**数据压缩博士申请课题**

**CompressDB: Enabling Efficient Compressed Data Direct Processing for Various Databases**

**一、主要工作**

1.我们直接在压缩数据上开发高效的数据操作，例如插入、删除和更新。

We develop efficient data manipulation operations, such as insert, delete, and update, directly on compressed data.

除了之前的随机访问支持外，我们还启用了数据查询和数据操作。

Along with previous random access support, we enable both data query and data manipulation.

2.我们开发了CompressDB，一个集成到文件系统中的存储引擎。

We develop CompressDB, a storage engine that is integrated into file systems.

CompressDB可以在不修改数据库的情况下无缝支持各种数据库系统。

CompressDB can support various database systems seamlessly without modifying the databases.

3.我们将操作下推到存储系统，避免了不必要的数据在内存和磁盘之间移动，从而提高了压缩数据的处理效率。

We enable operation pushdown to storage systems, which avoids unnecessary data movement between memory and disks, thus improving processing efficiency on compressed data.

**二、实现思路**

由于大数据通常存储在磁盘中，因此我们的想法是基于规则压缩在存储层对压缩后的数据进行随机更新。

Since large data are usually stored in disk, our idea is to develop random update over compressed data in storage layer based on rule compression.

在元素级别，我们引入了数据洞的概念，以允许在大数据块中进行更新。

in element level, we introduce the concept of data holes to allow updates in large data blocks.

在规则层，我们开发了哈希和计数数据结构，以有效地定位规则。

In rule level, we develop hashing and counting data structures for efficiently locating rules.

在DAG级别，我们限制了DAG的深度，以将规则拆分和合并的成本保持在一个小范围内。

In DAG level, we limit the depth of the DAG to retain the cost of rule split and merge within a small range.