Sri Lanka Institute of Information Technology

IT3021 Data Warehousing and Business Intelligence

Assignment 1

Submitted to Sri Lanka Institute of Information Technology

Rangana PWM

IT19987880

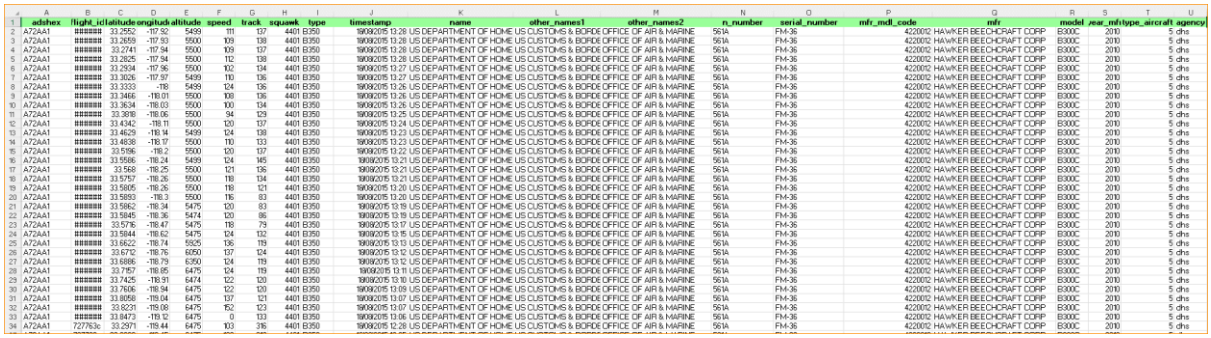
SLIIT
Discover Your Future

Table of Contents

1 Data Set 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.

Data Set Link :- https://github.com/BuzzFeedNews/2016-04-federal-surveillance-planes



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	adshex	flight_id	latitude	longitude	altitude	speed	track	squawk	type	timestamp	name	other_names1	other_names2	n_number	serial_number	mfr_mdl_code	mfr	model	year_mfr	type_aircraft	agency
2	A72AA1	######	33.2552	-117.92	5499	111	137	4401	B350	19/09/2015 13:28	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
3	A72AA1	######	33.2659	-117.93	5500	109	138	4401	B350	19/09/2015 13:28	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
4	A72AA1	######	33.2741	-117.94	5500	109	137	4401	B350	19/09/2015 13:28	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
5	A72AA1	######	33.2825	-117.94	5500	112	138	4401	B350	19/09/2015 13:28	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
6	A72AA1	######	33.2934	-117.96	5500	102	134	4401	B350	19/09/2015 13:27	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
7	A72AA1	######	33.3026	-117.97	5499	110	136	4401	B350	19/09/2015 13:27	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
8	A72AA1	######	33.3333	-118	5499	124	136	4401	B350	19/09/2015 13:26	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
9	A72AA1	######	33.3466	-118.01	5500	108	136	4401	B350	19/09/2015 13:26	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
10	A72AA1	######	33.3634	-118.03	5500	100	134	4401	B350	19/09/2015 13:26	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
11	A72AA1	######	33.3818	-118.06	5500	94	129	4401	B350	19/09/2015 13:25	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
12	A72AA1	######	33.4342	-118.11	5500	120	137	4401	B350	19/09/2015 13:24	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
13	A72AA1	######	33.4629	-118.14	5499	124	138	4401	B350	19/09/2015 13:23	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
14	A72AA1	######	33.4838	-118.17	5500	110	133	4401	B350	19/09/2015 13:23	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
15	A72AA1	######	33.5196	-118.2	5500	120	137	4401	B350	19/09/2015 13:22	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
16	A72AA1	######	33.5586	-118.24	5499	124	145	4401	B350	19/09/2015 13:21	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
17	A72AA1	######	33.568	-118.25	5500	121	136	4401	B350	19/09/2015 13:21	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
18	A72AA1	######	33.5757	-118.26	5500	118	134	4401	B350	19/09/2015 13:21	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
19	A72AA1	######	33.5805	-118.26	5500	118	121	4401	B350	19/09/2015 13:20	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
20	A72AA1	######	33.5893	-118.3	5500	116	83	4401	B350	19/09/2015 13:20	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
21	A72AA1	######	33.5862	-118.34	5475	120	83	4401	B350	19/09/2015 13:19	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
22	A72AA1	######	33.5845	-118.36	5474	120	86	4401	B350	19/09/2015 13:19	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
23	A72AA1	######	33.5716	-118.47	5475	118	79	4401	B350	19/09/2015 13:17	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
24	A72AA1	######	33.5844	-118.62	5475	124	132	4401	B350	19/09/2015 13:15	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
25	A72AA1	######	33.6622	-118.74	5525	136	119	4401	B350	19/09/2015 13:13	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
26	A72AA1	######	33.6712	-118.76	6050	137	124	4401	B350	19/09/2015 13:12	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
27	A72AA1	######	33.6886	-118.79	6350	124	119	4401	B350	19/09/2015 13:12	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
28	A72AA1	######	33.7157	-118.85	6475	124	119	4401	B350	19/09/2015 13:11	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
29	A72AA1	######	33.7425	-118.91	6474	122	120	4401	B350	19/09/2015 13:10	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
30	A72AA1	######	33.7606	-118.94	6475	122	120	4401	B350	19/09/2015 13:09	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
31	A72AA1	######	33.8058	-119.04	6475	137	121	4401	B350	19/09/2015 13:07	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
32	A72AA1	######	33.8231	-119.08	6475	152	123	4401	B350	19/09/2015 13:07	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
33	A72AA1	######	33.8473	-119.12	6475	0	133	4401	B350	19/09/2015 13:06	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs
34	A72AA1	727763c	33.2971	-119.44	6475	103	216	4401	B350	19/09/2015 12:28	US DEPARTMENT OF HOME	US CUSTOMS & BORDE	OFFICE OF AIR & MARINE	561A	FM-36	4220012	HAWKER BEECHCRAFT CORP	B300C	2010	5	dhs

Figure 1- First few rows of Fed1.CSV dataset file

2 ER – Diagram for Data Set

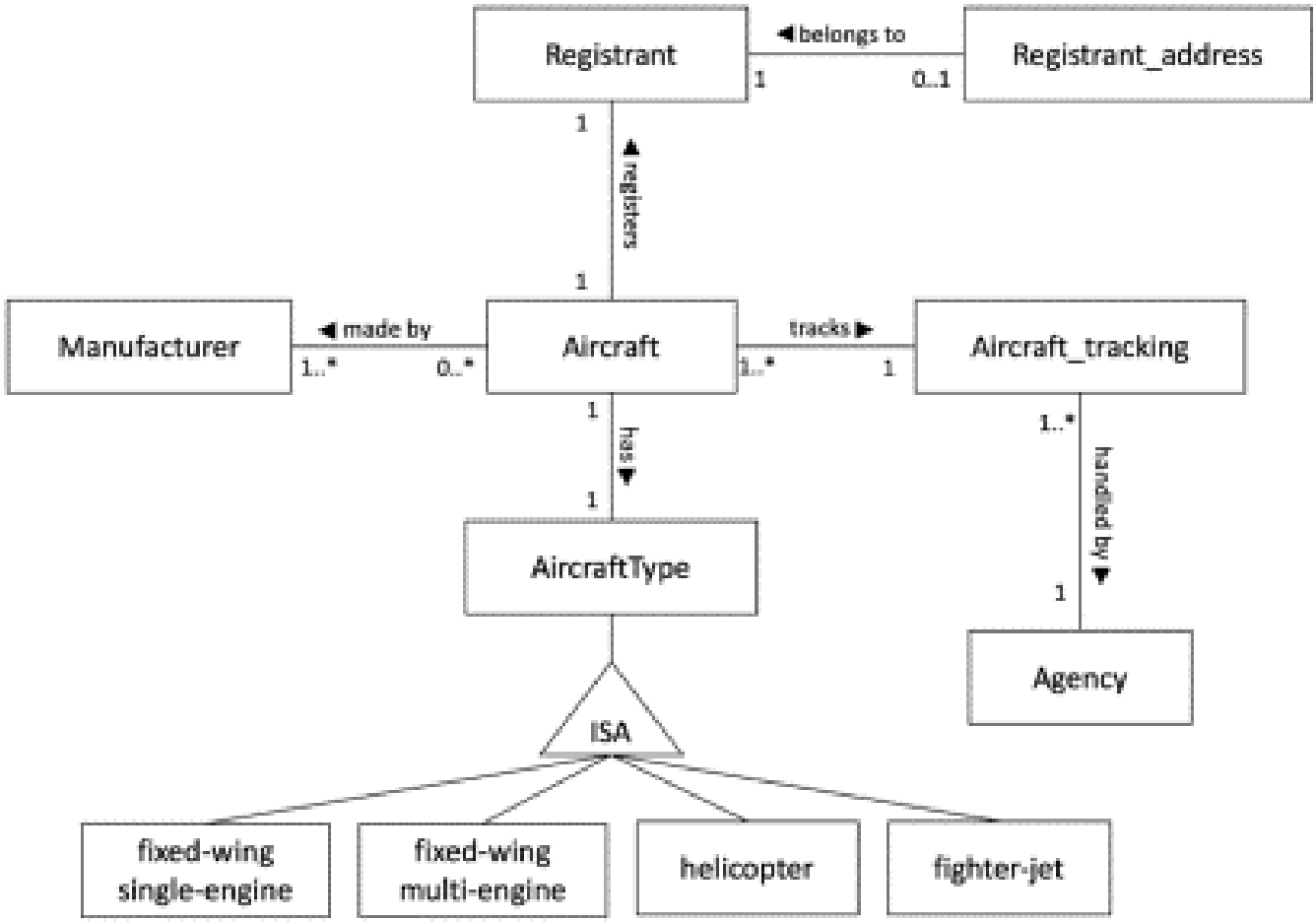


Figure 2- ER diagram

3 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.

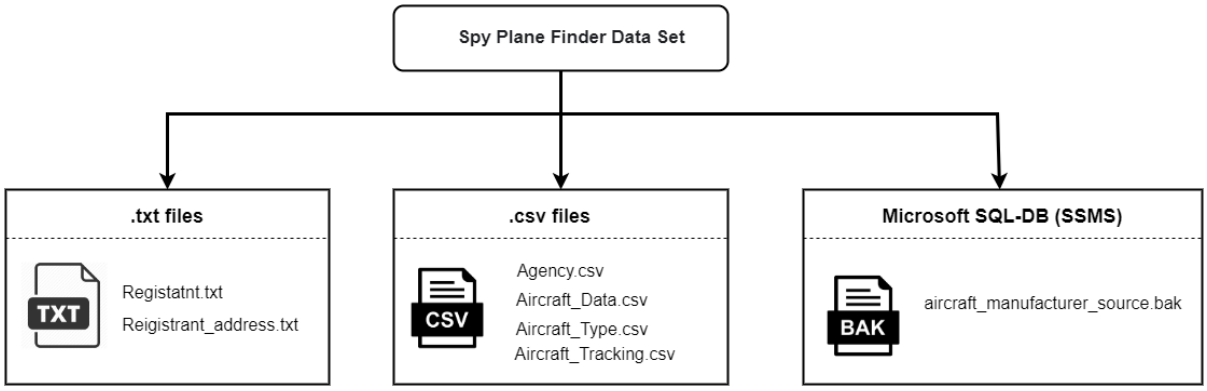
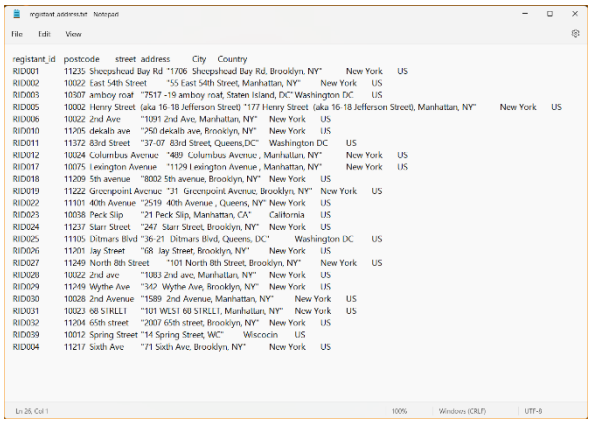


