

运营商数据治理实践

大数据时代下的数据治理

DTCC

2015中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2015

大数据技术探索和价值发现



自我介绍

- 郭岳 （三少）
 - Oracle 10G OCM (2009年)
 - 中国移动通信集团浙江有限公司
技术保障部
 - 中国移动集团专利评审专家
 - 新浪微博：正牌三少
 - 微信：qq379622
 - 邮箱：guoyue@zj.chinamobile.com



ORACLE®
10g Certified Master



第一部分

数据治理的概念

第二部分

浙江移动的数据治理计划

第三部分

实施阶段成果

第四部分

后续展望



大数据时代诞生推动数据资产管理发展

“大数据时代”的诞生：

半个世纪以来，随着计算机技术全面融入社会生活，信息爆炸已经积累到了一个开始引发变革的程度。它不仅使世界充斥着比以往更多的信息，而且其增长速度也在加快。信息爆炸的学科如天文学和基因学，创造出了“大数据”这个概念。同时，信息的价值得到充分的发掘，企业运用数据创造了更大，更多的价值。

因此，数据被业界公认为是企业的最宝贵资产之一。



大数据时代的来临使得数据的价值逐步显性化，也被各种组织和企业更加重视。如何应用手中的数据资产，指导其日常运营，战术和战略制定，提供更好的产品和服务、降低成本，控制风险成为组织和企业的首要问题。数据治理当前已经成为IT业界的一门新兴学科被广泛研究。



当数据成为数据资产

纷繁多样的数据不断在产生，如何使数据变成企业的数据资产，数据资产通常被认为必须具备**可控制**、**可量化**、**可变现**的特性。

可控制

企业自身生产或者取得的数据，必须是高度清晰的，可以充分被企业所掌控，完全被企业所拥有和进一步应用的数据。



可量化

数据资产货币化，以货币形式对数据资产进行评估和测量，有利于企业内以同一口径衡量数据价值

可变现

企业通过数据挖掘和分析、数据租售、信息销售等方式，持续将自身掌握的数据资产变现，使数据资产保值和增值。

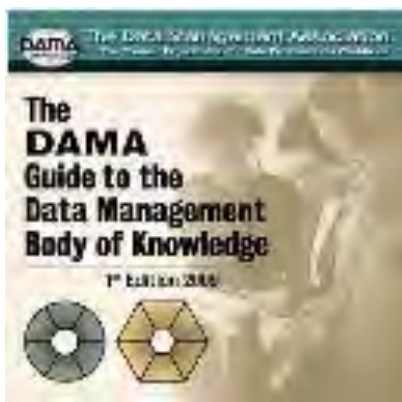


数据资产管理与DAMA

数据资产管理（Data Asset Management，简称DAM）正是规划、控制、和提供数据这种企业资产的一组业务职能，包括开发、执行和监督有关数据的计划、政策、方案、项目、流程、方案和程序，从而控制、保护、交付和提高数据资产的价值。

-- 摘自DMBOK

在国外随着数据管理行业的成熟和发展，数据资产管理作为一门专业管理领域被人们广泛研究和总结，国外一些数据资产领域的专家和学者成立了数据资产管理专业论坛和组织**DAMA**，总结出数据资产管理相关理论指导体系。



同时仿照项目管理PMBOK编写了关于数据资产管理的专业著作**DMBOK**，详细阐述了数据资产管理的相关理论、体系和知识。

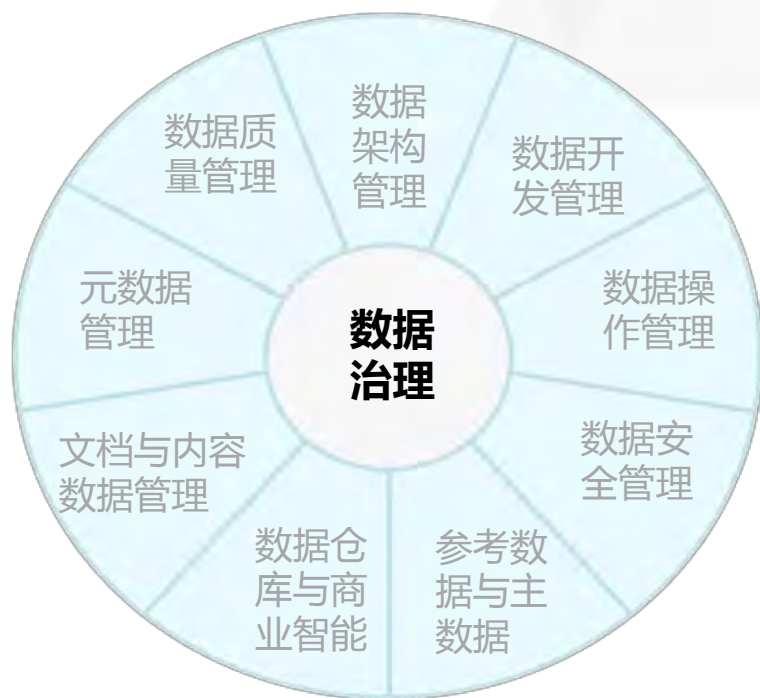


DAMA数据资产管理体系



DAMA-数据治理是数据资产管理的核心

数据治理 = 数据资产管理的核心 = 数据资产（管理）²



数据治理是数据资产管理框架的核心职能

- ❑ 数据治理作为数据资产管理框架的核心职能，它与其他数据资产管理职能交互并互相影响。
- ❑ 数据治理是对数据资产管理行使权力和控制的活动集合（规划、监控和执行），它指导其他数据管理职能如何执行。
- ❑ 数据治理是在高层次上执行数据管理制度。
- ❑ 数据治理，定义、审批、沟通企业高层次的数据战略、政策、标准、架构、流程和度量体系，同时，负责追踪并保证数据管理体系的合规性和一致性。
- ❑ 它还负责发起、追踪和监控数据资产管理的可交付成果，以及数据资产管理中的问题管理



