# THE AGGREGATORS: GROUP PROJECT

Austin Beller | Debasis Chatterjee | Monica Joyner | Levi Peckham | Jo Vivian IST 722 | SECTION M400 | Team 1

# Contents

Contents	2
Project Document	3
Project Charter	
Key Dates	4
Project Plan	5
Outline of Functional Requirements	6
Business Processes and Value	6
Fudgeflix	6
Fudgemart	6
Data Profiling Exercise	8
High-Level Dimensional Modeling Worksheet	9
Bus Matrix	9
Attributes and Metrics	10
Detail-Level Dimensional Modeling Worksheet	10
Data Warehouse on SQL Server	11
Star Schema	11
DW and Stage Implementation	11
ETL In SSIS	12
Source to Target Maps	12
SSIS Package	14
Source to Stage	14
Stage to DW	18
Business Intelligence	21
MOLAP	21
Visualizations	22

# **Project Document**

# Project Charter

Purpose, Background and Goals	Succe	ess Criteria			
Purpose: Apply IST722 course knowledge by implementing a DWH/BI system based on a case study  Background: In the context of the case, there are two OLTP sources and a third source containing supplemental data  Goals:  1. The timely delivery of project artifacts 2. The successful analysis, design, and implementation of a functional DWH/BI system 3. Team cohesion and collaboration with shared work responsibilities.	<ul> <li>Identify 4+ business as part of function</li> <li>Justify value of busin</li> <li>Implement 2 busine</li> <li>Implement meaning intelligence accord</li> <li>DWH is illustrative of Satisfy project delived</li> <li>Present DW/BI initiation information sharin</li> <li>Demo illustrates the</li> </ul>	functional requirements of Fudgemart, Inc.  He business processes to model in the DWH of functional requirements lue of business processes in the DWH of the meaningful, actionable business ence according to Kimball methods lustrative of 4 DWH characteristics roject deliverables criteria DW/BI initiative at executive level of attion sharing sustrates the satisfaction of functional ments based on business processes			
Products/Deliverables	Team Roles	Risks and Issues			
<ul> <li>Project document</li> <li>High-level dimensional modeling worksheet</li> <li>Detail-level dimensional modeling worksheet</li> <li>DWH on SQL Server</li> <li>SSIS ETL (Source-Target map, SSIS packages, dataflows, etc.)</li> <li>BI (SASS cubes, reporting, documentation)</li> <li>Presentation &amp; Demo</li> </ul>	Team Leader/Business Lead: Monica Joyner Project Manager: Jo Vivian Business Analyst: Austin Beller/Levi Peckham Data Architect/ETL Architect: Debasis Chatterjee BI Architect: Levi Peckham	<ul> <li>Time constraints</li> <li>DWH/BI knowledge         and experience         constraints</li> <li>New team required to         gel at an accelerated         pace</li> </ul>			

### **Key Dates**

### **Weekly Touchpoints**

- 01/29/2020
- 01/26/2020
- 02/02/2020
- 02/09/2020
- 02/16/2020
- 02/23/2020
- 03/01/2020
- 03/08/2020
- 03/15/2020

### Milestone Delivery Dates

- Milestone 1:
  - Draft of project document Project charter, project plan, outline of functional requirements, overview of business processes and value, primary team role assignments
- Milestone 2:
  - Draft of high-level dimensional modeling worksheet bus matrix, attributes and metrics, documentation of issues
- Milestone 3:
  - Draft of detail-level dimensional modeling worksheet and SQL implementation

