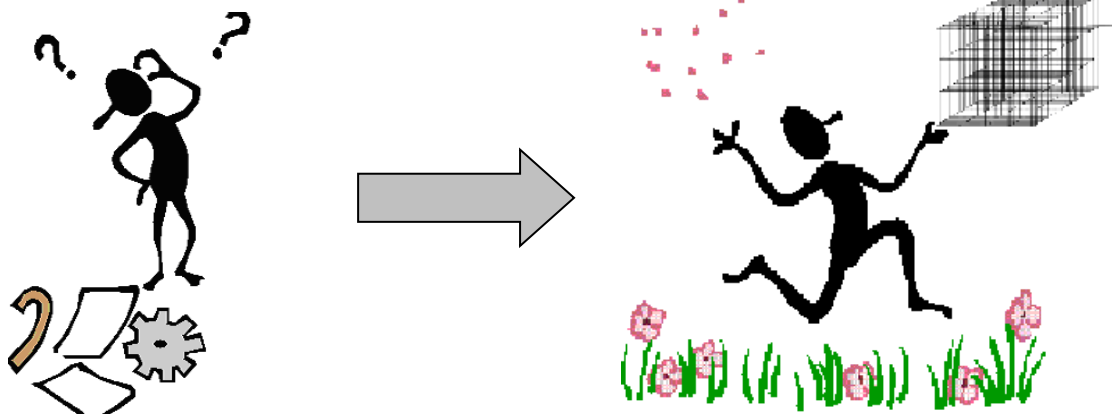


# Gestroomlijnde informatievoorziening met OLAP



Han Long Li  
februari - augustus 2003

## **Stageverslag**

Vrije Universiteit  
Faculteit der Exacte Wetenschappen  
Studierichting Bedrijfswiskunde en Informatica

Vermogensplanet  
Burgmeester Stramanweg 105A  
Amsterdam



## Voorwoord

De weg naar een gestroomlijnde informatievoorziening van een organisatie is te vergelijken met de weg naar boeddhistische perfectie. Het oude China telde heel veel boeddhisten die hard werkten om de innerlijke perfectie te ontdekken in hun ziel. Met andere woorden, de Boeddha in zichzelf te ontwaken. Ze vergelijkten het hard werken en het hard leren van de boeddhistische leer als het reinigen van hun ziel. Een beroemde monnik was het niet eens met deze denk- en werkwijze. Hij schreef een gedicht dat tot op de dag van vandaag algemeen wordt geaccepteerd door de boeddhisten. De belangrijkste stelling van het gedicht luidt als volgt:

*“Waarom reinig je de Boeddha? Boeddha is al schoon van zichzelf.”*

Voor boeddhisten komt inspanning tot uiting in twee soorten veranderingen:

Aan de ene kant de **kwantitatieve** verandering: het hard werken, het leren in de boeddhistische leer en het reinigen van je ziel. Elke inspanning die je hiervoor doet, uit zich in een kwantitatieve verandering; je kent een extra regel in de leer.

Aan de andere kant de **kwalitatieve** verandering: als iemand de wijsheden van Boeddha beheerst, dan moet zijn ziel wel schoon zijn. Waarom zal je dan nog moeite doen zijn ziel te reinigen?

Graag wil ik van de mogelijkheid gebruik maken de heer Jos van den Broek en de heer Raymond van Rootzelaar te bedanken voor hun enorme inzet bij het redigeren van mijn scriptie. Het feilloos schrijven in het Nederlands is voor mij, met de Chinese taal als mijn moedertaal, nog steeds moeilijk. Ik draai bijvoorbeeld nog geregeld die verdraaide Nederlandse zinsconstructies om. Zin voor zin, woord voor woord, is deze scriptie door beide heren nagekeken en waar nodig aangepast. Zonder hen zou mijn scriptie nooit kunnen zijn geworden tot wat het nu is. Verder gaat mijn dank natuurlijk ook uit naar professor Eiben (externe begeleider), de heer Keijzer (interne begeleider), de heer Kerstel (dagelijkse begeleiding), de heer Ter Bekke (een goede collega) en alle andere collega's die hebben bijgedragen aan het succes van mijn stage.

Na mijn stage mag ik blijven werken voor VP. Ik zal in mijn manier van werken zeker zoveel mogelijk de kwalitatieve verbetering van het werk voorstaan. Hierbij zal ik bovenstaande boeddhistische wijsheid altijd in mijn achterhoofd houden.

Han Long  
Amsterdam, 2003



# Inhoudsopgave

1	Inleiding .....	1
1.1.	Probleemstelling .....	1
1.2.	Vermogensplanet .....	2
1.3.	Doelstelling opdrachtgever .....	5
1.4.	Doelstelling stage .....	6
1.5.	Hoofdstuk indeling van de scriptie .....	8
2	Inleiding OLAP .....	9
2.1.	Wat is OLAP? .....	9
2.1.1.	Definitie .....	9
2.1.2.	OLAP concepten .....	10
2.1.3.	OLAP als onderdeel van BI .....	14
2.2.	Wanneer is gebruik van OLAP interessant? .....	15
2.2.1.	Beoogde voordelen van het gebruik van OLAP .....	15
2.2.2.	OLAP versus Ad-hoc <i>query</i> .....	16
2.2.3.	OLAP versus Spreadsheets .....	16
2.2.4.	OLAP versus data mining .....	17
2.3.	Wat zijn de beperkingen van OLAP? .....	18
2.4.	Samenvatting .....	19
3	Opbouw van een OLAP kubus .....	20
3.1.	Data warehouse .....	20
3.2.	Werkwijze van de OLAP-tool .....	22
3.3.	Eisen van de onderliggende data voor OLAP .....	23
3.4.	Samenvatting .....	24
4	Logische opzet van OLAP datamodellen .....	25
4.1.	Wanneer kan een model worden uitgebreid? .....	25
4.2.	Semantische kubus versus hyperkubus .....	28
4.3.	Keuze van aggregatieniveau .....	29
4.4.	Samenvatting .....	29
5	Situatie analyse aan het begin van het OLAP project .....	30
5.1.	Toestand van de data opslag .....	30
5.1.1.	Data opslag binnen het VP platform .....	30
5.1.2.	Data opslag buiten het VP platform .....	31
5.2.	Toestand van de informatievoorziening .....	32
5.3.	Rollen van de mensen binnen dit informatiesysteem .....	33
5.4.	Een mogelijke oplossing met OLAP .....	34
5.5.	Samenvatting .....	35
6	BI ondersteuning voor Vermogensplanet .....	36
6.1.	Data opslag .....	36
6.2.	OLAP oplossingen in het algemeen .....	36
6.2.1.	Identificeren van de algemene informatievragen .....	37
6.2.2.	Keuze van de te implementeren kubus .....	39
6.2.3.	Welke beslissingen ondersteunt de monitoringkubus? .....	39
6.3.	De implementatiestap .....	40
6.3.1.	Communiceren met behulp van definities .....	40
6.3.2.	Allocatie van de capaciteit .....	42
6.4.	Samenvatting .....	44
7	Project evaluatie .....	45
7.1.	Momentopname na levering .....	45
7.1.1.	Data opslag .....	45
7.1.2.	Informatievoorziening .....	45
7.1.3.	De rollen .....	46
7.2.	Gerealiseerde toegevoegde waarde .....	46
7.3.	Extra vragen bij het project oversluiten .....	47
7.3.1.	Verschillen tussen het <i>oversluiten</i> en het WP project .....	48
7.3.2.	Wat zijn de extra eisen van <i>oversluiten</i> ? .....	48
7.3.3.	Waarom kan de juni-kubus niet aan deze extra eisen voldoen? .....	49

7.3.4.	Een tijdelijke oplossing binnen juni-kubus voor <i>oversluiten</i> .....	49
7.3.5.	Een structurele oplossing .....	50
7.4.	Finetuning van de monitoringkubus .....	50
7.4.1.	Herkomst vertaling .....	50
7.4.2.	Standaardisering via <i>divebooks</i> .....	51
7.4.3.	Scheiding in projecten .....	51
7.5.	Realiseren van andere algemene kubussen .....	51
7.6.	OLAP draaiboek .....	52
7.6.1.	Formuleren <i>businessrequirements</i> .....	53
7.6.2.	Vertalen <i>businessrequirements</i> .....	53
7.6.3.	Toetsen beoogde resultaat (“verifieerbaarheid”) .....	54
7.6.4.	Vertalen onderhoudsbehoeften (“onderhoudbaarheid”) .....	54
7.7.	Samenvatting .....	55
8	Knelpunten en aanbevelingen .....	56
8.1.	Aggregatie niveau .....	56
8.2.	Communicatie .....	56
8.3.	Multi-dimensionaal besef bij de techniek .....	57
8.4.	Multi-dimensionale besef bij de business .....	57
8.5.	Capaciteit .....	57
8.6.	Aanbevelingen .....	58
8.7.	Samenvatting .....	59
9	Theorie versus praktijk .....	60
10	Conclusie .....	62
10.1.	Samenvatting stage resultaat .....	62
10.2.	Antwoord op het centrale vraagstuk .....	63
Appendix A.	Proces en data bij leadgeneration .....	64
Appendix B.	Doelstelling van de opdrachtgever .....	65
Appendix C.	Tijdsplanning stage .....	66
Appendix D.	Ster-schema oorspronkelijk VP platform .....	69
Appendix E.	Schema tabellen VP platform .....	70
Appendix F.	Monitoringkubus WP en Oversluiten .....	71
Appendix G.	Data voor monitoringkubus WP en Oversluiten .....	73
Appendix H.	Status schema juni-project .....	79
Appendix I.	Takenlijst Juni-project .....	80
Appendix J.	Multi-dimensionale besef bij technici .....	81
Literatuurlijst	.....	82

# 1 Inleiding

Ik heb mijn stage mogen lopen bij Vermogensplanet (VP). De doelstelling van mijn stage was het stroomlijnen van de informatievoorziening binnen VP.

Tijdens mijn stage kwam ik tot de ontdekking dat de technische afdeling van VP kampt met capaciteitstekorten. Mijn stageproject werd hierdoor ook geraakt met als gevolg een vertraging van anderhalve maand.

Een deeloplossing om de techniek structureel te kunnen ontlasten kan worden gezocht in het stroomlijnen van de informatievoorziening – het doel van mijn stage. Vooral bij de informatiebehoeften die regelmatig terugkomen in de bedrijfsvoering kan automatisering een enorme capaciteitsbesparing opleveren.

Deze scriptie is gebaseerd op mijn stageproject waarin de **On-Line Analytical Processing** (OLAP) analyses werden toegepast in het informatiesysteem van VP. De oplevering van het OLAP datamodel dat een deel van de informatiebehoefte bedient, kan gezien worden als een stap naar de kwalitatieve verbetering in de informatievoorziening.

## 1.1. Probleemstelling

Sinds het ontstaan van geautomatiseerde databases is de hoeveelheid opgeslagen data exponentieel toegenomen. De interpretatie van deze data, ofwel het halen van kennis uit de opgeslagen data, is hierbij sterk achtergebleven. Dit wordt ook wel aangeduid met de *knowledge gap*<sup>1</sup>. Deze *knowledge gap* wordt als gevolg van de lage kosten voor data opslag, almaar groter.

Business Intelligence (BI) is een set van applicaties die bijdraagt aan het verkleinen van deze *knowledge gap*. Een Engelstalige definitie van BI luidt:

*“Business intelligence is a broad category of applications and technologies for gathering, storing, analyzing and providing access to data to help enterprise users make better business decisions.”*

[whatis.com,2003]

Om data te analyseren, heeft BI het volgende instrumenten:

- *Data warehousing*
- On-Line Analytical Processing (OLAP)
- *Data mining*

---

<sup>1</sup> Kenniskloof

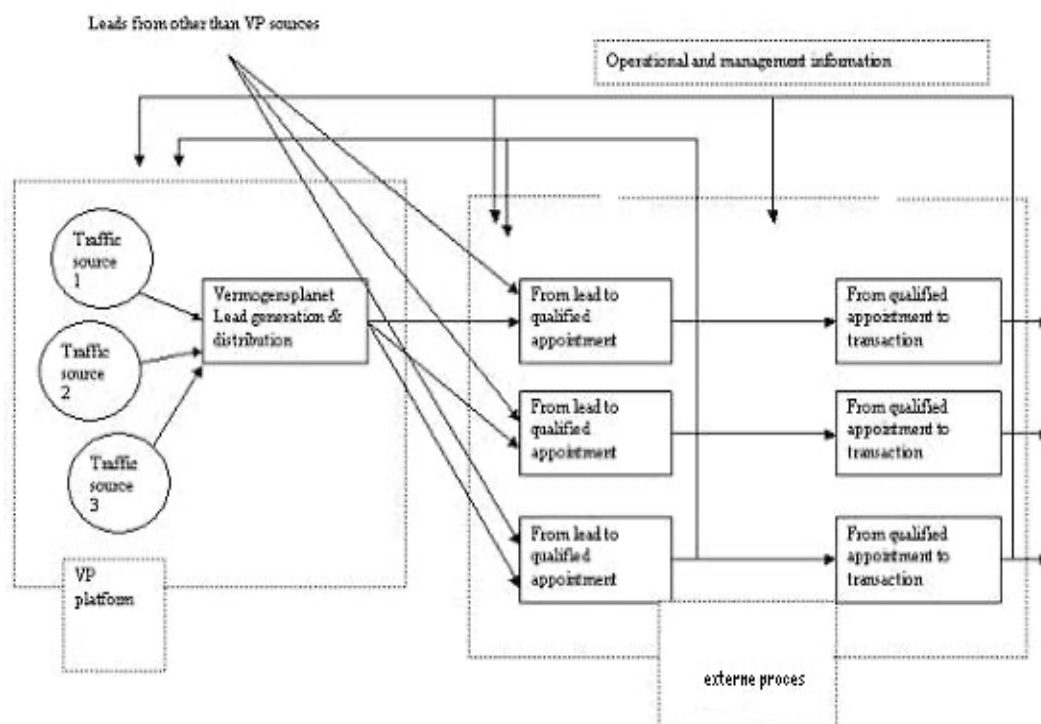
## De centrale probleemstelling van deze scriptie is:

Hoe kan de *knowledge gap* worden verkleind met Business Intelligence (BI); in het bijzonder met gebruikmaking van OLAP? Vertaald naar mijn stage: Hoe kan de informatie behoefte in een organisatie als Vermogensplanet worden gestroomlijnd?

### 1.2. Vermogensplanet

VP is een snel groeiend dynamisch bedrijf met ongeveer 15 werknemers. Het heeft lokaties in Amsterdam en Enschede. VP treedt op als online intermediair voor financiële producten. Hierbij genereert en verkoopt VP *leads*.

De positie van VP binnen het distributienetwerk van *leads* wordt weergegeven in figuur 1-1 [Marc Kerstel]. VP genereert *leads* en verkoopt deze door aan haar partners. Een persoon die interesse toont voor een bepaald financieel product en daarbij toestemming geeft voor verder contact wordt beschouwd als een *lead*.



figuur 1-1 Positionering VP

Het afhandelen van *leads* gebeurt met de tussenkomst van een callcenter. Het callcenter maakt afspraken met de *leads*. De financiële partner pakt vervolgens de door het callcenter gemaakte afspraken op door een financieel adviseur op te zetten. Uiteindelijk



worden de offertes getekend door de klanten die de financiële producten willen afnemen, dit levert transacties op.

Meestal kost het een aantal dagen voordat contact is opgenomen met een *lead*. De doorlooptijd van de afspraak hangt mede af van de agenda's van de adviseurs en die van de *leads*. Vaak wordt na een eerste afspraak plus bedenktijd van de potentiële klant, een vervolgspraak gepland. Bij elkaar kan de doorlooptijd van het gehele traject oplopen tot een aantal maanden.

De provisie aan VP kan worden berekend aan de hand van het:

- aantal *leads*
- aantal gemaakte afspraken
- aantal transacties

Voor VP houdt het operationele proces op wanneer een *lead* het VP platform (zie figuur 1-1) verlaat.

VP kan worden vergeleken met een productiebedrijf. VP "produceert" *leads*. Voor de productie zijn net als in een gewoon productiebedrijf, *grondstoffen* nodig. De *grondstof* om *leads* te produceren is de bezoekers op de website van VP. Deze kunnen worden ingekocht.

Een website die beheerd wordt door VP kan worden beschouwd als de *bedrijfsruimte*. Op de door VP beheerde websites staan rekenmodules voor de verschillende financiële producten. Deze rekenmodules zijn vergelijkbaar met de *machines* van een productiebedrijf; ze *produceren leads* voor VP.

Een bezoeker die op een door VP beheerde website komt, kan daar op zoek zijn naar bepaalde informatie. Door met behulp van de rekenmodules berekeningen uit te voeren krijgt de bezoeker inzicht in zijn eigen financiële situatie. Wanneer de bezoeker informatie over zijn bereikbaarheid achterlaat, is een *lead geproduceerd*. VP kan *leads* als *eindprodukt* verkopen. In een productiebedrijf koopt men de *grondstof* in, bij VP worden bezoekers ingekocht. De *traffic source*<sup>2</sup> (ofwel *leadbronnen*) zijn de leveranciers van VP. Om bezoekers te trekken naar de websites, sluit VP overeenkomsten af met deze leveranciers. Deze overeenkomsten worden onderverdeeld in:

- Cost per *click* (CPC)
- Cost per *lead* (CPL)

Bij een **CPC** overeenkomst wordt bijvoorbeeld een banner geplaatst op de website van de leverancier. Het aantal klikken op de banner wordt geregistreerd en wordt gefactureerd aan VP. Bij een **CPL** overeenkomst wordt in de meeste gevallen de rekenmodule geplaatst op de website van de leverancier. De leverancier wordt betaald aan de hand van het aantal *leads* dat wordt opgeleverd vanuit de betreffende rekenmodule.

---

<sup>2</sup> Zie figuur 1-1

Een rekenmodule wordt binnen VP een *Customer Journey* (CJ) genoemd.

Een CJ bevat:

- een welkomstbericht
- een invoerscherm voor de berekening
- een invoerscherm voor de contact gegevens
- het resultaat scherm

Het achterliggende idee van een CJ is dat binnen een CJ geen *zijpaden* mogelijk zijn. De klanten kunnen afhaken of zij kunnen de CJ afmaken. *Zijpaden* die de klanten naar een ander propositie leiden voordat een klant de CJ afmaakt zijn er dus niet.

De complexiteit van het interpreteren van bezoekersgedrag wordt beperkt met de CJ's. Aan de hand van de CJ stelt VP vast welk soort informatie de bezoeker op een website zoekt.

Een organisatie die al in het bezit is van contact gegevens van haar klanten, kan VP benaderen wanneer zij de online marketing wil uitbesteden. VP heeft als taak ervoor te zorgen dat deze klanten online zijn te benaderen door de opdrachtgever. Hiervoor bouwt en onderhoudt VP zogenaamde besloten websites, waarbij de klant wordt voorzien van een *login* en een *userID* zodat de klant op de besloten website kan inloggen. De opdrachtgevers kunnen via deze besloten websites zaken doen met haar klanten. Op dit moment heeft VP meer dan 60 websites onder haar beheer; hiervan zijn er enkele besloten. VP werkt samen met enkele tientallen leveranciers en heeft rond de tien afnemers.

VP genereert héél veel data met de door hen beheerde online campagnes. VP wil graag meer waarde uit de data analyse halen. Een conceptueel figuur van het *lead*genererende proces met bijbehorende datastore, kan worden gevonden in Appendix A.

In figuur 1-1 is goed zichtbaar dat er ook andere partijen zijn die *leads* leveren aan de afnemers van VP. Het alléén maar goed zijn in het genereren van *leads* is niet voldoende om de concurrentie voor te blijven. VP heeft de ambitie om goed te zijn in een aantal disciplines waardoor VP meer waarde toevoegt aan haar afnemers.

Deze disciplines zijn:

1. **Techniek** - met name internet technologie

De websites van VP zijn modulair opgebouwd. VP kan hierdoor zéér snel een website in elkaar zetten waarin ook nog eens rekening wordt gehouden met de gegevensanalyse.

2. **Online marketing** - waaronder vormgeving en *search engine* positionering  
Dit garandeert VP een goed bezoekerspijl op haar websites.

3. **Proces- en projectorganisatie**

Afhankelijk van de wens van de afnemer, kan VP ook de processen van een project besturen. Behalve het *lead*-genereren neemt VP dan ook de navolging van

de *leads* voor verantwoording; bijvoorbeeld door het inschakelen van een callcenter.

#### 4. **Business Intelligence (BI)**

Bij bovenstaande punten 2 en 3 worden veel beslissingen genomen. Deze beslissingen kunnen worden ondersteund door BI.

BI kan twee soorten ondersteuning geven:

- **Feedback**

Is de genomen beslissing goed? Is de overeenkomst met de leverancier bijvoorbeeld een winstgevende overeenkomst?

- **Feedforward**

Welke beslissing is geschikt voor de gegeven situatie? Welk soort klanten kan bijvoorbeeld het best worden benaderd met een bepaalde propositie?

Een combinatie van deze disciplines maakt VP sterk. Zomaar een website maken en wat *leads* genereren, is op zichzelf niet zo moeilijk. Daarentegen is het maken van een website die véél *leads* genereert en het beleidsmatig de kwantiteit en kwaliteit van *leads* beheren een vak apart. De kwaliteit van een *lead* wordt uitgedrukt in de kans dat een *lead* een transactie wordt. Dit is tenslotte de opbrengst van de afnemers van VP. De websites leveren analytische data die zowel VP, als haar afnemers in staat stellen hun klanten beter te kunnen bedienen. Voor partners die zelf geen moeite willen doen kan VP het gehele online kanaal op zich nemen.

Momenteel werkt VP met partners die hypotheeken en levensverzekeringen verstrekken. De combinatie van de huidige disciplines van VP kan zij steeds hergebruiken met verschillende *content* of *businesspropositions*. VP kan hierdoor bij wijze van spreken ook *leads* genereren voor een energiebedrijf. Om de BI component te versterken, is door VP mijn afstudeerproject in het leven geroepen.

### **1.3. Doelstelling opdrachtgever**

Hoofddoel van dit afstudeerproject is de toegevoegde waarde van data analyse voor VP te verhogen met Business Intelligence (BI). Dit moet vanuit twee invalshoeken benaderd worden:

1. Operationele bedrijfsvoering  
dat wil zeggen: data analyse door en voor VP medewerkers
2. Diensten aan derden  
dat wil zeggen: door VP medewerkers voor hun zakelijke partners.

De opdrachtgever heeft waarde toevoeging aan het eigen bedrijf en de partners als hoofddoel. In Appendix B is het originele hoofddoel van mijn stage terug te vinden.

## 1.4. Doelstelling stage

De doelstellingen van de opdrachtgever (gebruik maken van alle drie de BI instrumenten; beschreven in Appendix B), waren te breed om binnen de stageperiode te verwezenlijken.

De doelstellingen van de stage zijn daarom verworden tot een onderdeel van de totale initiële doelstelling. De uiteindelijke doelstelling is afgebakend samen met mijn externe begeleider: professor Eiben.

De drie BI instrumenten zijn:

- *Data warehousing*
- OLAP
- *Data mining*

### **Data warehousing**

Implementatie van de juiste *data warehousing* was een voorwaarde voor het succes van mijn deelname binnen het BI project tijdens mijn stage. Mijn taak hierin was het meedenken met de opzet van het *data warehouse*. Het *data warehouse* is de basis voor de data analyses. Op basis van de OLAP theorie moest ik de onderliggende data opslag toetsen aan een aantal voorwaarden. Voornaamste criteria hierbij was of de opslag de toekomstige analyses ondersteund. De technische afdeling was verantwoordelijk voor de voortgang van dit project.

### **OLAP**

Na de afbakening met professor Eiben, werd OLAP de focus van mijn stage. Hiertoe werd besloten aangezien het stroomlijnen van de informatievoorziening met behulp van OLAP, op dat moment de meeste relevantie voor VP had.

Een aantal algemene informatiebehoeften moest worden herkend, waarna de ondersteuning van deze behoeften kon worden geautomatiseerd. Op den duur moesten de businessgebruikers in staat zijn de gewenste informatie te verkrijgen zonder tussenkomst van de techniek.

### **Data Mining**

In een situatie waarin de data infrastructuur nog niet is geregeld is het niet mogelijk resultaten te garanderen bij het gebruik van *data mining*. Aangezien de data infrastructuur binnen VP nog in de kinderschoenen stond, werd besloten *data mining* buiten beschouwing van de stage-opdracht te laten.

Met de afbakening waarbij de focus op OLAP werd gevestigd, werden de volgende concrete sub-doelen en succescriteria vastgesteld:

1. Een blauwdruk of draaiboek opstellen voor een OLAP procedure voor interne business users.

VP wil van het huidige ad-hoc (of eigenlijk nihile) OLAP gebruik groeien naar een standaard procedure. Deze moet leiden vanaf het formuleren van een business

vraagstuk, tot het leveren en bekijken van een bijbehorende OLAP kubus.

Let wel: het gaat hier om standaard procedures! Het gaat niet om standaard analyses uit de tabel.

