

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

## **Membangun Data Warehouse dan Data Mining dari Sistem Manajemen Mata Kuliah: Studi Kasus dari Sistem Informasi Manajemen Matakuliah FUTA (Federal University of Technology Akure Nigeria)**

### **Deskripsi (*Abstraksi*)**

Paper ini memberikan laporan tentang pengembangan data warehouse untuk manajemen bisnis menggunakan Studi kasus dari Sistem Informasi Manajemen Matakuliah FUTA. Paper ini medeskripsikan proses dari desain data warehouse dan pengembangan menggunakan Layanan Analisa Server Microsoft SQL. Ini juga pengembangan dari sebuah *data cube* juga sebagai aplikasi dari OLAP ( *Online Analytical Processing* ) tools dan Data Mining tools in data analysis.

Tujuan dari paper ini adalah untuk mennjukkan keuntungan dari data warehouse dan untuk menyadarkan perusahaan di Nigeria agar mulai membangun fasilitas ini dalam ERMS ( *Enterprise Resource Management Systems* ) dengan tujuan untuk mengefektifkan pengambilan keputusan bisnis yang akan membuat perkembangan perusahaan-perusahaan tersebut sangat cepat.

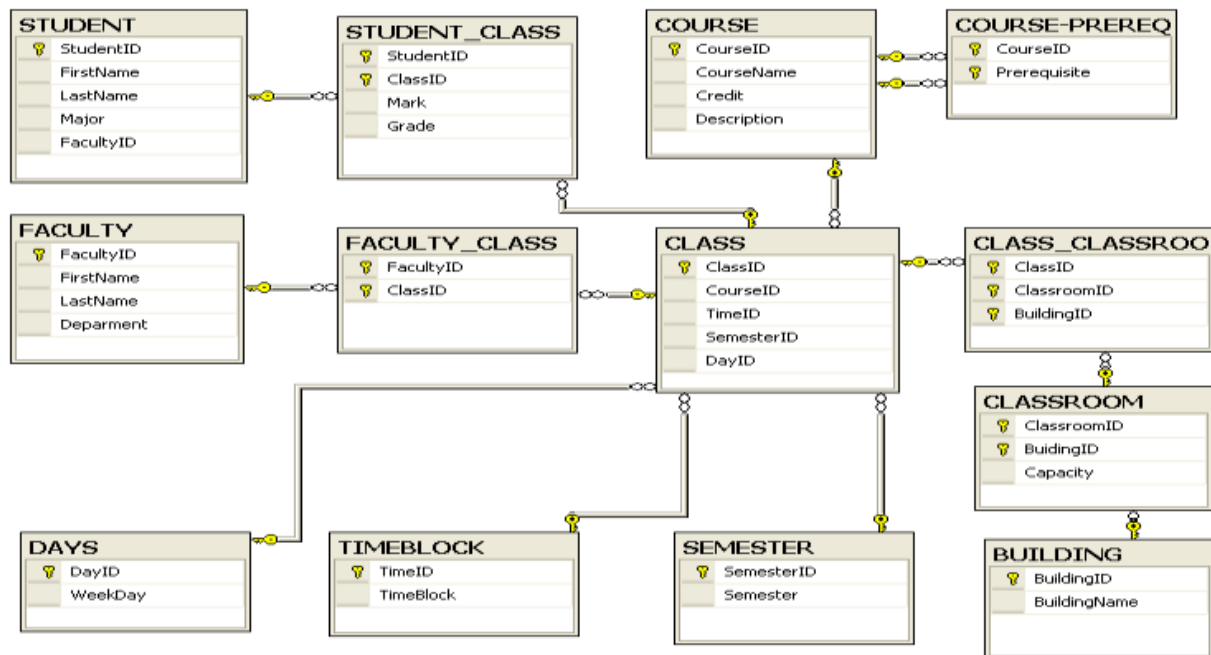
### **Masalah**

Studi kasus yang digunakan sebagai model dari proyek ini adalah sistem manajemen matakuliah di FUTA. Sekolah meregistrasi siswa setiap semester dan siswa mengambil matakuliah dan ujian. Hal ini dimanage oleh sistem OLTP ( *Online Transaction Processing* ). Representasi yang telah di simplekan mengenai desain dari sistem OLTP ditunjukkan dalam Figure 1.