DAMA数据治理关注的主要方面

专业数据管理组织成员负责管理数据政策、标准和规程。管理和实施数据架构，保护数据资产和利益相关者的利益，并提供数据服务。

数据政策是的数据治理管理初衷的说明和基本规划；数据规范和标准是企业由数据管理组织制定，并由数据治理委员会评审通过的关于企业数据的基础准则。

数据管理制度确保对数据资产进行有效控制和使用的业务职责问责制度。

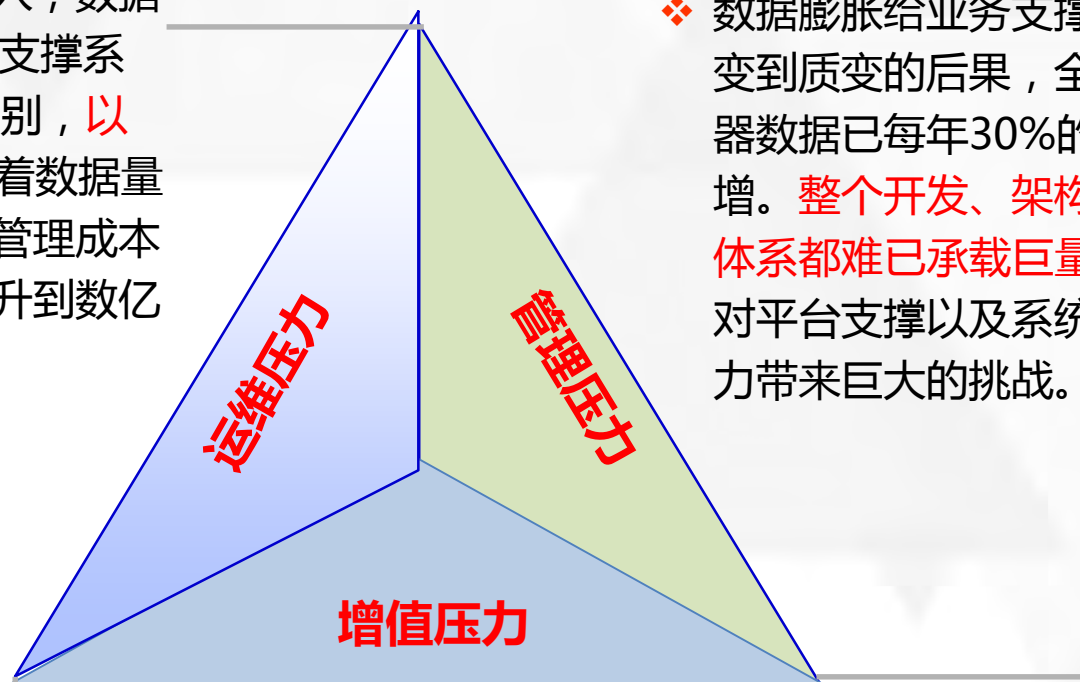
识别、管理、解决企业在日常中遇到的不同类型的数据问题，包括质量问题，标准问题、安全和保密问题等等



大数据时代下数据治理压力

大数据时代的业务支撑面临“运维、管理、增值”三大压力，通过数据治理，实现数据管理的降本增效，提升能力与价值

- ❖ 随着信息化技术的不断深入，数据存储将迈入EB级别。目前支撑系统的总数据量已达到PB级别，**以每年50%的速度递增**，随着数据量的不断增长，资源投入及管理成本也不断上升，从千万级上升到数亿级，提升几十倍。



- ❖ 数据膨胀给业务支撑带来量变到质变的后果，全球服务器数据已每年30%的速度递增。**整个开发、架构、运维体系都难以承载巨量的数据**，对平台支撑以及系统管理能力带来巨大的挑战。

- ❖ 数据是企业最核心的价值，面对越来越激烈竞争，越来越多的企业开始重视对数据管理及运营，用以提升业务发展能力。面对巨量数据，**如何有效进行数据的保值和增值。**