2. Een aantal concrete OLAP kubussen aanmaken gestuurd door de interne business user (Marc Kerstel), waarbij ook het traject van businessvraagstuk tot OLAP kubus beschreven moet worden.

Assistentie bij concrete data analyse vragen. Deze oefeningen vormen *stepping stones* om doel 1 te realiseren.

3. Identificeren van een aantal standaard analyses die de interne business user herhaaldelijk wil uitvoeren en in werking stellen van de bijbehorende procedures om de kubussen soepel te kunnen leveren.

Hierdoor wordt het gebruik van VP's OLAP tool (en in het algemeen: data analyse) geïntegreerd in de kern van het businessproces.

4. Kwantificeren van de waarde van data en data analyse voor VP (intern gebruik).

Voor het succesvol integreren van data analyse activiteiten in de kern van het business proces is het belangrijk om de gebruikers en het management te overtuigen van de toegevoegde waarde hiervan.

5. Kwantificeren van de waarde van data en data analyse voor VP (extern gebruik).

VP wil op termijn data gerelateerde diensten aanbieden aan haar zakenpartners. Voor een goede propositie is een overtuigend verhaal nodig. Inzicht in de waarde van zulke activiteiten is ook nodig als basis voor het berekenen van de provisie.

6. Ontwerp en realisatie (met medewerking van techniek) van een data opslag en data stroom die bovenstaande doelen goed ondersteunen.

Het doel van de hele "data huishouding" is om goede management informatie te kunnen leveren. Tot nu toe was er echter weinig sturing van de business kant.

Met andere woorden, de integratie tussen business users en technici is niet optimaal.

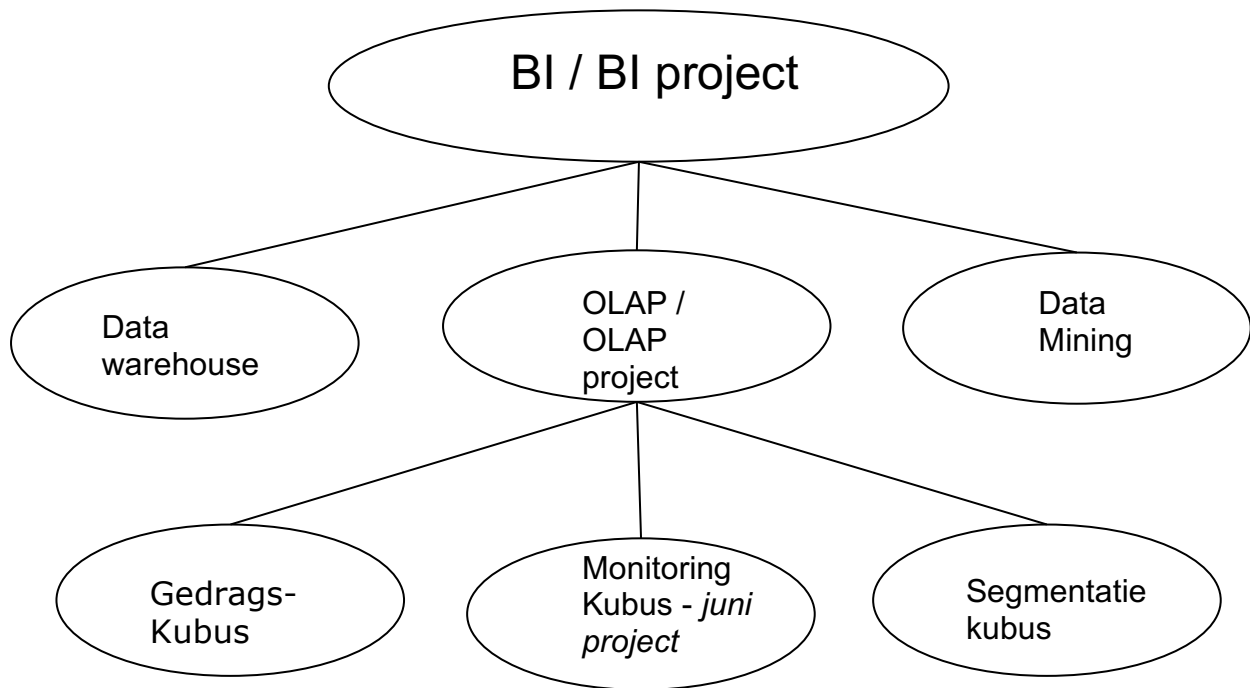
De bovenstaande doelen kunnen als volgt worden samengevat:

- Data huishouding helpen opzetten  
Afspraken maken met de techniek over de manier waarop de data wordt opgeslagen (doel 6).
- OLAP introduceren binnen VP  
Herkenning van de standaard informatiebehoefte en deze automatiseren met de OLAP-tool van VP (doel 3). Opleveren van een aantal "handmatige" OLAP datamodellen aan de business gebruikers (doel 2). Na het realiseren van het automatiseren van de analyses met de OLAP-tool kan een OLAP draaiboek worden opgesteld (doel 1). Dit draaiboek is een blauwdruk (lees houvast) voor VP om soortgelijke OLAP projecten te implementeren.
- Aantonen van de toegevoegde waarde van BI (doel 4 en 5).

De doelen zijn afhankelijk van elkaar; het resultaat van het ene doel kan de input zijn voor het andere. De tijdsplanning van het project is terug te vinden in Appendix C.

Met de in deze paragraaf geformuleerde doelstellingen, begon ik mijn stage. Ik werkte mee aan een volledige BI oplossing voor VP. Mijn stageproject, maakte onderdeel uit van dit BI project. De focus van mijn stage lag op de automatisering van de standaard informatiebehoeftes met behulp van OLAP (beschreven in hoofdstuk 6.2). Dit werd het OLAP project genoemd. Eén van de OLAP kubussen is in juni 2003 geïmplementeerd. Het project om deze kubus uit te werken werd dan ook het juni-project genoemd binnen VP.

Het volgende figuur illustreert de relaties tussen de projecten.



**figuur 1-2 Project relaties tijdens stage**

### **1.5. Hoofdstuk indeling van de scriptie**

Met hoofdstuk 2 tot en met 4 wordt een theoretisch kader geschapen van de OLAP concepten. Hoofdstuk 5 tot en met hoofdstuk 7 beschrijft het praktijk gedeelte van deze scriptie. Daarin wordt beschreven hoe de OLAP kubus binnen VP wordt geïmplementeerd. In hoofdstuk 5 wordt een momentopname van de beginsituatie gemaakt. Hoofdstuk 6 bespreekt het geleverde produkt en de tot stand koming van deze oplossing gekozen. Hoofdstuk 7 evalueert het project en geeft aan welke zaken VP binnen afzienbare tijd moet realiseren om het gebruik van OLAP ten volste te kunnen benutten.

In hoofdstuk 8 tot en met 10 zijn de aanbevelingen en de conclusies beschreven.

Hoofdstuk 8 bespreekt de knelpunten van VP met betrekking tot de BI projecten en geeft een aantal aanbevelingen voor de aanpak hiervan. Hoofdstuk 9 maakt vergelijkingen tussen de theorie en de praktijk. Uiteindelijk wordt in hoofdstuk 10 het resultaat van de stage samengevat en de centrale probleemstelling beantwoord.

## 2 Inleiding OLAP

In dit hoofdstuk wordt de theorie van OLAP toegelicht met de nadruk op de multi-dimensionale kijk die OLAP biedt op de data. Volgens de theorie valt OLAP onder **B**usiness **I**ntelligence (BI) en **D**ecision **S**upport **S**ystems (DSS).

Met behulp van voorbeelden zullen de verschillende basisconcepten van OLAP worden uitgelegd, zoals: *drill-down*, *roll-up* en *multi-* of *cross*-tabuleren.

Na het lezen van dit hoofdstuk dient de lezer de volgende vragen te kunnen beantwoorden:

- Wat is OLAP?
- Wanneer wordt het gebruik van OLAP interessant?
- Wat zijn de beperkingen van OLAP?

### 2.1. Wat is OLAP?

#### 2.1.1. Definitie

Wanneer iemand het over OLAP heeft, kan hij of zij het volgende bedoelen:

- OLAP concepten
- OLAP talen
- Een OLAP pakket

#### Een OLAP definitie:

*“OLAP is a category of business software that gives users access to analytical content such as time series and trend analysis views and summary-level information, as well as insight into data organized into multiple dimensions.”*

[OLAP Council]

Deze definitie is een verzameling van verschillende OLAP betekenissen. “*Category of business software*” slaat op de OLAP producten/pakketen en de OLAP talen. “*Data organized into multiple dimensions*” duidt de OLAP concepten aan.

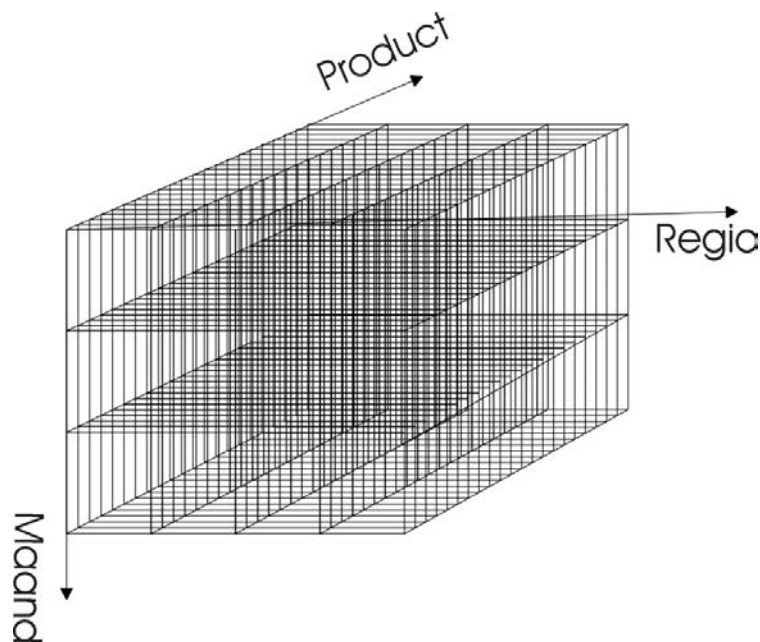
OLAP wordt veel toegepast in de praktijk. Er is een breed aanbod van softwarepakketten die de gebruikers helpen OLAP analyses te maken. Deze producten (OLAP talen en pakketen) kunnen onderling veel van elkaar verschillen; de OLAP concepten blijven echter constant. De OLAP concepten zorgen voor de multi-dimensionale representatie van de data. In de volgende paragraaf worden de OLAP concepten nader toegelicht aan de hand van een getallenvoorbeeld.

### 2.1.2. OLAP concepten

Het OLAP data model wordt vaak de OLAP kubus genoemd. Dit is te danken aan het feit dat de OLAP concepten in de literatuur vaak worden uitgelegd aan de hand van een kubus-metafoor. Aan de hand van de volgende voorbeelden wordt de multi-dimensionale denkwijze van OLAP toegelicht. Hierbij wordt gebruik gemaakt van de cijfers en rapporten van een fictief landbouw bedrijf dat met meerdere producten in meerdere regio's opereert.

Na het lezen van deze paragraaf moet de lezer de volgende multi-dimensionale operaties kunnen begrijpen:

- **Slice en dice**  
Overzicht creëren door het aantal dimensies te beperken
- **Multi- en cross-tabuleren**  
Weergeven van de interactie en dwarsverbanden tussen meerdere dimensies
- **Drill-down, roll-up**  
Verkennen in de hiërarchieën van de dimensies
- **Hoger dimensionale representatie (hoger dan 3 dimensies) in een twee dimensionale tabel.**



figuur 2-1 Weergave kubus metafoor

figuur 2-1 beeldt de kubus metafoor voor OLAP concepten uit. Deze kubus kan worden bekeken vanuit het perspectief dat er heel veel tabellen achter elkaar zijn geplaatst. Binnen de tabellen bevinden zich waarden als "omzetcijfers". Deze waarden worden vaak aangeduid met *measures* of *summaries*. De assen worden dimensies genoemd. Deze dimensies hebben ook waarden, zoals "melk" een waarde is van de dimensie "produkt".



De managers van het fictieve landbouw bedrijf willen graag de *measures* van:

- het aantal verkochte produkten
- de kosten van de verkochte produkten
- de opbrengsten in verschillende tijdsperiodes en regio's

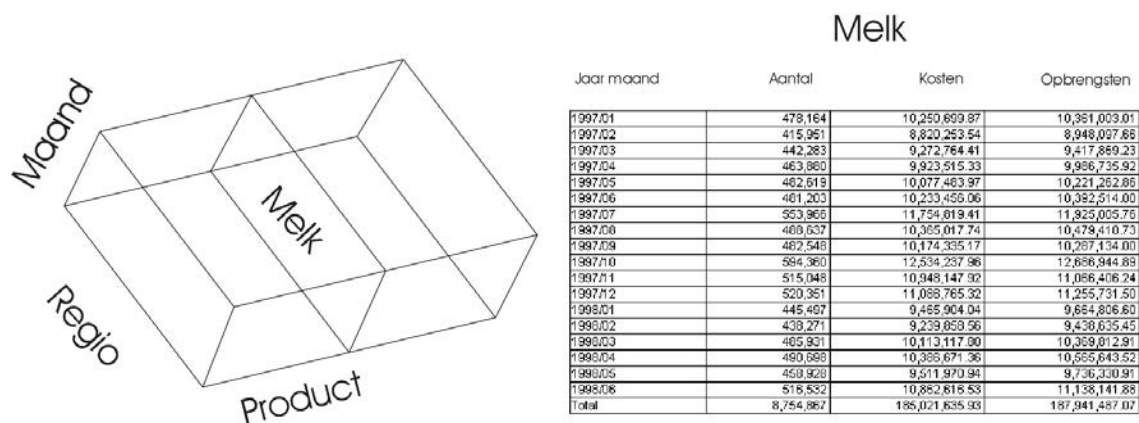
overzichtelijk representeren. Daarvoor maken ze gebruik van een OLAP kubus.

### **Slice en Dice**

Wanneer de managers het overzicht van de *measures* van het produkt "melk" willen vergelijken in verschillende "maanden" en "regio's", zullen ze bij "melk" de tabellen "maanden" en "regio's" moeten selecteren. De daarmee verkregen tabel is in feite een *slice* (lees doorsnede) van de 3 dimensionale kubus. Een *slice* is een weergave van *measures* via de invalshoek van een dimensie.

Een aantal voorbeelden over het opvragen van een bepaalde *slice* in de kubus, wordt hieronder gegeven.

figuur 2-2 beschrijft de *measures* van het produkt "melk", weergegeven in verschillende "maanden" (dimensie maand is de invalshoek).



**figuur 2-2 Voorbeeld slice**

figuur 2-3 is een weergave van een *slice* van de "produkt" dimensie, waarbij de overige dimensies variabel zijn gelaten.

Soort	Aantal	Kosten	Opbrengsten
Butter	5,378,559	114,761,980.49	115,767,097.83
Cheese	6,628,460	140,308,438.74	142,778,809.50
Cottage Cheese	670,272	13,586,713.43	14,483,879.00
Cream	2,009,097	42,371,432.38	43,567,580.60
Cream Cheese	474,122	9,769,791.76	10,294,551.24
Egg	1,656,342	34,520,173.07	35,853,083.67
Ice Cream	6,410,989	116,280,997.12	138,082,441.96
Milk	8,754,867	185,021,635.93	187,941,487.07
Yogurt	3,402,977	58,093,525.02	72,995,538.64
Total	35,385,685	714,714,687.94	761,764,469.51

figuur 2-3 Alle produkt soorten

### Multi- en cross-tabuleren

De *measures* vergelijken vanuit twee dimensies is ook mogelijk. Dit wordt *multi-tabuleren* genoemd. Een voorbeeld hiervan is figuur 2-4. Hierbij worden produkten uitgezet in maanden.

Soort	YearMo	Aantal	Kosten	Opbrengsten
Butter	1997/01	298,278	6,521,609.46	6,559,305.10
	1997/02	236,196	4,975,894.99	4,989,390.75
	1997/03	275,838	5,988,626.62	6,031,787.51
	1997/04	270,279	5,858,402.97	5,890,045.97
	1997/05	323,602	6,875,860.05	6,911,066.25
	1997/06	317,782	6,851,998.73	6,895,540.84
	1997/07	306,577	6,582,369.84	6,606,319.92
	1997/08	324,505	7,082,083.61	7,087,878.52
	1997/09	319,880	6,900,998.50	6,926,442.35
	1997/10	337,625	7,161,267.74	7,244,270.13
	1997/11	292,131	6,180,076.77	6,202,973.65
	1997/12	276,245	5,886,731.03	5,912,687.10
Cheese	1998/01	277,380	5,989,784.82	6,082,268.95
	1998/02	253,390	5,338,350.85	5,412,380.95
	1998/03	307,385	6,558,268.69	6,664,158.59
	1998/04	291,511	6,062,357.59	6,155,460.49
	1998/05	315,670	6,885,310.81	6,781,349.11
	1998/06	354,285	7,461,987.42	7,613,771.65
	1997/01	349,266	7,417,318.71	7,505,222.95
	1997/02	327,981	7,039,277.05	7,128,671.40
	1997/03	330,139	6,899,678.19	7,037,087.26
	1997/04	342,132	7,232,489.62	7,346,410.07
	1997/05	369,289	7,748,317.55	7,830,122.82
	1997/06	378,991	8,225,982.93	8,350,755.86
	1997/07	417,819	8,725,043.04	8,887,364.74

figuur 2-4 Produkt soort & maand

De volgende tabel is een voorbeeld van een tabel waarin via twee dimensies de *measures* kunnen worden gemonitored. Deze operatie wordt *cross-tabuleren* genoemd. De dimensie "maanden" is hierbij in de verticale kolom geplaatst en de dimensie "produkt soort" is horizontaal geplaatst.

Units	Butter	Cheese	Cottage Cheese	Cream	Cream Cheese	Egg	Ice Cream	Milk	Yogurt	Total
<b>Total</b>										
1997/01	298,278	349,266	34,980	106,761	27,064	94,620	365,131	478,164	183,416	1,937,680
1997/02	236,196	327,981	27,846	98,275	24,087	81,882	292,104	415,951	170,526	1,674,848
1997/03	275,838	330,139	36,590	101,613	19,462	82,647	312,084	442,283	175,453	1,776,109
1997/04	270,279	342,132	35,264	109,296	27,564	83,893	343,883	463,880	184,877	1,861,068
1997/05	323,602	369,289	35,183	106,164	26,501	90,989	356,248	482,619	184,284	1,974,879
1997/06	317,782	378,991	34,068	120,477	27,606	95,480	359,783	481,203	172,395	1,987,785
1997/07	306,577	417,819	51,702	128,028	28,991	99,242	406,337	553,966	212,102	2,204,764
1997/08	324,505	382,259	39,971	106,940	27,272	93,918	379,411	488,637	197,555	2,040,468
1997/09	319,880	385,211	42,599	117,236	24,231	94,726	390,312	482,548	207,633	2,064,376
1997/10	337,625	416,557	45,034	127,549	28,513	108,043	413,850	594,360	215,372	2,286,903
1997/11	292,131	415,858	39,684	118,533	27,474	99,208	368,949	515,048	203,617	2,080,502
1997/12	276,245	379,155	39,656	111,104	29,535	98,749	360,941	520,351	199,895	2,015,631
1998/01	277,380	330,280	31,748	102,086	26,744	89,557	342,876	445,497	172,738	1,818,906
1998/02	253,390	321,354	30,328	101,802	23,598	84,212	301,873	438,271	177,857	1,732,885
1998/03	307,385	372,530	36,023	108,270	22,875	87,392	340,481	485,931	186,976	1,947,863
1998/04	291,511	359,793	38,662	112,873	26,942	83,918	351,802	490,698	193,336	1,949,535
1998/05	315,670	358,605	33,252	105,619	24,575	88,394	346,176	458,928	179,270	1,910,489
1998/06	354,285	391,241	37,682	126,471	31,088	99,472	378,748	516,532	185,675	2,121,194
<b>Total</b>	5,378,559	6,628,460	670,272	2,009,097	474,122	1,656,342	6,410,989	8,754,867	3,402,977	35,385,685

figuur 2-5 Cross-tabel van maanden en produkten

Het verschil tussen een *multi-tabel* en een *cross-tabel* is dat met een *multi-tabel* in principe meer dan twee dimensies kunnen worden gerepresenteerd. Bij een *cross-tabel* is dit onmogelijk. De eindgebruiker heeft de vrijheid te kiezen welke tabel hij of zij meer geschikt vindt voor de te maken analyse.

Nu de kubus metafoor is besproken, is het duidelijk dat deze drie dimensionale kubus geprojecteerd kan worden in twee dimensies. Door *slices* binnen de kubus aan te roepen verkrijgen gebruikers de tabellen die zij wensen.

### Drill-down en Roll-up

Vaak hebben de dimensies onderlinge relaties met elkaar in de zogenaamde *dimensie hiërarchie*. Bijvoorbeeld "dagen", "maanden" en "jaren". De volgende basis-OLAP operaties zijn bedoeld voor het opereren in de dimensies met onderlinge hiërarchie.

#### - Drill-down

Dit doe je wanneer je op een tabel staat met "maanden" als een van de dimensies, waarbij je in een van de "maanden" het verloop van de *measures* op "dag" niveau wil volgen.

Het één niveau dieper gaan in de hiërarchie wordt een *drill-down* genoemd. In het bovenstaande voorbeeld, wordt zeggen zowel "maanden" als "dagen" iets over de dimensie "tijd". Beide doen dit echter op verschillende hiërarchische niveaus.

#### - Roll-up

Dit is precies het tegenovergestelde van een *drill-down*. Dit is bijvoorbeeld het geval wanneer een gebruiker die gebruik maakt van de dimensie "maanden", de gegevens op jaarbasis wil bekijken.

*Drill-down* en *roll-up* zijn verkapte vormen van *slice*. Wanneer in een maand een *slice* wordt genomen van de dimensie "datum" is dit een *drill-down*. In het geval dat de "maand" waarden variabel worden gelaten en een *slice* van de dimensie "jaar" wordt genomen, is sprake van een *roll-up*.

### **Hoger dimensionale representatie in een twee dimensionale tabel**

Een multi-dimensionale kubus bestaat in feite uit een verzameling tabellen die samengevoegd een multi-dimensionale *look and feel* geven. De kubus metafoor vergelijkt alleen een drie dimensionale kubus. De OLAP concepten kunnen worden gebruikt voor hogere dimensies. Het OLAP-pakket dat door VP wordt gebruikt kan bijvoorbeeld volgens de handleiding 128(!) dimensies aan. Door *slice-* of *multi-tabellen* te selecteren, kunnen de hogere dimensies in de geselecteerde tabellen worden geprojecteerd.

### **2.1.3. OLAP als onderdeel van BI**

In elke organisatie worden veel beslissingen genomen. Een goede beslissing is altijd de beste respons op een gegeven situatie. Voor een schaakspeler bestaat bijna nooit een absoluut goede zet. Het is daarom voor hem belangrijk een aantal te mogelijke zetten vooruit te denken als respons op de te verwachten zetten van de tegenstander. Bij het schaakspel is alle benodigde informatie voor de beide spelers op het schaakbord aanwezig. Van een geoefend schaakspeler mag worden aangenomen dat hij in staat is de situatie te overzien.

De meeste organisaties willen het effect weten van de genomen beslissing op de doelstelling van de organisatie; bijvoorbeeld winstmaximalisatie. Met deze informatie kan een organisatie leren in de toekomst betere zakelijke beslissingen te nemen. Kennis opgedaan met het evalueren van genomen beslissingen, verschaft belangrijke businessinzichten. In een organisatie komt het echter regelmatig voor dat de juiste informatie niet op het juiste tijdstip en de juiste plaats aanwezig is. Er is dan sprake van een zogenaamde informatie-asymmetrie. Dit maakt het moeilijk een uitspraak over de genomen beslissingen te doen en belemmert derhalve de vorming van zakelijke inzichten.

In het algemeen worden veel pogingen gedaan informatie-asymmetrie te minimaliseren door het inrichten van goede informatiesystemen. Met OLAP kan een gedeelte van de informatievoorziening worden gestroomlijnd teneinde toekomstige beslissingen beter te kunnen ondersteunen. Het is daarom niet verwonderlijk dat OLAP wordt gezien als een onderdeel van de beslissings ondersteunende systemen (DSS).

OLAP, *queries & reports*, *data mining* worden tezamen ook wel aangeduid met de term Business Intelligence(BI). BI op haar beurt is weer een onderdeel van DSS. De doelstelling van alle BI systemen is de organisatie te helpen bij het nemen van beslissingen door het analyseren van data.

## 2.2. Wanneer is gebruik van OLAP interessant?

De afgelopen jaren zijn de kosten voor data opslag drastisch lager geworden. Hierdoor werd het aantrekkelijker om grotere hoeveelheden data voor mogelijk toekomstig gebruik vast te leggen. De exponentiële toename van de te interpreteren data die hiervan het gevolg was, vormt voor veel organisaties een bottleneck.

Doordat veel data in genestelde tabellen is opgeslagen, wordt niet zelden het overzicht in de datahuishouding verloren. Daarbij blijft de hoeveelheid data alsmaar stijgen.