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper



**Figure 1.** Logical Design OLTP

Setiap hari, operasi dari sekolah sangat bergantung pada sistem OLTP tersebut. Staffnya menggunakan OLTP sistem untuk memanage sistem informasi. Kebutuhan akan pengambilan keputusan yang strategis mengharuskan pengembangan dari data warehouse.

OLTP hanya dapat menangani sedikit data histori, yang dimana ini tidak cukup untuk membuat keputusan strategis. Dengan sebuah data warehouse dan OLAP, staff dan manajemen dapat melakukan operasi roll-up dan drill-down untuk penerimaan siswa baru dengan pertahun, persemester, permatakuliah atau kombinasi apapun yang diinginkan.

Sistem ini akan memberikan bantuan pengambilan keputusan yang fleksibel dan user-friendly. Universitas akan suka untuk menggunakan data-warehouse untuk menjawab pertanyaan-pertanyaan seperti : Apa yang trend dalam penerimaan maba, pendaftaran siswa, Jadwal mengajar dosen, banyaknya pendaftaran tahunan, dan lain lain. Tipe tipe pertanyaan tersebut membutuhkan banyak data di masa lampau untuk di generasi yang dimana sistem OLTP sekarang, tidak dapat melakukannya.

## Solusi Paper

Perancangan data warehouse dan data mining.

## Pengenalan

## Data Warehouse

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

Data Warehouse (DW) adalah kumpulan data yang diproduksi untuk menunjang pengambilan keputusan. DW juga adalah sebuah repository dari data dulu dan data terkini yang dapat digunakan untuk analisa. Data biasanya terstruktur untuk tersedia dalam bentuk siap untuk proses analisa (seperti Online Analytical Processing [OLAP], data mining, querying, reporting dan aplikasi penunjang pengambilan keputusan lainnya).

Operasi sehari-hari dari sebuah organisasi dilakukan menggunakan system OLTP. Sistem ini bagus untuk operasi pada umumnya dan sedikit melakukan pengambilan keputusan tapi, sistem kurang baik ketika melakukan penunjang pengambilan keputusan. Kurangnya data histori pada OLTP membuatnya sulit untuk memberikan informasi mengenai operasi dari bisnis. Di satu sisi, DW memberikan central repository dari historical data yang dimana hal ini menunjang sebuah platform terintegrasi untuk melakukan analisis historical terhadap data. Dengan DW dan OLAP, user dapat melakukan analisa data lebih baik dan mendapatkan pengetahuan yang lebih dari data penyimpanan.

Menurut Stephen Brobst and Joe Rarey (2003) Lima stage dari decision support diidentifikasi dalam DW:

### **Stage 1: Reporting**

Stage awal dari peluncuran data warehouse umumnya focus pada melaporkan dari sebuah sumber kebenaran dalam suatu organisasi. Tantangan terbesar pada stage 1 adalah integrase data

### **Stage 2: Analyzing**

Pada stage 2 peluncuran basis data, pembuat keputusan, tidak berfokus pada apa yang terjadi tapi lebih pada mengapa hal itu terjadi. Aktifitas analisa tergantung dengan drilling down kedalam angka dalam laporan untuk memilah dan mengotak-kotakkan data pada level yang detail. Performa sangat dibutuhkan pada implementasi stage 2 data warehouse, karena penyimpanan informasinya digunakan dengan lebih interaktif.

### **Stage 3: Predicting**

Data mining tingkat lanjut seringkali melakukan fungsi matematika yang lebih kompleks seperti logaritma, eksponensial, fungsi trigonometri, dan statistika tingkat lanjut untuk melakukan prediksi.

