

## Case Study



## Airbnb

## Master Data Management Empowers a Global Network of Airbnb Employees

**INDUSTRY**

Travel & Hospitality

**USE CASE**

Master Data Management /  
Knowledge Graph

**GOAL**

Create a data resource search engine that's easy to navigate with reliable results

**CHALLENGE**

Employees were shackled by siloed data and tribal knowledge, dampening productivity.

**SOLUTION**

Created the Dataportal using Neo4j to connect disparate data resources.

**RESULTS**

- Airbnb employees have a one-stop shop for every available data resource.
- Neo4j easily integrated with Airbnb's existing tech stack for easy programming

*As Airbnb's workforce rapidly expanded across the globe and their data landscape became increasingly difficult to traverse, productivity was hindered. They sought out Neo4j to assist in creating a fast, user-friendly master data management system known as the Dataportal.*

### The Company

Airbnb connects people to unique travel experiences with an online sharing marketplace for leasing or renting short-term lodging. Using a variety of search filters, a user can easily wade through the site's over 4 million renter lodging listings that span 65,000 cities and 191 countries. For the third quarter of 2018, Airbnb announced revenue of more than \$1 billion and are valued at \$31 billion.

### The Challenge

Once a struggling startup, Airbnb has grown into a household name for the online accommodations marketplace. With the company's success came rapid expansion of their workforce, which currently includes 3,500 employees spread across 20 offices worldwide.

In any large, complex organization, an ever-growing landscape of internal and external data resources – especially when scattered across various platforms – eventually becomes unmanageable and restrictive. After a year at Airbnb, Software Engineer, John Bodley, recognized that Airbnb's data was prohibitively siloed, inaccessible or lacked proper context.

With over 200,000 tables in their main Hive data warehouse spread across multiple clusters, 10,000 Superset charts and dashboards, 6,000 experiments in metrics, over 6,000 Tableau workbooks and charts, and over 1,500 knowledge posts – the vast amounts of wayward data was working against their operational advantage.

Bodley also noticed that employees were relying on tribal knowledge for answers to questions, which ultimately stifled productivity. "We often run an employee survey," he said, "and we consistently scored really poorly around the question: 'The information I need to do my job is easy to find.'"

He knew they needed to democratize data so any employee, regardless of data-literacy level, was empowered to find resources, fully confident the results were relevant and reliable.

### The Strategy

"At a very high level, we just want to search for something," Bodley said, "so how do we frame our data in a meaningful way for searching, ranking and relevance?"

His team set off developing the Dataportal, a self-service, integrated data-space that presents a contextual, holistic view of Airbnb data for employees to navigate their data landscape easily and quickly

Case Study



“Why did we choose Neo4j as our graph database? Our data represents a graph, so it felt logical to use a graph database as well to store the data. It’s nimble. We wanted a really fast, performant system. It’s popular. It’s the world’s number one graph database. And finally, it integrates really well.”

– John Bodley  
Software Engineer, Airbnb

whenever they need access or answers for their daily working needs.

Bodley and his team determined the tool needed four major features: search, context & metadata, employee-centric data and team-centric data. Connecting relationships between each data resource – and their associated metatypes – would be the key to providing necessary data links for a fully functional tool ready for employee consumption.

### The Solution

With various resources (e.g., data tables, dashboards, reports, users, teams, business outcomes, etc.) each featuring levels of context and connections, Bodley and his team quickly realized their entire data ecosystem is best represented as a graph. That led them to the [Neo4j](#) graph database.

“There’s four main reasons,” said Bodley. “One, it kind of felt logical, right? Our data represents a graph, so it felt logical to use a graph database as well to store the data. It’s nimble. We wanted a really fast, performant system. It’s popular, right? It’s the world’s number one graph database... And finally, it integrates really well.”

In terms of speed, the Dataportal is meant to be a data resource search engine, where fast, detailed and accurate interactions ultimately incentivize exploration. Neo4j offers the fastest way to search through millions of data connections per second.

In terms of integration, Airbnb had their own tech stack in place, including Elasticsearch and Python. “We use Flask as a lightweight Python web framework for the API, which is consistent with a number of open source Airbnb data tools like Airflow, The Knowledge Repository, and Superset,” said Bodley. “The single-page web app leverages React and Redux.”

Neo4j integrated well with all of Airbnb’s preferred programming languages, while also allowing them to enrich search rankings by taking advantage of the graph topology. Everyday they push data from Hive into the Neo4j [graph database](#) – connecting their siloed data from a relationships perspective – to facilitate quick, highly relevant contextual search results.

### The Results

With Neo4j, Airbnb was able to tie together their entire data ecosystem and make it searchable, relevant and trustworthy for even the newest, most clueless employee.

Instead of far-flung staffers relying on tribal knowledge, which creates breakdowns in the production of quality work, the Dataportal is Airbnb’s one-stop resource for finding all relevant data, especially in terms of employee- and team-centric information critical to daily performance.

And because Neo4j features high scalability, the Dataportal is also poised to facilitate future company growth, instantly connecting new hires and new projects in real time.

Neo4j is the leader in graph database technology. As the world’s most widely deployed graph database, we help global brands – including [Comcast](#), [NASA](#), [UBS](#), and [Volvo Cars](#) – to reveal and predict how people, processes and systems are interrelated.

Using this relationships-first approach, applications built with Neo4j tackle connected data challenges such as [analytics and artificial intelligence](#), [fraud detection](#), [real-time recommendations](#), and [knowledge graphs](#). Find out more at [neo4j.com](#).

Questions about Neo4j?

Contact us around the globe:  
[info@neo4j.com](mailto:info@neo4j.com)  
[neo4j.com/contact-us](https://neo4j.com/contact-us)