Database Management System for Railway Company: Passenger and Freight Trains

Cazzolato Kevin

2023

1 Problem description

The goal is to create a database for a railway transport company that manages passenger transport during the time slot 06:00-21:00 and freight transport during the time slot 21:00-06:00. Companies will be able to rent one or more wagons for a route, where they can freely place their goods.

For trains in the second time slot, the intention is to record the train code (uniquely identified by 5 numbers) and the maximum number of wagons that can be transported.

For the wagons, the intention is to record: the rental income and the wagon code (uniquely identified by 5 letters).

For companies renting the wagons, the intention is to record: location, social region, name, and contact phone number. It is also required to record the buyer, seller, and date of each transaction.

For trains in the first time slot, the intention is to record the train code (uniquely identified by 5 numbers), the maximum number of carriages, and the potential number of passengers.

Trains in the second time slot will have a variable number of carriages. The intention is to record for each carriage: the presence of a bathroom, the presence of air conditioning, and the type of carriage, which can be VIP, first class, or second class.

Trains, regardless of the operating time, will have a specific route covering multiple stations to facilitate the loading/unloading of goods and passengers. A route segment (stretch between two adjacent stations) will be covered by at most one train, but a route can be traveled by multiple trains simultaneously.

For the route segment, the intention is to record: the train that traveled it, the actual time, the scheduled time, the date, and the ticket price.

For the station, the intention is to record: the number of tracks, the station master, the phone number, latitude, and longitude.

Each station will carry out maintenance within its facilities (which can be of various kinds) and will be entrusted to technicians.

For maintenance, the intention is to record: the cost, the number of employees employed, the start date, the end date, and a description of the operation.

The client wants to record, for each employee: job role, first name, last name, phone number, hours per week according to contract, hourly wage, and the train in which they work (if they work on a train).

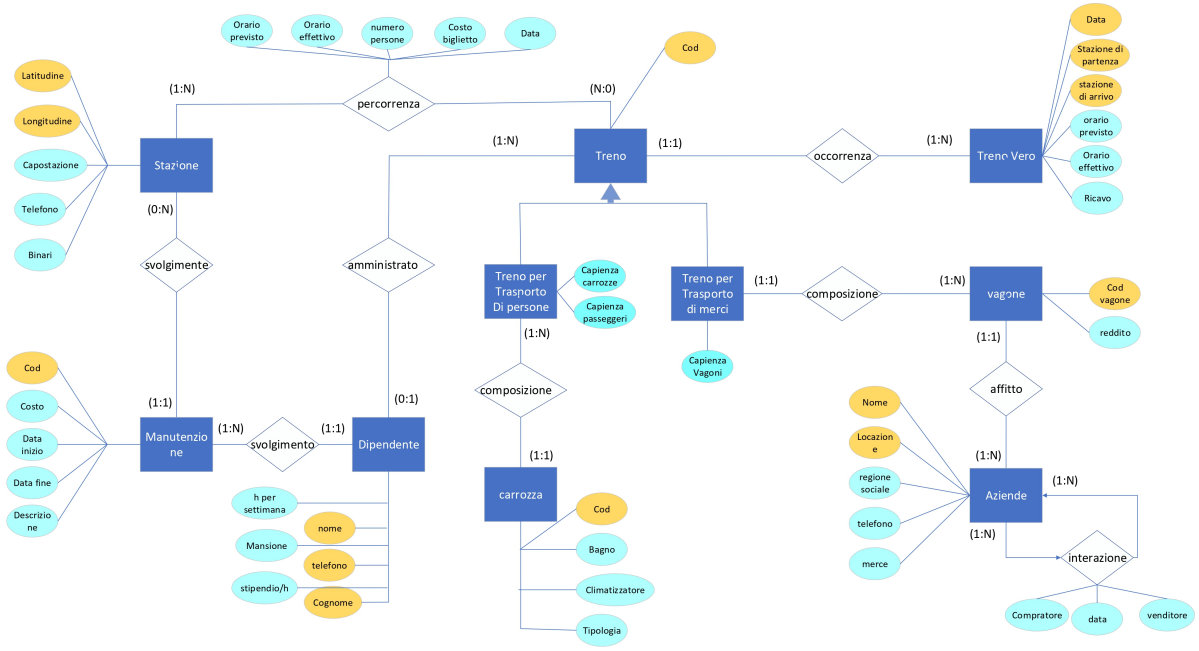
It is necessary to keep track of each train passing through each station. In particular, the intention is to record: date, scheduled arrival time, actual arrival time, departure station, arrival station, and revenue.