### **Stage 4: Ope-rationalizing**

Ope-rationalizing pada stage 4 dari evolusi, mulai membawa kita pada dunia Data warehousing aktif. Ketika stage 1 sampai 3 kita focus pada strategic decision making dalam sebuah

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

organisasi, stage 4 lebih berfokus pada tactical decision support. Ope-rationalizing umumnya berarti memberikan akses ke informasi atau langsung pada bagian decision making. Contohnya adalah : menjadwalkan dan mengirimkan paket

### **Stage 5: Active Warehousing**

Sebuah data warehouse mengirimkan informasi dan mengizinkan decision support melalui suatu organisasi dibandingkan dengan terperangkap pada proses strategis decision-making. Akan tetapi, tactical decision support tidak menggantikan strategic decision support. Makanya, sebuah aktif datawarehouse menunjang keberadaan bersama oleh kedua tipe pekerjaan

### **Data Mining**

Data mining sekarang umumnya digunakan oleh perusahaan dengan focus yang kuat pada pengguna, finansial, komunikasi, dan organisasi marketing. Data mining membuat perusahaan ini dapat menentukan hubungan antar factor internal seperti harga, pemosisian barang, atau skill staff dan factor external seperti indicator ekonomi, kompetisi, demografi pelanggan, dan data mining membuat mereka dapat menentukan hasil dari penjualan, kepuasan pelanggan, dan keuntungan perusahaan. Akhirnya data mining membuat mereka dapat “ drill down “ menjadi rangkuman informasi untuk melihat rangkuman data.

Contohnya, Blockbuster Entertainment menggali history rental video databasenya untuk merekomendasikan rental ke pelanggan. American Express dapat menyarankan produk ke cardholdernya berdasar pada analisa dari pengiriman bulanan mereka.

### **Bagaimana Data Mining Bekerja**

Ketika informasi teknologi skala besar telah berevolusi memisahkan transaksi dan analytical sistem, data mining memberikan hubungan antara kedua hal tersebut. Software data mining menganalisa hubungan dan pola dalam data transaksi yang tersimpan tergantung dari open-ended user queries. Umumnya ada 4 tipe hubungan yang dicari :

**Classes:** Data yang tersimpan digunakan untuk memperkirakan lokasi data. Contohnya, sebuah restoran dapat menggali data pembelian pelanggan untuk menentukan kapan pelanggan datang dan apa yang umumnya mereka pesan. Informasi ini dapat berguna untuk meningkatkan trafficking dengan mempunyai daily specials.

**Clusters:** Data item dibuat pergrup tergantung dari logical relationshipnya atau keinginan pengguna. Contoh : Data dapat digali untuk menentukan segmen pasar atau afinitas pelanggan.

**Associations:** Data dapat ditambang untuk mengidentifikasi asosiasi.

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

**Sequential patterns:** Data ditambang untuk mengantisipasi pola sifat dan trend, contohnya : seorang penjual peralatan bertamasya dapat memprediksi tasa akan dibeli dari pembelian pelanggan akan tempat tidur portable dan sepatu naik gunung.

### **Data mining terdiri dari 5 elemen utama:**

Extract, mengubah, and meload transaksi data ke sistem data warehouse.

Menyimpan and mengatur data dalam sistem database multidimensi.

Memberikan akses data ke bisnis analist dan professional informasi dan teknologi

Menganalisa data dari software aplikasi

Memberikan data dalam format yang berguna seperti graf atau table.

### **Mendesain Data Warehouse**

Diskusi selanjutnya secara garis membesar adalah proses dari desain data warehouse. Ini termasuk dengan logical design, OLAP design, dan data mining design.

### **The logical Design**

Tidak seperti ERD dan UML, proses data modelling, logical design dari Data warehouse didefinisikan sebagai pendekatan dimensional data modeling. Proses dimensioning design yang dilakukan pada proyek ini menganut pada methodology yang dipaparkan oleh Kimball dan Ross (2002).

Untuk mengurangi operasi join yang memperlambat query, normalisasi tidak menjadi panutan prinsip pada Desain Data Warehouse. Sebuah skema adalah koleksi dari objek database, termasuk table, view, index, dan sinonim. Ada banyak cara untuk menyusun schema objek dalam schema model yang didesain untuk data warehousing. Berikut ini adalah 2 tipe schema yang umumnya digunakan pada dimensional data modeling.

- **Star schema**

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

The star schema mungkin adalah schema data warehouse termudah. Disebut star schema karena ERDnya berbentuk seperti bintang, yang dimana pusatnya tersebar dari table tengah. Tabel tengah dari bintang ini terdiri dari sebuah fact table yang besar dan titik dari bintang ini adalah dimension tables.