第一部分

数据治理的概念

第二部分

浙江移动的数据治理计划

第三部分

实施阶段成果

第四部分

后续展望



数据治理的战略目标

理解企业和利益相关者的数据需求，促进数据架构被更广泛和深入理解

有效控制数据资产管理成本，降低数据管理和运维压力

保证数据资产的有效利用和价值最大化，使数据资产保值和增值。



获取、存储、保护和确保数据资产的整合以及其完整性

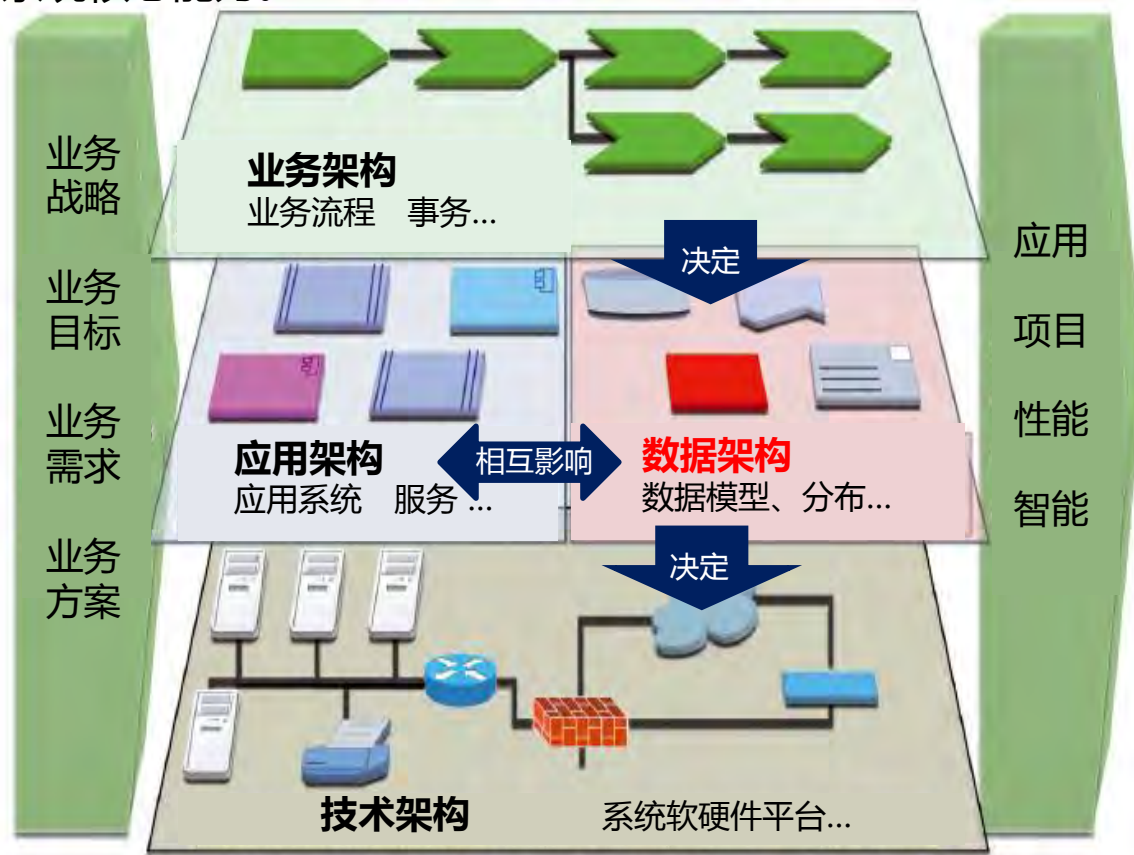
不断提高数据和信息的质量，包括准确性，一致性、及时性

确保数据隐私和保密，防止数据和信息通过不同途径未经授权或不恰当地被使用



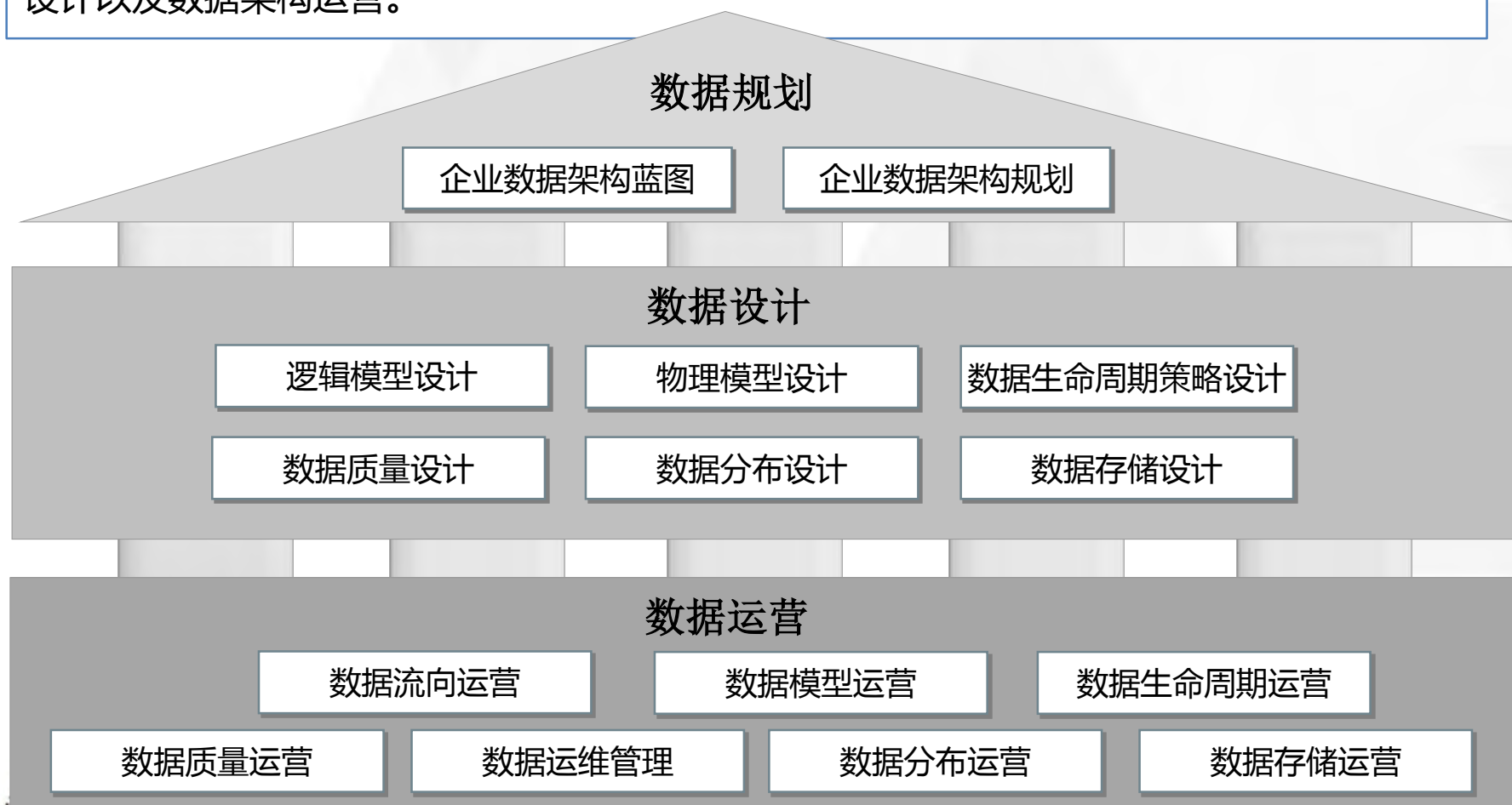
浙江移动IT域数据治理以数据架构为突破点

- ❑ 在TOGAF企业架构框架体系中，数据架构是企业架构中承上启下的核心组成部分，数据架构为中心辐射业务和应用，推动核心掌握能力提升。
- ❑ 浙江移动数据治理以数据架构为主要突破口，以数据架构辐射技术架构和应用架构，重新掌握IT系统核心能力。



浙江移动数据治理架构

根据企业架构框架理论以及国际上对数据治理经验为指引，结合浙江移动业务支撑自身系统业务特点，对数据治理体系相关职能进行细化补充和裁剪合并，形成浙江移动数据资产管理体系。浙江移动数据资产管理体系大致分为三个层面：数据架构规划、数据架构设计以及数据架构运营。



总体目标

数据治理的总体目标是理解企业的信息需求，规范数据的生成以及使用，持续改进数据质量，保证数据安全，最大化数据价值。通过数据资产管理可以提升核心能力掌控，加强对开发商管控力度，提升系统开发和维护质量，减少系统建设、实施及运维等各方面对于开发商的依赖。

IT域数据治理

IT域作为浙江移动的存量系统，已经上线运行多年，在浙江移动数据管理体系下，准备采用从下往上的方式，逐渐掌握IT系统的数据架构

- 建立业务系统物理模型基线和管理规则
- 制定数据生命周期管理规则
- 厘清数据分类建立业务支撑域宏观数据视图
- 建立部门内数据标准，参与数据开发管理
- 掌控系统数据逻辑模型，掌握数据质量管理
- 全面打通业务-应用-数据三者关系，降低系统开发难度和风险
- 进行系统架构解耦，降低系统建设的硬件和软件投资成本

DT域数据治理

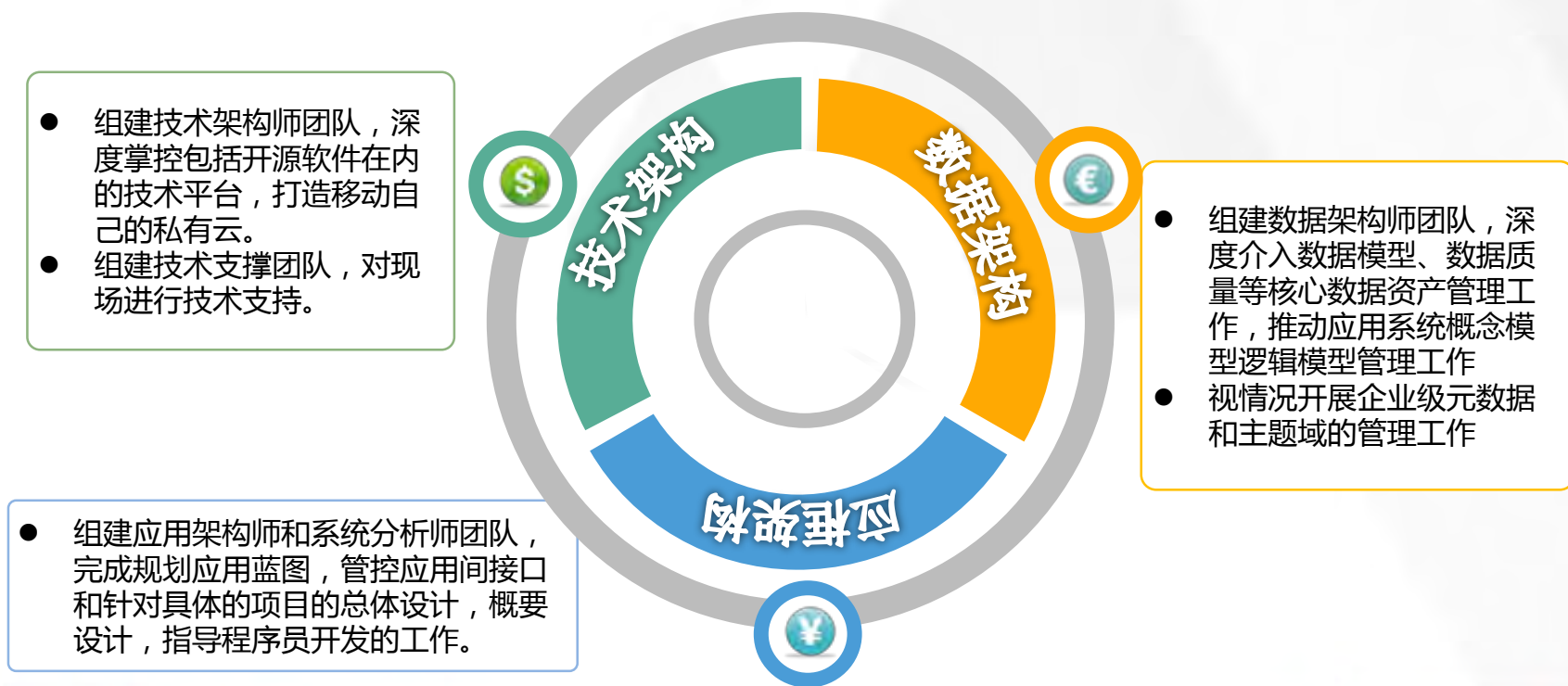
DT域作为浙江移动准备新建的增量系统，在浙江移动的数据管理体系下，准备采用从上往下的方式，全面掌握DT系统的数据架构

