

数据中台解决方案

目 录

| | | |
|-------|----------------|----|
| 1. | 大数据管理的现状 | 1 |
| 2. | 数据中台的概述 | 6 |
| 2.1 | 数据中台概念..... | 8 |
| 2.2 | 数据中台目标 | 9 |
| 3. | 数据中台体系 | 10 |
| 3.1 | 数据连接 | 13 |
| 3.2 | 数据建模 | 14 |
| 3.3 | 统计报表..... | 14 |
| 3.4 | 决策仪表盘..... | 14 |
| 3.5 | 数据大屏 | 15 |
| 4. | 数据中台核心领域..... | 15 |
| 4.1 | 大数据模型..... | 20 |
| 4.2 | 大数据生命周期..... | 20 |
| 4.3 | 大数据标准 | 22 |
| 4.4 | 主大数据 | 24 |
| 4.5 | 大数据质量 | 25 |
| 4.6 | 大数据服务..... | 27 |
| 4.7 | 大数据安全 | 27 |
| 5. | 数据中台保障机制..... | 28 |
| 5.1 | 制度章程..... | 28 |
| 5.1.1 | 规章制度..... | 28 |
| 5.1.2 | 管控办法..... | 29 |
| 5.1.3 | 考核机制..... | 29 |
| 5.2 | 数据中台组织 | 30 |
| 5.2.1 | 组织架构..... | 30 |
| 5.2.2 | 组织层次..... | 31 |
| 5.2.3 | 组织职责 | 32 |
| 5.3 | 流程管理 | 34 |

| | | |
|-------|----------------|----|
| 5.4 | IT技术应用 | 35 |
| 5.4.1 | 支撑平台 | 35 |
| 5.4.2 | 技术规范 | 37 |
| 附件A | 大数据管理规范..... | 39 |
| 附件B | 大数据质量评估办法..... | 57 |
| 附件C | 大数据质量管理流程..... | 61 |

关于数据中台的理解

1. 大数据管理的现状

数据中台不是产品，而是为了让数据更快、更省用起来的一些列产品组件而成的数据产品矩阵与解决方案。企业在数据中台解决方案规划时，要基于目前数据在采、存、管、治、用各个环节的痛点，进行针对性的降本提效建设。数据中台是不是YYDS，能解决业务痛点的，才是王道，说不定，几年之后又出现了新的名词，现有的产品体系是否可以更快的升级适应呢。

▶ 企业数据化管理的痛点



1. 指标口径不统一

产品部门和财务部门一起开会给老板汇报，APP 下单用户数产品1021W, 财务1000W, 产品说我的数据是数据团队出的，财务说我的也是，那数据为什么不一致呢？原因数据开发A 给运营出的报表，

按照业务的口径以设备ID去重，数据开发B，给财务出的报表是按照userID(注册会员id)统计，存多设备登录的情况

2. 数据质量差

指标表现异常，业务第一反应就是“是不是数据不准啊”，这时作为数据部门如何能够有底气来反驳这种DISS呢？数据业务系统同步到数仓，ETL加工，再输出到报表应用，会经过多个步骤，每一个步骤都有可能会出现任务的异常、延迟以及人为的bug，监控覆盖足够健全，业务反馈问题时，数据开发就可以自信的说，今天数据无异常(没有收到报警)，而不是我先确认下。

3. 数据重复建设

缺少统一的数仓建设和管理规范，CaseByCase地响应业务需求，往往会导致数据的重复建设，例如，数据开发A接到产品的大盘流量报表需求，直接基于ODS的明细数据进行ETL，加工出自己的为了满足这一报表需求的APP层表，数据开发B，接到会员营销的需求，报表指标不尽相同，小A的APP层表无法直接使用，于是自己又加工了新的数据表，由此，导致相同指标多个模型出现，但又无法复用，造成重复建设

4. 数据找不到

业务发展加上数据的重复建设，数据表的数量在10W+，缺少工具的指引，尤其是新用户很难找到需要的数据在哪个表里，处理逻

辑是不是自己需要的

5. 数据成本增长快

随着业务需求发展，数据处理所需要的存储和计算成本也线性或指数增长，对于DAU千万级的互联网公司，每个月大数据集群的资源成本可能也在百万~千万级，是真正的成本中心了。往往一线数据开发很多只关注新增业务，不去梳理历史任务，或者一些低效的SQL任务占据了大量的资源。

6. 数据报表开发周期长

定制化的数据可视化报表开发需要数据开发、接口开发、前端开发，产品迭代、活动上线节奏非常快，都需要对应的报表监控支持，单个报表的开发周期往往在1~2周，对开发资源的依赖导致需求响应周期长，很多时候报表上线了，活动结束了。