#### Presentation Date

• 03/19/2020

# Project Plan

Project	ct Namo				Business Lead	Monica Joyner		Data Architect	Debasis Chatterjee	
Project Name Project Deliverable Data Warehouse/Business Intelligence (DW/		//RI) Solution for Eudgema				Jo Vivian		ETL Architect	Monica Joyner	
Scope Statement		(7 bi) Solution for Fuagema		Business Analyst		Austin Beller		BI Architect Levi Peckham		
Start Date 1/9/2020		End Date	3/19/2020		Overall Proje			On Schedule/On Budget		
Start	Date	1/3/2020	Liid Date	3/13/2020		Overall 1 10j	cct i rogic.		On sened	ne, on baaget
At		- 10					Duration			
Risk?		Task Name	Assigned To	Start Date	End Date	Due Date	(Days)		Status	
			Project	Milestone 01 - Projec	t Document					
N	Assemble a Project	Charter	Monica	1/19/2020	1/22/2020	1/30/2020	3	Complete		
N	Devise a Project Pla	n	Jo	1/19/2020	1/23/2020	1/30/2020	4	Complete		
N	Outline Functional R	Requirements	Levi, Austin, Debasis	1/19/2020	1/23/2020	1/30/2020	4	Complete		
N Overview which business processes you will model from those functional requirements and explain their business value		Levi, Austin, Debasis	1/19/2020	1/25/2020	1/30/2020	6	Complete			
			Project Milestone 0	2 - High level dimension						
N	Bus Matrix		Levi, Austin, Debasis	1/26/2020	2/15/2020			In Progress/O		
N	Attributes and Metr		Levi, Austin, Debasis	1/26/2020	2/15/2020			In Progress/O	n time	
N		Peer Review worksheet	Jo, Monica	2/15/2020	2/16/2020	2/20/2020	-	Not Started		
N	Outline issues		Jo, Monica	2/15/2020	2/17/2020			Not Started		
			t Milestone 03 - Detail lev				1			
	T	Worksheet		2/21/2020	2/29/2020	3/12/2020	8	Not Started		
N	Contains fully com	pleted dimensions and facts	Austin, Debasis, Monica	2/21/2020	2/28/2020	3/12/2020	7	Not Started		
N	N Identifies Sources for your source-to-target map		Austin, Debasis, Monica	2/21/2020	2/28/2020	3/12/2020	7	Not Started		
N	N Should be used to generate a SQL schema for your data warehouse		Austin, Debasis, Monica	2/21/2020	2/28/2020	3/12/2020	7	Not Started		
N	Peer Review works	sheet	Levi	2/21/2020	2/29/2020	3/12/2020	8	Not Started		
	Data V	Varehouse on SQL Server		3/1/2020	3/8/2020	3/12/2020	7	Not Started		
N	Follow Consistent Co	onventions	Monica	3/1/2020	3/8/2020	3/12/2020	7	Not Started		
N	N Adopts techniques learned in the course, such as use of staging and enterprise bus		Monica	3/1/2020	3/8/2020	3/12/2020	7	Not Started		
			Projec	ct Milestone 04 - Final						
	In	itial ETL done in SSIS		3/13/2020	3/19/2020			Not Started		
N	Source-to-target n		Monica	3/13/2020	3/19/2020	3/26/2020		Not Started		
N	SSIS package to sta		Jo	3/13/2020	3/19/2020	3/26/2020		Not Started		
N		ad to DW from stage	Jo/Austin	3/13/2020	3/19/2020	3/26/2020		Not Started		
N	ETL Documentation			3/13/2020	3/19/2020	3/26/2020		Not Started		
N	Source to target		Monica	3/13/2020	3/19/2020	3/26/2020		Not Started		
N	Screenshot of da		Jo	3/13/2020	3/19/2020	3/26/2020		Not Started		
N	Explanation of E		Monica	3/13/2020	3/19/2020	3/26/2020		Not Started		
N	Data quality of su	,	-	3/13/2020	3/19/2020	3/26/2020		Not Started		
	Business Intelligence			3/13/2020	3/22/2020			Not Started		
_	N SSAS cubes on analysis services server		Levi	3/13/2020	3/22/2020	3/26/2020	9	Not Started		
	N BI Dashboard or application in Power BI and/or Excel		Levi	3/13/2020	3/22/2020			Not Started		
	N BI Documentation		Levi/Monica	3/13/2020	3/22/2020	3/26/2020		Not Started		
	N Goal of analytics		Levi/Monica	3/13/2020	3/22/2020	3/26/2020	9	Not Started		
N	Type of BI	acoutation and Dama	Levi	3/13/2020	3/22/2020			Not Started		
		esentation and Demo	TBD	<b>3/23/2020</b> 3/23/2020	<b>3/25/2020</b> 3/25/2020	3/26/2020	1	Not Started		
N		e presentation overviewing DW and BI	TBD	3/23/2020		3/26/2020		Not Started Not Started		
N	5-minute demo of	ы аррисации	TBD	3/23/2020	3/25/2020	3/26/2020				
N Demo in class			טמון	3/23/2020	3/25/2020	3/20/2020	4	Not Started		

