

旅游价值与文化特色数据处理规范

数字文化旅游平台规范

数字文化旅游平台规范建设课题组

二〇一六年五月

目 录

1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.1 旅游价值	1
3.2 文化特色	2
4. 数据采集	2
4.1 数据采集管理	2
4.2 数据采集方式	2
4.3 数据交换管理	3
4.4 互联网数据抓取	4
5. 数据组织	5
5.1 数据目录管理	5
5.2 元数据管理设计思路	6
5.3 元数据管理	6
5.4 数据元管理	6
5.5 数据质量管理	6
6. 数据存储	7
6.1 三权分立的安全机制	7
6.2 多种身份验证方式	7
6.3 资源限制	8
6.4 自主访问控制	8
6.5 标记与强制访问控制	8
6.6 存储加密	8
6.7 导出数据加密	9
6.8 独立的加密引擎	9
6.9 密钥自管理功能	9
6.10 客体重用	9
6.11 可靠性	9
6.12 故障恢复	9
6.13 基于时间点还原	10
6.14 备份压缩	10
6.15 数据复制	10
6.16 数据库审计	10
6.17 数据库的备份策略	11
6.18 数据组织方式	12
7 数据计算	12

1 范围

本规范规定了文化旅游资源的旅游价值与文化特色数据处理要求。

本规范适用于“数字文化旅游共性支撑技术研发与区域资源集成应用示范”课题应用，其他相关领域也可参考使用。

在不同的实际旅游目的地进行旅游价值和文化特色的数据处理工作时，处理要求根据具体情况和需求可以少于或多于本规范内容，但应达到相关技术要求。

2 规范性引用文件

下列文件中的条款通过本规范的引用而成为本规范的条款。凡是注明日期的引用文件，其随后所有的修改版（不包括勘误的内容）或修订版均不适用于本规范。凡是不注明日期的引用文件，其最新版本适用于本规范。

GB 2260-1995	中华人民共和国行政区划代码
GB 2659-1994	世界各国和地区名称代码
GB 3304-1991	中国各民族名称代码
GB 2261-1991	人的性别代码
GB 3100-1993	国际单位制及其应用
GBT 3358.1-2009	统计学词汇及符号 第1部分：一般统计术语与用于概率的术语
GB/T 7408-2005	数据元和交换格式 信息交换 日期和时间表示法（idt ISO 8601:2000）
GB/T 7408-1994	数据元和交换格式 信息交换：日期和时间表示法
GB/T 18391.3-2009	数据元的规范与标准化（idt ISO/IEC 11179.3-2003）
GB/T 26162.1-2010	信息与文献 文件管理 第1部分：通则（idt ISO 15489-1:2001）
	文化旅游资源兴趣点及道路采集规范
	文化旅游资源数据描述规范

3 术语和定义

下列术语和定义适用于本规范。

3.1 旅游价值

旅游价值是旅游客体对旅游主体的意义，是对旅游主体带来的效用和满足。旅游价值由旅游主体、旅游客体及旅游价值实现的路径和平台三部分构成。旅游客体主要是指旅游产品，包括依赖于旅游资源而生产的各种旅游产品以及资源脱离性的旅游活动和产品（如主题公园）。旅游主体则主要包括旅游者、旅游企业、政府、当地居民等其他利益相关者。旅游者是其余旅游价值主体获得旅游价值的载体。旅游企业追求的旅游价值主要是通过提供服务和产品而实现的。政府追求的旅游价值是宏观经济的发展和社会进步。相关利益者旅游价值满足程度也会影响到整个旅游价值系统。在旅游客体对旅游主体产生效用的过程中还必须借助

一些媒介和渠道，如优越的政策平台以及良好的市场环境。旅游价值主体、旅游价值客体和旅游价值实现媒介构成了整个旅游价值系统。

3.2 文化特色

文化特色指某社会群体或社会阶级在长期实践生活中形成的独特的语言、文学、艺术及一切意识形态在内的精神产品。这种特色的文化包括了有别于其他地区人群的较为独特的知识、信仰、艺术、伦理道德、法律、风俗和其他习得的能力和习惯。可形成旅游者到访旅游的吸引力构成因素。