De belangrijkste toegevoegde waarde van OLAP, is de mogelijkheid data in een multi-dimensionale vorm te representeren. Dit schept overzicht over de data.

De data in een zodanige vorm neerzetten dat het geschikt is voor de multi-dimensionale analyse, is het proces van het omzetten van data naar informatie. In deze paragraaf worden de voordelen van OLAP verder uitgewerkt. Daarnaast worden vergelijkingen gemaakt tussen OLAP en veel gebruikte andere informatievoorzieningen.

### 2.2.1. Beoogde voordelen van het gebruik van OLAP

Een aantal criteria gaan algemeen op voor informatiesystemen. Gedacht moet worden aan juistheid, precisie en accuratesse. Een goed OLAP project voldoet aan deze algemene criteria en heeft daarmee de volgende voordelen ten opzichte van andere informatiesystemen:

- **Flexibiliteit van de rapporten**

De management rapportages kunnen worden gerepresenteerd in verschillende vormen; meestal een combinatie van de dimensies.

- **Vergroot IT-mandaat van de gebruikers**

De gebruikers kunnen vaak moeilijk communiceren met de techniek omdat hun denkwijze verschilt in abstractie niveau. In die zin kan OLAP als een communicatie wijze worden gebruikt. Gebruikers hoeven alleen aan te geven welke dimensies en *measures* ze willen zien en de techniek moet vervolgens in staat zijn dit te implementeren. De kans op mislukking van het project door communicatie stoornissen, wordt hiermee beperkt.

- **Mogelijkheid om veel voorkomende *queries* met een data kubus op te vangen**

Dit biedt een mogelijkheid om een groot gedeelte van de informatievraag te standaardiseren en te automatiseren.

- **OLAP dwingt de gebruikers in eerdere stadia van ontwikkeling al na te denken over hetgeen ze willen zien.**

Vaak zijn projecten voor aanvang, al gedoemd te mislukken doordat de eisen en wensen van de gebruikers op een te laat moment worden vastgesteld. In extreme gevallen worden ze helemaal niet vastgesteld. Het opbouwen van een OLAP data kubus is sterk afhankelijk van het formaat van de data opslag. Te laat aanleveren van de eisen en wensen heeft zéér grote gevolgen voor de bruikbaarheid van de onderliggende data.

- **Dimensies en *measures* sluiten beter aan op de businessgedachten**  
Voor de businessgebruiker hebben genomen beslissingen invloed op de cijfers. Wanneer hij of zij een bepaalde beslissing wil toetsen: bijvoorbeeld de kwaliteit van de leverancier van een deelstof, kan de naam van deze leverancier als een dimensie worden gebruikt. De *measure* is bijvoorbeeld percentage "defect eind produkt".

### 2.2.2. OLAP versus Ad-hoc query

Een definitie van *queries & reports* volgens [Alter - 2001]:

*"Query (language): Special-purpose computer language used to provide immediate, online answers to user questions.*

*Report (generator): Program that makes it comparatively easy for users or programmers to generate reports by describing specific report components and features."*

Het verschil tussen OLAP en ad-hoc *queries* ligt in de flexibiliteit. Bij OLAP wordt de data al gestructureerd in de data kubus. Alle informatievragen over de dimensies en de *measures* van de kubus kunnen in feite worden gerepresenteerd met de besproken basisoperaties van de multi-dimensionale concepten. In feite kan met een OLAP kubus een groot aantal informatievragen worden beantwoord. Daarentegen beantwoordt een ad-hoc *query* meestal maar één vraag.

Hierdoor kan de vraag worden gesteld of ad-hoc *queries* nog zinnig zijn, wanneer met OLAP-datamodellen de informatievragen veel flexibeler kunnen worden beantwoord.

Waarom kan OLAP niet de ad-hoc queries volledig vervangen?:

- **Ontwikkelkosten om data in de kubus te stoppen zijn hoog**  
Een kubus moet worden gemoduleerd en geïmplementeerd. Als een *query* maar eens per maand nodig is, wegen de ontwikkelkosten mogelijk niet tegen de baten op.
- **OLAP geeft een geaggregeerde view op de data**  
Vragen over detail informatie moeten toch met een ad-hoc *query* worden beantwoord. Hier wordt in hoofdstuk 3 dieper op ingegaan.

Als vuistregel kan worden gesteld dat veel voorkomende informatievragen, met een OLAP-datamodel kunnen worden opgevangen. Minder voorkomende informatievragen kunnen beter met behulp van ad-hoc *queries* worden beantwoord.

### 2.2.3. OLAP versus Spreadsheets

Zoals het voorbeeld van paragraaf 2.1.2 zichtbaar heeft gemaakt, representeert de OLAP tool de data in de vorm van tabellen. Dit heeft twee praktische redenen:

- De multi-dimensionale kubus wordt nu eenmaal gepresenteerd op het platte beeldscherm of op papier.
- Gebruik van het spreadsheet is goed ingeburgerd in de wereld van de analisten. Gebruik maken van de tabellen verlaagt de drempel om naar multi-dimensionale analyse over te stappen.

Deze spreadsheet tabellen zijn overgenomen in OLAP, die daarmee vervolgens multi-dimensionale concepten representeert. Een spreadsheet kan ook de multi-dimensionale concepten van OLAP overnemen; neem als voorbeeld de *pivot table* functie van MS-Excel. Dit is in feite een multi-dimensionele tabel. Om die reden zou MS-Excel ook als OLAP produkt kunnen worden beschouwd. De pivot table functie is alleen nog niet geïntegreerd binnen overige spreadsheet functionaliteiten. Nog altijd moeten de gegevens van het worksheet worden geïmporteerd naar een pivot table.

Kan een spreadsheet de OLAP analyse vervangen?

Om deze vraag te kunnen beantwoorden heeft Eric Thomsen en zijn team [Eric Thomsen, 2002] een test gedaan. Het kostte een ervaren macro programmeur acht uur om een multi-dimensionale *look and feel* te programmeren in een spreadsheet (getest met Lotus 1-2-3, Quattro Pro en Excel). Een OLAP pakket kan binnen enkele minuten een kubus construeren. Het afhandelen van de dimensies is voor een OLAP-pakket slechts het optellen van waarden. In hoofdstuk 3 wordt de afhandeling van data binnen OLAP besproken. In tegenstelling tot OLAP, vereist een spreadsheet bij iedere nieuwe dimensie of dwarsverband van dimensies, een nieuwe macro.

OLAP pakketten zijn primair ontwikkeld (gespecialiseerd) voor representatie van multi-dimensionale data. Daarentegen is een spreadsheet meer generiek ontworpen voor het uitvoeren van (complexe) berekeningen. Het is mogelijk om met behulp van een spreadsheet multi-dimensionale data te respresenteren. Zonder OLAP-achtige *add-ins* is het gebruik hiervan erg arbeidsintensief.

#### **2.2.4. OLAP versus data mining**

*Data mining* is een ander instrument uit de BI "toolkit". Een definitie van *data mining* is:

*"The process of extracting valid, previously unknown, comprehensible and actionable information from large databases, and using it to make crucial business decisions."*

[TCCB - 2001]

*Data mining* heeft de volgende karakteristieken:

- Het zoekt verbanden en patronen
- Het opereert daarbij in grote hoeveelheden data
- Het ondersteund daarmee het maken van beslissingen door de toekomst te "voorspellen".

OLAP en *data mining* hebben twee verschillen. Dit zijn:

- **Niveau van aggregatie**

Zoals in de voorbeeld kubus van paragraaf 2.1.2 te zien was, zijn de *measures* als het aantal verkochte produkten, opgeteld. OLAP maakt een geaggregeerde vorm van de data, zodat deze makkelijk te interpreteren is voor de managers. Bij Data Mining moet vaak in niet geaggregeerde detail data, verbanden worden gezocht. Daarnaast zijn de verfijnde structuren in dimensies, *slices* en *cross* dimensies van OLAP niet terug te vinden in Data Mining pakketten.

- **Soort beslissingen dat wordt ondersteund**

De beslissingen die worden ondersteund zijn bij OLAP **beschrijvend** en bij data mining **verklarend**. In figuur 2-6 worden de verschillende beslissingsfuncties weergegeven. OLAP ondersteunt met name de beschrijvende (*descriptive*) beslissingen. Data Mining ondersteunt met name patroon herkenning en verklarende (*explanatory*) beslissingen zoals *regressie* en *clustering*. Met andere woorden: OLAP legt situaties vast. Data Mining zoekt verbanden en probeert te verklaren waarom situaties tot stand zijn gekomen.

<b>Descriptive</b>	<b>Explanatory</b>	<b>Predictive</b>	<b>Decision making</b>
Aggregation	Regressions	Pattern extension	Goal seeking
Allocation	Clustering	Model	Simulation
Ratios	Association	Management	
Scaling	Split finding	Model result	
Weighting	Probalistic	Merging	
Type transforms	Networks		

figuur 2-6 Decision functions [Erik Thomsen, 2002]

### 2.3. Wat zijn de beperkingen van OLAP?

OLAP heeft een aantal beperkingen waarmee tijdens de ontwikkeling van OLAP kubussen rekening moet worden gehouden. De oplossing ligt vaak in het gebruik van andere informatievoorzieningen naast het gebruik van OLAP. In deze paragraaf wordt een inventarisatie van deze beperkingen behandeld.

In tegenstelling tot *data mining* kan OLAP analyse moeilijk worden toegepast voor het verklaren van problemen. Dit is in de vorige paragraaf aan bod gekomen. OLAP wordt gebruikt door de eindgebruiker zelf. De meeste eindgebruikers kunnen echter niet denken in hogere dimensies. Als van meer dan vijf dimensies het effect in een *multi*-tabel beschreven dient te worden, is dat natuurlijk met OLAP mogelijk. Het is echter **niet** overzichtelijk zodat een gebruiker vrijwel niet in staat zal zijn deze data te interpreteren.



*Data mining* daarentegen is in staat de hogere dimensies (hoge aantal attributen) te berekenen; dit kan een eindgebruiker zelf niet.

Als bij de *requirement* fase al wordt aangegeven dat de onderliggende verbanden moeten worden verklaard, is OLAP waarschijnlijk niet geschikt voor de analyse.

Bij aggregatie gaat altijd informatie verloren. Via de dimensie "dag" kan als voorbeeld bijvoorbeeld de *measure* "omzet" bekeken worden. Wanneer het laagste aggregatieniveau van de kubus "dag" is, is het onmogelijk de "omzet" via "uren" te zien. Stel dat op bepaalde dagdelen meer omzet wordt gemaakt en dat deze informatie nodig is om een beslissing te kunnen nemen met betrekking tot het uitbreiden van het personeel. Deze informatie is echt verloren gegaan wanneer de onderliggende data registratie niet op het juiste aggregatieniveau is ingericht. Het is daarom van belang dat de juiste aggregatieniveaus vooraf gedefiniëerd worden.

## **2.4. Samenvatting**

In dit hoofdstuk is de kubus metafoor voor OLAP concepten besproken. De lezers weten nu wat het multi-dimensionale concept is. In paragraaf 2.2 werden de voordelen van OLAP besproken en werd OLAP vergeleken met een aantal andere informatievoorzieningen. OLAP is bedoeld voor organisaties die op een overzichtelijke wijze hun groeiende hoeveelheid opgeslagen data, willen kunnen interpreteren. In paragraaf 2.3 werden een tweetal belangrijke beperkingen van OLAP besproken. In een vroeg stadium van de ontwikkelfase van een informatievoorziening moet men rekening houden met deze beperkingen.

### 3 Opbouw van een OLAP kubus

Dit hoofdstuk beschrijft hoe een OLAP kubus wordt opgebouwd. Volgens de literatuur werkt OLAP het best samen met een *data warehouse*.

*“To facilitate OLAP, it is useful to work with the data warehouse (...) and with a set of OLAP tools.”*

[ET&JEA - 2001, pagina 147]

In paragraaf 3.1 worden zowel de kenmerken als de karakteristieken van het *data warehouse* beschreven. In paragraaf 3.2 wordt een getallen voorbeeld gegeven waarmee de opbouw van een OLAP kubus wordt verduidelijkt. Het is óók mogelijk een kubus te construeren zonder data warehouse. Tijdens mijn stage zijn handmatig een aantal kubussen geconstrueerd zonder ondersteuning van een *data warehouse*. Voorwaarde hiervoor is dat de onderliggende data aan bepaalde eisen voldoet. Het zal de lezer verderop in dit hoofdstuk duidelijk worden waarom een data warehouse OLAP aanvult.

#### 3.1. Data warehouse

In de literatuur zijn veel definities terug te vinden van een *data warehouse*. Vaak slaken aanbieders van *data warehouse* oplossingen populaire kreten als: “centrale data opslag”, “subject oriëntatie”, “non-volatiel”, “denormalisatie”, “opslaan van historische waarde” en “grote hoeveelheden data”.

Deze termen zijn vaak slechts een deelkarakteristiek van een *data warehouse*.

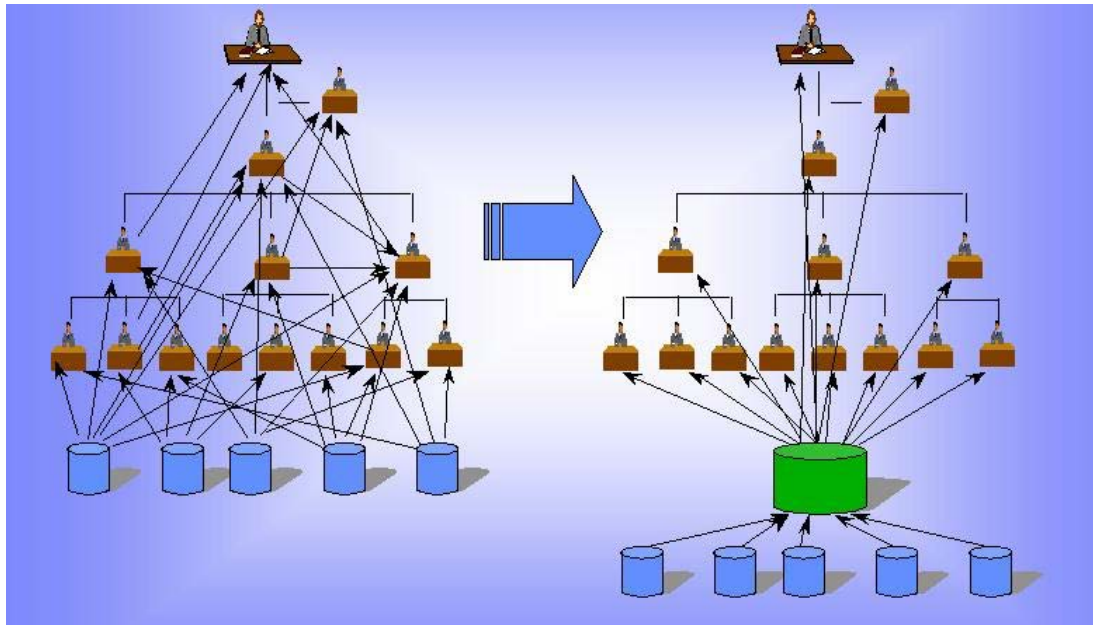
De volgende definitie van *data warehousing* drukt de essentie van een *data warehouse* beter uit:

*“A copy of transaction data specifically structured for query and analysis.”*

[Kimball, 2002]

De magie van *data warehousing* ligt in het omzetten van versplinterde data in een gecentraliseerde structuur. In figuur 3-1 is dit voordeel grafisch weergegeven. Het grote centrale vat in de rechter figuur, geeft het zogenaamde *data warehouse* aan. De kleine vaatjes in de figuur verwijzen naar de brondata. Deze zijn meestal afkomstig van operationele of externe databases.

De pijlen van de brondata naar het *data warehouse* worden vaak ETL's genoemd. ETL staat voor **E**xtracting, **T**ransforming en **L**oading. Dit wil zeggen dat de operationele data wordt ge-extraheerd, getransformeerd en wordt geladen in het *data warehouse*. Bij ETL wordt de data vaak opgeschoond. Daarom is de essentie van een *data warehouse* dat een kopie van de transactie data wordt gemaakt, waarna het wordt gestructureerd zodat de benodigde analyses kunnen worden uitgevoerd.



figuur 3-1 Data warehouse architectuur [Furstenberg]

De karakteristieken van het data warehouse kunnen het best worden uitgelegd door ze te vergelijken met de eigenschappen van een operationele database ofwel OLTP (**O**n-**L**ine **T**ransaction **P**rocessing).

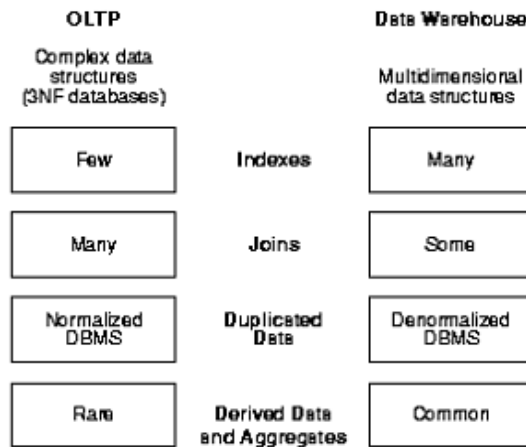
Operational Data Store	Data warehouse
Subject-oriented data	Subject-oriented data
Integrated data	Integrated data
Voilatile data	Nonvoilatile data
Only current data	Current and historical data
Short time between data refreshing	Longer time between data refreshing
Detailed data only	Detailed and summary data
Used for shour-term decisions	Used for planning purpose

figuur 3-2 Data warehouse algemeen

[TT&JEA - 2001, pagina 45]

Kijkend naar de karakteristieken van het *data warehouse*, is het nu duidelijk waar de gebruikte termen van de aanbieders vandaan komen. Hoewel ze gedeeltelijk gelijk hebben, zijn deze termen in feite karakteristieken van het *data warehouse* en niet zozeer de definitie hiervan.

Het gebruik van populaire kreten doet het echter beter richting de potentiële klant dan het simpelweg aangeven dat *data warehousing* de klant een kopie van de transactie data biedt.



figuur 3-3 Data warehouse op data verwerkingsniveau

[Trutek - 2003]

### 3.2. Werkwijze van de OLAP-tool

In deze paragraaf wordt een getallenvoorbeeld gegeven dat duidelijk maakt hoe een kubus kan worden geconstrueerd.

De onderstaande voorbeeld tabel geeft de brondata weer van de OLAP kubus uit hoofdstuk 2. Stel dat de onderliggende data van de voorbeeld kubus wordt opgeslagen in een database en dat bij elke transactie de onderstaande tabel wordt bijgewerkt. Stel tevens dat een verkoper in een winkel in Amsterdam 5 pakken "melk" verkoopt tegen een verkoopprijs van 10 cent per pak. Dit levert in totaal 50 cent op. Deze 5 pakken "melk" hebben in totaal 20 cent gekost. Deze gebeurtenis wordt genoteerd in de eerste regel van figuur 3-4.

Wat een OLAP-tool doet met deze onderliggende data is niets anders dan optellen. Stel dat je de verkochte aantallen "melk" wilt weten. Het OLAP pakket zoekt daarvoor in de dimensie "produkt soort" op, waar "melk" staat en telt de *measure* "aantal" bij elkaar op. Hetzelfde gaat op voor de "melk" opbrengst van "2001/06". Het OLAP pakket zoekt automatisch de waarden op waarbij de produkt soort "melk" is en "Jaar maand", "2001/06" is. De uitkomst die het OLAP pakket geeft is de som van de *measure* "opbrengsten" van alle gevonden records.

Produkt soort	Jaar maand	Regio	Aantal	Kosten	Opbrengsten
Melk	2001/06	Amsterdam	5	20	50
Melk	2001/06	Amsterdam	8	30	70
Melk	2001/06	Utrecht	4	16	40
Melk	2001/06	Amsterdam	6	24	50
Kaas	2001/07	Rotterdam	4	10	50
Kaas	2002/03	Amsterdam	5	20	60
...	...	...	...	...	...

figuur 3-4 Voorbeeld brondata voorbeeld kubus (hoofdstuk 2)

Een *slice* creëren is ook niets anders dan optellen. Wanneer een *slice* wordt gemaakt van alle waarden in de dimensie "produkt soort", zoekt het OLAP pakket alle unieke waarden van deze dimensie op. Vervolgens worden de waarden van de *measures* bij elkaar opgeteld. figuur 3-5 is een *slice* van alleen de dimensie "produkt soort" uit figuur 3-4.

Produkt soort	Aantal	Kosten	Opbrengsten
Melk	23	90	210
Kaas	15	30	110
...	...	...	...

figuur 3-5 *Slice* in dimensie produkt van tabel figuur 3-4

### 3.3. *Eisen van de onderliggende data voor OLAP*

In de vorige paragraaf is de werkwijze van een OLAP-tool besproken. De *measures* worden opgeteld wanneer de waarden van de dimensie, of een combinatie van een dimensie, uniek is. De optelsom kan gezien worden als een samengevatte *measure*-waarde van de dimensie. Vandaar dat een *measure* binnen sommige OLAP-pakketen een *summary* wordt genoemd. Gedenormaliseerde data is in feite al multi-dimensionaal. Gedenormaliseerde data kan worden gecreëerd door middel van het met elkaar *joinen* van meerdere genormaliseerde tabellen. *Joinen* levert echter in de praktijk vaak problemen op.

Stel dat Amsterdam en Rotterdam verschillende databases zijn. Voordat een OLAP analyse kan worden uitgevoerd, zal de data eerst met een *tool* moeten worden gekoppeld. Wanneer vervolgens de databases van Amsterdam en Rotterdam verschillende formaten hebben, ontstaan er problemen. Het attribuut "kaas" in de database van Amsterdam kan bijvoorbeeld in Rotterdam zijn opgesplitst in twee attributen: "jonge kaas" en "oude kaas". Het alleen oppakken of *joinen* van de data dump is dan ook vaak niet voldoende. Aangezien de OLAP-tool alleen de unieke waarden kan optellen, krijg je in de dimensie "produkt soort" verkeerde waarden als niet vooraf de data wordt bewerkt.

Een *data warehouse* wordt gecreëerd om data centraal op te slaan, ETL zorgt voor een gedenormaliseerde opslag en het opschonen van de data. De OLAP multi-dimensionale structuur komt overeen met de onderliggende structuur van het *data warehouse*, vandaar dat OLAP ook vaak een geaggregeerd *data warehouse* wordt genoemd. Vanuit het *data warehouse* kunnen namelijk meerdere geaggregeerde OLAP kubussen worden verkregen.

In de literatuur wordt geregeld een situatie geschetst waarin het *data warehouse* de onderliggende data voor de OLAP analyses verzorgt. Dit komt doordat het schonen en integreren van data op elementair niveau van records moet gebeuren.

We illustreren de uitdaging van integratie van data aan de hand van een voorbeeld waarin Amsterdam en Rotterdam verschillende attributen "kaas" hanteren.

Stel dat er geen *data warehouse* is en dat we toch de gegevens van Rotterdam willen integreren met die van Amsterdam. In dat geval kan een script worden geschreven dat de volgende stappen uitvoert: vervang de "jonge kaas" en de "oude kaas" door "kaas", en plak vervolgens de gegevens aan de database van Amsterdam.

Hierdoor zijn we één aggregatieniveau kwijtgeraakt. Het lukt namelijk niet meer met deze kubus om de verkoopcijfers van "jonge kaas" en "oude kaas" te vergelijken. Om een aggregatieniveau dieper te gaan, zal derhalve met ad-hoc *queries* moeten worden gewerkt in de transactie database van Rotterdam.

Wanneer dit probleem met behulp van een data warehouse wordt aangepakt, kan de ETL van het warehouse een allesomvattende tabel creëren. Deze allesomvattende tabel voorziet OLAP van brondata zonder informatie kwijt te raken.

Produkt soort	Jaar maand	Regio	Produkt detail	Andere variabelen	Measures ...
Kaas	2001/06	Amsterdam		...	...
Kaas	2001/06	Rotterdam	Jonge kaas	...	...
Kaas	2001/06	Rotterdam	Oude kaas	...	...
...	...	...	...	...	...

**figuur 3-6 Allesomvattende tabel voorbeeld**

Via de dimensie "produkt detail" kan onderscheid worden gemaakt tussen "jonge kaas" en "oude kaas" in Rotterdam. Via de "Produkt soort" kunnen op een hoger aggregatie niveau de *measures* van bijvoorbeeld "kaas" worden bekeken. De tussenkomst van deze allesomvattende tabel maakt het mogelijk op een later tijdstip te aggregeren. Hierdoor treedt een minimaal informatie verlies op wanneer de twee transactie databases worden gecombineerd.

### **3.4. Samenvatting**

In dit hoofdstuk zijn de verschillende technische aspecten van OLAP ten opzichte van de onderliggende data behandeld, zoals: de afhandeling en de eisen van OLAP ten opzichte van de onderliggende data. OLAP eist een juist kwaliteits- en aggregatieniveau van de onderliggende data. Teneinde de nodige analyses uit te kunnen voeren, zijn vaak gegevens nodig uit meerdere transactie databases. De karakteristieken van het data warehouse zijn besproken in paragraaf 3.1. Deze karakteristieken matchen met de eisen van OLAP ten opzichte van de onderliggende data opslag. Om deze reden wordt OLAP in de literatuur vaak beschreven naast een *data warehouse*.

