~~-------------------- dane~~

1. ~~Ze strony~~ [~~http://stat-computing.org/dataexpo/2009/the-data.html~~](http://stat-computing.org/dataexpo/2009/the-data.html) ~~pobrać:~~
   1. ~~Pliki z danymi – przynajmniej 2 lata (uwaga, mają po 500MB)~~
2. ~~Ze strony~~ [~~http://stat-computing.org/dataexpo/2009/supplemental-data.html~~](http://stat-computing.org/dataexpo/2009/supplemental-data.html) ~~pobrać:~~
   1. ~~Airports~~
   2. ~~Carriers~~
   3. ~~Plane-data~~
3. ~~Zaprojektować Hurtownię Danych w schemacie gwiazdy:~~
   1. ~~FactFlights na podstawie danych z pkt. 1~~
   2. ~~DimAirport, DimCarrier, DimPlane na podstawie danych z pkt. 2~~
   3. ~~DimDate – wygenerować (np. pętla WHILE,~~ [~~https://www.codeproject.com/Articles/647950/Create-and-Populate-Date-Dimension-for-Data-Wareho~~](https://www.codeproject.com/Articles/647950/Create-and-Populate-Date-Dimension-for-Data-Wareho)~~)~~

~~-------------------- projekt~~

1. ~~Utworzyć solucję w VS zawierającą 2 projekty:~~
   1. ~~Integration Services Project~~
   2. ~~SQL Server Database Project~~
2. ~~Utworzyć Hurtownię Danych – kod zgrać do projektu SQL SD~~
3. Utworzyć proces ETL zasilający HD z plików ściągniętych w pkt. 1, 2

~~-------------------- wymiary~~

1. ~~Przeładowanie pełne (TRUNCATE TABLE)~~
2. ~~Nadać klucze surogatowe (IDENTITY), założyć na nich PK~~
3. ~~Na kluczach biznesowych założyć UNIQUE~~
4. ~~Osobny Connection Manager dla każdego pliku CSV używanego w wymiarach~~
   1. ~~typy danych ustawione w CM~~
   2. ~~nazwy kolumn ustawione w CM (znaleźć na stronie)~~
5. ~~Osobna paczka dla każdego wymiaru~~
6. ~~Dodać procedurę w bazie, która dodaje do wymiaru element nieznany (Id = -1), osobna procedura dla każdego z wymiarów. Procedurę wywoływać w paczce za pomocą Execute SQL Task~~

-------------------- fakty

1. Utworzyć jedną tabelę faktów FactFlights, założyć indeks kolumnowy (COLUMNSTORE)
2. Utworzyć jedną paczkę przeładowującą plik CSV do tabeli faktów
3. Utworzyć **jeden** CM do plików CSV z faktami, CM powinien być konfigurowalny (ścieżka w zmiennej lub parametrze)
4. W DataFlow dla faktów wykonać następujące transformacje:
   1. Pozbyć się z kolumn wartości „NA”
   2. Wszystkie kolumny z datami i godzinami zapisane jako INT przekonwertować na DATE lub TIME (z tym jest dużo pracy)
   3. Za pomocą Lookup’ów pobrać klucze surogatowe wymiarów
5. Pozbyć się wartości NULL na kluczach (podmienić na -1)

-------------------- master

1. ~~Utworzyć paczkę „Master”, która~~
   1. Wywołuje paczki ładujące wymiary
   2. W pętli ForEach pobiera wszystkie pliki z faktami ze wskazanego w parametrze folderu

-------------------- uwagi ogólne

1. Dodać logowanie
2. ~~Dane tekstowe zmienić na UNICODE, pobierając ustawić język na English (United States)~~
3. ~~Latitude/Longtitude pobrać jako NUMERIC 14/10~~
4. ~~W pliku csv przed załadowaniem podmienić ręcznie lotnisko w linii 1253 (najlepiej w notatniku albo N++):~~
   1. ~~Z: DBN,"W. H. \"Bud\" Barron ","Dublin","GA","USA",32.56445806,-82.98525556)~~
   2. ~~Na: DBN,"W. H. Bud Barron ","Dublin","GA","USA",32.56445806,-82.98525556~~

-------------------- raporty

1. W bazie danych utworzyć schemat [raporty]
2. W schemacie dodać 4-5 zapytań analitycznych do hurtowni które podsumowują załadowane dane w różnych przekrojach. W zapytaniach użyć funkcji okienkowych.