



GÖTEBORGS UNIVERSITET

Utveckling av OLAP med Stjärnschema

Development of OLAP with Star Schema

Johan Lindell | j.lindell81@gmail.com

Kandidatuppsats i Informatik

Rapport nr. 2010:049
ISSN: 1651-4769

Abstract:

Title: Development of OLAP with Star Schema
Author: Johan Lindell
Keywords : OLAP, data warehouse, star schema

Objective: Based on selected scenarios to identify the potential effect that the use of OLAP can lead to. Through a discussion of the benefits of OLAP, after having presented the scenarios that arose after the case study, give the reader a greater understanding of how different industries can benefit from OLAP in their respective fields and show how powerful OLAP can be for organizations.

Research question: *What are the possible effects OLAP can bring in business?*

Method: A qualitative case study method called adopted in this study when a number of scenarios were developed. The data collection was done using a qualitative approach.

Theory: The chapter begins with an explanation of OLAP and its components. Explains OLAP, data warehouse, data mining, data quality, data warehouse data structure, star schema, and snowflake schema.

Results: A number of scenarios were developed based on existing activities and sectors of society. Star schema, data tables, data, examples, and finally SQL queries based on these star schemas are presented in this section.

Conclusion: The effect of OLAP brought varied depending on which branch to be deployed OLAP in. As long as businesses made use of their advantages OLAP can benefit their business.

This thesis is written in swedish.

Sammanfattning :

Titel: Utveckling av OLAP med Stjärnschema
Författare: Johan Lindell
Nyckelord: OLAP, datalager, stjärnschema

Syfte: Att utifrån utvalda scenarios identifiera vilka möjliga effekter som användandet av OLAP kan leda till. Genom en diskussion om nyttan med OLAP, efter att ha presenterat de scenarios som uppkom efter fallstudien, ge läsaren en ökad förståelse över hur olika brancher kan dra nytta av OLAP i deras respektive områden och visa hur kraftfullt OLAP kan vara för organisationer och verksamheter.

Forskningsfråga: *Vilka möjliga effekter kan OLAP tillföra i en verksamhet?*

Metod: En kvalitativ metod kallad fallstudie antogs till denna studie då ett antal scenarios arbetades fram. Datainsamlingen skedde med hjälp av en kvalitativ metod.

Teori: Kapitlet inleds med en förklaring till OLAP och dess komponenter. Här förklaras OLAP, datalager, data mining, datakvalité, datalagrets datastruktur, stjärnschema, samt snöflingeschema.

Resultat: Ett antal scenarios arbetades fram med utgångspunkt från existerande verksamheter och samhällssektorer. Stjärnschema, faktatabeller, dataexempel samt avslutningsvis SQL-frågor baserade på dessa stjärnscheman presenteras i detta avsnitt.

Slutsats: Effekterna som OLAP tillförde varierade beroende på vilken branch som OLAP nyttjades i. Så länge som verksamheterna utnyttjade sina fördelar så medförde OLAP nytta till respektive verksamhet.

Förord

Denna kandidatuppsats (15hp) är skriven vid Institutionen för tillämpad informationsteknologi vid Göteborgs Universitet under vårterminen 2010. Uppsatsen är även det sista momentet innan jag tar min kandidatexamen inom Systemvetenskap, IT, Människa och Organisation.

Det har varit en intressant och rolig resa med denna uppsats, inte bara att jag fick arbeta med Business Intelligence som tillsammans med affärssystem är vad jag tycker är intressant inom IT utan även för alla de situationer som uppkom under arbetet med uppsatsen. Det har varit en resa med uppgångar och nedgångar som jag har lärt mig mycket av.

Jag vill rikta ett stort tack till Bertil Lindberg på Ericsson för att han fick in mig på området OLAP från första början och väckte mitt intresse för detta Business Intelligence. Ett stort tack skall även mina studiekamrater ha för alla de råd och tips som de gav mig, även ett tack till min familj och vänner som har bidragit med goda råd och stöttning under hela uppsatsen.

Slutligen vill jag rikta ett extra stort tack till min handledare Faramarz Agahi för hans vägledning, kritik och tips som han gav mig under hela arbetet.

Innehållsförteckning

1	Introduktion	7
1.1	Bakgrund.....	7
1.2	Syfte	7
1.3	Forskningsfråga.....	7
1.4	Avgränsning	7
1.5	Disposition	8
2	Metod.....	9
2.1	Val av vetenskaplig metod.....	9
2.2	Typ av forskning	9
2.3	Forskningsmetod.....	10
2.4	Datainsamlingsmetod	10
2.4.1	Dokumentundersökning	10
2.4.2	Dagboksskrivande & Diktering.....	10
2.5	Dataanalyseringsmetod.....	11
2.6	Studiens arbetssätt	11
3	OLAP.....	12
3.1	On-Line Analytical Processing.....	12
3.2	Datalager	12
3.3	Datakvalité.....	13
3.4	Data mining.....	14
3.5	Datalagrets datastruktur	14
3.5.1	Stjärnschema.....	15
3.5.2	Snöflingeschema	15
4	Resultat	17
4.1	Scenarierna	17
4.1.1	Polis-scenariot.....	17
4.1.2	Ehandels-scenariot	17
4.1.3	Detaljhandels-scenariot.....	17
4.2	Resultat från fallstudien	18
4.2.1	Polis-scenariots OLAP	18
4.2.2	Enhandel-scenariots OLAP	24
4.2.3	Detaljhandels-scenariots OLAP	30
4.3	Visuell hjälp.....	34
5	Diskussion	36
5.1	Studiekritik.....	38

6	Slutsats.....	39
7	Källförteckning.....	40

1 Introduktion

Detta kapitel redogör bakgrunden till studien samt dess problemområde. Syftet med studien tas upp i detta kapitel samt dess forskningsfråga och avgränsning.

1.1 Bakgrund

Data som sparas i en organisations databas kan innehålla mycket information på många olika nivåer som organisationen går miste om ifall denna data inte bearbetas till information. Vanliga relationsdatabaser uppfyller inte de krav som beslutstagande användare eller analytiker har då det är allt för komplext och tidskrävande att få ut den information man söker, samt den data som dessa relationsdatabaser innehåller inte är lämpliga att göra analyser på då de oftast inte innehåller den historiska data som analytiker och beslutstagare behöver (Mishra et al, 2007). Det är även svårt för analytiker som arbetar med mycket information att bearbeta detta då dessa blir disorienterade på grund av den kognitiva överbelastning som detta leder till (Cuzzocrea & Mansmann, 2009).

Under mitten av 1990-talet uppkom det en ny gren inom informationsteknologi kallad Business Intelligence (BI) som innehöll verktyg och system såsom On-Line Analytical Processing (OLAP), datalager och data mining för att möta det krav på analytiska verktyg som slutanvändaren ställde, bland annat att det inte skulle krävas specialister inom informationsteknologi för att göra de sökningar systemen krävde (Turban et al, 2007 – Moody & Kortnik 2000). Med hjälp av Business Intelligence, datalager och OLAP kunde beslutstagande analysera den data som organisationens databaser innehöll för att omvandla datan till information som stödjer vid beslutstagande, då speciellt med ett verktyg som OLAP som stödjer den intelligenta delen av Business Intelligence genom att identifiera relationer mellan aktiviteter och de faktorer som påverkar aktiviteterna (Turban et al, 2007).

1.2 Syfte

Syftet med denna studie är att utifrån utvalda scenarios identifiera vilka möjliga effekter som användandet av OLAP tillföra en verksamhet eller en organisation. Genom en diskussion om nyttan med OLAP, efter att ha presenterat de scenarios som uppkom efter fallstudien, är syftet att ge läsaren en ökad förståelse över hur olika brancher kan dra nytta av OLAP i deras respektive områden och visa hur kraftfullt OLAP kan vara för olika organisationer eller verksamheter.

1.3 Forskningsfråga

Vilka möjliga effekter kan OLAP tillföra i en verksamhet?

Med denna öppna frågeställning var avsikten med studien att identifiera vilka möjliga effekter OLAP kan medföra i en verksamhet, men även vara öppen för vidare diskussion och frågor över användningsområdet OLAP som eventuellt kan uppkomma.

1.4 Avgränsning

Avgränsningen för denna studie innebär att ingen empirisk undersökning har skett, istället har en fallstudie genomförts. Istället för att gå ut i näringslivet och intervjua folk eller studera riktiga fall har denna studie koncentrerat sig på att göra en fallstudie med hjälp av scenarios då studien inte ville vara bunden till en riktig organisation. Risker med att binda sig till en organisation skulle kunna

innebära dataproblem där antingen inte rätt data skulle finnas att tillgå, att organisationen skulle få bestämma den data i de scheman som skapats, alternativt skulle det kunna ha hänt att studien i slutändan ändå skulle få använda sig av fiktiva tabeller då vissa organisationer kanske inte skulle ge medgivande till att deras tabeller skulle få presenteras i en studie. Detta utifrån författarens egna erfarenheter. Verksamheten som beskrivs i forskningsfrågan är begränsad till de brancher och samhällssektor som beskrivs i de scenarios som arbetades fram.

1.5 Disposition

Rapporten är indelad i fem avsnitt: *Metod*, *OLAP*, *Resultat*, *Diskussion* samt *Slutsats*. *Metodavsnittet* innehåller den valda forskningsmetoden tillsammans med datainsamlingsmetoden och dataanalyseringsmetoden, förklaring till varför vissa metoder användes finns förklarat här. *OLAP* är studiens teoretiska ramverk, här ges information om det problemområde som studien vill belysa samt ge information till läsaren om den övergripande bilden omkring OLAP. *Resultat* är den del av studien där läsaren får ta del av resultatet som studien kom fram till under fallstudien tillsammans med förklaringar kring de tabeller som användes. *Diskussionsavsnittet* handlar om OLAP och nyttan med detta utifrån de resultat som kom fram från fallstudien. Slutligen kommer *Slutsats* där en sammanfattning ges till läsaren utifrån den samlade studien.

2 Metod

Kapitlet innehåller här vald metod tillsammans med beskrivning över de metoder valdes att användas under studien. Kapitlet avslutas med att studiens arbetsätt beskrivs.

2.1 Val av vetenskaplig metod

Valet av vetenskaplig metod har skett med hjälp av internetsidan "Association for Information Systems" (AIS) (<http://home.aisnet.org>) som är en global organisation för experter och specialister inom informationssystem. Sidans mål är bland annat att skapa en standard för etik, utbildning, utförande samt för att stimulera lärandet av informationsteknologin genom att försöka samla så mycket information och standarder som möjligt på ett och samma ställe (AIS). Studien har tagit till sig de förslag och råd som ges av Michael Myers som är författare till den kvalitativa forskningsdelen för AIS. Innan en studie görs bestämmer man vilken typ av forskningsmetod som skall användas, AIS ger en bra övergripande modell över hur man går tillväga med detta som är gjord av författarna Straub, Gefen och Boudreau från AIS.

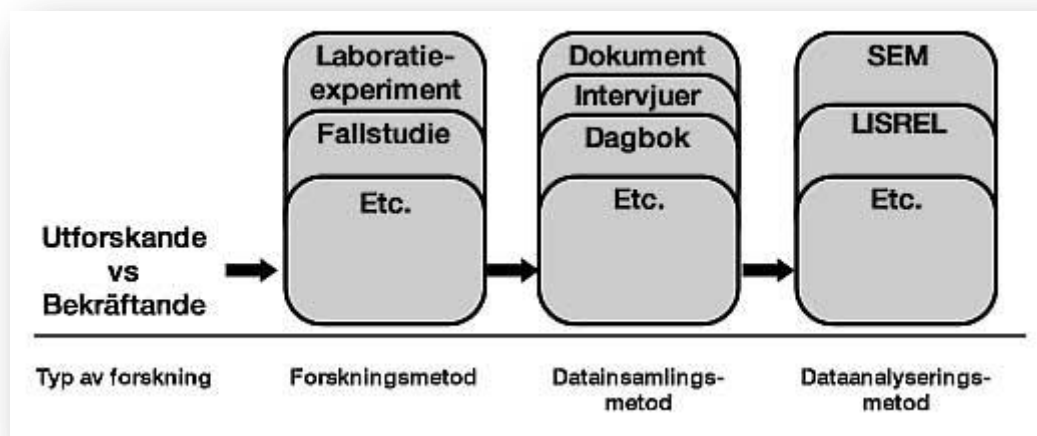


Fig. 1 Forskningsmodell från AIS

2.2 Typ av forskning

En studie börjar med att man bestämmer sig ifall man vill använda sig av utforskande eller bekräftande forskning eller induktivt och deduktion forskning. Att arbeta induktivt innebär att forskaren arbetar genom "upptäckandets väg" och på så vis arbetar genom att låta resultatet av studien vara grunden för en formulerad teori som skapas i slutet (Patel & Davidson, 2003). Deduktivt innebär att forskaren istället följer "besvisandets väg" och utifrån en redan existerande teori skapar sig en hypotes som sedan prövas empiriskt (Patel & Davidson, 2003). Denna studie har ingen förankring mot en tidigare teori samt det finns ingen metod som ligger till grund för att de arbetsätt eller modellering som har använts, därav bestämde sig författaren för att den induktiva metoden var den mest intressanta för studien.