- 规划企业级的大数据架构蓝图和数据规划
- 在浙江移动范围内优化数据标准，建立数据与应用的关联关系，提升对数据价值的认识，统一数据分布。健全元数据、主数据管理，增强数据的一致性
- 加强数据的治理和管控工作，做好数据质量管理、安全管理工作
- 建立统一的企业数据资产管理体系，构建共享全景数据视图
- 完成源系统数据、数据引入直至数据建模的三本数据字典

浙江移动IT域数据资产管理的重点是重掌核心能力

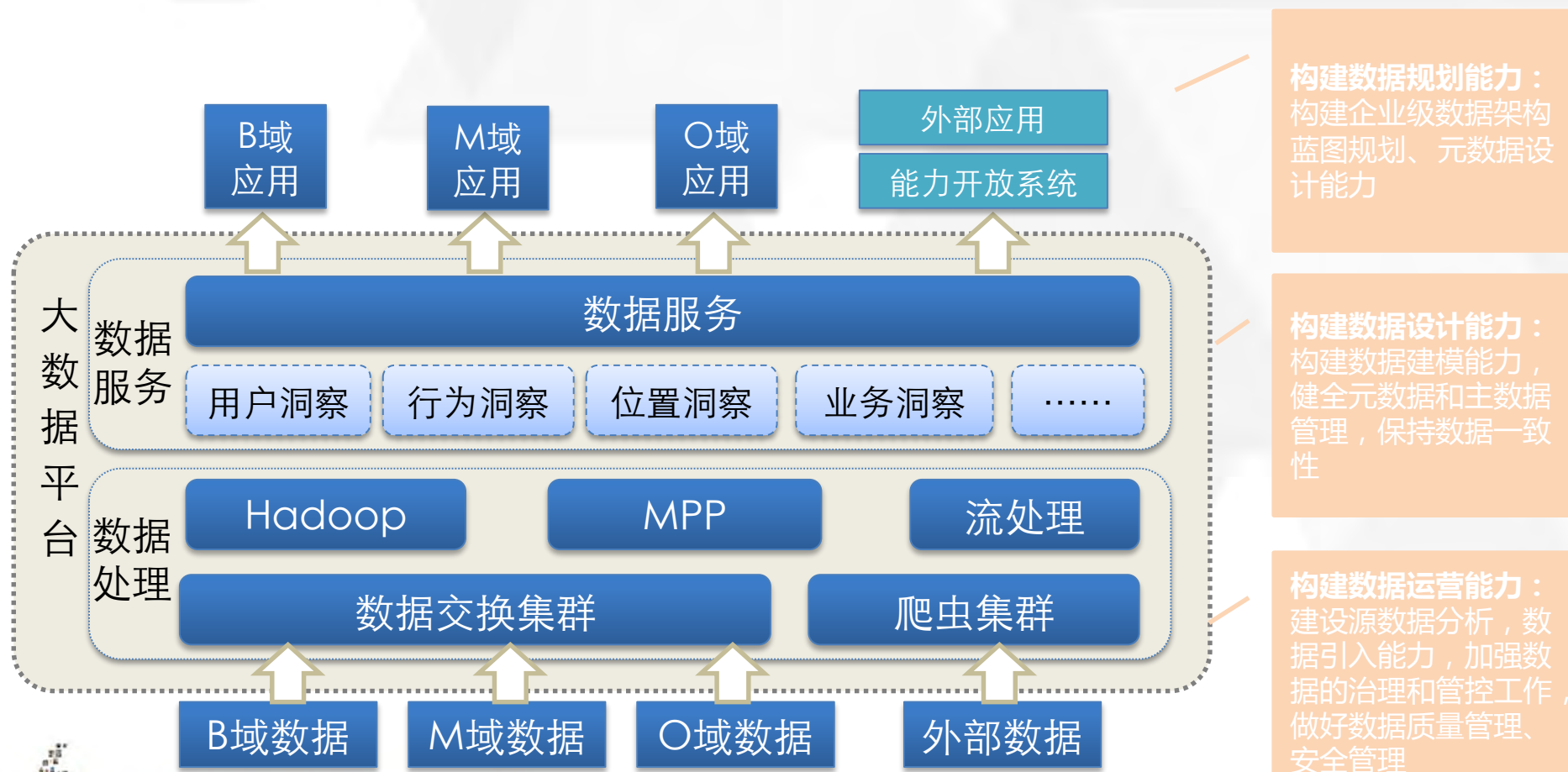
“重掌核心能力”是浙江移动IT域数据治理的重点。

- 以数据治理以及TOGAF企业架构体系为指导，浙江移动通过以数据架构管控为中心，打开核心系统黑盒，解耦核心数据架构，籍此幅射应用架构；进而进一步主导数据架构、应用架构和技术架构的演进，摆脱现有的束缚，让局方重新取得核心能力掌控。



浙江移动DT域数据治理的重点是构建核心能力

DT域作为浙江移动需要新建的增量系统，浙江移动要至上至下建设数据规划、数据设计和数据运营能力，从系统建设的开始就完全掌握数据架构的核心能力。



第一部分

数据治理的概念

第二部分

浙江移动的数据治理计划

第三部分

实施阶段成果

第四部分

后续展望



浙江移动IT域数据治理工作开展

结合浙江移动IT核心系统现状，浙江移动2014年度数据治理工作将重点围绕**数据生命周期管理**、**数据模型管理**两个方面展开。



数据流向梳理

除此以外还针对特定核心业务流程进行数据流向梳理，着重于揭示业务流程、应用以及数据模型三者之间的联系，展现数据资产所形成的数据流如何实现业务流程，提升开发和维护质量，提升系统核心掌控能力。



浙江移动IT域数据治理工作取得主要成绩

按数据治理总体规划，浙江移动2014年数据治理项目的工作重点集中于**数据生命周期管理**以及**物理模型梳理**两部分。第一阶段以理现状、建体系、优架构入手，阶段性成果如下：