4. 数据采集

数据采集包括三种采集方式：数据交换和互联网数据采集、数据接口集成。

4.1 数据采集管理

4.1.1 数据采集定义

数据采集管理配置主要通过可视化图形界面进行配置，主要配置内容有数据集、发送和接收端、自定义服务、运行等内容。系统提供统一配置界面，以目录资源结构的方式配置数据交换与服务共享，包括了数据集管理和实例管理。在数据集中，支持对数据库、文件和自定义服务的配置。也可以根据数据集采集的不同方式来确定采集任务的实施，如果是人工采集，直接导入前期整理的数据即可。如果是利用数据交换平台和数据抓取工具采集数据，利用配置界面配置相应的采集任务。它可以定义多种类型的数据采集方式，如数据库交换、文件交换、消息交换、Web Services 交换等。

4.1.2 采集任务管理

数据采集任务定义后，可以通过采集任务管理对已配置的任务进行编辑、修改、删除操作等。管理员可以查询到配置的其发送和接送的相关信息，可以监控当前采集客户端具有配置其发送和接收的相关信息。

4.1.3 任务监控

采集任务监控对正在执行的数据采集任务进行实时的在线跟踪监测，对任务的网络连接状态、数据发送量、数据接收量、数据采集交换数据量、数据采集状态等任务运行数据信息进行监控，并通过图形化的方式进行展示出来，以方便管理人员进行管理与日常监控。

采集任务监控对于 ETL 数据采集过程的每一步，提供可视化的在线监控系统。通过监控系统，可以查看所有正在运行的转换任务、故障历史等。监控记录中详细描述了数据采集的成功数目、失败数目，转换时长、日志明细等。

4.2 数据采集方式

4.2.1 数据交换

针对数据来源一般是多个业务系统，数据整合难度大，错误数据、不完整数据、重复数据等“脏数据”

较多，数据量大，转换任务并发量大等问题，大数据平台要提供完善的分布式数据转换工具与基于 web 的在线任务执行与监控系统，通过建立抽取模型将现有各个系统的数据进行整合、汇总、清洗、转换，并可以实时在线监控交换任务的数据输入量、输出量、成功个数、失败个数、失败原因等。

数据交换支持弹性部署。针对高并发、大数据量的数据转换任务，可以将单个转换任务分配给多个子服务器并行执行，对于转换过程中的每一步操作都有相关的日志记录、转换明细等，并支持事务回滚、数据更新等操作。

4.2.2 互联网数据采集

针对互联网数据量大、中文分词和向量空间模型导致效率低下等问题，要支持自动化集群管理、任务分配、负载均衡体系与基于机器学习算法建立的决策树模型的分布式爬虫技术，配合丰富实用的图形化界面，自动生成数据抽取和格式化规则，强力锁定目标数据的内容结构，应对数据结构的变化。通过定向、分类的数据抽取，将相关的数据统一收集到数据管理中心，进行分布式存储。

4.2.3 数据接口集成

针对采购资源及联盟资源，由于商务条件限制，部分数据只能进行数据接口集成。通过对方提供的数据接口，采用授权方式，实时读取数据，并且对数据进行整理后，通过数据服务及可视化接口展示列表，但需要确定条件的数据时，再进行下载。

4.3 数据交换管理

(1) 交换中心管理

(2) 交换传输管理

交换传输是数据交换的枢纽，通过交换传输管理，在前置交换功能之间构成信息交换通道，要求根据部署的交换流程，实现交换信息的打包、转换、传递、路由、解包等功能。交换传输管理通过消息总线模式，实现分中心前置交换信息库之间的信息处理和稳定可靠、不间断的信息传递，并提供目录注册与传输功能，要求包含数据适配、数据传输、数据转换等环节。