Figure 3 - Split Dataset

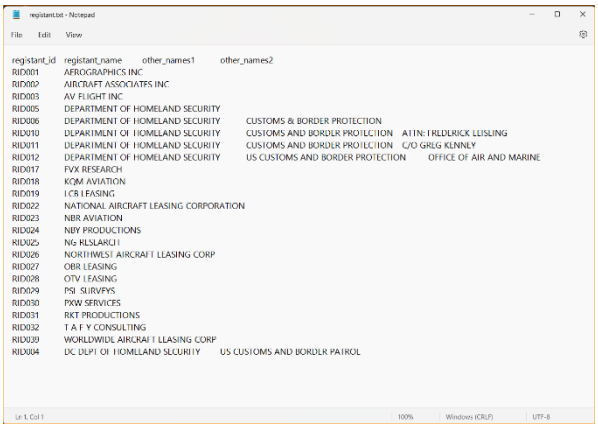
All the snapshots of this partitioned source files are attached below

3.1 .TXT files



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 road "7517 -19 amboy road, 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 "1091 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 "1179 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 11101 40th Avenue "2519 40th Avenue , Queens, NY" New York US
RID023 10038 Peck Slip "21 Peck Slip, Manhattan, CA" California US
RID024 11237 Starr Street "247 Starr Street, Brooklyn, NY" New York US
RID025 11105 Ditmars Blvd "36-21 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 STRLLT "101 WLST 68 STRLLT, 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" Wiscocin US
RID004 11217 Sixth Ave "71 Sixth Ave, Brooklyn, NY" New York US

Figure 4- Registant_Address .txt file



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 LUISLING
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 KQM AVIATION
RID019 LCB LEASING
RID022 NATIONAL AIRCRAFT LEASING CORPORATION
RID023 NBR AVIATION
RID024 NBY PRODUCTIONS
RID025 NG RLSLARCH
RID026 NORTHWEST AIRCRAFT LEASING CORP
RID027 OBR LEASING
RID028 OTV LEASING
RID029 PSI SURVEYS
RID030 PXW 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

Figure 5 - Registant.txt file

3.2 .CSV files

	A	B	C
1	agency_id	agency_code	agency_name
2	AGC001	dhs	Department of Homeland Security
3	AGC002	fbi	Federal Bureau of Investigation

agency

Ready

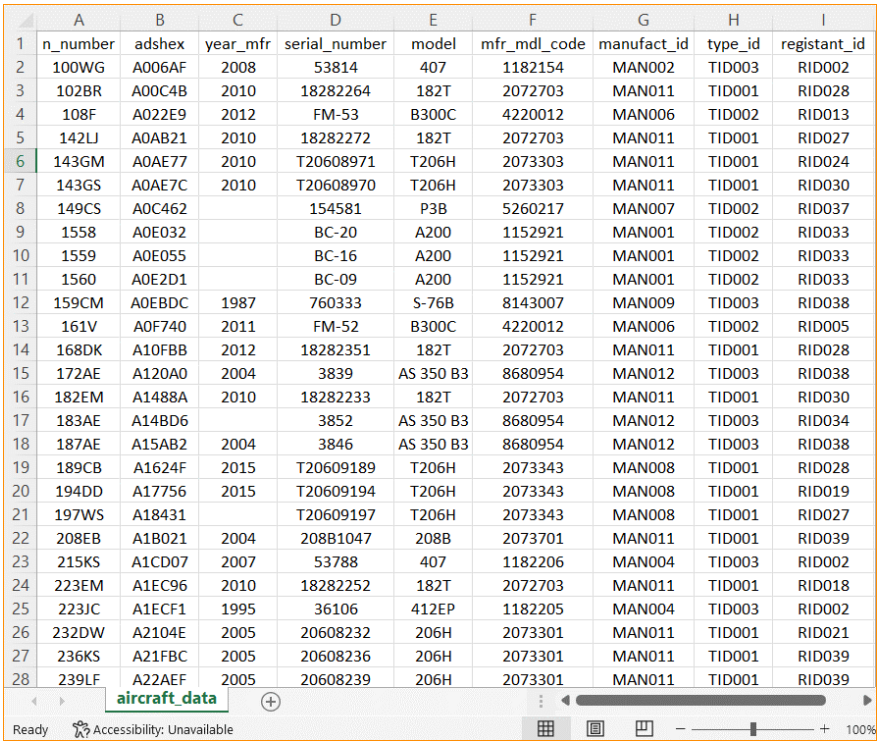
Figure 6 - Agency.csv file

	A	B
1	type_name	type_id
2	fixed-wing single-engine	TID001
3	fixed-wing multi-engine	TID002
4	helicopter.	TID003
5	fighterjer	TID004

aircraft_types

Ready Accessibility: Unavailable

Figure 7 - Aircraft_type.csv



	A	B	C	D	E	F	G	H	I
1	n_number	adshex	year_mfr	serial_number	model	mfr_mdl_code	manufact_id	type_id	registant_id
2	100WG	A006AF	2008	53814	407	1182154	MAN002	TID003	RID002
3	102BR	A00C4B	2010	18282264	182T	2072703	MAN011	TID001	RID028
4	108F	A022E9	2012	FM-53	B300C	4220012	MAN006	TID002	RID013
5	142LJ	A0AB21	2010	18282272	182T	2072703	MAN011	TID001	RID027
6	143GM	A0AE77	2010	T20608971	T206H	2073303	MAN011	TID001	RID024
7	143GS	A0AE7C	2010	T20608970	T206H	2073303	MAN011	TID001	RID030
8	149CS	A0C462		154581	P3B	5260217	MAN007	TID002	RID037
9	1558	A0E032		BC-20	A200	1152921	MAN001	TID002	RID033
10	1559	A0E055		BC-16	A200	1152921	MAN001	TID002	RID033
11	1560	A0E2D1		BC-09	A200	1152921	MAN001	TID002	RID033
12	159CM	A0EBDC	1987	760333	S-76B	8143007	MAN009	TID003	RID038
13	161V	A0F740	2011	FM-52	B300C	4220012	MAN006	TID002	RID005
14	168DK	A10FBB	2012	18282351	182T	2072703	MAN011	TID001	RID028
15	172AE	A120A0	2004	3839	AS 350 B3	8680954	MAN012	TID003	RID038
16	182EM	A1488A	2010	18282233	182T	2072703	MAN011	TID001	RID030
17	183AE	A14BD6		3852	AS 350 B3	8680954	MAN012	TID003	RID034
18	187AE	A15AB2	2004	3846	AS 350 B3	8680954	MAN012	TID003	RID038
19	189CB	A1624F	2015	T20609189	T206H	2073343	MAN008	TID001	RID028
20	194DD	A17756	2015	T20609194	T206H	2073343	MAN008	TID001	RID019
21	197WS	A18431		T20609197	T206H	2073343	MAN008	TID001	RID027
22	208EB	A1B021	2004	208B1047	208B	2073701	MAN011	TID001	RID039
23	215KS	A1CD07	2007	53788	407	1182206	MAN004	TID003	RID002
24	223EM	A1EC96	2010	18282252	182T	2072703	MAN011	TID001	RID018
25	223JC	A1ECF1	1995	36106	412EP	1182205	MAN004	TID003	RID002
26	232DW	A2104E	2005	20608232	206H	2073301	MAN011	TID001	RID021
27	236KS	A21FBC	2005	20608236	206H	2073301	MAN011	TID001	RID039
28	239LF	A22AEF	2005	20608239	206H	2073301	MAN011	TID001	RID039

aircraft_data

Ready Accessibility: Unavailable 100%

Figure 8 - Aircraft_data.csv file

A	B	C	D	E	F	G	H	I	J
adshex	flight_id	latitude	longitude	altitude	speed	track	squawk	timestamp	agency_id
A006AF	7f6369c	35.04171	-106.678	5700	75	275	0	11/04/2015 21:40	AGC002
A006AF	7e345bb	36.2514	-115.181	2700	56	18	1024	11/12/2015 22:00	AGC002
A006AF	7e92528	36.2414	-115.379	5981	0	278	4414	11/11/2015 07:45	AGC002
A006AF	7f22072	36.23011	-115.258	2900	0	287	1200	11/07/2015 07:38	AGC002
A006AF	7e3ef7b	35.1657	-106.78	5925	118	95	0	11/05/2015 03:25	AGC002
A006AF	7e3ef7b	35.17121	-106.813	5900	121	67	0	11/05/2015 03:25	AGC002
A006AF	7e345bb	35.0396	-106.654	5700	59	270	0	11/04/2015 21:39	AGC002
A00C4B	7428d8d	42.5227	-83.2468	4050	101	120	0	09/08/2015 22:05	AGC002
A00C4B	7428d8d	42.2064	-83.4369	5425	51	255	0	10/08/2015 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	-83.0741	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	-83.1034	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	-83.064	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

aircraft_tracking

Figure 9 - Aircraft_tracking.csv file

3.3 .bak file

	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
16	BEECH	MAN001
17	BELL HELICOPTER TEXTRON CANADA	MAN002
18	PILATUS	MAN003

Figure 10 -aircraft_manufacturer_source.bak file

To describe the sources, I have used Database Diagrams.

Source	Source type	Description
Agency.csv	csv	Includes the details of all agencies who owns tracked aircrafts
Aircraft_Data.csv	csv	Includes the details of aircrafts
Aircraft_Tracking.csv	csv	Include the tracking information of aircrafts owned by agencies
Aircraft_Types.csv	csv	Include the details of aircraft types
Registant .txt	txt	Include the details of Registered name
Registant_Address.txt	txt	Include the details of Address of Registered company
Aircraft_Manufacturer_ Source.bak	bak	Includes the details of Aircraft manufacturers

Table 1- Description of Source files

Detailed source file columns descriptions are given in the table below

3.3.1 Agency Table

Table	Column Name	Data Types	Description
Agency	Agency_id	nvarchar(6)	Unique agency id
	Agency_code	nvarchar(5)	Unique 3 letter code given for each agency
	Agency_name	nvarchar(50)	Name of the agency

Table 2 - Details of Agency Table

3.3.2 Aircraft_Data Table

Table	Column Name	Data Types	Description
Aircraft_Data	adshex	nvarchar(10)	Unique id given for each aircraft
	n_number	nvarchar(20)	Unique registration number given when the aircraft was registered
	year_mfr	int	Manufactured year
	serial_number	nvarchar(20)	Serial number of the air craft
	model	nvarchar(20)	Model of the aircraft
	mfr_mdl_code	int	Model code assigned by the manufactuer
	manufact_id	nvarchar(6)	ID of the manufacturer
	Type_id	nvarchar(6)	Corresponding ID of the type of the aircraft
	Registant_id	nvarchar(6)	ID of the registant who registered the aircraft

Table 3 - Details of Aircraft_Data Table

3.3.3 Aircraft_Type Table

Table	Column Name	Data Types	Description
Aircraft_Type	type_id	nvarchar(6)	Unique id of the aircraft type
	type_name	nvarchar(50)	Name of the aircraft type

Table 4 - Details of Aircraft Type table

3.3.4 Aircraft_manufacturer_source Table

Table	Column Name	Data Types	Description
Aircraft_ manufacturer_source	manufact _id	nvarchar(6)	Unique id of the manufacturer
	manufact_name	nvarchar(50)	Name of the manufacturer

Table 5 - Details of Aircraft manufacturer table

3.3.5 Registant Table

Table	Column Name	Data Types	Description
Registant	registant _id	nvarchar(6)	Unique id of the regitant
	registant _name	nvarchar(50)	Name of the registant
	other_names1	nvarchar(60)	Other names for the registant
	other_name2	nvarchar(60)	Other names for the registant

Table 6 - Details of Registant Table

3.3.6 Registant _Address Table

Table	Column Name	Data Types	Description
Registant_Address	registant _id	nvarchar(6)	Unique id of the regitant
	Postalcode	nvarchar(50)	Name of the registant
	city	nvarchar(60)	Other names for the registant
	address	nvarchar(60)	Other names for the registant
	country	nvarchar(10)	Country of the registant

Table 7 - Details of Registant Address Table

3.3.7 Aircraft_Tracking Table

Table	Column Name	Data Types	Description
Aircraft_Tracking	flight _id	nvarchar(6)	Flight id
	adshex	nvarchar(50)	Unique id of aircraft
	latitude	float	Geographic location in digital degrees
	longtitude	float	Geographic location in digital degrees
	altitude	int	Altitude in feet
	speed	int	Ground speed in knots
	track	int	Compass bearing in degrees,with 0 corresponding to north

	squawk	int	Four digit code transmitted bt the transponder
	timestamp	datetime	Extract the time when the aircraft was being tracked
	Agency_id	Nvarchar(6)	Federal agency operating the aircraft

Table 8 - Details of Aircraft Tracking Table

3.3.8 Names and definitions of columns

Names of columns are defined in the table given below.