1.核心系统现状梳理

根据运营管理室对业务支撑系统的分类结果，项目组在业务支撑域的27套核心系统的范围内，开展现状梳理工作。

取得
成果

2.建立数据资产运营管理体系

项目组编写了数据资产运营的审查标准和管理流程文档,数据生命周期和数据资产变更管理流程已经在实施。

3. 优化架构设计

从提高系统处理效率，降低系统消耗。优化客户感知角度出发，协助开发优化完善系统的应用架构。



IT域核心系统数据架构现状梳理

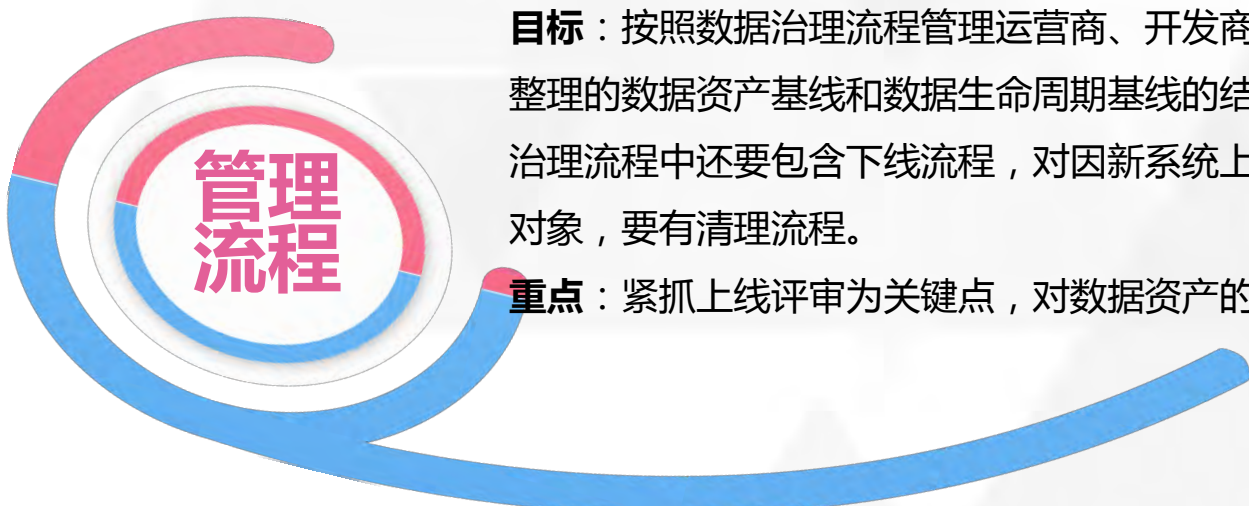
现状梳理进展:共分析27套库1G以上的表近3000余张(占数据库使用空间的97%以上)的数据生命周期保存策略,有近千张表存在生命周期保存策略问题,占总数的50%左右。

物理模型梳理进展:完成营业、账务、计费系统的物理模型梳理,其中有问题的疑似垃圾对象占总数50%左右。



建立IT域数据治理流程、标准和体系

数据治理流程



The diagram for the Data Governance Process features a central circle with the text '管理流程' (Management Process) in pink. This circle is surrounded by two concentric rings, the outer one being pink and the inner one blue. A large, thick blue arrow originates from the bottom of the rings and curves upwards and to the right, pointing towards the 'Data Governance Review Standards' section.

管理流程

目标：按照数据治理流程管理运营商、开发商、第三方厂家，以保证项目组整理的数据库资产基线和数据生命周期基线的结果和生产系统完全一致。数据治理流程中还要包含下线流程，对因新系统上线或业务变化导致无用的数据对象，要有清理流程。

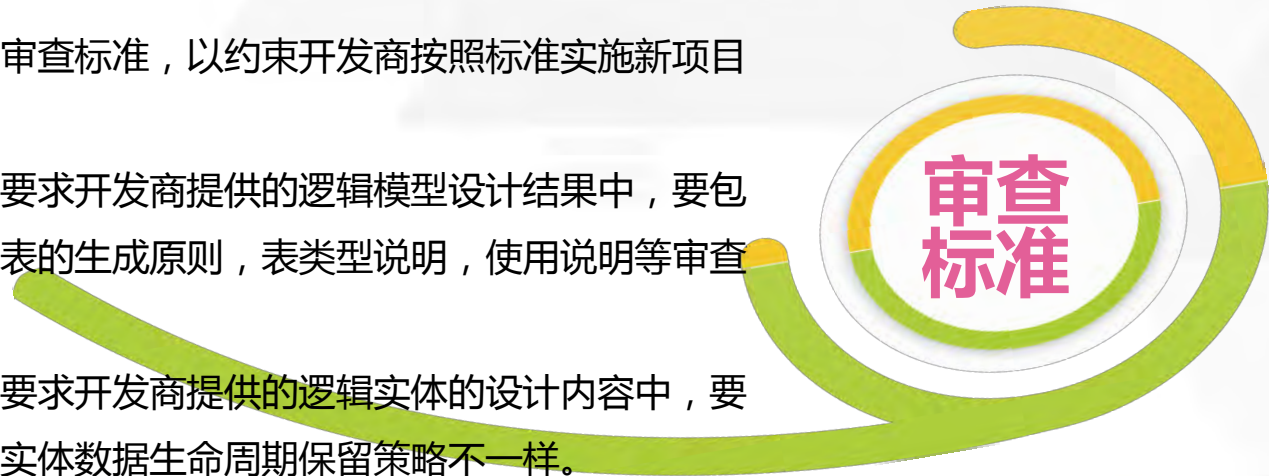
重点：紧抓上线评审为关键点，对数据库资产的变更发起、实施进行检查。

数据治理审查标准

目标：制订并发布数据治理审查标准，以约束开发商按照标准实施新项目和完成新业务上线。

重点：数据治理审查标准，要求开发商提供的逻辑模型设计结果中，要包含数据表的分区分表规则，表的生成原则，表类型说明，使用说明等审查者需要关心的元素。

数据生命周期的审查标准，要求开发商提供的逻辑实体的设计内容中，要考虑实体分类，不同类别的实体数据生命周期保留策略不一样。



The diagram for the Data Governance Review Standards features a central circle with the text '审查标准' (Review Standards) in pink. This circle is surrounded by two concentric rings, the outer one being yellow and the inner one green. A large, thick green arrow originates from the bottom of the rings and curves upwards and to the left, pointing towards the 'Data Governance Process' section.

