

## Promovendus maakt verwerking Big Data stukken sneller

27 OKTOBER 2015 17:14 | [RICHARD KEIJZER](#) | [NIEUWS](#)



Promovendus Sándor Héman ontwikkelde een methode om een omvangrijke database te comprimeren waardoor de gegevens veel sneller van de harde schijf naar de rekenkern van een computer worden overgebracht. Ook ontwikkelde hij efficiënte algoritmes om wijzigingen te kunnen aanbrengen in deze gecompriëerde data.

Op 28 oktober verdedigt hij zijn proefschrift hierover op woensdag 28 oktober aan de Vrije Universiteit. Het onderzoekswerk is gedaan bij het Centrum voor Wiskunde en Informatica (CWI) te Amsterdam.

### Dubbele taak

Veel organisaties verzamelen tegenwoordig enorme hoeveelheden digitale informatie, en slaan deze op in een databasemanagementsysteem. Deze database wordt continu gewijzigd en aangevuld, maar ook moet op elk moment geanalyseerd kunnen worden. Dit zijn twee verschillende bewerkingen, die verschillende eisen stellen aan de architectuur van het databasesysteem.

### Traagheid

Het transport van data van disk naar de computer vergt doorgaans veel tijd. Via compressie kan die tijd sterk worden bekort, zodat een flessenhals wordt vermeden. "In mijn onderzoek heb ik de negatieve gevolgen van zo'n flessenhals gereduceerd, door de omvang van een database te verkleinen door middel van data compressie", zegt Héman.

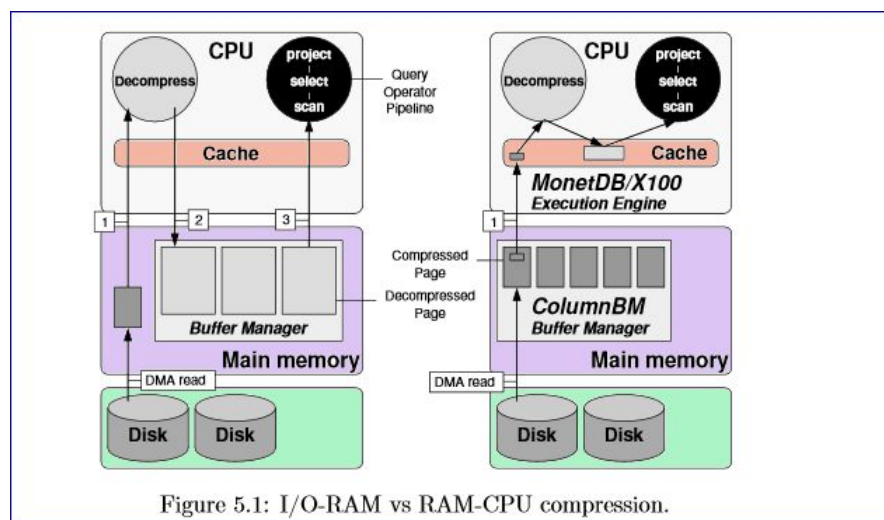
### Wijzigingen apart behandelen

Als er wijzigingen in een gecompriëerde database moeten worden aangebracht, is het omslachtig om de data eerst te moeten decomprimeren, wijzigen, hercomprimeren, en weer opnieuw weg te schrijven. Hiertoe heeft Héman technieken ontwikkeld die wijzigingen differentieel bijhouden, zoals errata bij een boek, maar dusdanig dat deze tijdens het lezen direct en efficiënt worden toegepast.

### Toepasbaarheid

Hémans onderzoek is relevant voor elk domein waar grote hoeveelheden data geanalyseerd dienen te worden. Dit zijn bijvoorbeeld zoekmachines die het hele web indexeren en doorzoeken, wetenschappelijke toepassingen, zoals analyses van astronomische data of genomen, of commerciële toepassingen waarbij de interesses van grote hoeveelheden klanten in kaart moeten worden gebracht.

Een overzicht van de oude en nieuwe werkwijze:



Share

27

Tweet

## Lees ook

[CWI herstelt fundamentele fout in Java](#) 25 FEBRUARI 2015