

Sri Lanka Institute of Information Technology



IT3021
Data Warehousing and Business Intelligence

Spy Plane Finder Data Warehouse Solution
Assignment 1
Document

Submitted By:
Name: Jayasinghe D.T.
ID number: IT19075754

Contents

1. DATASET SELECTION	3
2. PREPARATION OF DATA SOURCES	4
A. TABLE OF AIRCRAFT_TYPE - MYSQL DATABASE (PHPMYADMIN)	4
B. TABLE OF AIRCRAFT_DATA - MYSQL DATABASE (PHPMYADMIN)	5
C. REGISTRANT.TXT (TEXTEDIT).....	5
D. REGISTRANT_ADDRESS.TXT (TEXTEDIT).....	6
E. AGENCY.CSV (NUMBERS MACAPP).....	6
F. AIRCRAFT_TRACKING.CSV (NUMBERS MACAPP).....	6
G. MANUFACTURER TABLE – MICROSOFT SQL DATABASE (SSMS)	7
3. SOLUTION ARCHITECTURE.....	9
4. DATA WAREHOUSE DESIGN & DEVELOPMENT.....	10
5. ETL DEVELOPMENT	13
I. EXECUTE.....	13
II. TRANSFORM.....	13
III. LOAD	14



1. Dataset Selection

This dataset about flights of spy planes operated by the FBI and the Department of Homeland Security (DHS) which is the data comes from more than four months of plane tracking data provided by the website Flightradar24, plus the Federal Aviation Administration's aircraft registration database. The original source files can be found using the links provided below.

Source Links:

1. <https://github.com/BuzzFeedNews/2016-04-federal-surveillance-planes>
2. <https://www.kaggle.com/jboysen/spy-plane-finder>

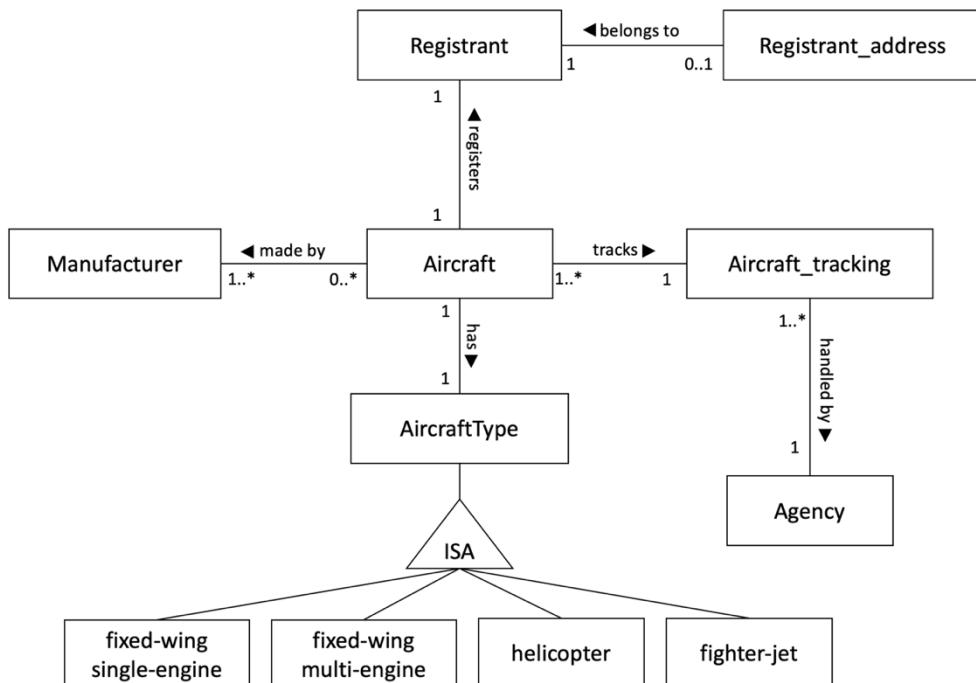
This dataset has more than 397,042 amounts of data.

feds1																				
index	flight_id	latitude	longitude	altitude	speed	track	squawk	type	timestamp	name	other_name1	other_name2	n_number	serial_number	mfr_mdl_code	mfr	model	year_mfr	type_aircraft	agency
A72AA1	7.2791E+12	33.2552	-117.91699	5499	111	137	4401	B350	2015-08-18T07:58:54Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.2659	-117.9208	5500	109	138	4401	B350	2015-08-18T07:58:39Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.2741	-117.93099	5500	109	137	4401	B350	2015-08-18T07:58:28Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.2851	-117.945	5500	112	138	4401	B350	2015-08-18T07:58:13Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.2934	-117.96699	5499	102	134	4401	B350	2015-08-18T07:57:58Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.3028	-117.96699	5499	110	136	4401	B350	2015-08-18T07:57:43Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.3333	-118.001	5499	124	136	4401	B350	2015-08-18T07:56:57Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.3466	-118.0499	5500	108	136	4401	B350	2015-08-18T07:56:42Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.3634	-118.034	5500	100	134	4401	B350	2015-08-18T07:56:12Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.3818	-118.06699	5500	94	129	4401	B350	2015-08-18T07:55:41Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.4242	-118.112	5500	120	137	4401	B350	2015-08-18T07:54:21Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.4629	-118.142	5499	124	138	4401	B350	2015-08-18T07:53:43Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.4838	-118.16699	5500	110	133	4401	B350	2015-08-18T07:53:08Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.5196	-118.203	5500	120	137	4401	B350	2015-08-18T07:52:17Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.5586	-118.23699	5499	124	145	4401	B350	2015-08-18T07:51:12Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	
A72AA1	7.2791E+12	33.568	-118.246	5500	121	136	4401	B350	2015-08-18T07:51:11Z	US DEPARTMENT OF HOMELAND SECURITY	US CUSTOMS & BORDER PROTECTION	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5 dhs	

First few rows of Fed1.csv dataset

I selected this dataset because it consists of a large CSV file along with few small CSV files. Furthermore, I have partitioned the main large CSV file into small sub source files.

When I am selecting the dataset, I make sure to Restrain from selecting a data set which we discussed during practical sessions such as customer/order scenarios as well as Adventure Works data sets. The guidelines also told to look for a novel scenario. So that, I use modified edition of Spy plane finder dataset. That scenario is also cause for talk in the US popular news as well. This dataset has sufficient number of records and attributes to analyse.



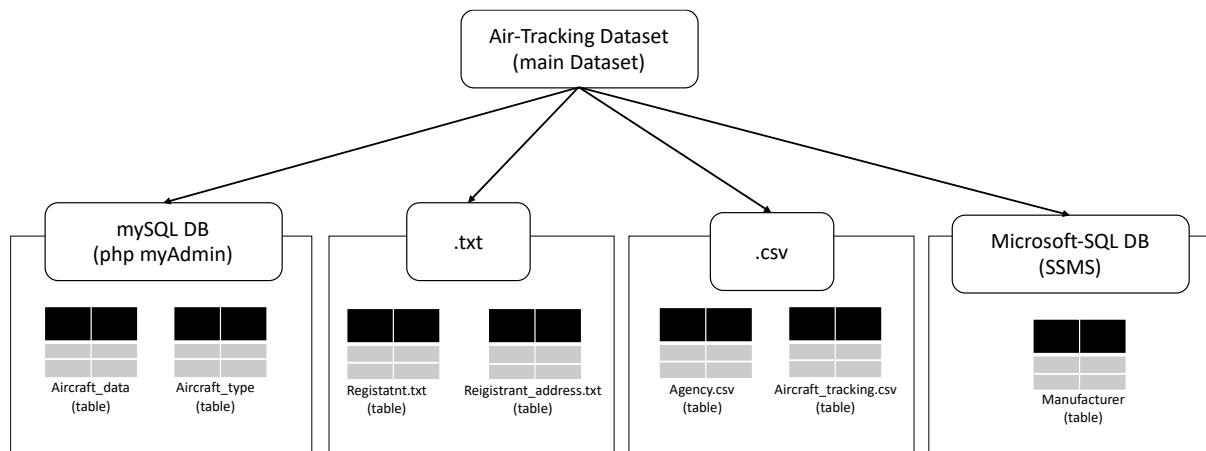
The data set was initiated with sufficient data, according to the assignment principles. First, I got approval for the dataset by one of module related lecturers.

EER diagram for aircraft tracking dataset

2. Preparation of data sources

More than 4 months of aircraft transponder detections from the plane tracking website Flightradar24 had obtained by BuzzFeed news, covering August 17 to December 31, 2015 UTC, containing all data displayed on the site. The United States, Alaska, Hawaii and Puerto Rico are the countries mainly focus on. Fightradar24 receives data from its network of ground-based receivers, supplemented by a feed from ground radars provided by the Federal Aviation Administration (FAA) with a 5min delay.

Then I have partitioned my FAA.csv dataset into different source types like in the diagram given below.



