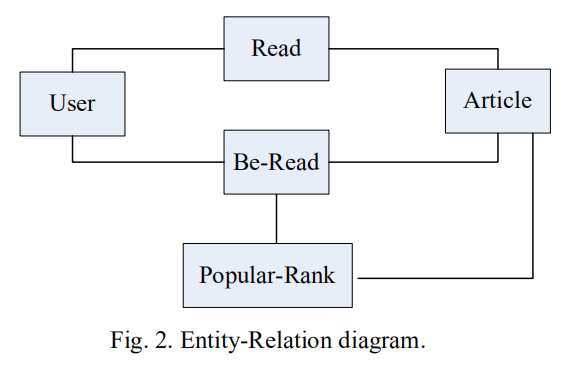
**Distributed Databases**

Data to be managed and processed include structured data (5 relational tables) and unstructured data (text, images, and video). Their inter-relations are illustrated in Fig. 2



(1)User table

id, timestamp, uid, name, gender, email, phone, dept. grade, language, region,role, preferTags, obtainedCredits

User table is fragmented based on region attribute, where region=”Beijing” allocated in DBMS1 and region= “HongKong” allocated in DBMS2

(2)Article table

id, timestamp, aid, title, category, abstract, articleTags, authors, language, text, image, video

Article table is fragmented based on article category attribute, where

category=”science” allocated in DBMS1 and DBMS2, category=“technology” allocated in DBMS2

(3)Read table---无副本

id, timestamp, uid, aid, readTimeLength, readSequence, readOrNot, aggreeOrNot, commentOrNot, commentDetail, shareOrNot

Read table is fragmented based on User table without replica, and with the same allocation schema as User table

(4)Be-Read table

id, timestamp, aid, readNum, readUidList, commentNum, commentUidList, agreeNum, agreeUidList, shareNum, shareUidList

Be-Read table contains two fragments based on Article table with duplication, where category=”science” allocated to DBMS1 and DBMS2,

category=“technology” allocated to DBMS2.

(5)Popular-Rank table

id, timestamp, temporalGranularity, articleAidList

// temporalGranularity= “daily”, “weekly”, or “monthly”

Popular-Rank table contains two fragments based on Article table with duplication, where category=”science” allocated to DBMS1 and DBMS2,

category=“technology” allocated to DBMS2

**3. Databased Insert and Query Operations**

1) Bulk load(批量加载)User table, Article table, and Read table into the data center

2) Query users, articles, users’ read tables (involving the join of User table and Article table) with and without query conditions[有或没有查询条件]

3) Populate(填充) the empty Be-Read table by inserting newly computed records into the Be-Read table.

4) Query the top-5 daily/weekly/monthly popular articles with articles details (text, image, and video if existing) (involving the join of Be-Read table and Article table)

**4.Data Center**

**5.Project Requirement**

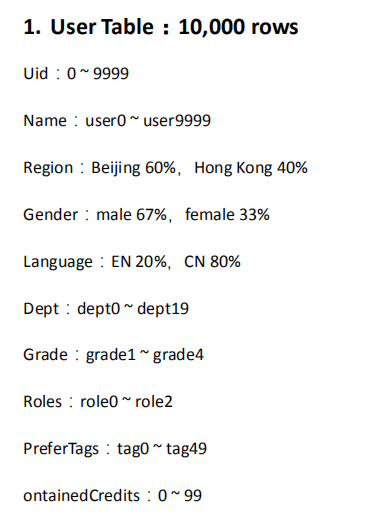
Implement a data center in a distributed context (Fig. 1) with the following functionalities.

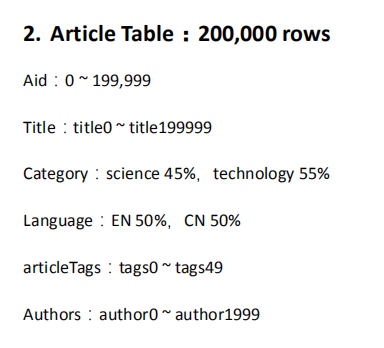
1) Bulk data loading with data partitioning and replica consideration(考虑数据分区和复制的大容量数据加载)

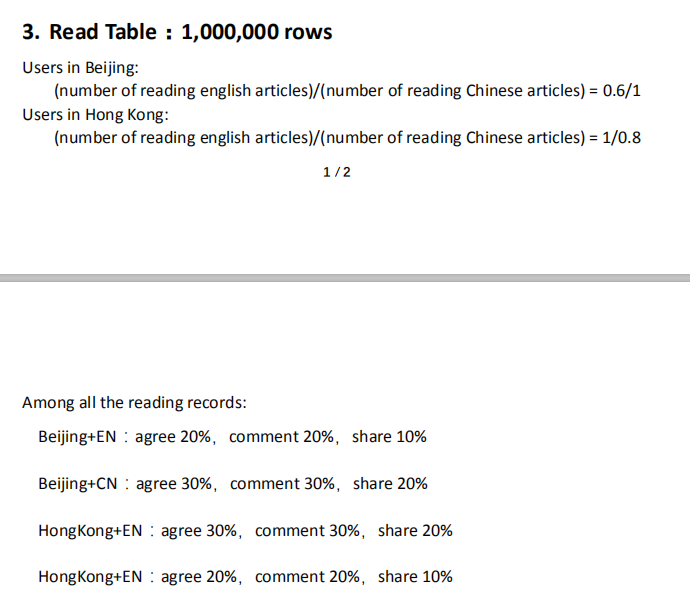
2) Efficient execution of data insert, update, and queries

3) Monitoring the running status of DBMS servers, including its managed data (amount and location), workload, etc. (监控DBMS服务器的运行状态，包括其托管数据(数量和位置)、工作负载等。)

#### Data used in project







#### ProjectScenarios

E-Bibliotherapy

#### 系统实现

（1）The system composes of two layers: client interface and data center

（2）实现一个命令行用户界面描述可能的用户操作和数据的交互。

**（3）数据中心由两组Redis cache和MongoDB对组成，表示在Docker容器中运行的不同位置以及在虚拟机中运行的HDFS**

**This project will set up an electronic management system to provide users with a better reading and destressing experience. This system is designed to provide a smooth user experience and support various user operations**

#### 所需技术

1. Hadoop—在虚拟机上
2. 数据存储---MySQL---standard storage in tables, primary and foreign keys to establish relationship, similar syntax to operate, and support for scalability and high-performance (Upwork, n.d.).

MongoDB---NoSQL databases, it has collections and JSON-like documents

1. cache—provides faster querying speed when dealing with a lot of users performing similar operations at the same time. **Redis**
2. Hadoop HDFS用于存储非结构化数据，例如文本，图像和视频