

**EXAMENSARBETE** Big Data Business Rules**STUDENT** Richard Stenlander**HANDLEDARE** Matteo Casalino (Amadeus), Peter Exner (LTH), Pierre Nugues (LTH)**EXAMINATOR** Björn Regnell (LTH)

# Parallell evaluering av Business Rules

POPULÄRVETENSKAPLIG SAMMANFATTNING **Richard Stenlander**

Business Rules är regler som definierar en del av en affärsverksamhet. I stora IT-system är det viktigt att snabbt kunna hitta den regel som skall appliceras, en uppgift som blir allt mer tidskrävande när antalet regler ökar.

Tänk dig att du har en kortlek som saknar ett kort. Ett sätt att hitta kortet som saknas är att gå igenom kortleken och sortera korten efter valör. I slutet kommer en valör vara representerad av tre kort, och då är det enkelt att se vilket kort som saknas. Detta problem är såpass enkelt att en person kan lösa det effektivt helt själv, men vad händer om vi istället har 100 kortlekar, och ett kort saknas? Vad händer när antalet kortlekar blir miljontals? Det är problem som dessa som uppstått med teknologins framfart. Mängden data har ökat explosionsartat och det är inte längre hållbart att lösa vissa problem på samma sätt som tidigare.

Detta var fallet för Amadeus, ett multinationellt företag verksam i reseindustrin som sett en signifikant ökning av mängden data. Amadeus använder sig av ett Business Rules-system för att hantera logiken i sin affärsverksamhet. Business Rules är regler som t.ex. säger: i scenario X gör Y. Med mer data kan fler sådana regler skapas, vilket utvecklar affärsverksamheten och för den framåt. Men när antalet regler ökar, tar det längre tid att hitta den bästa regeln att tillämpa. Amadeus Business Rules-team var således intresserade av att undersöka om regelsökningen kunde effektiviseras.

Låt oss nu återgå till vårt kortleksexempel, och föreställa oss att vi har fler personer som kan gå igenom det ökade antalet kortlekar. Har vi t.ex. en person per kortlek, så skulle den totala tiden för att hitta det kort som saknas knappt påverkas när antalet kortlekar ökar. Detta är den fundamentala idén bakom nya teknologier såsom Apache Spark, som har utvecklats för att lösa problem där data-

mängden har börjat bli överväldigande.



Cliparts from [www.pizabay.com](http://www.pizabay.com) under the Creative Common License.

En illustration av effektivisering baserad på parallellisering.

Alla uppgifter lämpar sig inte lika bra för parallell processering som detta kortexempel. Är det t.ex. svårt att dela upp ett problem till mindre, oberoende, delproblem kan parallell processering vara ineffektivt. Första frågan i detta examensarbete är således om det är möjligt med parallell evaluering av Business Rules.

Efter denna diskussion, tar vi i detta examensarbete fram en prototyp av Amadeus Business Rules-system, gjord i Apache Spark. Prototypen gör det möjligt att dela upp reglerna till små delmängder, och låta ett flertal processorer behandla varsin delmängd samtidigt. I slutsteget kan sedan de regler som hittats av varje processor jämföras, för att hitta den absolut bästa regeln för den givna situationen. I examensarbetet testas vi prestandan av prototypen och jämför mot originalversionen.

Slutsatserna som kan dras är att det är möjligt med regelsökning i en parallell miljö, och att Amadeus kan uppnå prestandavinster för redan existerande datamängder. Testerna visar även att prototypen skalar mycket bra när man ökar antalet regler ännu mer, vilket är intressant då mängden data sannolikt kommer fortsätta att växa.