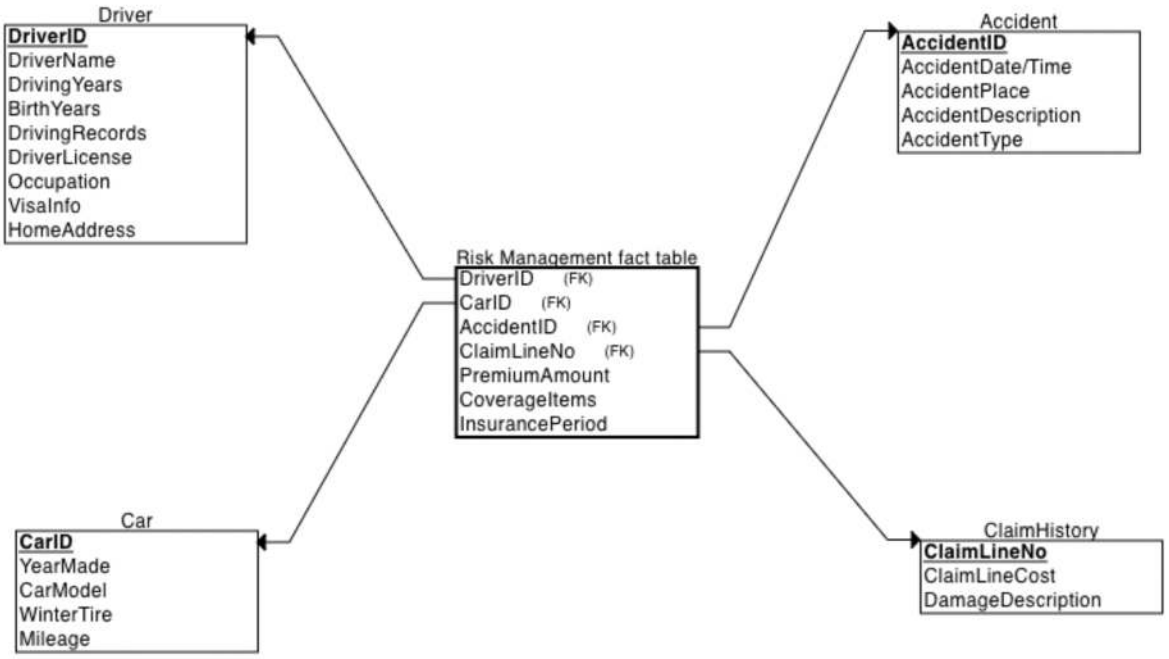
Assignment 3 - Data Warehouse Modeling

Database Design Management and Applications (McMaster University)

1. **Sources for information gathering**

I reviewed the three links provided in the assignment outline, and I also went through Travelers Canada's website, which is an insurance company that I personally got my car insurance from. I learned that many factors can affect the risk of auto insurance and insurance rates, such as demographic information including age, gender, location and occupation, credit score, and driver's license. The information of the vehicle like car model, make, mileage, year of the vehicle, and condition of the vehicle would also be the references to determine the risk. Also, based on my own experience of negotiating with auto insurance companies, they would also take my nationality and driving record into consideration to decide whether to proceed with my application for insurance. A data warehouse is an essential need for an auto insurance company because the amount of data from clients every year is huge for the company, especially when the company has a vast number of clients. Also, many auto insurance companies would store historical data about clients' usage of the insurance, which would be taken into further analysis for the company. So, data warehouses are necessary for auto insurance companies to store and classify various data, and then it would be more convenient for them to retrieve and analyze information, which can help them manage risks.

