

## Databas 2 - Lab 7 del 2

### Fråga 1

- a) För det första scenariot skulle jag vilja rekommendera en dokumentbaserad databas. När det kommer till dokumentbaserad databaser är jag framför allt bekant med MongoDB. Baserat på vad jag vet om MongoDB så är den bra när du ska hantera många olika typer av värden, något som behövs i det här fallet. Största anledningen till varför jag rekommenderar en dokumentbaserad databas är på grund av förmågan att använda ett nästlat dokument. Något som kan användas för att lagra "Vanliga Köp".
- b) Framför allt baserat på mängden data som lagras i det andra scenariot så skulle jag rekommendera en Key-value databas. Största anledningen till det är att mycket av den data som lagras är ja/nej frågor (bool). Den typen av information kan ändras ofta och därför kan det vara bra att använda en typ av databas som är skalbar. Key value behöver inte ha placeholder values såsom null vilket gör det lättare och mer effektivt att använda när du kanske inte har all information tillgänglig om en "person" lagrad i databasen. Du behöver till exempel inte lagra ifall användaren har en cykel om det är information ni på företaget inte är medvetna om.

### Fråga 2

- a) 

```
[[{$group: {
  _id: "$majorcategory",
  count: {
    $sum: 1
  }
}}, {$sort: {
  count: -1
}}, {$limit: 3}]
```
- b) Man kan eventuellt ta bort "FALSEALARM" ärendet beroende hur man tolkar frågan. Det påverkar inte kommande stegen däremot. Du grupperar ärendena baserat på "station" sedan räknar du ut hur många varje grupp innehåller med "\$sum". Det sista du gör att sortera med \$sort för att få ut att station 003 har flest ärenden.

```
[[{$match: {
  majorcategory:{$nin: ["FALSEALARM"]}
}}, {$group: {
  _id: "$station",
  count: {
    $sum: 1
  }
}]
```

```
}}, {$sort: {  
  count: -1  
}}]
```

### Fråga 3

- a) Data Warehouse - ETL

Data lake - ELT

- b) ETL står för Extract, Transform, Load. ETL är en process för att hämta data till ett Data Warehouse. Extract är när du hämtar datan du behöver. Transform är när du ändrar på datan beroende på behoven. Load är när du lagrar datan i en databas.

### Fråga 4

- 1) Det första steget som du måste genomföra innan du kan analysera data är att ta reda på vad du vill ta reda på. Det spelar ingen roll hur mycket data du har tillgång till om du inte vet vad du skall använda den till.
- 2) Samla in datan. Nu när du vet vilka frågor du vill ha svar på så är det dags att samla in datan. Ta reda på om någon av den datan som du redan har tillgång till kan användas för att få svar på önskat fråga. Om du inte redan har tillgång till datan så måste du lista ut hur du skall få tag på önskad data och genomföra insamlingen.
- 3) Analysera datan. Med en fråga att besvara och rätt data insamlad är det dags att analysera den. Kolla efter trender i datan genom att manipulera den på olika sätt. under analys steget kan det komma fram att du måste gå tillbaka till dom tidigare stegen för att samla in mer data eller kanske även ändra på frågan du ställde i steg 1.

### Fråga 5

- a) Baserat på min uppfattning av MySQL databaser så är dom bra att använda ifall datan som lagras är relativt stabil. Det vill säga att datan inte oftast ändras eller utökas. Om mängden data som lagras inte växer exponentiellt så kan vara en bra ide att använda MySQL. ACID är också exklusivt till MySQL och fungerar inte på NoSQL databaser. För att sammanfatta så är MySQL bättre om du prioriterar att datan skall vara stabil och pålitlig. MySQL är även billigare i regel att använda än noSQL.
- b) För att representera NoSQL databaser så har jag valt att använda dokument baserade databaser som MongoDB. Många av fördelarna som tas upp stämmer även in på andra typer av NoSQL databaser. NoSQL databaser är i regel modernare och uppdateras mer regelbundet än MySQL. NoSQL borde användas ifall datan som lagras oftast ändras och utökas eftersom NoSQL tillåter utökning i vilken typ av data som kan lagras. Det går i regel snabbare och är mindre prestandakrävande att hämta data från en noSQL databas än en MySQL databas.