(3) 交换前置机管理

(4) 交换数据服务

要求提供基础数据服务，如数据传输，数据转换，数据清洗，数据比对，数据加载，数据共享等功能。

(5) 交换配置管理

要求实现基础信息的配置，包括源、目标数据源的配置、标准数据库表结构配置及其编码表的配置等。

(6) 交换策略管理

要求实现数据清洗管理、数据传输管理、增量分离管理、转换映射管理、编码对比管理功能。

(7) 数据规则管理

许多数据操作，如数据清洗，数据比对，数据转换等，都需要通过一定的规则才能完成。规则管理模块

要求提供定义规则的用户界面，使用者可以很方便的以图形方式定义自己的规则。

(8) 运行调度管理

运行调度管理的功能要求提供直观的界面，使用户通过简单方便的操作，对于不同的数据源设置不同的从数据交换前置服务器复制数据到数据中心数据交换库、并进行增量分离的调度策略，以保证数据的及时更新。

4.4 互联网数据抓取

处理平台应提供网络爬虫任务配置管理功能，对于互联网数据，数据采集员可根据用户需求，灵活定义网络爬虫任务，配置数据采集规则，并根据数据采集策略对数据采集到的数据进行过滤处理。

4.4.1 分布式任务调度系统

分布式任务调度系统是对任务进行并行处理，利用多个功能部件或多个处理机同时工作来提高系统性能或可靠性的任务调度管理系统，提高任务处理的处理速度和效率。

与传统的任务调度方式，主要有以下优点

多线程任务同时运行，充分利用系统资源实现并行计算

提高任务执行的效果和执行速度，满足海量数据采集和处理的需要。

分布式的任务调度框架可实现横向扩展的需要。

可以避免出现单个节点失效而使整个系统崩溃的危险

4.4.2 采集任务监控

平台提供采集任务监控系统，对数据采集任务进行实时监控，保证任务正常运行，任务监控主要分为监控模块和反监控模块两个部分。

反监控模块

部分访问网站（特别是社会化媒体）会禁止爬虫访问，怎么让他们不能监控到我们的访问时爬虫软件，保证在对方网站不关闭外部访问的情况下，能够正常的获取数据，这就是需要引入反监控机制。

在数据抓取的过程中，对方网站也会因运维会调整策略，说到底这是一个战争，躲在电脑屏幕后的敌我双方，爬虫必须要能感知到对方的反监控策略进行了调整，通知管理员及时调整反监控策略。

监控模块

系统监控指不管什么系统都可能出问题，如果对方服务器宕机、网页改版、更换地址等我们需要第一时间知道，这时监控系统就起到作用，出现了问题及时发现并通知管理员调整页面爬取规则。

系统监控还要保证任务调度服务、爬虫服务能够正常运行，监控系统服务运行状态。

4.4.3 数据文件解析

根据指定资源格式的定义，进行数据库字段级别的解析，对信息源里包含的数据进行结构化处理，变成

规范的组织形式。将 Web 文档作为信息源，从半结构化的 Web 文档中抽取数据，转化为结构化数据

数据加工处理

对互联网采集到的数据，根据业务需求进行加工处理，对数据一致性、无效性、缺失值进行检测处理，主要包括以下残缺数据、错误数据、重复数据几种类型。

数据接口集成

联盟资源、采购资源在系统实现技术以及实现时间上的差异，造成了在不同的信息系统中存有大最异构数据源。从数据的存储方式上看，这些数据源中有结构化数据源，如 DBMS，有半结构化数据源如 HTML、XML 文件，也有非结构化数据源，如文本文件等：从数据的存储格式上看，在本质上表示相同信息的数据，在不同的数据源中被定义为不同的存储格式，如字段名称定义、字段类型定义的差异等。异构数据源的存在给实现不同信息系统之间数据的互访带来了很大的不便。为此，解决异构数据源方便、灵活的相互访问成为必须解决的问题。

就领域应用而言，各行业各部门应用系统的开发和建设大多是各自完成的，每个部门都有自己的数据库服务器或者数据库文件以及采用了不同时期体系结构的应用程序。在用传统的基于 CORBA、DCOM 或 JAVA、API 等技术的企业应用集成技术来处理多个厂商跨不同平台的产品的时候，存在很大的难度。而 Web Services 作为一种新生的技术、基于 Web Services 的集成技术作为一种新的面向函数和方法的集成应用技术，在很大程度上解决了这些问题，因此，基于 Web Services 的企业集成应用成为目前研究的热点。