Column name	More about description
adshex	Unique identifier for each aircraft, corresponding to its "Mode-S" code, in hexademical format.
flight_id	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 Fightradar24, surveillance planes were typically detected several times per minute.
latitude, longitude	Geographic location in digital degrees.
altitude	Altitude in feet.
speed	Ground speed in knots
track	Compass bearing in degrees, with 0 corresponding to north.
squawk	Four-digit code transmitted by the transponder
timestamp	Full UTC timestamp
name	Name of aircraft registrant.
other_names1, other_names2	Other names for the registrant, if listed.

n_number	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.
serial_number	Identifying number assigned to the aircraft by its manufacturer.
mfr_mdl_code	Code designating the manufacturer and model of the aircraft.
mfr	Manufacturer.
model	Aircraft model.
year_mfr	Year in which aircraft was manufactured.
type_aircraft	fixed-wing single-engine, fixed-wing multi-engine, helicopter, fighter-jet

Table 9 - Names of the columns and their definitions

4 Solution Architecture

4.1 Architectural diagram

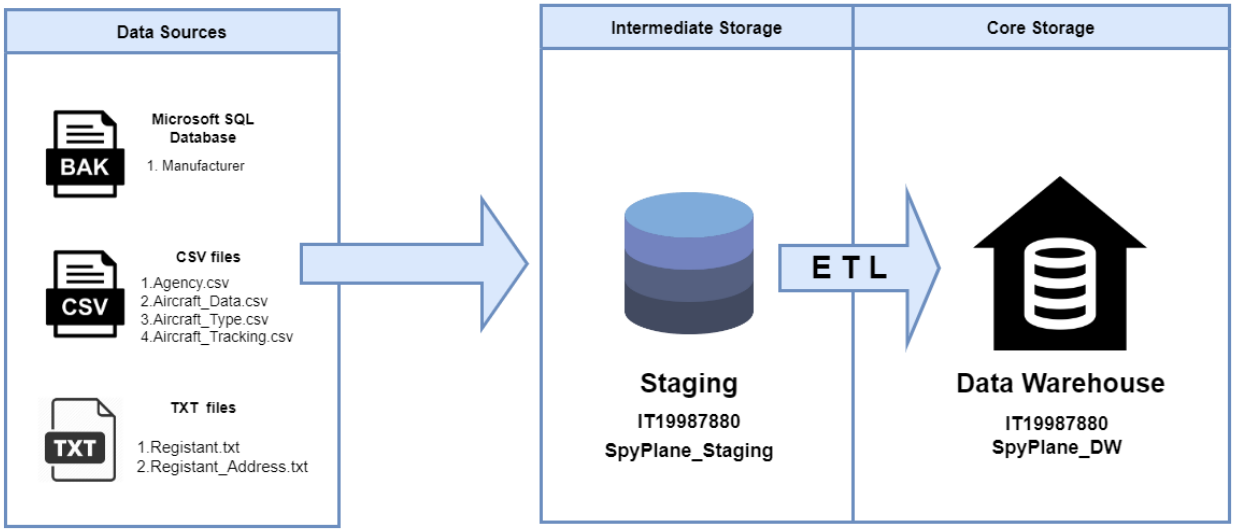


Figure 11 - Archetectural Diagram

Staging is the first step for the data warehousing .I created the Staging Database named as "**IT19987880 SpyPlane_Staging**" .Inside the Staging databasee, I manually created tables which are given below

- **Stg_Agency**
- **Stg_Aircraft_Data**
- **Stg_Aircraft_tracking**
- **Stg_Aircraft_Types**
- **Stg_Manufacturer**
- **Stg_Registant**
- **Stg_Registant_Manufacturer**

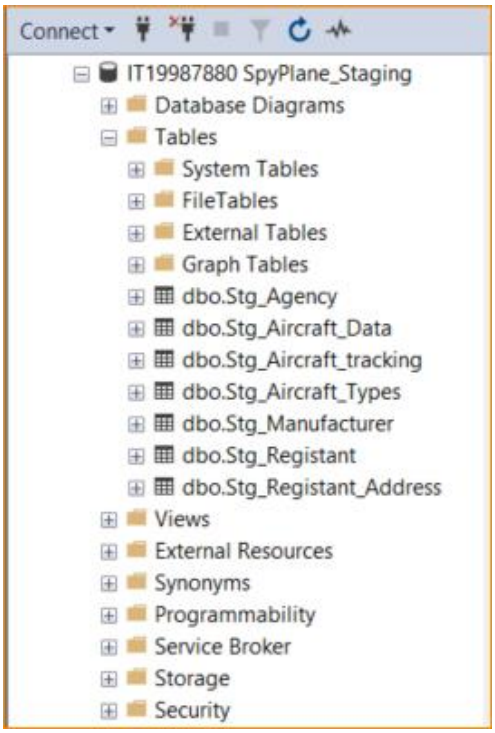


Figure 12 - Stagged tables

5 Data warehouse design & development

5.1 Dimension Table Design

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 **'IT19987880 SpyPlane_DW'**. Inside the data warehouse database, I manually created tables which are given below.

- **Dim_Agency**
- **Dim_Aircraft_Data**
- **Dim_Aircraft_Type**
- **Dim_Manufact**
- **Dim_Registrant**
- **Dim_Date**
- **Fact_Aircraft_Tracking**

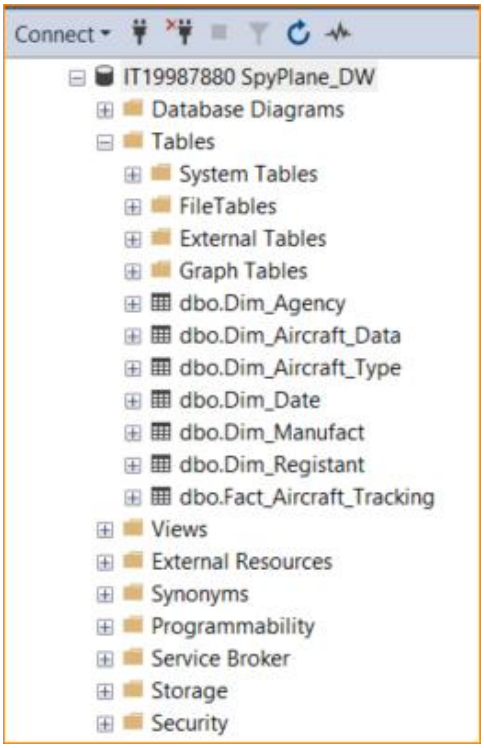


Figure 13 - created table list of Data warehouse

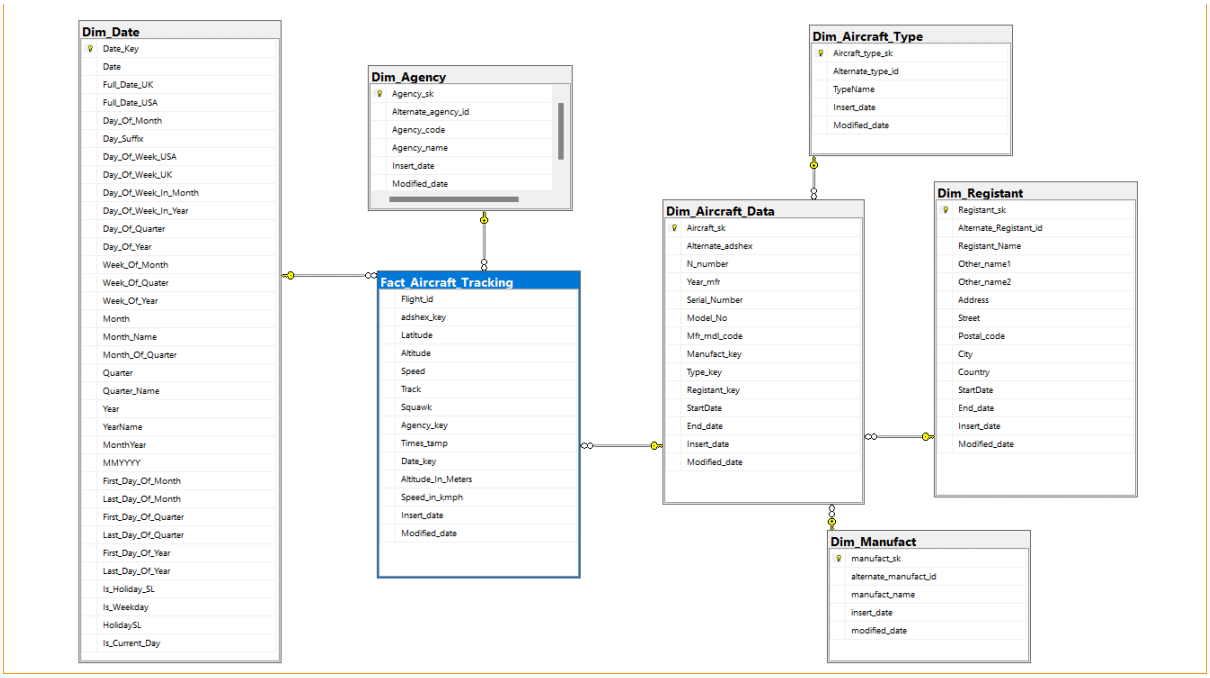


Figure 14 - Snowflake database diagram

5.1.1 Dim_Agency

DESKTOP-L0DEI0J.I...- dbo.Dim_Agency | SQLQuery3.sql - D...0DEI0J\pwmra (66))

Column Name	Data Type	Allow Nulls
Agency_sk	int	☐
Alternate_agency_id	nvarchar(50)	☑
Agency_code	varchar(50)	☑
Agency_name	varchar(50)	☑
Insert_date	datetime	☑
Modified_date	datetime	☑
		☐

Figure 15 - Dim_Agency table

5.1.2 Dim_Aircraft_Data

DESKTOP-L0DEI0J.I....Dim_Aircraft_Data | SQLQuery3.sql - D...0DEI0J\pwmra (66))

Column Name	Data Type	Allow Nulls
Aircraft_sk	int	☐
Alternate_adshex	nvarchar(10)	☑
N_number	nvarchar(20)	☑
Year_mfr	nvarchar(4)	☑
Serial_Number	nvarchar(20)	☑
Model_No	nvarchar(20)	☑
Mfr_mdl_code	int	☑
Manufact_key	int	☑
Type_key	int	☑
Registant_key	int	☑
StartDate	datetime	☑
End_date	datetime	☑
Insert_date	datetime	☑
Modified_date	datetime	☑

Figure 16 - Dim_Aircraft_Data table

5.1.3 Dim_Aircraft_Type

DESKTOP-L0DEI0J.IT....Dim_Aircraft_Type | SQLQuery3.sql - D...0DEI0J\pwmra (66))

Column Name	Data Type	Allow Nulls
Aircraft_type_sk	int	☐
Alternate_type_id	varchar(50)	☑
TypeName	varchar(50)	☑
Insert_date	datetime	☑
Modified_date	datetime	☑

Figure 17 - Dim_Aircraft_Type table

5.1.4 Dim_Date

DESKTOP-L0DEI0J....DW - dbo.Dim_Date | SQLQuery3.sql - D...0DEI0J\pwmra (66))

