

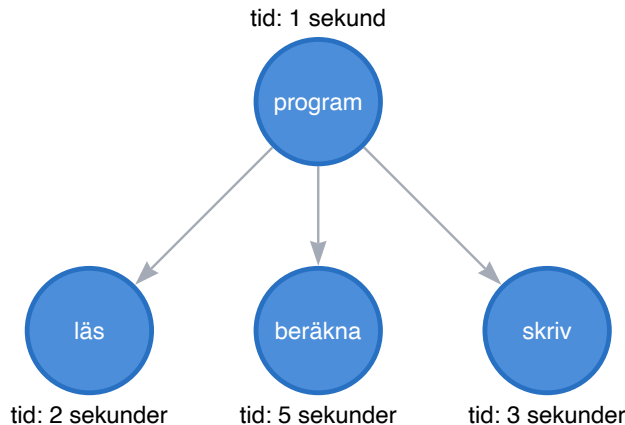
EXAMENSARBETE Modellering Profiling Data in a Graph Database for Performance Analysis**STUDENT** Richard Lundberg, Marcus Rettig**HANDLEDARE** Jonas Skeppstedt (LTH), Simon Priisalu (Neo4j), Jaroslaw Palka (Neo4j)**EXAMINATOR** Per Andersson (LTH)

Modellering av profileringsdata i en grafdatabas för prestandaanalys

POPULÄRVETENSKAPLIG SAMMANFATTNING **Richard Lundberg, Marcus Rettig**

Att identifiera långsam kod i mjukvara är en viktig, men tidskrävande uppgift. Vi undersöker hur profileringsdata kan modelleras i en grafdatabas för att underlätta processen att identifiera flaskhalsar och regressioner.

Programkod brukar delas in i mindre bitar, så kallade *funktioner*. Tiden det tar att köra hela programmet är summan av hur mycket tid som spenderas i varje funktion. För att optimera programmet vill man därför hitta de mest tidskrävande funktionerna och försöka förbättra dessa.



Genom att köra ett program med ett så kallat *profileringsverktyg* kan man göra mätningar kring hur funktioner anropar varandra och hur mycket tid som spenderas inuti dessa. Resultatet brukar kallas för *profileringsdata* och kan visualiseras i en trädliknande struktur, till exempel som i bilden. På företaget Neo4j görs hundratals sådana mätningar varje dag vilket resulterar i tusentals träd, där varje träd består av tusentals funktioner.

Vi har undersökt hur man kan analysera stora mängder profileringsdata genom att lagra den i en grafdatabas. De mest populära databaserna just nu representerar data i tabellformat, men en grafdatabas lagrar istället data som noder och relationer. Till exempel kan noder representera personer i ett socialt nätverk och relationerna kan representera vänskap mellan två personer. En grafdatabas passar även utmärkt för de trädliknande strukturer som beskrevs ovan eftersom sådana strukturer är problematiska att representera i tabellformat.

I en grafdatabas kan användaren ställa frågor, till exempel: "I vilka testkörningar tog denna funktionen mer än två sekunder?", eller "Hur lång tid tog denna funktionen idag jämfört med förra veckan?". Svaret på den första frågan kan identifiera funktioner som ofta presterar dåligt och därför påverkar prestandan för hela systemet, så kallade flaskhalsar (eng. bottlenecks). Svaret på den andra frågan kan identifiera funktioner som går långsammare idag än de gjorde för en vecka sedan, så kallade regressioner (eng. regressions).

Vårt arbete visar att en grafdatabas är ett lämpligt verktyg för att modellera profileringsdata. Vår föreslagna databasstruktur går att använda för flertalet analysmetoder och ökar tillgängligheten till data insamlad över tid.