7. 数据需求响应慢

对于无SQL的业务人员很多探索性的数据分析依赖于数据开发的SQL取数，一般SQL取数都是由数仓兼职进行，时间排期就有限，只能按照提需时间或者紧急需求的申请通道进行处理，临时取数的时效性要求更高，经常出现数据输出了，业务意见拍脑袋做完决策了。可能有人问可不可以安排全职取数，对于有个人追求的程序员，一直做SQL取数，估计很快就要离职了。

8. 数据服务难追踪

数据部门会输出很多的API 接口，由于历史久远文档不完善加上业务不断调整变化，导致接口和应用链路断层，接口出问题只能由业务反馈后处理。梳理出流量小的接口要做下线，却找不到应用端的人确认，只能先下线看下，有人反馈再处理。

9. 数据输出效率影响运营频率

精细化运营背景下，用户运营每个营销场景需要最精准的确定目标人群，比如会员生日关怀、迪士尼目标用户群体投放等，业务需要先找数据部门获取目标用户的id 信息，再进行投放，数据部门的响应周期和效率制约了运营活动的投放频次，即数据每周可以处理3~7次人群调取，那运营活动肯定不能超过这个频率。

根据行业信息化发展的现状，结合当今行业数据中台的要求，大型集团或政务管理部门现阶段大数据管理方面存在以下的不足：

(1) 大数据多头管理，缺少专门对大数据管理进行监督和控制的组织。信息系统的建设和管理职能分散在各部门，致使大数据管理的职责分散，权责不明确。组织机构各部门关注大数据的角度不一样，缺少一个组织从全局的视角对大数据进行管理，导致无法建立统一的大数据管理规程、标准等，相应的大数据管理监督措施无法得到落实。组织机构的大数据考核体系也尚未建立，无法保障

大数据管理标准和规程的有效执行。

(2) 多系统分散建设，没有规范统一的省级大数据标准和大数据模型。组织机构为应对迅速变化的市场和社会需求，逐步建立了各自的信息系统，各部门站在各自的立场生产、使用和管理大数据，使得大数据分散在不同的部门和信息系统中，缺乏统一的大数据规划、可信的大数据来源和大数据标准，导致大数据不规范、不一致、冗余、无法共享等问题出现，组织机构各部门对大数据的理解难以应用一致的语言来描述，导致理解不一致。

(3) 缺少统一的主大数据，组织机构核心系统间的人员等主要信息并不是存储在一个独立的系统中，或者不是通过统一的业务管理流程在系统间维护。缺乏对集团公司或政务单位主大数据的管理，就无法保障主大数据在整个业务范围内保持一致、完整和可控，导致业务大数据正确性无法得到保障。

(4) 缺乏统一的集团型大数据质量管理流程体系。当前现状中大数据质量管理主要由各组织部门分头进行；跨局跨部门的大数据质量沟通机制不完善；缺乏清晰的跨局跨部门的大数据质量管控规范与标准，大数据分析随机性强，存在业务需求不清的现象，影响大数据质量；大数据的自动采集尚未全面实现，处理过程存在人为干预问题，很多部门存在大数据质量管理人员不足、知识与经验不够、监管方式不全面等问题；缺乏完善的大数据质量管控流程和

系统支撑能力。

(5) 大数据全生命周期管理不完整。目前，大型集团或政务单位，大数据的产生、使用、维护、备份到过时被销毁的大数据生命周期管理规范 and 流程还不完善，不能确定过期和无效大数据的识别条件，且非结构化大数据未纳入大数据生命周期的管理范畴；无信息化工具支撑大数据生命周期状态的查询，未有效利用元大数据管理。

2. 数据中台的概述

通过灵活的数据交互和探索分析能力，以及 OEM 白标集成的方式，全面满足行业应用软件的数据分析需求。它提供多源数据整合、报表统计、数据可视化、自助式BI分析、以及数据填报等功能，帮助用户挖掘数据的潜在价值，为管理者制定决策提供数据支撑。

多数据源整合，为决策分析提供完整数据支持

通过数据查询设计器，有效整合您分散在企业内外的各种数据。包括数据库、云端数据、本地存放的文件数据以及 JSON/OData 等程序数据。既能通过拖拽操作完成跨源的数据建模，也支持直接编写查询语句。最终，通过数据模型访问控制和行级数据安全管控，分享给 BI 分析或报表统计者使用。

自助式 BI, 强大的交互分析和自由数据探索能力

自助式 BI, 让最终用户毫无约束的与数据交互，任意探索数

据背后的真正原因，发觉价值，为决策制定找到有效的数据支撑。

在 **Dashboard** 的设计和分析阶段，均提供图表联动、数据钻取、数据切片器、**OLAP** 等交互式分析功能，仅需通过极少的操作便能找到最有价值的数 据，并能与企业成员进行协同讨论分析。