异构数据源

异构数据是一个含义丰富的概念，不仅指不同的数据库系统之间的数据是异构的，如 Oracle 和 IBM DB2 数据库：而且还包括不同结构的数据之间的异构，如结构化的 IBM DB2 数据库数据和半结构化的 XML 数据。

信息集成、模式集成、数据集成这三个概念相互联系，不同的概念所研究的侧重点不同。

信息集成是一个广义的概念，它包含很多分支，例如信息抽取，数据清洗，模式集成和数据集成等。模式集成，就是把两个或多个模式集成为一个模式，主要用于数据模式设计。有待集成的模式可能是由不同的数据模型表示的，集成过程中，一般先把它们用公共模型表示，然后识别它们之间存在的各种冲突并加以处理，最后合并成新的模式 这一过程既可手动完成，也可机器辅助完成。数据集成，屏蔽各种异构数据间的差异，对各种异构数据提供统一的操作，使集成后的异构数据对用户来说是统一的和无差异的。狭义上来说，在讨论数据集成中，并不考虑模式集成等其它问题并假设全局模式已经存在(可能由模式集成而来，可能根据特定的应用需求设计，可能根据数据模式推导)。

5. 数据组织

数据管理包括数据目录管理、元数据管理、数据元管理、数据质量管理等模块组成。

5.1 数据目录管理

数据目录定义允许各种类型的用户进行在线定义数据，包括普通用户、部门管理员等，统一由平台管理

员审核。数据目录定义可以定义数据的名称、来源、更新周期、责任人、主题、语言、描述等，定义的数据目录可以选择是否可被下载，可以下载的数据会被打包成 Excel 和 XML 两种格式供用户下载使用。另外，数据目录定义时可以选择“是否允许在线浏览”，允许在线浏览的数据可以被用户在线分页查看数据明细，并可以进行在线数据统计、分析、挖掘和数据可视化。

5.2 元数据管理设计思路

元数据管理平台需实现元数据的类别管理、属性管理和内容管理，需实现元数据的采集、加工处理、审批、入库、数据发布的全生命周期管理，通过接口提供标准的元数据服务，支撑各类应用系统建设和改造。对分类标准、制度方法文件（包括制度布置文件、问题解答文件和法规性文件）等内容进行编辑和装载。

5.3 元数据管理

平台涉及多种信息资源、业务数据库、非结构化数据、业务模型、业务规则，以及信息资源的各种属性特征。因此，必须对这些特征元素进行统一管理，用户可以从技术和业务角度掌握全部信息资源的生产、存储、转换和同步等所有相关活动，提升信息共享，帮助技术人员和业务人员理解每一个数据的来龙去脉。

元数据定义是对数据资源的规范化描述，是按照相关行业的信息标准，从信息资源中抽取出共性特征，组成的一个特征元素集合，是用一组属性描述定义、标识、表达、转换规则和允许值的数据单元，这种规范化描述可以准确和完备的说明信息资源的各项特征，帮助信息资源使用者准确定位资源、准确规范的使用资源。

元数据定义是对数据目录对应的表结构的定义，包括元模型(表)基本信息和元数据(列)信息。元模型信息包括表的名称、中文描述、数据存储方式，其中数据存储方式有两种：大数据中心和远程数据库。元数据信息包括列的名称、数据类型、长度、标签、隐藏设置等。隐藏设置是对数据隐私加密的一种方式，用户可以通过设置元数据的隐藏规则从而实现对数据的加密，比如将元数据为姓名的这一列定义隐藏规则为第二到第三个字符加密，则实际数据中这一列所对应的所有数据的第二到第三个字符将都会被隐藏加密。

5.4 数据元管理

数据源定义可以定义数据源的名称、地址、用户名、密码、连接池信息等，并支持在线测试。数据源按照用途分为私用和公开两种，按照类型分关系型数据库数据源(MySQL、Oracle、SQL Server)和非关系型数据库数据源(HBase)，定义完成的数据源会在大数据平台中自动实例化成最优的数据库连接池，为数据的获取做好准备。