All the snapshots of this partitioned source files are attached below.

a. Table of aircraft_type - MySQL database (phpMyAdmin)

type_name	type_id
fixed-wing single-engine	TID001
fixed-wing multi-engine	TID002
helicopter.	TID003
fighterjer	TID004

*This **type** is aircraft model, if identified. It has 4 different types which is there in the table. As the picture shows this table has only 4 rows. It's primary key which is consider as the unique ID for this table is **type_id**. That row has no duplicates.*

b. Table of aircraft_data - MySQL database (phpMyAdmin)

n_number	adshex	year_mfr	serial_number	model	mfr_mdl_code	manufact_id	type_id	registrant_id
100WG	A006AF	2008	53814	407	1182154	MAN002	TID003	RID002
102BR	A00C4B	2010	18282264	182T	2072703	MAN011	TID001	RID028
108F	A022E9	2012	FM-53	B300C	4220012	MAN006	TID002	RID013
142LJ	A0AB21	2010	18282272	182T	2072703	MAN011	TID001	RID027
143GM	A0AE77	2010	T20608971	T206H	2073303	MAN011	TID001	RID024
143GS	A0AE7C	2010	T20608970	T206H	2073303	MAN011	TID001	RID030
149CS	A0C462		154581	P3B	5260217	MAN007	TID002	RID037
1558	A0E032	BC-20	A200		1152921	MAN001	TID002	RID033
1559	A0E055	BC-16	A200		1152921	MAN001	TID002	RID033
1560	A0E2D1	BC-09	A200		1152921	MAN001	TID002	RID033
159CM	A0EBDC	1987	760333	S-76B	8143007	MAN009	TID003	RID038
161V	A0F740	2011	FM-52	B300C	4220012	MAN006	TID002	RID005
168DK	A10FB2	2012	18282351	182T	2072703	MAN011	TID001	RID028
172AE	A120A0	2004	3839	AS 350 B3	8680954	MAN012	TID003	RID038
182EM	A1488A	2010	18282233	182T	2072703	MAN011	TID001	RID030
183AE	A14BD6		3852	AS 350 B3	8680954	MAN012	TID003	RID034
187AE	A15AB2	2004	3846	AS 350 B3	8680954	MAN012	TID003	RID038
189CB	A1624F	2015	T20600489	T206H	2073343	MAN008	TID004	RID028

This is a database source type like the **data_type** which is created using phpMyAdmin. This table has 188 rows and 9 columns. Its primary key is **adshex**. Moreover, **type_id** act as a foreign key to the **aircraft_type** table which is there in the same database called '**aircraft_data_db**'. Definitions for some names of columns are given after the table instructions.

c. registrant.txt (TextEdit)

```
registrant_id registrant_name other_names1 other_names2
RID001 AEROGRAPHICS INC
RID002 AIRCRAFT ASSOCIATES INC
RID003 AV FLIGHT INC
RID005 DEPARTMENT OF HOMELAND SECURITY
RID006 DEPARTMENT OF HOMELAND SECURITY CUSTOMS & BORDER PROTECTION
RID010 DEPARTMENT OF HOMELAND SECURITY CUSTOMS AND BORDER PROTECTION      ATTN: FREDERICK LEISLING
RID011 DEPARTMENT OF HOMELAND SECURITY CUSTOMS AND BORDER PROTECTION      C/O GREG KENNEY
RID012 DEPARTMENT OF HOMELAND SECURITY US CUSTOMS AND BORDER PROTECTION OFFICE OF AIR AND MARINE
RID017 FVX RESEARCH
RID018 KOM AVIATION
RID019 LCB LEASING
RID022 NATIONAL AIRCRAFT LEASING CORPORATION
RID023 NBR AVIATION
RID024 NBY PRODUCTIONS
RID025 NG RESEARCH
RID026 NORTHWEST AIRCRAFT LEASING CORP
RID027 OBR LEASING
RID028 OTV LEASING
RID029 PSL SURVEYS
RID030 PWX SERVICES
RID031 RKT PRODUCTIONS
RID032 T A F Y CONSULTING
RID039 WORLDWIDE AIRCRAFT LEASING CORP
RID004 DC DEPT OF HOMELAND SECURITY      US CUSTOMS AND BORDER PATROL
```

This is table of registrant who is responsible for register the aircrafts. It has 39 rows and 4 columns. **Registrant_id** is act as the primary key in this registrant table. Other_name1 and other_name2 can have null values. Registrant details had no address table so that I took addresses from another table to make this data set more efficient. The sample address set had huge number of records. In that case, I choose few from the dataset. Its shows in below.

d. registrant_address.txt (TextEdit)

registrant_id	postcode	street	address	City	Country
RID001	11235	Sheepshead Bay Rd	"1706 Sheepshead Bay Rd, Brooklyn, NY"	New York US	
RID002	10022	East 54th Street	"55 East 54th Street, Manhattan, NY"	New York US	
RID003	10307	amboy roaf	"7517 -19 amboy roaf, Staten Island, DC"	Washington DC US	
RID005	10002	Henry Street	(aka 16-18 Jefferson Street)"177 Henry Street (aka 16-18 Jefferson Street), Manhattan, NY"	New York US	
RID006	10022	2nd Ave	"1991 2nd Ave, Manhattan, NY"	New York US	
RID010	11205	dekalb ave	"250 dekalb ave, Brooklyn, NY"	New York US	
RID011	11372	83rd Street	"37-07 83rd Street, Queens, DC"	Washington DC US	
RID012	10024	Columbus Avenue	"489 Columbus Avenue , Manhattan, NY"	New York US	
RID017	10075	Lexington Avenue	"1129 Lexington Avenue , Manhattan, NY"	New York US	
RID018	11209	5th avenue	"8002 5th avenue, Brooklyn, NY"	New York US	
RID019	11222	Greenpoint Avenue	"31 Greenpoint Avenue, Brooklyn, NY"	New York US	
RID022	11161	40th Avenue	"2519 40th Avenue , Queens, NY"	New York US	
RID024	10038	Peck Slip	"21 Peck Slip, Manhattan, CA"	California US	
RID024	11105	Sixth Street	"247 Sixth Street,Brooklyn, NY"	New York US	
RID025	11105	Ditmars Blvd	"36-23 Ditmars Blvd, Queens, DC"	Washington DC US	
RID026	11201	Jay Street	"68 Jay Street,Brooklyn, NY"	New York US	
RID027	11249	North 8th Street	"101 North 8th Street, Brooklyn, NY"	New York US	
RID028	10022	2nd ave	"1083 2nd ave, Manhattan, NY"	New York US	
RID029	11249	Wythe Ave	"342 Wythe Ave, Brooklyn, NY"	New York US	
RID030	10028	2nd Avenue	"1589 2nd Avenue, Manhattan, NY"	New York US	
RID031	10023	68 STREET	"101 WEST 68 STREET, Manhattan, NY"	New York US	
RID032	11204	65th street	"2007 65th street, Brooklyn, NY"	New York US	
RID039	10012	Spring Street	"14 Spring Street, WC"	Wisconsin US	
RID004	11217	Sixth Ave	"71 Sixth Ave, Brooklyn, NY"	New York US	

Addresses of **registrants** are there in above registrant_address.txt file. This data set has columns such as **registrant_id**, **postcode**, **street**, **address**, **city**, **country**.

e. Agency.csv (Numbers macApp)

agency_id	agency_code	agency_name
AGC001	dns	Department of Homeland Security
AGC002	fbi	Federal Bureau of Investigation

Agency is Federal agency operating the aircraft, recorded by BuzzFeed News. There are two rows and **agency_id** is the primary key in this agency table. **Agency_id**, **agency_code** and **agency_name** are the columns of the table **agency**.

f. Aircraft_tracking.csv (Numbers macApp)

adshek	flight_id	latitude	longitude	altitude	speed	track	squawk	timestamp	agency_id
A006AF	716369c	35.04171	-106.67799	5700	75	275	0	11/4/2015 21:40	AGC002
A006AF	7e345bb	36.2514	-115.18099	2700	56	18	1024	11/12/2015 22:00	AGC002
A006AF	7e92528	36.2414	-115.37899	5981	0	278	4414	11/11/2015 7:45	AGC002
A006AF	7122072	36.23011	-115.258	2900	0	287	1200	11/7/2015 7:38	AGC002
A006AF	7e3ef7b	35.1657	-106.78	5925	118	95	0	11/5/2015 3:25	AGC002
A006AF	7e3ef7b	35.17121	-106.81299	5900	121	67	0	11/5/2015 3:25	AGC002
A006AF	7e345bb	35.0396	-106.65399	5700	59	270	0	11/4/2015 21:39	AGC002
A00C4B	7428d8d	42.5227	-83.24679	4050	101	120	0	9/8/2015 22:05	AGC002
A00C4B	7428d8d	42.2064	-83.4369	5425	51	255	0	10/8/2015 23:52	AGC002
A00C4B	7428d8d	42.3986	-83.2864	5475	155	109	4414	12/24/2015 22:31	AGC002
A00C4B	7428d8d	42.3637	-87.9723	5100	100	180	4414	12/20/2015 5:20	AGC002
A00C4B	7428d8d	42.3702	-83.2579	4650	172	265	4414	12/24/2015 19:14	AGC002
A00C4B	7428d8d	42.3531	-83.1286	3300	91	271	4707	8/28/2015 23:33	AGC002
A00C4B	7428d8d	42.35751	-83.1088	3250	82	326	4707	8/28/2015 23:33	AGC002
A00C4B	7428d8d	42.35521	-83.0964	3250	59	346	4707	8/28/2015 23:33	AGC002
A00C4B	7428d8d	42.3462	-83.07409	3250	106	62	4707	8/28/2015 23:32	AGC002
A00C4B	7428d8d	42.3432	-83.0773	3250	118	66	4707	8/28/2015 23:32	AGC002
A00C4B	7428d8d	42.3352	-83.1259	3250	127	31	4707	8/28/2015 23:31	AGC002
A00C4B	7428d8d	42.3322	-83.1301	3250	112	19	4707	8/28/2015 23:31	AGC002
A00C4B	7428d8d	42.32351	-83.1401	3250	120	356	4707	8/28/2015 23:31	AGC002
A00C4B	7428d8d	42.28601	-83.1205	3250	144	278	4707	8/28/2015 23:30	AGC002
A00C4B	7428d8d	42.2847	-83.1087	3275	123	218	4707	8/28/2015 23:29	AGC002
A00C4B	7428d8d	42.2884	-83.10339	3250	121	213	4707	8/28/2015 23:29	AGC002
A00C4B	7428d8d	42.30731	-83.0591	3250	138	123	4707	8/28/2015 23:28	AGC002
A00C4B	7428d8d	42.3139	-83.06399	3250	137	117	4707	8/28/2015 23:28	AGC002
A00C4B	7428d8d	42.3199	-83.1344	3250	81	340	4707	8/28/2015 23:27	AGC002
A00C4B	7428d8d	42.30331	-83.1472	3250	103	285	4707	8/28/2015 23:26	AGC002
A00C4B	7428d8d	42.2991	-83.14349	3250	100	279	4707	8/28/2015 23:26	AGC002
A00C4B	7428d8d	42.2942	-83.13	3250	82	258	4707	8/28/2015 23:26	AGC002
A00C4B	7428d8d	42.2896	-83.10689	3250	84	199	4707	8/28/2015 23:25	AGC002
A00C4B	7428d8d	42.2904	-83.10002	3250	104	167	4707	8/28/2015 23:24	AGC002

This table of **aircraft_tracking** has data around 370, 000. This table has lot of measurable values. For an example **latitude, longitude, altitude** as well as **speed, track, squawk** also can take as measurable values. This is the source which contain huge number of records after partitioning. If you have difficult column names it defined below.

g. Manufacturer Table – Microsoft SQL database (SSMS)

	manufact_name	manufact_id
1	BEECH	MAN001
2	BELL HELICOPTER TEXTRON CANADA	MAN002
3	PILATUS	MAN003
4	BELL	MAN004
5	BOMBARDIER INC	MAN005
6	HAWKER BEECHCRAFT CORP	MAN006
7	LOCKHEED	MAN007
8	TEXTRON AVIATION INC	MAN008
9	SIKORSKY	MAN009
10	AMERICAN EUROCOPTER CORP	MAN010
11	CESSNA	MAN011
12	EUROCOPTER	MAN012
13	PIPER	MAN013
14	AMERICAN EUROCOPTER LLC	MAN014
15	BEECHCRAFT CORP	MAN015

This is the table where contains all the manufacturer names as well as ids. Those manufacturers many numbers of aircrafts. It is a table which is there inside **AircraftManufacturer** database. **Manufact_id** is the primary key and **manufact_name** is another column which is there in this **manufacturer** table.

Source file descriptions are given in the table below.

A	B	C	D	E	F
1 Source	Source Type	Object Name	Scheme	Object Type	Description
2 Agency.csv	csv				Includes the details of all agencies who owns tracked aircrafts.
3					
4 Aircraft_data_DB	MySQL Database	aircraft_data	InnoDB	MySQL DB Table	Includes the details of all the aircrafts.
5		aircraft_type	InnoDB	MySQL DB Table	Includes the details of aircraft types.
6					
7 AircraftManufacturer_DB	Microsoft SQL Database	manufacturer	dbo	Microsoft SQL Table	Includes the details of aircraft manufacturers.
8					
9 Registant.txt	txt				Includes the details of all the registant who have registered track aircrafts.
10					
11 Aircraft_tracking.csv	csv				Includes the tracking information of aircrafts owned by agencies.
12					
13 Registant_address.txt	txt				Includes the addresses of all the registant.
14					

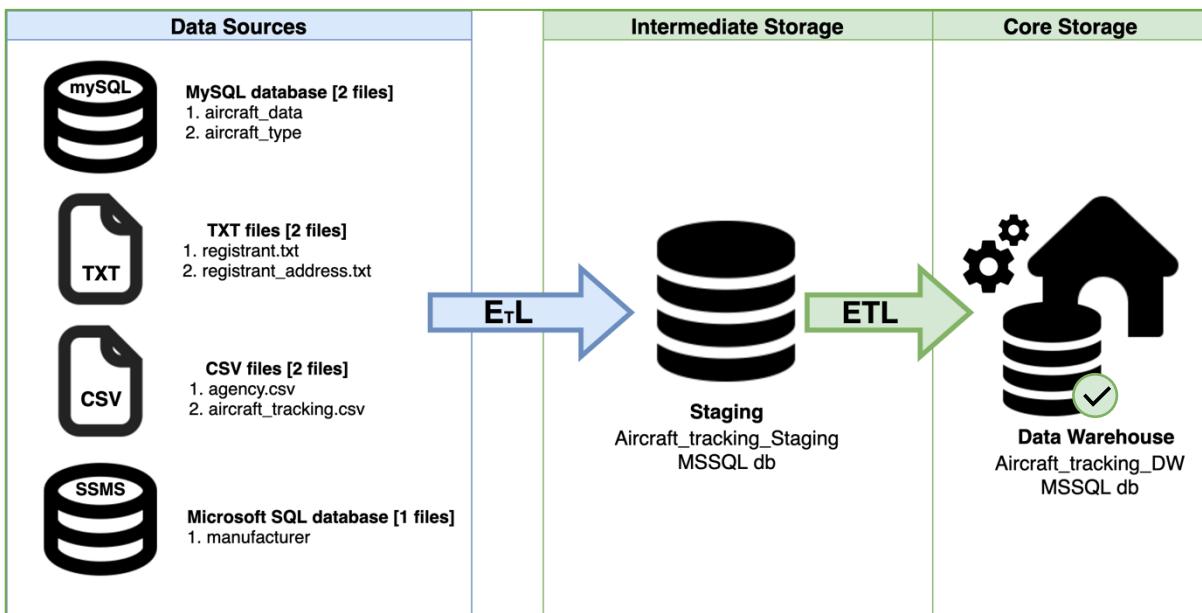
Detailed source file columns descriptions are given in the table below.

A	B	C	D	E	F	G	H
1 Source	Source Type	Table Name	Column Name	Data Type	Link Table	Link Column	Description
2 Agency.csv	csv	Agency	agency_id	nvarchar(5)			Unique agency ID
3			agency_code	nvarchar(5)			Unique 3 letter code given for each agency
4			agency_name	nvarchar(50)			name of the agency
5							
6 Aircraft_data_DB	MySQL Database	aircraft_data	adshex	nvarchar(10)			Unique ID given for each aircraft
7			n_number	nvarchar(20)			Unique registration number given when the aircraft was registered
8			year_mfr	int			manufactured year
9			serial_number	nvarchar(20)			serial number of the aircraft
10			model	nvarchar(20)			model of the aircraft
11			mfr_mdl_code	int			model code assigned by the manufacturer
12			manufact_id	nvarchar(5)	Manufacturer	manufact_id	ID of the manufacturer
13			type_id	nvarchar(5)	Aircraft_type	type_id	Corresponding ID of the type of the aircraft
14			registant_id	nvarchar(5)	Registant	registant_id	(ID of the registant who registered the aircraft
15							
16							
17 aircraft_type			type_id	nvarchar(5)			Unique ID of the aircraft type
18			type_name	nvarchar(50)			4: fixed-wing single-engine, 5: fixed-wing multi-engine, 6: helicopter
19 AircraftManufacturer_DB	Microsoft SQL Database	manufacturer	manufact_id	nvarchar(5)			Unique ID of the manufacturer
20			manufact_name	nvarchar(50)			name of the manufacturer
21							
22 Registant.txt	txt	Registant	registant_id	nvarchar(5)			Unique ID of the registant
23			registant_name	nvarchar(50)			Name of aircraft registant
24			other_names1	nvarchar(50)			Other names for the registant, if listed
25			other_names2	nvarchar(50)			Other names for the registant, if listed
26							
27 registant_address.txt	txt	Registant_address	registant_id	nvarchar(5)			Unique ID of the registant
28			postocode	nvarchar(20)			postalcde of the address of the registant
29			city	nvarchar(20)			city which the given address belongs to
30			address	nvarchar(50)			address of the registant
31			country	nvarchar(10)			country of the registant
32							
33 Aircraft_tracking.csv	csv	Aircraft_tracking	flight_id	nvarchar(50)			Unique ID for each "flight segment," in hexadecimal format. A flight segment is a continuous series of transponder definitions for one aircraft. There may be more than one segment per flight, if a plane disappears from FlightAware's coverage for a period
34			adshex	nvarchar(10)	Aircraft_data	adshex	Unique ID of the aircraft
35			latitude	float			Geographic location in decimal degrees
36			longitude	float			Geographic location in digital degrees
37			altitude	int			Altitude in feet
38			speed	int			Ground speed in knots
39			track	int			Compass heading in degrees, with 0 corresponding to north
40			regisrk	int			For aircraft code transmitted by the transponder
41			timestamp	datetime			Exact time when the aircraft was being tracked
42			agency_id	nvarchar(5)	Agency	agency_id	Federal agency operating the aircraft
43							

Names of columns are defined in the table given below.

adshex	<i>Unique identifier for each aircraft, corresponding to its “Mode-S” code, in hexadeciml format.</i>
flight_id	<i>Unique identifier for each “flight segment,” in hexadecimal format. A flight segment is a continuous series of transponder detections for one aircraft. There may be more than one segment per flight, if a plane disappears from FlightRadar24’s coverage for a period — for example when flying over rural areas with sparse receiver coverage. While being tracked by FlightRadar24, surveillance planes were typically detected several times per minute.</i>
latitude, longitude	<i>Geographic location in digital degrees.</i>
altitude	<i>Altitude in feet.</i>
speed	<i>Ground speed in knots</i>
track	<i>Compass bearing in degrees, with 0 corresponding to north.</i>
squawk	<i>Four-digit code transmitted by the transponder</i>
timestamp	<i>Full UTC timestamp</i>
name	<i>Name of aircraft registrant.</i>
other_names1, other_names2	<i>Other names for the registrant, if listed.</i>
n_number	<i>Aircraft registration number, sometimes called a “tail number.” For U.S.-registered planes, these begin with the letter “N,” followed by up to five alphanumeric characters.</i>
serial_number	<i>Identifying number assigned to the aircraft by its manufacturer.</i>
mfr_mdl_code	<i>Code designating the manufacturer and model of the aircraft.</i>
mfr	<i>Manufacturer.</i>
model	<i>Aircraft model.</i>
year_mfr	<i>Year in which aircraft was manufactured.</i>
type_aircraft	<i>fixed-wing single-engine, fixed-wing multi-engine, helicopter, fighter-jet</i>

3. Solution architecture



The data warehouse is the core of the BI system. The data warehouse is a database built for the purpose of data analysis and reporting. According to my scenario my data warehouse named as '**Aircraft_tracking_DW**'.

