规格、客户合同
- 高度机密数据:核心算法、未公布的财务报告

这种分类有助于公司实施分层的数据保护策略, 确保敏感信息的安全。

d) 时间维度

这个维度考虑数据的时效性和历史价值, 影响数据的存储策略和使用方式。

分类:

- 实时数据:需要立即处理的当前数据
- 近期数据:最近一段时间(如 3 个月内)的数据
- 历史数据:较旧但仍有分析价值的数据
- 归档数据:很少使用但需要长期保存的数据

一家电子商务公司可能这样管理其销售数据:

- 实时数据:当前库存水平、实时订单状态
- 近期数据:过去 30 天的销售趋势
- 历史数据:过去几年的季节性销售模式
- 归档数据:5 年前的交易记录(出于法规遵从需要保存)

这种分类有助于公司优化数据存储策略, 如将实时数据存储在快速访问的系统中, 而将归档数据转移到成本较低的存储介质。

e) 数据质量维度

这个维度评估数据的可靠性和完整性, 影响数据在决策中的使用。

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 13 of 25

分类:

- L3 数据: 经过严格验证的高质量数据
- L2 数据: 质量良好但可能存在小问题的数据
- L1 数据: 质量一般, 使用时需谨慎的数据
- 待确认数据: 质量未知, 需要进一步验证的数据

一家市场研究公司可能这样评估其调研数据:

- L3 数据: 大样本、多轮验证的消费者调查结果
- L2 数据: 单一来源的行业报告数据
- L3 数据: 社交媒体抓取的用户评论
- 待确认数据: 新开发的数据采集渠道收集的信息

这种分类有助于分析师了解数据的可靠性, 在制作报告时合理使用不同质量等级的数据。

f) 使用频率维度

这个维度反映数据被访问和使用的频繁程度, 有助于优化存储和缓存策略。

分类:

- 高频数据: 每天多次访问
- 中频数据: 每周或每月访问几次
- 低频数据: 每季度或每年访问几次
- 冷数据: 极少被访问, 但仍需保留的数据

一家保险公司可能这样分类其数据:

- 高频数据: 客户基本信息、当前保单状态
- 中频数据: 月度理赔统计、代理人业绩数据
- 低频数据: 年度精算报告、客户满意度调查结果
- 冷数据: 已结束的保单历史记录

这种分类可以帮助公司设计分层存储架构, 将高频数据存储在快速但昂贵的存储系统中, 而将罕用数据转移到更经济的存储介质。

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 14 of 25

3.6.2.2.3 制定数据分类标准的步骤

了解了常见的分类维度后，我们来看看如何制定一个全面而实用的数据分类标准。这个过程需要多方参与，反复讨论和优化。以下是一个详细的步骤指南：

步骤 1: 成立数据分类工作组

首先，我们需要组建一个跨部门的工作组，包括：

- 业务代表:了解各部门的数据需求和使用场景
- IT 代表:了解技术实现的可能性和限制
- 数据治理专家:提供数据管理的最佳实践
- 合规代表:确保分类标准符合法律和合规要求
- 信息安全专家:提供数据安全相关的建议

步骤 2: 调研现有分类实践

在制定新标准之前，我们需要了解企业现有的分类方法：

- 收集各部门现有的数据分类方法
- 调研行业最佳实践和标准
- 考虑相关法规要求

财务部门可能已经有一套成熟的数据分类方法，可以作为整体分类标准的一部分。

步骤 3: 确定分类维度

基于前期调研和企业需求，选择适合的分类维度：

- 选择对企业最重要的 3-5 个维度
- 确保选择的维度能覆盖主要的管理需求
- 考虑维度之间的关系，避免重复或冲突

一家金融机构可能选择以下维度：

- 业务领域(如零售银行、公司银行、投资银行)
- 数据敏感度(公开、内部、机密、高度机密)
- 数据类型(结构化、半结构化、非结构化)

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 15 of 25

- 数据质量(L3、L2、L1、待确认)

步骤 4: 定义分类类别

为每个选定的维度定义具体的分类类别:

- 确保类别定义清晰，易于理解
- 类别之间应该相互排斥，不重叠
- 类别应该能够覆盖所有可能的情况
- 控制类别数量，通常每个维度不超过 5-7 个类别

对于"数据敏感度"维度，我们可以定义如下类别:

- 公开数据:可以自由传播的信息，如产品目录
- 内部数据:仅供公司内部使用，如内部通讯录
- 机密数据:只有经过授权的人员才能访问，如客户财务信息
- 高度机密数据:最敏感的信息，如并购计划

步骤 5: 制定分类规则

为每个类别制定明确的分类规则:

- 描述如何判断数据属于哪个类别
- 提供具体的示例和反例
- 考虑特殊情况和边界条件

对于"机密数据"类别，规则可能包括:

- 包含客户个人识别信息的数据
- 未公开的财务报表和预测
- 正在申请的专利信息
- 重要的商业合同

步骤 6: 制定分类操作指南

创建一个详细的操作指南，帮助数据所有者和使用者正确分类数据:

- 分类流程图:展示如何逐步判断数据类别
- 决策树:帮助快速确定数据的分类
- 常见问题解答(FAQ):解答分类过程中可能遇到的疑问

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 16 of 25

3.6.2.3 数据资产清单全景图

完成数据源梳理和分类标准制定后，我们就具备了建立全面数据资产清单的基础。数据资产清单就像是企业数据资产的“资产负债表”，它全面展示了企业拥有哪些数据、这些数据的特征是什么、价值如何等信息。一个高质量的数据资产清单，是实现有效数据治理和价值挖掘的关键工具。

3.6.2.3.1 数据资产清单的重要性

在深入讨论如何建立数据资产清单之前，我们先来看看为什么它如此重要：

- 1) 全局视图:提供企业数据资产的全景图，帮助管理层了解数据资产的全貌。
- 2) 决策支持:为数据相关的战略决策提供依据，如数据治理投资、数据集成项目等。
- 3) 合规管理:帮助识别和管理敏感数据，支持隐私保护和监管合规。
- 4) 效率提升:使数据使用者能够快速找到所需的数据，避免重复工作。
- 5) 价值发现:通过系统化整理数据资产，有助于发现潜在的数据价值。
- 6) 风险管理:帮助识别数据相关的风险，如数据质量问题、安全隐患等。

3.6.2.3.2 数据资产清单的关键要素

一个完善的数据资产清单应该包含以下关键信息：

a) 基本信息

- **数据集名称:**准确描述数据集内容的名称
- **唯一标识符:**每个数据集的唯一 ID
- **描述:**对数据集内容和用途的简要说明
- **数据所有者:**负责该数据集管理的部门或个人
- **数据管理员:**日常维护数据的负责人

b) 技术信息

- **数据源:**数据的来源系统或平台
- **存储位置:**数据的物理或逻辑存储位置

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 17 of 25

- **数据量:**数据集的大小和记录数
- **更新频率:**数据的更新周期
- **数据格式:**如 CSV、数据库表、JSON 等

c) 分类信息

- **业务分类:**所属的业务领域或功能
- **敏感度分类:**数据的机密级别
- **其他分类:**根据企业定义的分类标准添加

d) 质量信息

- **质量评级:**数据的整体质量评分
- **完整性:**数据的完整程度
- **准确性:**数据的准确程度
- **一致性:**跨系统数据的一致程度

e) 使用信息

- **主要用户:**数据的主要使用部门或系统
- **使用目的:**数据的主要用途
- **访问方式:**如何获取和访问数据
- **使用限制:**使用数据时需要注意的限制条件

f) 合规信息

- **数据主体:**涉及的个人或实体类型
- **保留期限:**数据需要保留的时间
- **合规要求:**适用的法律法规要求
- **同意管理:**是否需要数据主体的同意

g) 关联信息

- **上游数据:**该数据集的数据来源
- **下游数据:**使用该数据集的其他数据集或系统
- **相关文档:**与该数据集相关的文档链接

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 18 of 25

以一家银行的"客户主数据"为例，其数据资产清单条目可能如下：

基本信息：

- 名称: 零售客户主数据
- ID: CUS001
- 描述: 包含所有零售银行客户的核心信息
- 所有者: 客户关系管理部
- 管理员: 张三(数据治理团队)