### **Outline of Functional Requirements**

- 1. Every time a customer orders an item that is logged to a worksheet which keeps track of how many times that item is sold
- 2. Every time a customer orders a product an email is sent to the customer reiterating the order then it emails when the item ships and when the item arrives
- 3. Employees and supervisors should be able to keep track of time.
- 4. Customers will be able to view what movies are available to rent
- 5. When a movie is rented it is then marked as out until no more copies are available which should then state the movie is unavailable.
- 6. When the account holder selects a plan, a message should be sent to mark which plan has been selected to the account holder thanking them.
- 7. Movies will be tracked via shipped and possible return date if the movie is close to the end of renting an email will be sent to the account holder.
- 8. All customers will have to submit a credit card before being able to order any products
- 9. Products will be ranked by favorability and rating, when a product falls below a set standard then the vendor will be notified the product is under review.
- 10. All employees must have a title.
- 11. Employees hours and payroll will be tracked and monitored by management.
- 12. All plans will be tier rated based on price with highest including streaming and lowest just offering shipping of DVD's.

### **Business Processes and Value**

### Fudgeflix

#### BP<sub>1</sub>

Track (USA?) Regional Sales Fulfillment by Year per Department

ff\_plans

ff\_account\_titles

FR: Weekly report by state indicating number of orders/qty

BI: Regional Heatmap

### BP 2

Track Regional Fill Rate by Quarter by Sate

ff plans

FR: Monthly report by state indicating increases/decreases in plan types

### Fudgemart

### BP<sub>3</sub>

Track Employee Hiring and Length of Employment Duration — we aspire to have happy (!) full-time employees who are paid fairly, equally, and want to stay w/us

fm\_employees

fm\_jobtitles\_lookup

fm\_departments\_lookup

fm\_employee\_timesheets

FR: Monthly report by department, supervisor, full-time status, hire/term date

### BP 4

Track Regional Distribution of Orders by Department — we may wish to increase products or entirely eliminate departments based on sales; also optimize warehouse resources/distribution

fm\_orders

fm\_products

fm\_departments\_lookup

FR: Daily report of products sold aggregated by department per USA region (group of USA states)

## Data Profiling Exercise

	Data Profiling for FudgeMart Companies, Inc.							
Database	Table Name	Type of Data?	Row Count One row in this table means	<b>▼</b> Business Keys	Notes / Issues ?			
ExternalSources	fm_customer_survey		25 Customer survey ordered by email	none	Must be classified			
ExternalSources	fm_customer_tweets		625 Various tweets from customers	none	Must be classified			
fudgeflix_v3	ff_account_billing		1294	fk_ab_account_id, fk_ad_plan_id	Only 35 unique AccountID for 1,294 rows?			
fudgeflix_v3	ff_account_titles		2368	fk_at_account_id, fk_at_title-id				
fudgeflix_v3	ff_accounts		35 per account_id	fk_account_zipcode, fk_plan				
fudgeflix_v3	ff_cast		70543	fk_cast_people_id, fk_cast_title_id				
fudgeflix_v3	ff_directors		7136	fk_director_people_id,fk_director_title_id				
fudgeflix_v3	ff_genres		485	pk				
fudgeflix_v3	ff_people		25347	pk				
fudgeflix_v3	ff_plans		7 per plan_id	pk	4 plan_current == 1, 3 plan_current == 0			
fudgeflix_v3	ff_title_genres		27841 genre per title_id	fk_tg_genre_name, fk_tg_title_id				
fudgeflix_v3	ff_titles		7185 movie/season	pk	season == TV show			
fudgeflix_v3	ff_zipcodes		41968 unique zip code	pk				
fudgemart_v3	fm_creditcards		49 Credit card information	pk	Different spaces for different credit cards			
				fk_customer_creditcards_creditcard_id,				
fudgemart_v3	fm_customer_creditcards		54 Credit card id	fk_customer_creditcards_customer_id	none noted			
				fk_customer_creditcards_creditcard_id,				
fudgemart_v3	fm_customer_product_reviews		1039 Contains reviews in stars	fk_customer_creditcards_customer_id	Redundancy of multiple products with multiple reviews			
fudgemart_v3	fm_customers		25 Contains Costumer contact information	pk	Some null in rows			
fudgemart_v3	fm_departments_lookup		6 Department title	pk	none noted			
				fk_fudgemart_employee_timesheets_fudgemart_empl				
fudgemart_v3	fm_employee_timesheets		6275 Total hours plus clock in and out information	oyees	none noted			
fudgemart_v3	fm_employees		33 Contains basic employee information	fk_employee_department, fk_employee_jobtitle	none noted			
fudgemart_v3	fm_jobtitles_lookup		4 Job Title	pk	none noted			
fudgemart_v3	fm_order_details		10466 Displays product id and department id	fk_order_details_items, fk_order_details_orders	possible redundancy order by			
				fk_order_shipvia, fk_orders_creditcard_id,				
fudgemart_v3	fm_orders	1	3516 Displays order and shipping information	fk_orders_customers	redundancy issue of possible double orders			
6 1 2			52 1	fk_fudgemart_products_fudgemart_vendors,				
fudgemart_v3	fm_products		53 Name of product and department location	fk_fudgemart_products_product_department	none noted			
fudgemart_v3	fm_shipvia_lookup		4 Name of Company which ships order	pk	none noted			
fudgemart_v3	fm_vendors		9 Contact information of Vendor	pk	Some null in rows			

