**NEO4J Queries**

**MODULE: B9DA102 Data Storage Solutions for Data Analytics**

**GROUP B (Slot 4)**

**Subramaniam Kazhuparambil (10524303)**

**Rahul Ramchandra Uppari (10523807)**

**Deeksha Sharma (10522688)**

**Mohit Singh (10525046)**

* **VEHICLE DIMENSION INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///vehicle.csv" AS row CREATE (v:Vehicle) SET v = row {vehicleKey:row.VehicleKey, vehicleID:row.VehicleID, vehicleBodyType:row.VehicleBodyType, vehicleDamageExtent:row.VehicleDamageExtent, vehicleMovement:row.VehicleMovement, speedLimit:toInteger(row.SpeedLimit), driverlessVehicle:row.DriverlessVehicle, parkedVehicle:row.ParkedVehicle, vehicleYear:toInteger(row.VehicleYear), equipmentProblem:row.EquipmentProblem} return v

* **DRIVER DIMENSION INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///driver.csv" AS row CREATE (d:Driver) SET d = row {personKey:row.PersonKey, personID:row.PersonID, driverLicenseState:row.DriverLicenseState, severityOfInjury:row.InjurySeverity, driverDistractedBy:row.DriverDistractedBy, driverSubstanceAbuse:row.DriverSubstanceAbuse, driverAtFault:row.DriverAtFault, nonMotoristSubstanceAbuse:row.NonMotoristSubstanceAbuse} return d

* **CRASH DATE DIMENSION INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///crashdate.csv" AS row CREATE (d:Date) SET d = row {dateKey:row.CarCrashDateKey, date:row.CrashDate, dayOfWeek:row.DayOfWeek\_, dayType:row.DayType, dayOfMonth:toInteger(row.DayOfMonth\_), month:toInteger(row.Month\_), quarter:row.Quarter\_, year:toInteger(row.Year\_)} return d

* **CRASH CONDITIONS DIMENSION INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///crashconditions.csv" AS row CREATE (c:CrashConditions) SET c = row {conditionKey:row.CrashConditionsKey, localCaseNumber:row.LocalCaseNumber, streetName:row.StreetName, streetType:row.StreetType, trafficControl:row.TrafficControl, collisionType:row.CollisionType, weather:row.Weather, surfaceCondition:row.SurfaceCondition, light:row.Light} return c

* **CASE DETAIL INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///casedetail.csv" AS row CREATE (c:CaseDetail) SET c = row {caseKey:row.CaseKey, reportNumber:row.ReportNumber, agencyName:row.AgencyName, acrsReportType:row.ACRSReportType} return c

* **CRASH ANALYSIS FACT INSERT QUERY**

LOAD CSV WITH HEADERS FROM "file:///crashfact.csv" AS row CREATE (c:CrashFact) SET c = row {vehicleKey:row.VehicleKey, personKey:row.PersonKey, crashConditionsKey:row.CrashConditionsKey, caseKey:row.CaseKey, carCrashDateKey:row.CarCrashDateKey, timeOfDay:row.TimeOfDay} return c

* **CONSTRAINTS**

CREATE CONSTRAINT ON (d:Driver) ASSERT d.personKey IS UNIQUE

CREATE CONSTRAINT ON (v:Vehicle) ASSERT v.vehicleKey IS UNIQUE

CREATE CONSTRAINT ON (c:CaseDetails) ASSERT c.caseKey IS UNIQUE

CREATE CONSTRAINT ON (c:CrashConditions) ASSERT c.conditionKey IS UNIQUE

CREATE CONSTRAINT ON (d:Date) ASSERT d.dateKey IS UNIQUE

* **RELATIONSHIPS**

MATCH (v:Vehicle), (f:CrashFact) WHERE v.vehicleKey = f.vehicleKey CREATE (v)-[r:vehicleCrashed]->(f) return v, f

MATCH (d:Driver), (f:CrashFact) WHERE d.personKey = f.personKey CREATE (d)-[r:personInvolved]->(f) return d, f

MATCH (c:CaseDetails), (f:CrashFact) WHERE c.caseKey = f.caseKey CREATE (c)-[r:caseReport]->(f) return c, f

MATCH (d:Date), (f:CrashFact) WHERE d.dateKey = f.carCrashDateKey CREATE (d)-[r:crashOccurredOn]->(f) return d, f

* **SELECT QUERIES (to compare with SQL)**

MATCH p = (c:CrashConditions{weather:"RAINING"})-->() RETURN COUNT(c)

MATCH (d:Driver) WHERE d.driverSubstanceAbuse = "ALCOHOL PRESENT" OR d.driverSubstanceAbuse = "ALCOHOL CONTRIBUTED" return d

MATCH (d:Date), (f:CrashFact), (p:Driver) WHERE d.dateKey = f.carCrashDateKey AND f.personKey = p.personKey AND d.date = "01/01/15" RETURN p.personID

MATCH (c:CrashConditions) RETURN DISTINCT c.weather