2.3 Forskningsmetod

Fallstudie valdes som forskningsmetod för denna studie. Anledningen till att en fallstudie valdes som metod för denna studie var för att försöka arbeta fram olika sätt för att utveckla ett stjärnschema och för att författaren själv skulle uppleva de svårigheter och viktiga aspekter som kan uppkomma vid en sådan process, detta för att sedan fokusera sig på nyttan med OLAP för olika verksamheter. Fallstudien genomfördes med hjälp av teoretiska scenarios där frågeställningen som scenarierna bygger på kan vara ställda utav verksamhetens ledning eller beslutstagande individer inom verksamheten. Frågeställningarna från dessa brancher är fiktiva och har ingen äkta grund från näringslivet.

En fallstudie går ut på att som forskare undersöka en enda undersökningsenhet för att på så sätt få djupgående kunskaper om ämnet, genom att endast koncentrera sig på en enhet istället för större delar skaffar sig forskaren mer detaljerade kunskaper om ämnet (Patel & Davidson, 2003 - Cornford & Smithson, 2006). Styrkan med att använda sig av fallstudie som metod är att man kan få mycket riklig data (Cornford & Smithson, 2006) men när man vill diskutera generaliserbarheten med de resultat beror oftast på de val av fall man har gjort (Patel & Davidson, 2003).

2.4 Datainsamlingsmetod

Tredje steget enligt AIS-modellen är datainsamlingsmetoden. Enligt AIS kan flera datainsamlingsmetoder användas till en studie då det är upp till forskaren och det resultat som denne vill analysera (AIS). Valet av datainsamlingsmetod måste dock vara i proportion med tiden och de medel som studien har för att bäst besvara den frågeställning som forskaren har valt att undersöka (Patel & Davidson, 2003).

För denna studie valdes dokumentundersökning och dagboksskrivande som de primära datainsamlingsmetoderna med diktering som sekundär, detta för att studien var begränsad till de scenarios som uppkom från fallstudien, dessa hade ingen riktig kvantitativ data såsom grafer eller siffror som resultat utan det var datan som kom utifrån den generaliserade arbetsmetodiken som var intressant att utforska. Denna datan ansågs komma bäst fram genom anteckningar och diktering från författaren själv.

2.4.1 Dokumentundersökning

Studien började med dokumentundersökning för att ge författaren en större insikt över problemområdet OLAP och stjärnschema samt för att förstå metodiken och funktionaliteten på en djupare nivå. Dokumentundersökningen skedde genom att använda sig av källor såsom biblioteket, Google scholar, vetenskapliga artiklar, fallstudier samt uppsatser. Studien har använt sig av artiklar och dokument som behandlar OLAP och stjärnstrukturen från fler än ett synsätt, artiklarna som har behandlat ämnet har därav inte endast varit av den positiva aspekten utan även negativa för att försöka få en sådan objektiv syn som möjligt (Patel & Davidson, 2003). Dokumentundersökningen har dock inte haft en avgörande roll i denna studie utan har mest varit en källa för information och objektiv utvärdering men har haft en sådan betydande roll att den har varit en del av metodiken för datainsamlingen.

2.4.2 Dagboksskrivande & Diktering

Tillsammans med dokumentundersökningen har studien använt sig av självrapportering såsom dagboksskrivande och diktering för att dokumentera egna tankar och idéer som sedan användes i

analyseringen och diskussionen. Från början var tanken att endast använda sig av dagboksskrivning för att dokumentera hur arbetet gick till och när generaliseringar kunde göras eller vissa aspekter utifrån arbetet kunde identifieras. Det tog dock inte lång tid innan det var uppenbart att ibland fanns inte tiden till för att skriva ner alla tankar och idéer som kom upp under utvecklandet av scenariorna och därav använde sig författaren sig av diktering som en sekundär källa. Under fallstudien utfördes dikteringen med hjälp av företaget Apples mobiltelefon iPhone med hjälp av applikationen iTalk (Från företaget Griffin Technology) för att efter fallstudien spelas upp igen och de viktigaste aspekterna skrevs ner tillsammans med dagboken. Det är därför viktigt att poängtera att allting som sades under diktationen har inte transkiberats till dagboksskrivandet, dagboksskrivandet har inte heller tagits med i denna uppsats då det endast är resultatet som presenteras i denna studie och skulle inte tillföra läsaren något utöver det som finns redovisat i studien.

2.5 Dataanalyseringsmetod

Till sist enligt AIS-modellen kommer dataanalyseringsmetoden. Då fallstudien handlar om att skapa ett antal stjärnscheman kan en kvantitativ analyseringsmetod användas men eftersom studien även är fokuserad på effekten av OLAP i verksamheter och gå in på djupet i ämnet var det intressant med analyseringsmetoden som används inom kvalitativ analyseringsmetod. Enligt Cornford & Smithson (2006) kan man kombinera kvantitativ och kvalitativ analys genom att låta dessa komplettera varandra. Därför har studien använt sig av en kvantitativ analyseringsmetod för att finna en generell arbetsmetodik och sedan använt sig av kvalitativ analyseringsmetod genom att analysera dagboksskrivandet och dikteringen (Patel & Davidson, 2003). Tillsammans har dessa två analyseringsmetoder varit grunden för resultatet och diskussionen om effekten av OLAP i verksamheter samt resultatet redovisas enligt kvalitativ metodik där citat har ersatts med bilder från scenariorna (Patel & Davidson, 2003).

2.6 Studiens arbetssätt

Datainsamlingen för denna studie skedde i två steg med ett tredje steg parallellt så som det beskrivs i figuren till vänster. Dokumentundersökningen användes löpande under hela studien för att hela tiden utforska vad det finns för olika aspekter beroende på vilket problem som studien kom att stöta på. Litteratur såsom böcker från biblioteket, Google scholar och vetenskapliga artiklar från tidskrifter har använts som informationskälla.

Datainsamlingen kom från fallstudien samt den litterella undersökningen med fokus på

stjärnschemat och uppbyggnaden av denna.

Genom att lära sig genom litteraturen och sedan applicera den till fallstudien var förhoppningen att hitta viktiga aspekter för studien.

Dataanalysering skedde i samförstånd med datainsamlingen då allt arbete som gjordes under fallstudien analyserades och jämföras med tidigare resultat.

Dagboksskrivandet och dikteringen skedde parallellt med alla aktiviteter.

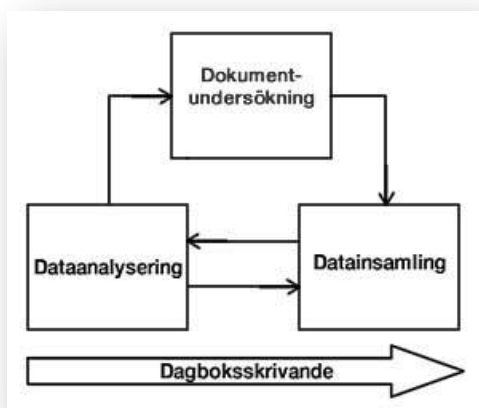


Fig. 2 Modell över studiens arbetssätt

3 OLAP

Detta kapitlet handlar om att ge läsaren en överblick över OLAP och hur denna teknik fungerar. Kapitlet börjar med att beskriva OLAP och de grundläggande komponenterna för att utnyttja denna teknik.

3.1 On-Line Analytical Processing

On-Line Analytical Processing (OLAP) är ett verktyg som kan analysera stora mängder historisk data för att besvara analytiska frågor som en användare har, exempelvis hur försäljning har utvecklats på en specifik geografisk plats, eller visa relationer i datan som annars inte är uppenbara (Turban et al, 2007). Vanligtvis är data lagrat i en vanlig relationsdatabas, men det hade varit mycket svårt för en vanlig användare att hantera de komplexa sökningar som måste göras för att få ut den relevanta information på liknande sätt som med OLAP då vanliga relationsdatabaser är normaliserade. Normalisering innebär att hundratals tabeller är kopplade till varandra för att minska redundans i databasen, detta gör att det blir för komplext för slutanvändaren utan att ha tillgång till experter inom området att få ut informationen från databasen (Moody & Kortink 2000). OLAP har inte detta problemet då uppbyggnaden av ett OLAP-system innebär att tabeller slås ihop och istället endast skapa en *dimension* som innehåller tabeller av gemensam natur, genom att slå ihop tabellerna till en stor dimension istället för många små tabeller minskar de nödvändiga sammanslagningar av tabeller i relationsdatabasen som slutanvändaren måste göra för att få ut den information som är intressant.

Hjärtat i ett OLAP-system är den multidimensionella kub som används för att utföra de analyser som användaren vill göra. Genom att bygga upp kuben av att koppla dimensioner till mätbar data, eller till *faktatabeller*, skapar kuben ett flertal dimensioner som en analytiker kan välja att utföra sina undersökningar genom (Turban et al, 2007 – Moody & Kortink 2000). Om en kub exempelvis innehåller dimensionerna tid, produkt och plats kan analyser göras på olika nivåer med hjälp av alla tre dimensioner eller några specifikt det är upp till användaren hur denne vill använda kuben och leta efter relevant information.

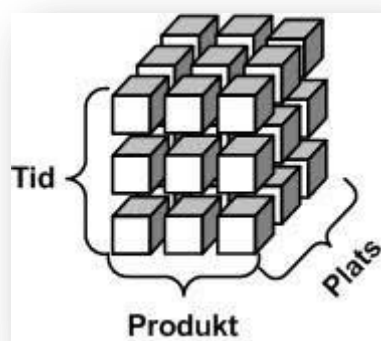


Fig. 3 OLAP-kub

3.2 Datalager

Ett datalager (på engelska Data Warehouse) som även kan kallas för informationslager har som uppgift att lagra den data som kommer från organisationens servrar, då oftast med historisk data för att kunna utföra de analyser som användaren vill göra (Turban et al, 2007). En organisation har oftast ett stort antal databaser innehållande en enorm mängd information som man vill ha med i sin analys men mycket ofta är det även så att man får data från andra företag eller andra samarbetspartners, dessa kan i sin tur leverera datan från helt olika system som den mottagande organisationens analysprogram inte kan bearbeta (Mishra et al, 2007). Datalager är konfigurerade för att möta detta problem med hjälp av ETL som fångar upp datan och omvandlar den till ett generellt format. ETL står för (Turban et al, 2007 – Moody & Kortnik 2000):

Extrahera – Med dess hjälp extraherar datalagret datan från operativa databaser, enskilda filer eller externa databaser. Denna process möjliggör att informationen som vill användas i datalagret kan komma från en stor mängd informationskällor.

Transforma – Transformerar datan till ett format som gör att man kan skicka in datan till en redan existerande databas eller till datalagret och som gör att datan i datalagret bli konsistent. Konsistens är en viktig aspekt för att analyseringar skall kunna utföras och för att informationen som förmedlas till användaren är likadan på alla ställen för att öka förståelsen för informationen.

Ladda – Laddar upp datan till datalagret.

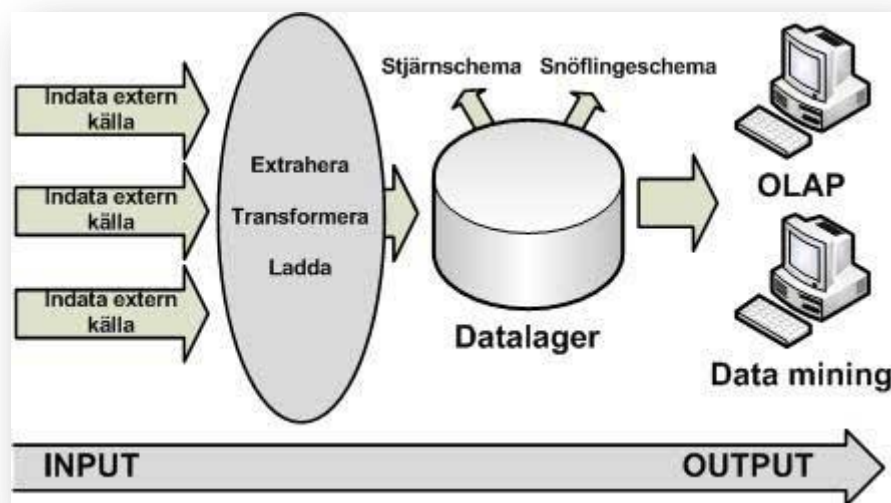


Fig. 4 Dataprocessen

Den data som används i ett datalager är oftast fokuserat till att analysera historisk data och inte transaktionsdata som skickas mellan organisationer såsom ordrar eller beställningar (Rizzi et al, 2006) därför är det viktigt att den data som finns i systemet och som skall analyseras är korrekt. Det är viktigt att bibehålla en hög standard på datan som kommer in i datalagret så denna inte är felaktig, datan som finns i datalagret skall inte heller kunna ändras eller påverkas efter att det har lagts in av användaren (Turban et al, 2007). Det är även viktigt att säkerhetsställa att endast en version av informationen används och inte flera olika versioner som kan ge fel information till slutanvändaren (Turban et al, 2007 - Rizzi et al, 2006). Slutligen är det viktigt att all data i ett datalager är konsistent och har en koppling till tid, alla dessa kriterier uppfyller ett datalager (Turban et al, 2007 - Rizzi et al, 2006).

3.3 Datakvalité

Eftersom OLAP behandlar historisk data som ofta kan komma från många olika källor är det viktigt att datakvalitén är hög. Med datakvalité menas att datan är pålitlig och uppdaterad, kort sagt att informationen som en beslutstagande får ut skall vara verklighetstroget och beslutstagarna skall kunna lita på informationen för att stödja deras beslutstagande (Turban et al, 2007 – Moody & Kortnik 2000 - Rizzi et al, 2006). När man pratar om datakvalité i samband med OLAP är det ofta konsistent som kommer på tal, informationen som läggs in i ett datalager måste vara av likadan sort

som överallt annars i systemet annars kommer inte denna information kunna utnyttjas eller i värsta fall ge användaren fel information som i sin tur leder till fel beslut eller värre följder (Turban et al, 2007). Ett exempel på när fel information når beslutstagande som kan ge ödestigande är rymdfärjan Challenger som exploderade kort efter start i USA 1986, explosionen orsakade sju dödsoffer bland besättningen. Orsaken till explosionen var ett teknisk fel men på grund av dålig datakvalité som gjorde att beslutstagarna fick missvisande information från systemet togs beslutet om uppskjutning vilket kostade besättningsmännen deras liv (DSSResources.com). En metod som kan användas för att hjälpa med att öka datakvalitén i ett datalager är data mining.

3.4 Data mining

Data mining är en process för att finna mönster från en stor samling data med hjälp av algoritmer vilket gör processen automatisk. För att göra detta måste datan som skall undersökas redan vara i en konsistent form, exempelvis så blir det svårt för en mining algoritm att skilja mellan 2010-01-01 och 10e maj 2010 (Turban et al, 2007) samtidigt som dessa avvikelser kan hittas med hjälp av data mining, detta gör att data mining och datakvalité kompletterar varandra.

3.5 Datalagrets datastruktur

Det sätt en vanlig databas skiljer sig mot ett datalager är hur man sätter upp de tabeller man är intresserad av att analysera (Turban et al, 2007). I en vanlig databas är det viktigt att man ser till att datan kan ges åtkomst till många användare och samtidigt snabbt (Rizzi et al, 2006). Datan behöver samlas på endast en plats i databasen och ändras datan så ändrar man endast den tabellen datan ligger i. Normalisering är ett sätt för att se till att man undviker redundans och ser till att dessa kriterier uppfylls samtidigt som vissa kvalitativa kriterier uppfylls (Riccardi, 2001 – Padron-McCarthy & Risch, 2005). I ett datalager däremot är inte normalisering optimalt eftersom det innebär mycket komplicerade sammanslagningar av tabeller som behövs göras för att få fram den data man är ute efter. Sökningarna går mycket långsammare med normalisering då ett datalager innehåller mycket olik data eftersom den innehåller historisk data, det är svårt för en användare att hitta de tabeller som man behöver och ofta kan en vanlig användare inte utföra dessa sökningar från databasen som detta kväver (Shin & Sanders, 2006 – Martyn, 2004). Datalagret löser detta problem genom att den innehåller två olika sorters tabeller: *faktatabell* och *dimensionstabell*.

En *faktatabell* innebär såsom namnet antyder på att tabellen innehåller fakta som är associerad till en specifik process inom organisationens arbetsflöde som man vill analysera (Pardon-McCarthy & Risch, 2005). Faktatabellen innehåller oftast mätbar data såsom transaktioner eller händelser (Turban et al, 2007). En *dimensionstabell* beskriver faktan med exempelvis plats, tid och produkt (Pardon-McCarthy & Risch, 2005). Detta gör att ett datalager innehåller väldigt mycket information men i väldigt få tabeller vilket gör att det arbete som krävs för att få fram den information man vill ha om man slå ihop tabellerna blir mycket enklare och kräver inte lika mycket kunskaper från användaren, ihopslagning av tabeller kallas för denormalisering (Shin & Sanders, 2006). De vanligaste sätten att strukturera datan i ett datalager är genom två olika scheman: *stjärnschema* och *snöflingescema*.

3.5.1 Stjärnschema

Ett stjärnschema är en till många relation där man placerar en faktatabell i mitten som sedan är kopplade till flera dimensionstabeller (Pardon-McCarthy & Risch, 2005) där dess namn kommer från att relationerna ser ut som en stjärna som visas i figuren. Det som gör stjärnschemat till en favorit



Fig. 5 Stjärnschema

bland många utvecklare är att den endast har en faktatabell som i sin tur är kopplad till olika dimensionstabeller och därav gör det enkelt att förstå kopplingarna (Pardon-McCarthy & Risch, 2005), men det är inte det enda skälet till att stjärnschemat är det mest använda när det gäller att skapa relationerna för en OLAP-kub. Ett av de största skälen till att stjärnschema används frekvent inom skapandet av OLAP-kuber är att tabellerna är enkla att underhålla ifall tabeller ändras sig eller om man behöver lägga till tabeller i framtiden (Pardon-McCarthy & Risch, 2005) då genom att endast lägga till en tabell eller enkelt att ändra i tabellen utan att behöva oroa sig för

att hela schemat blir förstört och oanvändbart. Stjärnschema är det schema som fallstudien kommer att använda sig av.

3.5.2 Snöflingschema

Om man som utvecklare fortfarande vill normalisera tabellerna kan man självklart göra detta och då finns det snöflingschemat som man oftast gör det med. Med snöflingschemat definierar man hierarkin genom att använda sig av multidimensionella tabeller, fördelen med detta är att användaren direkt får en hierarki i sitt datalager men metoden rekommenderas dock inte. Anledningen är att det blir svårare för användaren att förstå tabellerna samt det är mycket mer krävande funktioner som behövs att göras för att lyckas göra de analytiska undersökningar som man vill göra (Pardon-McCarthy & Risch, 2005). Man skapar en snöflingetabell genom att placera en faktatabell i mitten som vid stjärn-

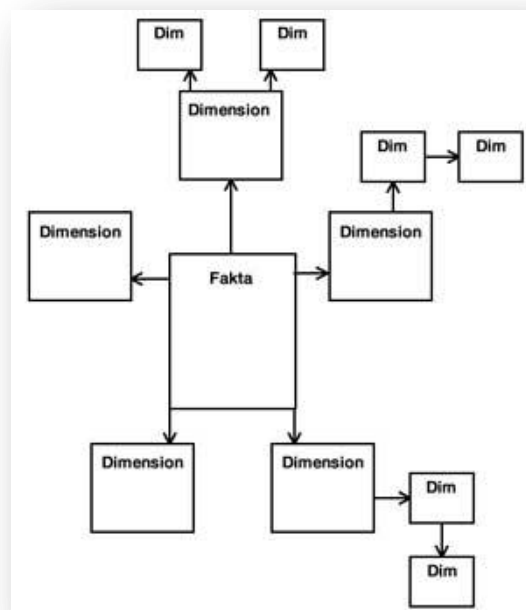


Fig. 6 Snöflingschema

schemat och koppla dimensionstabeller till denne, men till skillnad från stjärnschemat är dessa dimensionstabeller normaliserade på vanligt vis.

Ett argument till varför man vill använda sig av snöflingeschemat, förutom att ha hierarkin i sitt datalager vilket endast är användbart ifall man inte vill använda sig av ett OLAP -verktyg utan endast datan i sitt datalager, är oftast att man vill spara plats i sitt datalager, ett argument som är fullständigt rimligt då datalager behandlar historisk data som i sin tur kan leda till extremt många poster och tabeller inom vissa brancher eller företag. Även med väldigt många poster kan snöflingeschemat visa sig vara irrelevant då dimensionstabellerna oftast är mycket mindre än faktatabellerna (faktatabellerna normaliserar man inte) vilket gör att platsen man sparar i sin databas oftast inte är värt att offra för att undvika att använda sig av ett stjärnschema (Shin & Sanders, 2006 – Moody & Kortink, 2000).

4 Resultat

Presentationen av fallstudien som utfördes presenteras här. Genom att presentera varje scenario enskilt tillsammans med stjärnschema, datatabellsexempel samt SQL-frågor är förhoppningen att läsaren skall få en bra översikt över fallstudien och dess resultat. Både bild samt förklaring presenteras för att ge läsaren så mycket information som möjligt.

4.1 Scenarierna

Fallstudien utgick från tre stycken scenarios som presenteras här som studien har valt att kalla polis-scenariot, ehandels-scenariot samt detaljhandels-scenariot. Dessa scenarios har valts för att visa flexibiliteten med OLAP vid behandling av stora mängder data samt för att visa hur kraftfullt ett sådant verktyg kan vara för beslutstagande personer i en organisation.

För att visa vilken nytta OLAP kan tillföra verksamheter och vilken effekt detta kan bidra med kommer studien visa exempel på hur OLAP kan användas av respektive verksamhet/samhällssektor, dessa exemplen kommer att visas efter att tabellerna från stjärnschemat har förklarats.

Databasspråket SQL som har använts för att demonstrera hur information kan extraheras från OLAP har valts då det antagligen är ett av de mest kända språken för denna typ av process och kan tilltala majoriteten av läsarna. Författaren använde en MySQL databas för att kontrollera resultaten som uppkom från frågeställningarna.

4.1.1 Polis-scenariot

Ett scenario baserande på den information som en samhällssektor såsom polisen kan använda sig av. Scenariot är uppbyggt genom att endast representera inbrott vilket är endast en liten del av polisens arbetsuppgifter. Till skillnad från de andra två scenariorna är inte polisen en verksamhet vars syfte är att tjäna pengar utan finns för samhällets bästa. Det var intressant för studien att försöka visa att även om en organisations mål inte är att optimera en vinst så kan OLAP ändå användas effektivt.

4.1.2 Ehandels-scenariot

Ehandel på Internet var en aspekt som valdes att undersöka närmare då mycket information kan lagras i ehandels databaser tack vare att allt som händer på en webbplats kan lagras. En internet-baserad detaljhandel kan se vilka produkter som en besökare tittar på, vilka genrer som intresserar denne mest, vilken prisgrupp som denne köper mest av, vilka böcker som folk tittar på först och sedan vilken köpte de till slut med mera. Det enda som stoppar en ehandelsbutik från att spara allt en besökare gör är fantasin hos utvecklarna och plats i databasen. Detta förhållandesätt gör OLAP till ett utmärkt verktyg för att försöka finna vägar för direktreklam anpassat till besökaren eller visa besökaren böcker denne skulle vara intresserad av nästa gång denne loggar in på sidan men även enklare förfrågningar såsom vad årsförsäljningen har varit. Det scenario som har arbetats fram representerar en ehandel som säljer böcker via Internet.

4.1.3 Detaljhandels-scenariot

Sista scenariot representerar branchen detaljhandel, i detta fall en detaljhandel som säljer kläder i fysiskt belagda butiker och inte på Internet. En fysiskt belagd detaljhandel har inte lika omfattande information om sina kunder som en ehandelsbutik har, en fysiskt belagd detaljhandel kan inte på samma sätt spara allt som en kund gör eller ens identifiera kunden då denne kan betala kontant för en vara och på så sätt vara anonym. De kan dock vara intresserade av att få svar på frågor som inte

är lika individuellt inriktad som en ehandel kan vara utan mer koncentrera sig på en bredade front, alternativt ta reda på hur deras varor och butiker står i förhållande till deras konkurrenter eller få reda på mer information kring hur deras produkter.

4.2 Resultat från fallstudien

4.2.1 Polis-scenariots OLAP

Polis-scenariot består av sex stycken dimensioner samt en faktatabell. Stjärnschemat är en beskrivning på information som polisen kan använda för att undersöka historisk data om inbrott eller stölder som har skett tidigare.

Dimensioner:

- **D_TIME:** Tidstabell innehållande all information om tid för att en analytiker skall kunna använda tidsaspekten vid sina analyser på ett effektivt sätt. Analytikern kan välja mellan att se informationen via år, månader, veckor eller dagar. Månader kan visas både med nummer samt namn samt dagar kan visas med vilket nummer i året, nummer i månaden, nummer i veckan eller med dagens namn. Tidstabellen innehåller inget tidsslag i form av timmar eller sekunder då detta sparas istället i faktatabellen.
- **D_EVENTS:** Dimensionen innehåller de händelser som är sammankopplade med det som stjärnschemat bygger på, i detta scenario handlar det om inbrott eller stöld. Events innehåller kategori och underkategori vilket gör att analytikern kan gå igenom de kategorier som stjärnschemat behandlar antingen på en abstrakt nivå eller på en mer detaljerad nivå. Analytikern kan även behandla informationen genom att använda sig av tabellen som har klassificering av hur allvarligt larmet var.
- **D_VEHICLE:** En dimension innehållande det fordon med medhavande personal som undersökte larmet. Dimensionen innehåller vilken typ av fordon, dess identifieringsnummer samt en kort beskrivning över fordonet. Tillsammans med fordon innehåller även dimensionen vilken personal som behandlade fallet med vilken typ av anställd det handlade om och dennes namn.
- **D_LOCATION:** Innehållande information om platsen för händelsen. Genom att först visa vilken region som brottet skedde i kan analytikern sedan gå ner i hierarkin för att undersöka vilken stad det handlade om, vilken stadsdel, adressen och för att slutligen se detaljerad information om platsen för en mer generell nivå, exempelvis om det handlade om ett hus eller lägenhet.
- **D_CRIMINAL:** De brottslingar som har blivit arresterade samt dömda för ett brott sparas i en tabell utan personuppgifter för att endast spara information av generell typ, detta för att inte inkräkta på dessa människors integritet. Dimensionen innehåller ålder, utbildning, årsinkomst, nationalitet, plats för boende samt om dessa personer har ett tidigare straffregister. Denna information är endast till för att jämföra tillsammans med de andra dimensionerna som stjärnschemat innehåller för att försöka undersöka om relationer kan dras mellan olika sorters information.
- **D_VICTIM:** Är motsatsen till **D_CRIMINAL** där dimensionen sparar information på liknande sätt men även tillsammans med vilken typ av bostad som offret hade. Anledning till att denna tabell lades till var för att lyckas få ut information om typ av bostad skulle kunna leda

till ytterligare en informationskälla när det handlar om att hitta mönster eller relationer som annars inte är uppenbarliga

Faktatabell:

F_CRIME: Innehåller databasnycklar till samtliga sex dimensioner för att koppla ihop dessa vid ett analysbehov. Faktatabellen innehåller förutom nycklarna detaljer kring händelsen, detta för att identifiera tillvägagångssättet som brottslingen använde sig av. Vad som stals finns även sparad samt om någon arrestering skedde på plats eller strax efteråt. Tabellen sparar även vilken tid som larmet kom in till polisen samt den tid som uttryckningsfordonet rapporterade in att dessa var på plats, detta för att se hur lång tid som det tog för uttryckningsfordonet att komma till larmplatsen.

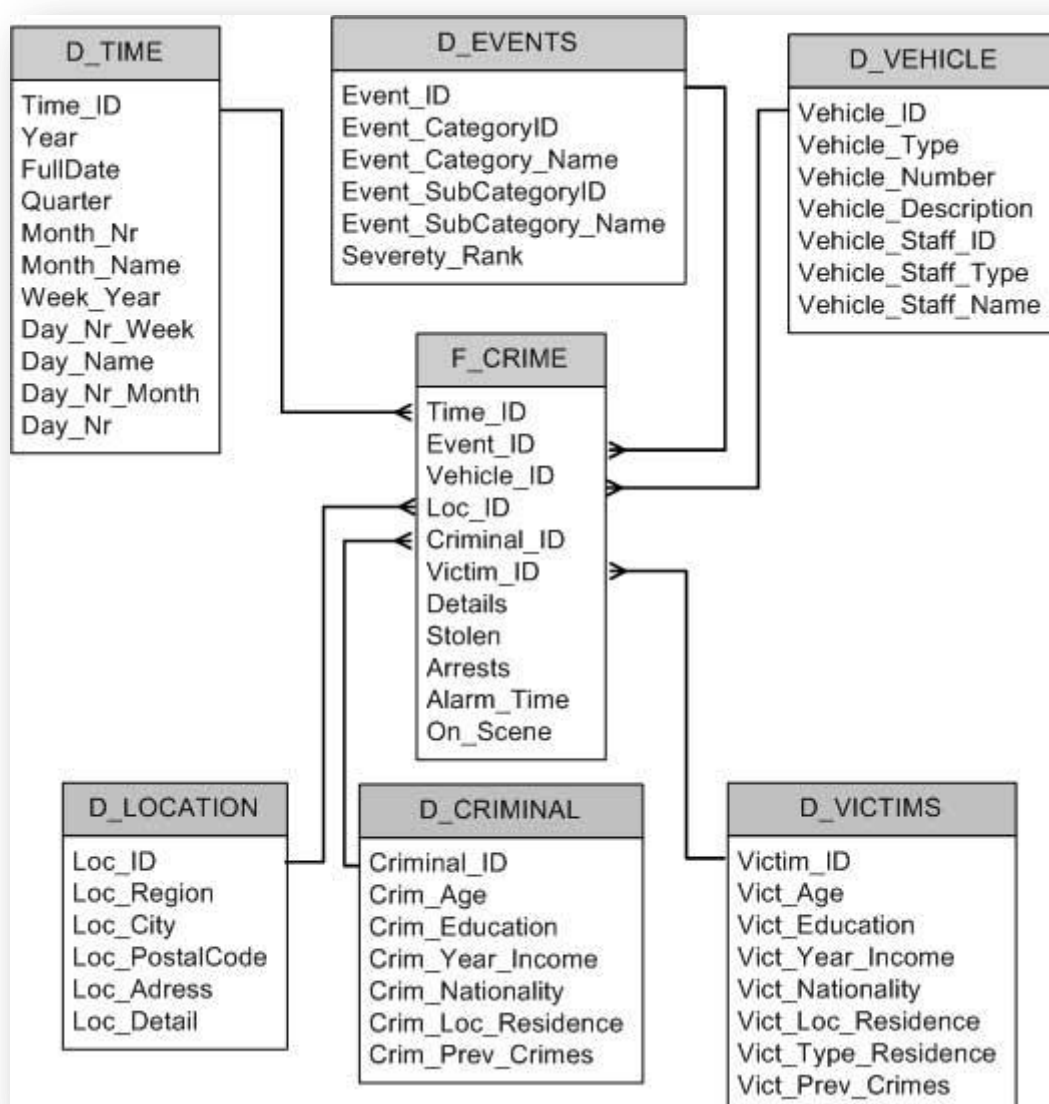


Fig. 7 Polis-scenariots stjärnschema

Dataexempel: Härnäst följer ett exempel på vad dessa tabeller kan innehålla för data. All data är fiktiv.

Time_ID	Year	FullDate	Quarter	Month_Nr	Month_Name	Week_Year	Day_Nr_Week	Day_Name	Day_Nr_Month	Day_Nr
363	2009	20091229	Q4	12	December	52	4	Thursday	29	363
364	2009	20091230	Q4	12	December	52	5	Friday	30	364
365	2009	20091231	Q4	12	December	52	6	Saturday	31	365
366	2010	20100101	Q1	1	January	1	7	Sunday	1	1
367	2010	20100102	Q1	1	January	1	1	Monday	2	2

D_TIME

Event_ID	Event_CategoryID	Event_Category_Name	Event_SubCategoryID	Event_SubCategory_Name	Severity_Rank
E4	EC1	Burglary	ESC11	Empty House	3
E5	EC1	Burglary	ESC12	Empty Apartment	3
E6	EC1	Burglary	ESC13	Empty Other	3
E7	EC2	Burglary Encounter	ESC21	Residence in House	1
E8	EC2	Burglary Encounter	ESC22	Residence in Apartment	1

D_EVENT

Vehicle_ID	Vehicle_Type	Vehicle_Number	Vehicle_Description	Vehicle_Staff_ID	Vehicle_Staff_Type	Vehicle_Staff_Name
V10	Police Car	72	Car on patrol	V55	Police Officer	Myrberg Christoffer
V11	Police Car	72	Car on patrol	V56	Police Officer	Lindgren Emelie
V12	Police Car	72	Car on patrol	V57	Police Officer	Arvidsson Magnus
V13	Police Car	73	Car on patrol	V58	Police Officer	Boyd Carl
V14	Police Car	73	Car on patrol	V59	Police Officer	Eriksson Emelie

D_VEHICLE

Loc_ID	Loc_Region	Loc_City	Loc_PostalCode	Loc_Address	Loc_Detail
L12	Vastra Gotaland	Goteborg	Backa	Adress1	Parking lot
L13	Vastra Gotaland	Goteborg	Backa	Adress2	Appartment
L14	Vastra Gotaland	Goteborg	Backa	Adress3	House
L15	Vastra Gotaland	Goteborg	Brunnsbo	Adress4	Mail
L16	Vastra Gotaland	Goteborg	Brunnsbo	Adress5	Parking lot

D_LOC

Criminal_ID	Crim_Age	Crim_Education	Crim_Year_Income	Crim_Nationality	Crim_Loc_Residence	Crim_Prev_Crimes
C8	18	None	>50 000	Swedish	Orgryste	Yes
C9	44	High	>50 000	Swedish	Orgryste	Yes
C10	38	High	<200 000	Swedish	Centrum	No
C11	18	None	None	Swedish	Centrum	No
C12	22	None	<100 000	Swedish	Centrum	Yes

D_CRIMINAL

Victim_ID	Vict_Type	Vict_Education	Vict_Year_Income	Vict_Nationality	Vict_Loc_Residence	Vict_Type_Residence	Vict_Prev_Crimes
V22	Single	High	<200 000	Swedish	Centrum	House	No
V23	Single	High	>50 000	Swedish	Centrum	Appartment	Yes
V24	Family	High School	<150 000	Swedish	Tuve	House	No
V25	Single	High School	<250 000	Swedish	Tuve	House	No
V26	Family	None	<150 000	Swedish	Lundby	Appartment	No

D_VICTIM

Time_ID	Event_ID	Vehicle_ID	Loc_ID	Criminal_ID	Victim_ID	Details	Stolen	Arrests	Alarm_Time	On_Scene
363	E7	V10	L13	C8	V23	Broken Door	None	Yes	00:21	00:31
363	E4	V13	L13	N/A	V26	Broken Window	Radio	No	04:12	06:17
363	E4	V13	L16	N/A	V26	Broken Window	Wallet	No	04:12	06:17
363	E6	V13	L16	N/A	V26	Broken Window	Phone	No	04:12	06:17
363	E6	V13	L16	N/A	V26	Broken Window	GPS	No	04:12	06:17

F_CRIME

Fig. 8 Polis-scenariots datatabeller

Exempel på användning av OLAP för Polis-scenariot

Med OLAP kan polis-scenariot exempelvis se om en viss bostadstyp har en högre tendens för att locka tjuvar än andra bostadstyper. Med hjälp av OLAP kan de jämföra bostadstyper som har haft inbrott det senaste året (2009) och även om vart de var belagda har någon betydelse.

```
SELECT    D_TIME.Year, D_VICTIM.Vict_Type_Residence AS Residence,
            D_VICTIM.Vict_Loc_Residence AS Location, count(D_Victim.Vict_Loc_Residence) as
            Count
FROM      D_TIME, D_VICTIM, F_CRIME
WHERE     D_VICTIM.Victim_ID = F_CRIME.Victim.ID AND
            D_TIME.Time_ID = F_CRIME.Time_ID AND
            D_TIME.Year = '2009'
GROUP BY D_VICTIM.Vict_Type_Residence AND D_VICTIM.Vict_Loc_Residence
ORDER BY Count DESC LIMIT 10;
```

Resultat för vilket område som hade flest inbrott 2009

Year	Residence	Location	Count
2009	Appartment	Centrum	123
2009	Appartment	Orgryte	120
2009	House	Tuve	113
2009	Appartment	Tuve	112
2009	House	Orgryte	110
2009	House	Tolered	78
2009	House	Torslanda	46
2009	House	Stampen	44
2009	Appartment	Stampen	12
2009	Appartment	Torslanda	11

Man kan sedan ändra på förfrågningarna lite för att få en annan bild över området, exempelvis om man vill titta på hur gamla personerna var som gjorde inbrott och som sedan kunde gripas av polisen.

```
SELECT    F_CRIME.Criminal_ID AS Crim_ID, D_CRIMINAL.Crim_Age AS CrimAge,
            D_LOC.Loc_PostalCode AS Location
FROM      D_LOC, D_CRIMINAL, F_CRIME
WHERE     D_CRIMINAL.Criminal_ID = F_CRIME.Criminal_ID AND
            D_LOC.Loc_ID = F_CRIME.Loc_ID AND
            F_CRIME.Arrests = 'Yes'
ORDER BY F_CRIME.Criminal_ID ASC LIMIT 8;
```

Resultat för hur gamla förövarna var samt vart brottet skedde

Crim_ID	CrimAge	Location
C10	32	Orgryte
C11	23	Centrum

C12	20	Tuve
C13	19	Centrum
C14	18	Centrum
C14	18	Tuve
C15	43	Lundby
C16	19	Centrum

Crim_ID är med i denna tabell för att identifiera de fall där det är samma person som har gjort inbrottet, är det samma Crim_ID så är det samma förövare men brottet har gjorts på annan plats. Information såsom denna kan vidareutvecklas för att försöka identifiera mönster hos återkommande förövare. OLAP kan även användas för att visa en övergripande bild över hur många arresteringar har gjorts under året genom att analysera all den historiska data som har samlats.

```

SELECT      D_TIME.Year, D_TIME.Month_Name AS Month, F_CRIME.Arrests,
              COUNT(F_CRIME.Arrests) AS Count
FROM        D_TIME, F_CRIME
WHERE       D_TIME.Time_ID = F_CRIME.Time_ID
GROUP BY   D_TIME.Year, D_TIME.Month, F_CRIME.Arrests
ORDER BY   D_TIME.Year, D_TIME.Time_ID, F_CRIME.Arrests DESC LIMIT 8;

```

Resultat för hur många brott som skedde samt hur många som ledde till arrestering

Year	Month	Arrests	Count
2008	October	No	15
2008	November	Yes	12
2008	November	No	10
2008	December	Yes	13
2008	December	No	12
2009	January	Yes	14
2009	January	No	15
2009	February	Yes	9

Hur polisen kan dra nytta av OLAP

OLAP är en teknik för att analysera stora mängder historisk data vilket en samhällssektor såsom polisen kan ha väldigt mycket nytta av. Stora mängder data kan sparas av polisen vilket i sin tur kan hjälpa med att exempelvis försöka förutse trender eller återkommande händelser såsom i det första exemplet med inbrott. Exemplet visar på att OLAP kan hjälpa polisen med att försöka förutse vilka stadsdelar som har flest inbrott samt vilken typ av bostad som oftast får inbrott. Med hjälp av information såsom denna kan polisen i förebyggande syfte koncentrera sig på sådana områden men även informera allmänheten om de ökade riskerna för att dessa själva kan försöka förebygga brott.