## 4 Logische opzet van OLAP datamodellen

In het vorige hoofdstuk zijn de technische aspecten van OLAP behandeld. Het is een algemeen bekend gegeven dat een OLAP-kubus zéér makkelijk in gebruik is. Het moduleren van een OLAP datamodel is echter niet simpel. In dit hoofdstuk worden de logische aspecten van de datamodellen besproken.

Een aantal criteria spelen een rol bij het kiezen van het juiste datamodel:

- Uitbreidbaarheid van het een model
- Het vraagstuk kleine kubussen versus hyperkubussen
- Keuze van aggregatieniveau

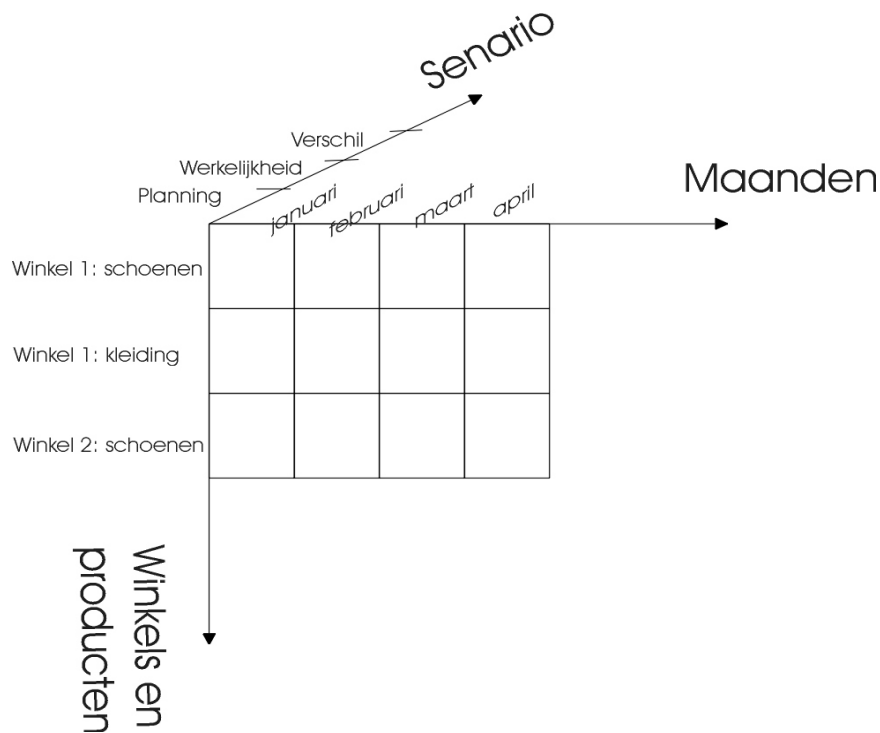
Deze punten zullen worden besproken aan de hand van een voorbeeld.

### 4.1. *Wanneer kan een model worden uitgebreid?*

Het komt regelmatig voor dat een bestaande kubus niet aan een extra informatievraag kan voldoen. In die gevallen wordt de oplossing vaak gezocht in de uitbreiding van de kubus. Een uitbreiding van een datakubus is het toevoegen van een dimensie of een *measure* in het bestaande datamodel. In deze paragraaf bekijken we wanneer een datamodel uitbreidbaar is en wanneer dit niet het geval is.

Stel dat er een aantal winkels zijn die “schoenen” en “kleding” verkopen. De winkels hebben hun omzet in “maanden” en “produkten” uiteengezet. In feite is dit een OLAP kubus met drie dimensies en één *measure*. De dimensies zijn: “winkel”, “produkt” en “maand”. De *measure* is omzet. Voor al deze winkels is aan het begin van dit kwartaal een planning gemaakt.

De managers willen weten hoe groot het verschil is tussen het werkelijke scenario en het geplande scenario. Dit heeft als gevolg dat de dimensie “scenario” als extra dimensie in de OLAP kubus moet worden toegevoegd. De bestaande OLAP-kubus wordt daarmee succesvol uitgebreid van drie naar vier dimensies. Dit voorbeeld wordt grafisch weergegeven in figuur 4-1.



**figuur 4-1 Juiste uitbreiding bestaande kubus**

Niet in alle gevallen verloopt het toevoegen van extra dimensies of *measures* even gemakkelijk. Stel dat de manager van de winkels het aantal "uren" dat een bepaalde werknemer heeft gewerkt óók wil toevoegen aan het bestaande model. De manager verwacht dat de ratio "omzet per gewerkt uur van een werknemer", hem een indicatie kan geven over de efficiëntie van een winkel. De werknemers werken voor verschillende winkels.

Kunnen de gewerkte uren van een bepaalde werknemer worden toegevoegd aan het bestaande model?

Winkelnaam	Werknemersnaam	Aantal uren per week	Week nummer
Winkel 1	Piet	36	Week 1
Winkel 1	Piet	20	Week 2
Winkel 1	Piet	36	Week 3
Winkel 1	Klaas	20	Week 1
Winkel 1	Klaas	20	Week 2
Winkel 2	...	...	...

**figuur 4-2 Gewerkte uren**



figuur 4-2 geeft het aantal gewerkte uren van de werknemers aan waaronder de gewerkte uren van Piet en Klaas. Piet heeft een 36-urige werkweek en Klaas heeft een 20-urige werkweek.

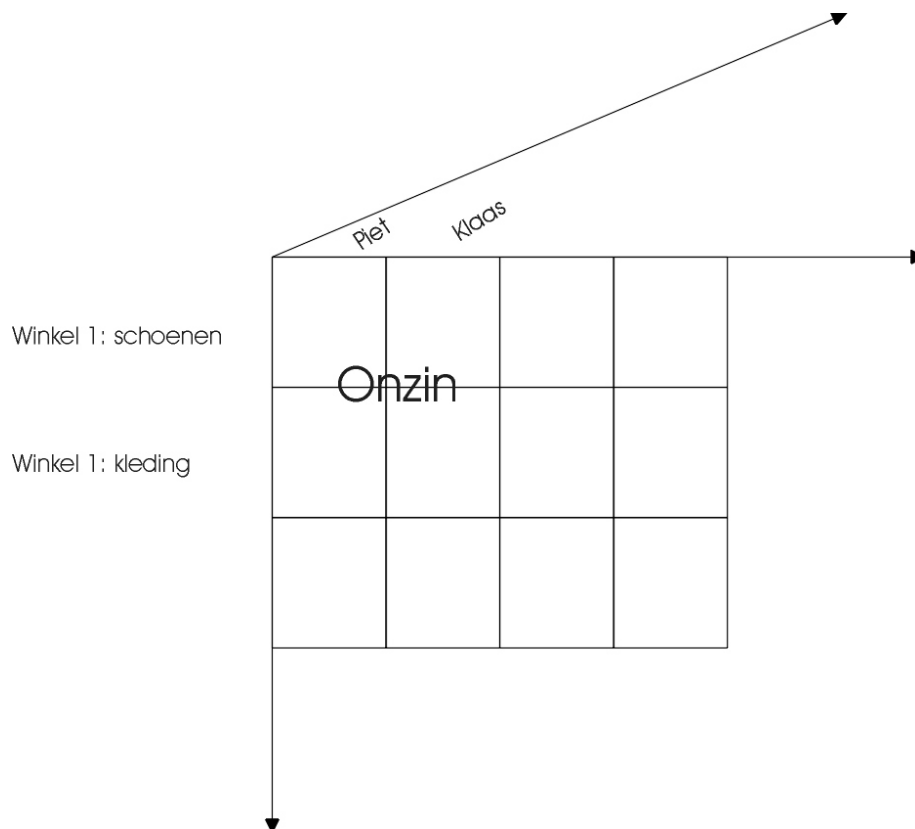
Het model bestaat uit de dimensies: "maanden", "produkt" en "winkels" en heeft de *measure* "omzet". We willen nu de uren van de werknemers toevoegen aan de kubus. Hierbij komen we twee problemen tegen.

- **De gewerkte uren zijn per week week geregistreerd**

Het is moeilijk om vanuit de weeknummers de maanden te bepalen. Soms bestaat een maand uit vier weken en soms uit vijf. Via de dimensie maand kun je helemaal geen gewerkte uren zien.

- **De toevoeging heeft niet hetzelfde dimensiestructuur als die van de bestaande kubus.**

Dit probleem wordt aan de hand van het volgende figuur uitgelegd. We nemen daarbij aan dat het probleem van punt 1 is opgelost door middel van registratie in zowel "maand" en "week".



**figuur 4-3 Onjuiste uitbreiding bestaande kubus**

Zoals figuur 4-3 weergeeft, kunnen we hiermee geen kubus construeren. De reden is dat het niet bekend is hoeveel tijd Piet besteedt aan de verkoop van "schoenen" ten opzichte van de tijd die hij besteedt aan de verkoop van "kleding". Daarom is het aantal uren via de dimensie produkt eigenlijk niet in te vullen.

Waarom kan een model niet altijd probleemloos worden uitgebreid?

Het antwoord op het **eerste voorbeeld**, waarbij de dimensie "scenario" werd toegevoegd, laat zien dat de geplande omzet dezelfde dimensiestructuur heeft als de omzet (de bestaande *measure*). In een dergelijk geval kan het datamodel zonder probleem worden uitgebreid.

Bij het **tweede voorbeeld** is de hoeveelheid gewerkte uren alleen via de dimensie "tijd", "winkel" en "werknemer" te zien, maar niet via "produkt" soort. Het gewerkte aantal uren heeft in dit geval een andere dimensiestructuur, waardoor de kubus niet zinvol wordt.

Als een uitbereiding wordt geforceerd van een dimensie of *measure* op een niet vergelijkbare dimensiestructuur, dan ontstaat een hyperkubus. In de volgende paragraaf wordt dit uitgelegd.

## 4.2. Semantische kubus versus hyperkubus

Een hyperkubus wil zeggen dat de dimensiestructuur van een bepaalde *measure* anders is dan de dimensiestructuur van andere *measures*.

OLAP kan alleen aan informatie vragen voldoen als de benodigde informatie in de dimensies en measures is opgenomen.

De keuze tussen een hyperkubus en een kubus in de semantische ruimte, hangt niet af van het modelleren van de *businessrequirements*. Deze hangt af van het onderhoud en de informatiedichtheid van de kubus. Zie figuur 4-4.

Kubussen in semantische ruimte	Hyperkubus
Makkelijker met onderhoud: wijzigingen in de dimensiestructuren hebben invloed op alle <i>measures</i> .	Moeilijker onderhoud: wijzigingen in de dimensiestructuren hebben alleen invloed op een gedeelte van de <i>measure</i> , er moet eerst geïnventariseerd worden wat de effecten zijn van de wijzigingen.
Subjectgeoriënteerd: een kubus beantwoordt de vragen over een van de business deelvragen of deelprocessen. Hierdoor zijn de <i>slices</i> overzichtelijk en geven ze meer informatie.	Niet subjectgeoriënteerd: een kubus beantwoordt alle vragen, de vragen zijn niet gerangschikt. Hierdoor zijn de <i>slices</i> minder overzichtelijk, maar geven ze nog wel informatie. Vaak moet er eerst nog een interpretatie plaatsvinden.

**figuur 4-4 Kubussen in semantisch ruimte versus hyperkubus**

Hierdoor ligt het voor de hand dat de mogelijke OLAP oplossingen in de semantisch ruimte prevaleren boven oplossingen met een hyperkubus.

Binnen VP worden moduleringstechnieken toegepast zodat de kans op een hyperkubus

minimaal is. In hoofdstuk 6 worden deze moduleringstechnieken besproken. Het is de bedoeling, dat voor elke groep van deze informatievragen een aparte OLAP-kubus wordt geconstrueerd. Hierdoor zijn de kubussen subject-georiënteerd. In paragraaf 6.3.1 wordt een van deze technieken besproken: het mini-ster schema. Dit schema zorgt ervoor dat de dimensiestructuren hetzelfde blijven voor de *measures*.

#### **4.3. Keuze van aggregatieniveau**

Een ander belangrijk punt van de dimensionele analyse, betreft het detail-niveau van de data representatie. Het is mogelijk met OLAP van een gedetailleerd niveau (lage aggregatie) naar een minder gedetailleerd niveau (hoge aggregatie) te gaan, maar niet andersom. Daarom moet tijdens de *user-requirement* fase van de ontwikkeling al worden uitgezocht op welk detail-niveau de informatie gerepresenteerd dient te worden. Een manier om het aggregatieniveau te kunnen bepalen via communicatie met de gebruikers wordt weergegeven in figuur 6-2.

#### **4.4. Samenvatting**

In dit hoofdstuk zijn de logische aspecten van het OLAP-datamodel besproken.

- De keuze tussen een hyperkubus en meerdere kleine kubussen in de semantische ruimte
- De uitbreidbaarheid van een kubus
- De keuze van aggregatieniveau

Deze keuzes zijn van essentieel belang wanneer een gebruiker OLAP analyses wil uitvoeren.

## 5 Situatie analyse aan het begin van het OLAP project

Een goede situatie analyse is de helft van de oplossing! VP is een klein dynamisch bedrijf met een vlakke organisatiestructuur. Het organiseren en uitvoeren van de projecten wordt gedaan zonder enige bureaucratische regelgeving. Het aantal gelijktijdige projecten en pilots die worden georganiseerd is groter dan het aantal werknemers dat binnen VP rondloopt. Op elk moment kan een nieuw project worden gestart. Een eenduidig definitie van een project ontbreekt binnen VP. Op dit moment wordt een project bepaald door de volgende factoren:

- **De partij met wie VP samenwerkt**  
Het genereren van hypotheek*leads* voor een geldverstrekker, hierna wordt deze geldvertrekker VP genoemd.  
wordt gezien als een project.
- **De content of propositie die VP te bieden heeft**  
Lijfrente wordt als een andere project gezien dan hypotheek  
of belastingvoordeelscan.

De projectleiders willen stuurinformatie met betrekking tot hun eigen projecten. Het aantal projecten en de veranderlijke cultuur binnen het bedrijf leggen een extra last op de informatievoorziening. In dit hoofdstuk wordt de beginsituatie beschreven van mijn stage. Deze momentopname is vastgelegd vanuit de volgende invalshoeken:

- Toestand van de data opslag
- Toestand van de informatievoorziening
- Rollen van personen binnen dit informatiesysteem

Met de eerste twee punten wordt de status van de technische kant van het informatiesysteem vastgelegd. De beschrijving van de rollen van personen legt de eisen en wensen van de "spelers" in dit informatiesysteem vast.

### 5.1. Toestand van de data opslag

#### 5.1.1. Data opslag binnen het VP platform

VP was twee jaar geleden een *datawarehousing* project gestart. Het *data warehouse* werd destijds geïmplementeerd door twee studenten die bezig waren met hun afstudeer opdracht. Het onderhoud van dit systeem is complex doordat in de loop der tijd op het oorspronkelijke systeem uitbreiding op uitbreiding plaats heeft gevonden onder druk van de snel veranderende bedrijfsprocessen. Bijkomend nadeel was dat het *data warehouse* was opgeleverd voordat de business met haar eisen en wensen kwam.

Bij het moduleren van dit *data warehouse* werd gebruik gemaakt van een zogenaamd *ster*- schema. In Appendix D is het oorspronkelijk ontwikkelde *ster* schema opgenomen. Bij een *ster* schema staat de zogenaamde feitentabel centraal. De rijen van de

feitentabel worden beschreven met de attributen van de dimensietabel [Thomas J. Kelly]. In appendix D staat “*Actionfact*” centraal als feitentabel. Een klant wordt bijvoorbeeld beschreven in de dimensietabel van “cliënt”. Door het ontwerp in een *ster* schema is het *data warehouse* ge-denormaliseerd, dus multi-dimensionaal.

Het project waar het oorspronkelijke *data warehouse* voor was bedoeld, ging uiteindelijk niet door. Het *data warehouse* werd vervolgens snel omgezet om andere projecten te kunnen ondersteunen. Men was bezig met het operationeel draaiend houden van het bedrijf.

Vervolgens werden weer meer andere projecten gestart, waardoor het systeem wederom aangepast moest worden. Zolang dit werkte en het de operationele websites ondersteunde, was dit géén probleem. Het omzetten van het oorspronkelijke ontwerp van het *data warehouse* veranderde het *ster* schema in een zogenaamd *snowflake* schema.

In Appendix E is een overzicht van de onderliggende tabellen database opgenomen. Als de herkomstdatabase gelinkt kan worden met de rest van het schema, kan het schema van Appendix E een zogenaamd *extended snowflake* schema worden genoemd. Dit schema maakt gebruik van meerdere lagen dimensietabellen en meerdere feitentabellen [Thomas J. Kelly]. Bij een dergelijke *extended snowflake* schema is de structuur complexer dan voorheen. Ook de omvang van de tabellen is toegenomen. De cliëntdimensie van het oorspronkelijke *ster* schema bevatte 21 attributen, nu bevat deze meer dan 90!

Een bezoeker kan meerdere CJ's afleggen, bij CJ's kunnen verschillende gegevens worden ingevuld. Deze worden allemaal opgeslagen in de cliënt tabel. De genestelde structuur en de toegenomen omvang zorgden voor een grotere complexiteit die uitmondde in een langere doorlooptijd van een informatie aanvraag.

Het oorspronkelijke *data warehouse* en de uitbereidingen op dit *data warehouse* wordt binnen VP het VP platform genoemd.

### 5.1.2. Data opslag buiten het VP platform

Een aantal van de websites onder het beheer van VP, sloeg de data niet binnen het VP platform op. Deze websites werden binnen VP PHP-sites genoemd, genaamd naar de scripttaal waarmee deze zijn ontwikkeld. Een twintigtal website-databases zorgden voor de data opslag van de PHP-sites.

De volgende knelpunten werden met data afkomstig van PHP-sites onderkend:

- **Geen centrale data opslag**  
Elke website had een eigen database
- **Performance problemen bij ad-hoc queries**  
Tijdens mijn stageperiode werd eenmaal een ad-hoc *query* aangevraagd met

betrekking tot deze websites. De *query* heeft 8 manuren gekost aangezien de grote hoeveelheid data constant time-out problemen op de server veroorzaakte.

- **Geen schone data**

De data bevatte slechte tekens. Bij de data analyse veroorzaakte dit binnen het OLAP pakket veel foutmeldingen.

- **Geen goed gedefinieerde bezoeker of sessie**

Een bezoeker was niet uniek identificeerbaar omdat tijdens het bezoeken geen sessie wordt gedefinieerd. Een *refresh* van de website werd bijvoorbeeld gezien als een extra bezoeker.

## 5.2. Toestand van de informatievoorziening

VP was zich voor mijn stage al bewust van de nadelen van haar informatiesysteem. Besloten werd het OLAP-pakket Di-Atlantis van Dimensional Insight aan te schaffen. Dit pakket werd na de aanschaf maanden niet gebruikt. De OLAP kubus moest worden gemoduleerd en worden gevuld met data en dat vergt veel inspanning. Deze was niet geleverd.

Een van mijn stage doelstellingen was te onderzoeken of het mogelijk is meer waarde uit het OLAP-pakket te halen.

Het pakket kan met een drop-down menu de *URL* en een tijdsrange die de gebruiker wil monitoren, uitkiezen. Vervolgens kan een gebruiker bijvoorbeeld op de links van een aantal *leads* klikken. De *query*-tool zal vervolgens het aantal *leads*, gegenereerd door de betreffende website, laten zien in de gegeven periode.

Deze informatievoorziening heeft de volgende nadelen:

- **Geen kwaliteitsgegevens beschikbaar over een *lead***

Het is namelijk alleen mogelijk de kwantiteit van de *leads* te meten.

- **Rapportage is niet dynamisch weergegeven**

Wanneer data uit een andere periode van dezelfde website is gewenst, dient een nieuwe *query* te worden uitgevoerd.

- **Tijdrovend**

Wanneer een gebruiker een aantal websites naast elkaar wil vergelijken, zal het resultaat van de *query* moeten worden opgeschreven of moeten worden gekopieerd in een spreadsheet.

- **Traag**

De *query* moet vaak meerdere website-databases doorzoeken, hierdoor is het opvragen van informatie via de *queries* traag. Vaak neemt dit meer dan 10 minuten in beslag.

- **Geen overzicht of aggregatie mogelijk**

Zoals in paragraaf 1.2 kon worden gelezen, beheert VP een groot aantal websites. Vaak is het juist noodzakelijk informatie met betrekking tot deze websites samengevat te rapporteren.

VP was zich bewust van de nadelen van het informatiesysteem en had het OLAP-pakket Di-Atlantis aangeschaft bij Dimensional Insight. Dit pakket werd na de aanschaf maanden gebruikt. De OLAP kubus moest worden gemoduleerd en worden gevuld met data en dat vergt veel inspanning. Deze was niet geleverd.

Een van mijn stage doelstellingen was te onderzoeken of het mogelijk is meer waarde uit het OLAP-pakket te halen.

### **5.3. Rollen van de mensen binnen dit informatiesysteem**

Het kostte businessgebruikers veel tijd de gewenste rapportage te verkrijgen. Volgens de marketing manager van VP, Marc Kerstel, was hij hier per maand wel drie volle werkdagen aan kwijt. Vaak ook, werd de gewenste vorm van rapportage niet ondersteund door de *query*-tool. In dat geval moest een ad-hoc *query* worden aangevraagd bij de technische afdeling of moest de gebruiker zijn beslissingen nemen zonder de beschikbaarheid van deze gegevens.

De business realiseerde zich dat VP maar een beperkt aantal parameters heeft die gestuurd kunnen worden.

- Het selecteren van leverancier of afnemer
- de *look and feel* van een website
- de inhoud van een CJ.

Men is er van overtuigd dat de beslissingen in deze parameters het bedrijfsresultaat beïnvloeden. Vandaar dat deze parameters binnen VP worden aangeduid als stuurparameters.

Als een CJ een optimistische rekenmodule heeft, is de kans dat een bezoeker een *lead* wordt groot. Dit zegt echter nog niets over de kwaliteit van het na-traject waarin de *lead* wordt opgevolgd. Onvoldoende *feedback* op de genomen beslissingen, speelt het nemen van beslissingen op basis van ervaring en intuïtie onbedoeld in de kaart.

Het tijdig en op een overzichtelijke wijze representeren van de effecten die de genomen beslissingen in de stuurparameters hebben, is de kern van de informatievoorziening voor VP. De beslissingen op tactisch en op operationeel niveau moesten daarom worden ondersteund door een gestroomlijnde informatievoorziening die de *feedback* in zowel kwantitatieve als in kwalitatieve grootheden weergeeft.

Voor de kwaliteitsbepaling van de *leads*, is externe data noodzakelijk. De kwaliteit van een *lead* is evenredig aan de kans dat een *lead* tot een transactie kan leiden. Hoe hoger de kans op een transactie, hoe hoger de kwaliteit van een *lead*.

De kwaliteit van een *lead* is eigenlijk één van de belangrijkste *measures* van VP. De inkomsten van VP hangen direct danwel indirect af van de kwaliteit van een *lead*. Men wil graag weten wat voor effecten een beslissing in de stuurparameter van VP kan hebben op de kwaliteit.

De *query*-tool is niet geschikt voor het bepalen van de invloed van de verschillende websites, CJ's, uitingen en/of proposities en de leverancierskeuze op de kwaliteit van de *leads*.

Omdat de *query*-tool niet geschikt is, zijn de volgende wensen ten opzicht van de nieuwe BI-project opgesteld:

Eigenaar	Wensen
Businessgebruiker	<ul style="list-style-type: none"> <li>- Snellere rapportage</li> <li>- Meer flexibele rapportage</li> <li>- Meer precieze meting van de kwaliteit en de kwantiteit van de <i>leads</i></li> <li>- <i>Feedback</i> in resultaten bij genomen beslissingen</li> </ul>
Techniek	<ul style="list-style-type: none"> <li>- Betere definities</li> <li>- Minder tijd voor ad-hoc <i>queries</i> zodat meer tijd beschikbaar komt voor de automatisering van standaard vragen</li> </ul>

**tabel 5-1 De wensen en hun eigenaren**

#### **5.4. Een mogelijke oplossing met OLAP**

De wensen van de businessgebruiker en de techniek wezen al snel naar een OLAP oplossing. Stel je een OLAP kubus voor met de volgende eigenschappen:

- Server en cliënt architectuur die de gebruikers in staat stelt via een client programma zonder tussenkomst van de techniek bij zijn informatie te komen.
- Kubus die alle stuurparameters als dimensies en resultaat gegevens (kwaliteit en kwantiteit van een *lead*) als *measure* bevat. Hierdoor krijgen de gebruikers *feedback* op de genomen beslissingen.
- Externe data wordt geïntegreerd met interne data. Voor VP is dit noodzakelijk voor het bepalen van de kwaliteit van de *leads*.
- De OLAP kubus wordt automatisch ververs met nieuwe informatie.

