Case Study

NewYork–Presbyterian Hospital

# NYP Advances Analysis to Track Infections with Neo4j

**INDUSTRY**
Life Sciences

**USE CASE**
Graph Data Science

**GOAL**
Track the spread of infection in the hospital

**CHALLENGE**
Connecting the when and where of events: spatial data and time series data

**SOLUTION**
Use Neo4j to connect space and time data to track infections

**RESULTS**
- Created an effective way to monitor the spread of infection and address risks
- Designed a reusable network structure that captures the time and place of events
- Used graph algorithms to analyze event data across space and time

*NewYork–Presbyterian Hospital wanted to track the spread of infections in their facilities, but the data types that describe where events happen and when are difficult to blend. Their analytics team used Neo4j to relate all their event data, enabling them to track infections and take strategic action to contain them.*

## The Company

NewYork–Presbyterian Hospital (NYP) is a nonprofit academic medical center in New York City that works in collaboration with Columbia University College of Physicians and Surgeons and Weill Cornell Medical College. NYP is ranked number 5 in the U.S. and number 1 in New York according to *US News and World Report*'s 2019 survey.

## The Challenge

Stopping the spread of infection is a critical concern for hospitals.

Michael Zelenetz, Analytics Project Leader at NewYork–Presbyterian Hospital, set out to devise a way to track the spread of infection. Zelenetz and his team sought to identify connecting points so hospital staff could focus their efforts to reduce or eliminate the infection.

But tracking contagions requires knowing both *where* an infected patient was and *when* the patient was there. "It was important to us to be able to look at events across both time and space," said Zelenetz. "We wanted to be able to ask questions like, 'Who got an infection during their stay?' and 'Who was sharing a room with them at the time they contracted that infection?' We had difficulty modeling the relationship of space and time."

Maps pinpoint where events happened, but not when. A time series captures the sequence of events, but not where they happened.

Tracking infections requires combining this data. "The problem we kept running into was that we wanted to know what happened to patients during their visit and where they were at the time that a certain event happened," said Zelenetz.

## The Solution

Graph technology offers a flexible way to connect all the dimensions of an event: what, when and where it happened.

The team proposed using Neo4j to graph space and time. They wanted to log every event that took place in the hospital – from the time a patient was admitted to all of the tests they undergo and their release from the hospital.

Case Study

"The reason that we started doing this was to be able to look at how infections spread through the hospital. That's why getting time and location was very important to us."

*– Michael Zelenetz, Analytics Project Leader, NewYork–Presbyterian Hospital*

First Zelenetz looked for ways to create a time tree to model time series data in Neo4j. He used code he found in a Neo4j blog post to create a time tree. He then created a location tree to model all the rooms in the facility.

Data modeling is a key part of any Neo4j project. Visualizing their initial model revealed a massive number of relationships, but that alone did not meet their goals. Zelenetz decided to add an event entity to connect the time and location trees.

The event entity includes time, space, place and duration, with timestamps for the beginning and end of an event.

Including an event entity made the data model flexible; the graph they created could serve multiple use cases and ultimately position NYP's analytics team to analyze everything that happened in its facilities.

Zelenetz fed the graph dataset he created into a community detection graph algorithm. The algorithm grouped events into various specialties such as oncology and pediatrics, validating the team's work.

## The Results

NYP achieved its goal of creating a graph-based capability to track the spread of infection across the hospital, using graph data science to pinpoint critical connection points and focus containment efforts.

"The reason that we started doing this was to be able to look at how infections spread through the hospital," said Zelenetz. "That's why getting time and location was very important to us. We had to understand that patient A was in the same room as patient B, at the same time that they contracted an infection. We can look and see which patients were nearby, being cared for by the same provider and who was collocated on the unit. This allowed us to do some more advanced analysis to figure out how diseases move around and who might be at risk."

Even better, using the flexible graph data model the team created, the analytics team is positioned to analyze all sorts of events that happen in the hospital, from tests to surgeries. This moves toward Zelenetz's initial goal: to create a social network of everything that happens in NYP's facilities.

While the NYP graph data model connects events as visits, a possible next step is to map to the level of individual patients so their care can be tracked across its inpatient and outpatient facilities – a common data problem.

"One of the problems that we have is that we have individual patients in multiple systems," said Zelenetz. "They may see an outpatient provider who's affiliated and then have an inpatient visit." Graph analysis at the patient level could track hospital readmissions, a key metric for healthcare.

*This case study is based on a talk by Michael Zelenetz, Analytics Project Leader, NewYork–Presbyterian Hospital, at GraphConnect 2018.*