# High-Level Dimensional Modeling Worksheet

### Bus Matrix

Business Process	Fact	Fact Grain				Justom	et of or	) her Da	e doar	-/ / e/ e/
Name	Table	Туре	Granulairty	Facts	/(	75/ Q	5°/0	'/ Si	10	4
Movie Recommendation	FactFFRecommendation	transaction	one row per subscription	Average Rating, Runtime, YearsOld	х				х	7
Price Plan Coverage	FactFFSubscriptions	transaction	one row per subscription	Plan Price, Months Opened	х	х			х	7
Fudgemart Sales/Revenue	FactFMSales	transaction	one row per order	Quantity, Wholesale Price, Retail Price, Revenue, Profit	х	х				7
Order Fulfillment	FactFMFulfillment	transaction	one row per order	FulfillmentInLagDays, ShippedVia	х		х	х		7
Movie/Product Recommendation	FactItemAnalysis	transaction	one row per customer	Retail Price, Plan Price	х	х			х	1

### **Attributes and Metrics**

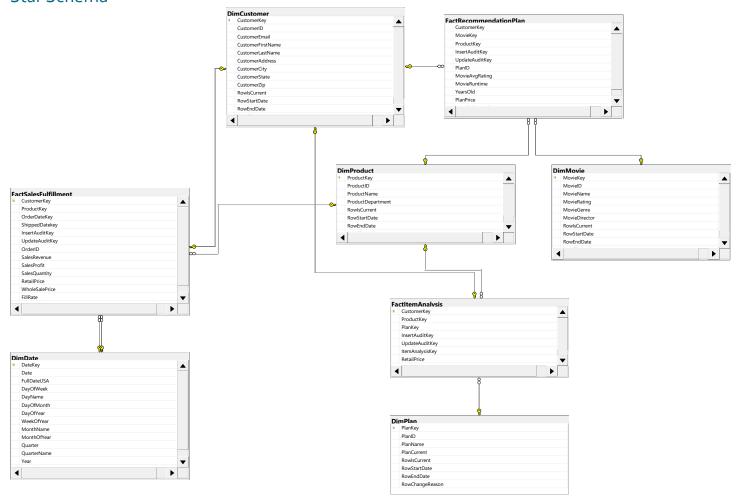
Dimension /	Attribute /	•	Alternate	Sample
Fact Table	Fact Name	Description	Names	Values
FactFudgeflixRecommendation	MovieAvgRating	Average customer rating		3.96
	MovieRuntime	Movie runtime in seconds		10240
	YearsOld	Current Year - Movie Release year		15
FactFudgeflixSubscriptions	IsPlanCurrent	Binary: 0 for no, 1 for yes		0
	PlanPrice	Subscription plan price in \$		\$ 12.99
	Months Opened	Current date - date account opened in months		4
FactFudgemartSales	Revenue	RetailPrice * Quantity		\$ 35.96
	Profit	(RetailPrice - WholesalePrice) * Quantity		\$ 24.00
	Quantity	# of items of a product		4
	RetailPrice	Retail price of a product in \$		\$ 8.99
	WholesalePrice	Wholesale price of a product in \$		\$ 2.99
FactFudgemartFulfillment	FulfillmentLagInDays	ShippedDate - Order Date		1
	ShippedVia	Company used to ship order (mini dimension)		JiffyEx
FactItemAnalysis	RetailPrice	Retail price of a product in \$		\$ 8.99
·	PlanPrice	Subscription plan price in \$		\$ 12.99
DimCustomer	CustomerFirstName	Customer's first name		Lisa
	CustomerLastName	Customer's last name		Karforless
	CustomerAddress	Customer's address		1555 W Lane Ave
	CustomerZipcode	Customer's zip code		43221
DimOrderDate	OrderDateID	ID for date		20040926
J.III O'UC' DUC	Day	15 for date		26
	Month			September
	Year			2004
DimshippedDate	ShippedDateID	ID for date		20040926
	Day			26
	Month			September
	Year			2004
DimProduct	ProductName	name of product		Straight Claw Hammer
	ProductDepartment	department it's in in fudgemart, for fudgeflix it would be fudgeflix		Hardware
DimMovie	MovieName	Name of Movie		Meet Joe Black
<del>-</del>	MovieRating	Viewer Rating of movie		PG-13
	MovieGenre	Genre of Movie		20th Century Period Pieces
Disc Code a scientia se Dia se	Discollance	Coloniation along and		Davis Davis
DimSubscriptionPlan	PlanName	Subscription plan name		Basic Rental
	PlanCurrent	Flag gor if the plan is current		1

# Detail-Level Dimensional Modeling Worksheet

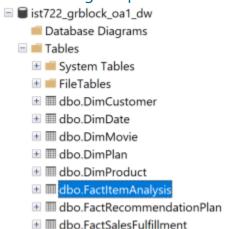
See Excel file IST722\_M400\_Team1\_Detailed\_Dim\_Model\_v2

