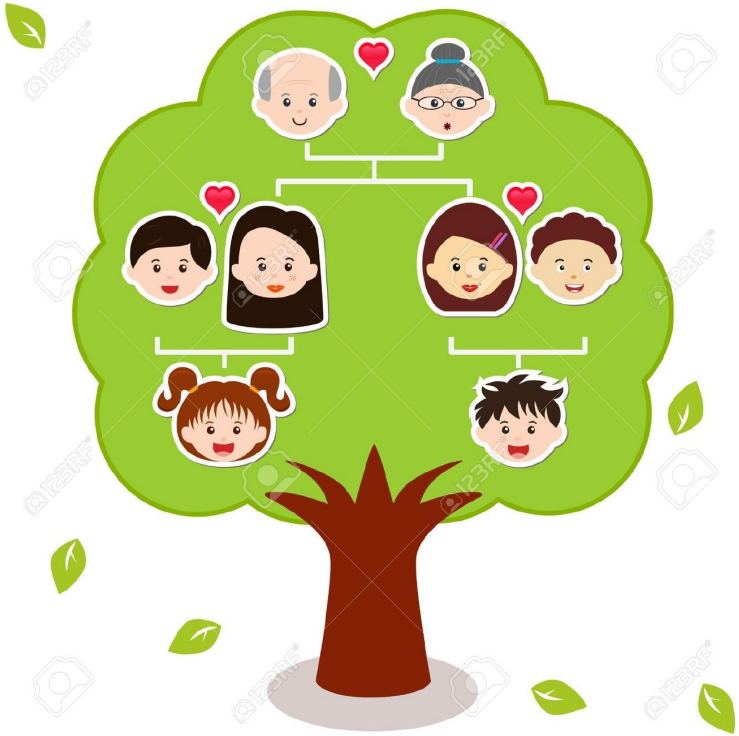
Big Taco Mining

Consulting Services for FamilySearch

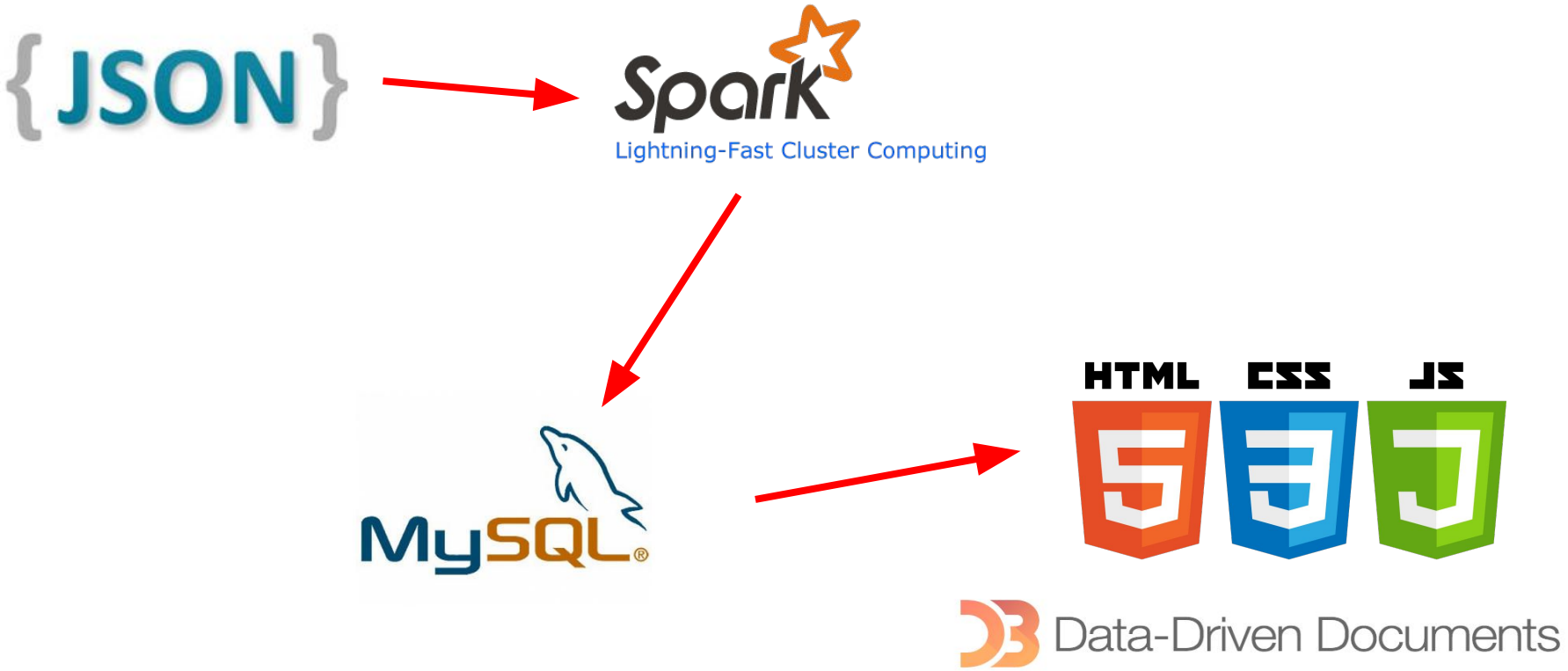
FamilySearch



Introduction to Project

- determine what percentage of an area's population at any given period of time does FamilySearch has record of.

Final Approach



Date - Different Formats

- About 1887
- June 1998
- 01, January, 1943
- 09-12-1967
- 01-02-90

Place - Different Names

- Sometimes there is country
- Sometimes there is only state
- Sometimes there is only city

- USA
- United States
- United States of America
- US
- Brooklyn