Column Name	Data Type	Allow Nulls
Date_Key	int	☐
Date	datetime	☑
Full_Date_UK	char(10)	☑
Full_Date_USA	char(10)	☑
Day_Of_Month	varchar(2)	☑
Day_Suffix	varchar(4)	☑
Day_Of_Week_USA	char(1)	☑
Day_Of_Week_UK	char(1)	☑
Day_Of_Week_In_Month	varchar(2)	☑
Day_Of_Week_In_Year	varchar(2)	☑
Day_Of_Quarter	varchar(3)	☑
Day_Of_Year	varchar(3)	☑
Week_Of_Month	varchar(1)	☑
Week_Of_Quater	varchar(2)	☑
Week_Of_Year	varchar(2)	☑
Month	varchar(2)	☑
Month_Name	varchar(9)	☑
Month_Of_Quarter	varchar(2)	☑
Quarter	char(1)	☑
Quarter_Name	varchar(9)	☑
Year	char(4)	☑
YearName	char(7)	☑
MonthYear	char(10)	☑
MMYYYY	char(6)	☑
First_Day_Of_Month	date	☑
Last_Day_Of_Month	date	☑

Figure 18 - Dime_Date table

5.1.5 Dim_Manufact

DESKTOP-L0DEI0J.I...dbo.Dim_Manufact | DESKTOP-L0DEI0J....DW - dbo.Dim_Date

Column Name	Data Type	Allow Nulls
manufact_sk	int	☐
alternate_manufact_id	nvarchar(8)	☑
manufact_name	nvarchar(50)	☑
insert_date	datetime	☑
modified_date	datetime	☑

Figure 19 - Dim_Manufact table

5.1.6 Dim_Registrant

DESKTOP-L0DEI0J.I...dbo.Dim_Registant | DESKTOP-L0DEI0J.I...dbo.Dim_Manufact

Column Name	Data Type	Allow Nulls
Registant_sk	int	☐
Alternate_Registant_id	nvarchar(6)	☑
Registant_Name	nvarchar(60)	☑
Other_name1	nvarchar(60)	☑
Other_name2	nvarchar(60)	☑
Address	nvarchar(100)	☑
Street	nvarchar(60)	☑
Postal_code	nvarchar(20)	☑
City	nvarchar(50)	☑
Country	nvarchar(50)	☑
StartDate	datetime	☑
End_date	datetime	☑
Insert_date	datetime	☑
Modified_date	datetime	☑

Figure 20 - Dim_Registant table

5.1.7 Fact_Aircraft_Tracking

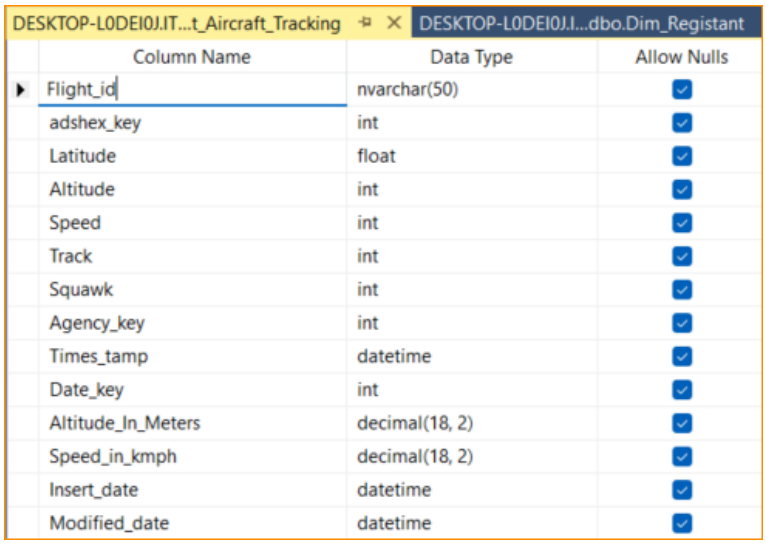


Figure 21 - Fact_Aircraft_Tracking

5.2 Calculation

When creating the fact table, I have added two extra attribute columns to the fact table which will be derived from the data which are already in the fact table .

Those two equations are shown below.

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

Those data types are derived attributes as well.

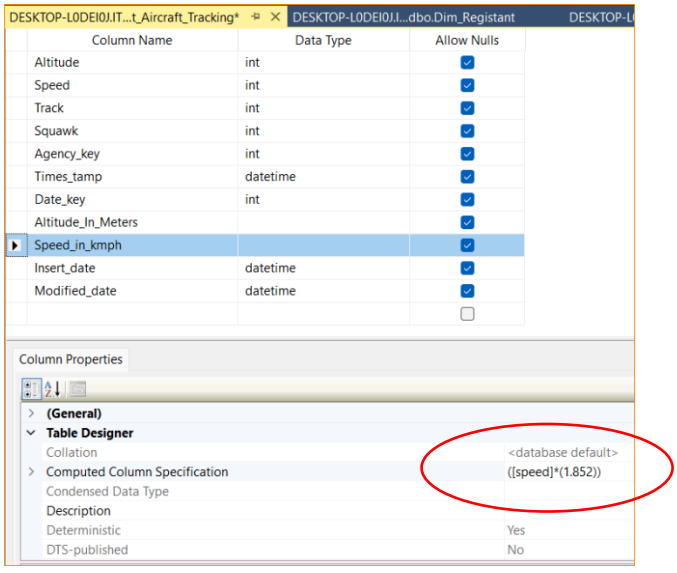


Figure 23 - calculation for speed in kmph

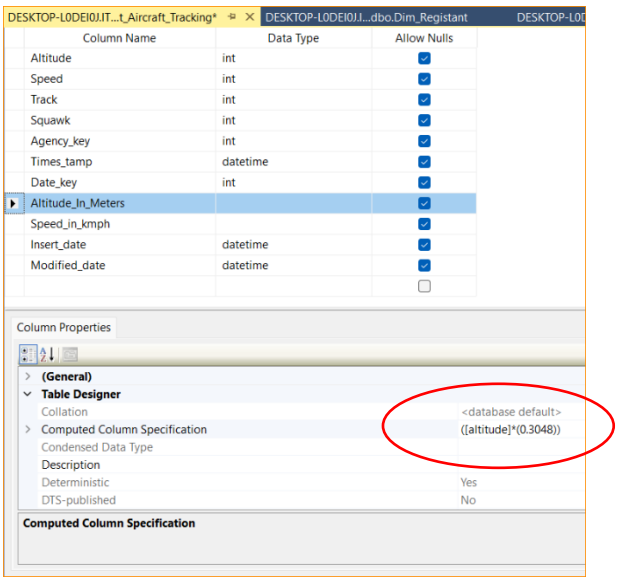


Figure 22 - Calculation for altitude in meters

5.3 Assumptions:

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

Dim_Date table is a static table. in that case I named the data inside the **Dim_Date** table as static data. All the data in the **Dim_Date** 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.

5.4 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

5.5 Static Table

Here in my scenario **Dim_Date** 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 **Dim_Date** table used in my DWBI lab sessions.

5.6 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

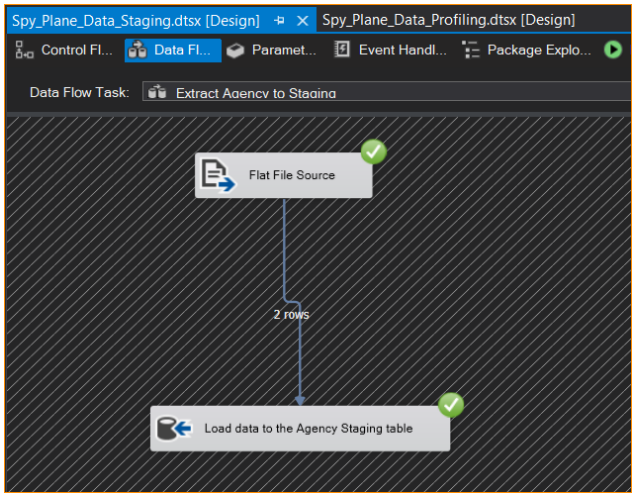
6 ETL Development

Snap shots of ETL process are included below.

