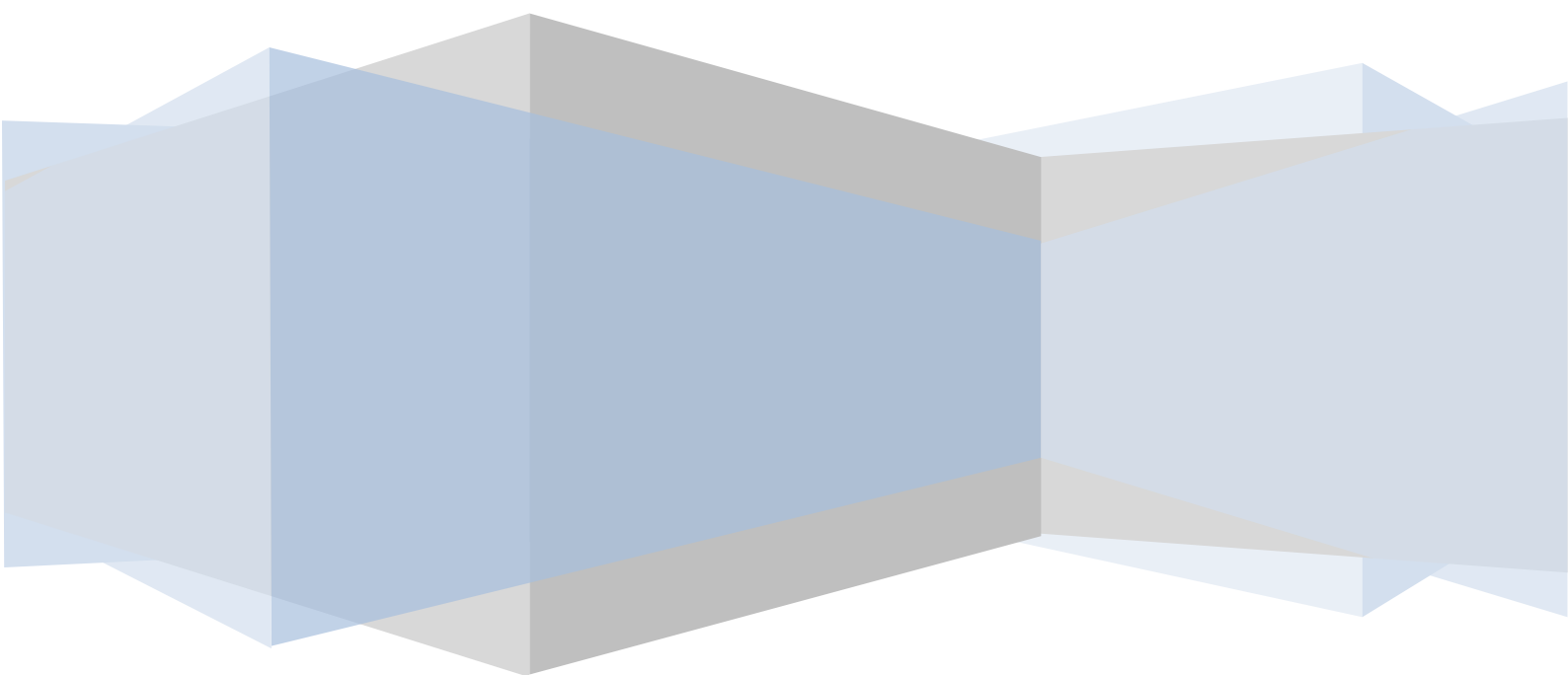


重庆工业职业技术学院

数据存储标准

规范 V1.0



目录

1 范围	4
2 术语和定义	4
2.1 数据	4
2.2 结构化数据	4
2.3 半结构化数据	4
2.4 非结构化数据	4
3 数据分类和类型	4
3.1 数据分类	4
3.2 结构化数据类型	5
3.3 半结构化数据类型	5
3.4 非结构化数据类型	5
4 数据命名规则	5
4.1 基本原则	5
4.2 结构化数据命名规则	5
4.3 非结构化数据命名规则	6
5 数据存储	6
5.1 要求	6
5.2 存储方式	6
5.3 存储格式	6
6 数据入库	7
6.1 入库方式	7

6.2 入库记录管理	7
6.3 数据检查	7
7 存储备份	7
7.1 备份内容	7
7.2 备份方法	8
7.3 备份管理	8
8 存储安全	8
8.1 数据存储安全	8