数据可视化，让数据之间的关系更加清晰明了、便于理解

具有多屏自适应能力的仪表板，内置丰富的数据可视化类型，而且开放的数据可视化插件功能，几乎可以将任意的可视化组件库集成到产品中，比如：Echarts, D3, ChartJS, 三维模型等，充分满足数据可视化大屏等场景的需要。

在线报表，最终用户和实施人员都能设计报表

创新的在线报表设计功能，类似微软 Office 产品的使用体验，功能丰富却极易上手。从此，对软件公司而言，客户定制化报表需求再也不用修改软件源代码；对企业 IT 部门而言，让业务部门自主设计所需报表的畅想变得可能。

嵌入式分析，满足软件项目的深度集成和 OEM 合作需要

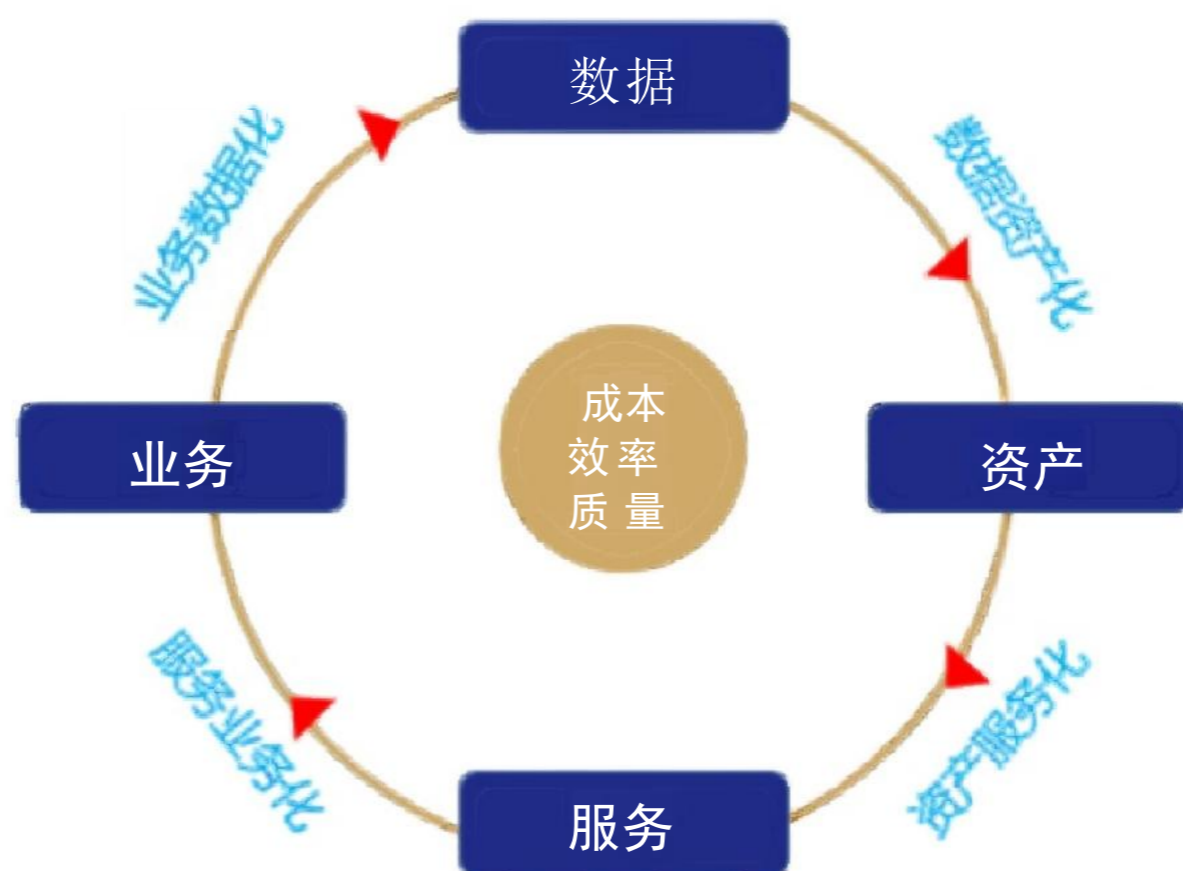
几乎可以将 **Wyn Enterprise** 的全部功能集成到您自己的软件产品中。对您的用户而言，他们完全感觉不到我们的存在；对您的技术人员而言，我们提供静默安装、软件界面定制、单点登录、报表/仪表板的设计器和查看器的集成等开箱即用的功能，大大节省系统集成所花的时间。