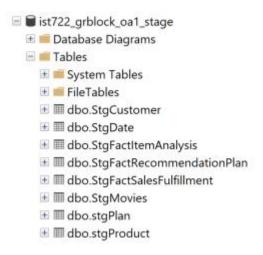
### Data Warehouse on SQL Server

Star Schema



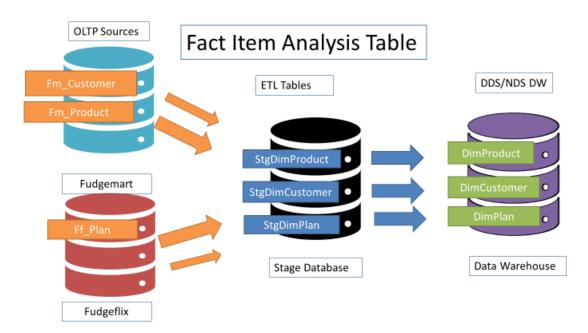
### DW and Stage Implementation

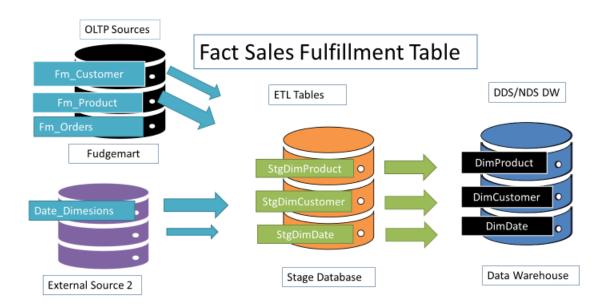




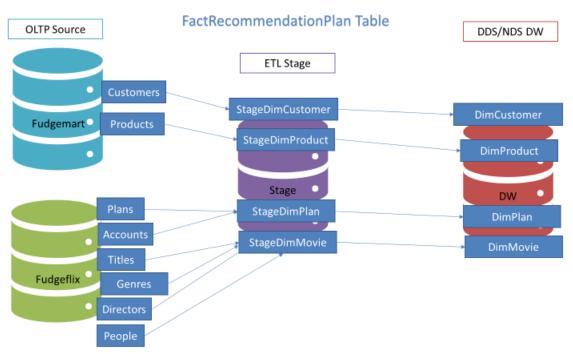
### **ETL In SSIS**

### Source to Target Maps



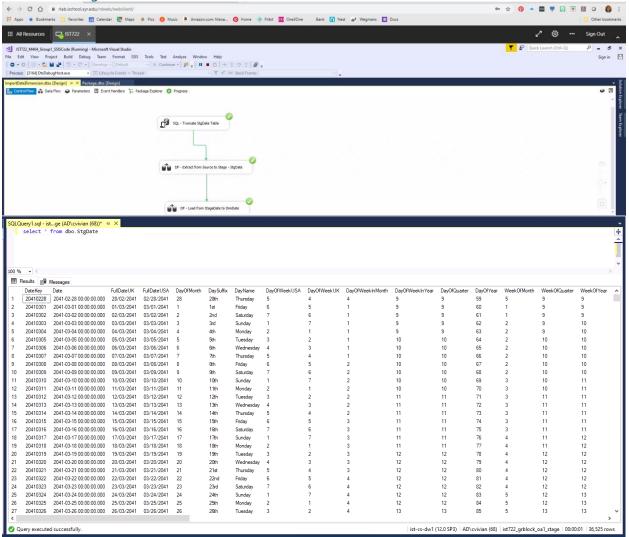