审查标准



IT域数据生命周期管理降低扩容成本

通过建立多层次的数据生命周期管理体系（包括需求管理、策略管理和技术管理），实现系统历史数据按既定数据保存策略的自动化或半自动化迁移，降低单位数据的存储成本，提升系统效率。

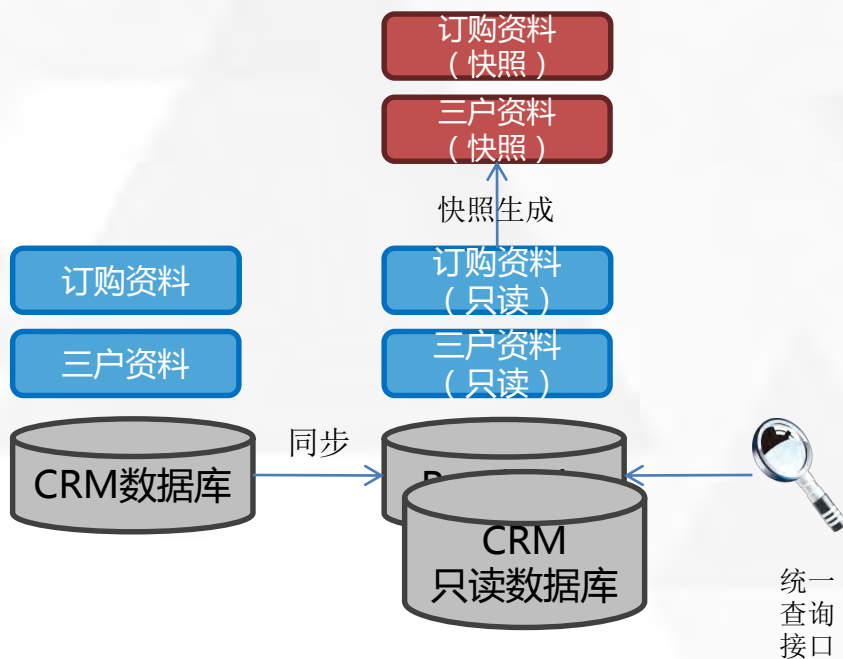
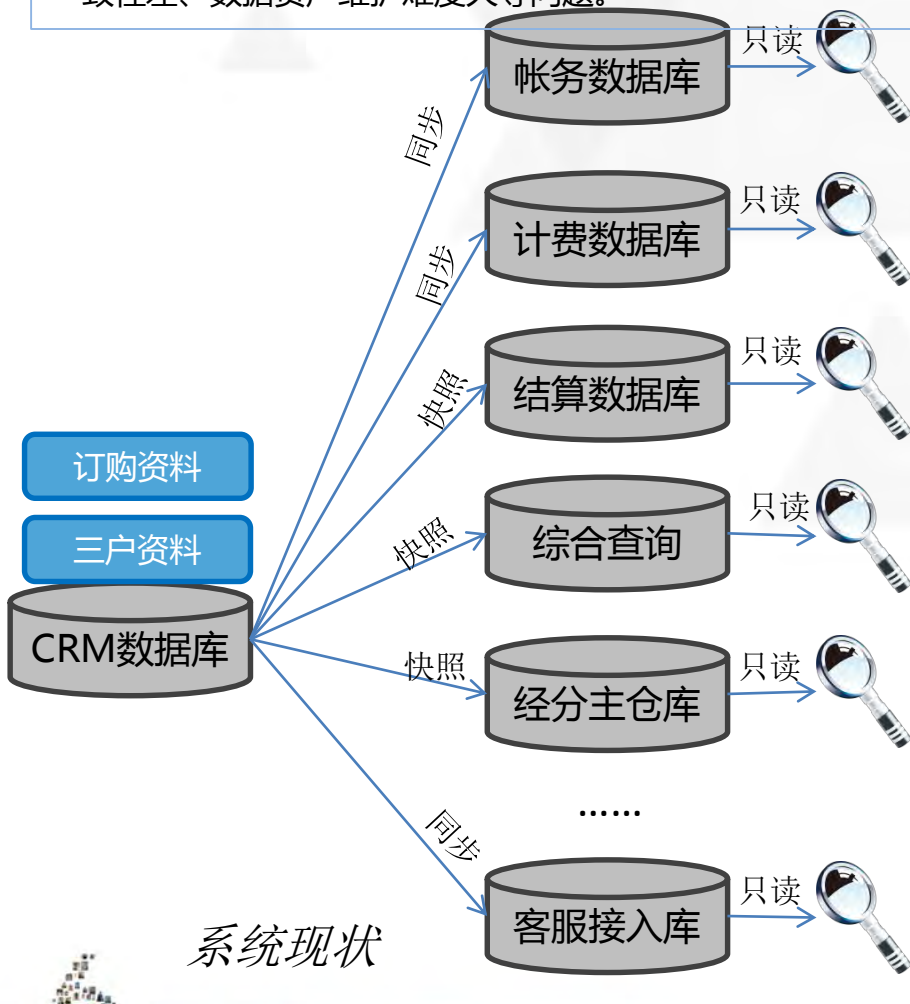
策略实施过程中通过数据迁移清理实施，一次性节省存储几十TB，存储费用估算约近百万元；此后，每月定期清理数据量存储量上升到近10TB，按年计算节省存储扩容费用估算以百万元计。



