## 数据模型

在Sparrow中，所有业务数据对象（实体）都抽象（或者说包装）为树形目录结构中的目录节点。这样的抽象，旨在解决信息管理类软件系统的业务数据异构与业务操作异构性。

* 解决业务数据异构性

业务数据对象的异构性通常体现在描述该数据对象属性的不同。比如，一个学生数据对象可能具有描述其性别的属性，然而一个课程数据对象则不具备性别属性。相对于MongoDB等NoSQL的Scheme-less解决方案，本项目探讨使用关系型数据库PostgreSQL的解决方案。

PostgreSQL支持JSON与JSONB字段，其实际上是在二维表的Scheme-based约束下的一种Scheme-less扩充。Sparrow将业务数据对象统一设计为目录节点，并构建数据对象目录树，其结构如下：

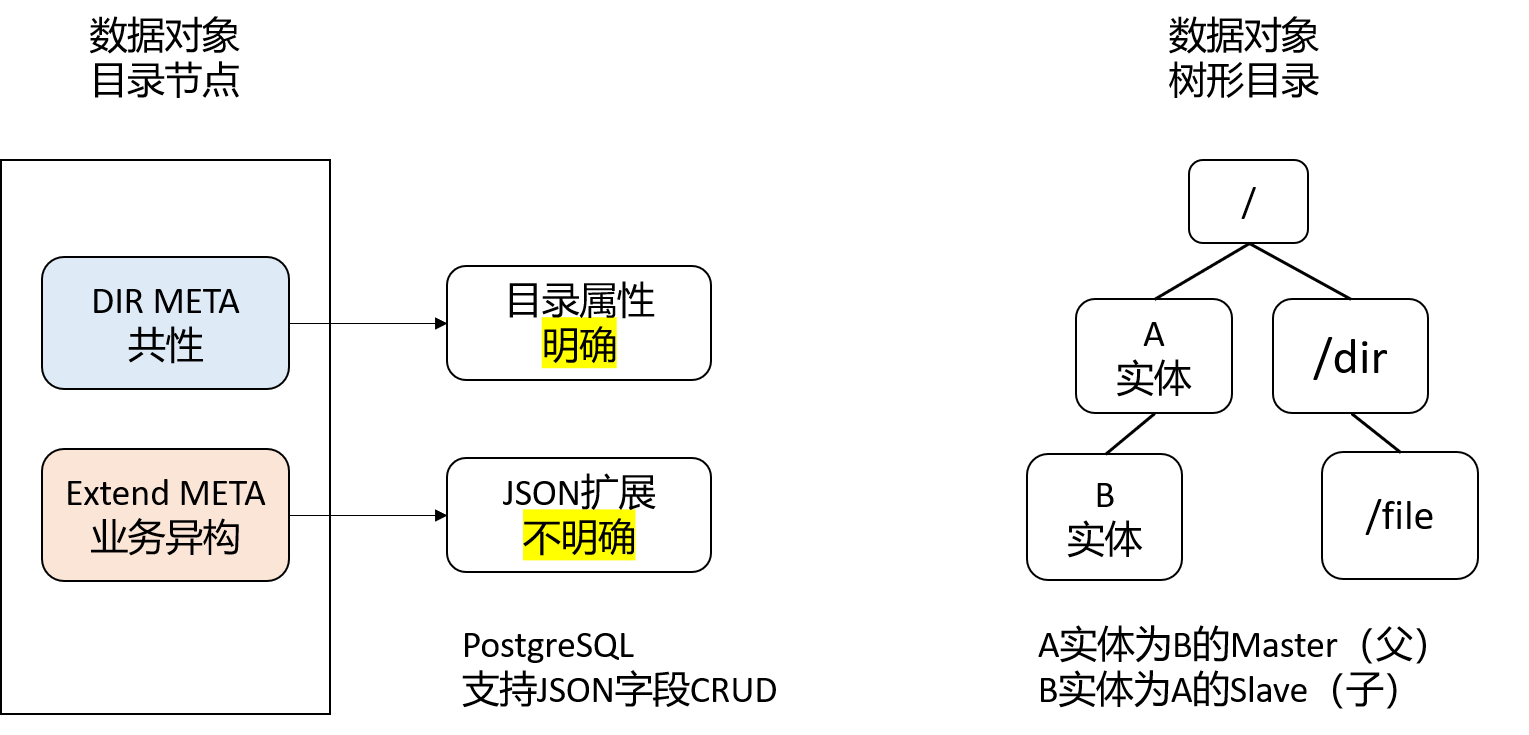


图 \* 解决业务数据异构性

上图所示，一个数据对象目录节点，包含目录属性（DIR META）和业务JSON扩展属性（Extend META）。目录属性是共性的，即非业务相关的，主要表达目录属性，包括目录唯一标志符、目录类型（普通目录节点或者文件节点）、创建者、是否开放访问等。业务扩展属性，则是利用PostgreSQL对JSON数据类型的支持，将异构的业务数据对象系列为相应JSON存储在相应的JSON扩展属性中。

下表\*描述数据对目录节点的属性及二维表结构：

表 \* 数据对象目录节点

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 名称 | 列名 | 约束 | 说明 |
| 0 | Id | 行Id | 自增 | 便于运维 |
| 1 | SpaId | 目录-标志符 | 字符串{36} | UUID，主键（如何缩短） |
| 2 | ReferredSpaId | 关联目录SpaId | 字符串{36} | 在访问控制上，访问本节点，等同于访问ReferredSpaId指向的目录节点 |
| 3 | CustId | 租户Id | 字符串{36} | 多租户支持，本数据目录节点所属租户 |
| 4 | Name | 名称 | 字符串{25} | 用户指定，或由按业务规则系统生成 |
| 5 | Type | 类型 | 字符串{8} | sysroot（系统根目录）  custroot（租户根目录）  userroot（用户根目录）  dataroot（业务数据根目录）  custdir（租户数据根目录），  busdir（租户-业务数据目录）  orgdir（租户-组织机构数据目录）  bindir（租户-垃圾数据回收目录）  dir（目录）  file（文件） |