Sebuah star schema di karakteristik dengan 1 atau lebih fact tables yang mempunyai informasi utama dalam data warehouse, beberapa dimension table yang lebih kecil (atau lookup tables), yang dimana memiliki informasi mengenai isi dari sebuah atribut dalam fact table. Star schema memfasilitasi respon cepat dari query. Inti dari nilai detail disimpan dalam fact table. Info dimensional dan hierarki disimpan di dimension table.

- **Snowflake schema**

Lebih kompleks dari star schema, jika ingin tau lebih silahkan meriset sendiri dari internet.

## **Pendekatan Dimensional Data Modeling**

Pendekatan dimensional cukup berbeda dari pendekatan normalisasi diikuti ketika mendesain sebuah database untuk operasi sehari-hari.

### ***Data Hierarchies in Dimensional tables***

Setiap dimensi mengandung setidaknya 1 hierarki. Hierarki ini mengizinkan user untuk menganalisa agregasi data menggunakan OLAP. Ini dapat membuat item sejenis digabungkan dan dirangkum pada untuk analisa tingkat tinggi selama tetap mempertahankan kemampuan untuk melakukan drill down ke detail produk yang lebih spesifik.

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

### The Dimensional Data showing Hierarchies

<table><tr><th colspan="2">DimStudent</th></tr><tr><td>PK</td><td><u>StudentID</u></td></tr><tr><td></td><td>F i s r t N a m e LastName Sex Major Sex</td></tr></table>	DimStudent		PK	<u>StudentID</u>		F i s r t N a m e LastName Sex Major Sex	<table><tr><th colspan="2">DimFaculty</th></tr><tr><td>PK</td><td><u>FacultyID</u></td></tr><tr><td></td><td>F i s r t N a m e LastName Department</td></tr></table>	DimFaculty		PK	<u>FacultyID</u>		F i s r t N a m e LastName Department	<table><tr><th colspan="2">DimClassroom</th></tr><tr><td>PK</td><td><u>ClassroomID</u></td></tr><tr><td></td><td>C l a s s r o o m BuildingName Capacity</td></tr></table>	DimClassroom		PK	<u>ClassroomID</u>		C l a s s r o o m BuildingName Capacity
DimStudent																				
PK	<u>StudentID</u>																			
	F i s r t N a m e LastName Sex Major Sex																			
DimFaculty																				
PK	<u>FacultyID</u>																			
	F i s r t N a m e LastName Department																			
DimClassroom																				
PK	<u>ClassroomID</u>																			
	C l a s s r o o m BuildingName Capacity																			
<table><tr><th colspan="2">DimSchedule</th></tr><tr><td>PK</td><td><u>ClassID</u></td></tr><tr><td></td><td>a c a d y e a r Semester WeekDay TimeBlock</td></tr></table>	DimSchedule		PK	<u>ClassID</u>		a c a d y e a r Semester WeekDay TimeBlock	<table><tr><th colspan="2">DimCourse</th></tr><tr><td>PK</td><td><u>CourseID</u></td></tr><tr><td></td><td>C o u r s e N a m e Credit Description</td></tr></table>	DimCourse		PK	<u>CourseID</u>		C o u r s e N a m e Credit Description							
DimSchedule																				
PK	<u>ClassID</u>																			
	a c a d y e a r Semester WeekDay TimeBlock																			
DimCourse																				
PK	<u>CourseID</u>																			
	C o u r s e N a m e Credit Description																			

### Fact Table

Fact table memiliki atribut dimension dan pengukuran. Dimension atribut adalah Foreign key dari atribut yang lain, yang disebut degenerate dimension (<<dd>>). Measures adalah nilai yang akan di agregasi ketika grup query dijalankan.

ClassFact Table	
PK	<u>StudentID</u>
PK	<u>CourseID</u>
PK	<u>ClassID</u>
PK	<u>ClassroomID</u>
PK	<u>FacultyID</u>
	<u>Grade</u>
	<u>Mark</u>
	<u>Registered</u>