2 Glossary of terms

Term	Definition	Synonyms	Link
Companies	organizations that buy and/or sell goods	firm	Wagons, route
Train	Railway convoy with a variable number of carriages/wagons. It can be used for freight or passenger transport		route, company, wagon, carriage, employees
Wagon	Railway convoy for transporting goods		train, company
Rental income	income received by the client for renting the requested wagon		companies, wagon
Route	Path taken by one or more trains (e.g., "Venice-Trieste") composed of multiple route segments		Train, stations,Route Segment
Route Segment	Path covering two adjacent stations		Route, stations, trains
Stations	Place where trains stop, composed of one or more tracks		Trains, employees, maintenance
Carriages	Railway convoy for transporting passengers		Train
Maintenance	set of operations necessary to maintain the station's functionality and efficiency		stations, employees
Employees	subordinates/workers		trains, maintenance, stations

Observation: To avoid excessive verbosity, sentence grouping is omitted here, as differentiation of them has already been highlighted in the problem description by indentation.

3 Entity relationship schema



NOTE: Entities in orange represent the primary key.

4 Data Dictionary

Entity	Description	Attributes	Identifier
Station	transit station for trains	Station master, Phone, Tracks, Latitude, Longitude	Latitude, Longitude
Maintenance	restructuring/repair operations	cost, start date, end date, description, maintenance code	maintenance code
Employees	subordinate workers	hours per week, job role, hourly wage, name, phone, last name	name, phone, last name
Carriage	related to passenger transport	Bathroom, Air conditioning, type, carriage code	carriage code
Passenger transport train	train operating in the time slot 6:00-21:00	Passenger capacity, Carriage capacity, Train code	Train code
Freight transport train	train operating in the time slot 21:00-6:00	Wagon capacity, Train code	Train code
Wagon	related to the freight transport train that companies can rent	income, wagon code	wagon code
Actual Train	occurrences of trains and related information	Scheduled time, actual time, date, departure station, arrival station, train code	date, departure station, arrival station, train code
Companies	companies that buy/sell goods	social region, phone, Name, location	Name, location
Actual Train	occurrence of each train	Date, Departure station, Arrival station, Scheduled time, Actual time, revenue	date, Departure station, Arrival station

Relation	Description	Composition	Attributes
Travels	train travels on a route segment	Train code, latitude, longitude	Scheduled time, actual time, number of passengers, ticket cost, date
Carries out (*maintenance-station*)	maintenance carried out in stations	maintenance code, latitude, longitude	
Carries out (*maintenance-employees*)	maintenance performed by employees	name, phone, last name, maintenance code	
Managed by	train managed by employees	name, phone, last name, train code	
Composition (*carriage-passenger train*)	composition of the passenger train with carriages	carriage code, train code	
Composition (*freight train-wagon*)	composition of the freight train with wagons	train code, wagon code	
Rental	wagon rental by companies	name, location, wagon code	
Interaction	interaction between seller and buyer companies	name, location , name, location	buyer, date, seller
Records	data records related to a train and its route segment	date, revenue, time	

5 Constraints

- Freight trains will operate exclusively in the time slot 21:00-06:00.
- Passenger trains will operate exclusively in the time slot 06:00-21:00.
- There cannot be more passengers/carriages/wagons than permitted.