2.1 数据中台概念

数据中台是指将大数据作为组织资产而展开的一系列的具體化工作，是对大数据的全生命周期管理。

数据中台的思想

让数据更快、更省地用起来



关于数据中台的定义和概念，已经被讲烂了，结合近三年的数据中台实践，总结一下就是“让数据更快、更省地用起来”的一种思想、架构。也就是，数据中台所做的一切，最终的目标都是数据价值的挖掘和应用输出，为了达到这一目标，涉及数据的采、存、管、治、用各个环节和流程，可以用来“降本增效”的产品，都归属于数据中台产品体系。

在数据中台概念清晰之前，各个互联网公司其实也都做了很多的基础建设工作，只是没有明确地定义为数据中台而已。每个公司都在实践中寻找解决数据应用实践方法，例如构建指标体系解决指标口径不一致的问题；建设自助取数工具，业务自助取数不求人，开

发人力释放专注于数仓模型建设；开发配置化的BI 可视化产品，减少可视化报表对接口开发、前端开发人力的依赖；建设精准营销(DMP)平台，业务自助圈选目标用户进行精准触达，提升运营活动频率等。所以，个人理解，数据中台概念的出现，只是提供了一套完整的解决方案和思想，把原来的不成体系的“野路子“，扣上”中台“的帽子后，成了有方法论、战略的指引和支撑正规军了。

可以把数据中台类比成汽车工厂，如果发动机、轮胎等零配件已经生产完毕，可以很快组装出一辆汽车。而Hadoop 生态，集群建设，就像水电煤等基础设施，提供工厂运行所需能源支持，大数据平台，数据开发工具就像是机床设备，提供制造零配件的工具能力，而数据仓库的建设，则像是用机床加工好各自零配件，并且提供快捷的仓库索引目录，能够最短时间找到所需配件。

数据中台体系是指从组织架构、管理制度、操作规范、IT 应用技术、绩效考核支持等多个维度对组织的大数据模型、大数据架构、大数据质量、大数据安全、大数据生命周期等各方面进行全面的梳理、建设以及持续改进的体系。

2.2 数据中台目标

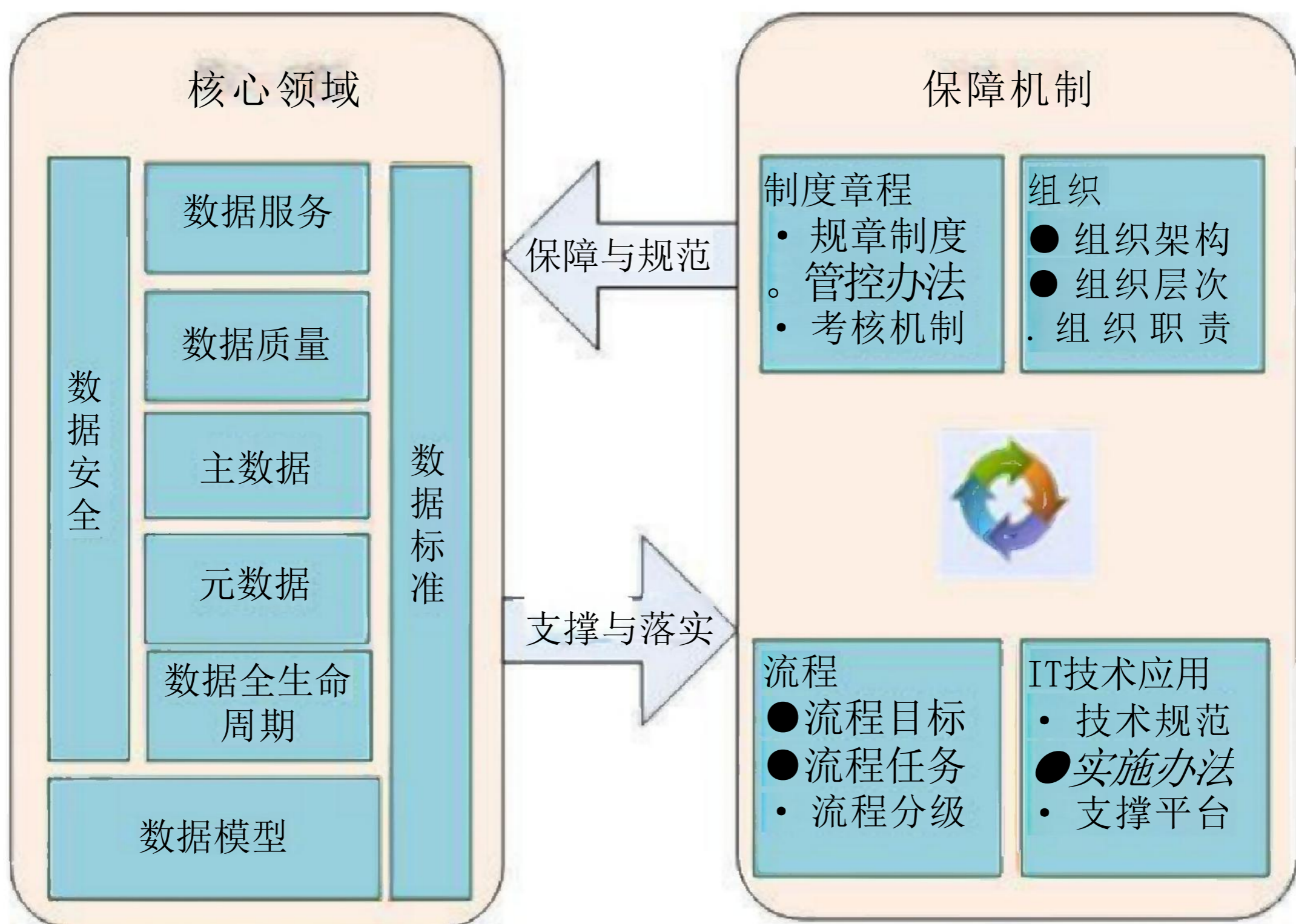
数据中台的目标是提高大数据的质量(准确性和完整性),保证大数据的安全性(保密性、完整性及可用性),实现大数据资源在各组织机构部门的共享;推进信息资源的整合、对接和共享,从而提