Inside the data warehouse database, I manually created tables which are given below.

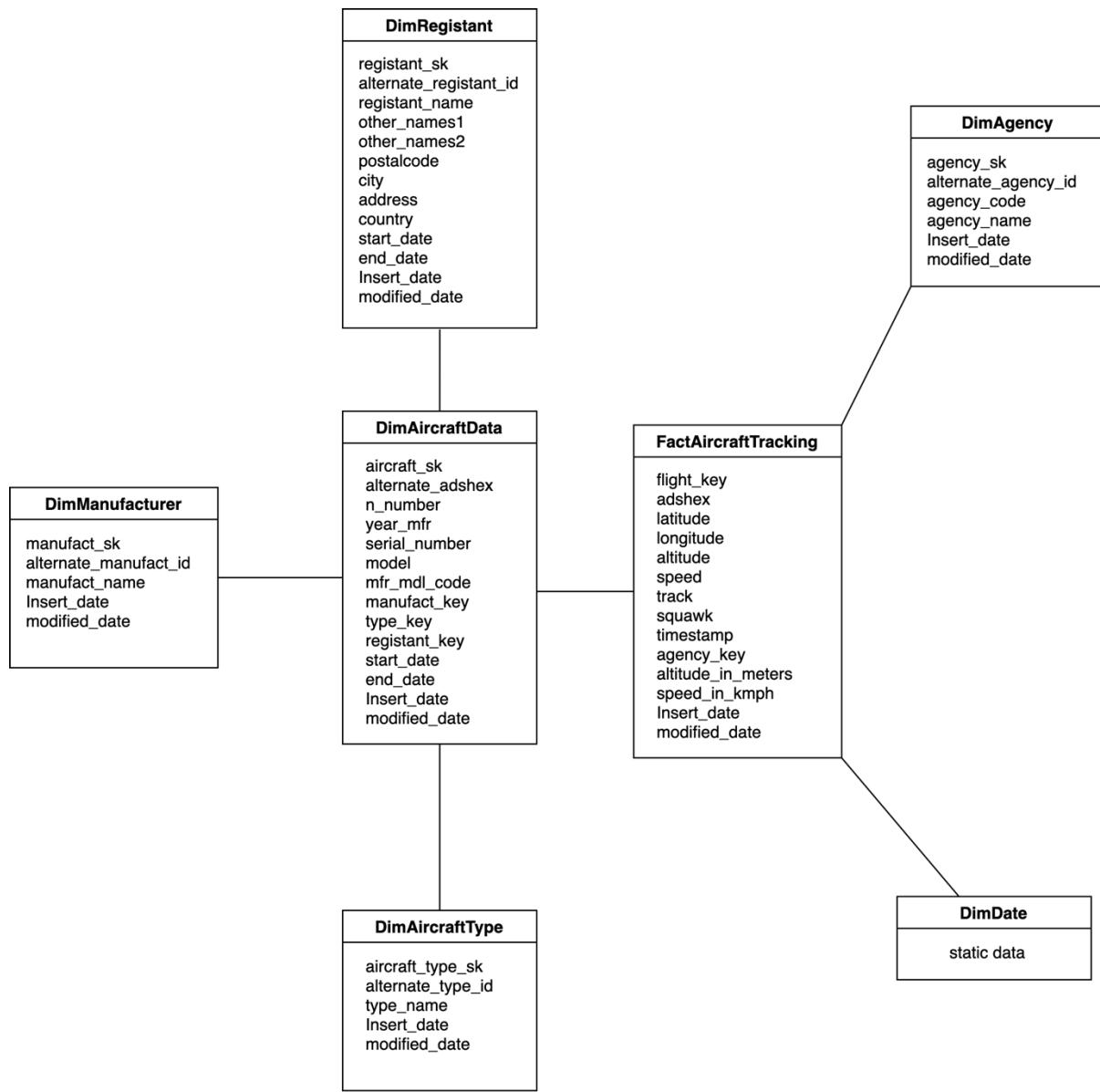
- **DimAgency**
- **DimAircraftData**
- **DimAircraftType**
- **DimManufacturer**
- **DimRegistrant**
- **DimDate**
- **FactAircraftTracking**

Staging layer mainly doing executing and loading tasks, transforming is not a main task there. Our staging database is named as '**Aircraft_tracking_staging**'.

According to the given architecture above you can get an idea about the BI solution to the data warehouse design.

I have used 4 data sources for my scenario. Those data sources are briefly described above and picture given above also says.

4. Data warehouse design & development

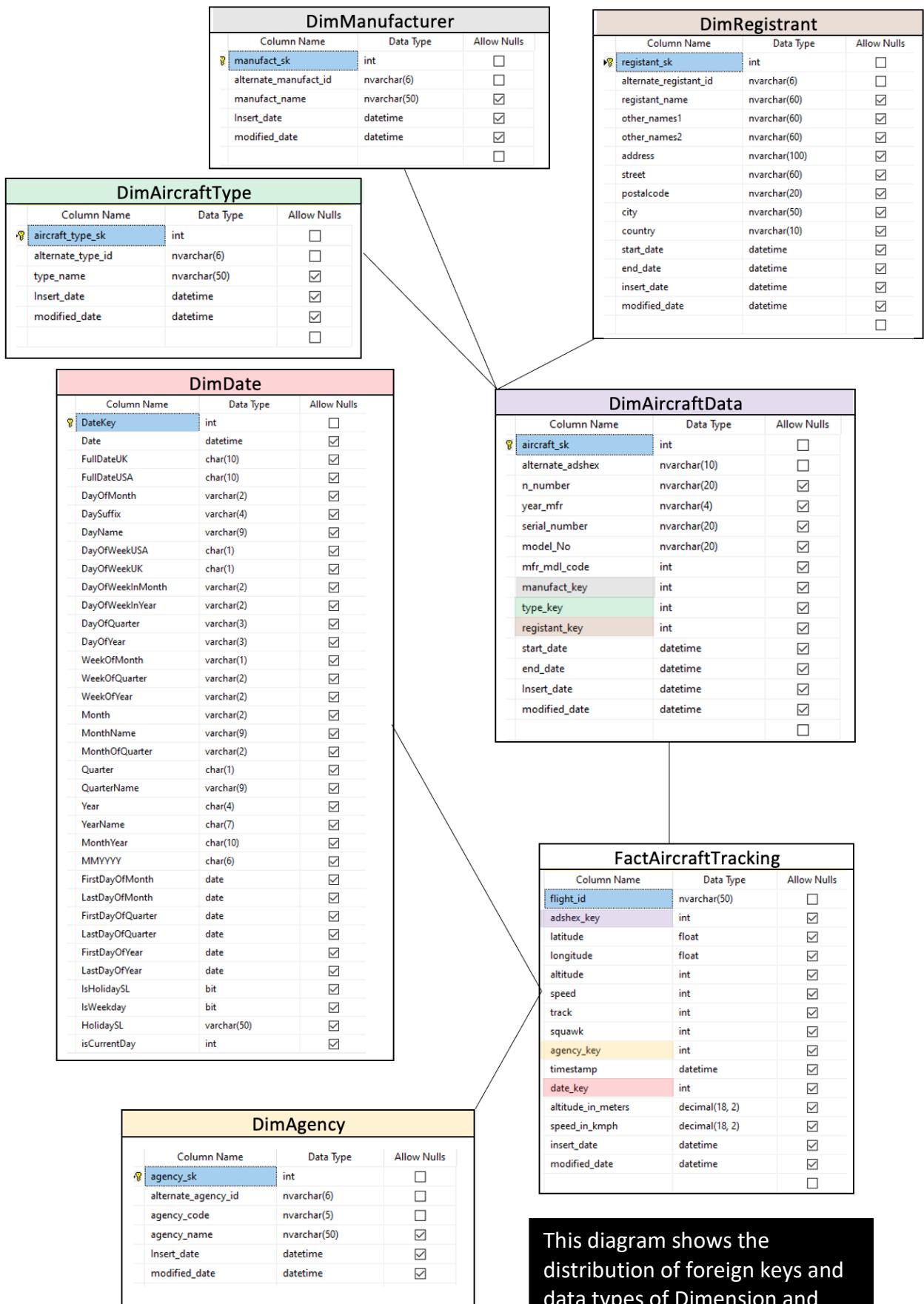


Assumptions:

I have taken DimRegistant as a slowly changing dimension. I need to keep track of the historical dates assuming that, registrant addresses are change time to time.

DimDate table is a static table. in that case I named the data inside the DimDate table as static data. All the data in the Dimdate table is given in the diagram below with its datatype.

As you can see, this is a snowflake schema which has normalized dimension tables. Snowflake schemas will use less space to store dimension tables but are more complex.



This diagram shows the distribution of foreign keys and data types of Dimension and fact Tables

This table given below described the features of dimension tables and fact table.