- New York
- NY

Place - Different Spellings

- Translations
- Unicode

México

VS

Mexico

Seoul

VS

Seul

Place - Different Languages

- Where was the record entered?

Spain

VS

Espana

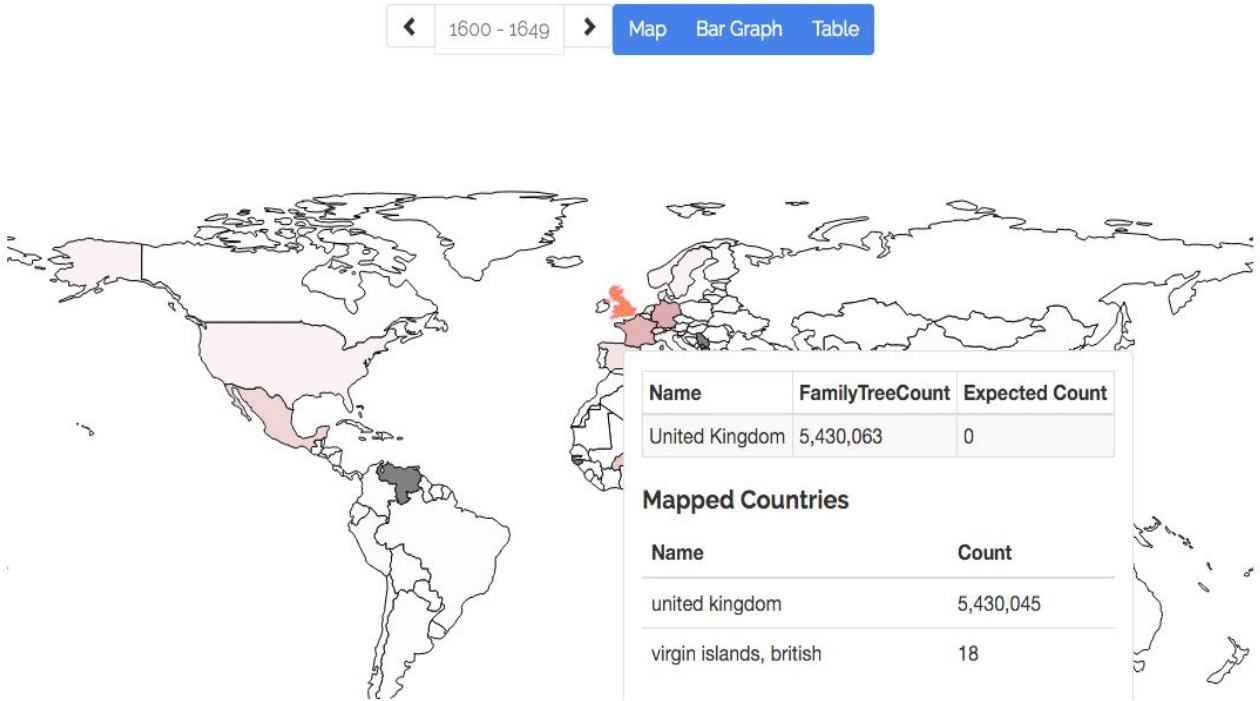
European Countries like “Belgium”

Dutch name	French name
Aalst	*Alost*
Aarlen	Arlon
Aat	Ath
Antwerpen	*Anvers*
Bastenaken	Bastogne
Bergen	Mons
Borgworm	Waremme

Solution

- Look for substrings

- Country Name
- States Names
- 2 characters Abbreviations US, MX
- 3 characters Abbreviations USA, MEX

Final Visualization

https://big-data-family-tree.herokuapp.com/