Nama : Julio Anthony Leonard

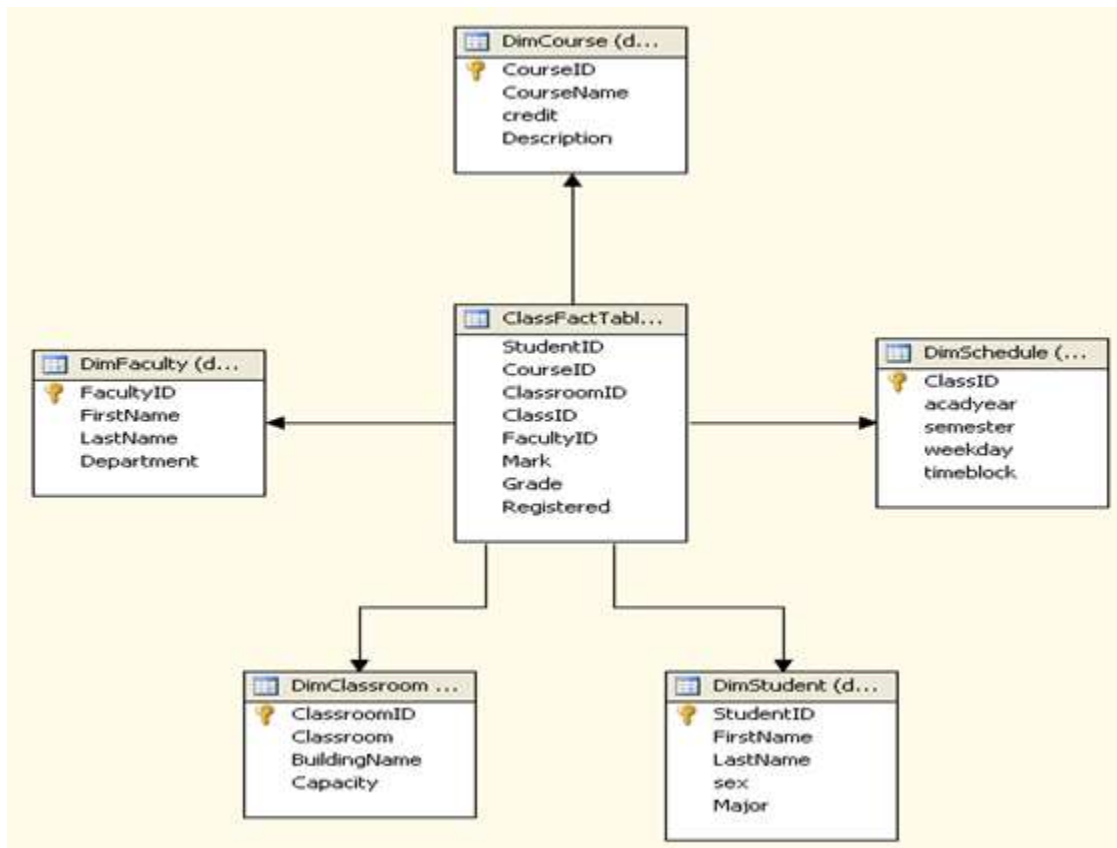
NRP : 5113100148

Tugas Resume Paper

Registered adalah measure, dan yang lainnya adalah (<<dd>>)

## The Star Schema

Data warehouse umumnya di organisasikan dengan sebuah fact table yang besar ditengah, dan banyak dimension table yang kecil. Konfigurasi ini kita sebut star schema. Sebuah star schema telah digunakan dalam riset ini. Berikut aalah diagram dari hubungan antara fact table dan dimensional table.



## Data Warehouse Implementation

### Transporting Data from OLTP Database to Data Warehouse



Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

SQL server melayani alat SSIS untuk membantu mentransport data kedalam dan keluar database. Ini termasuk pembuatan dari struktur data dimension table, diikuti dengan pembuatan proyek servis terintegrasi menggunakan Business Intelligence package pada SQL server. Setelah membuat proyek, sebuah sumber data kemudian dibuat diikuti dengan pembuatan sumber data OLEDB dan destinasi. Table berikut menunjukkan bahwa dimensional table dan fact table dibuat dalam proyek ini.

