

## INFO-H-419 – Data Warehouses

First session examination

### Question 1: Data Warehouse Design (8 points)

A data warehouse of a train company contains information about train segments. It consists of six dimensions, namely, departure station, arrival station, trip, train, arrival time, and departure time, and three measures, namely, number of passengers, duration, and number of kilometers.

1. Design a conceptual schema for the data warehouse. Propose dimension attributes and dimension hierarchies.
2. Translate the conceptual schema into a relational schema. Clearly indicate primary and foreign keys in your tables.
3. For the relational schema obtained in the previous question, write in SQL the following queries.
  - (a) Total number of kilometers made by Alstom trains during 2018 departing from French or Belgian stations.
  - (b) Total duration of international trips during 2018, that is, trips departing from a station located in a country and arriving at a station located in another country.
  - (c) Total number of trips that departed from or arrived at Paris during July 2018.
  - (d) Average duration of train segments in Belgium in 2018.
  - (e) For each trip, average number of passengers per segment, that is, take all the segments of each trip, and average the number of passengers.

### Question 2: View Materialization (4 points)

Consider the graph in Fig. 1, where each node represents a view and the numbers are the costs of materializing the view. Assuming that the bottom of the lattice is materialized, determine using the View Selection Algorithm the five views to be materialized first.

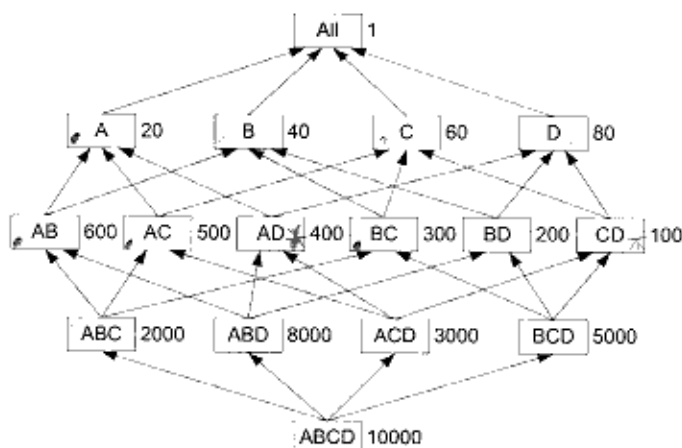


Figure 1: A data cube lattice

**Question 3: Indexing (4 points)**

Consider the Employee and the Department tables below.

Employee					
Employee Key	Employee Name	Title	Address	City	Department Key
e1	Peter Brown	Dr.	...	Brussels	d1
e2	James Martin	Mr.	...	Wavre	d1
e3	Ronald Ritchie	Mr.	...	Paris	d2
e4	Marco Benetti	Mr.	...	Versailles	d2
e5	Alexis Manoulis	Mr.	...	London	d3
e6	Maria Mortsel	Mrs.	...	Reading	d3
e7	Laura Spinotti	Mrs.	...	Brussels	d4
e8	John River	Mr.	...	Waterloo	d4
e9	Bert Jasper	Mr.	...	Paris	d5
e10	Claudia Brugman	Mrs.	...	Saint-Denis	d5

Department		
Department Key	Department Name	Location
d1	Management	Brussels
d2	Production	Paris
d3	Marketing	London
d4	Human Resources	Brussels
d5	Research	Paris

1. Show how a bitmap index on attribute Employee.Title would look like. Compress the bitmap values using run-length encoding.
2. Show how a bitmap join index for tables Employee and Department on attribute *city* is given next.

*location*

**Question 4: Aggregate Computation (4 points)**

A cube Sales(A, B, C, D, Amount) has to be fully materialized. Suppose that the cube contains 64 tuples and that that every GROUP BY with  $k$  attributes has  $2^{k+2}$  tuples. Assume that sorting takes typically  $n \log(n)$  units of time while a full table scan takes  $n$  units of time.

1. Compute the cube using the PipeSort algorithm.
2. Compute the gain of applying the PipeSort compared to the cost of computing all the views from scratch.