OLAP kan även användas i lärande syfte för att se på historien av vad som har hänt för att sedan utnyttja detta i sina utbildningar såsom andra exemplet visar med ålder på förövarna. Denna information kan vidareutvecklas genom att lägga till vart förövarna kommer ifrån, vart dessa har vuxit upp, deras historia om denna finns m.m. för att på så vis avgöra vilken typ av person som är

mest trolig att genomföra brott eller om det finns ett samband mellan ålder och plats. Om polisen kan använda en sådan information för att kunna klassificera möjligheten för att dessa individer har en hög sannolikhet för att begå brott kan förebyggande åtgärder ske i tid.

OLAP kan även användas i syfte för att se på organisationen som helhet genom att använda sig av historisk data för att avgöra effektiviteten med olika processer såsom det sista exemplet i scenariot visar på. Exemplet visar på hur många brott som blivit uppkärlade men information om hur länge det tar för vissa larm att bli utredda, hur lång tid det tog att komma till larmplatsen, hur många högprioriterade larm blir utredda m.m. kan även analyseras. OLAP kan hjälpa polisen med att se på sin egen organisation och på så sätt försöka förbättra den.

4.2.2 Ehandel-scenariots OLAP

Ehandel-scenariot innehåller fyra stycken dimensioner tillsammans med en faktatabell.

Stjärnschemat är fokuserat på vad som har sålts i en internetbaserade butik. Stjärnschemat är uppbyggt med försäljning av böcker i fokus.

Dimensioner:

- **D_TIME:** Tidsdimensionen är likadan som tabellen för polis-scenariot. I en internetbutik är kunden ofta inloggad på sidan redan innan denne vill handla någonting om en registrering innan på webbplatsen har gjorts. Detta gör att en ehandelsbutik inte har några problem med att spara informationen som de vill ha om denna kund. Dimensionen innehåller endast datum och inga tider, exakt tid för varje besök eller varje produkt som besökts skulle snabbt eskalera och orsaka enorma datamängder så detta har studien haft i åtanke. Dagar, veckor, månader samt år går att se på alla nivåer som en analytiker skulle önska.
- **D_PRODUCT:** Behandlar informationen om de produkter som organisationen har till försäljning. Kategori och subkategorier finns för att analytikern skall kunna särskilja mellan de böcker som ehandelsbutiken har till försäljning samt bokens titel och författare. Beskrivning om boken har ett hårt eller mjukt omslag, pocket eller ljudbok finns med i beskrivningen av boken. Bokens förläggare beskrivs i produktdimensionen. Bokens kostnad samt om boken tillhör ett specialerbjudande finns lagrat samt i syfte för att skicka boken till kund finns även storlek och vikt sparad.
- **D_CUSTOMER:** Är den stora dimensionen i detta stjärnschema med information om kunden och dess namn. Kön på kunden finns även med för att se om det finns kopplingar mellan kön och genre eller kön och en viss bok, detta för att förutse köpmönster eller ge förslag till liknande kunder. Kategori på kund finns sparad för att analytikerna skall göra analyser med hur ofta en kund besöker ehandelsbutiken som sparas i nummerform samt med namn. Dimensionen sparar även de böcker som kunden har besökt samt vilken bok som de sedan besökte, med denna hjälp kan analytikerna eller webbplatsansvarig se om den reklam som finns på sidan är till intresse för kunderna eller vilka böcker som kunder oftast tittar på tillsammans. Kunden tilldelas även vilken eller genrer som besöks eller köps oftast som sedan kan ligga till grund för den direktinformation som en ehandelsbutik kanske vill visa för kunden. Hur många gånger som kunden har loggat in finns sparad för att se utveckling över tid samt kundens valuta. Vilken betalning som

kunden föredrar samt dennes leveransinformation sparas för att undersöka om kopplingar kan göras med hjälp av denna information. Kostnaden för leverans till kundens land och den tid som detta tar sparas i tabellen som är baserad på land och stad som kunden redan har registrerat sig med.

- D_SUPPLIER: Dimension för leverantör av de böcker som säljs i stjärnschemat. Dimensionen innehåller namnet på företaget som levererar produkterna samt kontaktpersonen som en ehandelsbutik kan ha där. Kontaktpersonens kontaktuppgifter finns registrerat samt vilken genre som denne handlar med.

Faktatabell:

- F_SALES: Innehåller dimensionsnycklarna för att koppla dessa dimensioner tillsammans med priset på varan och hur många exemplar som köptes vid detta tillfälle. Detta räknas ihop till en totalsumma för att sedan räkna ut den skatt som skall dras av vid försäljning, i detta stjärnschema räknades endast momsen bort för enkelhetens skull. Om produkten har rabatt dras detta av här för att tillslut räkna ihop allting, inklusive fraktkostnaden till en totalsumma. Till slut sparar tabellen när leveransen borde ske och när den faktisk skedde för att undersöka om störningar i denna process finns eller om förbättringar finns att göra.

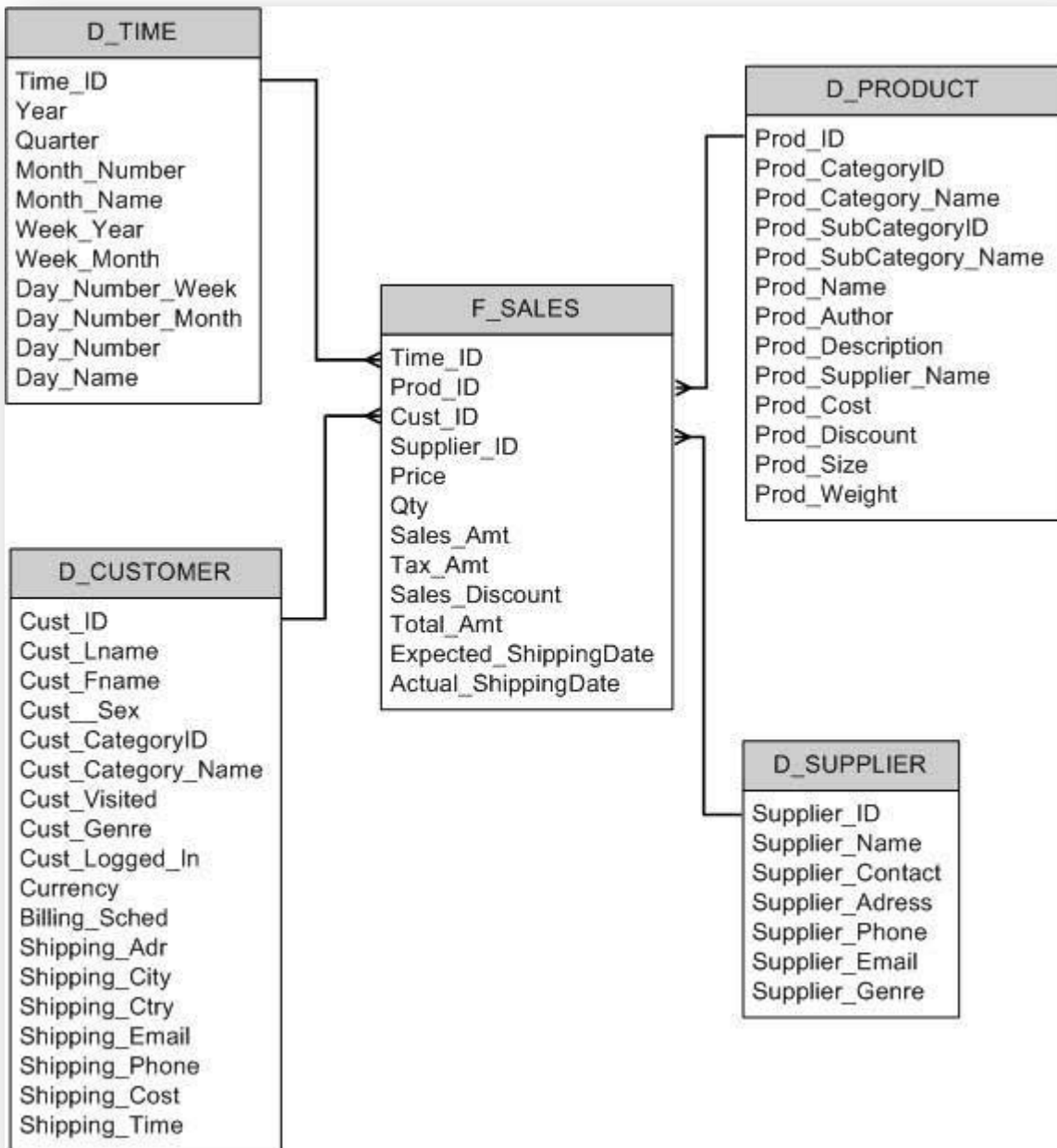


Fig. 9 Ehandels-scenariots stjärnschema

Dataexempel: Härnäst följer ett exempel på vad dessa tabeller kan innehålla för data. All data som finns i tabellerna är fiktiva.

Time_ID	Year	FullDate	Quarter	Month_Nr	Month_Name	Week_Year	Day_Nr_Week	Day_Name	Day_Nr_Month
363	2009	20091229	Q4	12	December	52	4	Thursday	29
364	2009	20091230	Q4	12	December	52	5	Friday	30
365	2009	20091231	Q4	12	December	52	6	Saturday	31
366	2010	20100101	Q1	1	January	1	7	Sunday	1
367	2010	20100102	Q1	1	January	1	1	Monday	2

D_TIME

Prod_ID	Prod_CategoryID	Prod_Category_Name	Prod_SubCategoryID	Prod_Subcategory_Name	Prod_Name
P121	PC1	Books	PSC28	Science	Title#1
P122	PC1	Books	PSC28	Science	Title#2
P123	PC1	Books	PSC28	Science	Title#3
P124	PC1	Books	PSC28	Science	Title#4
P125	PC1	Books	PSC29	Science Fict. & Fantasy	Title#5

D_PRODUCT

Prod_Description	Prod_Supplier_Name	Prod_Cost	Prod_Discount	Prod_Size	Prod_Weight
Hardcover	Vintage	78,87	0	M	0,6
Hardcover	Vintage	71,47	0	M	0,6
Hardcover	PublicAffairs	79,47	10	L	1,2
Hardcover	PublicAffairs	78,87	10	S	0,45
Hardcover	Eos	67,32	0	S	0,45

Cust_ID	Cust_LName	Cust_FName	Cust_Sex	Cust_Category_ID	Cust_Category_Name	Cust_Visited	Cust_Genre	Cust_Logged_In
C1	Lind	Coffy	M	1	Frequent	P2	Fantasy	46
C2	Alm	Amor	M	1	Frequent	P56	Art	44
C3	Berg	Miranda	F	3	Often	P123	Culture	29
C4	Boyd	Linda	F	2	More than often	P103	Fashion	35
C5	Falt	Marie	F	5	Not so often	P132	Science	10

D_CUSTOMER

Cust_Currency	Billing_Sched	Shipping_Adr	Shipping_City	Shipping_Ctry	Cust_Shipping_Email	Shipping_Phone	Shipping_Cost	Shipping_Time
SEK	VISA	Adress 1	Gothenburg	Sweden	lind@coffy.com	+46 31 122334	30	5
SEK	VISA	Adress 2	Malma	Sweden	amor@olm.com	+46 40 345678	30	5
SEK	MASTERCARD	Adress 3	Stockholm	Sweden	berg@miranda.com	+46 8 100010	30	5
SEK	VISA	Adress 4	Gothenburg	Sweden	boyd@linda.com	+46 31 122334	30	5
SEK	CHEQUE	Adress 5	Lund	Sweden	falt@marie.com	+46 46 122222	30	5

Supplier_ID	Supplier_Name	Supplier_Contact	Supplier_Address	Supplier_Phone	Supplier_Email	Supplier_Genre
SU1	Vintage	Larsson Per	Adress 1	+46 31 123456	larsson@per.com	Science
SU2	PublicAffairs	Brett Peter	Adress 2	+46 31 123457	brett@oeter.com	Science
SU3	Tor Books	Sullivan Angela	Adress 3	+46 31 123458	sullivan@angela.com	Fantasy
SU4	Eos	Larke Anna	Adress 4	+46 31 123459	larke@anna.com	Fantasy
SU5	Del Rey	Jemisin Adrian	Adress 5	+46 31 123460	jemisin@adrian.com	Science Fiction

D_SUPPLIER

Time_ID	Prod_ID	Cust_ID	Supplier_ID	Price	Qty	Sales_Amt	Tax_Amt	Sales_Discount	Total_Amt	Expected_ShippingDate	Actual_ShippingDate
363	P125	C1	SU4	159	1	159	9,54	0	198,54	20091229	20091229
363	P2	C1	SU3	79	1	79	4,74	0	113,74	20091229	20100105
363	P10	C1	SU4	79	1	79	4,74	0	113,74	20091229	20091229
364	P123	C4	SU2	149	1	149	8,94	14,9	173,04	20091230	20100101
364	P32	C4	SUB	229	1	229	13,74	0	272,74	20091230	20100101

F_SALES

Fig. 10 Ehandels-scenariots databeller

Exempel på användning OLAP för Ehandels-scenariot

En ehandelsbutik kan använda sig av OLAP för att se sin totala försäljning under alla åren, sorterat på år och sedan månad.

```

SELECT      D_TIME.Year, D_TIME.Month_Name AS Month, sum(F_SALES.Total_Amt) AS Total
              Sales
FROM        D_TIME, F_SALES
WHERE       D_TIME.Time_ID = F_SALES.Time_Id
GROUP BY   D_TIME.Month_Name
ORDER BY   D_TIME.Time_ID LIMIT 5;

```

Resultat för ehandelsbutikens totala försäljning under 2007

Year	Month	Total Sales
2007	April	198 540
2007	May	133 740
2007	June	87 340
2007	July	173 040
2007	August	272 740