**a) Dim Student**

Table - dbo.DimStudent					
Summary					
	StudentID	FirstName	LastName	sex	Major
	1	Ojo	Adewale	M	COMPUTER
	2	Aina	Olubukola	F	MATHEMATICS
	3	Funmilayo	Olaiya	F	BIOLOGY
	4	Fabunmi	Oni	M	CHEMISTRY
▶	5	Kunle	Ajayi	M	Physics
	6	Ronke	Dawodu	F	COMPUTER
*	NULL	NULL	NULL	NULL	NULL

**b) Dim Faculty**

Table - dbo.DimFaculty				
Table - dbo.DimSchedule				
Table - dbo.DimSchedule				
	FacultyID	FirstName	LastName	Department
	1	Kolawole	Olu	Computer Science
	2	Ajewole	Ajayi	Computer Science
	3	Oni	Dada	Computer science
✎	4	Wale	Ige	Biology
	5	Femi	Akinsola	Physis
	6	Ojo	Funmilayo	Maths
*	NULL	NULL	NULL	NULL

**c) Dim Course**

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

Table - dbo.DimCourse					Table - dbo.DimFaculty	Table - dbo.DimSchedule
	CourseID	CourseName	credit	Description		
▶	1	COMPUTER ARC...	3	COMPUTER COU...		
	2	DATABASE MAN...	3	DATABASE		
	3	SYSTEM ANALYS...	3	ANALYSIS		
	4	COMPUTER NET...	3	NETWORKS DES...		
	5	Solid Electronics	3	elctronic		
	6	Physics	3	physics		
	7	Solid Electronics	3	elctronic		
	8	SYSTEM ANALYS...	3	ANALYSIS		
*	NULL	NULL	NULL	NULL		

#### d) Dim Schedule

Table - dbo.DimSchedule						Table - dbo.DimSchedule	Table - dbo.DimStudent	Summary
	ClassID	acadyear	semester	weekday	timeblock			
▶	1	2009	FISRT	Monday	8-10			
	2	2009	SECOND	Monday	10-12			
	3	2009	SECOND	Tuesday	12-2			
	4	2010	FIRST	Wednesday	2-4			
	5	2010	SECOND	Wednesday	12-2			
	6	2010	SECOND	Monday	2-4			
	7	2011	FIRST	Friday	8-10			
	8	2011	SECOND	Thursday	10-12			
*	NULL	NULL	NULL	NULL	NULL			

#### e) Dim Classroom

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

Table - dbo.DimClassroom		Table - dbo.DimCourse		Table - dbo.DimFaculty	
	ClassroomID	Classroom	BuildingName	Capacity	
▶	1	201	ETF	300	
	2	102	BIG LT OBANLA	300	
	3	201	ETF	300	
	4	102	BIG LT OBANLA	300	
	5	104	ETF	250	
	6	102	BIG LT OBANLA	300	
	7	101	SMALL LT OBANLA	200	
	8	201	ETF	300	
*	NULL	NULL	NULL	NULL	

## Fact Tables

Table - dbo.ClassFactTable		Table - dbo.DimStudent		Diagram - FB4...tabaseDiagram		Summary		
	StudentID	CourseID	ClassroomID	ClassID	FacultyID	Mark	Grade	Registered
	1	1	1	1	1	91	A	1
	1	2	2	2	2	95	A	1
	1	3	3	3	3	91	A	1
	2	1	1	1	1	88	B	1
	2	2	2	2	2	80	B	1
	2	3	3	3	3	90	A	1
▶	3	1	1	1	1	75	C	1
	3	2	2	2	2	90	A	1
	4	1	1	1	1	90	A	1
	4	2	2	2	2	80	B	1
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