技术信息：

- 数据源: CRM 系统
- 存储位置: Oracle 数据库 CUST_MASTER 表
- 数据量: 约 1000 万条记录, 50GB
- 更新频率: 实时更新
- 数据格式: 关系型数据库表

分类信息：

- 业务分类: 客户管理
- 敏感度分类: 机密
- 数据类型: 结构化数据

质量信息：

- 质量评级: 4.5/5
- 完整性: 98%
- 准确性: 95%
- 一致性: 与信用卡系统存在 2% 的不一致

使用信息：

- 主要用户: 营销部, 风控部, 客服中心
- 使用目的: 客户画像, 风险评估, 服务个性化
- 访问方式: 通过数据服务层 API 访问
- 使用限制: 禁止导出完整客户列表

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 19 of 25

合规信息:

- 数据主体: 个人客户
- 保留期限: 客户关系终止后 7 年
- 合规要求: 遵守 GDPR 和本地银行业监管要求
- 同意管理: 需要客户明确同意才能用于营销目的

关联信息:

- 上游数据: 客户开户系统, 网上银行注册系统
- 下游数据: 客户细分系统, 营销活动管理系统
- 相关文档: 《客户数据管理规范》, 《数据质量报告-202306》

这个例子展示了一个全面的数据资产清单条目, 涵盖了数据的各个重要方面。

3.6.2.3.3 建立数据资产清单的步骤

如何系统地建立数据资产清单, 这个过程需要多方协作, 并且可能需要持续的努力才能完成。以下是一个分步骤的指南:

步骤 1: 准备工作

a) 组建团队:

- 数据治理专家:负责整体协调
- IT 代表:提供技术支持
- 业务代表:提供业务视角
- 法务代表:确保合规性

b) 确定范围:

- 决定优先处理哪些数据域或系统
- 设定清晰的完成目标和时间表

c) 选择工具:

- 可以使用电子表格(如 Excel)开始
- 考虑专业的数据目录工具(如 Collibra, Alation)

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 20 of 25

一家中型制造企业可能决定先从生产数据和客户数据开始，计划在 3 个月内完成这两个领域的数据资产清单。

步骤 2: 信息收集

a) 利用前期工作:

- 参考数据源梳理的结果
- 应用已制定的数据分类标准

b) 技术扫描:

- 使用数据发现工具扫描数据库和文件系统
- 收集技术元数据(如表结构, 数据量等)

c) 业务调查:

- 设计调查问卷, 收集业务相关信息
- 与关键人员进行访谈, 了解数据的使用情况

d) 文档审查:

- 审查现有的系统文档, 数据字典等
- 收集相关的业务和技术文档

例如, 对于客户数据, 可能需要与 CRM 系统管理员进行技术对接, 同时与销售和客服部门进行访谈, 了解数据的业务用途和重要性。

步骤 3: 信息整理和验证

a) 数据整合:

- 将从不同渠道收集的信息整合到一起
- 确保信息的一致性和完整性

b) 初步填写:

- 根据收集到的信息, 初步填写数据资产清单

c) 信息验证:

- 与数据所有者和主要使用者确认信息的准确性
- 解决信息不一致或缺失的问题

d) 质量评估:

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 21 of 25

- 进行初步的数据质量评估
- 记录已知的数据质量问题

在整理客户数据信息时，可能发现销售部门和客服部门对数据使用目的的描述有所不同，需要进一步沟通协调，达成一致的描述。

步骤 4: 丰富和完善

a) 补充高级信息:

- 添加数据血缘关系信息
- 评估数据的业务价值
- 明确数据的法律和合规要求

b) 标准化描述:

- 使用统一的术语和格式
- 确保不同数据集的描述风格一致

c) 链接相关资源:

- 关联相关的业务术语表
- 链接数据质量报告，使用指南等文档

d) 审核和批准:

- 由数据治理委员会或相关负责人审核
- 正式确认数据资产清单的有效性

对于客户主数据，可能需要明确其与交易数据，营销数据的关系，评估其在客户服务和风险管理中的价值，并确保其处理符合 GDPR 等隐私法规的要求。

步骤 5: 持续维护和更新

a) 建立更新机制:

- 设定期定期审查的时间表(如每季度)
- 建立变更通知机制，及时捕捉数据变化

b) 指定负责人:

- 为每个数据域或系统指定维护负责人
- 明确职责和工作流程

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 22 of 25

c) 自动化支持:

- 利用数据目录工具的自动更新功能
- 建立与源系统的元数据同步机制

d) 用户反馈:

- 建立用户反馈渠道，收集使用中发现的问题
- 定期调查用户满意度，持续改进

e) 考核指标:

将数据质量（错误率下降目标）、资产利用率（高价值数据调用次数）纳入部门绩效考核。

可以设立一个数据资产管理小组，负责每月审查数据资产清单的更新情况，处理用户反馈，并协调必要的更新工作。

3.6.3 结果输出

盘点结果输出如下文档：

- 数据源清单
- 数据资产清单
- 数据质量报告

3.7 常见问题与解决方案

问题 1：数据分散在员工个人设备

方案：制定数据存储规范，强制上传至企业云盘，配合审计工具监控。

问题 2：跨系统数据冲突

方案：建立主数据管理（MDM）平台，统一字段定义。

问题 3：盘点耗时过长

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 23 of 25

方案：采用增量盘点（仅扫描变动数据）+自动化工具，缩短周期 50%以上。

4.0 RESPONSIBILITY (责任)

在数据盘点的整个过程中，需要多种角色的参与，每种角色的职责如下：

- **数据 Owner**: 任命数据盘点的业务对接人（一般为数据管家）；组织、推动数据盘点工作；组织数据质量问题闭环管理。
- **数据管家**: 代表数据 Owner 组织协调数据盘点工作；梳理本部门的数据的分类分级制度；配合数据保管团队梳理数据源清单，整理系统的功能清单；配合数据治理团队梳理数据资产清单；协调部门内进行数据质量问题闭环处理。
- **数据保管者**: 整理数据源清单和数据表清单；整理数据字典；配合数据管家及数据治理团队进行数据查看和数据抽样工作；配合数据治理团队整理数据资产清单。
- **数据治理团队**: 统筹数据盘点工作；协助数据管家梳理数据的分类分级制度；协助数据保管者整理数据源清单；整理数据资产清单。

5.0 ASSOCIATED DOCUMENTS (相关文件)

6.0 PROCESS (流程)

7.0 Reaction Plan(应急计划)

8.0 OTHERS (其他)

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
	SHEET (页码)	Page 24 of 25

9.0 RECORD RETENTION (记录保存)

Record No (记录表格编号)	Record title (记录的标题)	Owner (责任部门)	Retention Time (保存时间)	Storage Location (存储位置)	Medium (存储方式)	File/Index method (归档/索引方法)

10.0 Revisions History & Change Details (版本历史&变化信息) :

Keep at least 3 revisions (至少保留 3 个版本)

Rev (版本)	Originator: (制作人)	Prepared Date (编写日期)	Revision summary (文件更改概述)

Category of distribution(文件分发类型)

Hard copy (硬版本)	S/N	Dept./WC (部门/工作坊)	QTY. (数量)
	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		

WORK INSTRUCTION 工作指示

	DOCUMENT TITLE (文件名称): Data Inventory Work Instruction	DOCUMENT NO (文件编号): 不要编辑
		RELEASE DATE (发行日期) 不要编辑
		SHEET (页码) Page 25 of 25

Paperless (无纸化)	S/N	Dept./WC (部门/工作坊)	Process name/Station Code (If applicable) (工序名称/工作站代码) (如适用)
	1		
	2		
	3		
	4		
	5		
	6		

JABIL Confidential, For Internal Use Only,
 Photography Is Prohibited .
 公司机密,仅限内部使用,禁止拍照。