Genom att utnyttja sina fördelar med att kunna spara allt som en kund gör på en sida kan OLAP hjälpa en ehandelsbutik med mycket mer som kan vara direkt kopplat till varje enskild kund. Om exempelvis en ehandelsbutik vill börja ge personliga köptips till sina kunder kan de använda sig av OLAP med detta. Om en kund köper en specifik bok kan OLAP användas för att se vilka andra böcker som kunder var intresserade av efter att de köpte just den boken.

```

SELECT      F_SALES.Cust_ID AS CustID, F_SALES.Prod_ID AS ProdID, D_CUSTOMER.Cust_Visited
              AS Vis_ProdID
FROM        D_TIME, F_SALES, D_PRODUCT, D_CUSTOMER
WHERE       F_SALES.Time_ID = D_TIME.Time_ID AND
              F_SALES.Cust_ID = D_CUSTOMER.Cust_ID AND
              F_SALES.Prod_ID = D_PRODUCT.Prod_ID
ORDER BY   D_TIME.Time_ID, F_SALES.Prod_ID, D_CUSTOMER.Cust_Visited LIMIT 6;

```

Resultat för vilka böcker andra kunder var intresserad av efter ett köp

CustID	ProdID	Vis_ProdID
C10	P131	P11
C12	P131	P11
C11	P131	P19
C13	P18	P22
C14	P28	P54
C15	P101	P44

Genom att koppla tidsdimensionen till faktatabellen kan en ehandelsbutik använda sig av OLAP för att kontrollera hur väl leveransprocessen fungerar till slutkunden.

```

SELECT      F_SALES.Time_ID AS Expected, D_TIME.Time_ID AS Actual,
              F_SALES.Expected_ShippingDate AS FullD_Exp, F_SALES.Actual_ShippingDate AS
              FullD_Act
FROM        D_TIME, F_SALES

```

WHERE F_SALES.Actual_ShippingDate = D_TIME.FullDate AND NOT
 F_SALES.Actual_ShippingDate = F_SALES.Expected_ShippingDate
ORDER BY F_SALES.Time_ID, D_TIME.Time_ID LIMIT 6;

Resultat för vilka leveranser som inte skedde i tid

Expected	Actual	FullID_Exp	FullID_Act
363	364	20091229	20091230
363	365	20091229	20091231
363	367	20091229	20100102
363	367	20091229	20100102
364	365	20091230	20091231
364	365	20091230	20091231

Hur en ehandelsbutik kan dra nytta av OLAP

En av de största fördelarna med att använda sig av en ehandelsbutik är att nästan allting som sker inom verksamheten sker elektroniskt. All kontakt med kund eller leverantör kan ske elektroniskt vilket öppnar upp en stor värld när det kommer till att lagra information. Detta gör att OLAP kan vara ett mycket värdefullt verktyg då den data som OLAP behöver för att kunna tillföra värde till verksamheten får en ehandelsbutik nästan gratis och automatiskt tack vare sitt affärsområde. Detta gör att en ehandelsbutik kan ha full översikt över sin verksamhet såsom första exemplet visar med den totala försäljningen men denna översikt kan innebära så mycket mer. Istället för att endast få en översikt över den totala försäljningen kan en ehandelsbutik se ner på varje enskild produkt vilken typ av kund som köper produkten, vilka länder som oftast köper den boken, vilka länder som oftast köper av ehandelsbutiken m.m. En ehandelsbutik kan använda sig av OLAP och på så sätt få information om praktiskt taget allt som involverar deras verksamhet.

Förutom att kunna se på sin verksamhet kan OLAP även hjälpa en ehandelsbutik med att försöka förutse köpmönster från sina kunder såsom andra exemplet visar på. En ehandelsbutik kan som tidigare sagt spara allting som en kund gör på webbplatsen, detta förutsatt att dessa är registrerade och inloggade så webbplatsen kan identifiera dessa. Med hjälp av detta kan en ehandelsbutik registrera vilka böcker som en kund brukar köpa men även vilka böcker som denne tittar på efteråt. Med hjälp av OLAP kan exempelvis ehandelsbutiken jämföra kunder som har köpt likadana böcker, eller som är intresserade av likadana böcker för att sedan ge förslag direkt till den enskilda kunden baserat på vad andra kunder har köpt.

OLAP kan även användas för att se till så att ehandelsbutikens arbetsprocesser fungerar som de skall. Sista exemplet visar på hur OLAP kan hjälpa ehandelsbutiken med att kontrollera så leveranserna till kund har skett i tid vilket antagligen är en av de viktigaste processerna för en ehandelsbutik. Det är viktigt för en ehandelsbutik att leveranserna sker inom den tidsram som de har lovat till kunder då att inte få sina produkter i tid kan medföra att kunderna byter ehandelsbutik. OLAP kan användas för att kontrollera så att inga flaskhalsar finns inom verksamheten där exempelvis den kan användas till att kontrollera om vissa produkter har en tendens att alltid skickas sent, om vissa länder brukar få sena produkter, om leverantörerna håller sitt avtal med leverans till

kund m.m. Detta kan en ehandelsbutik använda sig av för att öka sitt förtroende gentemot sina kunder.

4.2.3 Detaljhandels-scenariots OLAP

Detaljhandels-scenariot innehåller fem stycken dimensionstabeller och en faktatabell som är bestående av mätbar fakta. Fokus i detta schemat har varit på att jämföra detaljhandelns personliga produkter och butiker med konkurrenters butiker, produkter och priser för att se över sin verksamhet ur ett konkurrenskraftigt perspektiv. Faktatabellen samlar all den mätbara datan som den fysiska butiken kan extrahera utifrån sin försäljning.

Dimensionerna:

- D_TIME: Är likadan som i de andra scenariorna utan att någonting har ändrat sig för att kunna analysera dimensionerna ur olika tidsperspektiv med största möjliga valmöjligheter i form av år, kvartal, månad, vecka och dagar.
- D_PRODUCT: Innehåller all information som en analytiker behöver för att jämföra sina produkter med konkurrenternas produkter. Kategori, underkategori och pris finns även med för att undersöka ifall dessa skiljer sig mycket från konkurrenternas nivå.
- STORE_ID: Information gällande butikerna som en detaljhandel kan ha. Sorterade på samma sätt som konkurrensdimensionen samt innehållande land då butiker kan sträcka sig utanför Sveriges gränser. Region, stad och detaljerat för att se på abstrakt respektive detaljerad nivå eller jämförelse på likadana nivåer med konkurrenternas butiker.
- D_COMPETITOR: En dimension innehållande all den information som en detaljhandel har om sina konkurrenter som säljer samma eller liknande produkter. Dimensionen innehåller kategorier på liknande sätt som produktdimensionen med kategori och underkategori samt produktens namn. Priset på produkten från konkurrenternas butiker finns med i denna tabell för att jämföra med detaljhandelns egna prissättning och kostnaden för att köpa in just dessa produkter. Möjlighet till att jämföra produkter med detaljhandelns egna sortiment kan ge planeringsmöjligheter för att köpa in nya produkter som kan ge bättre omsättning.
- D_SUPPLIER: Leverantörsdimension innehållande alla leverantörer för alla märken som organisationen har i sitt sortiment. Att ha information om leverantören i en dimension gör att detaljhandeln kan använda sitt analysprogram och jämföra sina egna leverantörer för att försöka få så bra förmåner som möjligt vid inköp.

Faktatabell:

- F_SALE: Likadan som i tidigare scenario där en faktatabell samlar den mätbara historiska datan.

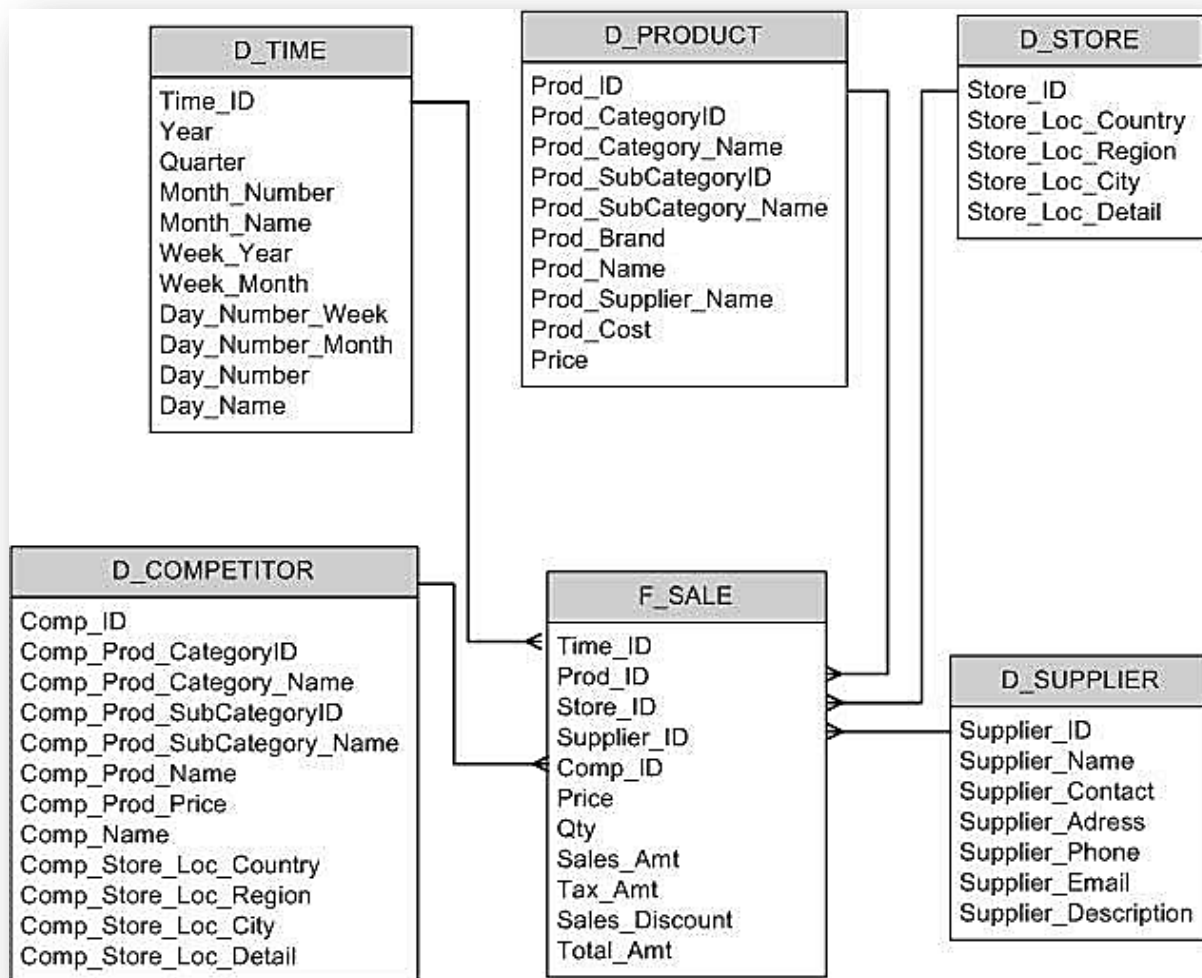


Fig. 11 Detaljhandels-scenariots stjärnschema

Dataexempel: Härnäst följer ett exempel på vad dessa tabeller kan innehålla för data. All data är fiktiv, den faktalösa tabell som detta stjärnschema har presenteras inte då inget av intresse fanns att visa med tabellen.

Time_ID	Year	FullDate	Quarter	Month_Nr	Month_Name	Week_Year	Day_Nr_Week	Day_Name	Day_Nr_Month
363	2009	20091229	Q4	12	December	52	4	Thursday	29
364	2009	20091230	Q4	12	December	52	5	Friday	30
365	2009	20091231	Q4	12	December	52	6	Saturday	31
366	2010	20100101	Q1	1	January	1	7	Sunday	1
367	2010	20100102	Q1	1	January	1	1	Monday	2

D_TIME

Prod_ID	Prod_CategoryID	Prod_Category_Name	Prod_SubCategoryID	Prod_Subcategory_Name	Prod_Brand	Prod_Descrip	Prod_Supplier_Name	Prod_Cost	Price
P22	PC1	Upper Body	PSC2	Sweater Long Sleeve	East West	SW_Long Sleeve 1	Supplier1	229	699
P23	PC1	Upper Body	PSC2	Sweater Long Sleeve	East West	SW_Long Sleeve 2	Supplier1	149	449
P24	PC1	Upper Body	PSC2	Sweater Long Sleeve	Riley	SW_Long Sleeve 3	Supplier2	199	499
P25	PC1	Upper Body	PSC3	Shirt Long Sleeve	Riley	SH_Long Sleeve 1	Supplier2	79	249
P26	PC2	Lower Body	PSC4	Jeans	Lee	Jeans 1	Supplier3	499	799

D_PRODUCT

Store_ID	Store_Loc_Country	Store_Loc_Region	Store_Loc_City	Store_Loc_Detail
ST4	Sweden	Vastra Gotaland	Gothenburg	Centrum
ST5	Sweden	Vastra Gotaland	Gothenburg	Centrum
ST6	Sweden	Vastra Gotaland	Gothenburg	Backaplan
ST7	Sweden	Stockholms Lan	Stockholm	Centrum
ST8	Sweden	Stockholms Lan	Stockholm	Centrum