Met een kubus zoals hierboven beschreven (aangenomen wordt dat datahuishouding ok is), kunnen de wensen van de businessgebruikers (zoals beschreven in tabel 5-1) worden vervuld. Niet alleen is het hiermee mogelijk snellere en meer flexibele rapporten te genereren, het is zelfs mogelijk de business-processen *real-time* te volgen. De snelheid is met name te danken aan het feit dat de OLAP-kubus altijd de geaggregeerde vorm is van de oorspronkelijke dataset.

Een voorbeeld met een onderliggende data van meer dan 20 *Megabyte* in een testcase toonde aan dat wanneer de dimensies en *measures* zijn geselecteerd, na aggregatie een data kubus van minder dan 1 *Megabyte* overblijft.



VP kan alleen de opvolging van de *leads* meten wanneer zij terugkoppeling krijgt van de volgende schakel in de distributieketen. Doordat de externe data (= terugkoppeling) wordt geïntegreerd in de kubus, kan VP de kwaliteit van de *leads* meten.

Het implementeren van een OLAP kubus vereist goede definities, deze definities worden vooraf vastgesteld. Aan de hand hiervan kan de techniek beter verifiëren of de juiste datamodellen worden opgebouwd.

De flexibiliteit wordt zichtbaar doordat de gebruiker via verschillende (combinaties van) invalshoeken zijn business kan monitoren. Doordat de eindgebruiker zelf veel *views* en representaties kan creëren vervalt een groot gedeelte van de informatievraag naar ad-hoc *queries* waardoor de technische afdeling meer tijd beschikbaar krijgt voor hun primaire taken.

## **5.5. Samenvatting**

Aan de hand van een momentopname van het informatiesysteem kon een inventaris worden opgesteld van de wensen van VP:

- Bepaling van de kwaliteit van de *leads*
- *Feedback* van het effect van de genomen beslissingen op de bedrijfsvoering
- Informatievoorziening die minder handelingen kost dan de huidige informatievoorziening
- Data opslag geschikt maken voor analyse

In paragraaf 5.4 is een voorbeeld BI-oplossing gegeven (op basis van OLAP) die aan al deze wensen VP tegemoet kan komen.

## 6 BI ondersteuning voor Vermogensplanet

Er wordt een OLAP kubus geleverd tijdens mijn stage. Dit hoofdstuk beschrijft hoe dit produkt tot stand is gekomen. Ook wordt in dit hoofdstuk een momentopname vastgelegd van de status van de informatievoorziening na levering van het eindprodukt. Door de momentopname na levering te vergelijken met de aan het begin gemaakte momentopname (zie hoofdstuk 5), wordt de toegevoegde waarde van mijn project duidelijk gemaakt. De toegevoegde waarde is gesplitst in de interne en in de externe toegevoegde waarde.

### 6.1. Data opslag

De huidige data wordt opgeslagen in relationele databases. Welke technologie gebruikt wordt voor het opslaan van onderliggende data is niet van belang [Erik Thomsen - 2002]. Zolang de data voldoet aan de eisen van OLAP, kunnen OLAP kubussen worden geconstrueerd. De eisen met betrekking tot de onderliggende data zijn besproken in hoofdstuk 3.

De technische afdeling diende een aantal garanties af te geven om het succesvolle verloop van het BI-project binnen VP te kunnen waarborgen.

Deze garanties waren:

- De gegevens moesten centraal worden opgeslagen
- Er moest een super tabel komen die de benodigde dimensies en *measures* bevat voor de OLAP-kubus. Dit wil zeggen dat de BI-er moest definiëren wat benodigd was door middel van het bestuderen van het proces en het interviewen van businessanalisten. Het concept van deze super tabel lijkt veel op de allesomvattende tabel in paragraaf 3.3
- De super tabel moest zowel schone data als externe data bevatten en moest tevens multi-dimensionaal zijn. Multi-dimensionaal is in dit geval equivalent aan ge-denormaliseerd.
- De super tabel moest incrementeel worden aangevuld. Dat wil zeggen dat de super tabel automatisch wordt gevuld met nieuwe gegevens uit de operationele databases.
- De ontwikkelwijze van de super tabel diende incrementeel te zijn. De super tabel en de ETL die de super tabel automatisch genereert, diende rekening te houden met toekomstige nieuwe *requirements*.

### 6.2. OLAP oplossingen in het algemeen

De garanties besproken in de voorgaande paragraaf waren cruciale randvoorwaarden voor het BI-project. Door het bedrijfsproces en de strategie te bestuderen en de businessgebruikers te interviewen werd de algemene informatievraag binnen VP geïdentificeerd. Deze algemene informatievraag is de informatie die een manager

regelmatig nodig heeft om zijn beslissingen goed te kunnen nemen.

In paragraaf 6.2.1 worden deze informatievragen geïdentificeerd en gegroepeerd in drie kubussen:

- De monitoringkubus
- De bezoekersgedragskubus
- De segmentatiekubus

Hierdoor wordt de kans op een hyperkubus<sup>3</sup> beperkt.

Het bleek niet mogelijk alle drie de soorten kubussen ter ondersteuning van de algemene informatievragen tegelijk te implementeren. Besloten werd daarom om één van de besproken kubussen uit te werken: de monitoringkubus.

### 6.2.1.      **Identificeren van de algemene informatievragen**

Door middel van het interviewen van gebruikers heb ik de volgende drie soorten standaard informatiebehoeften binnen VP kunnen vaststellen:

- **Monitoring informatie**  
Kosten- en winstcijfers worden gemonitord. Slechte overeenkomsten en proposities moeten op tijd worden gestopt. Een precieze en accurate ondersteuning van de monitoring kan helpen bij het nemen van de juiste beslissing.
- **Bezoekersgedrag informatie**  
Hoe reageert een bezoeker op een pagina of een uiting? Waar haken de bezoekers af? Het analyseren van het bezoekersgedrag beoogt de conversie van bezoeker naar *leads* te verhogen.
- **Segmentatie informatie**  
De echte klanten leren kennen, waardoor de juiste klanten benaderd kunnen worden met het juiste produkt of propositie. Voorwaarde hiervoor is dat VP weet welk profiel en de daaraan gekoppelde behoeften haar klanten hebben. Dit zal kunnen leiden tot een betere klantenbinding.

#### **Monitoringkubus**

De monitoringkubus ondersteund het nemen van grote aantallen beslissingen in het bedrijfsproces. Een kubus die het effect van deze beslissingen op de kwantiteit en kwaliteit van de *leads* weergeeft is daarom uitermate belangrijk voor VP. Een dergelijke OLAP-kubus wordt een monitoringkubus genoemd.

#### **Bezoekersgedragskubus**

Een andere kubus die interessant is, geeft het bezoekersgedrag weer. Zoals eerder gesteld, beperkt een CJ het aantal *zijpaden* van het bezoek op de site. Met de monitoringkubus kan alleen worden bekeken of een CJ of een website het goed doet ten opzicht van een andere CJ of website. Het effect van een verandering in de CJ of de

---

<sup>3</sup> De hyperkubus theorie is in beperkte vorm beschreven in hoofdstuk 4.

website zelf kan alleen op den duur worden gemerkt. Dit heeft een nadeel: na een verandering van CJ of website moet nog redelijk lang worden gewacht tot er genoeg data beschikbaar is voor het doen van een uitspraak over het resultaat hiervan. Het is van weinig belang te weten of een verandering goed of slecht is. Wél van belang is het te weten hoe de *lead*-genererende capaciteit van de CJ en de website kan worden vergroot. Dit kan alleen worden bereikt met het *online* analyseren van het bezoekersgedrag. De kubus die daarvoor bedoeld is wordt de bezoekersgedragskubus genoemd.

Een goed in kaart gebracht bezoekersgedrag heeft een hoge toegevoegde waarde. Als aan de hand van het gedrag van de bezoekers de juiste CJ's worden ontwikkeld, kan dit leiden tot een stijging in de conversie ratio van bezoekers naar *leads*. Hierdoor is VP in staat *leads* goedkoper te genereren. Een heel klein percentage van de bezoekers leidt immers tot *leads*. Het is niet moeilijk voor te stellen dat een verhoging van dat percentage met 0.1% een enorm positief verschil uitmaakt op de kostprijs van een *lead*. Met de bezoekersgedragskubus kan VP de CJ's optimaliseren.

### **Segmentatie kubus**

Het verkopen van financiële producten, vanaf de *lead* tot aan de handtekening van de klant, neemt vaak maanden doorlooptijd in beslag. Bij een hypotheek *lead*, beslaat de tijdsperiode van *lead* naar transactie gemiddeld 4 maanden. Geplaatst in de kubus theorie komt dit er in feite op neer dat je auto rijdt met alleen zicht op de achteruitkijkspiegel. Dit, terwijl het eigenlijk gewenst is vooruit te kijken; de potentiële klanten met de juiste propositie benaderen. Dit is een segmentatieprobleem.

Kennis hebben over je afnemers is altijd belangrijk bij het doen van zaken. Wanneer een bakker de verjaardagen van zijn klanten onthoudt is hij in staat klanten te binden door bijvoorbeeld een verjaardagscake goedkoper aan te bieden op de verjaardag de individuele klant. Naast klantenbinding heeft dit ook nog een mogelijk hogere omzet tot gevolg.

Voor VP gaat bovenstaand principe ook op. De klanten van VP kunnen worden gesegmenteerd, VP kan het juiste segment klant benaderen met de juiste propositie en daardoor beter de klanten bedienen. Klantensegmentatie vertaald in de termen van data is niet anders als het zoeken naar de verbanden tussen de data in hogere data dimensies. OLAP is niet de meest geschikte tool om deze taak uit te voeren. Wel kan met een OLAP kubus de data worden verkend zodat het een globale segmentatie vooraf mogelijk maakt. Data mining<sup>4</sup> is een beter geschikt middel om segmentaties te maken.

---

<sup>4</sup> Zie ook paragraaf 2.2.4

### 6.2.2.       Keuze van de te implementeren kubus

VP groeit; er worden veel opdrachten binnengesleept. Hierdoor ontstaan structurele capaciteitstekorten. Als gevolg hiervan moest een keuze worden gemaakt welk gedeelte van de algemene informatie behoefte als eerste moest worden ondersteund. Besloten werd de monitoringkubus te implementeren in juni 2003. Binnen het bedrijf werd dit in de wandelgangen aangeduidt met het "juni project"

De keuze voor de monitoringkubus lag min of meer voor de hand:

- Grootste aantal informatievragen in vergelijking met andere kubussen
- Lager risico ten opzichte van segmentatie  
Bij segmentatie wordt OLAP naast data mining uitgevoerd. *Data mining* is buiten de scope van mijn stage geplaatst. Zie paragraaf 1.4
- Hogere strategische waarde dan bezoekersgedragskubus  
De markt waarin VP opereert is niet doorzichtig. Meestal wordt alleen nog maar naar de kwantiteit van de *leads* gekeken. Een kwalitatief goede *lead* levert de afnemers inkomsten op, maar een kwalitatief slechte *lead* kan de afnemer geld kosten!  
VP vindt het belangrijk dat door de waardeketen heen wordt geoptimaliseerd. Daarom is het aantonen van de kwaliteit van *leads* belangrijk voor VP.

### 6.2.3.       Welke beslissingen ondersteunt de monitoringkubus?

Binnen VP zijn de volgende zaken (deelprocessen) vanuit het bedrijfsproces wederkerend:

- Marketing
- Inkoop
- *Lead* generatie
- Verkoop

De monitoringkubus moest de volgende businessvragen kunnen beantwoorden:

#### **Marketing**

- Op welke uiting wordt gereageerd?
- Op welke proposities worden gereageerd?
- Gaat de conversie (naar *leads* of naar transactie) omhoog of omlaag wanneer de **uiting** wordt gewijzigd?
- Gaat de conversie (naar *leads* of naar transactie) omhoog of omlaag wanneer de **propositie** wordt gewijzigd?

#### **Inkoop**

- Wat is de *lead*-kwaliteit van de *leadbron*?  
Ervaring leert dat deze per inkoopbron sterk kan verschillen
- Wat kost een CPC-overeenkomst per *lead*?

- Wat willen we bij een bepaalde leverancier maximaal uitgeven per *lead* of per *click*?
- Is de overeenkomst winstgevend?

### **Lead generatie**

- Welke propositie doet het het best op welke website?
- Welke CJ doet het goed?
- Op welke site kan het best een bepaalde uiting worden geplaatst?
- Hoe wordt de conversie beïnvloed als de CJ wordt gewijzigd?

### **Lead verkoop**

- Aan wie kunnen we de leads het beste verkopen?
- Welk soort overeenkomst (per lead of op provisie basis) is interessant?
- Hoe verloopt de opbrengst per afnemer?

## **6.3. De implementatiestap**

Nu de monitoringkubus was gedefinieerd, kwam de focus op de implementatie hiervan te liggen. Bij de implementatiefase van de monitoringkubus moest aandacht worden besteed aan onder andere de volgende zaken:

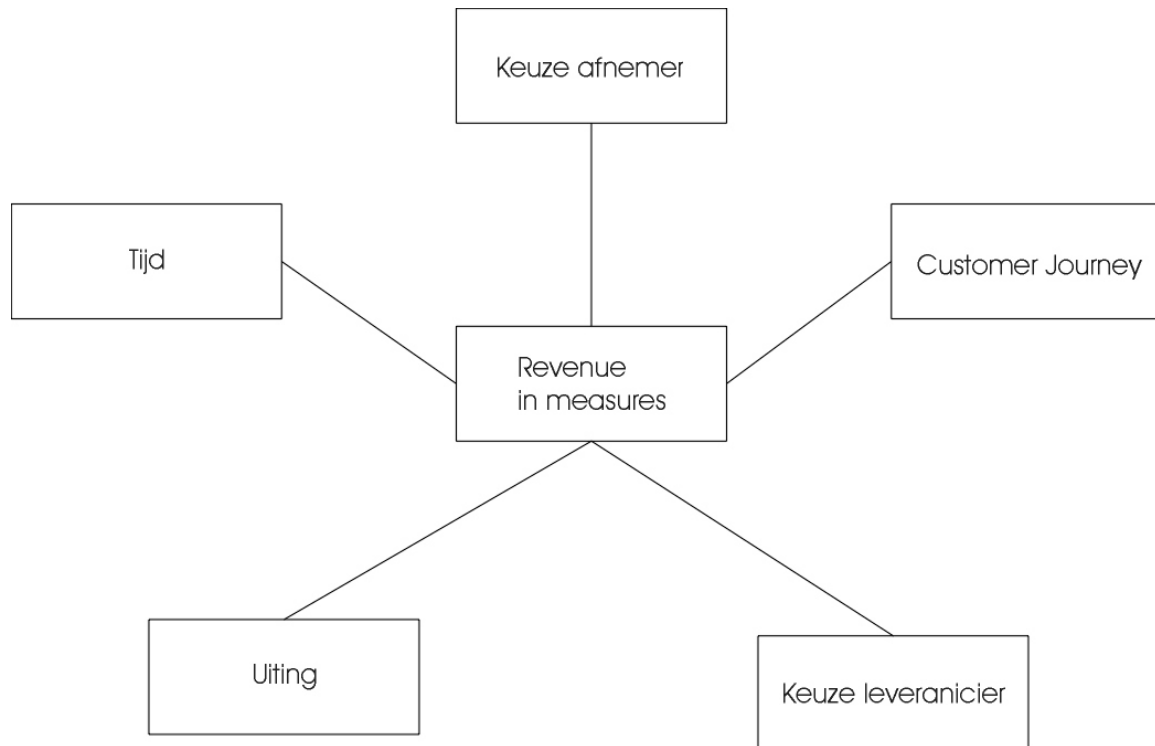
- **Communicatie**  
Zodat de gebruikers op de hoogte waren van de ontwikkelingen.
- **Allocatie van capaciteit van de technische afdeling**

### **6.3.1. Communiceren met behulp van definities**

De factoren die de *measures* (ofwel opbrengsten) zullen kunnen beïnvloeden zijn weergegeven in het mini-ster schema van figuur 6-1. Dit figuur is in feite al een multi-dimensionale representatie van de data. De middelste "Revenu in *measures*" representeert de inhoud van de kubus en de omliggende factoren zijn de assen van de dimensies.

Dit noem ik een mini-ster schema. Qua uiterlijk lijkt het op een ster schema van een *data warehouse*, eigenlijk is dit niet het geval. Bij een ster schema van een *data warehouse* staat de feitentabel centraal. Bovendien is deze veel gedetailleerder dan in deze mini-ster. Het beoogde nut van een ster schema is moduleren. Bij het mini-ster schema is het beoogde nut: "communiceren".

Dit schema vormt een krachtig communicatiemiddel. De businessanalisten behoeven hiermee niets van databases af te weten. Wel kunnen ze hiermee makkelijk aangeven waar ze, via welke invalshoek, op willen kunnen monitoren. Het afstemmen van het doel van de OLAP kubus met de gebruikers, was met de hulp van het mini-ster schema veel gemakkelijker geworden. Het schema verschaft immers in één oogopslag een overzicht van dit doel.



**figuur 6-1 Mini-ster schema**

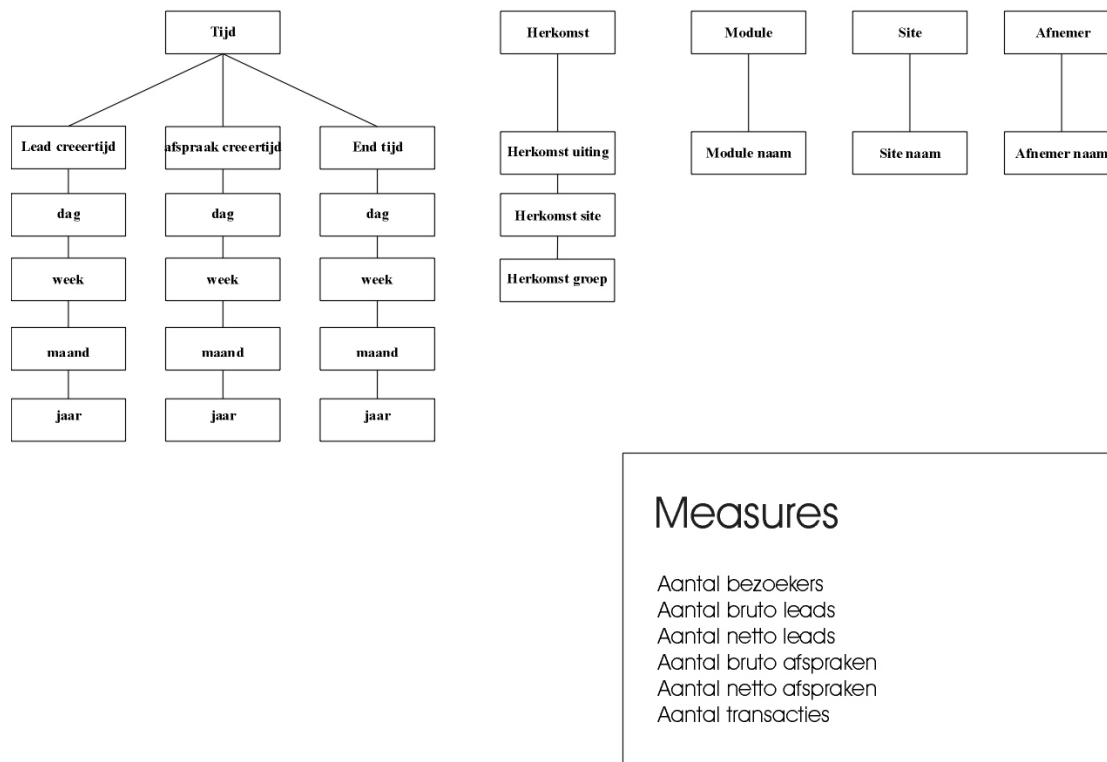
Voor het uitvoeren van dit OLAP-project was een mini-ster schema niet voldoende aangezien de invloed factor niet gelijk hoeft te zijn aan één dimensie van een OLAP kubus. Met de dimensie tijd alleen al, zijn er meerdere keuzes op verschillende aggregatieniveaus:

- gegevens per dag
- per week
- per maand
- of per jaar.

De keuze van het aggregatieniveau is de verantwoordelijkheid van de businessanalisten. De keuze op aggregatieniveau in de dimensies is in feite de vraag: "op welke detail niveau wil een businessanalist sturen?"

Daarom wordt verder gecommuniceerd een figuur dat de gebruikerswensen vastlegt - figuur 6-2. Dit figuur geeft een overzicht van de dimensiestructuur en de *measures* aan. De businessanalist mag daarin de gewenste dimensies en *measures* omcirkelen.

# Standaard kubus voor het monitoren



figuur 6-2 Dimensie en *measures*

De vastgelegde wensen zijn nog niet voldoende gedefinieerd om een probleemloze implementatie door de techniek te kunnen garanderen. In Appendix F ziet u een overzicht van dimensies en *measures* voor dit project. Appendix G beschrijft de definitie van de *measures*. Deze zijn wel te interpreteren door de techniek.

## 6.3.2. Allocatie van de capaciteit

Binnen VP is de capaciteit van de techniek een structurele bottleneck geworden. De allocatie van voldoende capaciteit was daarom van levensbelang voor de succesvolle afronding van het BI-project. Samen met de externe begeleider professor Eiben had ik het stappenplan van het zogenaamde juni-project gedefinieerd. Het juni-project diende twee projecten binnen VP te voorzien van een monitoringkubus: namelijk het "WP" project (WP) en het project "Oversluiten".

Het overzicht van de benodigde acties om de juni-kubus te construeren, wordt weergegeven in figuur 6-3. Dit overzicht maakte het mogelijk de initiële levering van de juni-kubus in welgeteld vier mandagen te realiseren.

De pijlen naar de VP operationele data, zijn de externe databronnen. Extern wil zeggen dat de data zich niet binnen het VP platform bevindt. Op basis van de data wordt via een automatische routine (ETL) de supertabel geconstrueerd. Deze supertabel bedient de



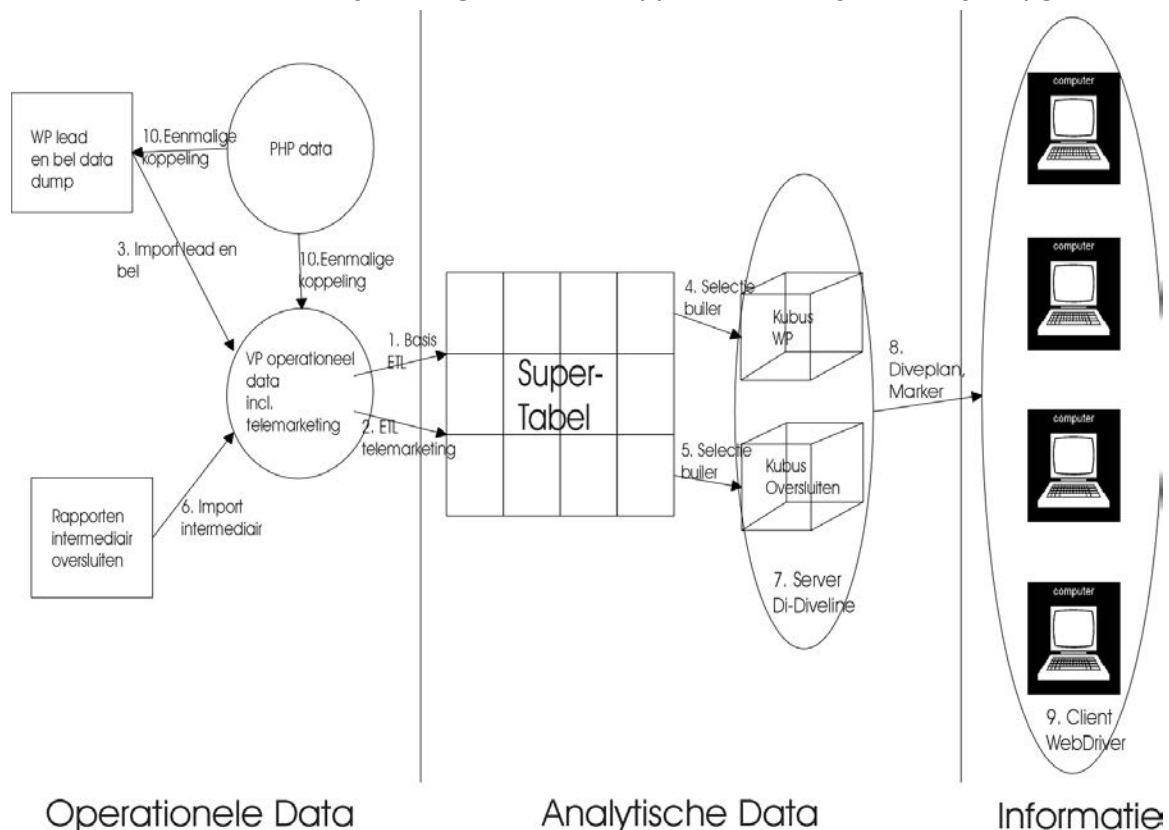
kubussen die worden opgeslagen op een server zodat de gebruikers via Internet en het cliëntprogramma, de kubussen kunnen bereiken.