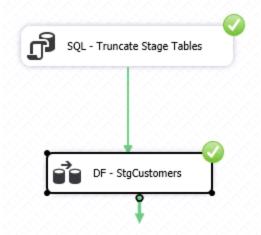
IST 722 | M400 | Team 1

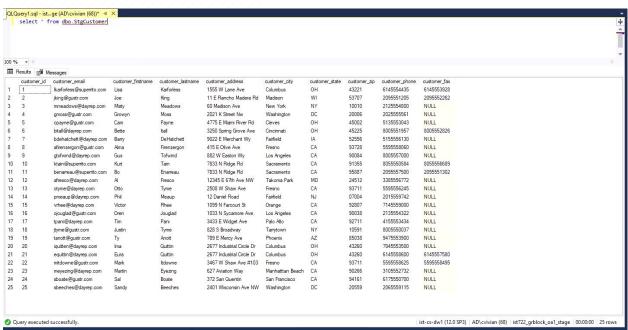


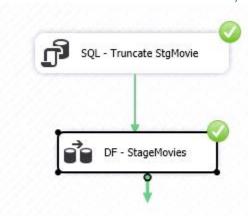
### SSIS Package

Source to Stage



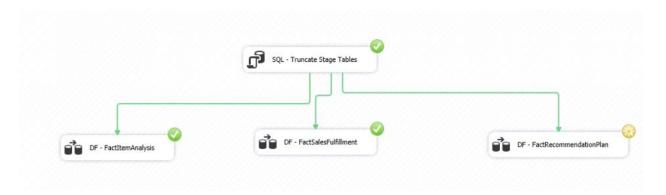




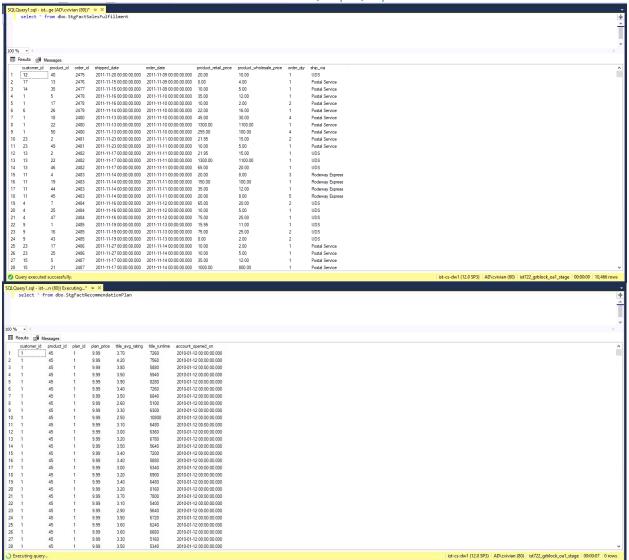


Query executed successfully.