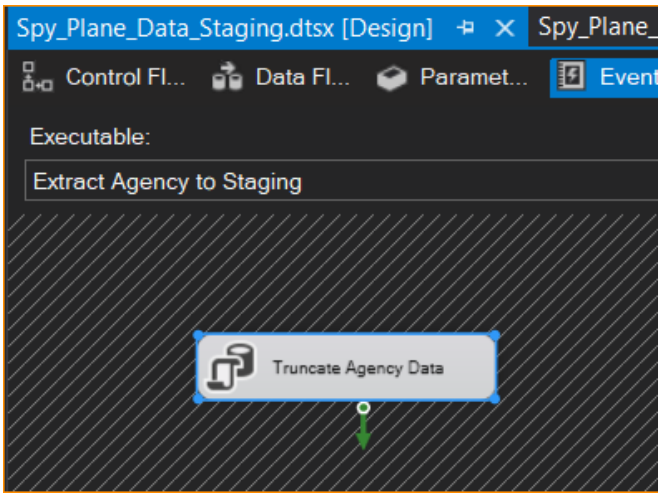
6.1 Staging Steps

6.1.1 Extract Agency data to staging

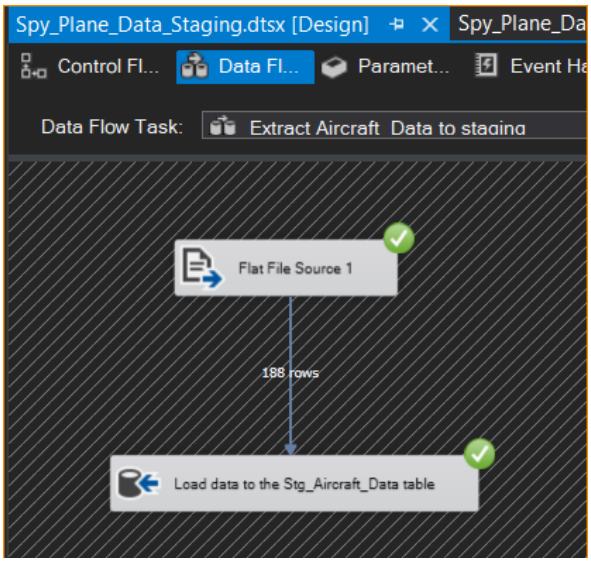
Dataflow



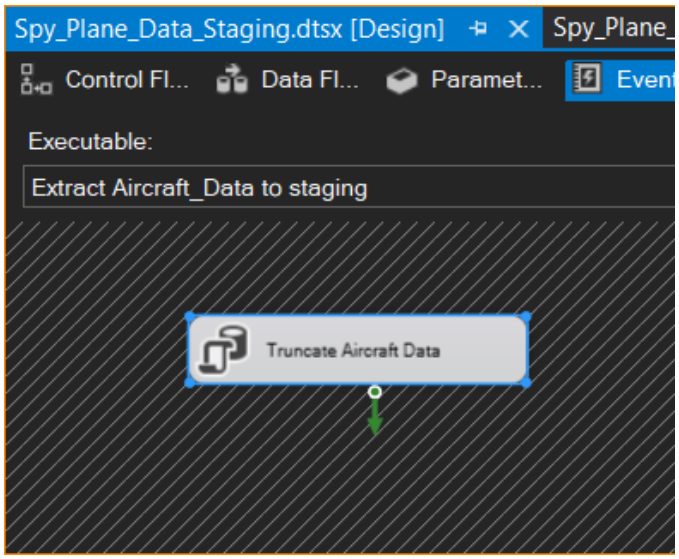
truncate



6.1.2 Extract Aircraft_Data to staging

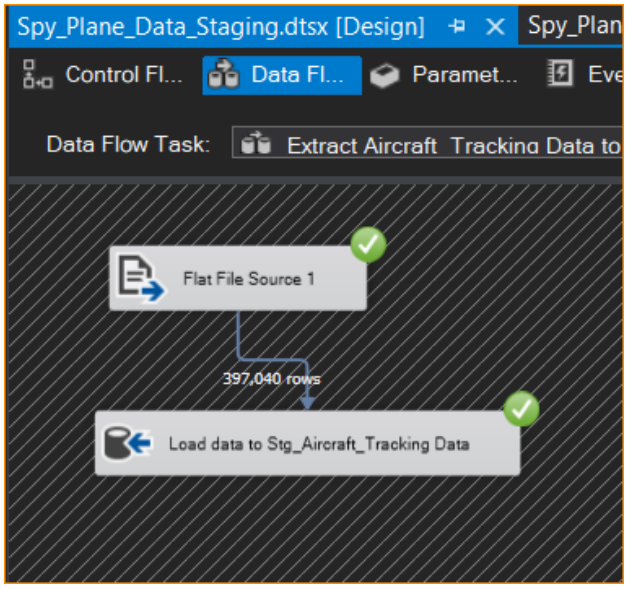


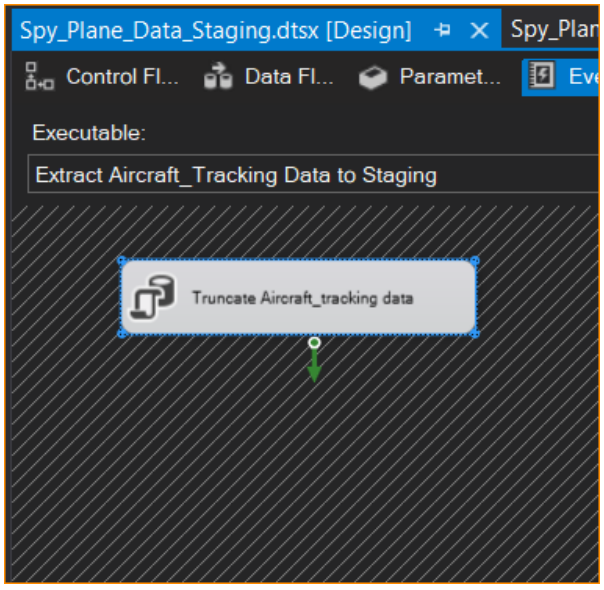
Dataflow



Truncate

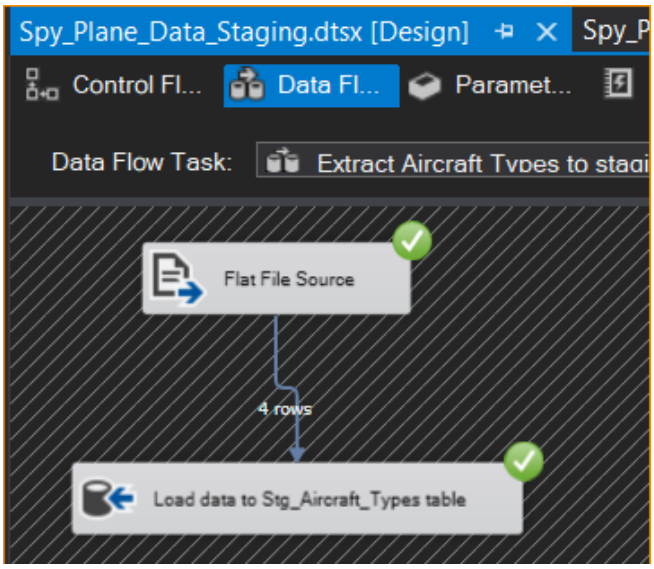
6.1.3 Extract Aircraft_Tracking to staging