| 6 | CreatorId | 创建者 | 字符串{36} |  |
| 8 | CreateAt | 创建时间 | 长整数 | 时间戳的长整数 |
| 9 | ModifyAt | 最新修改时间 | 长整数 | 时间戳的长整数 |
| 10 | Access | 访问状态 | 整数 | {0（关闭）, 1（正常）} |
| 11 | FileId | 外部  Minio  或者  阿里云OSS Id | 字符串{\*\*} | 当type为file时，其需要在Mino或者阿里云OSS中进行存储，OSS文档Id |
| 12 | MineType | 文件类型 | 字符串{8} | 文件的多媒体类型  {mine, e.g. jpeg} |
| 13 | Thumbnail | 缩略图路径 | 字符串{256} | 缩略图路径  Dir：  File： |
| 14 | ExtendType | 用户扩展类型 | 字符串{16} | 用户扩展数据类型 |
| 15 | ExtendData | 用户扩展字段 | JSON  or  JSONB | 支持用户自定义扩展，PostgreSQL支持JSON数据类型的CRUD与基本检索 |
| 16 | ExtendIndex | 外部  ElasticSearch Id | 字符串{\*\*} |  |

Sparrow使用外联表设计，实现数据对象目录树的构造。该Master-Slave表结构如下：

表 \* 目录树形结构Master-Slave

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 名称 | 列名 | 约束 | 说明 |
| 0 | Id | 行Id | 自增 | 便于运维 |
| 1 | MasterId | 父节点id | 字符串{36} | UUID，DirId |
| 2 | SlaveId | 子节点id | 字符串{36} | UUID，DirId |

* **Sparrow内置数据对象**

访问控制是信息管理系统的基础模块，涉及租户、用户与授权管理。这些内置的数据对象，同样包装为目录节点，并存在于目录树中。

* **系统管理员（Sparrow Root）**：Sparrow初始化后，默认存在一个全系统域root用户，拥有整个系统资源的全部权限，并在业务过程中，创建其他租户和用户。系统管理员是一个特殊的系统租户与用户。
* **租户（Customer Root）**：租户根据服务协议使用Sparrow各项服务；Sparrow系统支持多租户。