6 General considerations

The primary key for employees was decided to be a combination of name, last name, and phone number, but for convenience and increased security, it will be changed to a unique code in restructuring. The generalization Train with child entities Freight Transport Train and Passenger Transport Train is complete as it covers the main types of trains. It is assumed that maintenance will not interfere with regular train traffic. It was decided to directly include the revenue attribute, which can be calculated from the number of passengers and ticket price, as it is not necessary for the requested operations in (8) to maintain this attribute decomposed. It should be noted that the choice of primary key (Name and Location) for companies is well-placed, as, due to antitrust laws, two companies with the same name and location cannot coexist.

7 Volume Table

It is assumed that the client has 40 stations, 120 trains (70 for freight transport and 50 for passenger transport). Additionally, there are approximately 800 employees. Volumes will represent a one-year time frame; note that these are rough estimates and may not accurately reflect reality.

Concept	Type	Volume
Station	E	40
Maintenance	E	250
Employees	E	800
Carriage	E	400 (=50*8)
Passenger transport train	E	35
Freight transport train	E	45
Wagon	E	840 (=70*12)
Actual Train	E	2073600
Companies	E	420
Travels	R	2073600
Connection	R	120
Carries out 1	R	250
Carries out 2	R	250
Managed by	R	360 (=120*3)
Composition 1	R	400
Composition 2	R	840 (=70*12)
Rental	R	840
Interaction	R	302400 (=840*30*12)
Records	R	2073600

Note: The estimate of actual trains has been increased as follows: assuming each station is connected to a maximum of 3 stations, then there are 40*3 total routes. Assuming also that each route has exactly one train traveling it and that the travel time between two stations is 30 minutes, then:

$$120 * 2 * 15 = 108000$$

is the number of rows for actual passenger trains; similarly, for freight trains, there will be at most 5760 rows.

8 Operations of Interest

Operation	Type	Frequency
Estimate of expenses and revenues	Interactive	2/month
Route segment with lowest revenue	Interactive	1/month
Insertion of a maintenance operation within a station	Interactive	1 time every 2 months

9 Redundancy Analysis

It is observed that, although there is a cycle between Train, station, maintenance, and employee, this is not a redundancy as no concepts are expressed twice. Additionally, there seems to be redundancy in actual trains, as the information within them is all derivable from entities: Train, Route Segment, and Station. Therefore, the decision will be made on whether to maintain or eliminate this redundancy.

9.1 Operation 1:

9.1.1 In the absence of redundancies

Concept	Construct	Accesses	Type
Train	E	120	L
Travels	R	58800(=35*2*14*60)	L
Station	E	40	L
Maintenance	E	10	L
Wagon	E	315(=7*45)	L
Composition	R	415800(=45*7*2*11*60)	L

Note: The calculation includes an estimate of revenue from first and second time slot trains and an estimate of maintenance expenses. An average of about 7 wagons per train is estimated, and a travel time of approximately 30 minutes per route segment is also estimated.

9.1.2 With redundancies

Concept	Construct	Accesses	Type
Actual Train	E	345600	L
Maintenance	R	10	L

Observing the strong evidence against removing redundancies, it is decided to retain them for lower read access in operation 1). No redundancy analysis will be performed on other operations as the access flow would be significantly lower and therefore would not affect the choice of whether or not to retain them.

10 Elimination of Generalizations

It is noted that there is only one generalization: with parent Train and children Passenger Transport Train and Freight Transport Train. Since there is interest in studying the train types separately, it is decided to separate the generalization and treat Passenger Transport Train and Freight Transport Train as two distinct entities.