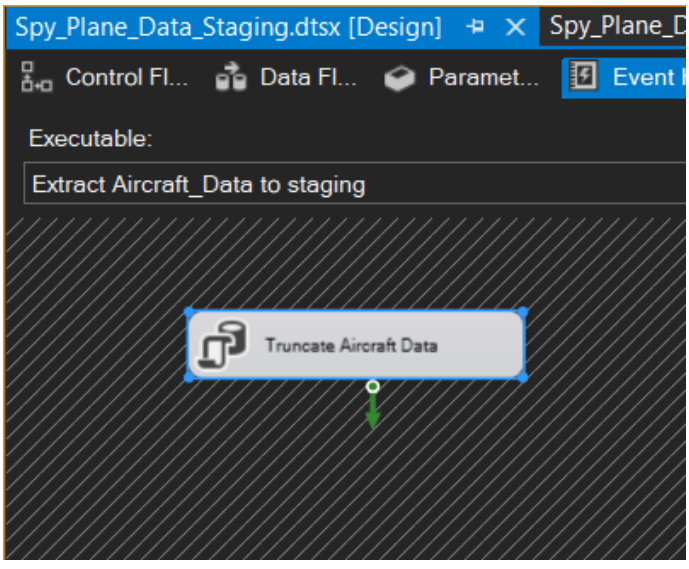




6.1.4 Extract Aircraft_Types to staging

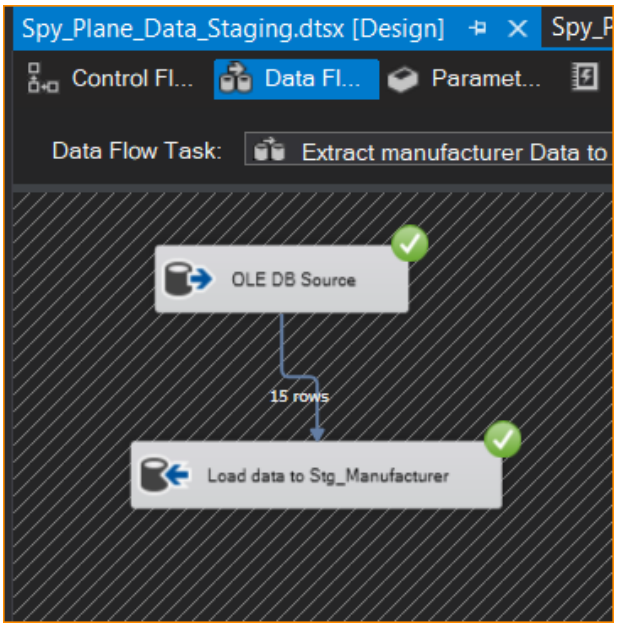
w Truncate



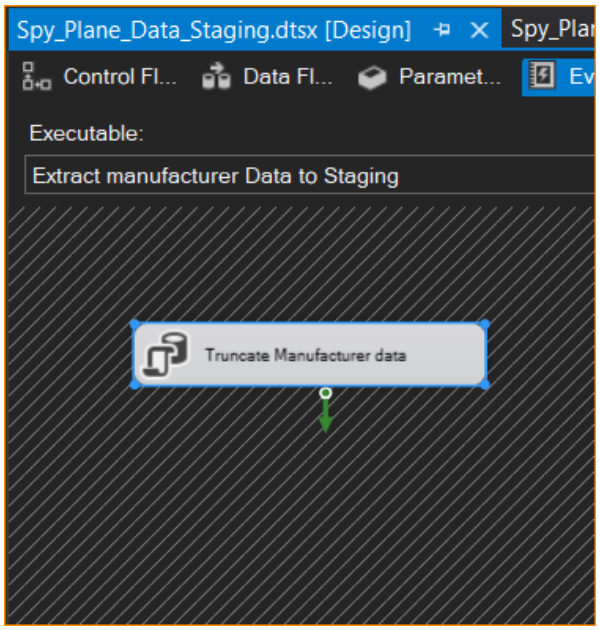


6.1.4 Extract Manufacturer data to staging

Dataflow

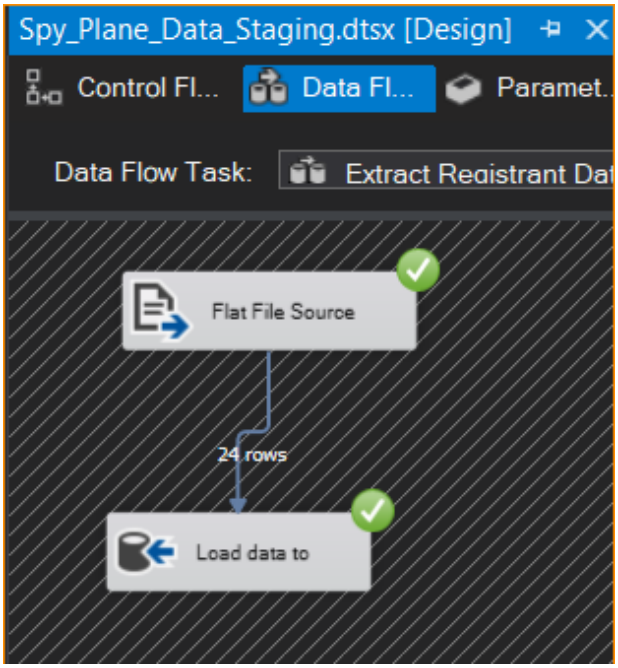


Truncate

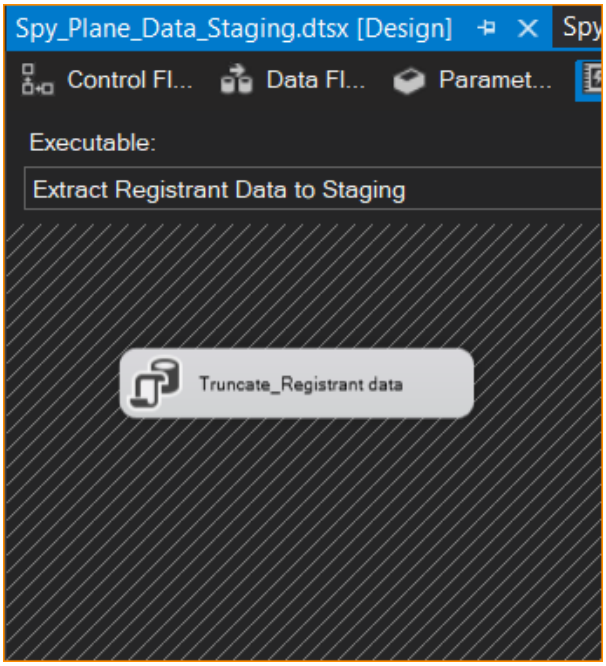


6.1.5 Extract Registant data to staging

Dataflow

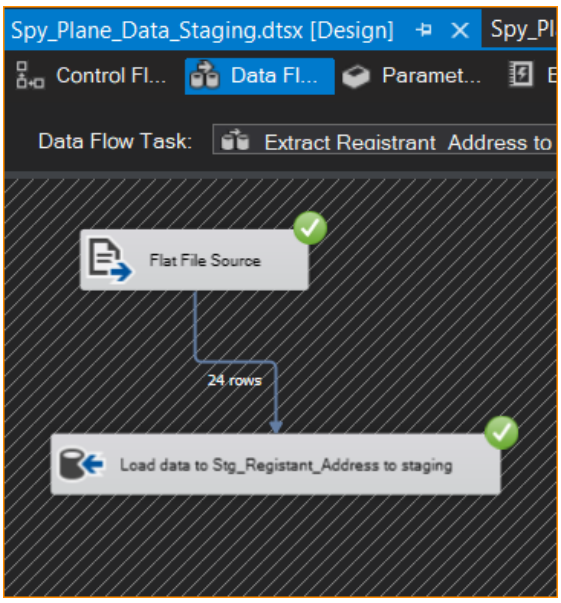


Truncate

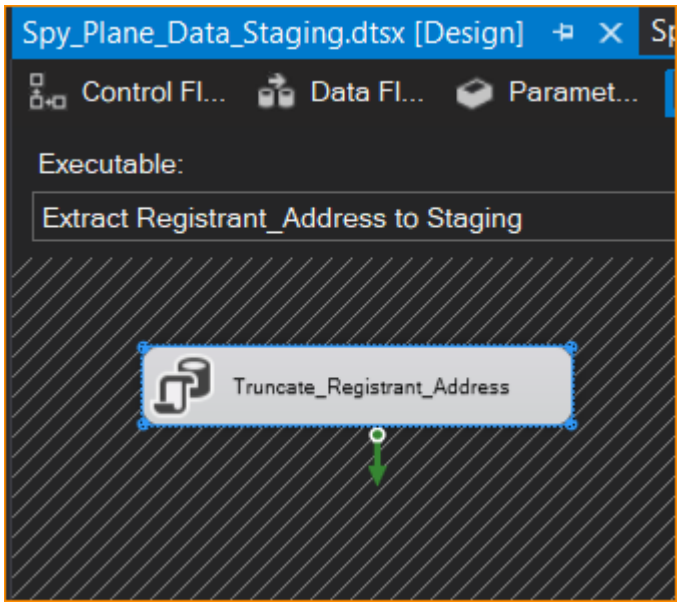


6.1.6 Extract Registant_Address data to staging

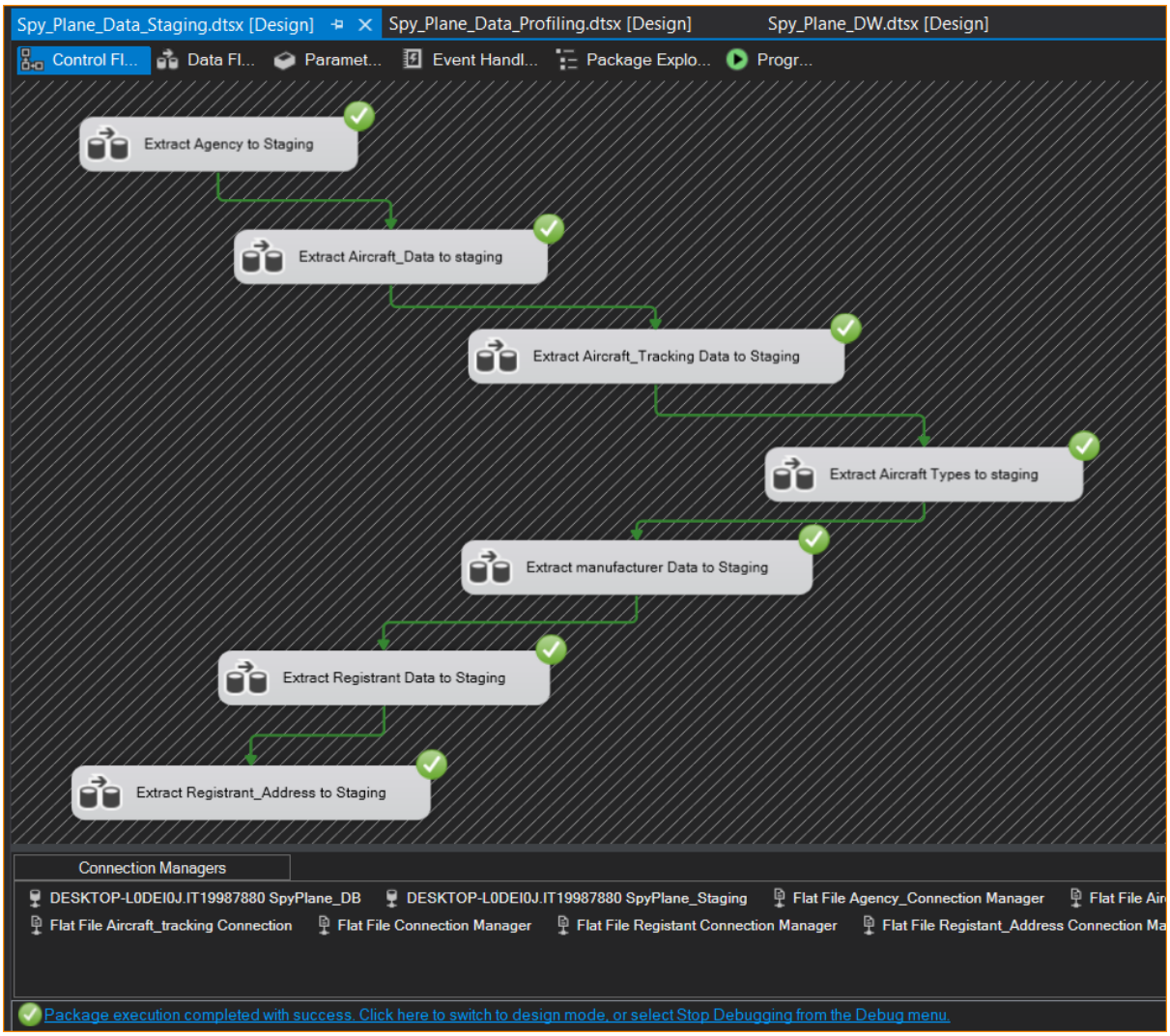
Dataflow



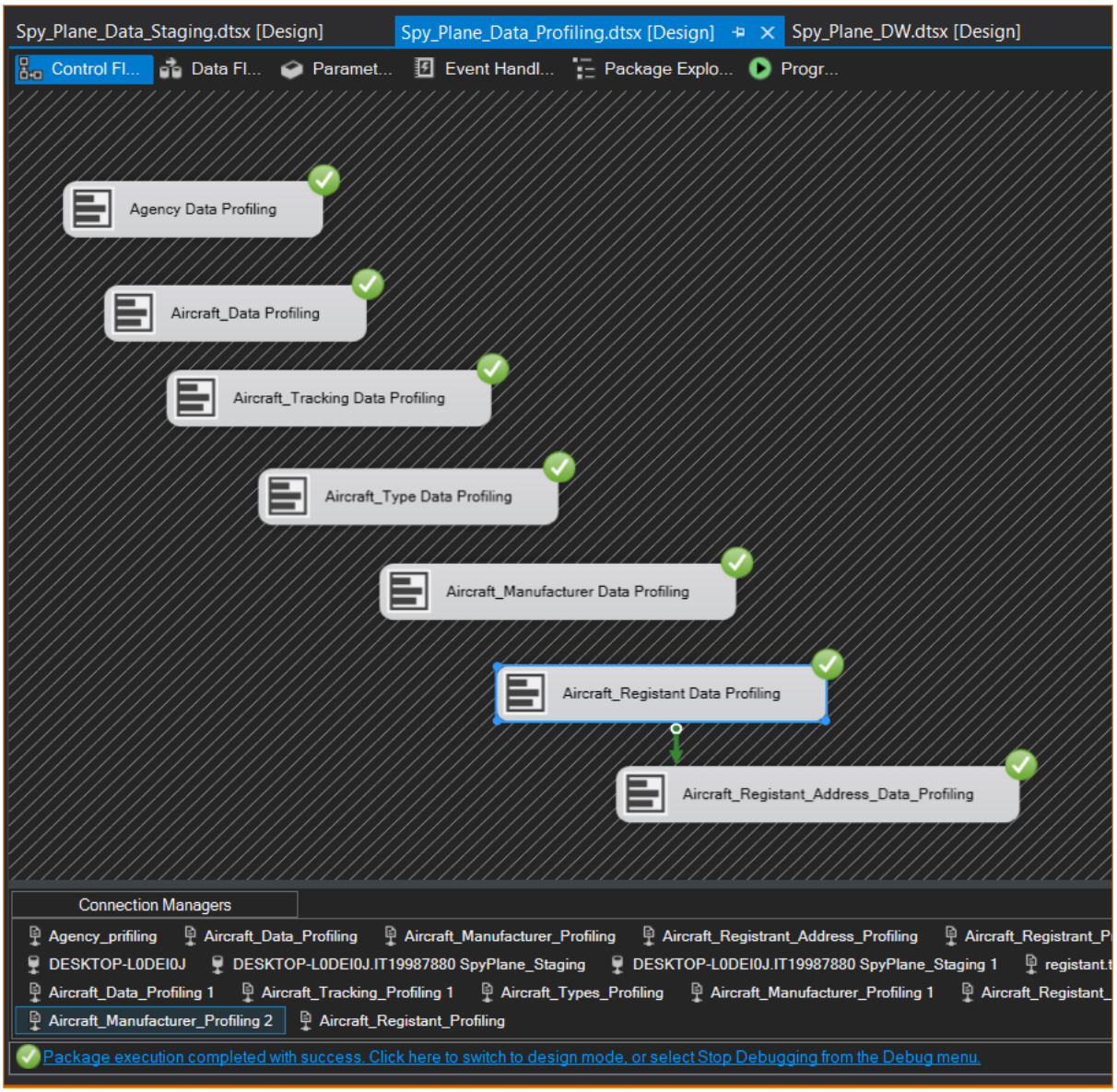
Truncate



6.1.7 Overall Staging Diagram



6.2 Data Profiling



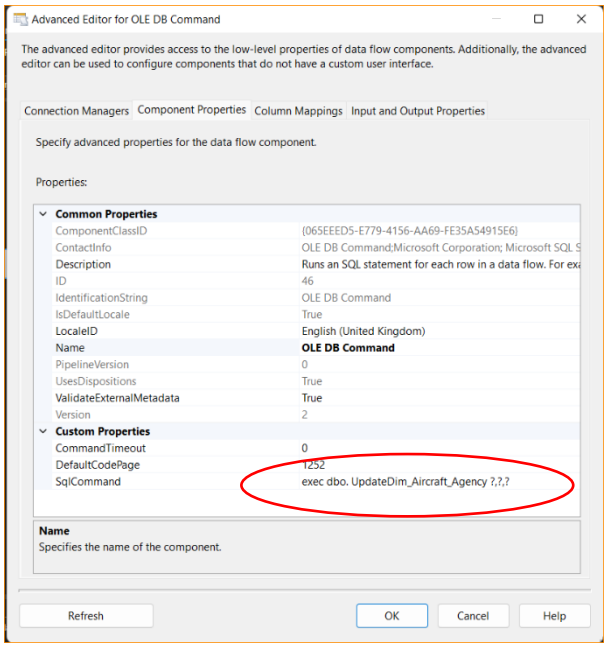
6.3 Data Transformation

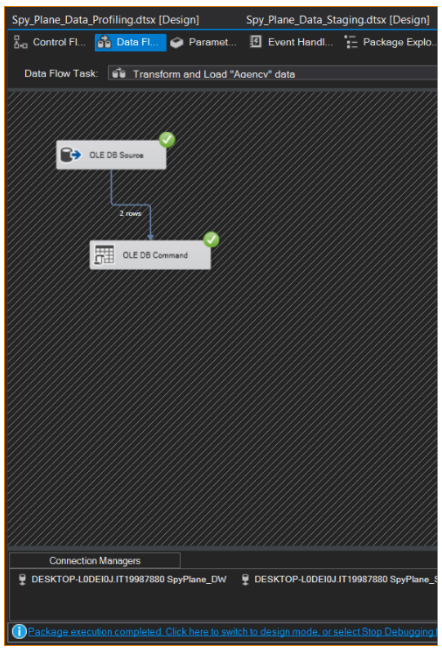
6.3.1 Transform and Load “Agency” details

First I have created a procedure called “**UpdateDim_Aircraft_Agency**” and executed in the **IT19987880 SpyPlane_DW** Database. I have mentioned the code below

```
CREATE PROCEDURE [dbo].[UpdateDim_Aircraft_Agency]
@agency_id nvarchar(50),
@agency_code nvarchar(6),
@agency_name nvarchar(50)

AS
BEGIN
if not exists (
select Agency_sk
from dbo.Dim_Agency
where Alternate_agency_id = @agency_id
)

BEGIN
insert into dbo.Dim_Agency(Alternate_agency_id ,Agency_code ,Agency_name,
Insert_date, Modified_date)
values( @agency_id, @agency_code,@agency_name, GETDATE(), GETDATE())
END;

if exists (
select Agency_sk
from dbo.Dim_Agency
where Alternate_agency_id = @agency_id
)

BEGIN
update dbo.Dim_Agency
set
Agency_code = @agency_code,
Agency_name = @agency_name,
Modified_date = GETDATE()

where Alternate_agency_id = @agency_id
END;
END;
```

OLE DB Command SSIS tool used to execute, **UpdateDim_Aircraft_Agency** procedure, it is used to insert data into position staging to **Dim_Agency** without data duplication



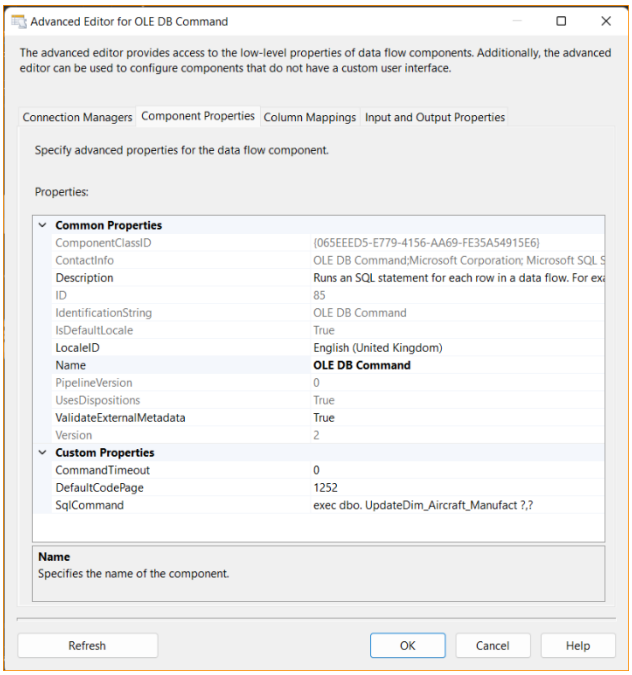


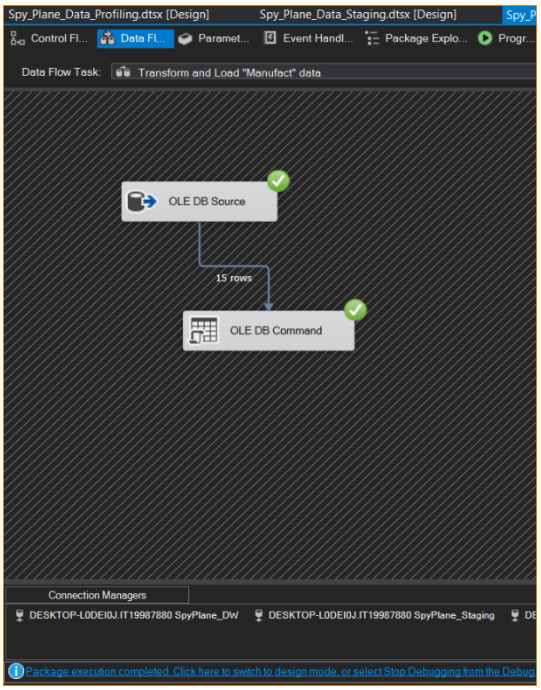
6.3.2 Transform and Load "Manufacturer" details