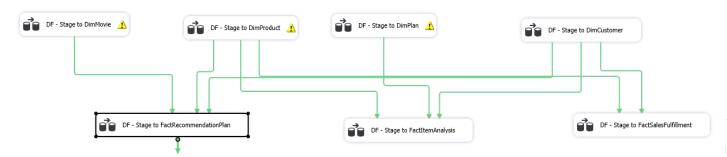
SQLQuery1.sql - ist...ge (AD\cvivian (69))\* 🖼 🗙 select \* from dbo.StgMovies 100 % title\_id title\_rating director\_people\_id title\_name tg\_genre\_name people\_name Pleasantville PG-13 13kZA 20th Century Period Pieces Gary Ross 80023 2 Apt Pupil 13kZs R 20th Century Period Pieces 158689 Bryan Singer 13kZV 3 Velvet Goldmine R 20th Century Period Pieces Todd Havnes 40456 4 Who Shot Pat? 1CZuz R 20th Century Period Pieces Robert Brooks 20005355 5 Digging to China 1EGtv PG 20th Century Period Pieces Timothy Hutton 44302 6 1EJQ4 The Newton Boys PG-13 20th Century Period Pieces Richard Linklater 20000542 Shadrach 1EJTS PG-13 20th Century Period Pieces Susanna Styron 11639765 8 Without Limits 1EJXv PG-13 20th Century Period Pieces Robert Towne 93295 9 A Bronx Tale 5bCa R 20th Century Period Pieces Robert De Niro 22588 10 Cat on a Hot Tin Roof 5gvP NR 20th Century Period Pieces Jack Hofsis 20001073 PG-13 Chaplin: The Movie 5ile 20th Century Period Pieces Richard Attenborough 11 3667 12 Chinatown 5IRm R 20th Century Period Pieces Roman Polanski 74213 5MyU 58351 13 Badlands PG 20th Century Period Pieces Terrence Malick 14 Circle of Friends 5mZM PG-13 20th Century Period Pieces Pat O'Connor 69199 15 Big Night 5TZE R 20th Century Period Pieces Campbell Scott 83499 16 Big Night 5TZE R 20th Century Period Pieces Stanley Tucci 93928 17 Bom on the Fourth o... 5YII R 20th Century Period Pieces Oliver Stone 89386 18 Dead Presidents 5z7n R 20th Century Period Pieces Allen Hughes 43789 19 Dead Presidents 5z7n R 20th Century Period Pieces Albert Hughes 43788 20 Devil in a Blue Dress 62SV R 20th Century Period Pieces Carl Franklin 31693 21 63mX R 20th Century Period Pieces Barry Levinson 54812 660W 22 The Doors R 20th Century Period Pieces Oliver Stone 89386 23 Drugstore Cowboy 6803 R 20th Century Period Pieces Gus Van Sant 95033 The English Patient 24 6CKF R 20th Century Period Pieces Anthony Minghella 64222 25 Eye of the Needle Richard Marguand 6Ego R 20th Century Period Pieces 20001834 26 Fall Time 6Fio R 20th Century Period Pieces Paul Wamer 159388 27 The Hindenburg 6fwD PG 20th Century Period Pieces Robert Wise 100318 6KD8 R 20th Century Period Pieces 28 Five Comers Tony Bill 8213 29 The Hudsucker Proxy 6koj PG 20th Century Period Pieces Joel Coen 18016 30 The Idolmaker 6n3Q PG 20th Century Period Pieces Taylor Hackford 38130