De definities van de kubus in Appendix G, geven aan wat de voorwaarden zijn om de kubus te kunnen construeren. Door de benodigde voorwaarden te beschrijven werden de gevolgen van het niet realiseren van een bepaalde stap in kaart gebracht in de termen van dimensies en *measures*. Een voordeel van deze manier van werken was dat het gemakkelijk is om op basis hiervan de status van het project te communiceren. Appendix H bevat een voorbeeld van de communicatie van de statussen. De tabel geeft een overzicht welke stappen door wie moeten worden uitgevoerd, in hoeverre deze stappen voltooid zijn en het verwachte resterende werk.

Het juni-project omvatte niet alleen het implementeren van de kubus. Andere onderwerpen die binnen het project vielen, waren:

- Verificatie of het mogelijk was het systeem te bouwen
- Validatie of het juiste systeem was geïmplementeerd
- Opleiden van de gebruikers in de multi-dimensionale denkwijze

Deze taken werden door mijzelf uitgevoerd. In Appendix I is mijn takenlijst opgenomen.



**figuur 6-3 Grafische weergave van het stappenplan van het juni-project**

#### **6.4. Samenvatting**

De totstandkoming van de juni-kubussen werd in dit hoofdstuk besproken. In paragraaf 6.1 werden de randvoorwaardelijke garanties ten opzichte van het juni-project besproken. In paragraaf 6.2 zijn de redenen van het te implementeren plan uitgelicht. BI keek eerst naar de verschillende informatiebehoeften binnen de organisatie, vervolgens werden deze behoeften gegroepeerd in een drietal OLAP kubussen. De op dat moment meest relevante kubus voor VP werd uitgekozen om als eerste geïmplementeerd te worden. De verdere implementatiefase werd beschreven in paragraaf 6.3.

## 7 Project evaluatie

Dit hoofdstuk evalueert het BI-project in zijn geheel (niet alleen het juni-project). Opnieuw wordt een moment opname vastgesteld; ditmaal na de levering van de monitoringkubus. Deze is terug te vinden in paragraaf 7.1. Wanneer de beide momentopnames uit deze scriptie met elkaar worden vergeleken, kan de gerealiseerde toevoegde waarde van het juni-project worden vastgesteld.

Een nieuwe oplossing neemt vaak nieuwe problemen met zich mee. De heer Beekmans (directeur van VP) vergelijkt dit met de volgende situatie: "Als je een persoon die honger heeft elke dag witte rijst geeft, zal hij dit prachtig vinden. Maar na een maand van alleen rijst eten, wil hij er toch een stukje groente of vlees bij."

Paragraaf 7.3 tot en met paragraaf 7.5 gaan over dat stukje "groente of vlees" dat nog niet is meegeleverd in het juni-project. De evaluatie wordt afgesloten met de beschrijving van het OLAP draaiboek voor VP. Dit kan worden gebruikt als checklist voor toekomstig op te zetten OLAP-projecten.

### 7.1. *Momentopname na levering*

Na de oplevering van de juni-kubus werd opnieuw een momentopname gemaakt. Deze momentopname werd wederom vastgelegd vanuit de invalshoeken:

- Toestand van de data opslag
- Toestand van de informatievoorziening
- Rollen van personen binnen dit informatiesysteem

#### 7.1.1. Data opslag

Vanaf februari tot mei 2003 heeft de technische afdeling de losstaande websites die ieder een aparte database hadden, omgebouwd zodat de gegevens centraal konden worden opgeslagen. Het ombouwen van de data opslag viel buiten de scope van het OLAP project. Centrale opslag was één van de voorwaarden voor het kunnen maken van een aantal analyses.

#### 7.1.2. Informatievoorziening

De informatievoorziening verloopt nu via de OLAP kubus. Men kan de vooraf gedefinieerde *measures* via verschillende invalshoeken of combinaties hiervan bekijken. Men gebruikt *divebook*<sup>5</sup> voor het maken van een aantal standaard *views*. De gegevens zijn visualiseerbaar. De informatie via de OLAP kubus is gestroomlijnd. Een automatisch ETL script zorgt ervoor dat de OLAP kubus dagelijks met nieuwe gegevens wordt

---

<sup>5</sup> Set van bookmarkers in de data. Dit is een functionaliteit van het gebruikte OLAP pakket.

aangevuld. De externe gegevens moeten tot op heden nog handmatig worden geïmporteerd in de transactie database. Vanaf deze database handelt het script de verdere verwerking automatisch af. De supertabel wordt ge-update en vervolgens wordt de kubus gevuld met vernieuwde informatie.

### **7.1.3. De rollen**

De juni-kubus betekent voor de gebruikers het volgende:

- Minder handelingen nodig bij het genereren van de benodigde informatie.
- Flexibel weergeven van de rapporten en de mogelijkheid tot het weergeven van dwarsverbanden.
- Mogelijkheid van de bepaling van de kwaliteit van de *leads*.
- *Feedback* van de beslissingen in de stuurparameters.

## **7.2. Gerealiseerde toegevoegde waarde**

De juni-kubus heeft een grote strategische waarde voor VP. Het bepalen van de kwaliteit van de *leads* stelt VP in staat haar toegevoegde waarde voor de partners beter inzichtelijk te maken. Verder kan VP dezelfde technologie reproduceren en wanneer gewenst, bij meerdere projecten en klanten toepassen.

De volgende toegevoegde waarden zijn gerealiseerd voor interne en externe gebruikers.

### **Toegevoegde waarde intern**

- Kwaliteit van *leads* kan worden bepaald
  - Informatievoorziening voor het monitoren is gestroomlijnd
  - Minder tijd is benodigd voor het verkrijgen van informatie
- Dit laatste bespaart de businessgebruiker drie volle werkdagen per maand.

### **Toegevoegde waarde extern**

De data terugkoppeling van de externe partij WP heeft een tweetal interessante attributen. Namelijk:

- het kantoor welke de *leads* heeft gevolgd
- de adviseur die de *leads* heeft behandeld

Deze attributen zijn geïntegreerd in de monitoringkubus. Hierdoor kan de geleverde kubus ook stuurinformatie aan WP verschaffen.

Voor WP zijn de volgende twee soorten informatie van belang:

- Stuur informatie van het callcenter is opvraagbaar bij VP
- De resultaten kunnen worden gemonitord op het niveau van "kantoor" en van "adviseurs".

### 7.3. **Extra vragen bij het project oversluiten**

Het OLAP project werd in februari gestart. Bij het construeren van de kubussen werden uitvoerig de *requirements* van het project WP bestuurd. Het project WP was destijds het belangrijkste project binnen VP. Eind april startte een ander project: het project "oversluiten". Oversluiten behandelt net als het project WP de hypotheek *leads*; klanten die over willen stappen van de ene naar een andere hypotheek.

Bij een historisch lage rente kan het soms aantrekkelijk zijn voor mensen hun bestaande hypotheek over te sluiten naar een nieuwe hypotheek. VP speelt hierop in door eind april het project Oversluiten te starten. Bij het oversluiten van een hypotheek komen meerdere factoren kijken. De klant moet meestal een boete betalen aan de hypotheekverstrekker omdat de hypotheek vervroegd wordt afgelost. Verder moet rekening worden gehouden met de verzekerings- en bemiddelingskosten. De klant wil graag inzicht krijgen in de financiële gevolgen wanneer hij/zij overweegt een bestaande hypotheek over te sluiten. Derhalve zullen de naast de verzekerings- en bemiddelingskosten, ook de terugverdien tijd van de boete en de extra bijkomende kosten van het oversluiten moeten worden berekend.

VP heeft CJ's ontwikkeld die de financiële gevolgen berekenen van het oversluiten. Deze zijn op een aantal websites geplaatst. Wanneer een klant de uitkomst van een CJ interessant vindt, kan hij of zij een *lead* via een website worden. De *lead* wordt verder geholpen door telemarketeers en hypotheek adviseurs.

De telemarketingfase heeft de volgende doelen:

- **Nakijken of een klant zijn gegevens compleet en correct heeft ingevuld**  
Deze stap wordt inventarisatie genoemd. De telemarketeer moet de gegevens inventariseren en alle gegevens die nodig zijn voor de berekening verzamelen aan de hand van een checklist.
- **Narekenen of een concurrerende offerte kan worden uitgebracht**  
Deze stap wordt acceptatiefase genoemd en wordt uitgevoerd door een adviseur met financiële kennis. Hierbij wordt naar twee zaken gekeken:
  - o Toetsen of de klant voldoet aan de krediet eisen
  - o Afwegen of onze offerte voldoet aan de wens van de klant

De beloning van VP hangt af van het aantal netto afspraken. Wanneer een klant wordt geaccepteerd, worden zijn gegevens verzonden naar een financiële partner van VP. Deze partner probeert een afspraak met de klant te maken en zal een adviseur op pad sturen. Een netto afspraak houdt in: een geplande en gehouden ontmoeting tussen een geaccepteerde *lead* en de adviseur.

In de volgende paragrafen gaan we verder in op de niet gerealiseerde informatievraag. De juni-kubus kan door het opnemen van extra *measures* een gedeelte van deze niet gerealiseerde vraag opvangen. Een structurele oplossing om dit gedeelte van de informatievoorziening te stroomlijnen wordt besproken in paragraaf 7.3.5.

### 7.3.1. Verschillen tussen het oversluiten en het WP project

Bij het oversluiten project werden de wensen ten op zichte van de informatievoorziening pas duidelijk tijdens de implementatiefase van de juni-kubus. Daardoor was er onvoldoende tijd om aan alle informatievragen van dit project, binnen het juni-project, te kunnen voldoen.

Het proces van oversluiten is voor VP een ander proces dan het proces van WP. Bij het project WP, stoppen VP's activiteiten wanneer een bezoeker een *lead* wordt. De *leads* worden naar WP gestuurd en worden nagebeld en verder vervolgd door WP zelf. Voor VP zijn de resultaten van *lead* naar transactie belangrijk om de kwaliteit van de *leads* te kunnen meten. Of het callcenter van WP efficiënt zijn werk doet is voor VP niet relevant aangezien dit callcenter niet onder VP's verantwoordelijkheid valt. Voor oversluiten is het efficiënte werken van het callcenter **wél** relevant voor VP aangezien zij hierbij de telemarketingfase voor hun rekening nemen.

### 7.3.2. Wat zijn de extra eisen van oversluiten?

De projectleider van Oversluiten wil weten hoe efficiënt gewerkt wordt in de telemarketingfase. Ook wil de projectleider weten waar zich binnen het proces de eventuele knelpunten bevinden. De volgende tabel verwoordt haar informatiebehoefte.

Week nummer	Aantal bezoekers	Aantal <i>leads</i>	Aantal geïnventariseerd	Aantal geaccepteerd	Aantal afspraken
17	1000	20	19	12	11
...	...	...	...	...	

tabel 7-1 Informatievraag bij oversluiten

Het schema lijkt op het schema met de dimensies en *measures* die ook bij het project WP zijn gebruikt. Het verschil ten opzichte van het WP-schema ligt in de definities van de *measures*. Bij het project WP is de ratio analyse voor het monitoren van de resultaten zéér belangrijk. Hier houdt het "aantal afspraken" in week 17 ook daadwerkelijk het aantal afspraken voortkomend uit *leads* in week 17 in. Dit is zinvol wanneer wordt gekeken naar de conversie van *leads* naar afspraken.

Voor oversluiten, wil de projectleider het aantal geïnventariseerde *leads*, het aantal geaccepteerde *leads* en het aantal afspraken die zijn gerealiseerd in week 17, zien. Hier kan bijvoorbeeld geen ratio analyse op los worden gelaten aangezien de 19 *leads* die geïnventariseerd werden in week 17, niet per sé alle *leads* zijn die in deze week zijn gegenereerd. Wanneer bijvoorbeeld een achterstand wordt ingehaald, zullen meer afspraken worden gerealiseerd dan het aantal in die week gegenereerde *leads*.

### 7.3.3.        **Waarom kan de juni-kubus niet aan deze extra eisen voldoen?**

In de oude situatie, werden de ratio's uitgerekend aan de hand van de onderliggende data die gelinkt is aan de *lead-ID*. De enige tijd dimensie in de kubus is het tijdstip waarop de bezoeker een *lead* wordt. Door de extra vragen tijdens het project oversluiten zijn meerdere tijd-dimensies ontstaan:

- Tijdstip dat de bezoeker een *lead* wordt
- Datum dat de *lead* geïnterpreteerd wordt
- Datum dat de *lead* geaccepteerd wordt
- Datum van de afspraak met de potentiële klant.

Wanneer deze tijdsdimensies alsnog toegevoegd moesten worden aan de bestaande kubus, zouden de *measures* niet meer in een semantische ruimte staan. De *juni*-kubus zou hierdoor een hyperkubus worden. De bestaande dimensiestructuur van de *juni*-kubus kon niet aan deze uitbreiding voldoen. De theorie van de hyperkubus in vergelijking met een kubus in semantische ruimte is in hoofdstuk 4 terug te vinden.

### 7.3.4.        **Een tijdelijke oplossing binnen juni-kubus voor oversluiten**

Gekeken werd naar het type beslissing dat wordt ondersteund door de informatie in het formaat van tabel 7-1. Deze typen zijn:

- Waar zijn de knelpunten binnen de telemarketingfase?
- Wanneer moet worden ingegrepen door bijvoorbeeld een extra persoon in te zetten?

Gelukkig kunnen deze vragen worden beantwoord met behulp van de volgende *measures* van de *juni*-kubus:

- Werkvoorraad inventarisatie
- Werkvoorraad acceptatie

Het doel van het opnemen van werkvoorraad in de *measures*, was om beter de ratio's te kunnen berekenen. De conversie van *lead* naar aantal afspraken, is bijvoorbeeld niet nauwkeurig als alleen de aantallen op elkaar worden gedeeld. Uit een gedeelte van de *leads* die nog niet zijn verwerkt, zijn ook nog afspraken te verwachten. Om deze reden is het aantal afgehandelde *leads* van belang bij het bepalen van de conversie ratio's. De werkvoorraden zijn een speciale selectie van *leads*. Daarom past deze in dezelfde dimensiestructuur.

De projectleider kan door het monitoren van de hoogte van de werkvoorraden de processen in de gaten houden en bijsturen wanneer hij of zij dat nodig acht.

Deze oplossing ondersteunt de beslissing, maar heeft twee nadelen:

- Sluit niet helemaal aan bij de gevraagde informatie.
- Biedt geen ondersteuning met historische waarden.

In de volgende paragraaf wordt een structurele oplossing voor de bovenstaande nadelen besproken.

### 7.3.5. Een structurele oplossing

In paragraaf 7.3.4 wordt aangegeven dat de informatie van de juni kubus wel mogelijkheden verschaft voor proces besturing. De daarbij genoemde nadelen zorgen ervoor dat de informatie vraag echter niet in zijn geheel beantwoord wordt.

In feite is er sprake van twee soorten proces monitoring:

- Monitoring van de resultaten
- Monitoring van de procesperformance

Het monitoren van de **resultaten** is van belang bij de geautomatiseerde processen. In het geval van bijvoorbeeld een website, kunnen de bezoekers gericht gestuurd worden naar de "betere" website wanneer de resultaten van de websites bekend zijn.

De (technische) performance van de systemen waarop de website draait, zijn primair van belang voor de systeembeheerder en zijn van veel minder belang voor management en businessanalisten.

Het monitoren van de **procesperformance** is van belang voor processen waar menselijke interactie noodzakelijk is, bijvoorbeeld bij een callcenter. Een manager wil bijvoorbeeld graag weten of efficiënt wordt omgegaan met de *leads*.

Resultaatmonitoring is niet voldoende voor elk project. Een aparte procesperformance-kubus zal in de toekomst moeten worden ontwikkeld zodat de business gebruikers zelf de keuze tussen deze twee soorten monitoring kunnen maken.

## 7.4. Finetuning van de monitoringkubus

De juni-kubus en de toegevoegde waarde ervan zijn besproken in hoofdstuk 6. Een eerste versie van de juni-kubus is opgeleverd. Een aantal wijzigingen zullen nog moeten worden toegevoegd zodat de gebruikers echt moeiteloos de informatie voorgeschoteld kunnen krijgen.

De volgende punten met betrekking tot de juni-kubus zullen op korte termijn moeten worden opgepakt:

- Naamvertaling van herkomstnamen en banners
- Standaardisering via divebooks
- Scheiding in projecten

### 7.4.1. Herkomst vertaling

In paragraaf 1.2 wordt het domein van monitoring besproken; VP heeft meer dan 60 websites, talrijke leveranciers en meerdere afnemers. Het is niet noodzakelijk dat de



verwijzingen naar leveranciers in de database logische namen hebben. Vaak bestaat de verwijzing in database als een gewone code. De naam van de leverancier (lees: de herkomst van de *lead*) Google spreekt direct tot de verbeelding en geeft aan waar de *lead* vandaan komt. De herkomstnaam "LWfd" zegt de meeste gebruikers echter niets. Om te zorgen dat alle herkomstnamen in de kubus voldoende informatie bevatten, zullen alleen logische namen moeten worden gebruikt.

Hetzelfde probleem keert ook terug bij de naamgeving van banners in de kubus. De kubus geeft wel *feedback* over het resultaat van een banner, maar door de moeilijk te interpreteren namen kunnen de gebruikers deze informatie niet meenemen in de beslissingen.

#### **7.4.2.        Standaardisering via *divebooks***

Een interessante view of visualisering van het datamodel kan opgeslagen worden in een zogenaamde *bookmark*. Een *divebook* is in feite een verzameling van de gerangschikte *bookmarkers*. De gebruikers kunnen met een paar klikken bij de *views* komen die ze wensen. Voor beide projecten, WP en oversluiten, is een aantal *views* in de *divebook* bij de eerste oplevering meegeleverd. De standaardisering van de *divebooks* is een iteratief proces. De gebruikers zullen met feedback en nieuwe wensen komen waarmee vervolgens het *divebook* kan worden ge-update.

#### **7.4.3.        Scheiding in projecten**

Niet alle stappen, vermeld in figuur 6-3, zijn voltooid. Vanwege de vervuilde data en de complexiteit van integratie hebben de gebruikers besloten in de toekomst geen PHP-data op te nemen in de analyse. Hierdoor werd de stap van het importeren van PHP-data naar het VP platform, overbodig.

Door tekort capaciteit van de technische afdeling, is ook de *builder* selectie stap niet geïmplementeerd. Hierdoor komen de kubussen van oversluiten en WP in één grote hyperkubus te staan. Oversluiten heeft in tegenstelling tot WP, *measures* over telemarketing. Hierdoor is een hyperkubus ontstaan met alle nadelige gevolgen van dien (zie ook hoofdstuk 4). Dit zal alsnog in een later stadium moeten worden rechtgetrokken.

#### **7.5.        Realiseren van andere algemene kubussen**

Een aantal algemene informatievragen werd geïdentificeerd in paragraaf 6.2.1. Van deze vragen is tot op heden alleen de monitoringkubus gerealiseerd. De overige kubussen zullen in de nabije toekomst moeten worden geïmplementeerd.

## 7.6. OLAP draaiboek

In deze paragraaf wordt een checklist beschreven bedoeld als ondersteuning van de data analyse bij nieuw op te starten OLAP projecten binnen VP. Bij de start van een project zal in de toekomst ruimte moeten worden vrijgehouden voor de OLAP analyses.

In het verleden bestond een directe interactie tussen de techniek en de business bij het invullen van de informatiebehoefte. De communicatie was echter niet optimaal. De business kreeg vaak rapportages die afweken van hun oorspronkelijke wensen. De BI afdeling zal als tussenschakel gaan fungeren bij OLAP-projecten. Teneinde deze rol succesvol in te kunnen vullen, zal de afdeling BI moeten voldoen aan de volgende criteria:

- denken in verschillende abstractie niveaus
- in kunnen schatten van de technische complexiteit van een informatievraag
- gespecificeerd vragen formuleren richting de techniek

Een OLAP-project binnen de VP organisatie kan worden onderverdeeld in de volgende fasen:

- **Inventariseren van de behoeften (*requirements*)**  
De business heeft behoefte aan ondersteuning van haar beslissingen. Lang niet altijd is de business echter in staat deze behoeften op een juiste manier te verwoorden. Een van de taken van de afdeling BI is het ondersteunen van de business bij het formuleren van de behoeften.
- **Opstellen van de business *requirements***  
Deze juist geformuleerde behoeften van de business worden vastgelegd in de business *requirements*. De uiteindelijke versie van dit document wordt in overeenstemming met de business "*bevroren*", zodat deze als basis kan dienen voor het verdere project.  
**Nota bene:** wijzigingen op de "*bevroren*" versie in een verdere fase van het project, hebben doorgaans verstrekkende gevolgen.
- **Opstellen van de functionele specificaties**  
De business *requirements* worden omgezet in functionele *requirements* (lees specificaties) zodat deze te implementeren zijn door de techniek. De functionele specificaties dienen als communicatiemiddel richting de techniek.  
Met het vertalen van de business *requirements* naar de functionele specificaties, wordt de kloof tussen de business en de techniek gedicht.
- **Implementatie**  
De techniek pakt vervolgens het project op en draagt zorg voor implementatie van het gevraagde.
- **Validatie en verificatie van het geleverde**  
In deze fase wordt door de afdeling BI constant getoetst of het geleverde produkt voldoet aan de beoogde doelstellingen zoals vastgelegd in de *businessrequirements*.

- **Onderhoud van het geleverde**

Na acceptatie van het geleverde eindprodukt door de business, zal de techniek zorg dragen voor het dagelijks onderhoud van het produkt. Gewenste wijzigingen op het produkt worden ingediend door de business bij de afdeling BI.

Samengevat is de afdeling BI verantwoordelijk voor:

- het helpen formuleren van de *businessrequirements*
- het vertalen van *businessrequirements* naar functionele specificaties
- het toetsen van het produkt op het beoogde resultaat
- het vertalen van mogelijke onderhoudsbehoeften op het produkt

### 7.6.1.        **Formuleren *businessrequirements***

De BI afdeling legt de *requirements* ten behoeve van de data analyse vast in een document dat in ieder geval de volgende onderwerpen behandelt:

- De doelstelling en het domein van het project.  
Wat wil de business analyseren? Welk type beslissingen wil men ondersteund zien?
- Inventarisatie van de *ins and outs* van het project.
- Kunnen bestaande gestroomlijnde oplossingen gebruikt worden voor het bereiken van de doelstelling?
- Welke informatiebehoeften dienen voor aanvang van het betreffende project gestroomlijnd te worden?
- Welke informatiebehoeften kunnen beter ad-hoc blijven?

### 7.6.2.        **Vertalen *businessrequirements***

Het vertalen van de *businessrequirements* naar de functionele specificaties speelt een cruciale rol binnen een project. De functionele specificaties moeten voldoen aan de volgende criteria [Wieringa – 1996]:

- **Communiceerbaarheid**  
De specificaties moeten met betrekking tot het produkt voor alle belanghebbenden dienen als communicatie kanaal.
- **Waarheid**  
De specificaties dienen alleen de specificaties te beschrijven en niet anders dan dat.
- **Volledigheid**  
De specificaties dienen alle, maar dan ook alle *requirements* te bevatten.
- **Haalbaarheid**  
De specificaties dienen gedrag te beschrijven dat haalbaar is in een product.
- **Verifieerbaarheid**  
Het dient mogelijk te zijn waar te nemen of een produkt voldoet aan de specificaties.

- **Onderhoudbaarheid**

De specificaties moeten onderhoudbaar zijn wanneer *requirements* na oplevering van het produkt veranderen.

De “communiceerbaarheid” en “volledigheid” kunnen worden bereikt met een aantal gebruikte tools voor het Juni-project. Het mini-ster schema in figuur 6-1 is bijvoorbeeld een goede tool voor de communiceerbaarheid. Hierentegen zijn de dimensies en measures van figuur 6-2 een goede tool om de volledigheid van de specificaties te testen.

In paragraaf 3.3 zijn de eisen van OLAP toegelicht. Wanneer aan deze eisen wordt voldaan, is het “haalbaar” een multi-dimensionale weergave te geven ter ondersteuning van de informatiebehoeften. “Verifieerbaarheid” en “onderhoudbaarheid” worden als verantwoordelijkheid van de BI afdeling apart behandeld in de volgende paragrafen.

### 7.6.3. Toetsen beoogde resultaat (“verifieerbaarheid”)

Bij de verifieerbaarheid draait het om het proces van toetsing van het op te leveren product gedurende verschillende stadia van de ontwikkeling. Een aantal toetsingsmomenten moet minimaal worden ingebouwd bij de volgende mijlpalen van de implementatie:

- **Implementatie ETL**

Na de implementatie van ETL is de super tabel te observeren. Hiermee kan de juistheid van de opgenomen dimensies en measures worden geverifieerd.

- **Integratie van externe data**

Bij integratie van externe data wordt de super tabel uitgebreid met extra dimensies en measures. Deze moeten ook worden geverifieerd.

- **Vlak voor de oplevering**

Getoetst moet worden of het geleverde produkt voldoet aan de beoogde doelstellingen zoals vastgelegd in de *businessrequirements*.

Bij deze contactmomenten dienen structureel de doelstellingen te worden getoetst aan de hand van de huidige status en de planning. Afwijkingen hierop moeten tijdig worden gerapporteerd aan het management.

### 7.6.4. Vertalen onderhoudsbehoeften (“onderhoudbaarheid”)

Onderhouds- en veranderingsbehoeften kunnen worden getriggerd door de volgende zaken:

- Een deel van de applicatie sluit niet (langer) aan bij de doelstelling en informatiebehoeften van het bedrijf.
- Verandering in de doelstellingen en de informatiebehoeften van het bedrijf die nog niet worden ondersteund door het huidige systeem.

- Verandering in de scope van het project in het geval er extra leveranciers of afnemers bij komen. De vertaaltabel van herkomst en uitingen moet worden bijgewerkt, zodat de gebruikers de gegevens goed kunnen interpreteren.

Het onderhoud moet traceerbaar zijn; de specificaties moeten altijd een volledige, correcte, en gedetailleerde historie van de aangebrachte veranderingen weergeven.

## **7.7. Samenvatting**

In dit hoofdstuk werd het OLAP-project geëvalueerd. Paragrafen 7.1 tot en met 7.3 evalueerden het juni-project. In paragraaf 7.1 werd de nieuwe momentopname van de informatiesysteem vastgelegd. Aan de hand van de vergelijking van deze momentopname met die van hoofdstuk 5 werd de toegevoegde waarde van het project in paragraaf 7.2 vastgesteld. Paragraaf 7.3 besprak verfijningen die doorgevoerd moeten worden op de opgeleverde kubus van het Juni-project, waardoor deze beter kan aansluiten op de behoeften. Paragraaf 7.5 maakte duidelijk dat het juni-project slechts een deel uitmaakt van het gehele OLAP-project. Een aantal andere kubussen moet nog worden gerealiseerd. In paragraaf 7.6 werd een OLAP draaiboek geïntroduceerd, deze dient als ondersteuning van de data analyse bij nieuw op te starten OLAP projecten binnen VP.

## 8 Knelpunten en aanbevelingen

Een hypothetische vraag: "Hoe zou ik mijn aandacht verdelen als ik dit project over mocht doen?" In hoofdstuk 7 wordt het project geëvalueerd. In dit hoofdstuk staat de aanpak en organisatie van een nieuw op te zetten OLAP project centraal.

Paragraaf 8.1 bespreekt een technische opzet waarmee meer informatievragen kunnen worden opgevangen. Paragraaf 8.2 besteedt aandacht aan de communicatie tussen de business en de techniek. Aangezien al veel boeken verkrijgbaar zijn over dit onderwerp in het algemeen, zal in deze paragraaf alleen de extra inspanning benodigd bij OLAP projecten worden uitgelicht. Paragraaf 8.3 behandelt een aantal aanbevelingen die toepasbaar zijn tijdens de ontwikkeling van een OLAP project. In de samenvatting van dit hoofdstuk zal bovenstaande hypothetische vraag worden beantwoord.

### 8.1. *Aggregatie niveau*

Het aggregatieniveau<sup>6</sup> van de supertabel is te hoog. Deze is namelijk gebouwd op statistisch niveau. Een *drill-down* naar de *leadID*'s is niet mogelijk. OLAP kan een hoger aggregatie niveau weergeven in *views*, maar kan niet een lager aggregatie niveau dan dat van de onderliggende data weergeven. De opbouw van een dergelijke supertabel dient eigenlijk vanaf *lead* niveau te zijn aangezien de gebruiker vaak naar dit niveau een *drill-down* wil kunnen uitvoeren. De huidige supertabel schiet wat dit betreft te kort.

### 8.2. *Communicatie*

Communicatie binnen een OLAP project is van essentieel belang aangezien OLAP-project vaak politiek gevoelig liggen. Gebruikers kunnen meebepalen wat de dimensies en de *measures* zijn. Daarnaast heeft iedere gebruiker zijn eigen idee hoe de best mogelijke OLAP kubus eruit moeten zien. Voor het succes van het project is de steun van de gebruikers onmisbaar.

Door de eisen die OLAP stelt aan de onderliggende data, kent OLAP een lang ontwikkelpad. Vaak zal bij de initiële levering van de kubus nog niet alles naar wens zijn. Er moet voortdurend gekeken worden of de kubus (tijdens de ontwikkeling) aan de verwachtingen van de gebruikers voldoet. De gebruikers dienen tijdig op de hoogte te worden gesteld wanneer de ontwikkeling afwijkt van de verwachtingen. Om de verwachtingen in kaart te brengen en door te spelen naar de techniek, moet veel aandacht worden besteed aan de communicatie. Bij een OLAP project kan tijdens de ontwikkelfase van de kubus gecommuniceerd worden in de termen van dimensies en *measures*.

---

<sup>6</sup> Deze term wordt uitgelegd in hoofdstuk 4.

### **8.3. Multi-dimensionaal besef bij de techniek**

Technici denken vaak in een genormaliseerde vorm van tabellen omdat ze gewend zijn te werken met relationele databases. Op een genormaliseerde tabel kan geen OLAP analyse worden gebaseerd<sup>7</sup>. Eenmaal tijdens mijn stage werd bijvoorbeeld externe data geanalyseerd in samenwerking met de leverancier van het OLAP-pakket. Een technicus gaf de voorkeur aan genormaliseerde data omdat het technisch makkelijk te leveren is en volgens hem dezelfde informatie levert.

Genormaliseerde data kan wel in het OLAP pakket worden gestopt, maar daarmee kan geen multi-dimensionale *look and feel* worden bereikt. In appendix J is een kopie van de correspondentie met deze technicus opgenomen. De dimensionale denkwijze is essentieel voor de technici bij de communicatie in termen van dimensies en *measures*.

### **8.4. Multi-dimensionale besef bij de business**

De businessgebruikers kennen vaak geen multi-dimensionale denkwijze; ze zijn gewend aan de beperkte representaties in MS-Excel en in platte tabellen. Het leren omgaan met *slices* en een *cross*-dimensionale kijk op hun informatie is niet alleen belangrijk om de tool te kunnen beheersen, ook zullen ze hierdoor beter hun wensen kunnen formuleren.

Tijdens het project rees de volgende informatie vraag:

*“Wat is de conversie ratio van leads naar afspraak deze week?”*

De conversie ratio is gedefinieerd als het aantal afspraken gedeeld door het aantal leads van een bepaalde week.”

In het tijdperk Excel verwachtten de gebruikers niet anders dan deze niet nauwkeurige weergave van de informatie. De gebruiker wil gewoon een indicatie krijgen van het verloop van het proces. De afspraken die bijvoorbeeld deze week worden gemaakt, kunnen voortkomen uit de *leads* van twee weken daarvoor. In platte tabellen is het onmogelijk dit uit te drukken. De gebruikers gaan af op niet nauwkeurige ratio's. De monitoringkubus, beschreven in hoofdstuk 6, geeft deze ratio's wél goed weer.

De business gebruikers moeten overzicht krijgen welke mogelijkheden multi-dimensionale analyse hen biedt.

### **8.5. Capaciteit**

Anderhalve maand lang was er geen capaciteit voor het realiseren van ETL naar de super tabel. De super tabel is de bron van de OLAP kubussen. Hierdoor moesten veel resource-verslindende ad-hoc *queries* worden gedraaid . Vervolgens werd de data van deze *queries* handmatig in een kubus gestopt. Als het juni-project anderhalve maand eerder

---

<sup>7</sup> Zie hoofdstuk 3

was opgeleverd, had op dit moment een geperfectioneerde kubus ontwikkeld kunnen zijn die de gebruikers beter had kunnen bedienen.

De reden van de vertraging was dat de technische afdeling andere projecten binnenkreeg die een hogere prioriteit kregen van het management. Hierdoor konden zij geen tijd vrij maken voor het stroomlijnen van de informatievoorziening.

Vaak lopen IT-projecten uit. Er is kennelijk iets mis met de communicatie over de projectplanning. De techniek is verantwoordelijk voor het realiseren van een produkt binnen een bepaalde tijdsperiode. De business op haar beurt is verantwoordelijk voor de input, namelijk de *businessrequirements*, de resources en voor een deel ook voor de prioriteiten.

De resources van de techniek zijn een tijdsbepalende factor voor het op te leveren eindprodukt. De techniek dient goed onderbouwd aan te geven waar eventuele bottlenecks in de resources worden verwacht. Als hierover structureel wordt gecommuniceerd, zullen in het geval van een verwachte afwijking maatregelen moeten worden getroffen. In het geval dat er onvoldoende capaciteit bij de techniek beschikbaar is, krijgt het management de keuze:

- de doorlooptijd van het project te verlengen
- de *requirements* af te zwakken
- de prioriteit ten opzichte van andere projecten te veranderen
- meer capaciteit beschikbaar te maken.

## **8.6. Aanbevelingen**

In de voorgaande paragrafen zijn de knelpunten besproken. Hieronder worden aanbevelingen gedaan om het effect van de knelpunten in het gebruik van OLAP te reduceren.

- **Creëren van een BI-functie**
  - o Tussenschakel van de business en techniek in de BI-projecten.
  - o Continue bewaking over de technische randvoorwaarden ten behoeve van de analyse.
  - o Monitoren en advieseren over het aggregatie niveau van de super tabel.
  - o Trainingen geven aan de business en de techniek op het gebied van multi-dimensionale communicatie.
- **Oppakken van een incrementele werkwijze**

In het geval van de juni-kubus veroorzaakte de oplevering van de kubus nieuwe vragen en wensen aan de businesskant. Met het feit dat nieuwe vragen en wensen ontstaan na oplevering, moet tijdens de ontwikkeling rekening worden gehouden. Een modulaire aanpak kan hieraan bijdragen.
- **Mimaliseren informatieverlies**

De onderliggende supertabel moet op een zo laag mogelijk aggregatieniveau



worden opgebouwd. Hierdoor wordt minder informatie verloren tijdens het aggregeren.

- **Managen van verwachtingen**

Binnen VP lopen veel projecten uit. Een verbetering op de huidige werkwijze kan bestaan uit het beter managen van de verwachtingen. De techniek moet de consequenties van de beslissingen die de businessgebruiker oplegt, goed in kaart brengen. De business wordt daardoor bij het nemen van beslissingen gedwongen de juiste overwegingen te maken. Wanneer de gebruiker op een later stadium van de ontwikkelfase bijvoorbeeld wijzigingen in de *requirements* wil doorvoeren, zal het de gebruiker duidelijk worden welke consequenties dit heeft voor bijvoorbeeld doorlooptijd en zijn andere projecten, uitgaande van een gelijkblijvende capaciteit van de techniek.

- **Structurele technische capaciteit voor BI alloceren**

BI is op te splitsen in BI-infrastructuren en BI-analyses. Een deel van de infrastructuur is gerealiseerd met het juni-project. BI kan geen resultaten van de analyses leveren zonder geschikte data huishouding. Onderhoud op bestaande infrastructuur moet een structurele plaats hebben binnen de werkzaamheden zodat de infrastructuur goed blijft aansluiten op de behoeften. Het **niet** structureel toewijzen van voldoende technische capaciteit ter ondersteuning van de verantwoordelijkheden van de afdeling BI, zal de efficiency van de afdeling BI zéér nadelig beïnvloeden.

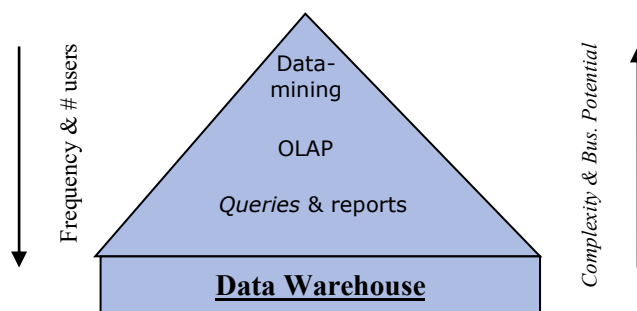
## **8.7. Samenvatting**

In de paragrafen 8.1 tot en met 8.5 zijn een aantal knelpunten behandeld. Dit mondde uit in een aantal aanbevelingen richting VP in paragraaf 8.6. Gewapend met deze inzichten kan ik de hypothetische vraag beantwoorden: "Hoe zou ik mijn aandacht verdelen als ik dit project over mocht doen?":

- Betere allocatie van capaciteit van de techniek
- Verhogen van het multi-dimensionale besef bij zowel de techiek als bij de business
- Inbouwen van meer toetsingsmomenten

## 9 Theorie versus praktijk

In het algemeen is de OLAP theorie, die besproken is in de hoofdstukken 2 tot en met 4, goed in de praktijk toe te passen. Zelf heb ik moeite een overzicht te krijgen van de plaatsing van de OLAP analyses binnen BI.



figuur 9-1 BI piramide

### Complexiteit en *businesspotential*

De rechter pijl duidt in bovenstaande afbeelding de rangschikking van de BI onderdelen aan naar complexiteit en *businesspotential*.

- De *queries & reports* staan onderaan in de rangschikking en hebben derhalve de minste complexiteit.
- Een OLAP kubus (midden in de piramide) is eigenlijk een multi-dimensionale "*look and feel*" van een verzameling *queries*. Dit is logischerwijs complexer dan de *queries & reports* laag.
- Een *data mining* project (bovenaan in de piramide) waarbij "uit het niets" naar verbanden binnen data wordt gezocht, is op zijn beurt weer complexer dan een OLAP project. *Data mining* beantwoordt maar één tot vier procent van de totale informatievragen. Dit kan voor veel organisaties net de "extra" winst inhouden. Het is daarom niet ondenkbaar dat men *data mining* ziet als het BI onderdeel met de meeste potentie.

De rangschikking van de **complexiteit** is in mijn ogen correct. Het ***businesspotential*** is naar mijn mening niet te rangschikken met de door de business gebruikte tools. Het *businesspotential* van de verschillende BI instrumenten is altijd afhankelijk van de situatie waarin de organisatie zich bevindt. Alleen met een goede kosten- en baten afweging is het mogelijk deze rangschikking te maken.

Het standaardiseren van OLAP processen heeft **voor VP** in de huidige situatie een hoger *businesspotential* dan het uitvoeren van relatief riskante *data mining* projecten. Daarmee gaat de rangschikking binnen de piramide voor de rechter pijl (complexity & ***businesspotential***) niet meer op!!!

Let wel: het is niet ondenkbaar dat voor andere organisaties louter ad-hoc queries voldoende kunnen zijn als informatievoorziening. Het ontwikkelen van een data warehouse kost namelijk naar verhouding veel geld.

### **Frequentie en aantal gebruikers**

De linker pijl in figuur 9-1, duidt in bovenstaande afbeelding op de rangschikking van BI onderdelen naar frequenties en het aantal gebruikers. Ik ben van mening dat deze pijl niet realistisch is.

*Data mining* (bovenaan in de piramide), beantwoordt maar één tot vier procent van de totale informatievragen en heeft daardoor de laagste gebruiksfrequentie van de drie BI tools.

OLAP-analyses zijn goed toepasbaar voor het:

- representeren van dynamische *views*
- automatiseren van de standaard informatiebehoefte
- 

Idealiter moeten informatievragen die op regelmatige basis terugkeren, worden opgevangen met OLAP-analyses. Hier gaat de 80%-20% vuistregel op; gestreefd moet worden 80% van de informatievragen te automatiseren. 20% mag blijvend op basis van ad-hoc en op basis van *data mining* worden ingevuld.

(ad-hoc) *Queries & reports* moeten juist dat gedeelte van de informatievragen opvangen waar het standaardiseren niet zinvol is (zie paragraaf 2.2.2). Niet zinvol is het bijvoorbeeld een informatievraag die maar eens per jaar voorkomt te standaardiseren in een OLAP kubus. Dit kost onevenredig veel resources.

## 10 Conclusie

### 10.1. Samenvatting stage resultaat

Mijn stage heeft voor VP de volgende hoofdzaken gerealiseerd:

- Afspraken met de techniek met betrekking tot de technische randvoorwaarden voor het gebruik van OLAP (*data warehouse*)
- OLAP kubus voor de projecten WP en oversluiten
  - o Meetbaarheid van de *lead*-kwaliteit
  - o *Feedback* op kwantiteit en kwaliteit van de genomen beslissingen

De volgende tabel geeft de status van de te realiseren sub-doelen en/of succescriteria.

	Sub-doel en/of succescriteria	Status	Toelichting
1	Opstellen OLAP draaiboek	bereikt	Zie beschrijving “OLAP draaiboek” in paragraaf 7.6.
2	Handmatig leveren van meerdere OLAP kubussen	bereikt	
3	Herkennen van de algemene informatie en deze standardiseren	gedeeltelijk bereikt	Beschreven in hoofdstuk 6. De voor de projecten WP en oversluiten opgeleverde OLAP kubus, voldoet aan één van de drie algemene informatiebehoeften; de monitoringbehoefte.
4	Kwantificeren van de <b>interne</b> toegevoegde waarde	gedeeltelijk bereikt	Tijdbesparend voor de businessanalisten is gekwantificeerd. Niet gekwantificeerd is dat dit ook leidt tot het nemen van betere beslissingen.
5	Kwantificeren van de <b>externe</b> toegevoegde waarde	niet bereikt	Benodigde data is niet door externe partij geleverd.
6	Bijdragen aan opzet van data opslag	gedeeltelijk bereikt	Randvoorwaarden geschapen voor de analyses (zie paragraaf 6.1). Aggregatieniveau van de supertabel was te hoog.

## **10.2. Antwoord op het centrale vraagstuk**

Het centrale vraagstuk van mijn onderzoek luidde: "Hoe kan de *knowledge gap* worden verkleind met Business Intelligence (BI); in het bijzonder met gebruikmaking van OLAP?"

Deze scriptie heeft aangetoond dat de OLAP analyses door het verschaffen van inzichten (dynamische *feedback* en *feedforward views*), de *knowledge gap* kunnen verkleinen. Voorheen konden deze inzichten onmogelijk worden weergegeven.

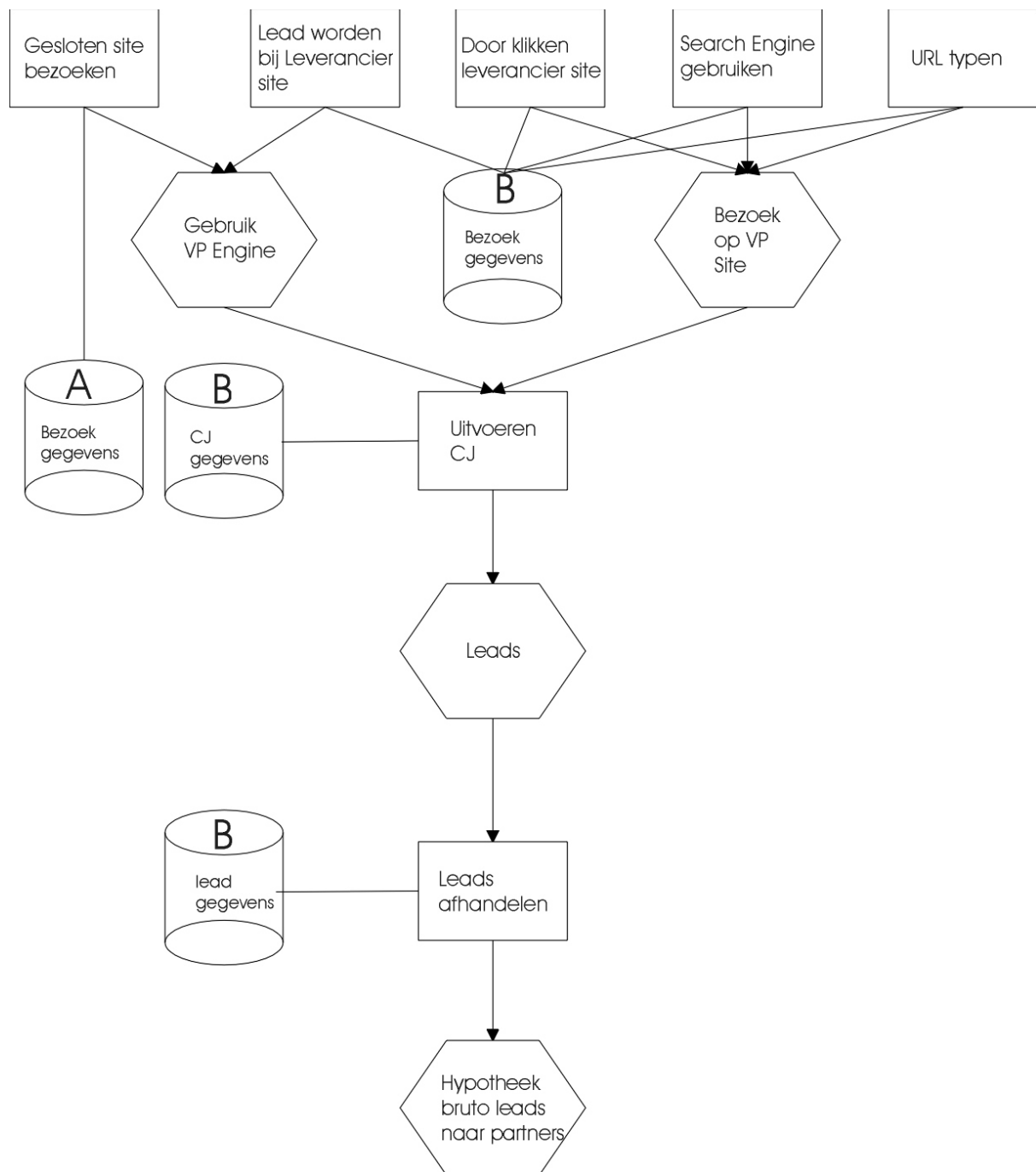
### ***Feedback***

Door middel van het plaatsen van de stuurparameters (beslissingen) en het bedrijfsresultaat (effect van de beslissingen) in respectievelijk de dimensies en de measures, wordt een OLAP kubus geschapen. Daarmee wordt de *feedback* gerealiseerd op de beslissingen in de dimensies en de combinaties daarvan. Een dergelijke OLAP kubus is binnen VP als produkt van mijn stage geïmplementeerd. De toegevoegde waarde van deze kubus is beschreven in hoofdstuk 7.

### ***Feedforward***

OLAP kan ook *feedforward* beslissingen ondersteunen. Deze functionaliteit is onderdeel van de overige algemene kubussen; deze zijn zoals eerder beschreven, nog niet geïmplementeerd.

## Appendix A. Proces en data bij leadgeneration



## Appendix B. Doelstelling van de opdrachtgever

Hoofddoel van dit project is de toegevoegde waarde van data analyse voor VP te verhogen. Dit moet vanuit twee invalshoeken benaderd worden:

- a) Operationele bedrijfsvoering d.w.z. data analyse door en voor VP medewerkers en
- b) Dienst voor derden, d.w.z. door VP medewerkers voor hun zakelijke partners.

De technische kant heeft drie lagen:

1. data opslag
2. on-line analytical processing (OLAP)
3. data mining

### Ad 1. Hoofddoel: optimaliseren van data opslag en data stromen.

In 2001 hebben twee afstudeerders een data warehouse (DW) gebouwd. Sindsdien zijn er heel veel gegevens bijgekomen, de uitbreidingen zijn steeds ad hoc en pragmatisch opgezet. De taak hier is om de huidige data repository te inventariseren en tegen de DW theorie te houden. Op basis van deze analyse moeten drie hoofdvragen beantwoord worden.

- Heeft de huidige data repository alle voor VP relevante functies van een data warehouse?
- Zo niet, welke dingen kunnen we nu niet maar met een echt DW wel?
- Hoe zouden de data opslag en de data stromen verbeterd kunnen worden?

### Ad 2. Hoofddoel: maximaliseren van de toegevoegde waarde van VP's OLAP pakket.

VP beschikt over een OLAP pakket. Hoe kan VP de maximale waarde uit dit pakket halen: mogelijkheden van het pakket tegen de (al dan niet latente) behoeften van de zakelijke gebruikers houden. Meehelpen met het opzetten van een server-client architectuur om toegang voor derden aan te kunnen bieden.

### Ad 3. Hoofddoel: goed onderbouwd advies over de mogelijke waarde van data mining.

- Business kant: zijn er businessvragen die (alleen) met data mining opgelost kunnen worden? Wat is de geschatte inspanning - waarde verhouding?
- Technische kant: In 2001 hebben twee afstudeerders een data mining algoritme (SPRINT) geïmplementeerd. SPRINT is gemaakt met het oog op zeer grote hoeveelheden data. Beken moet worden of de huidige databestanden ook door WEKA (zeer uitgebreid gratis data mining Java library) ge-mined kunnen worden of is SPRINT echt nodig. Als WEKA de data aankan, moet een selectie gemaakt worden uit de daarin aanwezige methodes die voor VP's businessvragen relevant zijn.
- "Proefproces": een data mining traject uitvoeren om het proces (eventuele knelpunten!) in kaart te brengen.