5.5 数据质量管理

数据质量是从数据整合、数据预处理、资源入库、资源监控、资源利用等数据处理流程环节入手，建立完善的数据生命周期管理与数据质量管控机制，是对数据从获取、清洗、转换、关联、存储、使用等生命周期的每个阶段里可能引发的各类数据质量问题，进行识别、度量、监控、预警等一系列质量管理的活动。数据质量管理是循环管理过程，其终极目标是通过可靠的数据，提升数据在政府、企业决策等业务中的使用价

值。

数据质量作为大数据平台的核心之一，具有多重属性，其基本质量特性主要包括：完整性、一致性、准确性和及时性等四个方面，要对数据质量进行较好地控制，就必须对数据的四个基本质量特性进行很好了解，从而在各个方面采取措施，杜绝数据质量问题的出现，使数据监控工作能够真正达到控制数据质量的目的。

1. 数据完整性检测

数据完整性检测主要包括数据目录对应的数据集是否存在；实际数据集是否存在；数据集中的主键、约束、数据类型相对于元数据是否有缺失等。

2. 数据一致性检测

数据一致性检测指实际数据的值是否和元数据定义的标签的验证规则一致。

3. 数据准确性检测

数据准确性检测主要指数据集中的数据是否有乱码；实际数据的数据类型是否和元数据中定义的相同。

4. 数据及时性检测

数据及时性检测指数据的更新周期是否符合数据目录中定义的更新周期。

5. 质量评分

针对数据集的缺失、元数据的不匹配、数据的不准确等数据质量问题，数据管理平台制定了客观的评分规则，对每个数据目录进行统一打分，并提供了完善的管理系统在线预览数据质量评分排名、有问题的数据、各项数据质量问题的明细等。

6. 数据存储

对全系统所使用的数据库进行统一管理和维护，主要完成数据库结构管理、目录文件结构的管理、数据备份与恢复、数据迁移回迁、数据归档、数据清除、数据存储策略定义、数据访问权限配置定义等功能。

6.1 三权分立的安全机制

数据存储管理在安全管理方面采用了三权分立的安全管理体制，把系统管理员分为数据库管理员 DBA、数据库安全管理员 SSO、数据库审计员 AUDITOR 三类。DBA 负责自主访问控制及系统维护与管理方面的工作，SSO 管理标记和强制访问控制，由 AUDITOR 管理系统审计，三者互相监督，互相配合来共同实施系统管理。这种方式可有效避免传统数据库授权体系中，系统管理员权限过于集中所带来的各种隐患。

6.2 多种身份验证方式

数据存储管理能够根据用户在系统中的身份确定该用户是否具有登录的权限和其在系统中的各种操作级别的角色，确定该用户能够做什么和不能够做什么。数据存储管理提供两种身份验证模式来保护对服务器访问的安全，即数据库身份验证模式和外部身份验证模式。数据库身份验证模式需要利用数据库口令，外

部身份验证模式支持基于操作系统(OS)的身份验证,并提供口令管理策略。

6.3 资源限制

资源限制是控制用户对数据存储管理服务器系统资源的使用情况,以尽可能减少人为的安全隐患。数据存储管理能对登录的安全属性与可访问资源进行限制,同时还可配置表的存储空间配额。系统管理员可借此功能对每个数据库用户单独配置最合适的管理策略,并能有效防止各种恶意抢占资源的攻击。

6.4 自主访问控制

数据存储管理提供了系统权限和对象权限管理功能,并支持基于角色的权限管理,方便数据库管理员对用户访问权限进行灵活配置。

数据存储管理的系统权限和对象权限的最大的区别在于系统权限不属于某个具体的数据库对象。系统权限指的是在数据库内建表权限、建视图的权限等,而对象权限则是定义对某个具体数据库对象的访问权限,如对某个表的插入、删除更新和查询权限等。

当某个用户拥有数据库对象上的某些操作权限及相应的转授权时,该用户可以不用申请和审批,自主地把这些操作权限部分或全部转授给其他用户,从而使得其他用户获得在这些数据库对象上的使用权。数据存储管理也提供级联回收方式来回收这种通过转授所获得的权限。

