

## ارائه روش نوین فشرده سازی ترکیبی برای بهینه سازی مصرف حافظه و سرعت دسترسی در پایگاه داده ابری

سید بهنام حسینی<sup>۱</sup>

کارشناسی ارشد، دانشکده علوم کامپیوتر، دانشگاه پویش، قم.

### چکیده

شرکتها و ارائه کنندگان خدمات ابری افزایش چشمگیری در میزان حجم داده ذخیره شده در فضای ابر عمومی و خصوصی را داشته اند. بنابراین هزینه ذخیره سازی داده ها در حال رشد است. زیرا آن ها از تنها عملکرد لایه ذخیره سازی برای ذخیره سازی تمام داده های ابری استفاده می کنند. این موضوع قابل توجه است که برای کاهش هزینه های ابری، داده ها را به دو دسته فعال (گرم) و غیرفعال (سرد) طبقه بندی می کنند. با توجه به پژوهش در حافظه اصلی پایگاه داده ها، آثار اخیر در مورد روش های متمرکز برای شناسایی داده های گرم و داده های سرد است. روش پیگیری تاپل شناسایی است و به شناسایی گرم و سرد شناخته می شود. در مقابل، ما یک رویکرد جدید به نام روش LOAD DATA INFILE که آهنگ هر دو تاپل ها در دسترسی ذخیره سازی ثانویه پایگاه های داده است را معرفی می کنیم. هدف ما به منظور افزایش عملکرد در شرایط سه بعد است که عبارتند از: فضای ذخیره سازی، زمان سپری شده پرس و جو، و مصرف CPU. به منظور تأیید اثربخشی روش مان، پیاده سازی دقیق آن را بر روی رویکرد LOAD DATA INFILE Approach (LDA) متوجه شدیم که سطرها را از یک فایل متنی داخل یک جدول با سرعت بسیار بالا با استفاده از معیار شناخته شده qps و TPC-H می خواند. نتایج تجربی نشان می دهد که عملکرد رویکرد LOAD DATA INFILE در مقایسه با روش prepare\_data، رویکرد بهتری را نسبت به دو بعد عملکرد دارد. به طور خاص، در عملکرد LOAD DATA کاهش فضای ذخیره سازی ۱۴ الی ۶۲ درصد به طور متوسط صورت گرفته و کاهش متوسط زمان پرس و جوی سپری شده نیز ۲۸۰ الی ۴۴۰ برابر نسبت به رویکرد سنتی پایگاه داده است.

کلمات کلیدی: رایانش ابری، ذخیره سازی ابری، حافظه اصلی پایگاه داده، داده گرم/سرد، مدیریت داده سرد.

## The presentation of new hybrid compression techniques to optimize memory usage and speed of access in cloud database

Seyd Behnam Hoseini<sup>1</sup>

<sup>1</sup>Pooyesh University, Qom, Iran.

### Abstract

Enterprises and cloud service providers face dramatic increase in the amount of data stored in private and public clouds. Thus, data storage costs are growing hastily because they use only one single high-performance storage tier for storing all cloud data. There,s considerable potential to reduce cloud costs by classifying data into active (hot) and inactive (cold). In the main-memory databases research, recent work focus on approaches to identify hot/cold data. Most of these approaches track tuple accesses to identify hot/cold tuples. In contrast, we introduce a novel LOAD DATA INFILE that tracks both tuples and columns accesses in secondary storage databases. Our objective is to enhance the performance in terms of three dimensions: storage space, query elapsed time and CPU usage. In order to validate the effectiveness of our approach, we realized its concrete implementation on LOAD DATA INFILE Approach (LDA) that reads rows from a text file into a table at a very high speed by using the well-known qps and TPC-H benchmark. Experimental results show that the proposed LOAD DATA approach outperforms prepare\_data in respect of two performance dimensions. In specific, LOAD DATA reduces the storage space by average of 14-62% and reduces the query elapsed time by average of 280-440 times compared to the traditional database approach.

**Keywords:** Cloud computing, Cloud storage, Main-memory database, Hot/cold data, Cold data management.

### تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۰/۰۱/۱۶

تاریخ اصلاحات: ۱۴۰۰/۰۳/۰۷

تاریخ پذیرش: ۱۴۰۰/۰۳/۲۳

تاریخ انتشار: ۱۴۰۰/۰۶/۳۱

### Keywords:

Cloud computing,

Cloud storage,

Main-memory database,

Hot/cold data,

Cold data management

\*ایمیل نویسنده مسئول:

Behnam.hoseini1989@gmail.com