Moe HBase 设计

1. RowKey：

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

1. 表模式设计：

Spider\_info表:

|  |  |  |
| --- | --- | --- |
| Rowkey | spiderName | time |
|  | Sanqin | Starttime |
|  |  | endtime |

News表：

|  |  |  |
| --- | --- | --- |
| Rowkey | Info | Pictures |
|  | url | Picture\_num |
|  | Path\_url | MD5(pic\_url):b\_picture[0] |
|  | Path\_text |  |
|  | Title |  |
|  | Publish\_time |  |
|  | Editor |  |
|  | Abstract |  |
|  | Source\_text |  |
|  | Source\_href |  |
|  | Picture\_urls |  |
|  | Content |  |
|  | keywords |  |