IT域数据分布优化—订购和三户数据同步

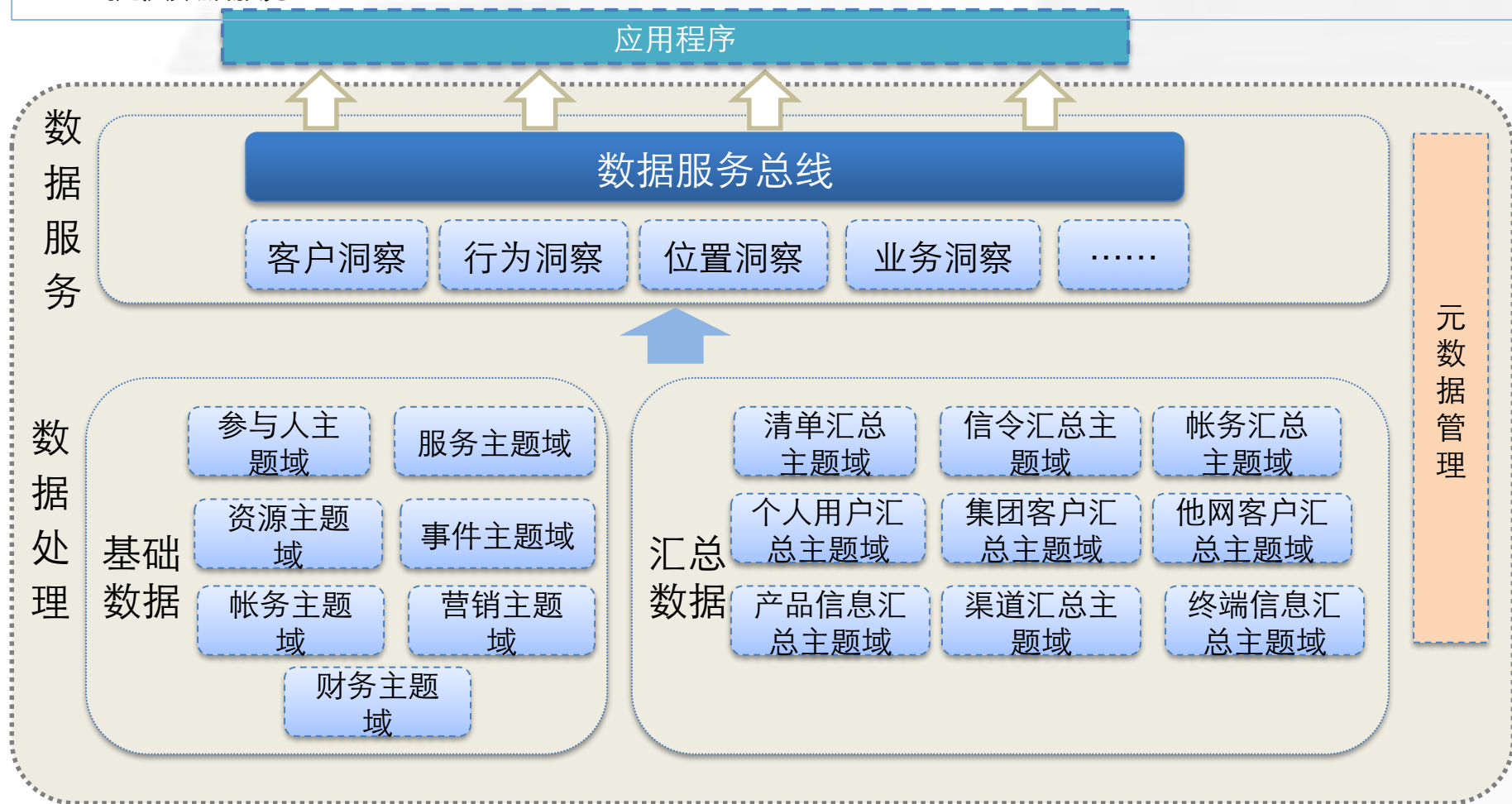
存在问题

现状描述：**三户资料**和**订购信息**是CRM系统的核心数据，它的生成和维护都在CRM系统。由于这些资料的重要，整个业务支撑系统中又由这些数据通过同步或者快照等方式生成的多份COPY或变种数据。带来如浪费资源（主机、存储、网络）、数据一致性差、数据资产维护难度大等问题。



DT域数据蓝图规划

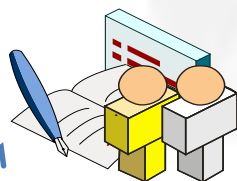
DT域的数据蓝图规划是在元数据管理的辅助下，以浙江移动企业的角度来统一数据描述语言，分不同主体域设计基础数据和汇总数据的逻辑模型，在基础数据和汇总数据的基础上，为应用程序在数据服务总线上提供数据服务。



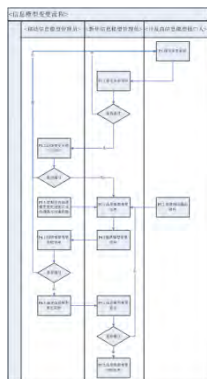
DT域数据基线整理

- 以IT系统整理的核心系统物理模型基线为基础，以大数据营销及变现为驱动力，通过理解数据蕴含的意义，判断出这些数据的价值，明确大数据的分布，建立大数据基线版本。
- 以建设数据基线为契机去建立和完善源系统数据、数据引入直至数据建模的三本数据字典，影响企业大数据蓝图规划和数据建模，夯实大数据的工作的基础

物理模型基线



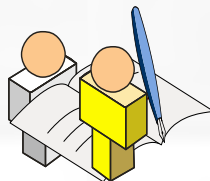
数据治理流程



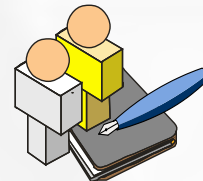
数据生命周期管理

系统名称	系统描述	系统用途	系统重要性	系统维护建议
1. 电话系统	用于客户电话接入	客户电话接入	高	定期维护，确保系统稳定运行
2. 飞信	用于客户短信接收	客户短信接收	高	定期维护，确保系统稳定运行
3. 无线音乐	用于客户音乐下载	客户音乐下载	高	定期维护，确保系统稳定运行
4. 手机视频	用于客户视频播放	客户视频播放	高	定期维护，确保系统稳定运行
5. 手机游戏	用于客户游戏娱乐	客户游戏娱乐	高	定期维护，确保系统稳定运行
6. 手机阅读	用于客户文档阅读	客户文档阅读	高	定期维护，确保系统稳定运行
7. 彩铃	用于客户彩铃播放	客户彩铃播放	高	定期维护，确保系统稳定运行
8. 手机MM	用于客户移动多媒体	客户移动多媒体	高	定期维护，确保系统稳定运行
9. 手机动漫	用于客户动漫播放	客户动漫播放	高	定期维护，确保系统稳定运行
10. MM商城	用于客户移动商城	客户移动商城	高	定期维护，确保系统稳定运行
11. 手机证券	用于客户证券交易	客户证券交易	高	定期维护，确保系统稳定运行
12. 139邮箱	用于客户邮箱服务	客户邮箱服务	高	定期维护，确保系统稳定运行
合计			232	

判断数据价值



大数据基线版本



序号	系统	优先级	实体数
1	电话系统	高	9
2	飞信	高	26
3	无线音乐	高	41
4	手机视频	高	13
5	手机游戏	高	13
6	手机阅读	高	27
7	彩铃	高	22
8	手机MM	高	14
9	手机动漫	高	20
10	MM商城	高	12
11	手机证券	高	29
12	139邮箱	高	6
合计			232

序号	系统	优先级	实体数
1	电话系统	中	16
2	外呼系统	中	14
3	客服接入	中	5
4	存管系统	中	12
5	营销管理	中	7
6	统一短信	中	4
7	PBOSS	中	17
8	渠道管理	中	30
9	渠道数据管理	中	13
10	公共配置	中	12
11	统一支付	中	15
12	计费管理	中	5
13	手机支付	中	5
合计			155

序号	系统	优先级	实体数
1	资源管理	低	7
2	账务管理	低	9
3	订单管理	低	2
4	终端管理	低	13
5	产品管理	低	31
6	12580	低	18
7	其他增值平台	低	0
8	账务处理	低	3
9	客户管理	低	7
10	统一权限	低	3
11	CBOSS	低	0
12	酬金管理	低	7
13	综合结算	低	0
合计			100



