

# Assignment #1

---

1. [**Big Data concept**] (15) Give one example of Big Data application you know. Use the detailed example to explain each of the five Big V's.
2. [**Relational Data Model**] (35) As of January 2017, the OpenFlights Airports Database (<https://openflights.org/data.html>) contains over 10,000 airports, train stations and ferry terminals spanning the globe. Each entry in the Airport table contains the following:  
-----

**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. See countries.dat to cross-reference to ISO 3166-1 codes.

**IATA** 3-letter IATA code. Null if not assigned/unknown.

**ICAO** 4-letter ICAO code.

**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** Hours 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 zone** Timezone in "tz" (Olson) format, eg. "America/Los\_Angeles".

**Type** Type of the airport. Value "airport" for air terminals, "station" for train stations, "port" for ferry terminals and "unknown" if not known. In airports.csv, only type=airport is included.

**Source** Source of this data. "OurAirports" for data sourced from OurAirports, "Legacy" for old data not matched to OurAirports (mostly DAFIF), "User" for unverified user contributions. In airports.csv, only source=OurAirports is included.

---

- a. (15) Consider the following terms: *relation schema*, *relational database schema*, *domain*, *attribute*, *attribute domain*, *relation instance*. Give what these terms are with the above Airport database. Give one small (4-5 tuples) instance of the Airport table.
  - b. (20) There are three databases in the OpenFlight dataset: Airport, Airline, and Route. Give the schema of these three databases and mark the primary keys, foreign keys and provide examples of functional dependencies you identified over the three tables. [You may draw a diagram to illustrate the schema, PKs, FKs and FDs]
3. **[Functional Dependencies] (30)** Recall Armstrong's axioms.
- Reflexivity rule: if  $Y \subseteq X$  then  $X \rightarrow Y$
  - Augmentation rule: if  $X \rightarrow Y$  then  $XZ \rightarrow YZ$
  - Transitivity rule: if  $X \rightarrow Y$  and  $Y \rightarrow Z$  then  $X \rightarrow Z$
- a. Give two examples for using Armstrong's inference rules to induce new FDs from the set of FDs you designed in question 2 (b).
  - b. Prove the following inference rules also hold, using FD definition and Armstrong's Axioms.
    - i. decomposition rule: **if  $X \rightarrow YZ$  then:  $X \rightarrow Y$  and  $X \rightarrow Z$**
    - ii. Psuedo transitivity: **if  $X \rightarrow Y$  and  $YW \rightarrow Z$  then:  $XW \rightarrow Z$**

4. **[Normalization] (20)** Given a relation  $R(A_1, A_2, A_3, A_4)$ , with three FDs  $A_2, A_3 \rightarrow A_4$ ;  $A_3, A_4 \rightarrow A_1$ ;  $A_1, A_2 \rightarrow A_3$ . Provide the 3NF and BCNF form of the schema and explain why.