Data mining consists of five major elements:

* Extract, transform, and load transaction data onto the data warehouse system.
* Store and manage the data in a multidimensional database system.
* Provide data access to business analysts and information technology professionals.
* Analyze the data by application software.
* Present the data in a useful format, such as a graph or table.
* **Size of the database**: the more data being processed and maintained, the more powerful the system required.
* **Query complexity**: the more complex the queries and the greater the number of queries being processed, the more powerful the system required.

Tweets may include one or more entities in their 140 characters of content and reference one or more places that map to locations in the real world.

Tweet entities are essentially the user mentions, hashtags, URLs, and media that may be associated with a tweet, and places are locations in the real world that may be attached to a tweet. Note that a place may be the actual location in which a tweet was authored, but it might also be a reference to the place described in a tweet.

Possible languages I could work with –

C++ – https://github.com/swatkat/twitcurl

Node.js – https://github.com/aivis/user-stream

Java – https://github.com/twitter/hbc

For simplicity of development, the key pieces of information that you'll need to take away from your newly created application's settings are its consumer key, consumer secret, access token, and access token secret.