A Cloud Guru/PluralSight Hands-On Lab:
Apply Appropriate Data Models in Cosmos DB for NoSQL

Imagine you are a data engineer working for a healthcare software company. You have access to two external datasets and then maintain your own, internal dataset:
1) **Providers:** A public dataset with virtually every provider in the United States, with their unique identifier, called an NPI. There are two types of providers: Individual and Organizations/Facilities. This data is updated incrementally on a weekly basis, but you only ingest the monthly, full-set updates.
2) **ADTs:** A subscription to near real-time messages called ADTs, each of which contain information on patient admission, discharge and transfer events. ADTs not only include the patient information, but also the facility that issued the ADT message, and a handful of physicians associated with the event. The events often include admission diagnoses, procedures performed and other health-related details. Every type of transfer can trigger an event; even small transfers from the ER to radiology and back to the ER can result in three ADT messages. So, a single admission to the ER can result in dozens of ADT messages.
3) **Patients:** A private, centralized, database of patients that you maintain for various healthcare related applications.

All of these data sets are currently stored in relational databases, but you plan to migrate them to Cosmos DB for NoSQL, starting with two microservices.

The primary entities involved in these datasets are listed, below, with a few key column names.

Providers dataset entities and a few key data columns:
Provider: NPI*, ProviderType (I for Individual or F for Organization/Facility), Name, PrimaryPhone
Address -- Street, Street2, City, State, Zip
Specialty -- SpecialtyCode, SpecialtyDescription, SubSpecialtyCode1...up to 12, each with a description

ADT dataset entities and a few key data columns:
Message -- UniqueMessageID, Message_DateTimeStamp
Encounter -- EventType (A, D, or T), UniqueEncounterID (there can be multiple messages with this same EncounterID), AdmissionDiagnosis, Encounter_StartDateTimeStamp
Facility NPI*, Name, City
Address -- Street, Street2, City, State, Zip
Patient -- InsuranceID, PrimaryCareProvider_NPI*, Name, DateOfBirth, YearOfBirth
Provider -- NPU*, Name,RoleInADTEvent (attending, surgeon, etc.)
Diagnosis -- DiagCode (1-50 or more)
Procedure -- ProcCode (1-50 or more)

Patients dataset entities:
Address -- Street, City, State, Zip (Up to 5 per patient)
Person – PatientID, InsuranceID, PCP_NPI, Name, PrimaryPhone, Email, DateOfBirth, YearOfBirth
Encounter (a medical visit/event) – UniqueEncounterID, Event_DateTimeStamp, NPI*, reasonForVisit (a diagCode), primaryProcCode

*NPI stands for National Provider Identifier and is widely accepted as a means to uniquely identify specific medical providers and organizations. There are some caveats and complexities in regards to NPI assignment, but for this exercise, assume that it is a unique identifier and that each provider or facility has only one NPI.