First I have created a procedure called "**UpdateDim_Aircraft_Manufact**" and executed in the **IT19987880 SpyPlane_DW** Database. I have mentioned the code below

```
CREATE PROCEDURE [dbo].[UpdateDim_Aircraft_Manufact]
@manufact_name nvarchar(50),
@manufact_id nvarchar(6)

AS
BEGIN
if not exists (
select manufact_sk
from dbo.Dim_Manufact
where alternate_manufact_id = @manufact_id
)

BEGIN
insert into dbo.Dim_Manufact(alternate_manufact_id ,manufact_name, insert_date,
modified_date)
values( @manufact_id, @manufact_name, GETDATE(), GETDATE())
END;

if exists (
select manufact_sk
from dbo.Dim_Manufact
where alternate_manufact_id = @manufact_id
)

BEGIN
update dbo.Dim_Manufact
set manufact_name = @manufact_name,
modified_date = GETDATE()
where alternate_manufact_id = @manufact_id
END;
END;
```

OLE DB Command SSIS tool used to execute, **UpdateDim_Aircraft_Manufact** procedure, it is used to insert data into position staging to **Dim_Manufact** without data duplication





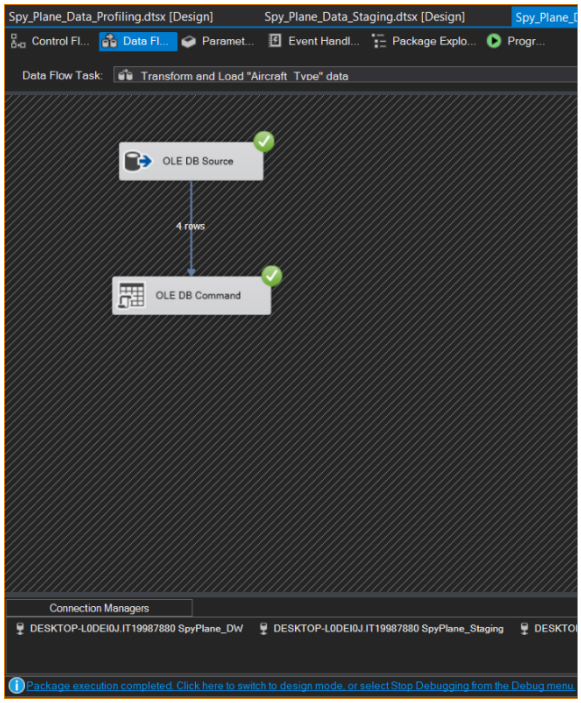
6.3.3 Transform and Load "Aircraft_Type" details

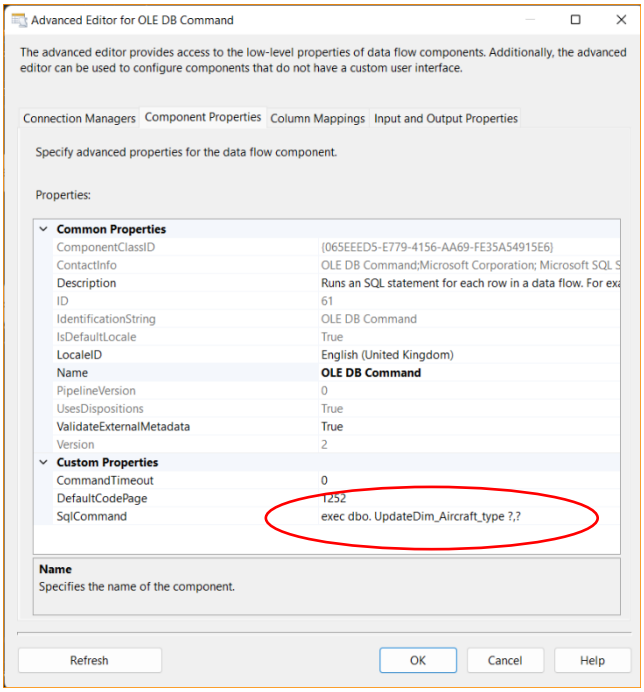
First I have created a procedure called "**UpdateDim_Aircraft_type**" and executed in the **IT19987880 SpyPlane_DW** Database. I have mentioned the code below

```
CREATE PROCEDURE [dbo].[UpdateDim_Aircraft_type]
@type_name varchar(50),
@type_id varchar(50)

AS
BEGIN
if not exists (
select Aircraft_type_sk
from dbo.Dim_Aircraft_Type
where Alternate_type_id = @type_id
)

BEGIN
insert into dbo.Dim_Aircraft_Type(Alternate_type_id , TypeName,
Insert_date,Modified_date)
values(@type_id, @type_name, GETDATE(), GETDATE())
END;

if exists (
select Aircraft_type_sk
from dbo.Dim_Aircraft_Type
where Alternate_type_id = @type_id
)
```

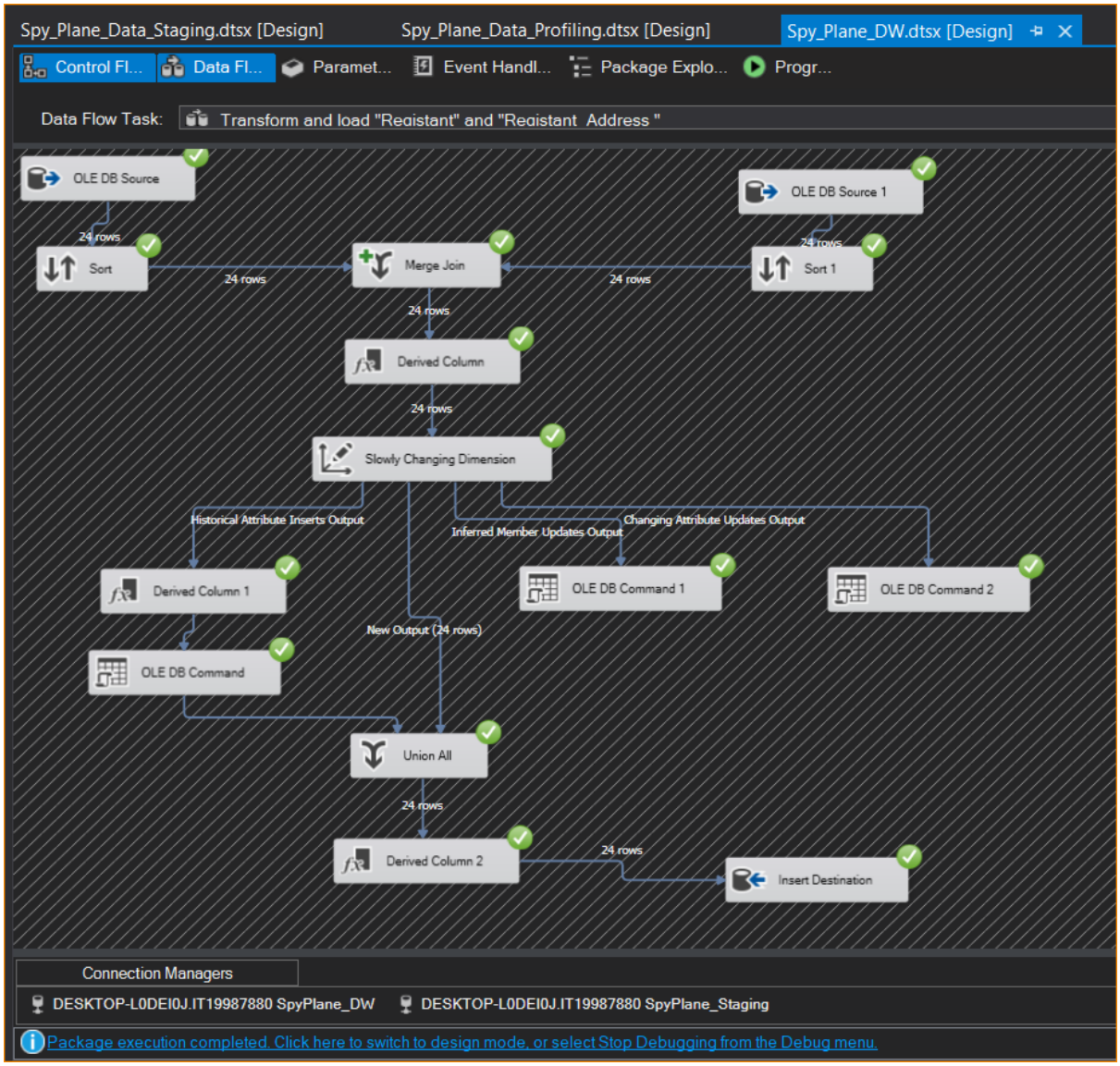
```
BEGIN
update dbo.Dim_Aircraft_Type
set TypeName = @type_name,
Modified_date = GETDATE()
where Alternate_type_id = @type_id
END;
END;
```

OLE DB Command SSIS tool used to execute, **UpdateDim_Aircraft_type** procedure, it is used to insert data into position staging to **Dim_Aircraft_Type** without data duplication