升集团公司或政务单位信息化水平，充分发挥信息化作用。

3. 数据中台体系

数据中台体系包含两个方面，一是大数据质量核心领域，二是大数据质量保障机制。

具体两者内容及相互关系可以参见下图：



数据汇聚

将异构数据源通过源和目标参数配置实现数据入湖、入仓，以及存储介质的转换，降低人肉脚本处理带来的风险和维护成本。构建统一的数据集散中心，打破数据孤岛。

资产沉淀

将数据提纯加工，形成可快速使用的数据模型，建立完善的数据共享机制与安全管控流程，构建数据复用能力。同时需要对资产进行常态化、周期性的质量管控与治理。

产品化能力

数据采集、资产管理、数据应用流程的平台化、配置化，基于工具实现数据的快速流转，提升数据输出的效率。

业务赋能

数据驱动决策、为产品智能化、运营精细化赋能。 一是赋能效率的提升，二是赋能过程的数据资产管控。

数据中台四大核心能力



通过灵活的数据交互和探索分析能力，以及 OEM 白标集成的方式，全面满足行业应用软件的数据分析需求。它提供多源数据整合、报表统计、数据可视化、自助式BI 分析、以及数据填

报等功能，帮助用户挖掘数据的潜在价值，为管理者制定决策提供数据支撑。



(1) 数据应用效率问题

自助BI 与可视化分析：以产品化的方式降低数据获取、数据分析、数据应用的成本，解决数据响应周期长、开发成本高、运营效率低问题

能力要求：集成数据建模、自助分析、数据可视化、数据治理、智能分析的一站式数智化决策分析平台，数据开发专注数仓模型建设，提供健全的模型、完善的资产元数据信息后，业务拖拽式、可视化的数据查询和分析，不需要数据开发介入。针对需要周期性使用的数据，可以保存成可视化Dashboard，自助进行可视化报表减少，释放接口和前端开发人力。比如： QuickBI、观远、帆软BI、tableau 等

智能营销平台 (DMP)：基于大数据计算和数据挖掘技术，构

建用户画像标签体系，用户圈选、精细化分层，进行差异化运营和营销触达，提升运营ROI。业务同学可基于平台实现从人群圈选、场景构建、触达投放、效果回收的闭环，同时，基于算法挖掘标签及模型推荐的人群组合，从基于人的经验运营，到基于大数据算法推荐的智能运营。

(2) 数据资产建设与治理问题

21年云栖大会，阿里云数据中台负责人强调，要在场景的驱动下，把数据中台的资产模块做的更厚实。



整体应用流程：

3.1 数据连接

通过数据连接引擎建立与Oracle、Mysql、SqlServer 等各种关系型数据库、API、ODBC 或者 Excel、JSON、XML 等数据文件

等不同来源的数据连接，支持跨源，用于数据建模。

3.2 数据建模

将填报采集的数据和已有业务系统如OA、ERP、CRM 数据库的数据，通过可视化数据模型设计器或者自定义SQL数据集，建立多表关联模型，并对模型进行字段重命名、二次计算、分组计算等操作，添加参数及筛选过滤，形成最终用于设计报表和仪表板的数据集。

3.3 统计报表

为业务层的日常运营提供必备的明细报表，以满足数据查询、报表打印等需要。包括典型的Excel 类中国式复杂报表、合同文档类报表、分组报表、明细报表、图表、分栏报表等各种报表类型。