D_STORE

Comp_ID	Comp_Prod_CategoryID	Comp_Prod_Category_Name	Comp_SubCategoryID	Comp_SubCategory_Name	Comp_Prod_Brand	Comp_Prod_Descrip	Comp_Prod_Price
CMP3	CPC1	Upper Body	CSC1	Sweater Long Sleeve	East West	SW_Long Sleeve 1	559
CMP4	CPC1	Upper Body	CSC2	Shirt Long Sleeve	Riley	SH_Long Sleeve 1	179
CMP5	CPC1	Upper Body	CSC1	Sweater Short Sleeve	East West	SW_Short Sleeve 1	589
CMP6	CPC2	Lower Body	CSC3	Jeans	Lee	Jeans 1	699
CMP7	CPC1	Upper Body	CSC2	Shirt Long Sleeve	Riley	SH_Long Sleeve 2	299

D_COMPETITOR

Comp_name	Comp_Store_Loc_Country	Comp_Store_Loc_Region	Comp_Store_Loc_City	Comp_Store_Loc_Detail
Åhléns	Sweden	Vastra Gotaland	Gothenburg	Centrum
Dressman	Sweden	Vastra Gotaland	Gothenburg	Centrum
H&M	Sweden	Vastra Gotaland	Gothenburg	Centrum
JC	Sweden	Vastra Gotaland	Gothenburg	Centrum
Johansson	Sweden	Vastra Gotaland	Gothenburg	Centrum

Supplier_ID	Supplier_Name	Supplier_Contact	Supplier_Address	Supplier_Phone	Supplier_Email	Supplier_Description
SP1	Supplier1	Larsson Lena	Adress1	+46 31 222222	larsson@lena.com	Sweater
SP2	Supplier2	Mange Per	Adress2	+46 31 222223	mange@per.com	Sweater
SP3	Supplier3	Ekholm Anton	Adress3	+46 31 222224	ekholm@anton.com	Jeans
SP4	Supplier4	Berg Lucas	Adress4	+46 31 222225	berg@lucas.com	Shirts
SP5	Supplier5	Holm Magnus	Adress5	+46 31 222226	holm@magnus.com	Shirts

D_SUPPLIER

Time_ID	Prod_ID	Store_ID	Supplier_ID	Price	Qty	Sales_Amt	Tax_Amt	Sales_Discount	Total_Amt
363	P225	ST4	SP4	229	1	229	68,7	22,9	206,1
363	P22	ST4	SP3	279	2	558	167,4	0	558
363	P210	ST4	SP4	179	1	179	53,7	0	179
364	P31	ST2	SP2	49	3	147	44,1	36,75	110,25
364	P322	ST2	SP8	229	1	229	68,7	0	229

F_SALE

Fig. 12 Detaljhandels-scenariots datatabeller

Exempel på användning av OLAP för detaljhandels-scenariot

OLAP kan även hjälpa detaljhandeln med att informera analytikern om köptrender. Exempelvis skulle OLAP kunna hjälpa detaljhandeln med att se hur försäljningen har gått med Jeans från Lee det senaste året.

```
SELECT      D_TIME.Month, D_PRODUCT.Prod_Brand AS Brand, SUM(F_SALE.Total_Amt) AS
Total_Sum
```


FROM F_SALE, D_PRODUCT, D_TIME
WHERE F_SALE.Prod_ID = D_PRODUCT.Prod_ID AND
 F_SALE.Time_ID = D_TIME.Time_ID AND
 F_SALE.Prod_ID = 'P26' AND
 D_TIME.Year = 2009
GROUP BY D_TIME.Month
ORDER BY D_TIME.Time_ID LIMIT 6;

Resultat för försäljning av Lee Jeans 2009

Month	Brand	Total_Sum
January	Lee	4031
February	Lee	2893
March	Lee	9588
April	Lee	8789
May	Lee	9632
June	Lee	4403

Analytikern kan sedan använda sig av OLAP för att kontrollera ifall det finns något sammanband mellan dessa försäljningsciffror med tidigare, genom att jämföra tidigare års försäljningsciffror kan analytikern kontrollera ifall dessa samband finns.

SELECT D_TIME.Year, D_PRODUCT.Prod_Brand AS Brand, SUM(F_SALE.Total_Amt) AS
 Total_Sum
FROM F_SALE, D_PRODUCT, D_TIME
WHERE F_SALE.Prod_ID = D_PRODUCT.Prod_ID AND
 F_SALE.Time_ID = D_TIME.Time_ID AND
 F_SALE.Prod_ID = 'P26' AND
 D_TIME.Month = 'March'
GROUP BY D_TIME.Year
ORDER BY D_TIME.Time_ID DESC LIMIT 6;

Resultat för försäljning av Lee Jeans under mars månad för varje år

Year	Brand	Total_Sum
2009	Lee	9588
2008	Lee	4020
2007	Lee	9192
2006	Lee	4044
2005	Lee	4204
2004	Lee	6754

Detaljhandeln kan även med hjälp av konkurrentdimensionen jämföra sina produkter och priser med konkurrenternas för att se hur de ligger till jämförelsevis.

SELECT D_PRODUCT.Prod_Descrip AS Description, D_PRODUCT.Prod_Brand AS Brand,
 D_PRODUCT.Prod_Price AS Store_Price, D_COMPETITOR.Comp_Prod_Price AS

```

Comp_Price, D_COMPETITOR.Comp_Name AS Comp_Name, D_PRODUCT.Prod_Cost
AS Prod_Cost
FROM D_PRODUCT, D_COMPETITOR
WHERE D_PRODUCT.Prod_Descrip = D_COMPETITOR.Comp_Prod_Descrip AND
D_PRODUCT.Prod_Brand = D_COMPETITOR.Comp_Prod_Brand
ORDER BY D_PRODUCT.Prod_Descrip, D_PRODUCT.Prod_Brand, D_COMPETITOR.Comp_Name
LIMIT 3;

```

Resultat för detaljhandelns jämförelse av produkter och priser mot konkurrenternas

Description	Brand	Store_Price	Comp_Price	Comp_Name	Prod_Cost
SW_Long Sleeve 1	East West	699	559	Åhléns	229
SH_Long Sleeve 1	Riley	249	179	Johansson	79
Jeans 1	Lee	799	699	JC	499

Hur en detaljhandel kan dra nytta av OLAP

Till skillnad från en ehandelsbutik kan inte en fysiskt belagd detaljhandel spara lika mycket information om sina kunder, oftast kan de inte ens identifiera kunderna eftersom betalning som sker med kontanter innebär att kunden inte registreras i systemet utan endast varan registreras. Detta är dock inget hinder från att använda sig av OLAP då en verksamhet kan lägga till andra dimensioner för att dra fördel över sin situation. Första och andra exemplet som scenariot visar på är liknande de från ehandels-scenariot där detaljhandeln kan använda sig av OLAP för att få en översikt över sina produkter eller butiker men OLAP kan erbjuda mycket mer än så.

Medan en ehandelsbutik hanterar sina kundkontakter elektroniskt kan en fysiskt belagd detaljhandel utnyttja sina fördelar med OLAP med att lägga till dimensioner, fördelar som en fysisk detaljhandel har utöver en elektroniskt är just att de har fysiska aspekter som de kan utnyttja. Dimensioner såsom sina konkurrenter och deras fysiska läge, som tredje exemplet visar med jämförelse av produkter och konkurrenter, är ett sätt att effektivt utnyttja sin situation med OLAP men andra dimensioner kan även tas med för att ytterligare stärka detta. Detaljhandeln kan exempelvis koppla en väderdimension för att analysera ifall vädret har en roll i köpbeteendet hos kunderna, eller koppla information om medelinkomsten i ett område för att analysera ifall det skulle vara lönsamt att öppna en ny butik på den platsen. De kan även använda sig av OLAP för att organisera i butiken genom exempelvis att sätta produkter som säljer mycket längst fram, för att sedan placera produkter som oftast säljs tillsammans med den produkten längre in i butiken. En annan infallsvinkel som OLAP kan erbjuda en detaljhandel är att lägga till en trafikdimension för att analysera vart det är mest trafik och vilka tidpunkter, med detta kan detaljhandeln avgöra vart det är mest optimalt att placera sitt lager för utkörning till butikerna. Genom att använda sig av information som OLAP kan ge en detaljhandel kan de använda detta till sin fördel och försöka komma med så effektiva lösningar som möjligt som gynnar deras situation.

4.3 Visuell hjälp

OLAP är ofta utrustade med visuella hjälpmedel som gör det enklare för användaren att identifiera datan och för att kunna ge så mycket information som möjligt på så kort tid som möjligt, dessa kallas för *dashboards* (Turban et al, 2007). Vanligt med ett dashboard är att det är kopplat till Key

Performance Indicators (KPI) som är till för att definiera och mäta de mål som en organisation har genom att ge en visuell bild över situationen (Turban et al, 2007 – Cuzzocrea & Mansmann, 2009). Skulle ett dashboard användas för denna fallstudie skulle ett dashboard kunna se ut exempelvis (KPI mättal för ehandels-butiken är satta till 140 000 i försäljning varje månad för exemplet) som följande sida visar. Exemplerna är tagna från de första förfrågningarna som gjordes hos respektive scenario.

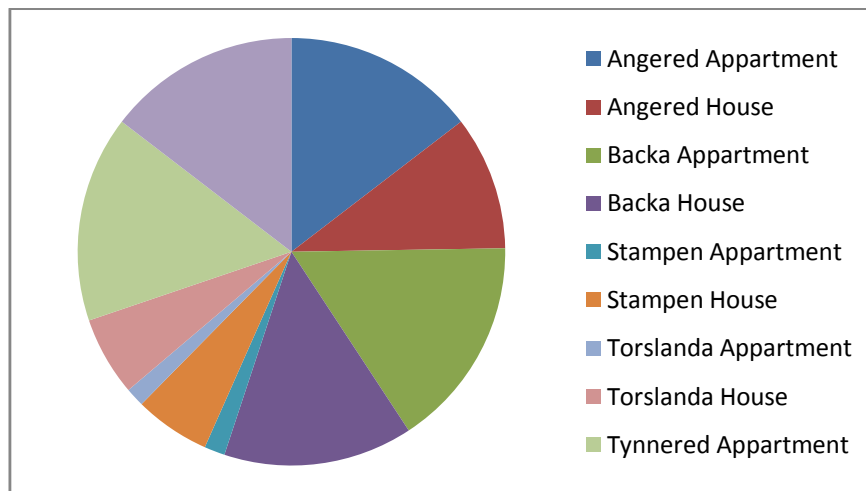


Fig. 13 Polis-scenariots dashboard

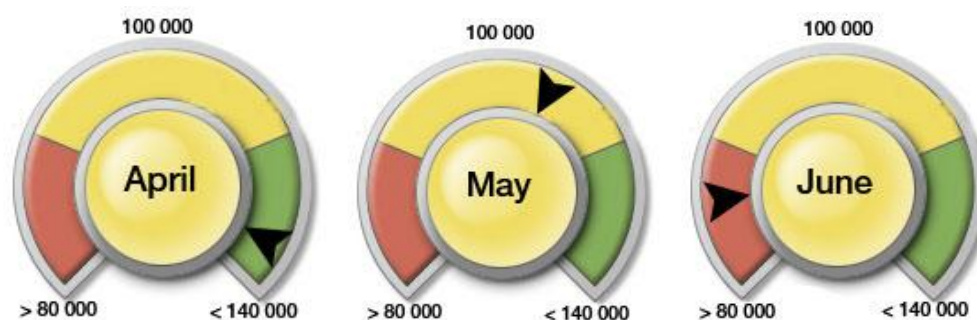


Fig. 14 Ehandels-scenariots dashboard

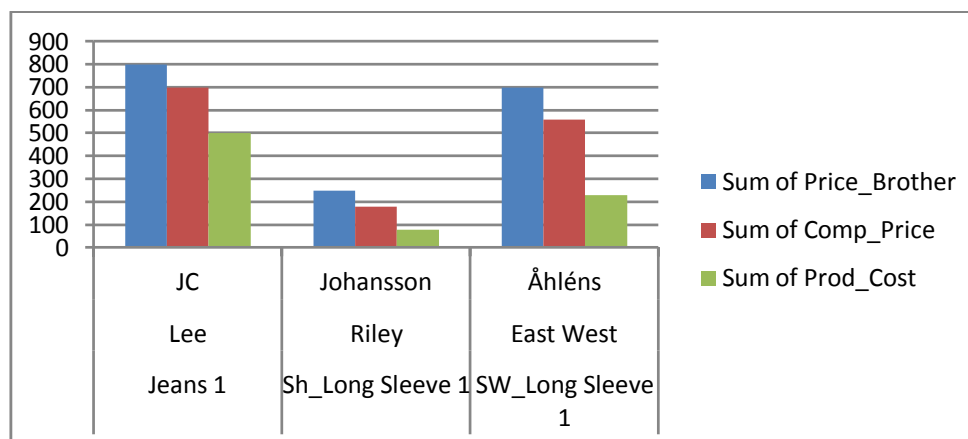


Fig. 15 Detaljhandels-scenariots dashboard

5 Diskussion

Här följer en diskussion om effekterna i verksamheten som följer användandet av OLAP och vad OLAP kan erbjuda verksamheten. Diskussionen har sin grund från fallstudien.