A	B	C	D	E	F	G	H	I	J
Dimension Name	Truncate Before Update?	Dimension Attributes	Derived Attribute?	Data Type	Null/ Not Null	Key Column?	Data Length	Derived Logic	Description
DimAgency	No	agency_sk	Yes	int	Not Null	Yes (PK)		Surrogate key	Dimension Surrogate key
		agency_name_id	No	nvarchar(6)	Not Null	No	6		
		agency_code	No	nvarchar(10)	Not Null	No	5		
		agency_name	No	nvarchar(50)	Null	No	50		
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
DimAircraftData	No	aircraft_sk	Yes	int	Not Null	Yes (PK)		Surrogate key	Dimension Surrogate key
		alternate_adshex	No	nvarchar(10)	Null	No	10		
		n_number	No	nvarchar(20)	Null	No	20		
		year_mfr	No	int	Null	No			
		serial_number	No	nvarchar(20)	Null	No	20		
		model	No	nvarchar(20)	Null	No	20		
		mfr_mdl_code	No	int	Null	No			
		manufact_key	No	int	Not Null	Yes (FK)			
		type_key	No	int	Not Null	Yes (FK)			
		registant_key	No	int	Not Null	Yes (FK)			
		start_date	Yes	datetime	Not Null	No			
		end_date	Yes	datetime	Not Null	No			
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
DimAircraftType	No	aircraft_type_sk	Yes	int	Not Null	Yes (PK)		Surrogate key	Dimension Surrogate key
		alternate_type_id	No	nvarchar(6)	Null	No	6		
		type_name	No	nvarchar(50)	Null	No	50		
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
DimManufacturer	No	manufact_sk	Yes	int	Not Null	Yes (PK)		Surrogate key	Dimension Surrogate key
		alternate_manufact_id	No	nvarchar(6)	Null	No	6		
		manufact_name	No	nvarchar(50)	Null	No	50		
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
DimRegistant	No	registant_sk	Yes	int	Not Null	Yes (PK)		Surrogate key	Dimension Surrogate key
		alternate_registant_id	No	nvarchar(6)	Null	No	6		
		registant_name	No	nvarchar(60)	Null	No	60		
		other_name1	No	nvarchar(60)	Null	No	60		
		other_name2	No	nvarchar(60)	Null	No	60		
		postocode	No	nvarchar(20)	Null	No	20		
		city	No	nvarchar(20)	Null	No	20		
		address	No	nvarchar(60)	Null	No	60		
		country	No	nvarchar(10)	Null	No	10		
		start_date	Yes	datetime	Not Null	No			
		end_date	Yes	datetime	Null	No			
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
FactAircraftTracking	No	flight_key	No	int	Not Null	Yes (PK)			
		adshex	No	nvarchar(10)	Null	No	10		
		latitude	No	float	Null	No			
		longitude	No	float	Null	No			
		altitude	No	int	Null	No			
		speed	No	int	Null	No			
		track	No	int	Null	No			
		squawk	No	int	Null	No			
		timestamp	No	datetime	Null	No			
		agency_key	No	int	Not Null	Yes (FK)			
		altitude_in_meters	Yes	decimal(18,2)					(altitude*(0.3048))
		speed_in_kmph	Yes	decimal(18,2)					(speed*(1.852))
		insert_date	Yes	datetime	Not Null	No		timestamp	sysdate
		modified_date	Yes	datetime	Not Null	No		timestamp	sysdate
DimDate	No	static_table							static table

Surrogate Key

Surrogate key is the key acts as primary key or unique id in dimensional model tables.

According to my scenario all the surrogate keys are,

- **agency_sk**
- **aircraft_sk**
- **aircraft_type_sk**
- **manfact_sk**
- **registant_sk**

Calculation

Here as calculation, I have included 2 attributes. Those are,

1. **altitude_in_meters** → ([altitude]*(0.3048))
2. **speed_in_kmph** → ([speed]*(1.852))

Those data types are derived attributes as well.

Static Table

Here in my scenario **DimDate** table is the only table that can consider as a static table.

Simply, it's a dimension table but it simply can create by executing a query which can find via internet. I used the same DimDate table used in my DWBI lab sessions.

Derived Attributes

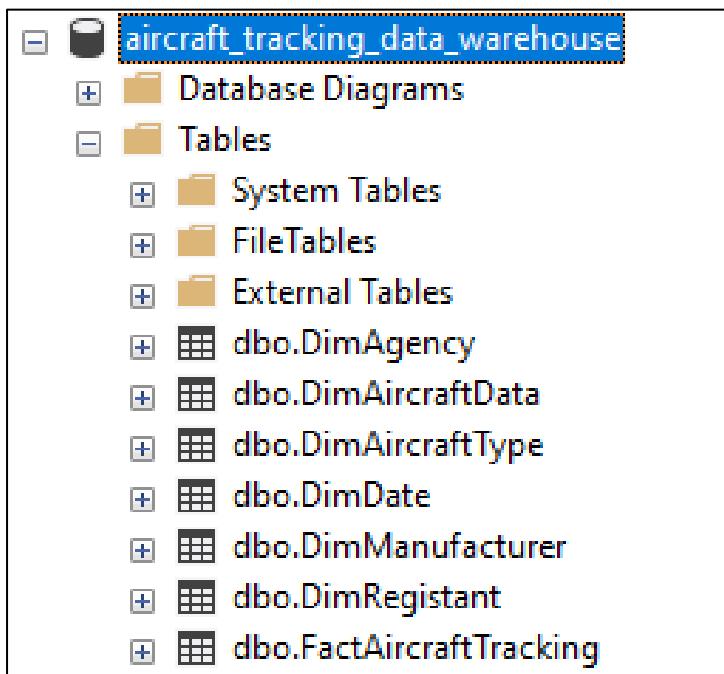
Those are the attributes which are not in the source files of the dataset. Simply, the attributes we created according to the dimension and fact table rules as well as for our preference.

surrogate keys, insert_date, modified_date, start_date, end_date, altitude_in_meters and speed_in_kmph are the attributes which can be consider as derived attributes.

5. ETL development

Snap shots of ETL process are included below.

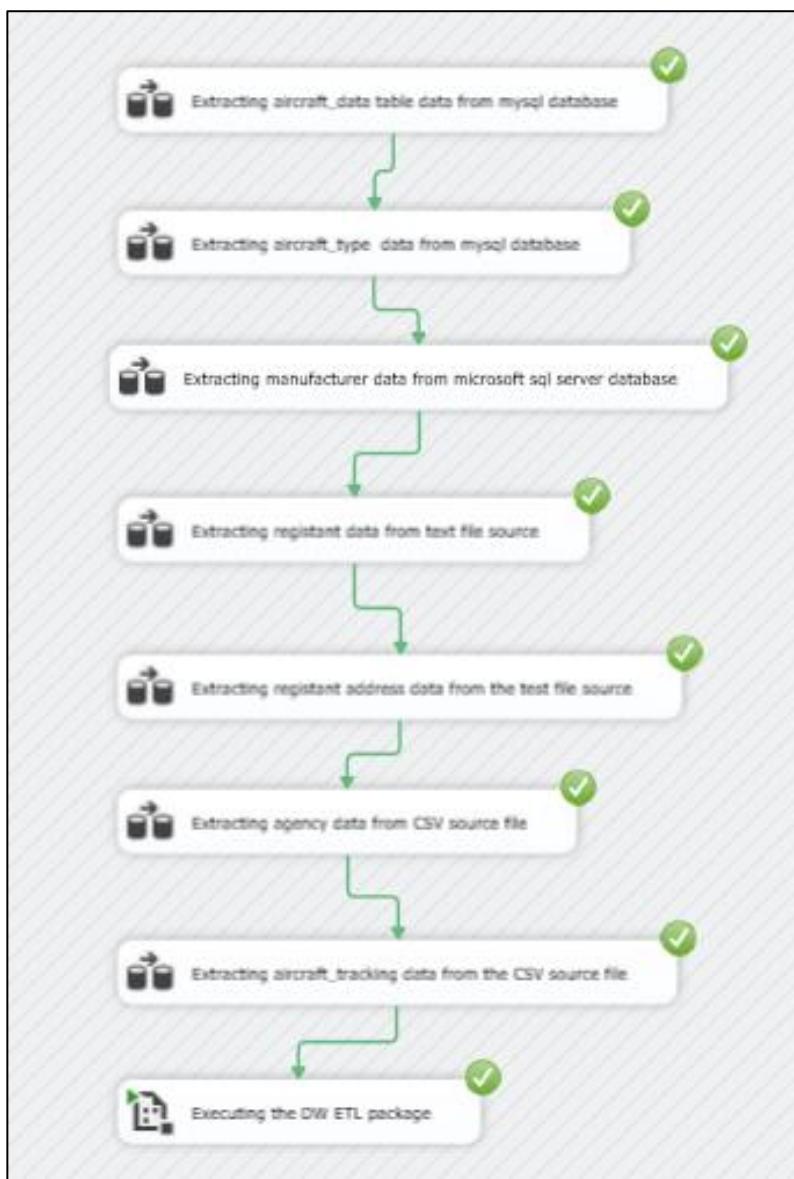
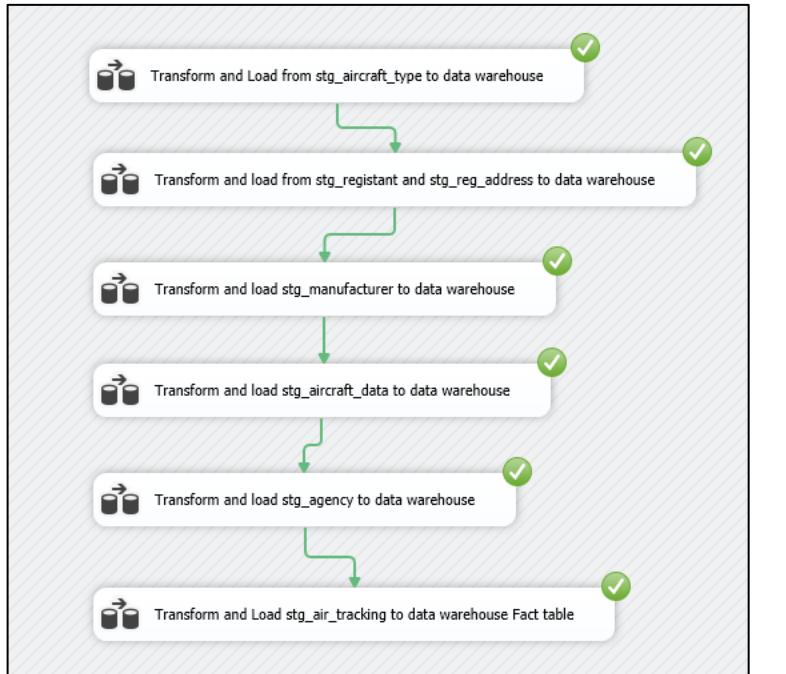
I. Execute



