☕DATA 514 Section 9 Worksheet

This doc outlines the running **Payroll/Regist example**, as well as how it should be implemented in the various **NoSQL** options we'll be discussing. We will add a new table to the running example, ParkingTickets.

As a reminder, these are the schema and example values. Note that Frances doesn't have a car; Magda has two, and nobody employed at UW owns the Aston Martin.

Payroll

UserID	**Name**	**Job**	**ParkingPermit**
123	Leslie	TA	C15
345	Frances	TA	NULL
567	Magda	Prof	E18
789	Quinn	Prof	NULL

Regist

UserID	**Car**	**LicensePlate**
123	Charger	123 AAA
567	Civic	MMM 1234
567	Ferrari	MMM 5678
789	Picklemobile	PIK 1024
007	Aston Martin	XYZ 0007

ParkingTickets

LicensePlate	**ParkingLot**	**Date**	**Amount**
MMM 1234	C15	2022-11-20	$10
MMM 1234	E01	2022-11-21	$15
PIK 1024	E18	2022-11-22	$10
XYZ 0007	C19	2022-11-01	$10
MMM 1234	E01	2022-11-22	$20

Application

This is a parking enforcement app which supports the following methods:

1. [infrequent] Listing the permitted parking lot and per-car tickets incurred by each user
    - method sig is `uid -> [ {car1, [{tix1}, {tix2}]}, {car2, []} ]`
2. [frequent] Counting how many tickets a license plate has ever had
    - method sig is `plate -> int` or it can be `plate -> [ {tix1}, {tix2} ]`
    - We use the output of this method to determine the citation amount.
3. [multiple times / sec] Determining whether a plate is allowed to be a specific lot
    - method sig is `(plate, lot) -> true/false` or it can be `plate -> lot`

Because NoSQL design is so intimately tied to its use-cases, there are three decisions which will need to be made for each system:

- **Method 3**: should we store the license plate and permitted lot as the key, with the value being true and the expectation that values which don't exist should default to false?
- **Method 2**: should we store a count of the tickets, or the actual tickets themselves? Note that the latter introduces the possibility of data anomalies (which may be deemed acceptable)
- How should we handle **cars without owners** (the Aston Martin) or **employees without cars** (Frances)? Is data loss acceptable or not?

Question 1: Wide Column Database

Describe how you would implement the Payroll/Regist application using a wide column database. Your description should contain enough detail that we understand the keyspace (eg, do you have different “types” of keys?), the column families and their contained columns, as well as whether you use explicit timestamps.

Question 2: Graph Database

Describe how you would implement the Payroll/Regist application using a graph database.