在数据存储管理中,可以对用户直接授权,也可以通过角色来授权。角色表示一组权限的集合,数据库管理员可以通过创建角色来简化权限管理进程。可以把一些权限授予一个角色,而一个角色又可以被授予多个用户,从而使基于这些角色的用户间接地获得权限。在实际的权限分配方案中,通常先由数据库管理员为数据库定义一系列的角色,然后再由数据库管理员将权限分配给基于这些角色的用户。

6.5 标记与强制访问控制

数据存储管理利用策略和标记来实现数据库的强制访问机制。该功能主要是针对数据库用户、表以及表上的元组,控制粒度达到了记录级。用户操作元组时,不仅要满足自主访问控制的权限要求,还要满足用户和元组之间标记的相容性。这样,就避免了出现管理权限全部由数据库管理员一人负责的局面,可有效防止敏感信息的泄露与篡改,增强系统的安全性。在数据存储管理中,安全管理员可以在每个数据库中定义多种安全策略,每个安全策略包括一组预定义的标记组件,一个标记可以定义多个等级、范围和组,用来表示现实生活中的不同安全特征。安全管理员将这些安全策略应用于表和用户上,就给用户、表和元组都指定了安全标记。在为用户应用策略时,同时可以授予策略特权。策略特权分为访问特权和行标记特权,用来设置读写数据是否受策略影响以及改变行标记的特权配置。用户访问表时,必须保证应用于表上的所有强制访问控制策略均应用于该用户,比较的先后顺序是等级、组和范围,否则访问被拒绝。若表上未应用任何强制访问控制策略,则用户只需满足自主访问控制条件即可。

6.6 存储加密

数据存储管理实现了对存储数据的透明存储加密、半透明存储加密和非透明存储加密。每种模式均可自

由配置加密算法。用户可以根据自己的需要自主选择采用何种加密模式。

(1) 透明存储加密

对于透明存储加密，密钥生成、密钥管理和加解密过程由系统自动完成，用户在数据操作过程中无需人工干预。该功能使对表有访问权限的用户能像对待普通数据一样操作加密数据，在不影响应用逻辑的同时保护数据库中的敏感数据的存储安全。

(2) 半透明存储加密

对于半透明存储加密，系统通过对 DDL 语句进行扩展提供设置加密的接口，另外提供系统函数辅助用户管理密钥，根据用户密钥以及加密设置自动完成加解密动作。通过该功能，用户能够利用自己提供的密钥对隐私数据加密，对其它用户隐藏私有数据，基本不影响应用逻辑，提高了存储加密的易用性。

(3) 非透明存储加密

对于非透明存储加密，系统以存储函数的方式对外提供接口，支持文本、二进制等各种数据类型的加密。

6.7 导出数据加密

数据存储管理提供导出数据加密功能，可以对备份文件以及迁出的文本文件进行加密保护。

6.8 独立的加密引擎

数据存储管理对外提供了独立的加密引擎管理接口，允许用户自由加载第三方的加密模块。用户可选用数据存储管理提供的内置或第三方的加密引擎，保证在利用系统加密功能的同时，加密强度和性能符合要求。

6.9 密钥自管理功能

数据存储管理实现了多级密钥管理机制，统一管理不同的用户的密钥，根据用户身份自动使用其密钥进行加解密，还提供简单方便的密钥更换操作接口。

6.10 客体重用

数据存储管理内置的客体重用机制使数据库管理系统能够清扫被重新分配的系统资源，以保证数据信息不会因为资源的动态分配而泄露给未授权的用户。

6.11 可靠性

任何一个系统都存在发生各种意外故障的可能性。数据存储管理的高可靠性可以避免或降低系统的意外故障对用户带来的损失。

数据存储管理主要通过提供以下功能实现了系统的高可靠性：(1) 故障恢复 (2) 多种备份与还原方式 (3) 基于时间点还原 (4) 备份压缩 (5) 数据复制 (6) 数据库集群