II. Transform



III. Load



Object Explorer

Connect ▾

DESKTOP-7C6K6F8 (SQL Server 13.0.4001.0 - MicrosoftA)

- [-] Databases
 - [+] System Databases
 - [+] Database Snapshots
 - [+] aircraft_manufacturer_source
 - [+] aircraft_tracking_data_warehouse
 - [+] aircraft_tracking_staging
 - [+] Database Diagrams
 - [+] Tables
 - [+] System Tables
 - [+] FileTables
 - [+] External Tables
 - [+] dbo.stg_agency
 - [+] dbo.stg_air_tracking
 - [+] dbo.stg_aircraft_data
 - [+] dbo.stg_aircraft_type
 - [+] dbo.stg_manufacturer
 - [+] dbo.stg_reg_address
 - [+] dbo.stg_resistant
 - [+] Views
 - [+] External Resources
 - [+] Synonyms
 - [+] Programmability
 - [+] Service Broker
 - [+] Storage
 - [+] Security

Staging database after finished loading

ETL process details are given in the table below.

Record counts validating check

```
use aircraft_tracking_data_warehouse;
select count(*) from [dbo].[DimAgency];
select count(*) from [dbo].[DimAircraftData];
select count(*) from [dbo].[DimAircraftType];
select count(*) from [dbo].[DimManufacturer];
select count(*) from [dbo].[DimRegistant];
select count(*) from [dbo].[FactAircraftTracking].
```

Results Messages

(No column name)	1	2
(No column name)	1	186
(No column name)	1	4
(No column name)	1	15
(No column name)	1	24
(No column name)	1	794080

Query executed successfully.

No of columns of data warehouse database

```
use aircraft_tracking_staging;
select count(*) from [dbo].[stg_agency];
select count(*) from [dbo].[stg_air_tracking];
select count(*) from [dbo].[stg_aircraft_data];
select count(*) from [dbo].[stg_aircraft_type];
select count(*) from [dbo].[stg_manufacturer];
select count(*) from [dbo].[stg_reg_address];
select count(*) from [dbo].[stg_registant].
```

Results Messages

(No column name)	1	2
(No column name)	1	397040
(No column name)	1	188
(No column name)	1	4
(No column name)	1	15
(No column name)	1	24
(No column name)	1	24

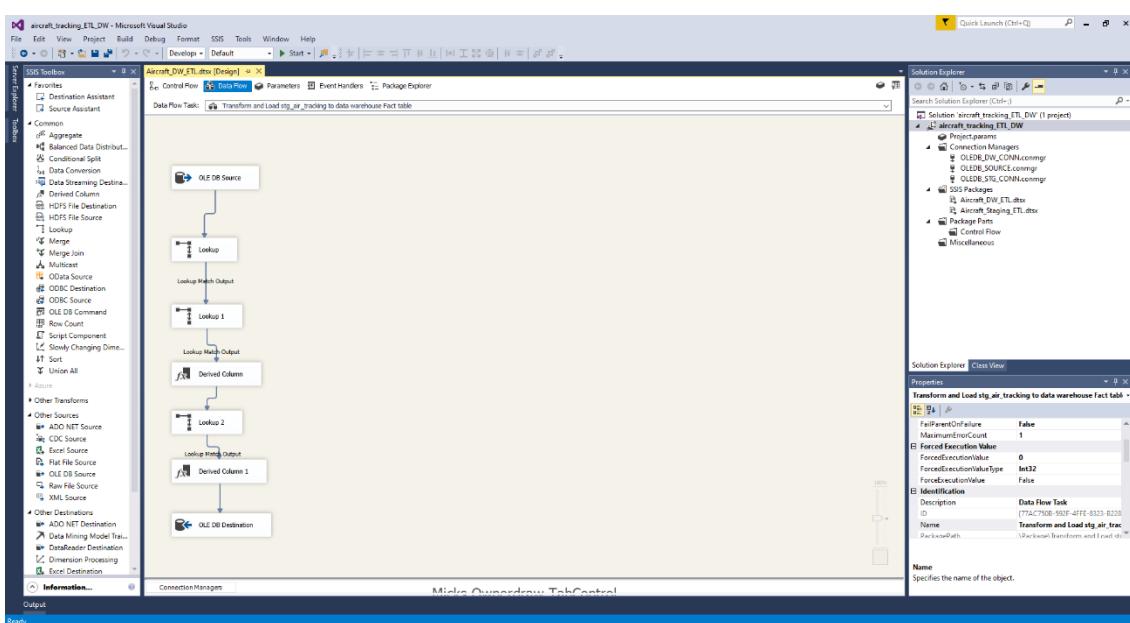
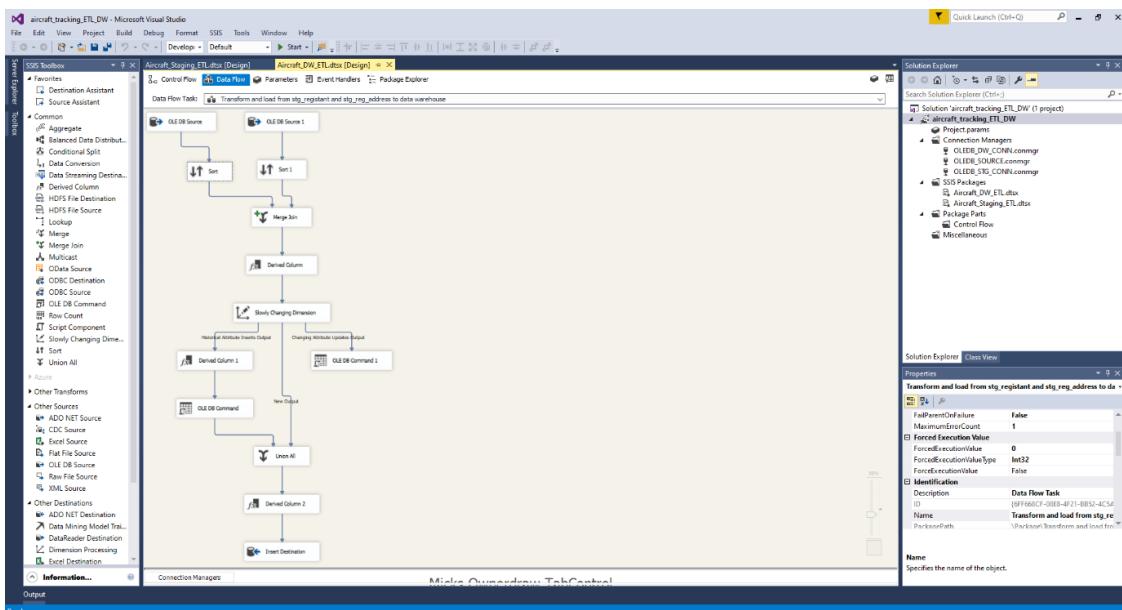
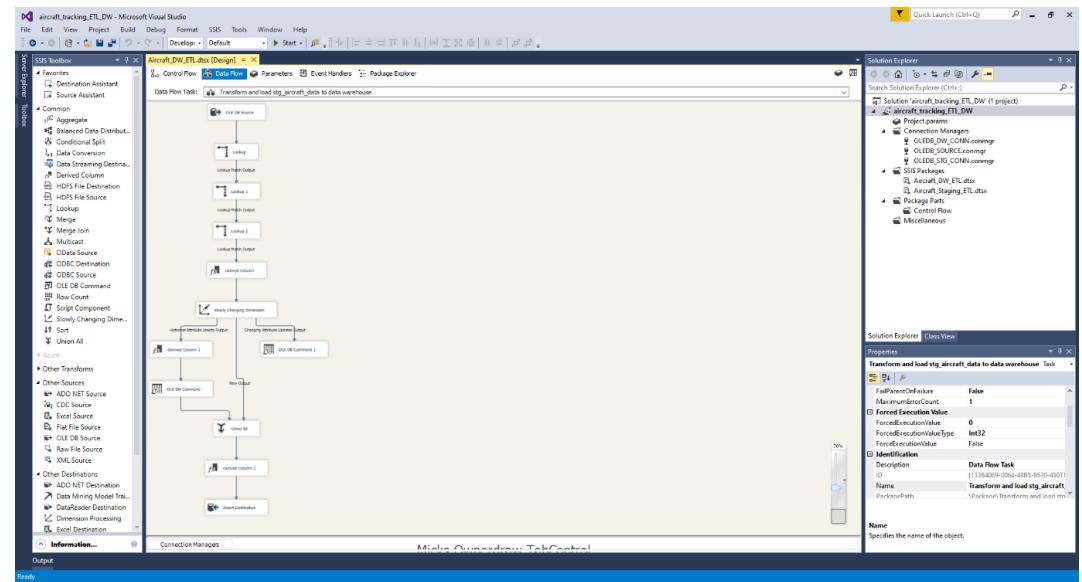
Query executed successfully.