3.4 决策仪表盘

以高度交互与自由数据探索的管理看板为基础，辅助决策的制定。通过灵活自由的布局页面，拖拽可视化组件，绑定数据，生成仪表盘。浏览者在查看时，可以通过条件过滤、条件格式化、排序、图表切换、调整数据绑定等组件，多维自由的从不同视角进行数据分析。仪表盘自带的联动和钻取分析，深入探查数据背

后的原因。指导管理层科学决策。

3.5 数据大屏

通过可视化大屏，对企业综合信息进行展示，满足企业内、外部信息共享与交流的需要。通过各种酷炫的可视化组件，将企业核心经营指标、KPI 指标、综合管控指标等直观的展示到总经理办公室、会议室或者会客大厅等，方便管理层掌握企业经营情况。

4. 数据中台核心领域

为了有效管理信息资源，必须构建集团级数据中台体系。数据中台体系包含数据中台组织、大数据构架管理、主大数据管理、大数据质量管理、大数据服务管理及大数据安全管理内容，这些内容既有机结合，又相互支撑。

(1) 数据应用效率问题

自助BI 与可视化分析：以产品化的方式降低数据获取、数据分析、数据应用的成本，解决数据响应周期长、开发成本高、运营效率低问题

能力要求：集成数据建模、自助分析、数据可视化、数据治理、智能分析的一站式数智化决策分析平台，数据开发专注数仓模型建设，提供健全的模型、完善的资产元数据信息后，业务拖拽式、可

可视化的数据查询和分析，不需要数据开发介入。针对需要周期性使用的数据，可以保存成可视化 Dashboard，自助进行可视化报表减少，释放接口和前端开发人力。比如： QuickBI、观远、帆软 BI、tableau 等

智能营销平台(DMP)：基于大数据计算和数据挖掘技术，构建用户画像标签体系，用户圈选、精细化分层，进行差异化运营和营销触达，提升运营ROI。业务同学可基于平台实现从人群圈选、场景构建、触达投放、效果回收的闭环，同时，基于算法挖掘标签及模型推荐的人群组合，从基于人的经验运营，到基于大数据算法推荐的智能运营。

(2) 数据资产建设与治理问题

21年云栖大会，阿里云数据中台负责人强调，要在场景的驱动下，把数据中台的资产模块做的更厚实。

数据血缘：数据入湖到输出应用经过多个环节，上游数据问题如何快速通知下游，下游数据逻辑排查如何向上追溯，以及数据治理表或路径下线，如何评估下游的影响并通知，都依赖于全链路数据血缘的建设。可以说，完善的血缘功能，可以极大提高数据开发的工作效率

成本优化：数据有自己的生命周期，比如活动期间的数据监控报表，活动下线后，报表可以下线释放资源。成本优化提供高耗任务、小文件、冷数据等不同治理维度的指标，及治理目标，从资产健康度评估维度，指导数据开发人员主动进行成本优化、数据治理，系统层面具备治理目标检测、一键治理、数据回收、彻底删除等治理功能，并且可以基于固化的治理规则，进行系统自动化治理。

(3) 数据开发流程的效率问题

目标：提供异构数据源数据同步可视化工具，通过源和目标参数配置实现数据入湖、入仓，以及存储介质的转换，降低人肉脚本处理带来的风险和维护成本。建设统一的数据开发平台，数据开发只需要关注数据处理逻辑，无需关注集群资源、任务调度，通过配置化的方式进行依赖关系配置，及任务运行周期，快速进行数据回溯、任务重启、停止

数据集成：业务数据库、操作日志、状态变更消息等数据源接入数据中心，如Biglog 同步、MySQL库表订阅、Kakfa 数据落HDFS

等。数据经过实时或离线ETL后，数据集成再将数据输入CK、Hbase、ES等供业务端应用

离线开发平台：批数据处理，一般为T+1或小时级的准实时数据，包括任务逻辑处理、依赖配置、调度配置、任务运维等功能。

实时开发平台：流数据处理，以FlinkSQL、StreamSQL为主要计算处理框架，实时处理消息队列等各种流式数据，输出实时报表、实时接口推荐等服务

随着批流技术组件的发展，批流一体化开发平台的建设也陆续在实践中。

(4) 数据服务快速输出

有人也把数据中台称之为DAAS，即数据即服务，数据如何快速输出业务端，赋能产品创新。API服务统一管理，建立完善的应用血缘关系，提供通用接口的配置化生成能力，降低对Java开发的依赖。

数据服务管理平台：数据中台思想下，数据服务输出是应用输出的最主要形式，数据服务管理平台一方面要具备将数据资产自助配置化输出的能力，即数仓清洗好的数据模型，数据开发或业务人员可以通过入参、出参的可视化配置生成API接口，不需要接口开发介入。同时也要把API资产化管理，API接口文档、应用调用情况做到可追踪、可监控。