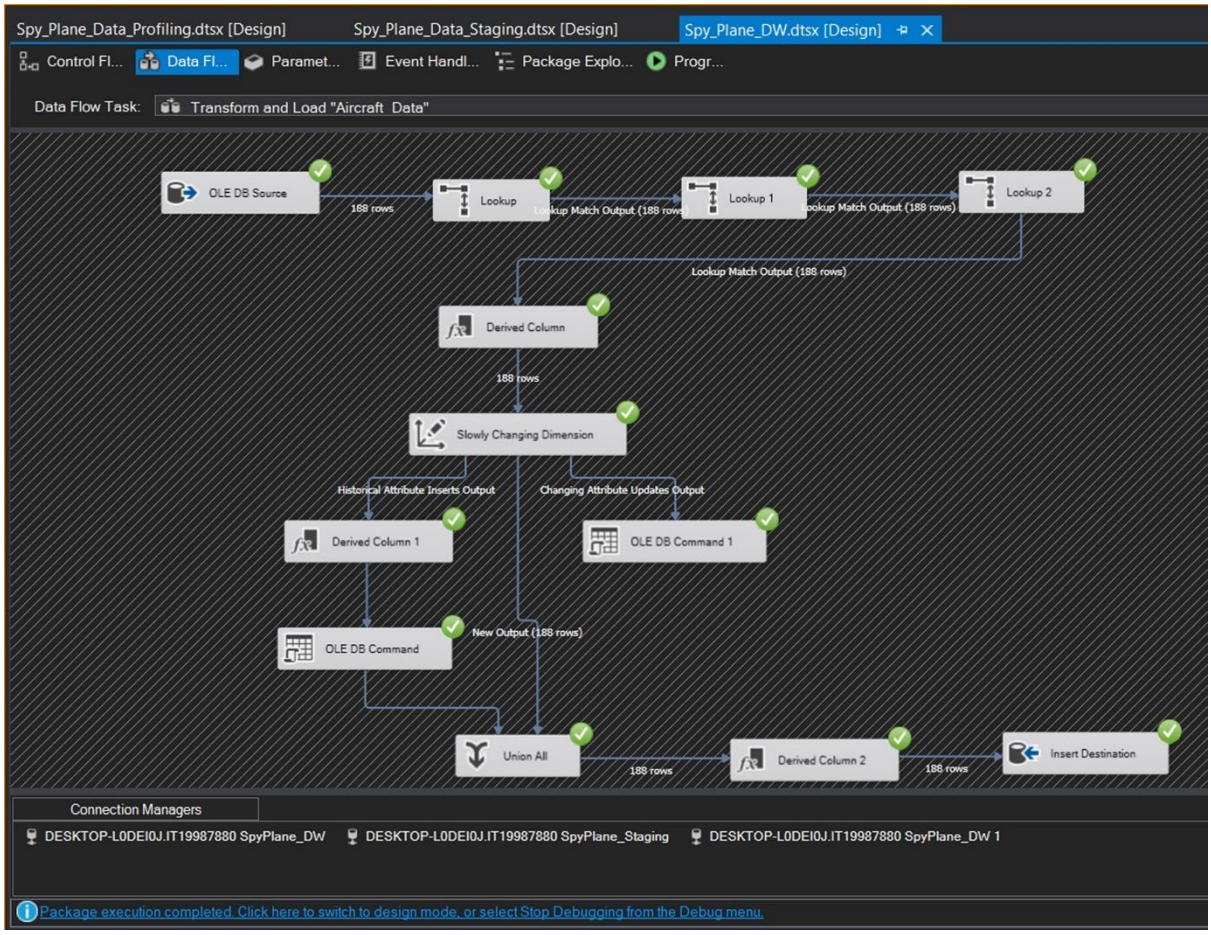




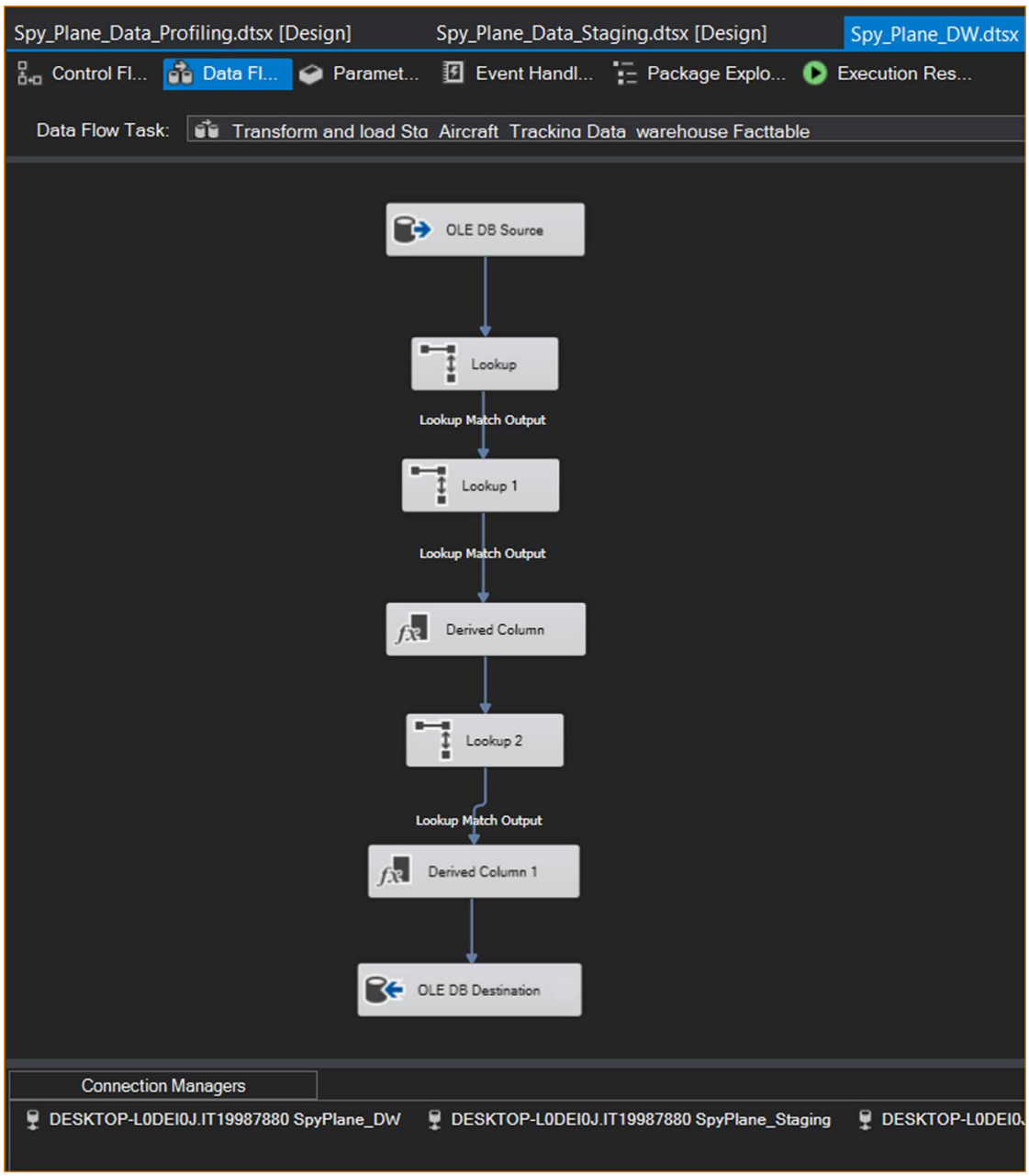
6.3.4 Transform and Load “Fact_Regitant” and “Registant_Address” details



6.3.5 Transform and Load “Aircraft Data” details



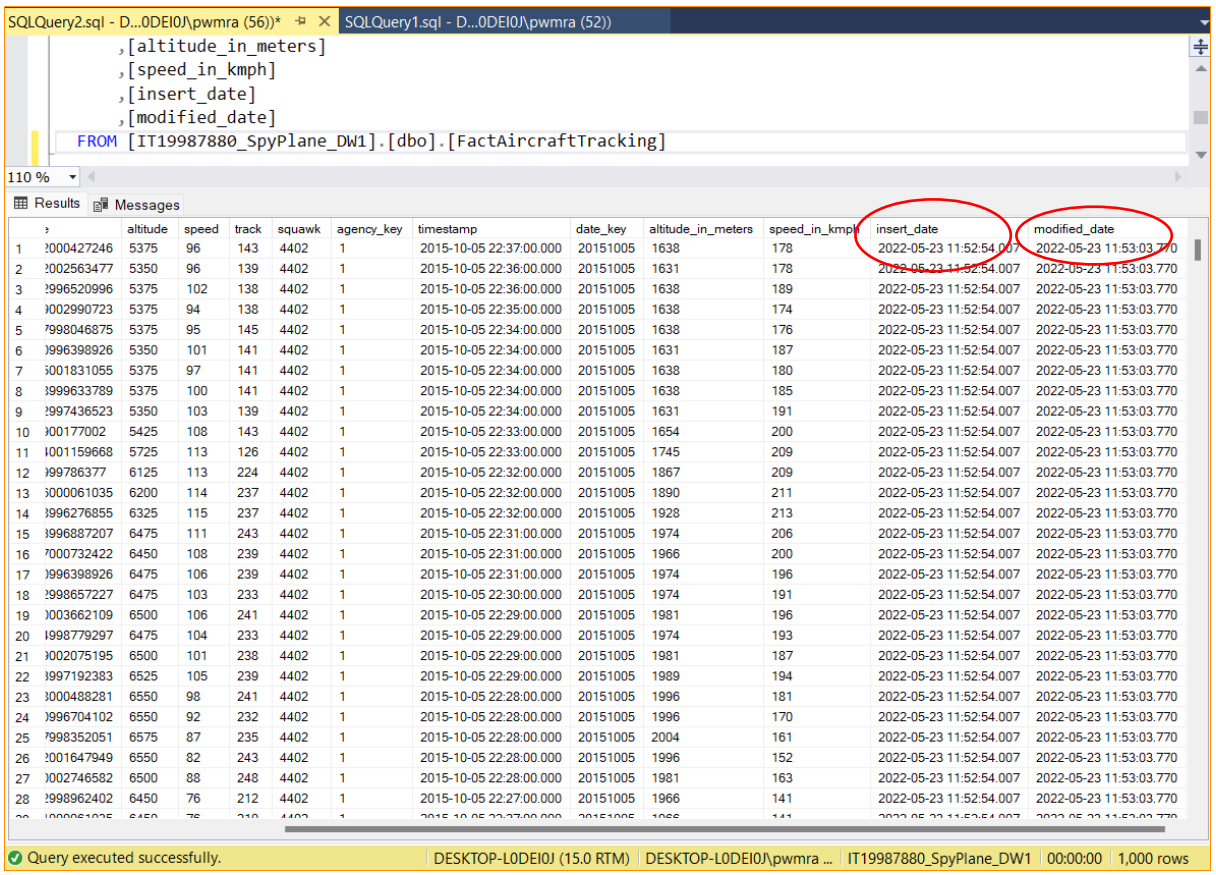
6.3.6 Transform and Load “Fact_Aircraft_Tracking” details



6.4 Overall Diagram of Data Warehouse

Final IT**19987880 SpyPlane_DW** Control flow

6.5 Accumulating Fact Table



	altitude	speed	track	squawk	agency_key	timestamp	date_key	altitude_in_meters	speed_in_kmph	insert_date	modified_date
1	5375	96	143	4402	1	2015-10-05 22:37:00.000	20151005	1638	178	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
2	5350	96	139	4402	1	2015-10-05 22:36:00.000	20151005	1631	178	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
3	5375	102	138	4402	1	2015-10-05 22:36:00.000	20151005	1638	189	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
4	5375	94	138	4402	1	2015-10-05 22:35:00.000	20151005	1638	174	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
5	5375	95	145	4402	1	2015-10-05 22:34:00.000	20151005	1638	176	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
6	5350	101	141	4402	1	2015-10-05 22:34:00.000	20151005	1631	187	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
7	5375	97	141	4402	1	2015-10-05 22:34:00.000	20151005	1638	180	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
8	5375	100	141	4402	1	2015-10-05 22:34:00.000	20151005	1638	185	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
9	5350	103	139	4402	1	2015-10-05 22:34:00.000	20151005	1631	191	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
10	5425	108	143	4402	1	2015-10-05 22:33:00.000	20151005	1654	200	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
11	5725	113	126	4402	1	2015-10-05 22:33:00.000	20151005	1745	209	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
12	6125	113	224	4402	1	2015-10-05 22:32:00.000	20151005	1867	209	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
13	6200	114	237	4402	1	2015-10-05 22:32:00.000	20151005	1890	211	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
14	6325	115	237	4402	1	2015-10-05 22:32:00.000	20151005	1928	213	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
15	6475	111	243	4402	1	2015-10-05 22:31:00.000	20151005	1974	206	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
16	6450	108	239	4402	1	2015-10-05 22:31:00.000	20151005	1966	200	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
17	6475	106	239	4402	1	2015-10-05 22:31:00.000	20151005	1974	196	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
18	6475	103	233	4402	1	2015-10-05 22:30:00.000	20151005	1974	191	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
19	6500	106	241	4402	1	2015-10-05 22:29:00.000	20151005	1981	196	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
20	6475	104	233	4402	1	2015-10-05 22:29:00.000	20151005	1974	193	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
21	6500	101	238	4402	1	2015-10-05 22:29:00.000	20151005	1981	187	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
22	6525	105	239	4402	1	2015-10-05 22:29:00.000	20151005	1989	194	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
23	6550	98	241	4402	1	2015-10-05 22:28:00.000	20151005	1996	181	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
24	6550	92	232	4402	1	2015-10-05 22:28:00.000	20151005	1996	170	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
25	6575	87	235	4402	1	2015-10-05 22:28:00.000	20151005	2004	161	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
26	6550	82	243	4402	1	2015-10-05 22:28:00.000	20151005	1996	152	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
27	6500	88	248	4402	1	2015-10-05 22:28:00.000	20151005	1981	163	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770
28	6450	76	212	4402	1	2015-10-05 22:27:00.000	20151005	1966	141	2022-05-23 11:52:54.007	2022-05-23 11:53:03.770