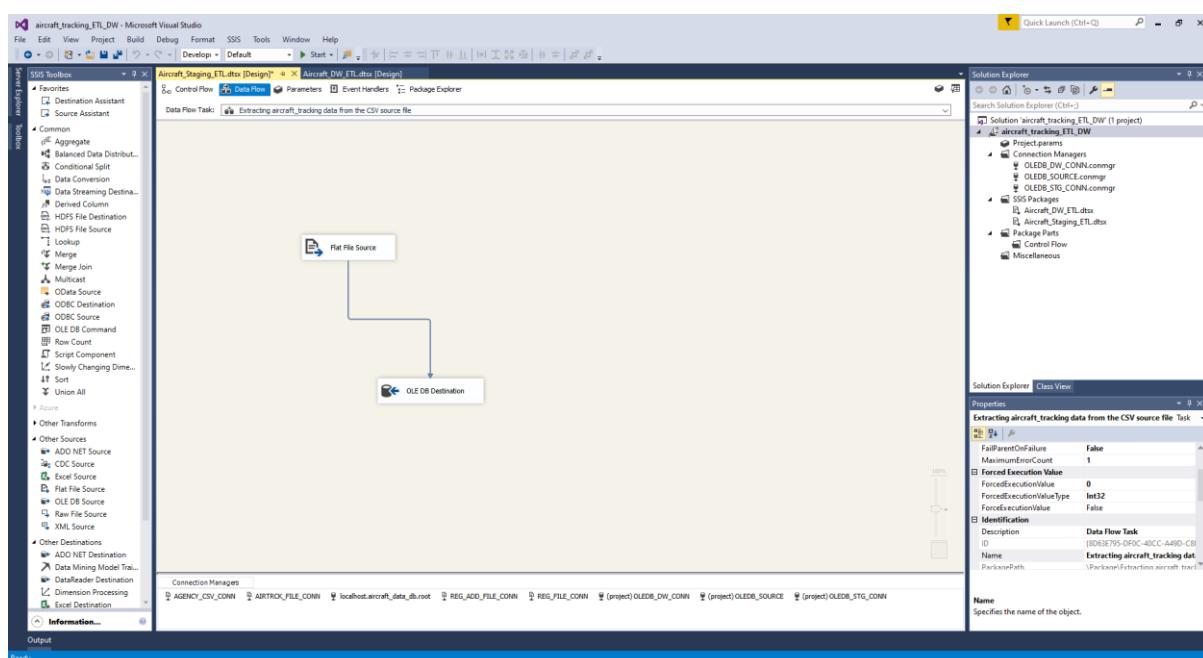
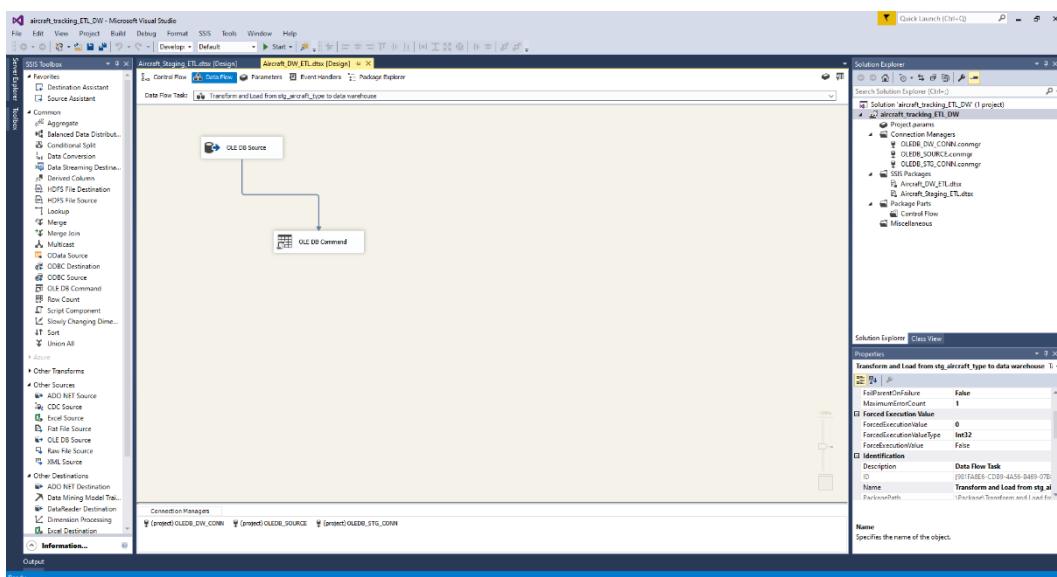
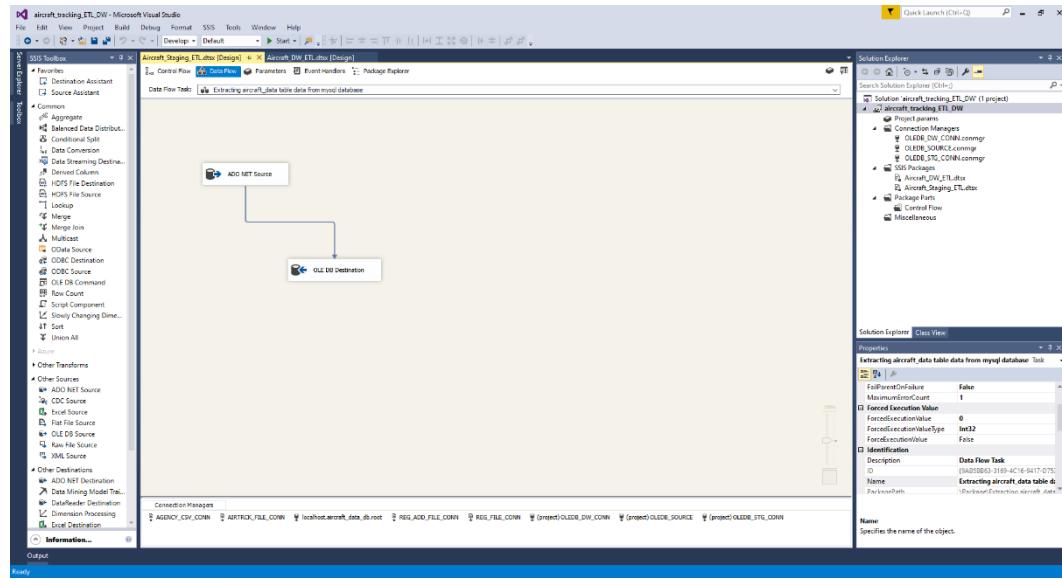
No of columns of staging database

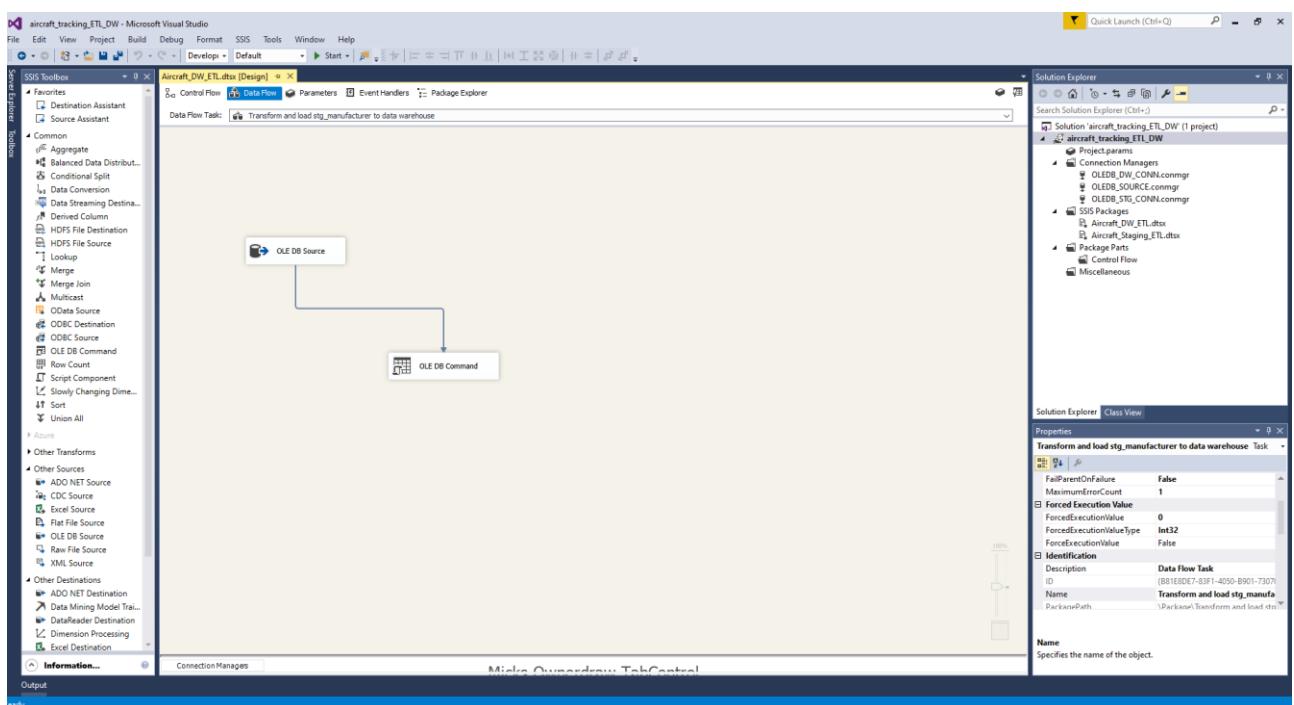
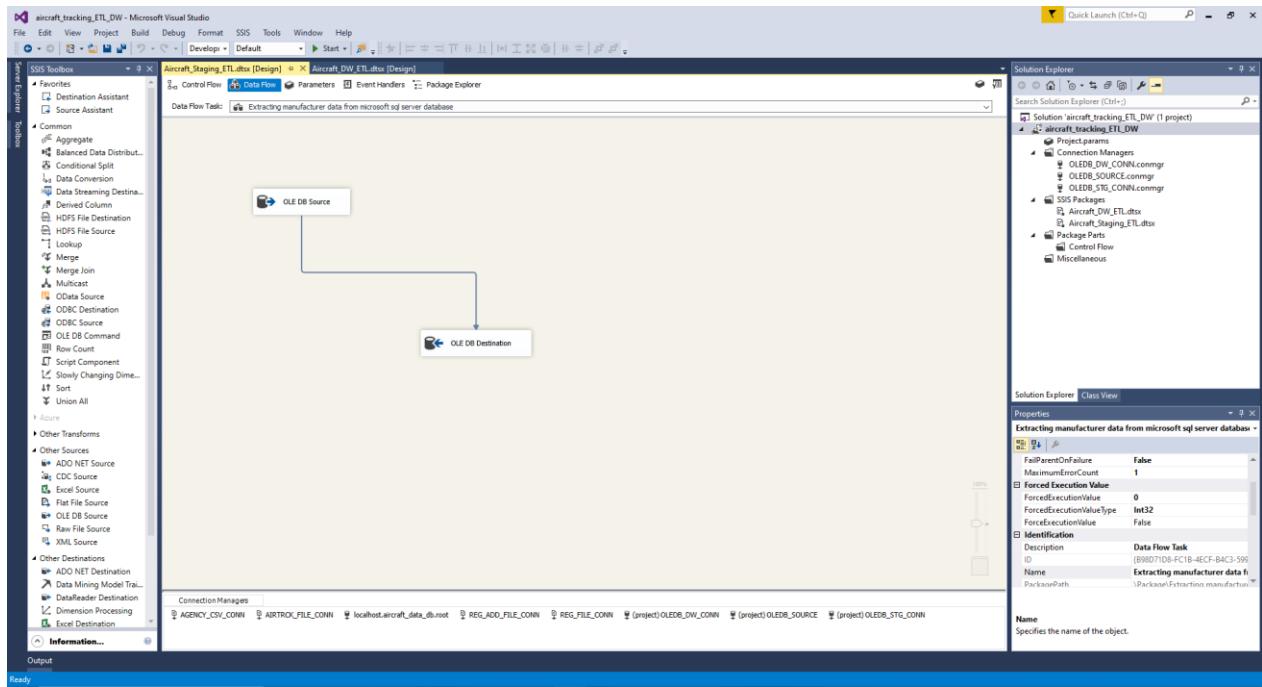
Detailed table of DimDate