4.1 大数据模型

大数据模型是大数据构架中重要一部分，包括概念大数据模型和逻辑大数据模型，是数据中台的关键、重点。理想的大数据模型应该具有非冗余、稳定、一致、易用等特征。逻辑大数据模型能涵盖整个集团的业务范围，以一种清晰的表达方式记录跟踪集团单位的重要大数据元素及其变动，并利用它们之间各种可能的限制条件和关系来表达重要的业务规则。大数据模型必须在设计过程中保持统一的业务定义。为了满足将来不同的应用分析需要，逻辑大数据模型的设计应该能够支持最小粒度的详细大数据的存储，以支持各种可能的分析查询。同时保障逻辑大数据模型能够最大程度上减少冗余，并保障结构具有足够的灵活性和扩展性

4.2 大数据生命周期

一般包括大数据生成及传输、大数据存储、大数据处理及应用、大数据销毁四个方面。

(1) 大数据生成及传输

大数据应该能够按照大数据质量标准和发展需要产生，应采取 措施保证大数据的准确性和完整性，业务系统上线前应该进行必要的安全测试，以保证上述措施的有效性。对于手工流程中产生的大数据在相关制度中明确要求，并通过事中复核、事后检查等手段保

证其准确性和完整性。大数据传输过程中需要考虑保密性和完整性的问题，对不同种类的大数据分别采取不同的措施防止大数据泄漏或大数据被篡改。

(2) 大数据存储

这个阶段除了关注保密性、完整性之外，更要关心大数据的可用性，对于大部分大数据应采取分级存储的方式，不仅存储在本地磁盘上，还应该在磁带上，甚至远程复制到磁盘阵列中，或者采用光盘库进行存储。对于存储备份的大数据要定期进行测试，确保其可访问其大数据完整。大数据的备份恢复策略应该由大数据的责任部门或责任人负责制定，信息化管理部门可以给予相应的支持。同时还需要注意因为部门需要或故障处理的需要，可能对大数据进行修改，必须在大数据管理办法中明确大数据修改的申请审批流程，审慎对待后台大数据修改。

(3) 大数据处理和应用

信息化相关部门需要对大数据进行分析处理，以挖掘出对于管理及业务开展有价值的信息，为保证过程中大数据的安全性，一般应采用联机处理，系统只输出分析处理的结果。但是实际中，因为相关大数据分析系统建设不到位，需要从大数据库中提取大数据后再对大数据进行必要的分析处理，在这个过程中就需要关注大数据提取操作是否可能对大数据库造成破坏、提取出的大数据在交付给

分析处理人员的过程中其安全性是否会降低、大数据分析处理的环境安全性等等。

(4) 大数据销毁

这个阶段主要涉及大数据的保密性。应明确大数据销毁的流程，采用必要的工具，大数据的销毁应该有完整的记录。尤其是对于需要送出外部修理的存储设备，送修之前应该对大数据进行可靠的销毁。

4.3 大数据标准

大数据标准是集团单位建立的一套符合自身实际，涵盖定义、操作、应用多层次大数据的标准化体系。

大数据标准的建立是集团单位信息化、数字化建设的一项重要工作，行业的各类大数据必须遵循一个统一的标准进行组织，才能构成一个可流通、可共享的信息平台。

数据中台对标准的需求可以划分为两类，即基础性标准和应用性标准。前者主要用于在不同系统间，形成信息的一致理解和统一的坐标参照系统，是信息汇集、交换以及应用的基础，包括大数据分类与编码、大数据字典、数字地图标准；后者是为平台功能发挥所涉及的各个环节，提供一定的标准规范，以保证信息的高效汇集和交换，包括元大数据标准、大数据交换技术规范、大数据传输协议、大数据质量标准等。

(1) 大数据分类与编码

大数据分类与编码标准是信息化建设中标准化的一项基础工作，该类标准规定平台汇集、交换相关信息统一的分类系统和排列顺序以及编码规则，目的是在不同系统和用户之间建立交通大数据的一致参照，对提高大数据采集、处理和大数据交换效率具有重要作用。大数据分类与编码标准的制定将有力推进平台标准化及交通信息化建设标准化的进程。

(2) 大数据字典

针对实际需求，定义大数据集，建立各个领域的大数据字典，规范大数据概念和大数据定义。在此基础上，形成完备的集团单位大数据集和大数据字典。

(3) 元大数据标准

元大数据标准是描述大数据资源的具体对象时所有规则的集合，它包括了完整描述一个具体大数据对象所需要的大数据项集合。针对各种信息资源分别制定适当的元大数据标准，可为信息的管理、发现和获取提供一种实际而简便的方法，从而提高大数据交换效率。

(4) 大数据交换标准

为了保证大数据共享和交换的顺利实现，必须明确定义和规范大数据交换的相关标准。大数据交换的标准规范是集团单位综合信息平台的核心标准。其中应当包括大数据交换内容、大数据交换格