11 Relationship Partitioning

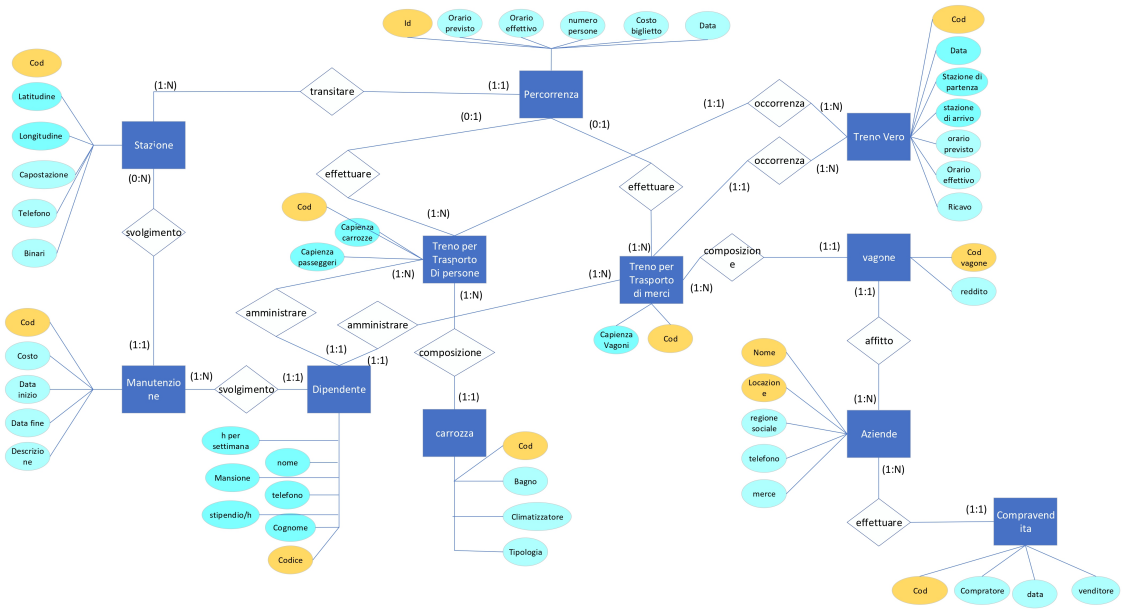
The relationship Travels is a many-to-many relationship between Train and Station. Therefore, a new entity "travels" is created, associated one-to-many with Train and one-to-many with Station. Additionally, a new entity "Transaction" is created, representing sales transactions between companies, which will be a one-to-many relationship between companies and transactions.

12 Primary Key Selection

- Passenger Transport Train: the primary key is chosen as *train code*
- Freight Transport Train: the primary key is chosen as *train code*
- Travels: the chosen identifier is *id* (row identifier)
- Station: the chosen identifier is a *station code*
- Maintenance: the chosen identifier is *maintenance code*
- Employee: an identifier *employee code* is introduced
- Carriage: the chosen identifier is *carriage code*
- Actual Train: an identifier *id* (row identifier) is introduced

- Wagon: the chosen identifier is *wagon code*
- Companies: the chosen identifier is *name, location*
- Transaction: the identifier *Code* is introduced