## Online Analytical Process (OLAP)

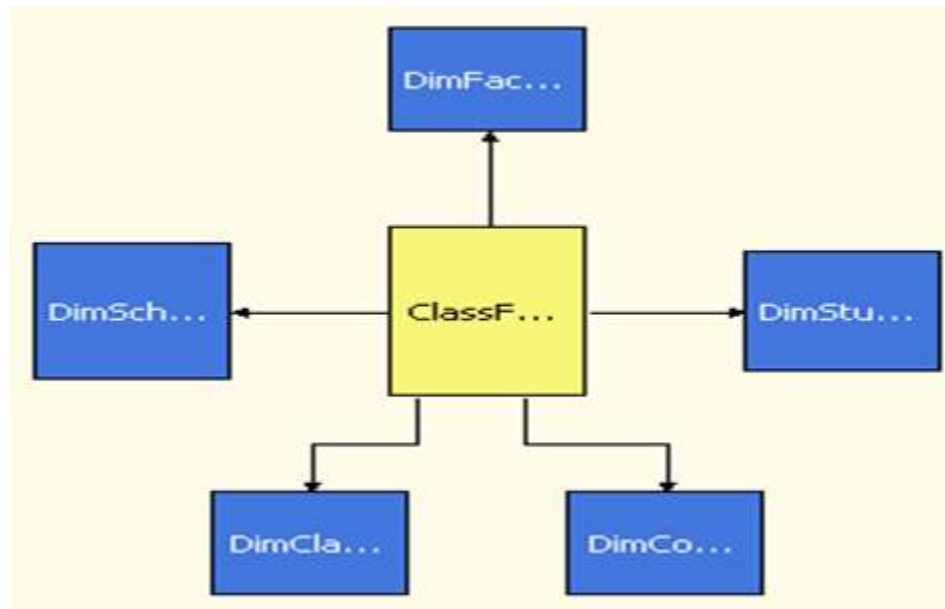
Setelah data warehouse telah diisi dengan data, langkah selanjutnya dalam pengembangan data warehouse adalah menyediakan user dengan data analysis tool seperti OLAP dan Data Mining untuk menganalisa proses bisnis untuk decision making. OLAP adalah sebuah servis yang secara otomatis mengambil sebuah himpunan dari rangkuman view (tabel) dan menyimpan rangkuman view ini kedalam disk. OLAP juga mengatur view ini dan mengupdate mereka ketika fact table telah memiliki data baru. Untuk membuat ini, kita harus

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

membuat Proyek Servis Analisa menggunakan Bussiness Intelligence Development Studio. Lalu, kita mendefinisikan sumber datanya, lalu membuat view data sourcenya kemudian membuat Cubenya. Berikut ini gambar OLAP Cube yang dibuat dengan SQL Server.



Gambar ini menunjukkan bagaimana sebuah star schema yang dijelaskan sebelumnya diubah menjadi bentuk internal, yang dimana SQL Business Intelligence dapat mengambil data yang diminta oleh user.

### Laporan Analisa Data

Laporan dibawah adalah hasil dari operasi roll up dan drill down dari olap yang dibuat dengan data untuk berbagai jenis hirarki dan dimensi.

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

### Gambar Analisa Data Berdasarkan Gender dan Course yang didaftarkan

Drop Filter Fields Here				
	Course Name ▾			
	COMPUTER ARCHITECTURE	DATABASE MANAGEMENT	SYSTEM ANALYSIS AND DESIGN	Grand Total
Sex ▾	Class Fact Table Count	Class Fact Table Count	Class Fact Table Count	Class Fact Table Count
F	2	2	1	5
M	2	2	1	5
Grand Total	4	4	2	10

### Gambar Analisa Data Berdasarkan Tahun Akademik dan Matakuliah yang didaftarkan

Drop Filter Fields Here				
	First Name ▾			
	Ajewole	Kolawole	Oni	Grand Total
Acadyear ▾	Class Fact Table Count	Class Fact Table Count	Class Fact Table Count	Class Fact Table Count
2009	4	4	2	10
Grand Total	4	4	2	10

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

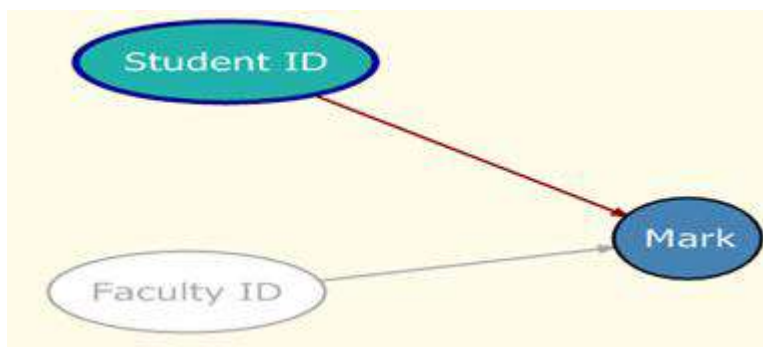
### Gambar Analisa Data Berdasarkan Ruangan, Jadwal dan Matakuliah yang didaftarkan