式、大数据传输方式、各类中心间大数据接口的标准化等方面。

(5) 大数据质量标准

由于大数据采集任务通常由其他二级平台完成，数据中台平台的标准方法主要集中在大数据的加工和管理上。应该重点开发的一个领域是大数据质量控制方法。应当从三个方面对大数据质量方法进行研究：“坏大数据”或“不可靠大数据”的识别，错误大数据的编辑方法，以及缺少值的处理。

4.4 主大数据

主大数据管理要做的就是从各部门的多个业务系统中整合最核心的、最需要共享的大数据(主大数据),集中进行大数据的清洗和丰富,并且以服务的方式把统一的、完整的、准确的、具有权威性的主大数据传送给集团单位范围内需要使用这些大数据的操作型应用系统和分析型应用系统。

主大数据管理的信息流应为:

- 1) 某个业务系统触发对主大数据的改动;
- 2) 主大数据管理系统将整合之后完整、准确的主大数据传送给所有有关的应用系统
- 3) 主大数据管理系统为决策支持和大数据仓库系统提供准确的大数据源。

因此对于主大数据管理要考虑运用主大数据管理系统实现,主

大数据管理系统的建设，要从建设初期就考虑整体的平台框架和技术实现。

4.5 大数据质量

大数据质量不高将影响大数据仓库应用程度不高。低下的大数据质量往往造成开发出来的系统与用户的预期大相径庭，大数据质量关系建设有关分析型信息系统成败，同时大数据资源是集团单位的战略资源，合理有效的使用正确的大数据能指导集团单位做出正确的决策，提高省综合竞争力。不合理的使用不正确的大数据(即差的大数据质量)可导致决策的失败，正可谓差之毫厘、谬以千里。

大数据质量管理包含对大数据的绝对质量管理、过程质量管理。绝对质量即大数据的真实性、完备性、自治性是大数据本身应具有的属性。过程质量即使用质量、存储质量和传输质量，大数据的使用质量是指大数据被正确的使用。再正确的大数据，如果被错误的使用，就不可能得出正确的结论。大数据的存贮质量指大数据被安全的存贮在适当的介质上。所谓存贮在适当的介质上是指当需要大数据的时候能及时方便的取出。大数据的传输质量是指大数据在传输过程中的效率和正确性。

高质量的交通运输行业大数据至少有如下几项要求：

一是正确性，在转换、分析、存储、传输、应用流程中不存在错误；

二是完整性，大数据应用或要求的所有记录、字段都存在；

三是一致性，体现在整个大数据的定义和维护方面，确保大数据在使用的整个过程中是一致的；

四是时效性，衡量指标是在指定的大数据与真实的业务情况同步的时间容忍度内，即指定的更新频度内，及时被刷新的大数据的百分比；

五是可靠性，提供大数据的大数据源必须能够可靠稳定地提供大数据。

大数据质量管理的规划和实施包括以下内容：

一是大数据质量管控体系的建立，包括大数据质量的评估体系，定期评估大数据质量状况；

二是在部门各个应用系统中的落实，包括每个应用系统中的大数据质量检查等；

三是在最开始建立大数据质量管理系统的時候，借助数据中台平台上，通过建立大数据质量管理的规则来集中化地建立大数据质量管理体系，发现问题并持续改进；

四是大数据质量管理与业务稽核的结合，通过业务规则的稽核来发现大数据质量深层次的问题，将大数据质量与业务一线结合起来，使业务人员对大数据质量问题有更加清晰和明确的认识。

完善的大数据质量管理是保障各项数据中台工作能够得到有效

落实，达到大数据准确、完整的目标，并能够提供有效的增值服务的重要基础。

4.6 大数据服务

大数据整理最终目的就是要服务于各部门单位、人员等，能更准确更快更方便的服务是大数据服务管理的目标。

大数据服务管理是指针对内部积累多年的大数据，研究如何能够充分利用这些大数据，分析行业业务流程优化业务流程。大数据使用的方式通常包括对大数据的深度加工和分析，包括通过各种报表、工具来分析运营层面的问题，还包括通过大数据挖掘等工具对大数据进行深度加工，从而更好的管理者服务。通过建立统一的大数据服务平台来满足针对跨部门、跨系统的大数据应用。通过统一的大数据服务平台来统一大数据源，变多源为单源，加快大数据流转速度，提升大数据服务的效率。

4.7 大数据安全

由于集团单位的重要且敏感信息大部分集中在应用系统中，大数据安全更是至关重要。如何保障大数据不被泄露和非法访问，是非常关键的问题。大数据安全管理主要解决的就是大数据在保存、使用和交换过程中的安全问题。

大数据安全管理主要体现在以下六个方面：

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/625131224323011241>