6.12 故障恢复

数据存储管理支持事务故障、系统故障和介质故障恢复，在归档模式下，即使介质损坏，也可利用归档

日志将数据恢复到故障发生前一刻。同时，数据存储管理还支持并行恢复，可有效利用多 CPU 和磁盘阵列等特性，减少恢复时间。

6.13 基于时间点还原

数据存储管理系统提供了更灵活的还原方式，在归档模式下，可根据用户要求恢复到指定的时间点。

6.14 备份压缩

数据存储管理在备份时，支持对备份文件进行压缩，能极大的减小备份文件占据的磁盘空间。

6.15 数据复制

数据库的复制技术，提供了一种机制，使对某一个站点上数据的修改操作能够传递到其他站点，从而保证站点间数据的一致性。

数据存储管理提供基于事务的同步复制和异步复制功能。同步复制即所有复制节点的数据是同步的，如果复制环境中的主表数据发生了变化，这种改变将以事务为单位同步传播和应用到其他所有复制节点。如果任何一个节点的数据更新操作失败，则涉及的相关事务全部回滚。异步复制是指在多个复制节点之间，主节点的数据更新需要经过一定的时间周期之后才反映到从节点，主、从节点之间的数据虽然最后保持一致，但是并非实时保持一致。如果复制环境中主节点要被复制的数据发生了更新操作，这种改变将在不同的事务中被传播和应用到其他所有从节点。这些不同的事务间可以间隔几秒，几分钟，几小时，也可以是几天之后。复制节点之间的数据在一段时间内是不同步的，但传播最终将保证所有复制节点间的数据一致。数据存储管理的数据复制功能支持一到多单向表\站点\库级复制、多到一单向表\站点\库级复制、级联复制、多主多从复制、环形复制、对称复制以及大数据对象复制。

数据存储管理的数据复制功能满足了用户对数据库中的数据进行分析和集中的要求，增强了数据的可用性、可靠性，同时也满足了基于数据复制的数据库集群方案的需求。

6.16 数据库审计

审计机制是数据管理系统安全管理的重要组成部分之一。一个灵活的审计子系统，可以通过它来记录系统级事件、个别用户的行为以及对数据库对象的访问。通过查看审计信息，数据库审计员可以知道用户访问的形式以及试图对该系统进行的操作。一旦出现问题，数据库审计员可分析审计信息，跟踪审计事件，查出原因。

系统允许在三个级别上进行审计设置：系统级、语句级和对象级。

(1) 系统级审计

系统级审计系统的启动与关闭，由系统强制产生，用户无法进行设置。

(2) 语句级审计

语句级审计主要针对影响特定类型数据库对象的特殊 SQL 或语句组的审计，如 `AUDIT TABLE` 将审计 `CREATE TABLE`、`ALTER TABLE` 和 `DROP TABLE` 等语句。

语句级审计分为两大类：全局的审计和局部的审计。全局审计的动作是全局的，不对应于具体的数据库，审计设置登记在 SYSTEM 库中的 SYSAUDIT 字典表内，后者审计的动作发生在具体的数据库内，审计设置登记在各个库的 SYSAUDIT 字典表内。

(3) 对象级审计

对象级审计主要针对特殊对象上的语句的审计，因为对应于特定数据库对象，因此总是局部的，其审计设置登记在数据库对象所在数据库的 SYSAUDIT 字典表中

6.17 数据库的备份策略

根据对数据库应用的理解，特提出了针对数据库的备份策略。

1. 对备份与恢复操作的复杂度必须可管理。

由于生产及管理系统数据量大，数据库数量多，需要针对不同的数据库有一套完整的统一的管理策略，通过精心的备份策略设计，保障数据备份的有效性和高可用性。同时，备份的速度快而且能够优化备份手段，实现短时间段内的数据备份。

2. 自动备份并避免人为失误的可能性。

建立起对计算机系统安全备份的管理制度，将迅速增大的日常数据安全与灾难恢复系统的建立结合起来，在数据管理上结合自动化存储设备，到自动化备份，实现无忧化管理，使得灾难恢复过程中有一套结合数据备份管理的有效手段。

3. 备份必须与事务量成正比，而不应与数据库的物理大小成正比，恢复所花的时间必须与需要恢复的数据成正比。

数据备份应面向系统恢复，保证主系统发生故障时的数据恢复能在指定时间内按事先设置好的步骤完成。保证主机系统内的其它系统设置信息也能顺利恢复。同时充分考虑断点恢复的有效手段以保证系统恢复后及时投入运行。