I de olika resultaten från de scenarios som arbetades fram i fallstudien presenterade studien hur olika SQL-satser kunde generera olika svar från olika tabeller, dessa även oftast från fler dimensioner än en. Exempelvis visades att informationen kunde summeras utifrån ehandelsbutiken och därifrån kunde vissa beslutsunderlag identifieras, men även att enkla tabeller presenterades bredvid varandra för detaljerad jämförelse och analys av informationen såsom för den fysiska detaljhandeln

Det var uppenbart under studien att möjligheten att kunna se på informationen på olika sätt, antingen på en övergripande nivå med exempelvis ehandels-scenariot som bland annat ville se sin försäljning per månad men även för detaljhandel-scenariot där analyseringen skedde på produktnivå var oerhört kraftfullt och intressant att kunna vrida och vända på datan för analyseringar. Detta är en av grundstenarna med OLAP att genom applicering av dimensioner kunna analysera data på så många olika sätt som möjligt, detta är även vad som gör OLAP så användbart för verksamheter då det tillför så många olika valmöjligheter. Genom att erbjuda så många valmöjligheter så innebär det att OLAP är väldigt flexibelt för verksamheter som ehandels- och detaljhandels-scenariorna visade, där scenariorna visade att OLAP kan tillföra mycket nytta till en verksamhet om de utnyttjar sina fördelar tillsammans med OLAP.

Det gäller dock för användaren att identifiera de dimensioner som denne kan använda för att på bästa sätt utnyttja OLAP-verktyget. Under utarbetandet av scenariorna var identifieringen av de dimensioner som skulle användas antagligen den svåraste aspekten för att lyckas få ut den information som var intressant. Det var ibland svårt för författaren att skaffa sig en minnesbild över de dimensioner som fanns och tabeller som dessa innehöll för att sedan koppla ihop dessa, det var antagligen en bit av kognitiv överbelastning som författaren råkade ut för men skulle kunna likna med problem som användare skulle kunna råka ut för när dessa vill göra ad hoc analyser. Ad hoc-analys i denna bemärkelse menar författaren att det kan uppkomma situationer när användaren plötsligt vill få svar på en fråga som denne bestämmer sig för snabbt. En likadan situation som författaren hamnade i där det var svårt att komma ihåg vilken data som fanns i vilken dimension kan användaren hamna i när denne vill få ut ett svar från OLAP som denne vanligtvis inte ställer systemet som ofta är fallet i ad hoc-analyser. Exempelvis om en användare brukar titta på försäljning och försäljningens utveckling över tid men plötsligt vill titta på försäljning och vart detta har skett. Detta kan innebära att istället för två dimensioner (faktatabell och tidsdimension för att summera totalsumma såsom i ehandels-scenariot) som användaren brukar använda sig av måste ytterligare dimensioner läggas till. Som de olika scenariorna visar kan plats vara en dimension för sig självt eller så är den med i en större dimension exempelvis i en dimension som visar butiker, det finns inga speciella regler för vilka tabeller som måste vara i vilka dimensioner utan det är upp till utvecklaren så länge tabellerna kan identifieras. När OLAP blir många dimensioner är det vara speciellt viktigt att användaren kan identifiera vart de tabeller som denne vill använda finns där den tror eller förväntar sig.

En annan sak som uppenbarade sig under fallstudien angående dimensioner var användandet av tidsdimensionen. Tidsdimensionen användes i alla tre scenarios som studien presenterade men det fanns aldrig någon anledning till att ändra på den, författaren försökte flera gånger men mångfalden av tidsdimensioner som en analytiker kan tänkas använda sig av gjorde att den fortsatte genom alla scenarios oförändrad. Tidsdimensionen är enligt författaren och studien en av den viktigaste dimensionen som en OLAP-kub är uppbyggd med och för mycket information skulle gå förlorad om denna dimension inte är uppbyggd rätt.

Fallstudien presenterade förutom de tabeller och data som kunde extraheras även hur ett dashboard eller visuellt hjälpmedel kan se ut i OLAP. Det visuella hjälpmedlet kan omvandla data som OLAP presenterar i form av bilder, tabeller, bildmarkeringar på kartor m.m. Det gör att det visuella hjälpmedlet inte bara presenterar informationen mycket snabbare till användaren, vilket gör det effektivare än med tabeller som endast visar siffror, den kognitiva överbelastningen genom att bara se siffror och data minskar även. Att bearbeta bilder istället för data är mycket enklare för användaren och mer information kan visas till användaren på samma gång. Det var en intressant aspekt att först extrahera informationen i tabeller för att sedan transformera dessa till bildformat, det var en mycket enkel process för författaren då det är enklare att arbeta med bilder än vad det är med siffror och bokstäver och de kan kombineras till många olika varianter. Cuzzocrea & Mansmann (2009) skriver att OLAP-system som endast erbjuder svar i form av SQL som de tabeller de olika scenariorna visar på inte kan mäta sig med visuella OLAP-system. Vissa dimensioner går helt enkelt inte att analysera med icke visuella OLAP-system, efter fallstudien kan författaren inte annat än att hålla med vad dessa två författare skriver.

I polis-scenariot visades hur ett OLAP kan ge information om vilka bostäder som fick inbrott och samtidigt vilket område detta hade skett flest gånger. Skulle denna data istället presenteras med hjälp av en karta och markeringar på kartan för var och en av adresserna som hade fått inbrott, skulle det visuella OLAP-systemet kunna visa användaren exakt vart dessa låg på kartan istället för med endast en adress på en informationsrad, detta skulle i sin tur kunna lyfta fram ytterligare ett mönster till en analytiker och göra OLAP mer funktionellt för analytikern. Ett annat exempel skulle vara att om man återgår till detaljhandels-scenariot där jämförelse av produkter mot konkurrenternas produkter skedde så skulle användaren med hjälp av ett visuellt OLAP-system kunna se på en karta vart butikerna låg, vilken produkt som sålde bäst i vilket område, vilka konkurrenter som finns i närheten, vilken årsinkomst medelinkomsttagaren har i det område som butikerna ligger med mera. Finns bara grunddata så verkar det bara vara fantasin som sätter gränser för vad ett OLAP kan erbjuda användaren i form av analytiska undersökningar.

Att använda sig av ett visuellt OLAP öppnar inte bara upp för fler dimensioner ur analytisk synpunkt då analytikerna exempelvis kan lägga till geografisk plats för sitt problemområde, det kan även öppna upp för en annan typ av användare. Alla användare är inte experter inom databaser eller arbetar som analytiker/beslutstagande, det kan vara vanliga användare som vill föra statistik över sina aktieköp de senaste åren och se hur dessa har utvecklats sig över tiden. Molntjänster idag är ett omtalat ämne som innebär att tekniska tjänster läggs ut på internet som gör att det endast behövs åtkomst till Internet för att användaren skall kunna ta del av dessa tjänster. Om man erbjuder ett visuellt OLAP-system på detta sätt till vanliga användare som innehåller all den hjälp som en användare skulle behöva för att själva lägga in sin data och skapa de dimensioner som krävs så skulle kanske inte OLAP vara lika begränsad till stora företag som det är idag. Om man med det visuella

hjälpmedlet kan göra det enkelt, informativt och webbaserat för användaren men samtidigt erbjuda all den funktionalitet som OLAP innehåller kan man på så sätt få vanliga användare att utnyttja OLAP oftare än vad som är utspritt idag, det gäller bara för användaren att ha grunddata som kan analyseras.

Den data som fanns i de olika scenariornas datatabellerna var den data som författaren förväntade sig att tabellerna skulle ha, men det är inte alltid så enkelt. Om författaren hade ändrat lite i tabellerna och ersatt tabellerna med data som var av lägre kvalitet och inte var konsistent till den övriga datan, exempelvis om datumen istället var ifyllda fel eller att nullvärden fanns med i tabellerna skulle det innebära att informationen som återgavs skulle vara felaktig. Om fel information ges till användaren för ofta kan det sluta med att användaren inte längre har någon tillförlitlighet till systemet och slutar helt enkelt att använda sig av systemet helt och hållet. Det finns ingen större anledning för användaren att använda sig av ett analytiskt system som inte återger en verklighetsbild som användaren kan bygga vidare på. Att ha bra kvalitet i sitt datalager är av högsta prioritet eftersom OLAP kan inte avgöra ifall informationen är rätt eller inte, det är upp till användaren eller organisationen att se till att denna information som finns i datalagret är korrekt.

Det kanske inte alltid är organisationen orsak till att datakvalitén är bristande i ett OLAP-system dock. Om man exempelvis går tillbaka till ehandels-scenariot där kunddimensionen sparade allt det kunden gjorde på sidan. Om man spekulerar och föreställer sig att istället för att det är den registrerade kunden som handlar på sidan så är det dennes barn som handlar ibland istället. Dessa barn är för unga för att ha ett kreditkort vilket resulterar i att de använder sig av sin fars inloggning och kreditkort för att kunna köpa böcker via en ehandelsbutik. När dessa barn har köpt böckerna från sin fars inloggning och med dennes kreditkort innebär det att det är fel data som sparas av OLAP-systemet, istället för att endast faderns uppgifter lagras så lagras även allt vad barnen gör på sidan. Det är fullständig rätt data från början eftersom inloggning borde resultera i att rätt person sitter bakom tangentbordet men eftersom kontot delas mellan personer i familjen så sparas fel information, vilket kan leda till att fadern får reklam om ungdomsböcker eller barnböcker som denne kanske inte är intresserad av då OLAP-analyserna blir felaktiga men med hjälp av rätt data.

5.1 Studiekritik

Författaren är medveten om att den miljö som har undersökts är samma miljö som författaren har befunit sig i vilket leder till en självutvärdering som inte skall göras i en vetenskaplig uppsats då med risk för att studien får en subjektiv präglning. Författaren har försökt att se på all den information och data som har uppkommit från fallstudien på ett objektivt sätt och har försökt undvika att låta det egna arbetet påverka studien, genom att hela tiden försöka se de scenarios som har arbetats fram som om dessa kom från den empiriska världen och riktiga företag.

6 Slutsats

Utifrån studiens insamlade material kan slutsatsen dras till att OLAP är ett kraftfullt verktyg som kan möjliggöra effekter i en verksamhet på många olika sätt. Genom OLAP-kuben kan stora mängder data analyseras från flera olika dimensioner och på så sätt tillföra mer värde till verksamheter som studien har visat på med hjälp av de scenarios som presenterats. Studien visar även på att det kan vara svårt för användaren att identifiera de möjliga dimensioner och hitta de tabeller som användaren är intresserad att använda sig utav.

En annan viktig aspekt visade sig vara datakvalité då felaktig data kan leda till fel information till användaren som i sin tur kan leda till ett oanvändbart OLAP. Det behöver dock inte alltid vara verksamhetens fel att datakvalitén är dålig. Rätt data kan sparas som rätt fakta men på grund av oförutsedda aspekter eller faktorer som organisationen inte kan påverka kan datan ändå bli fel genom att felaktig data registreras i databasen.

Utifrån studiens frågeställning som löd: *Vilka möjliga effekter kan OLAP tillföra i en verksamhet?* Kan slutsatsen dras till att effekterna varierade beroende på vilken branch eller affärsområde som OLAP nyttjas i. I alla de scenarios som arbetades fram fanns det dock positiva aspekter med användandet av OLAP och så länge som verksamheterna utnyttjade sina fördelar så medförde OLAP nytta till respektive verksamhet.

7 Källförteckning

Böcker

Cornford, T., & Smithson, S. (2006) *Project Research in Information Systems: A Student's Guide*. New York: Palgrave Macmillan

Löwgren, J., & Stolterman, E. (2008) *Design av informationsteknik – Materialet utan egenskaper*. Pozkal: Studentlitteratur

Padron - McCarthy, T., & Risch, T. (2005) *Databasteknik*. Lund: Studentlitteratur

Patel, R., & Davidson, B. (2003) *Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning*. Lund: Studentlitteratur.

Riccardi, G. (2001) *Principles of Database Systems with Internet and Javatm Applications*. USA: Addison Wesley.

Turban, E., Aronson, J.E., Liang, T-P., & Sharda, Ramesh. (2007) *Decision Support and Business Intelligence Systems*. New Jersey: Pearson

Artiklar

Cuzzocrea, A., & Mansmann, S. (2009) *OLAP Visualization: Models, Issues, and Techniques* IGI Global, Sid 1439-1446

Martyn, T. (2004) Reconsidering Multi-Dimensional Schemas. *SIGMOD Record*. Vol. 33, Nr. 1, sid 83-88.

Moody, D., & Kortink, M. (2000) From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. *DMDW2000* Vol. 28, Article 5, sid 5-1 – 5-12.

Mishra, D., Yazici, A., & Basaran, B.P. (2007) A Comparison of Design Models in Data Warehousing. *International Journal of Information Studies*. Vol . 1, Nr. 1 sid 77-90

Rizzi, S., Abelló, A., Lechtenböcker, J., & Trujillo, J. (2006) *Research in Data Warehouse Modeling and Design: Dead or Alive?* USA: ACM.

Shin, S.K., & Sanders, G.L. (2006) Denormalization strategies for data retrieval from data warehouses. *Decision Support Systems*. Vol . 42, Nr. 1, sid 267-282.

Webbplatser

Association for Information Systems (AIS) URL:

http://home.aisnet.org/associations/7499/files/Index_Markup.cfm

Hämtad 2010-05-10

Challenger Rymdfärjan: URL <http://dssresources.com/cases/spaceshuttlechallenger/index.html>

Hämtad 2010-05-17

Straub, D., Gefen, D., and Boudreau, M.-C. "The IS World Quantitative, Positivist Research Methods Website", 2004, URL: <http://dstraub.cis.gsu.edu:88/quant/>

Hämtad 2010-05-10.