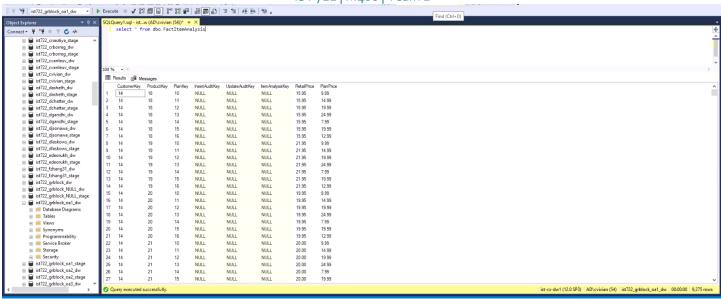


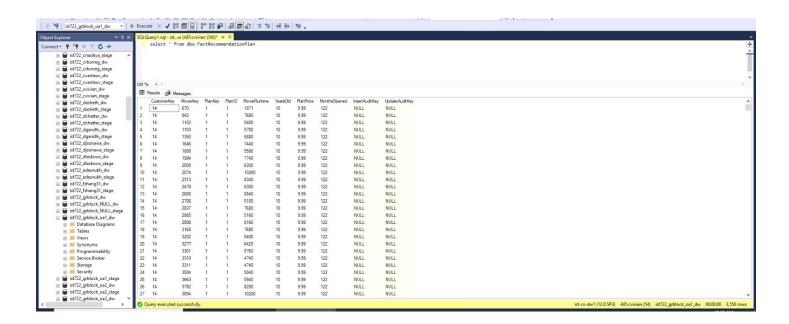


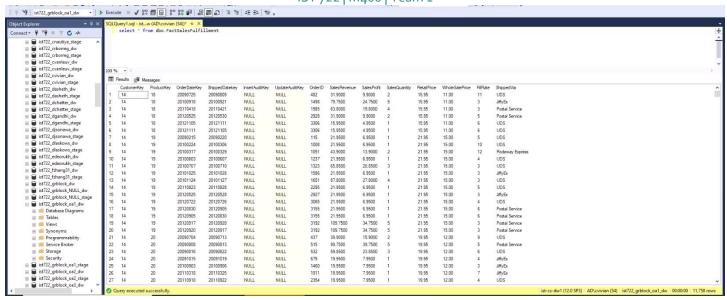


### Stage to DW



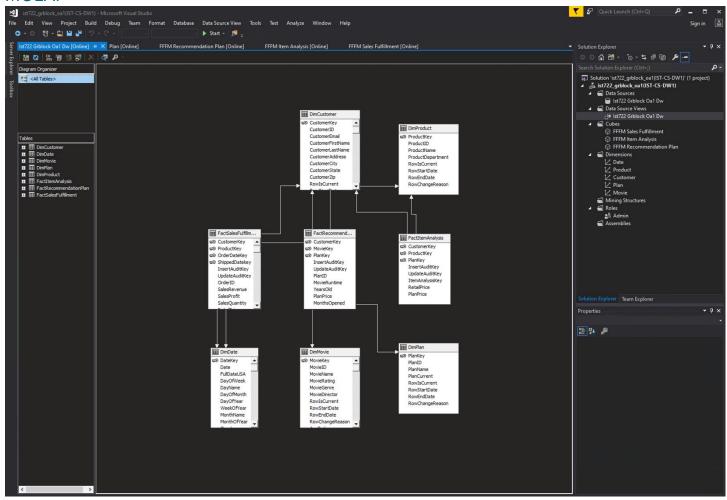


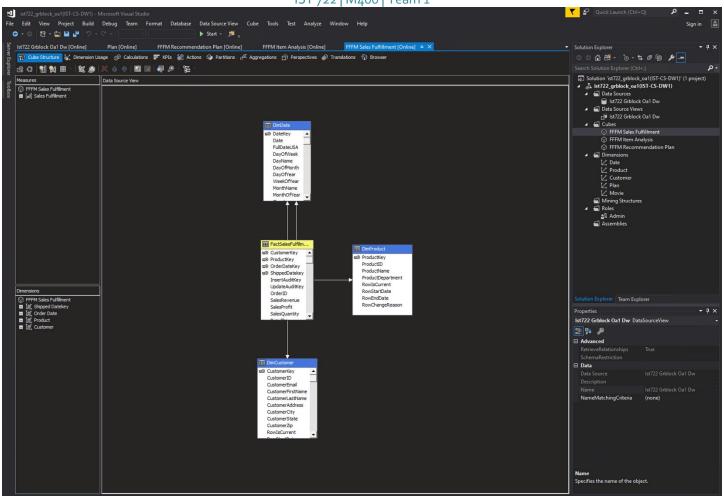




# **Business Intelligence**

### **MOLAP**





### Visualizations