## Appendix C. Tijdsplanning stage

week nummer	doel 1	doel 2	doel 3	doel 4 wagner	doel 4 oversluter	doel 5	doel 6
week 8		cubussen					verg. disc. agenda opties
week 9							
week 10				Wagner			
week 11							
week 12							
week 13							
week 14						req.	
week 15			Alg. Proc. Besch.		proc. Besch.		
week 16				analyse wagner oud.			rapport
week 17							
week 18		cubussen					
week 19		beschrijv.					
week 20						analyse	
week 21			Arrange cub.	analyse wagner nieuw	analyse oversluter		
week 22							
week 23							
week 24							
week 25	afsluiten rapportages over de uitvoering van de doelstellingen.						
week 26							
week 27	alg.						
week 28	blauwdruk						
week 29	stage verslag						
week 30							
week 31	uitloop van de verschillende projecten, in verband met de risico's						
week 32							
week 33							

### Toelichting

#### Doel 1:

Voor doel 1 heb ik in ieder geval resultaten van doel 2 tot en met doel 5 nodig. Daarom is het echte werk pas gepland op week 27, nadat ik over de afsluitend rapportage heb nageacht van de activiteiten voor de andere doelstellingen.

#### Doel 2:

Voor doel 2 zijn er al in het begin van mijn stageperiode een aantal OLAP-kubussen gemaakt. De ervaring van de uitvoering leidt tot een aantal voorwaarden die nodig zijn voor de succesvolle tegemoetkoming van de informatiebehoeften. Op dit moment mist er een belangrijke voorwaarde van de OLAP kubussen, dat is een goede datastructuur die de dimensionale analyses ondersteunt. Technisch afdeling belooft eind april het nieuwe platform gereed te maken. Daarom ga ik in week 18 pas verder. De verwachting is dat de data opslag van nieuwe stijl moet heel snel (binnen 8 uur) een OLAP kubus leveren. De leereffect zit met name op de businesskant. Nadat de kubussen klaar zijn is er nog een week voor gereserveerd om de kubussen goed te documenteren. Dit draagt ook bij doel 1.

#### Doel 3:

Met doel 3 moet ik de standaard informatiebehoeften herkennen. Dat kan ik doen door middel van het proces van VP goed in kaart te brengen en vanuit de processen de informatiebehoeften te herkennen. De output van doel 2 (afhankelijk van aantal



kubussen) kan als extra check dienen voor dit doel. Daarom ga ik in week 20 nadat ik de output heb van doel 2 gelijk de kubussen rangschikken. Daarmee kan ik doel 3 afsluiten.

#### *Doel 4:*

Voor doel 4 ga ik 2 projecten binnen VP bekijken, en daarmee de enigszins uitspraak doen over de toegevoegde waarde van de data analyse. Daarom is in de planning toegespitst in planning over project WP en project oversluiter. WP project was begonnen voordat ik bij VP begon. Project oversluiter ben ik vanaf begin erbij betrokken geweest.

#### *WP:*

Voor het maken van de kubussen is er in week 10 tot en met week 12 de data *requirement* over gestuurd over de WP project. Maar de rapportage is niet succesvol afgerond door de slechte data kwaliteit van de PHP-platform (bv. een bezoeker is niet uniek identificeerbaar). Inmiddels is de leiding mee eens dat ik alleen analyses uit kan voeren als de data een redelijk kwaliteit heeft. De techniek gaat in de komende weken het nieuwe platform bouwen. Dat wil zeggen over een paar weken zal ik goede data hebben. In week 21 ga ik de WP data in nieuwe platform analyseren. Week 16 zal ik de eerste terugkoppeling van de WP data krijgen. Ik heb week 16 en 17 gepland om toch in de WP externe data te kijken, hoewel het koppelen met interne data (door slechte kwaliteit) niet mogelijk is. Mijn doelstelling is om toch een uitspraak te doen over de analyses. De vraag is: welke deel van de data en hoe moeten we de externe WP data integreren in onze nieuwe platform?

#### *Oversluiter:*

VP's andere relevant project is oversluiter. Hiervoor wordt er een externe projectmanager ingehuurd. Ik ga de projectmanager helpen met de procesbeschrijving. Dat moet deze week gerealiseerd worden. Dan wacht ik aantal weken tot ik data krijg van oversluiter, dan begin ik de data te analyseren. Daar trek ik 4 weken vooruit. Het analyseren van de data loopt parallel met de WP analyses.

#### *Doel 5:*

Voor doel 5 ga ik alleen een project van VP uitwerken, dat is Davinci project. In dit project moet VP een rekenmodule analyseren. De informatie en data *requirement* is al geformuleerd en opgestuurd naar het externe bedrijf (Davinci). Een andere bedrijf (COOP) zorgt ervoor dat de data in kubussen komen. Nadat ik de kubussen krijg ga ik een analyse uitvoeren en dat afsluiten met een rapport. Analyse wordt gepland op week 20.

#### *Doel 6:*

Voor doel 6 zijn er veel discussies, vergaderingen en agendapunten uitwisselingen geweest binnen VP. Ik realiseer me dat dit de basis is voor mijn stage. Ik kan mijn stage alleen goed afsluiten wanneer er goede onderliggende dataopslag is die voor de kwalitatief goede data zorgt. Dit doel is afgesloten met een vergadering met een externe adviseur. Op dit moment implementeert de technisch afdeling de opzet die afgesproken is. Volgende week (16) ga ik dit doel afsluiten met een rapportje.

*Rapportage fase:*

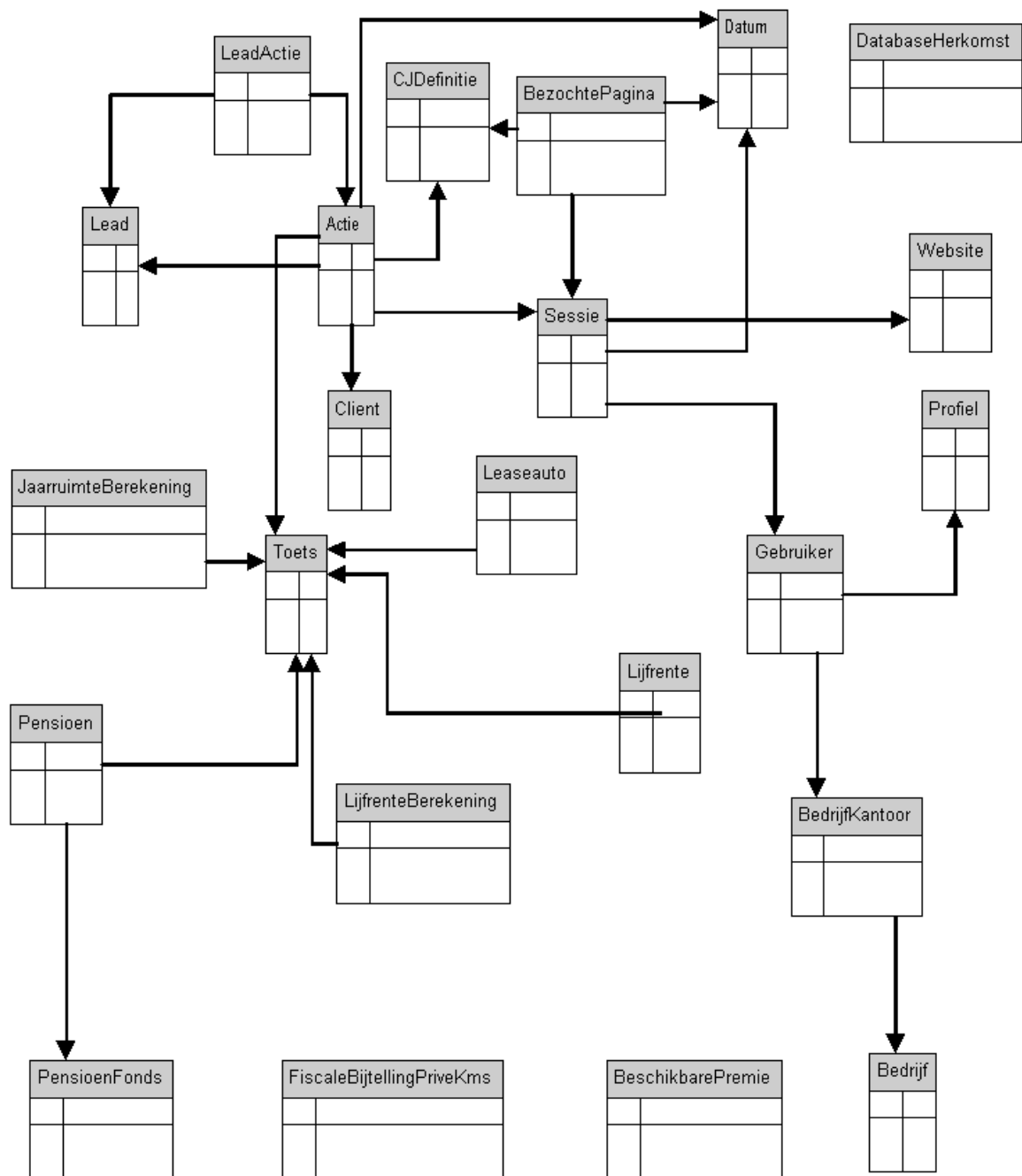
Week 25 en 26 is gepland voor rapportages van doel 2 t/m 5. Twee weken heb ik nodig om alles te structureren en de uitloop van de analyse fase op te vangen.

Week 29 en 30 is bedoel om de

## Appendix D. Ster-schema oorspronkelijk VP platform



## Appendix E. Schema tabellen VP platform



## Appendix F. Monitoringkubus WP en Oversluiten

Kubus monitoring van het project WP		
VP data		
Dimensies:	Measures:	Afgeleide Measures:
Bezoek Tijd	Aantal bezoekers	Conversie van bezoekers naar bruto <i>leads</i>
Herkomst site	Aantal bruto <i>leads</i>	
Herkomst uiting		
Herkomst groep		
CJ naam		
Site VP		
<i>Leads</i> afnemer		

Inclusief WP externe data		
Dimensies:	Measures:	Afgeleide Measures:
Adviseur	Aantal netto <i>leads</i>	Gerealiseerde conversie bezoekers naar bruto afspraken
Vestiging	Aantal bruto afspraken	Verwachte conversie bezoekers naar bruto afspraken
	Aantal netto afspraken	Gerealiseerde conversie <i>leads</i> naar bruto afspraken
	Aantal transacties	Verwachte conversie <i>leads</i> naar bruto afspraken
	Aantal keer gebeld	Aantal <i>leads</i> nog niet afgehandeld
	Aantal <i>leads</i> in behandeling	Aantal <i>leads</i> nog niet ingevoerd in " <i>leads</i> en bel"
	Aantal bruto afspraken in beh.	Gerealiseerde conversie bezoekers naar netto afspraken
		Verwachte conversie bezoekers naar netto afspraken
		Gerealiseerde conversie <i>leads</i> naar netto afspraken
		Verwachte conversie <i>leads</i> naar netto afspraken
		Gerealiseerde conversie bezoekers naar transactie
		Verwachte conversie bezoekers naar transactie
		Gerealiseerde conversie <i>leads</i> naar netto transactie
		Verwachte conversie <i>leads</i> naar netto transactie
		Gerealiseerde conversie netto afspraken naar netto transactie
		Verwachte conversie netto afspraken naar netto transactie

Basis		
Dimensies:	Measures:	Afgeleide Measures:
Bezoek Tijd	Aantal bezoekers	Conversie van bezoekers naar bruto <i>leads</i>
Herkomst site	Aantal bruto <i>leads</i>	
Herkomst uiting		
Herkomst groep		
CJ naam		
Site VP		
<i>Leads</i> afnemer		

Telemarketing		
Dimensies:	Measures:	Afgeleide Measures:
Afspraken afnemer	Aantal netto <i>leads</i>	Conversie van bezoekers naar netto <i>leads</i>
	Aantal bruto afspraken	Conversie van bruto <i>leads</i> naar netto <i>leads</i>
	Doorlooptijd callcenter	Conversie van bezoekers naar bruto afspraken
		Conversie van bruto <i>leads</i> naar bruto afspraken
		Conversie van netto <i>leads</i> naar bruto afspraken
		Aantal <i>leads</i> in voorraad

Intermediair		
Dimensies:	Measures:	Afgeleide Measures:
	Aantal netto afspraken	Conversie van bezoekers naar netto afspraken
	Aantal transacties	Conversie van bruto <i>leads</i> naar netto afspraken
	Doorlooptijd geheel	Conversie van netto <i>leads</i> naar netto afspraken
	Aantal afzeggingen afspraken	Conversie van bruto afspraken naar netto afspraken
	Aantal afspraken niet nagekomen	Conversie van bezoekers naar transactie
		Conversie van bruto <i>leads</i> naar transactie
		Conversie van netto <i>leads</i> naar transactie
		Conversie van bruto afspraken naar transactie
		Conversie van netto afspraken naar transactie
		Aantal afspraken in voorraad

## Appendix G. Data voor monitoringkubus WP en Oversluiten

### Data voor het monitoren van het project *oversluiten*

Basis				
Brondata & procedure	Definities		Voorwaarden	Datum gereed
Data uit de VP webserver, De basis ETL	<b>Naam attribuut</b>	<b>Opmerking</b>	VP server data OK	?
	Bezoek Datum		basis ETL OK	
	Herkomst uiting			
	Herkomst site			
	Herkomst groep			
	CJ naam	Naam van de customer journey		
	Site VP	Bv: uwhypotheekoversluiten.nl		
	Lead Afnemer	In het geval <i>oversluiten</i> is het altijd Robert (tot dus ver).		
	Sessie ID	Een unieke sessie ID is per definitie een bezoeker.		
	LeadID	Bezoekers die <i>leads</i> worden, krijgen een <i>leadID</i>		

Telemarketing				
Brondata & procedure	Definities		Voorwaarden	Datum gereed
Data uit de besloten site voor telemarketing,	Naam attribuut	Opmerking		
ETL besloten site telemarketing	Inventarisatie datum		telemarketing DB OK	?
	Inventarisatie duur		ETL telemarketing OK	
	Acceptatie datum	Voorlopig gebeurt inventariseren en accepteren tegelijk, de datum en duur zijn niet te scheiden in de brondata.		
	Acceptatie duur			
	Geinventariseerd (ja/nee)	De klant nagebeld, de gegevens gecontroleerd samen met de klant		
	Geaccepteerd (ja/nee)	Nadat geïnventariseerd is, kan de adviseur een offerte uitbrengen (ja), anders (nee).		
	Resultaat telemarketing	De uitkomst van de callcenter benadering. Zoals: Afspraak, niet te bereiken, valse <i>leads</i>		
	Afspraken afnemer	Welke intermediair krijgt de afspraak, B&B of D&S		



<b>Intermediair</b>				
<b>Brondata &amp; procedure</b>	<b>Definities</b>		<b>Voorwaarden</b>	<b>Datum gereed</b>
Datadump van intermediair BB,	<b>Naam attribuut</b>	<b>Opmerking</b>		
Datadump van intermediair D&S,	Benader datum	Intermediair belt de klant om afspraak te maken, datum benadering door intermediair.	ETL BB OK	?
ETL intermediair	Afspraak datum	Datum van geplande afspraken door intermediair.	ETL D&S OK	
	End resultaat	Uitkomst intermediair, bv: afzegging, offerte getekend etc.	Datadump BB regelmatig binnen	
			Datadump D&S regelmatig binnen	

## Data voor het monitoren van het project WP

Basis				
Brondata & procedure	Definities		Voorwaarden	Datum gereed
Data uit de VP webserver, De basis ETL	<b>Naam attribuut</b>	<b>Opmerking</b>	VP server data OK	?
	Bezoek Datum		basis ETL OK	
	Herkomst uiting			
	Herkomst site			
	Herkomst groep			
	CJ naam	Naam van de customer journey		
	Site VP	site in beheer van VP, bv: hypotheek.cc		
	<i>Lead</i> Afnemer	In dit geval WP		
	Sessie ID	Een unieke sessie ID is per definitie een bezoeker.		
	<i>Lead</i> ID	Bezoekers die <i>leads</i> worden, krijgen een <i>lead</i> ID		

WP				
Brondata & procedure	Definities		Voorwaarden	Datum gereed
Data uit WP "lead en bel" systeem,	Naam attribuut	Opmerking	VP server data OK	?
ETL WP externe data	Vestiging	De vestiging van WP die de klant behandelt	basis ETL OK	
	Adiseur	De naam van de adiseur van WP die de klant behandelt	ETL WP&Partners OK	
	Bel datum	De datum waarbij de klant wordt opgebeld		
	Aantal keer bellen	De klant wordt aantal keer benaderd als ze geen gehoor geven.		
	Herkomst voor WP	Lead en bel gehanteerd herkomst, bv: vermogensplanet.		
	Bel resultaat	De uitkomst van de callcenter benadering. Zoals: Afspraak, niet te bereiken, valse leads		
	Afspraak resultaat	De uitkomst van de benadering van de adviseur: offerte getekent, geen intresse na gesprek, etc		

Eenmalige koppeling PHP-data				
Brondata & procedure	Definities		Voorwaarden	Datum gereed
PHP-databases Goedkoop Hypotheek en Hypotheek rente met WP "lead en bel",	Naam attribuut	Opmerking		
Eenmalige koppeling van de data uit php-sites met "lead en bel" zodat deze ook gevolgd kunnen worden in de toekomst.	Lead Datum	Datum de bezoeker lead wordt, op php-site is overigens geen bezoeker uniek te identificeren	zowel VP als WP moet effort instoppen, de stappenplan uit bijlage uitvoeren.	?
	Herkomst site	Bestaat niet bij GH database		
	Herkomst groep	Bestaat niet bij GH database		
	CJ naam	CJ naam mits bestaan, soms de naam van de rekenmodule		
	Site VP	De site onder beheer van VP waar de klanten leads worden. (bv hypotheek.cc, GH, HR)		
	Lead Afnemer	In dit geval WP		
	LeadID	Lead ID van php databasen		
	attributen uit "lead en bel"	Attributen uit "lead en bel" zijn ongewijzigd, zie tabel van WP		

## Appendix H. Status schema juni-project

Actie	Subacties/voorwaarden	Geschatte resterende inspanning	Verandwoordelijke persoon	% Gereed	Oplever datum
1	A. VP server data OK B. Basis ETL implementeren + testen	b) 40uur	a) Rene b) Rene	A 100% B 20%	14/06
2	Telemarketing site voor oversluiten klaar A. Telemarketing DB bevat juiste attributen B. Oude gegevens Robert worden overbracht in de telemarketing DB C. ETL implementeren	b) 1uur c) 12 uur d) 8uur <sup>8</sup>	a) Rene+ Robert b) Rene + Han Long c) Ilona d) Rene	A. 100% Rene 0% Robert B 0% C 0% D 0%	
3	A. Regelmatige data levering van WP B. Attributen WP passen bij de informatiebehoefte C. Data import implementeren D. Basis ETL aanpassen	b) < 2 u c) 4 uur d) 8 uur	a) Marc b) Han Long c) Rene d) Rene	A 0% B 70% C 0% D 0%	
6	A. Regelmatige data levering intermediairs B. Attributen intermediairs passen bij de informatiebehoefte C. Data import implementeren D. Basis ETL aanpassen	b) < 2 u c) 4 uur d) 8 uur <sup>1</sup>	a) Ilona b) Han Long c) Rene d) Rene	A 90% B 70% C 0% D 0%	
4+5	A. Automatisch script om builder aan te roepen. B. Selectiecriteria gedefinieerd	a) 4 uur b) < 2 u	a) Rene b) Han Long	A 0% B 80%	
7	A. Diveline installeren B. Server via internet te bereiken	a) 1 uur	a) Rene b) Rene	A 0%	
8	A. Gebruikers specificatie B. Implementatie C. Documentatie	a) 4 uur b) 1-2 dagen c) 4-6 uur per project	a) Han Long + gebruiker b) Han Long c) Han Long	A 20% B 20% C 0%	
9	A. Cliënt WebDive installeren bij business gebruikers B. Gebruikers trainen	b) 16-24 uur uitgesmeerd	a) gebruiker mbv HanLong b) Gusztí + Han Long	A 0% B 0%	b) eind juni
10	Zie Hoofdstuk “Eenmalige koppeling oude PHP data”	Na juni?		0%	

<sup>8</sup> Zonder de “doorlooptijd” mee te nemen, met “doorlooptijd”: 12 uur

## Appendix I. Takenlijst Juni-project

1. Verifiëren of WP *lead* en bel dump aansluit bij informatiebehoefte.
  - a. Datamodel Suptertabel vs. Kubus WP
  - b. Inhoud WP datadump vs. Supertabel
  - c. Rapport van a en b
2. Verifiëren of de telemarketing DB aansluit bij informatiebehoefte.
  - a. Datamodel TM DB vs kubus *oversluiten*
  - b. Datamodel SuperTabel van telemarketing vs. Kubus
  - c. Inhoud SuperTabel van telemarketing vs kubus
  - d. Rapport a, b en c
3. Verifiëren of data van intermediairs aansluit bij informatiebehoefte.
  - a. Exploren data van intermediairs.
  - b. Verzoek op aanpassingen van intermediairs data.
  - c. Verifiëren of de aangepaste data van intermediair wel of niet aansluit bij informatiebehoefte.
4. Testen supertabel (al of niet externe gegevens afhankelijk van de progressie van de techniek.)
  - a. Datamodel SuperTabel vs. kubussen
  - b. Inhoud SupterTabel (en ETL) vs. Kubussen
  - c. Rapportage over a en b
5. Diveline of Divebook standaardiseren naar wensen van de gebruikers.
  - a. Vertaling datamodel van de supertabel naar de kubus met standaard bewerkingen via Diveline en/of Divebook.
  - b. Verifiëring de opgeleverde Diveboek of Diveline met de gebruikers wens
  - c. Rapport van a en b
6. Gebruikers training op weg naar zelfstandig gebruiken van Di-diver.
  - a. Gebruikersdag
  - b. Gebruikers helpen met gebruiken van de tool
  - c. Gebruikers helpen met het ontwikkelen van dimensionale analyse

## Appendix J. Multi-dimensionale besef bij technici

Doel:

Probeer een tabel te maken zodat:

- vanuit stap te zien is hoeveel mensen een stap zijn ingegaan
  - hoeveel mensen voor een bepaalde optie gekozen hebben
- jij noemt dit allebei als *measure* "aantal bezoekers". volgens mij gaat dat niet lukken.

zie bijvoorbeeld bijgaand tekstbestandje (nogiets.txt). Bouw hiervan vanuit de diver (File->Build model) even snel een modelletje.

Wat je wilt zien:

- 2 mensen kozen B, 1 koos A, 1 koos C
- iedereen haalde stap 1, slechts 2 haalden stap 2 en 3.

Dit laatste kun je bijvoorbeeld nog wel tevoorschijn halen, dat daarmee eerste is dan lastig te combineren.

Wat volgens mij beter gaat werken is als we het opzetten zoals in languit.txt.

Laat maar horen of dit je helpt.

Groetjes,

...

---

id	optie	optiecount	stap	stap count	stap finished	optie finished
----	-------	------------	------	------------	---------------	----------------

1	A	1	1	1	1	
---	---	---	---	---	---	--

1	A	1	2	1	1	
---	---	---	---	---	---	--

1	A	1	3	1		
---	---	---	---	---	--	--

2	B	1	1	1	1	
---	---	---	---	---	---	--

3	C	1	1	1		
---	---	---	---	---	--	--

4	B	1	1	1	1	
---	---	---	---	---	---	--

4	B	1	2	1	1	
---	---	---	---	---	---	--

4	B	1	3	1	1	1
---	---	---	---	---	---	---

---

id	optie	stap1_af	stap2_af	stap3_af	stap4_af	laatste stap	totaletijd
----	-------	----------	----------	----------	----------	--------------	------------

1	A	1	1	0	0	2	10
---	---	---	---	---	---	---	----

2	A	1	1	1	1	4	12
---	---	---	---	---	---	---	----

3	B	1	1	1	1	4	14
---	---	---	---	---	---	---	----

4	B	1	0	0	0	1	15
---	---	---	---	---	---	---	----

5	B	1	1	1	0	3	16
---	---	---	---	---	---	---	----

## Literatuurlijst

- [Erik Thomsen - 2002] Erik Thomsen, "OLAP solutions second Edition", ISBN: 0-471-40030-0
- [OLAP council] <http://www.olapcouncil.org/research/whtpaply.htm>, 2003
- [ET&JEA - 2001] Efraim Tuban en Jay E. Aronson, "Decision support systems and intelligent systems sixth edition", ISBN 0-13-032723, 2001
- [Trutek - 2003] <http://www.trutek.com/page43.html>, 2003
- [Thomas J. Kelly] Thomas J. Kelly, "Dimensional Data Modeling", <http://www.fe.up.pt/~jlborges/ADPIfiles/DimensionalDataModeling.pdf>
- [Alter - 2001] Information Systems: a management perspective, Addison Wesley Longman, 1999
- [Furstenberg] Tom A. Furstenberg, "Business Intelligence & Data warehousing", <http://wwwis.win.tue.nl/~houben/isa/isa-dw-TF.ppt>
- [Kimball - 2002] Ralph Kimball, "The data warehouse toolkit", 2002
- [TCCB - 2001] Thomas M. Connolly and Carolyn E. Begg, "Database Systems : A Practical Approach to Design, Implementation, & Management", 2001
- [Wieringa - 1996] R.J. Wieringa, "*Requirements Engineering: Frameworks for Understanding*", 1996