Moe HBase 设计

1. RowKey：

所有的表都使用MD5(info\_date)\_(spider\_time)作为主键

1. 表模式设计：

Spider\_info表:

|  |  |  |
| --- | --- | --- |
| Rowkey | spiderName | time |
|  | moeGov | Starttime |
|  |  | endtime |

Declaration(文件公告)表：

|  |  |  |  |
| --- | --- | --- | --- |
| RowKey | 列族 | | |
| MD5(info\_date)\_MD5(url) | **Info** | **file** | **File\_name** |
| Spider\_path | File\_num | File\_num |
| position | MD5(file\_url): | MD5(file\_url): |
| Author | MD5(file\_name1) | … |
| publish\_date | … |  |
| Info\_title | MD5(file\_namen) |  |
| Info\_index |  |  |
| Info\_date |  |  |
| Info\_institution |
| Document\_num |
| Info\_category |
| Info\_abstract |
| Info\_content |
| File\_url |
| QRCode\_url |
|  | b\_QRCode |  |  |

News(新闻动态)表:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RowKey | 列族 | | | |
| MD5(info\_date)\_MD5(url) | **Info** | **file** | **File\_name** | **picture** |
|  | url | file\_num | File\_num | Picture\_num |
| Spider\_path | MD5(file\_url): | MD5(file\_url): | 0: |
| Position |  |  | 1: |
| news\_yt | **…** |
| News\_title |  |
| news\_sub\_title |
| news\_fwzh |
| News\_time |
| News\_source |
| News\_editor |
| News\_content |
|  | Picture\_urls |  |  |  |
|  | File\_urls |  |  |  |
|  | QRCode\_url |  |  |  |
|  | b\_QRCode |  |  |  |

Data(数据)表：

|  |  |  |  |
| --- | --- | --- | --- |
| RowKey() | 列族 | | |
| MD5(info\_date)\_MD5(url) | **Info** | **File\_name** | **File** |
| url | File\_num | File\_num |
|  | Path | MD5(file\_url0) | MD5(file\_url0) |
|  | Position | … | … |
|  | date |  |  |
|  | Title\_yt |  |  |
|  | Data\_title |  |  |
|  | Sub\_title |  |  |
|  | title\_fwzh |  |  |
|  | Data\_content |  |  |
|  | ~~Remarks~~ |  |  |
|  | Editor |  |  |
|  | File\_urls |  |  |
|  | QRCode\_url |  |  |
|  | b\_QRCode |  |  |

Conference(发布会)表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rowkey | 列族 | | | | |
| MD5(info\_date)\_MD5(url) | Info | agenda | material | Picture\_news | news |
| url | url | materialNum | News\_num | News\_num |
| Path |  | url0:time | url0:time | url0:time |
| position |  | … | … | … |
| Date |  |  |  |
| Author |
| source |
| title |
| Time |
| site |
| compere |
|  | member |  |  |  |  |
|  | content |  |  |  |  |
|  | Picture\_text |  |  |  |  |
|  | Picture\_url |  |  |  |  |
|  | b\_picture |  |  |  |  |

create 'moe\_declaration', {NAME => 'info', VERSIONS => 3}, {NAME => 'file', VERSIONS => 3}, {NAME => 'file\_name', VERSIONS => 3}

create 'moe\_news', {NAME => 'info', VERSIONS => 3}, {NAME => 'file', VERSIONS => 3}, {NAME => 'file\_name', VERSIONS => 3}, {NAME => 'picture', VERSIONS => 3}

create 'moe\_data', {NAME => 'info', VERSIONS => 3}, {NAME => 'file', VERSIONS => 3}, {NAME => 'file\_name', VERSIONS => 3}

create 'moe\_conference', {NAME => 'info', VERSIONS => 3}, {NAME => 'agenda', VERSIONS => 3}, {NAME => 'material', VERSIONS => 3}, {NAME => 'picture\_news', VERSIONS => 3}, {NAME => 'news', VERSIONS => 3}