

**EXAMENSARBETE** Row vs. column data layout in a graph database query engine**STUDENT** Eric Sporre**HANDLEDARE** Henrik Nyman (Neo4j), Luigi Nardi (LTH)**EXAMINATOR** Jacek Malec (LTH)

# Alternativ minnesrepresentation för snabbare grafdatabaser

POPULÄRVETENSKAPLIG SAMMANFATTNING **Eric Sporre**

Grafdatabaser blir allt vanligare och är essentiella för allt mellan snabba rekommendationssystem till att tolka de ökända Panamadokumentet. I det här arbete undersöks en alternativ representation inom datorminnet som kan leda till snabbare resultat.

På moderna processorer sitter det flera temporära minnen i olika storlekar. De minsta är de absolut snabbaste medan de större är långsammare. Då de temporära minnena tenderar att vara små är det viktigt att anpassa den data som passerar genom processorn så att den passar om man vill uppnå hög prestanda. Om processorn är ute efter data som inte finns i sitt temporära minne tvingas den hämta den från antingen primärminnet (RAM) eller i värsta fall disk vilka båda är magnituder långsammare. Det kan jämföras med ifall man har alla ingredienser inför matlagning i kylan eller ifall man behöver åka till affären och handla.

Jag har arbetat med en alternativ sortering av block av data, tänk en tabell med rader och kolumner som nu behöver representeras på bara en rad. Normalt sorteras dessa block radvis men istället har jag försökt sortera dem utefter kolumnerna. Bilden visar ett mindre färgkodat exempel. Den bakomliggande tanken är att detta skulle kunna leda till att mer intressant data för processorn hamnar nära varandra då alla värden inom varje kolumn hör samman.

Den nya sorteringen testades för vissa mindre delar inom grafdatabasen och visade på lo-

vande resultat ifall den kombinerades med nya metoder som kan dra ytterligare nytta av sorteringen. Dessa nya metoder kan utnyttja processorn bättre och föredrar den nya sorteringen. Det krävs dock mer undersökning innan man definitivt kan övergå till den nya sorteringen då den bara testats inom en mindre del av grafdatabasen.

1	2	3	4
5	6	7	8

Tabell 1: En tabell med data som ska representeras på en rad istället för två

1	2	3	4	5	6	7	8
---	---	---	---	---	---	---	---

Tabell 2: Radvis sortering av datatabellen

1	5	2	6	3	7	4	8
---	---	---	---	---	---	---	---

Tabell 3: Kolumnvis sortering av datatabellen