源系统数据



数据引入



数据建模



第一部分

数据治理的概念

第二部分

浙江移动的数据治理计划

第三部分

实施阶段成果

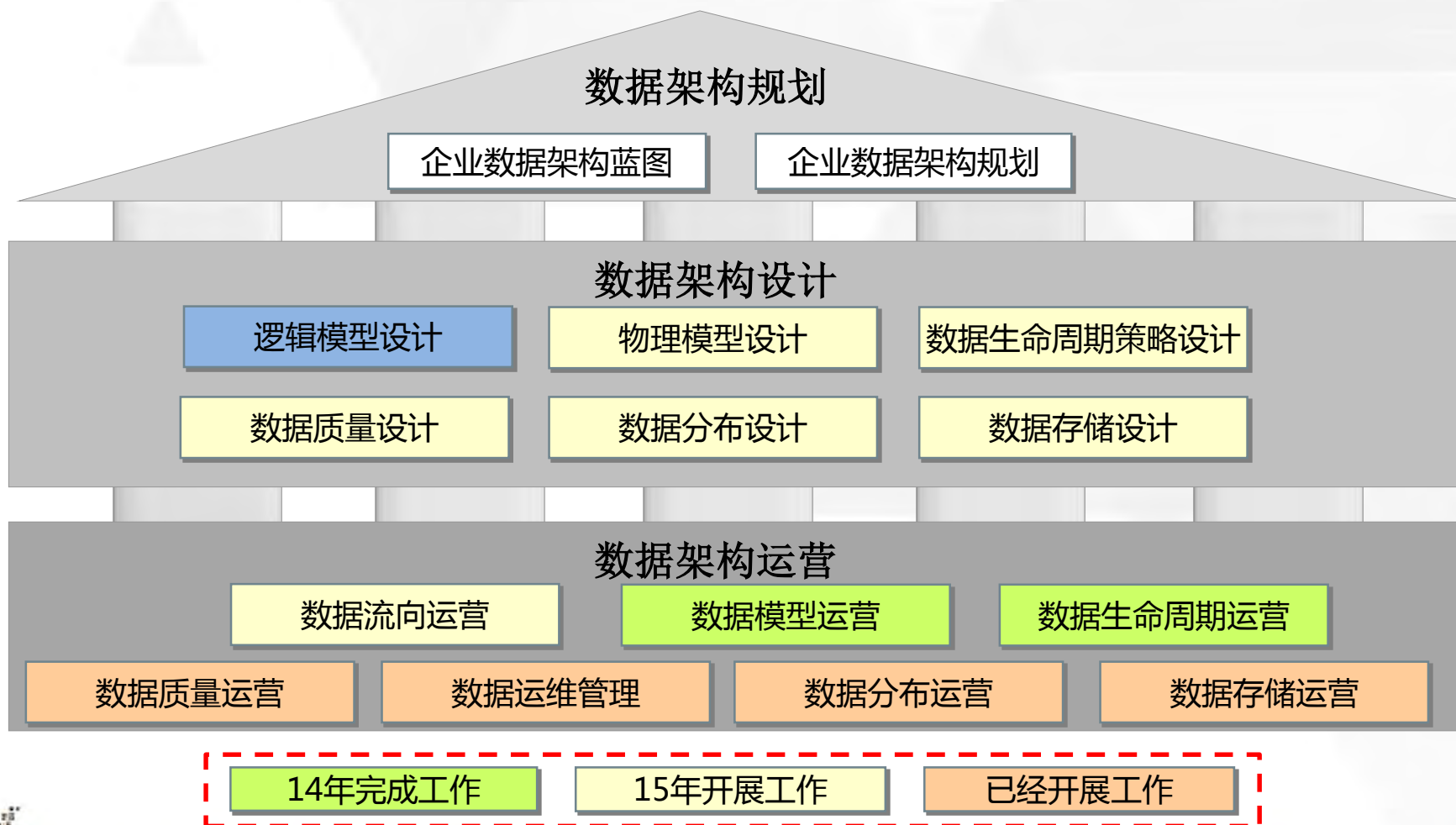
第四部分

后续展望



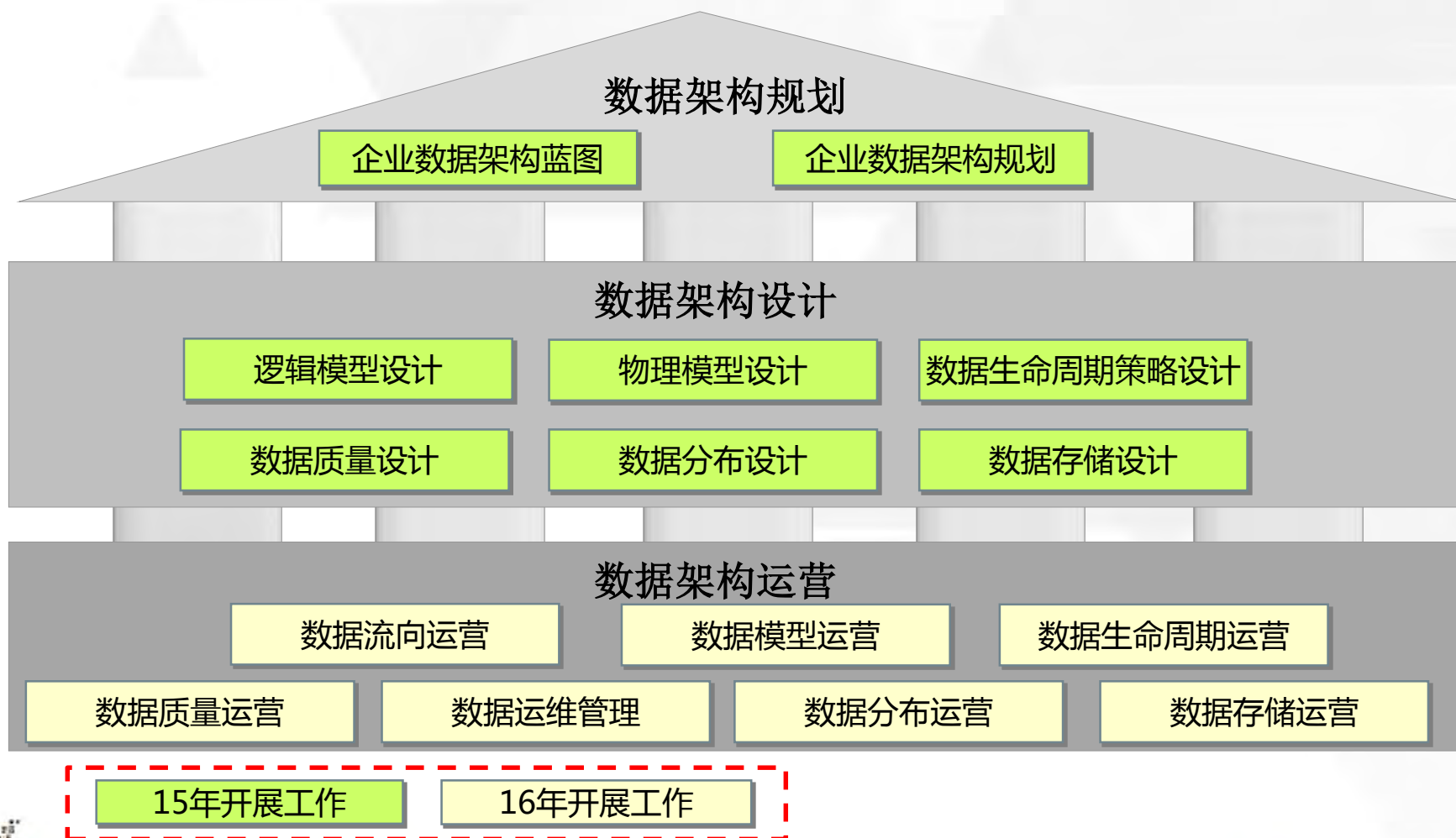
后续展望-IT域工作计划

浙江移动计划通过两年的努力，结合DAMA理论指导，在IT域建设完成浙江移动自己的数据治理体系。



后续展望-DT域工作计划

浙江移动计划通过两年的努力，在DAMA理论指导下，在DT域建设完成浙江移动自己的数据治理体系。



请关注我们的团队-微信公众号：三墩IT人

招兵买马：

- 工资不高，但稳定！
- 工作充实，但不会过劳死！
- 没有BAT的实力，但团队技术氛围好！
- 争做传统行业领先地位！



扫码关注：

- 交流技术
 - 了解运营商的技术内容
 - 获得运营商内部推荐机会
- 拿出微信扫一扫





THANKS