	DateKey	Date	FullDateUK	FullDateUSA	DayOfMonth	DaySuffix	DayName	DayOfWeekUSA	DayOfWeekUK	DayOfWeekInMonth	DayOfWeekInYear	DayOfQuarter	DayOfYear	WeekOfMonth	WeekOfQuarter	WeekOfYear	Month	MonthName
1	19900101	1990-01-01 00:00:00.000	01/01/1990	01/01/1990	1	1st	Monday	2	1	1	1	1	1	1	1	1	January	
2	19900102	1990-01-02 00:00:00.000	02/01/1990	01/02/1990	2	2nd	Tuesday	3	2	1	1	2	1	1	1	1	January	
3	19900103	1990-01-03 00:00:00.000	03/01/1990	01/03/1990	3	3rd	Wednesday	4	3	1	1	3	1	1	1	1	January	
4	19900104	1990-01-04 00:00:00.000	04/01/1990	01/04/1990	4	4th	Thursday	5	4	1	1	4	1	1	1	1	January	
5	19900105	1990-01-05 00:00:00.000	05/01/1990	01/05/1990	5	5th	Friday	6	5	1	1	5	1	1	1	1	January	
6	19900106	1990-01-06 00:00:00.000	06/01/1990	01/06/1990	6	6th	Saturday	7	6	1	1	6	1	1	1	1	January	
7	19900107	1990-01-07 00:00:00.000	07/01/1990	01/07/1990	7	7th	Sunday	1	7	1	1	7	2	1	2	1	January	
8	19900108	1990-01-08 00:00:00.000	08/01/1990	01/08/1990	8	8th	Monday	2	1	2	2	2	8	2	2	2	January	
9	19900109	1990-01-09 00:00:00.000	09/01/1990	01/09/1990	9	9th	Tuesday	3	2	2	2	2	9	2	2	2	January	
10	19900110	1990-01-10 00:00:00.000	10/01/1990	01/10/1990	10	10th	Wednesday	4	3	2	2	2	10	2	2	2	January	
11	19900111	1990-01-11 00:00:00.000	11/01/1990	01/11/1990	11	11th	Thursday	5	4	2	2	2	11	2	2	2	January	
12	19900112	1990-01-12 00:00:00.000	12/01/1990	01/12/1990	12	12th	Friday	6	5	2	2	2	12	2	2	2	January	
13	19900113	1990-01-13 00:00:00.000	13/01/1990	01/13/1990	13	13th	Saturday	7	6	2	2	2	13	2	2	2	January	
14	19900114	1990-01-14 00:00:00.000	14/01/1990	01/14/1990	14	14th	Sunday	1	7	2	2	2	14	3	2	3	January	
15	19900115	1990-01-15 00:00:00.000	15/01/1990	01/15/1990	15	15th	Monday	2	1	3	3	3	15	3	3	3	January	
16	19900116	1990-01-16 00:00:00.000	16/01/1990	01/16/1990	16	16th	Tuesday	3	2	3	3	3	16	3	3	3	January	
17	19900117	1990-01-17 00:00:00.000	17/01/1990	01/17/1990	17	17th	Wednesday	4	3	3	3	3	17	3	3	3	January	
18	19900118	1990-01-18 00:00:00.000	18/01/1990	01/18/1990	18	18th	Thursday	5	4	3	3	3	18	3	3	3	January	
19	19900119	1990-01-19 00:00:00.000	19/01/1990	01/19/1990	19	19th	Friday	6	5	3	3	3	19	3	3	3	January	
20	19900120	1990-01-20 00:00:00.000	20/01/1990	01/20/1990	20	20th	Saturday	7	6	3	3	3	20	3	3	3	January	
21	19900121	1990-01-21 00:00:00.000	21/01/1990	01/21/1990	21	21st	Sunday	1	7	3	3	3	21	4	3	4	January	
22	19900122	1990-01-22 00:00:00.000	22/01/1990	01/22/1990	22	22nd	Monday	2	1	4	4	4	22	4	4	4	January	
23	19900123	1990-01-23 00:00:00.000	23/01/1990	01/23/1990	23	23rd	Tuesday	3	2	4	4	4	23	4	4	4	January	
24	19900124	1990-01-24 00:00:00.000	24/01/1990	01/24/1990	24	24th	Wednesday	4	3	4	4	4	24	4	4	4	January	
25	19900125	1990-01-25 00:00:00.000	25/01/1990	01/25/1990	25	25th	Thursday	5	4	4	4	4	25	4	4	4	January	
26	19900126	1990-01-26 00:00:00.000	26/01/1990	01/26/1990	26	26th	Friday	6	5	4	4	4	26	4	4	4	January	
27	19900127	1990-01-27 00:00:00.000	27/01/1990	01/27/1990	27	27th	Saturday	7	6	4	4	4	27	4	4	4	January	
28	19900128	1990-01-28 00:00:00.000	28/01/1990	01/28/1990	28	28th	Sunday	1	7	4	4	4	28	5	4	5	January	
29	19900129	1990-01-29 00:00:00.000	29/01/1990	01/29/1990	29	29th	Monday	2	1	5	5	5	29	5	5	5	January	
30	19900130	1990-01-30 00:00:00.000	30/01/1990	01/30/1990	30	30th	Tuesday	3	2	5	5	5	30	5	5	5	January	
31	19900131	1990-01-31 00:00:00.000	31/01/1990	01/31/1990	31	31st	Wednesday	4	3	5	5	5	31	5	5	5	January	
32	19900201	1990-02-01 00:00:00.000	01/02/1990	02/01/1990	1	1st	Thursday	5	4	1	5	5	32	1	5	5	February	
33	19900202	1990-02-02 00:00:00.000	02/02/1990	02/02/1990	2	2nd	Friday	6	5	1	5	5	33	1	5	5	February	
34	19900203	1990-02-03 00:00:00.000	03/02/1990	02/03/1990	3	3rd	Saturday	7	6	1	5	5	34	1	5	5	February	
35	19900204	1990-02-04 00:00:00.000	04/02/1990	02/04/1990	4	4th	Sunday	1	7	1	5	5	35	2	5	6	February	
36	19900205	1990-02-05 00:00:00.000	05/02/1990	02/05/1990	5	5th	Monday	2	1	6	6	6	36	2	6	6	February	
37	19900206	1990-02-06 00:00:00.000	06/02/1990	02/06/1990	6	6th	Tuesday	3	2	1	6	6	37	2	6	6	February	
38	19900207	1990-02-07 00:00:00.000	07/02/1990	02/07/1990	7	7th	Wednesday	4	3	1	6	6	38	2	6	6	February	
39	19900208	1990-02-08 00:00:00.000	08/02/1990	02/08/1990	8	8th	Thursday	5	4	2	6	6	39	2	6	2	February	

DESKTOP-7C6K6F8 (13.0 SP1) | DESKTOP-7C6K6F8\Dimush... | aircraft_tracking_data... | 00:00:01 | 39,812 rows







-END-