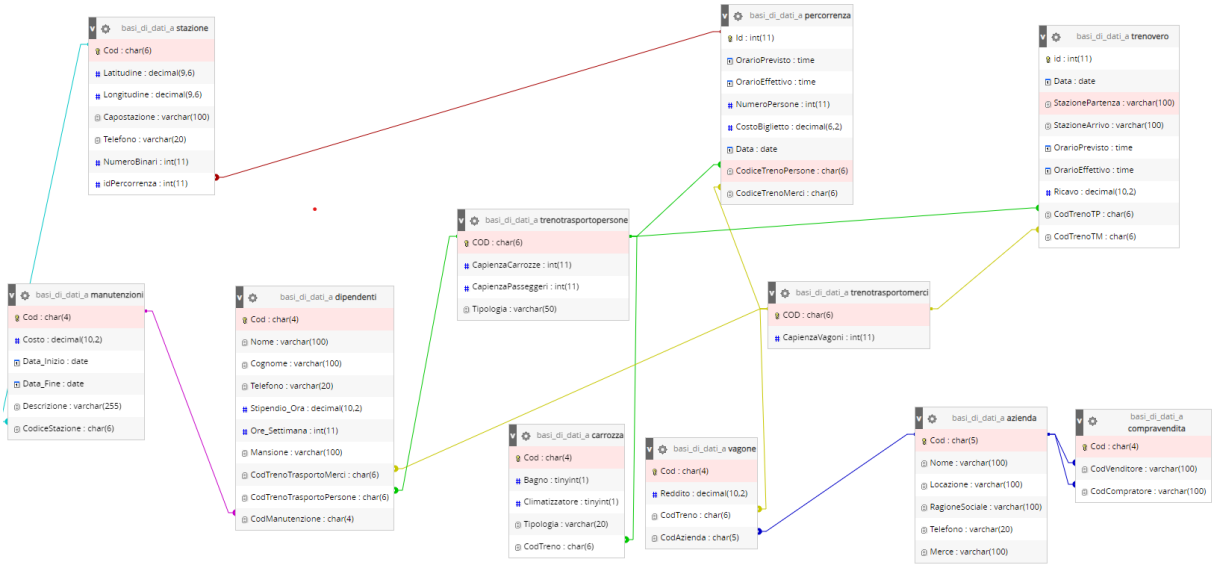
13 Restructured ER Diagram



14 Relational Model

- Passenger Transport Train (Code, Carriage Capacity, Passenger Capacity, Type)
- Freight Transport Train (Wagon Capacity, Code)
- Travels (Id, scheduled time, actual time, number of passengers, ticket code, date, $\overline{Passengertransporttraincode}$, $\overline{Freighttransporttraincode}$)
- Station (Code, Latitude, Longitude, Station Master, Phone, Tracks, $\overline{TravelID}$)
- Maintenance (Code, Cost, Start date, End date, Description, $\overline{Stationcode}$)
- Employee (Code, last name, first name, phone, hourly wage, job role, hours per week, $\overline{Maintenancecode}$, $\overline{Passengertraincode}$, $\overline{Freighttraincode}$)
- Carriage (Code, Bathroom, Air Conditioning, Type, $\overline{Traincode}$)
- Wagon (Code, income, $\overline{Traincode}$, $\overline{CompanyName}$, $\overline{CompanyLocation}$)
- Actual Train (Code, Date, Departure Station, Arrival Station, scheduled time, actual time, Revenue, $\overline{PassengerTraincode}$, $\overline{FreightTraincode}$)
- Companies (Name, Location, Social Region, Phone, Goods)
- Transaction (Code, date, $\overline{Buyer}$, $\overline{Seller}$)

15 Logical Schema



16 Normalization

The database is in the first normal form, as all columns are atomic.
The second normal form is also respected because it meets the first normal form, and every column in each entity fully depends on the primary key.
The database is not in the third normal form, as the revenue attribute does not fully depend on the primary key. However, it was decided to keep the database in this form for better clarity and usability, as indicated in the redundancy analysis.

17 Triggers

17.1 First Non-Expressible Constraint

Freight trains will travel routes exclusively between 21:00 and 06:00.

```
DELIMITER $$
CREATE TRIGGER Trg_FreightTrainHours
BEFORE INSERT ON Travels
FOR EACH ROW
BEGIN
    DECLARE insert_time TIME;
    DECLARE freight_train_code CHAR(6);

    SET insert_time = NEW.ScheduledTime;
    SET freight_train_code = NEW.FreightTrainCode;

    IF freight_train_code IS NOT NULL AND
        (insert_time < '21:00:00' OR insert_time > '06:00:00')
    THEN
        SIGNAL SQLSTATE '45000'
        SET MESSAGE_TEXT = 'Freight trains can only travel between 21:00 and 06:00.';
    END IF;
END $$
DELIMITER ;
```