4. 能够在运行 7x24 业务的同时做备份和局部恢复。

连续备份保证数据库恢复到较近时刻，再结合有效的断点恢复技术或手段，保证数据库恢复的成功。

5. 备份必须是可扩展的、可靠的，并且充分利用可用的存储硬件。

结合大型磁带库，磁盘阵列等存储设备，通过集中管理和策略管理，做到自动备份和恢复。

数据库备份与恢复的方法

数据库提供的备份与恢复方法很多，如脱机备份、联机备份、物理备份和恢复、逻辑备份和恢复、完全恢复、不完全恢复、RMAN 备份和恢复、备用数据库等。不管是哪种备份方式都要和硬件的配置结合起来，因此，首先要做的是根据硬件配置的实际情况选择备份和恢复方式。其中，数据库系统的恢复分为数据文件损坏、控制文件损坏、整个系统文件损坏等各种方式，在以上几种情况下，都可以通过备份的数据实现快速恢复。

数据库的数据保护主要是数据库的备份，当计算机的软硬件发生故障时，利用备份进行数据库恢复，以恢复破坏的数据库文件或控制文件或其他文件；另一种数据保护就是日志，数据库实例都提供日志，用以记录数据库中所进行的各种操作，包括修改、调整参数等，在数据库内部建立一个所有作业的完整记录；再一个就是控制文件的备份，它一般用于存储数据库物理结构的状态，控制文件中的某些状态信息在实例恢复和介质恢复期间用于引导数据库。

6.18 数据组织方式

数据中心数据的组织方式可从以下四个维度进行：数据存储层、数据资源目录层、服务目录层、元数据层等。

6.18.1 数据存储层

总体来说，数据中心管理的诸多数据在数据存在形式上大体可分为空间数据、属性数据、非结构化数据、切片缓存数据等四大类。其中空间数据和属性数据进行图属一体化管理，即矢量数据的管理技术（如下图）；非结构化数据如图件、Word、Excel 等以 FTP 和数据库编目的方式管理；切片缓存数据采用文件目录的方式进行管理。

6.18.2 数据资源目录层

入库后的数据在数据管理平台中将以统一的数据资源目录来进行管理，数据资源目录见附表 5。

6.18.3 数据服务目录层

数据服务目录层对信息资源进行梳理，按照统一的数据服务目录体系注册到数据注册管理系统中，建立信息资源服务目录规划，以便将共享资源进行集中有序组织，方便用户资源查找、数据检索和应用系统访问，数据服务目录见附表 6。

6.18.4 应用服务目录层

应用服务目录层通过开发数据申请、数据分发、数据仓库、数据查询、数据调用、功能仓库、WEB 应用等功能组件，基于 SOA 技术，在综合管理服务平台上注册和统一管理，为“一个平台、两个市场”提供开放式的应用服务支撑，满足“一个平台、两个市场”各业务应用系统快速开发需求，应用服务目录见附表 7。

6.18.5 元数据层

根据《核心元数据》标准，并在该标准基础上根据数据中心数据的实际情况作了一定扩展，设计了适合本数据中心的数据资源核心元数据、数据服务元数据和应用服务组件元数据等。

7 数据计算

数据计算包括离线计算、内存计算、流式计算等，数据存储包括关系数据库、列式数据库及分布式文件系统等。警务数据包括多种数据类型，如警综、大情报、PGIS 等结构化数据，也有交通卡口、图像网、网安网等图片、视频的非结构化数据，针对不同的数据类型使用不同的存储方式，保证数据的存储质量。

离线计算一般是批量处理数据库的过程，比如利用 Hadoop 的 MapReduce；内存计算是将数据是放在在

内存中，效率比较高，流式计算为应用提供高效分布式的触发式任务、定时任务等事件驱动程序处理能力。

数据架构体系平台设计用户容量最终可能达到准千万级，各级各部门的大量数据会在平台上存储和处理，需要采用当今主流的大数据处理技术应对海量数据处理和高并发访问。当今主流的大数据处理技术包括分布式文件系统、分布式数据库、分布式计算、分布式缓存等。

数字文化旅游平台规范