表 \* 租户数据对象

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 名称 | 列名 | 约束 | 说明 |
| 1 | Name | 名称 | 字符串{25} | 用户定义 |
| 2 | Type | 类型 | 字符串{8} | System：系统管理租户  Normal：普通租户 |
| 3 | CreateAt | 创建时间 | 长整数 | 时间戳的长整数 |
| 4 | CustDirId | 租户数据根目录 | 字符串{36} |  |

* **用户，指系统使用者。**

用户类型：

1. Sparrow Root 系统管理员，具有整个Sparrow的完全权限；
2. System User 系统用户，由Sparrow Root创建与管理，并依据授权进行系统访问；
3. Customer Root 租户管理员，具有租户所辖范围的完全权限；
4. Customer User 租户用户，由Customer Root创建与管理，并依据授权进行系统访问。

表 \* 用户数据对象

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 名称 | 列名 | 约束 | 说明 |
| 1 | Name | 登录名 | 字符串{25} | 用户定义 |
| 2 | Type | 类型 | 字符串{8} | sysroot，sysuser，custroot，custuser |
| 3 | CreateAt | 创建时间 | 长整数 | 时间戳的长整数 |
| 4 | HomeDirId | 用户数据根目录 | 字符串{36} |  |
| 5 | CustId | 租户Id  租户目录节点的SpaId | 字符串{36} | 多租户支持 |

Sparrow采用面向资源操作许可的访问控制策略。许可是访问主体（用户/群组）、访问客体（资源/目录节点）与操作的三元组。一个许可，表示用户可对该资源实施绑定的操作行为。

表 \* 许可数据模型

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 名称 | 列名 | 约束 | 说明 |
| 1 | Name | 名称 | 字符串{25} | 用户定义 |
| 2 | SpaId | 资源（目录）Id | 字符串{36} | 访问客体：目录或子目录树 |
| 3 | Type | 类型 | 字符串{8} | {simple（细粒度权限），supper（层次权限继承）} |
| 4 | PathRegex | 目录授权正则表达式 | 字符串{255} | 当type=simple时，DirPathRegex为SpaId的DirPath；  当type=supper时，DirPathRegex为SpaId的DirPath\*；  DirPath：从根节点到本节点的路径 |
| 5 | Setting | 成员设置（主体） | 字符串{8} | Positive：默认包含全部Members  Negative：默认不包含全部Members |
| 6 | Members | 成员列表 | JSON | 访问主体：  99个用户Id列表；JSON格式数组  当Setting为positive时，Members表示不包含这些用户（黑名单）；  当Setting为negative时，Members表示包含这些用户（白名单）； |
| 7 | Operations | 操作许可 | 字符串{16}二进制字符串 | 操作：  见：访问控制章节 |
| 8 | CreatorId | 创建者 | 字符串{36} | 表 2 UserId |
| 9 | CreateAt | 创建时间 | 长整数 | 时间戳的长整数 |
| 10 | Invalid\_at | 有效期至 | 长整数 | 时间戳的长整数  长期有效:  2999年12月31日23时29分 |
| 11 | CustId | 租户Id  租户目录节点的SpaId | 字符串{36} | 多租户支持 |