17.2 Second Non-Expressible Constraint

Passenger trains will travel routes exclusively between 06:00 and 21:00.

```
DELIMITER $$
CREATE TRIGGER Trg_PassengerTrainHours
BEFORE INSERT ON Travels
FOR EACH ROW
BEGIN
    DECLARE insert_time TIME;
    DECLARE passenger_train_code CHAR(6);

    SET insert_time = NEW.ScheduledTime;
    SET passenger_train_code= NEW.PassengerTrainCode;

    IF passenger_train_code IS NOT NULL AND
        (insert_time > '21:00:00' OR insert_time < '06:00:00')
    THEN
        SIGNAL SQLSTATE '45001'
        SET MESSAGE_TEXT =
        'Passenger trains can only travel between 06:00 and 21:00.';
    END IF;
END $$
DELIMITER ;
```

17.3 Third Non-Expressible Constraint

17.3.1

There cannot be more passengers than permitted.

```
DELIMITER $$
CREATE TRIGGER Trg_CheckPassengerCapacity
BEFORE INSERT ON Travels
FOR EACH ROW
BEGIN
    DECLARE passenger_capacity INT;

    SELECT PassengerCapacity INTO passenger_capacity
    FROM PassengerTransportTrain
    WHERE CODE = NEW.PassengerTrainCode;

    IF NEW.NumberOfPassengers > passenger_capacity THEN
        SIGNAL SQLSTATE '45002'
            SET MESSAGE_TEXT = 'The number of passengers exceeds the train capacity.';
    END IF;
END $$
DELIMITER ;
```

NOTE: The remaining triggers are omitted here as they are similar to the one above, with only the variable type changing.

18 Stored Procedures

18.1 Find Trains with Delays over 10 Minutes

```
DELIMITER $$
CREATE PROCEDURE FindDelayedTrains()
BEGIN
    SELECT *
    FROM ActualTrain
    WHERE TIMEDIFF(ActualTime, ScheduledTime) > '00:10:00';
END $$
DELIMITER ;
```

18.2 Inserting Trains into Actual Train

```
DELIMITER $$
CREATE TRIGGER Trg_InsertActualTrain
AFTER INSERT ON Travels
FOR EACH ROW
BEGIN
    DECLARE departure_station VARCHAR(100);
    DECLARE arrival_station VARCHAR(100);
    DECLARE revenue INT;

    SELECT DepartureStation INTO departure_station
    FROM Station
    WHERE TravelId = NEW.Id;

    SELECT ArrivalStation INTO arrival_station
    FROM Station
    WHERE TravelId = NEW.Id;

    SET revenue = NEW.NumberOfPassengers * NEW.TicketPrice;

    INSERT INTO ActualTrain (Date, DepartureStation, ArrivalStation,
    ScheduledTime, ActualTime, Revenue, PassengerTrainCode, FreightTrainCode)
    VALUES (NEW.Date, departure_station, arrival_station, NEW.ScheduledTime, NEW.ActualTime, revenue,
END $$
DELIMITER ;
```

19 Queries

19.1 Costs and Revenues of the Last Two Months

```
    SELECT (SELECT SUM(Cost) FROM Maintenance) AS Costs,
        SUM(actualtrain.Revenue) AS Revenue,
        SUM(actualtrain.Revenue)-(SELECT SUM(Cost) FROM Maintenance) AS Total
        FROM ActualTrain
        WHERE TIMESTAMPDIFF(month,CURRENT_DATE,actualtrain.Date)<=2
```

19.2 Route Segment with the Lowest Revenue

```
SELECT actualtrain.DepartureStation, actualtrain.ArrivalStation
FROM actualtrain
WHERE Revenue=(SELECT MIN(Revenue) FROM actualtrain)
```

19.3 Inserting Maintenance within a Station

```
INSERT INTO Maintenance (Code, Cost, StartDate, EndDate, Description, StationCode)
VALUES ('M011', 5250.00, '2023-07-10', '2023-07-15', 'Scheduled maintenance', 'S005')
```