Drop Filter Fields Here		Course Name ▾			
Building Name ▾	Timeblock ▾	DATABASE MANAGEMENT			
		Class Fact Table Count	Class Fact Table Count	Class Fact Table Count	Grand Total
BIG LT OBANLA	10-12	4			4
	Total	4			4
ETF	12-2			2	2
	8-10		4		4
	Total		4	2	6
Grand Total		4	4	2	10

### The Data mining (Pengetahuan menggunakan Data mining)

Data mining adalah sebuah proses yang membuat computer mencari korelasi antara data, dan memberikan hipotesis akurat agar pengguna dapat mempertimbangkannya. Beberapa masalah yang dapat diselesaikan dengan data mining antara lain:

- Categorization: Mengklasifikasikan himpunan kasus yang diberikan
- Clustering: Menemukan pengrupan yang natural dari himpunan kasus yang diberikan.
- Association rule: Menemukan item apa yang sering diproses bersama.



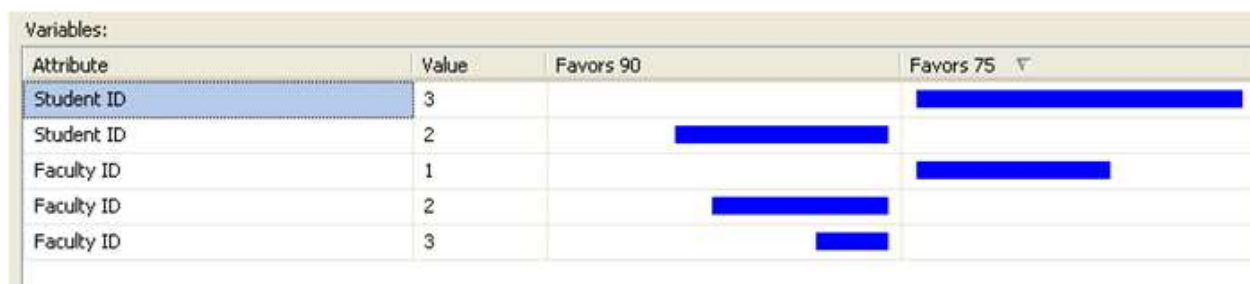
Gambar Data Clustering Mining yang menemukan hubungan antara nilai siswa dan fakultas

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

Angka ini menunjukkan hasil yang diperoleh dari model data-mining di atas. Ini menunjukkan kelompok kelas Nilai 90 dan kelompok Nilai 75. Kelas mahasiswa berpihak antara tanda 90-75 dan derajat keanggotaan kelas seperti ditunjukkan oleh bar. Juga ditampilkan fakultas dan derajat keanggotaan mereka dalam pemberian tanda di kisaran 90 atau 75 dan derajat keanggotaan mereka dari kelompok tersebut. Sebuah data-mining yang mengelompokkan siswa dan Kuliah Menampilkan kelas siswa berpihak 'antara tanda 90-75 dan derajat keanggotaan kelas seperti ditunjukkan oleh bar.



## Kesimpulan dan Diskusi

Pada paper ini, kita telah berhasil mendemonstrasikan proses desain dan pengembangan data warehouse dan aplikasi data mining menggunakan SQL server Business Intelligence Development tools menggunakan studi kasus dalam lingkungan akademik. Perlu diperhatikan bahwa teknik ini dapat diaplikasikan di organisasi manapun yang berharap untuk mengimplementasikan Bussiness Intelligence as part of their strategic decision support operations.

Kekuatan dari Data Warehousing dalam analisa data sangat hebat dan data mining dapat menemukan harta harun tersembunyi dalam data warehouse. Organisasi, umumnya di Nigeria dapat mulai mengimplementasikan proyek ini sebagai bagian dari pengambilan keputusan strategis mereka. Dengan ini, mereka dapat mulai melihat trends dari berbagai hal secara historis seperti operasional hari ke hari mereka diakumulasikan dalam 1 tahun. Mereka juga dapat meramalkan masa depan menggunakan Neural Networks, Regression Analisis, dan

Nama : Julio Anthony Leonard

NRP : 5113100148

Tugas Resume Paper

data mining lainnya yang terdapat pada SQL Server Business Intelligence. Data warehouse adalah solusi yang kita butuhkan untuk melontarkan bisnis kita atau organisasi kita ke level selanjutnya.