



UvA-DARE (Digital Academic Repository)

MonetDB/DataCell: leveraging the column-store database technology for efficient and scalable stream processing

Liarou, E.

Publication date
2013

[Link to publication](#)

Citation for published version (APA):

Liarou, E. (2013). *MonetDB/DataCell: leveraging the column-store database technology for efficient and scalable stream processing*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Samenvatting

Vandaag de dag moeten online analytische programmas kunnen omgaan met een snelle stroom van gegevens. Bijvoorbeeld, toepassingen in de mobiele sector proberen de stroom van gegevens te gebruiken voor advertenties en routing. In dezelfde lijn vereisen grootschalige Cloud infrastructures een continue monitoring om stabiliteit te waarborgen en cyberaanvallen te kunnen pareren. Wetenschappelijk databanken en web-log analyses vereisen een efficiënte verwerking voor decision support.

Het afhandelen van langlevende queries (continuous queries) en het snel analyseren van grote data stromen in combinatie en vergelijking met reeds opgeslagen informatie kan nog niet goed met de bestaande database en streaming technologie worden uitgevoerd. Database systemen missen de functionaliteit voor verwerking van continuous queries en data streaming systemen schalen niet. Dit nieuwe probleem vereist een oplossing die de beste eigenschappen van beide werelden combineert.

In dit proefschrift wordt een ontwerp besproken hoe data stromen kunnen worden verwerkt in een modern kolom-georiënteerde database systeem. In het bijzonder richten we ons hier op het open-source systeem MonetDB als platform voor ontwerp experimentatie. Het omvat aanpassingen in zowel het opslag deel, de verwerkingskern, als ook de optimizers. De mogelijkheden worden op een rij gezet en geanalyseerd om de beste richting te kunnen bepalen. De grootste winst wordt gehaald bij 'window-based' verwerking van de data stroom.

We laten aan de hand van de Linear Road Benchmark zien dat het prototype een uitstekende performance biedt en ook in vergelijking met state-of-the-art commerciële systemen. Deze resultaten maken het mogelijk om schaalbare gegevensverwerking te verkrijgen voor zowel stromende als persistente gegevens in een geïntegreerde, moderne datawarehouse omgeving.