Airlines Project Hadoop

Problem Statement:

A. Find list of Airports operating in the Country India
B. Find the list of Airlines having zero stops
C. List of Airlines operating with code share
D. Which country (or) territory having highest Airports
E. Find the list of Active Airlines in United state

Dataset Description:

In this use case, there are 3 data sets:

Final_airlines, routes.dat, airports_mod.dat
****

Air Ports data set i.e. airports_mod.dat

It contains the following fields:

Airport ID   Unique OpenFlights identifier for this airport.
Name   Name of airport. May or may not contain the City
name.
City   Main city served by airport. May be spelled
differently from Name.
Country   Country or territory where airport is located.
IATA/FAA   3-letter FAA code, for airports located in Country
"United States of America"
3-letter IATA code, for all other airports.
Blank if not assigned.
ICAO   4-letter ICAO code.
Blank if not assigned.
Latitude   Decimal degrees, usually to six significant digits.
Negative is South, positive is North.

Longitude   Decimal degrees, usually to six significant digits. Negative is West, positive is East.

Altitude   In feet.

Timezone   Hour offset from UTC. Fractional hours are expressed as decimals, eg. India is 5.5.

DST   Daylight savings time. One of E (Europe), A (US/Canada), S (South America), O (Australia), Z (New Zealand), N (None) or U (Unknown). See also: Help: Time

Tz database time Timezone in "tz" (Olson) format, eg. "America/Los_Angeles". zone

Air Lines Data set:

It contains the following fields:

Airline   Unique OpenFlights identifier for this airline. ID

Name   Name of the airline.

Alias   Alias of the airline. For example, All Nippon Airways is commonly known as "ANA".

IATA   2-letter IATA code, if available.

ICAO   3-letter ICAO code, if available.

Callsign   Airline callsign.

Country   Country or territory where airline is incorporated.

Active "Y" if the airline is or has until recently been operational, "N" if it is defunct. This field is not reliable: in particular, major airlines that stopped flying long ago, but have not had their IATA code reassigned (eg. Ansett/AN), will incorrectly show as "Y".

Routes Data set i.e. routes.dat:

It contains the following fields:

Airline   2-letter (IATA) or 3-letter (ICAO) code of the airline.

Airline ID — Unique OpenFlights identifier for airline (see Airline).

Source airport — 3-letter (IATA) or 4-letter (ICAO) code of the source airport.

Source airport ID — Unique OpenFlights identifier for source airport (see Airport)

Destination airport — 3-letter (IATA) or 4-letter (ICAO) code of the destination airport.

Destination airport ID — Unique OpenFlights identifier for destination airport (see Airport)

Codeshare — "Y" if this flight is a codeshare (that is, not operated by Airline, but another carrier),empty otherwise.

Stops — Number of stops on this flight ("0" for direct)

Equipment — 3-letter codes for plane type's generally used on this flight, separated by spaces