* **Sparrow数据对象目录树**

Sparrow为所有数据对象构建目录树的组织结构，下图展示一个案例：

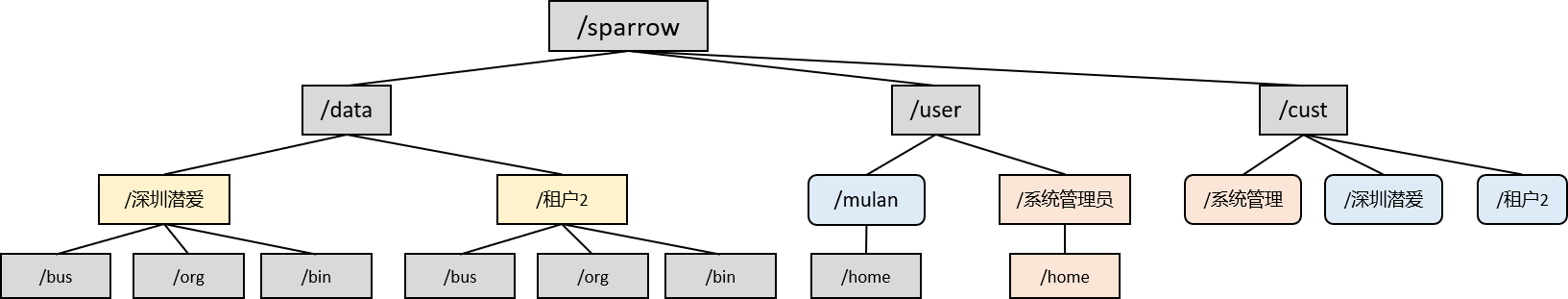


图 \* 数据对象目录树

在数据对象目录树中，/sparrow作为系统根目录，其下设置有/cust租户根目录，/user用户根目录，和/data数据根目录。

/cust目录下包含所有租户数据对象，其中/系统管理作为系统管理员对应的租户，/深圳潜爱，是深圳潜爱组织的业务系统对应的租户。

/user目录下包含所有的系统用户，如/系统管理员用户。每个用户目录下，包含一个/home个人数据目录。

/data下包含每个租户的数据根目录，如/深圳潜爱。在租户数据根目录下，包含/bus业务数据根目录，/org组织机构根目录和/bin数据对象回收根目录。

注意，目录树中的每个节点所包装的数据对象类型可能不同，如/sparrow/user/mulan节点包含的是系统用户数据对象，/sparrow/cust/深圳潜爱节点包含的是租户数据对象，而/sparrow/data/深圳潜爱，则仅仅是目录节点，并不包装任何业务数据对象。

在业务系统实施中，如深圳潜爱的残枝培育项目，其所有业务数据均挂载在/sparrow/data/深圳潜爱/bus目录下，其结构由具体业务决定，其一实施方案：

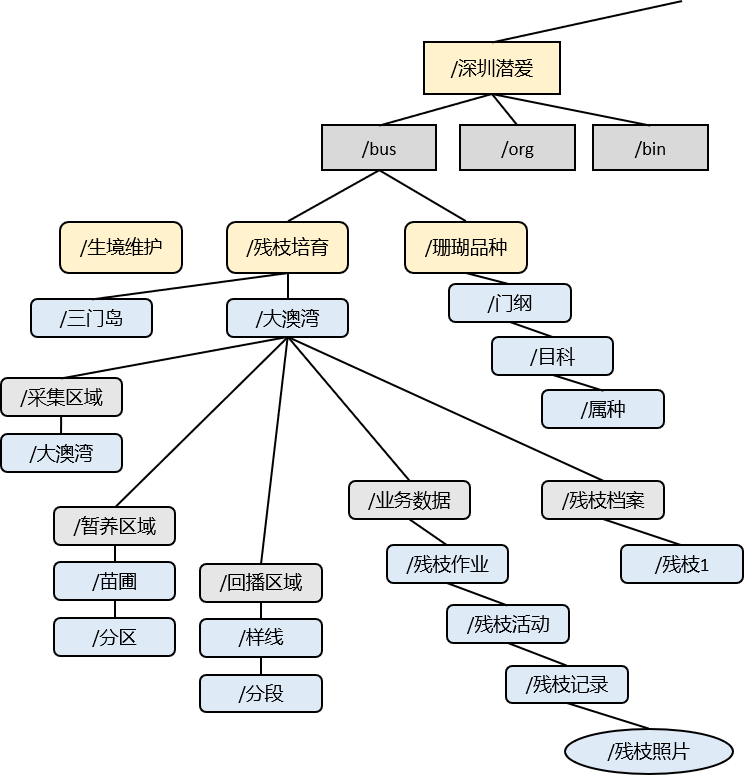


图 \* 深圳潜爱-残枝培育项目数据对象目录树方案

问题：目录节点的主键，采用字符串长度为36的UUID，导致叶子节点的路径过长。考虑，更短的唯一主键方案。