1 范围

本文件规定了可存储数据的基本类型和命名规范,提出了数据存储、备份、安全的操作规范。

2 术语和定义

下列术语和定义适用于本文件。

2.1 数据 data

对事实、概念或指令的一种形式化表示,适用于以人工或自动方式进行通信、解释或处理。

2.2 结构化数据 structured data

能够用简短的数字、文字或统一的结构加以表示的数据,或者能用平面(二维)行列表结构进行逻辑表达的数据。

2.3 半结构化数据 semi-structured data

具有结构性,但结构变化大,且难以用结构化数据的处理方法将其放进平面行列表格里的数据。

2.4 非结构化数据 unstructured data

数据结构不规则或不完整,没有预定义的数据模型,不能用数据库二维逻辑表来表现的数据。

3 数据分类和类型

3.1 数据分类

根据数据的结构化程度及特征，数据分为结构化数据、半结构化数据、非结构化数据。

3.2 结构化数据类型

结构化数据类型包括：字符串类型数据；整数、浮点数、定点数类型数据；二进制数类型数据；日期与时间类型数据；布尔型数据。

3.3 半结构化数据类型

半结构化数据类型包括：XML 类型数据；JSON 类型数据；HTML 类型数据。

3.4 非结构化数据类型

非结构化数据类型包括但不限于：所有格式的办公文档；各种文本、报表；图片、音频、视频。

4 数据命名规则

4.1 基本原则

4.1.1 具有唯一性，能单独成立。

4.1.2 字段名称的字符采用 26 字母、09 的自然数以及下划线‘_’组成。

4.1.3 不应使用数据库关键字命名，如 select、table、show、time、password 等。

4.1.4 宜使用普遍理解的缩略语，应尽量准确、简练。

4.2 结构化数据命名规则

4.2.1 表命名

表命名结构由部门简称首字母、‘_’、数据命名词组成。命名应全部采用小写字母，不宜使用大写字母。应填写表描述信息。

4.2.2 字段命名

4.2.2.1 命名全部采用小写字母和数字组成，应以字母开头，长度不超过 20 个字符。

4.2.2.2 应填写字段描述信息。

4.2.2.3 命名表的列，不应重复表的名称。

4.3 非结构化数据命名规则

非结构化数据以文件形式存储，命名由“数据部门简称_文件命名词_日期.文件后缀”组成。

5 数据存储

5.1 要求

5.1.1 应为单个机构提供独立逻辑存储空间。

5.1.2 应具备可弹性伸缩、高可用的分布式对象存储的能力。

5.1.3 应能为结构化、半结构化和非结构化类型数据提供不同级别的存放策略。

5.1.5 应创建不同类型的字段，根据用途和需求变化对字段进行增加、删除和修改操作，并提供数据导入/导出和数据迁移功能。

5.2 存储方式

5.2.1 结构化数据统一存入数据库中，如存入 MySQL、PGsql 等。

5.2.2 半结构化数据转换成结构化数据统一存入数据库中。

5.2.3 非结构化数据采用文件形式存储至对象存储中，并将文件基本属性、存储路径等描述信息以结构化数据存入数据库中。

5.3 存储格式

5.3.1 能够支持多种格式的内容数据存储。

5.3.2 同一条记录可容纳一个或多个内容数据。

6 数据入库

6.1 入库方式

6.1.1 文件导入

在线文件导入是通过政务数据平台在线文件上传功能,将数据以结构化数据文件形式录入到数据库中,一次导入可批量入库多条记录。

6.1.2 数据库同步

将关系型数据库中的数据,通过数据库采集功能同步至政务数据平台数据库中,一次可批量同步多条记录。

6.2 入库记录管理

6.2.1 对入库数据的文件格式或数据库类型进行检查,对不支持的文件格式或数据库类型显示提示信息。

6.2.2 记录入库数据的数量、时间、处理人员等处理过程相关信息。

6.2.3 监测和过滤错误数据,并记录数据入库产生的错误日志。

6.3 数据检查

6.3.1 入库前后数据类型、长度等属性应保持一致。

6.3.2 入库前后数据量保持一致。

6.3.3 入库前后数据信息项保持一致。

6.3.4 应符合数据库规定的编码类型。

7 存储备份

7.1 备份内容

7.1.1 应对数据库数据、数据库结构、数据库配置定义文件等进行备

份。

7.1.2 应对数据在采集、入库等过程的配置策略进行备份。

7.2 备份方法

应支持全量备份、增量备份和差分备份等备份方式。

7.3 备份管理

7.3.1 对备份对象、备份介质、备份时间、备份数据保存时间和备份方式等策略进行管理。

7.3.2 支持对备份策略进行添加、删除、修改等操作。

7.3.3 支持对备份作业状态和备份设备状态的监控。

8 存储安全

8.1 数据存储安全

8.1.1 应做好数据存储保密性，保障数据存储不被非授权主体访问。

8.1.2 应做好数据存储完整性，保证数据存储时不被非法篡改或破坏数据。

8.1.3 应做好数据存储可用性，保障访问存储数据时能保证数据是能够正常使用的。

8.1.4 应做好数据安全访问，保证合法的身份及访问授权。