1. E-R Model diagram



- Explanation:
    - This table is designed for auto insurance companies' risk management process. Beside all the foreign keys, the fact table contains the premium amount of the insurance policy, the items included in the insurance coverage, and the duration or period of the insurance policy. All the information could be determined by analyzing risks using the data from all dimensions.
    - There are four dimensions included in this table, namely Driver, Accident, Car, and ClaimHistory.
        - **Driver:** This dimension contains all the necessary personal information that is needed for risk analysis, which includes the basic information, such as the driver's name, driving years,

birthday, home address, occupation, and driver's license. Also, some additional information would be needed for further analysis, like driving records, which are used to analyze drivers' driving behavior and general risk. Based on my own experience of applying for insurance for my car, companies would charge a higher monthly premium for people who are not citizens like international students. Therefore, I also included an attribute for this information called "VisaInfo".

- **Car:** This dimension contains information about cars, which includes the year that the car is made, car model, and mileage. I also included an attribute called "WinterTire" because sometimes companies would check if drivers have winter tires for their cars as a part of the risk analysis, and if they do, the premium can be lowered a little.
- **Accident:** This dimension contains drivers' accident records including accident date and time, where accidents happened, a description for each accident, and the type of accident (e.g., minor, severe). This is an important section for the risk analysis. Some companies in the middle-risk market can even refuse to proceed to issue an insurance policy to a driver if there are too many accidents on the driver's record.
- **ClaimHistory:** This dimension contains drivers' historical claim records including claim line numbers (it is also the primary key of

this dimension), claim line costs, and damage description. This is for examining how many claims a driver previously went through so that the company can determine how frequently the driver uses the coverage.

2. Sample data

○ Driver

DriverID	DriverName	DrivingYears	BirthYears	DrivingRecords	DriverLicense	Occupation	VisaInfo	HomeAddress
D100000	Leyla Tanner	3	1995	No previous violation / accident; Good driving style	G2	Laywer	Citizen	107 Waterloo Ave, North York, ON M3H 3Y6
D100001	Mehmet Huffman	3	1995	No previous violation / accident; Good driving style	G	Accountant	Citizen	17 Northcote Ave, Toronto, ON M6J 3K2
D100002	Jerry Hampton	4	1992	Three red light violation; One severe accident; Reported stunning driving	G	Business owner	Citizen	149 Marlborough St, Brantford, ON N3S 4S8
D100003	Ahmad Combs	1	2001	One red light violation; No accident; Good driving style	G2	Undergraduate student	Visa student	274 Sherbrooke St, Peterborough, ON K9J 2N6
D100004	Kade Rogers	6	1990	One red light violation; Two minor collisions; Good driving style	G	Data engineer	Permanent resident	99 Forward Ave, London, ON N6H 1B8
D100005	Caitlyn Kirby	10	1985	Five red light violations; One minor collision; Good driving style	G	Senior business analyst	Citizen	6807 111 St NW, Edmonton, AB T6H 3G2

○ Car

CarID	YearMade	CarModel	WinterTire	Mileage
C100000	2018	Nissan Sentra	Yes	48000
C100001	2019	Honda Civic	Yes	45000
C100002	2018	Subaru BRZ	Yes	70000
C100003	2021	Mazda 3	No	12000
C100004	2016	Hyundai Elantra	Yes	92000
C100005	2010	Toyota Corolla	Yes	150000

○ Accident

AccidentID	AccidentDate/Time	AccidentPlace	AccidentDescription	AccidentType
A100000	N/A	N/A	N/A	N/A
A100001	N/A	N/A	N/A	N/A
A100002	10-29-2020, 12:30 AM	Near 2950 Bayview Avenue, Toronto, ON	Severe collision with another car due to stunning driving	Severe
A100003	N/A	N/A	N/A	N/A
A100004	12-11-2021, 6 PM	Near 4K Spadina Ave., Toronto, ON M5V 3Z2	Minor collision with another car due to icy and slippery road	Minor
A100005	06-15-2018, 7 PM	Near 9510 105 St NW, Edmonton, AB T5K 0Z5	Minor collision with another car due to icy and slippery road	Minor

- **ClaimHistory**

ClaimLineNo	ClaimLineCost	ClaimDescription
CH100000	0	N/A
CH100001	0	N/A
CH100002	10000	Full repair and maintenance for the car
CH100003	0	N/A
CH100004	1000	Repair for the car's front lights
CH100005	1000	Repair for the car's front lights

- **Risk Management Fact Table**

DriverID	CarID	AccidentID	ClaimLineNo	PremiumAmount	CoverageItems	Insurance Period
D100000	C100000	A100000	CH100000	6000	item 1; item 2; item 3	2022-1-1 to 2023-1-1
D100001	C100001	A100001	CH100001	5500	item 1; item 2; item 3	2022-1-1 to 2023-1-1
D100002	C100002	A100002	CH100002	7000	item 1; item 2; item 3	2022-1-1 to 2023-1-1
D100003	C100003	A100003	CH100003	5000	item 1; item 2; item 3	2022-1-1 to 2023-1-1
D100004	C100004	A100004	CH100004	4800	item 1; item 2; item 3	2022-1-1 to 2023-1-1
D100005	C100005	A100005	CH100005	3600	item 1; item 2; item 3	2022-1-1 to 2023-1-1

3. Sources of data and challenges

- **Internal:** The main internal source of data would be clients' personal information that is stored in insurance companies' systems, and some historical claims data from clients can also be loaded into data warehouses for risk analysis. Also, some companies would implement technologies to track how clients drive and where they usually go, which can generate a lot of data that can help examine clients' driving behaviour, and such information would be loaded into the data warehouse as well.
- **External:** This data warehouse needs a lot of input from external sources. New clients' personal and vehicle information would be provided by the drivers themselves, and sometimes the insurance companies would need to request for accident records data from police office systems. To

examine clients' driving behaviour, insurance companies would also need the driving records data from clients' previous insurers.

- **Challenges:** The process of collecting data can be time-consuming and inefficient since there is usually a large amount of data needed for determining the risks and premiums to issue a policy for each client. Sometimes there is also a risk that clients might provide incorrect information about themselves and their cars. For example, the premium of a car insurance policy varies a lot depending on clients' home addresses, and sometimes clients would provide inaccurate addresses in order to lower the monthly premium. Also, some insurance companies like Travelers Canada would use mobile systems that can track drivers' driving behaviour, and the implementation of such technology can be costly, especially for some small insurance companies. On the other hand, the ethical considerations for data collection can also be a challenge for auto insurance companies since they would collect a wide range of data from clients in order to determine the risks and premiums, and some clients might be concerned about providing all the personal information and getting their locations tracked by companies' systems.

4. **Possible risk analysis questions (with explanation)**
    - Verifying if clients' personal information aligns with the records in police office systems.
        i. This would determine whether the information clients provided is authentic. If it is not, the insurance application would be rejected.
    - Verifying clients' home addresses with the support of car tracking systems.
        i. This would help insurers accurately determine the premiums of clients' insurance policies.
    - Analyzing clients' credit rating based on their credit card usage and bank statements.
        i. This should be an important part of risk analysis. Clients' credit rating can help insurers decide if an insurance application should be accepted or rejected since a low credit score might indicate the risk that a client would not regularly pay the monthly premium.
    - Gathering the information about clients' driving behaviour from their previous insurers.
        i. This can help insurers estimate how frequently clients can possibly encounter accidents, and it can help determine risks.
    - Analyzing clients' previous claim records and identifying the number and types of accidents they experienced.
        i. This can help insurers estimate how frequently clients would use the insurance policies to make claims. It can help insurers

determine how much clients could cost through claims so that insurers can better determine the premiums.

5. Summary

I spent approximately 6 hours on this assignment. I have learned the basic structure of an auto insurance company's data warehouse and the aspects that companies might consider when analyzing the risk of each client. Also, through searching information online, I learned what kind of data auto insurance companies would usually collect for risk analysis and the possible sources of that data.