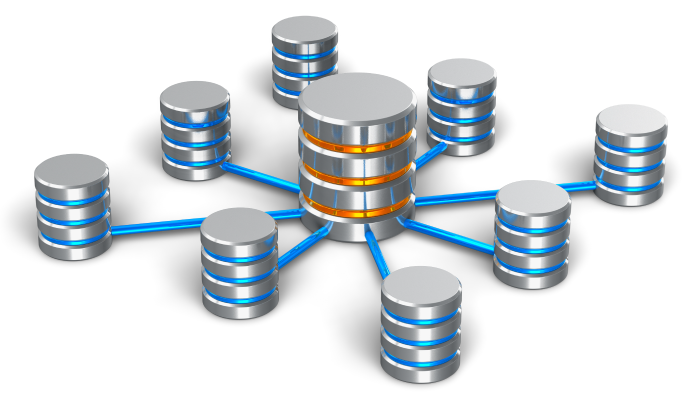
Bing, Elsa, Frederik, Jonathan

**NHL ETL Project**

**Final Report**

Extraction



We extracted a dataset on NHL player data from Kaggle. This dataset was in the form of a CSV (<https://www.kaggle.com/martinellis/nhl-game-data#player_info.csv>) listing 2,386 players. In addition to our Kaggle dataset, we scraped the headlines of NHL news articles published on Reuters website (<https://www.reuters.com/news/archive/icehockey-nhl?view=page&page=4&pageSize=10> ) querying close to 5,000 headlines.

Transform

Given the available dataset, no cleaning was required. We looped through our scraped headlines and counted how many articles featured each player of our dataset. After we found this number, we joined them in one dataset.

Load

With our data, we created a MySQL database. We did this as our data collected is primarily structured and relational (headlines related to the player names). We stored the following tables:

* Players;
* Headlines;
* Number of headlines per player.