

Identify Multi-level Hierarchies and Tree Structure of Group Chain Companies using Graph data and Relational Databases

Bijen Shah¹, Jack Lewis²

¹University of Westminster, ² The Data City

Background and Motivation

This industry project was completed in collaboration with The Data City. The company's mission is to map the UK's emerging economy, providing researchers, policy makers and investors with real time data on dynamic sectors and the companies within them. The company has a product which allows users to view information like industry, financials, employment, growth, etc. about companies. The companies listed in different jurisdictions throughout the country lacked information about company group structure unless they want the public to know. The Data City's product classifies these companies based on their properties into Real Time Industrial Classifications (RTICs) and their team wants to group companies based on their beneficiary company and group structures. They only have limited information on group structure and struggle with cascading a whole group structure into one parent company.

Data

- Open Ownership data of beneficiaries and company ownerships is sourced from UK's Companies House government agency.
- Used to Identify Multi-level Hierarchies and Tree Structure of Group Chain Companies.
- Uncover and Estimate 400K+ group of companies and their ultimate parent company using Open Ownership Big Data.
- Establish a robust ETL infrastructure to graph data into 1 hierarchical table using SQLite, DB Browser and Python Pandas.

Method

- Finding multi-level hierarchies and company group structures using graph data on a relational database like SQLite.
- Recursive querying on data stored.
- Storing data in a flat table with multi row-level relations between entities and persons defined in the data.
- Estimating company relationships and groups based on their other company listing details where beneficiaries are not explicitly defined.

Key Findings

The project has the following objectives and outcomes:

- Identify group of companies and their ultimate parent company (entities beneficiaries).
- Estimate groups of companies based on company details for person beneficiaries.
- Use the insights from The Data City's RTICs for bucketing companies.
- Build an infrastructure to simplify graph data into a table on a relational database.
- Orchestration pipeline to be built for the data transformations.
- Develop a process to use this information on a list for user acceptance testing and validation of bucketing of companies.

Value of the research

- The potential use of the company structures information is to visually help the users to create new journeys when they are using The Data City product.
- Allow the users to estimate the company's financials, employee count, growth, etc.
- Understand how future acquisitions are impacting the companies' details and who is benefiting from the business.
- Create new marketing strategies to cater to this new finding.
- I am proud that the learnings from this project were one of the many motivations of The Data City's global product which will be used for global comparisons of classifications of businesses.

Fig 1: The Data City Global Product

Fig 2: Companies in group chains starting from U - Ultimate Parent through P - Parent Companies and to C - Child Companies

