

Web APIs:

1. Stripe API: <https://stripe.com/docs/api/>
2. GitHub API: <https://docs.github.com/en/rest>
3. National Weather Service (NWS) Weather API <https://www.weather.gov/documentation/services-web-api>
4. Champaign-Urbana Mass Transit District (CUMTD) API <https://developer.cumtd.com/>
5. Queue@Illinois API <https://github.com/illinois/queue/blob/master/doc/API.md>

Common Features in Best in Class API design:

- 1.
- 2.
- 3.
- 4.
- 5.

Data Stores -- (SQL, NoSQL/Mongo/Redis):

- Optimized for data storage and lookup.
- Focus on queries for data:
  - SELECT \* FROM table WHERE date="2021-11-26";
  - find\_one( {date: {\$eq: "2021-11-26"}} )
  - ...
- Developer is responsible for processing the data.

Problems?

Big Data Processing

- Optimized for data processing on large data sets.
- Queries involve functions that perform logic as part of processing the data.
- Key Technology: \_\_\_\_\_.

MapReduce

- Developed as a research project out of Google.
- OSDI'04: *MapReduce: Simplified Data Processing on Large Clusters*
- **Big Idea:** Create a framework for processing data based on functions that can be "automatically parallelized".
  - Allows many nodes to contribute to processing the data without human design/programming.

<https://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf>



### MapReduce: Map Functions

- Input:
  
- Output:

### Reduce Function:

- Input:
  
- Output:

---

### Example #1: Word Count

<i>The</i>	<i>quick</i>	<i>brown</i>	<i>fox</i>	<i>jumps</i>	<i>over</i>	<i>the</i>	<i>lazy</i>	<i>dog</i>
[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]

Map:

Reduce:

### Example #2: Mutual Friends

Through asking about your friends about their friends, you have identified who are friends of whom ( $\rightarrow$  means “*is friends with*”):

A  $\rightarrow$  B, C  
B  $\rightarrow$  A, C, D  
C  $\rightarrow$  A, B, D  
D  $\rightarrow$  B, C

You want to identify all **mutual friends** to any set of two people. For example: {A, B}  $\rightarrow$  C, D.

Map:

Reduce: