From Nodes to Graphs: A Proof of Concept for Neo4j, MongoDB, and Redis

Abbas Siddiqui & Achyuth Kolluru

Introducing Our Proof of Concept

- We are interested in determining the optimal first store to deliver products to from BART
- We want to determine a path to deliver food to subsequent stores in each county
- We want to run a simulation where we can store a lot of orders from each store every day
- We want to provide many viewers with a live-update map of our deliveries

Overview of Data being used and Preprocessing

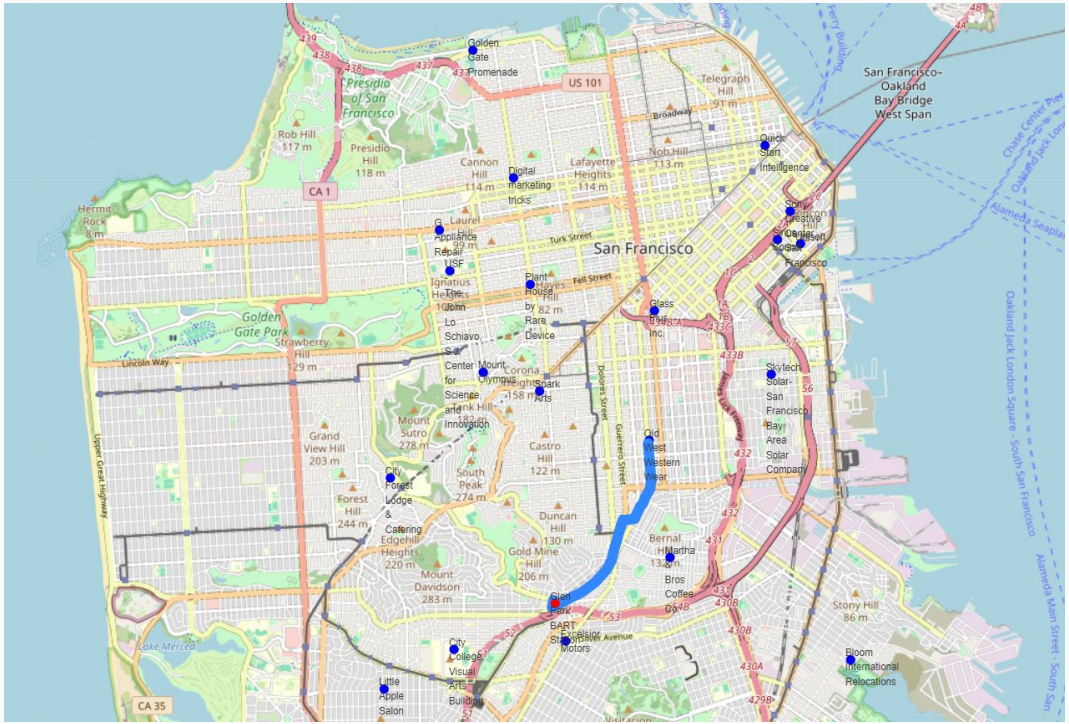
- Utilized FourSquare for sourcing recent data on high-traffic points (colleges, companies, parks, etc.) within each county.
- Identified BART terminals within each county.
- Assigned arbitrary population values to each point.
- Sampled 20 nodes per county.
- Computed inter-node distances and saved the three closest nodes to each other node.

Snapshot into County and Bart Station Tables

	name	county	coordinates	esti. pop.
40	Bus Stop #50773/28-San Leandro Bart	Alameda	(37.695409, -122.077991)	916
41	Logitech	Alameda	(37.552155, -122.063208)	1501
42	SpeeDee Oil Change and Tune-Up	Alameda	(37.708072, -121.924819)	1875
43	Envie Interactive	Alameda	(37.411732, -122.014626)	1185
44	California Express Motors	Alameda	(37.656344, -122.071653)	1725
45	University Union	Alameda	(37.654885, -122.055788)	1977
46	Speedee Oil Change & Tune Up	Alameda	(37.693297, -122.066176)	1816
47	Men's Wearhouse	Alameda	(37.692118, -121.927759)	902
48	Up 2 Speed Sports Therapy	Alameda	(37.698831, -121.917708)	2411
49	SafeAmerica Credit Union	Alameda	(37.695814, -121.932341)	815
50	Red Apple Pizza	Alameda	(37.696829, -122.139635)	884
51	My Union City Plumber Hero	Alameda	(37.581319, -122.017075)	2677
52	Simbol Materials	Alameda	(37.661614, -121.903984)	2481
53	Tint-N-Sound	Alameda	(37.769199, -122.238094)	1431
54	Design Energy Group	Alameda	(37.698278, -121.919474)	1097
55	Hayward BART Garage	Alameda	(37.670639, -122.087372)	1345
56	Phantom Motors	Alameda	(37.678241, -122.09081)	980
57	Randall E. Strauss	Alameda	(37.807695, -122.265135)	2697
58	Emcor Integrated Solutions	Alameda	(37.663021, -121.899895)	1437
59	Pho Play	Alameda	(37.694429, -122.051597)	1435

	name	county	coordinates	esti. pop.
0	West Dublin/Pleasanton BART Station	Contra Costa	(37.699709, -121.928243)	428
1	Montgomery St BART Station	San Francisco	(37.788862, -122.402116)	1812
2	San Antonio BART Station	Alameda	(37.788309, -122.245649)	1489
3	Fremont BART Station	Alameda	(37.557451, -121.976611)	1813
4	Santana Row BART Station	Santa Clara	(37.323154, -121.948811)	965

Overview of Data being used and Preprocessing (Cont.)



- Utilized Folium to render our maps, location points, and paths
- Utilized OSMNX to calculate our shortest paths between points

How we can use Neo4j?

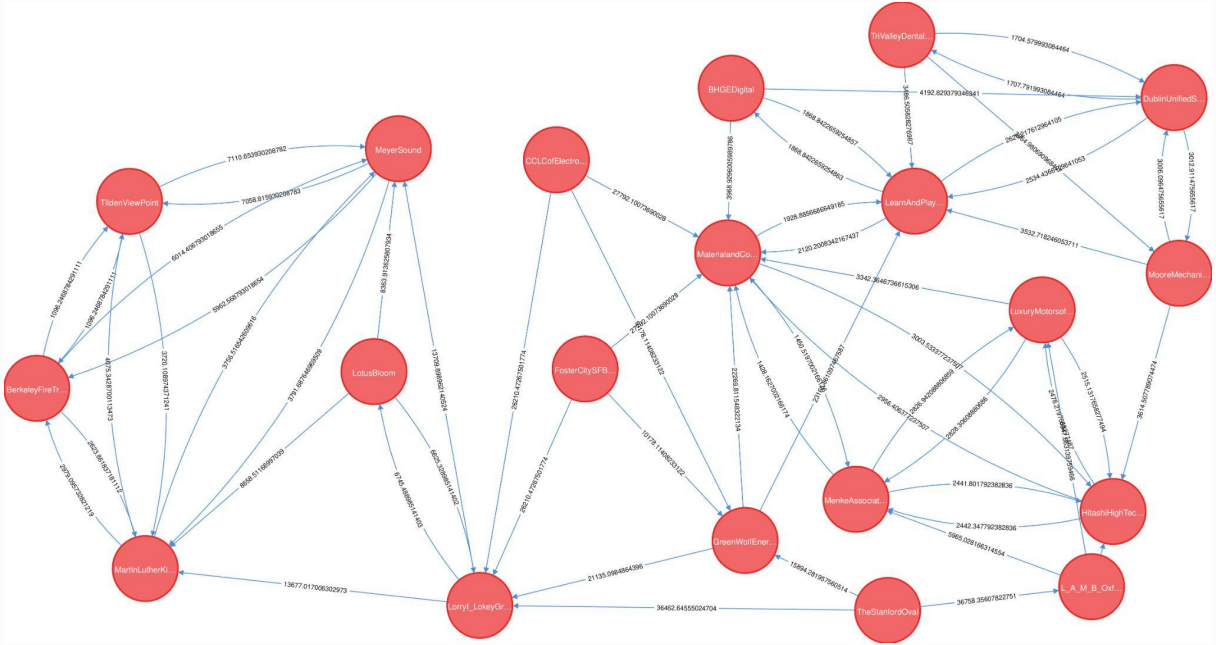
Neo4j

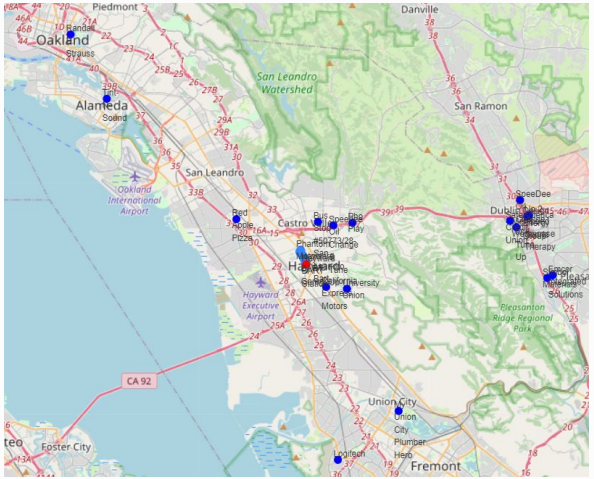
- Ideal for applications with complex relationships that can be represented as graphs.
- Efficient for queries involving like networks or graph algorithms.
- Fast travel of connections between data points.
- Easy to visualize

Why a relational database is not a good fit for this business example:

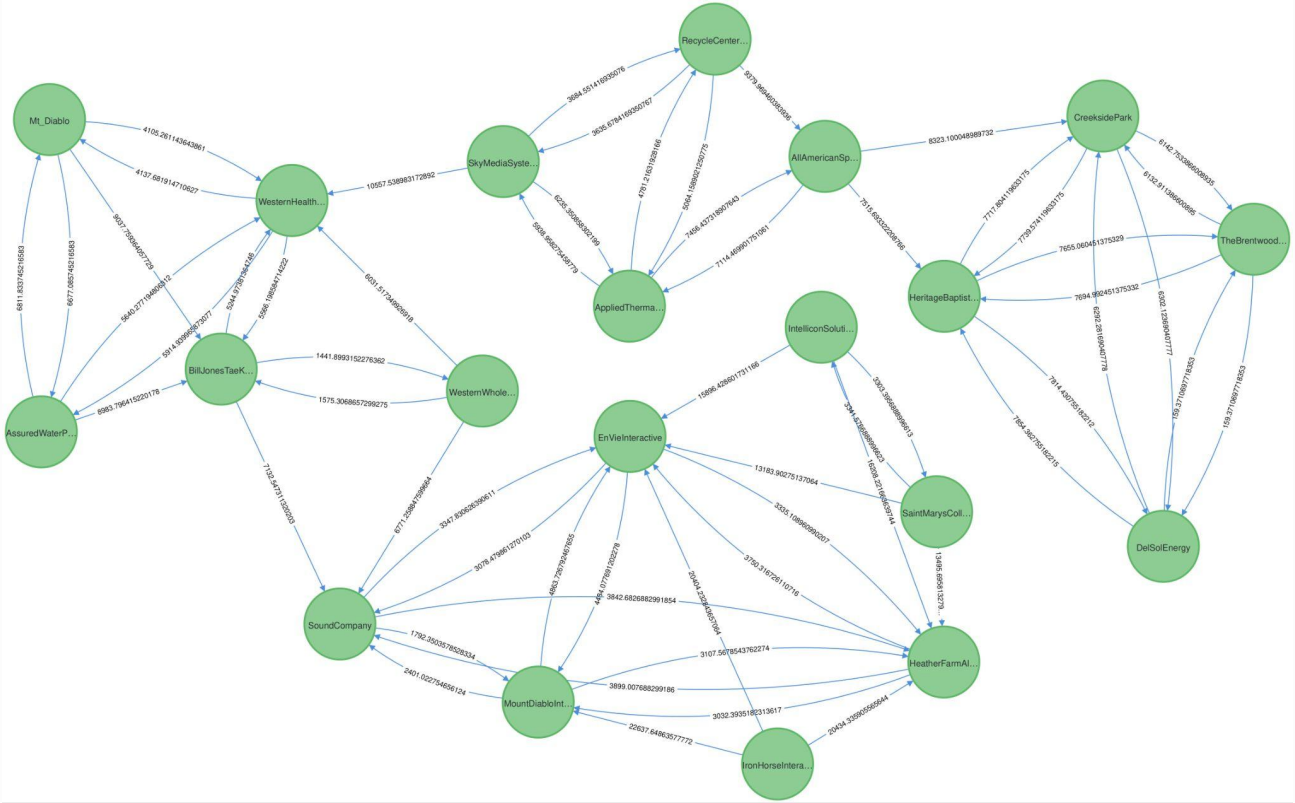
- Complex relationship networks which we want to run graph algorithms on
    - Neo4j can do this for us by default

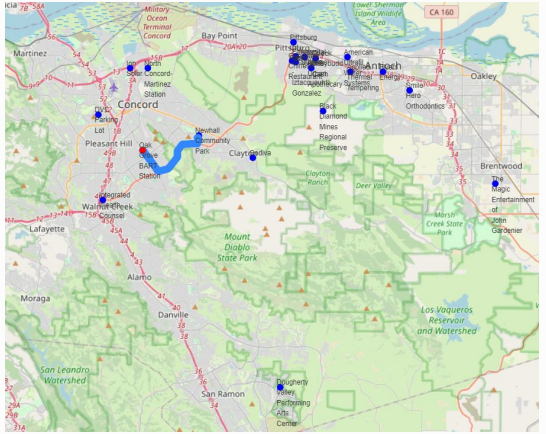
Primary Neo4j Graph (Alameda)



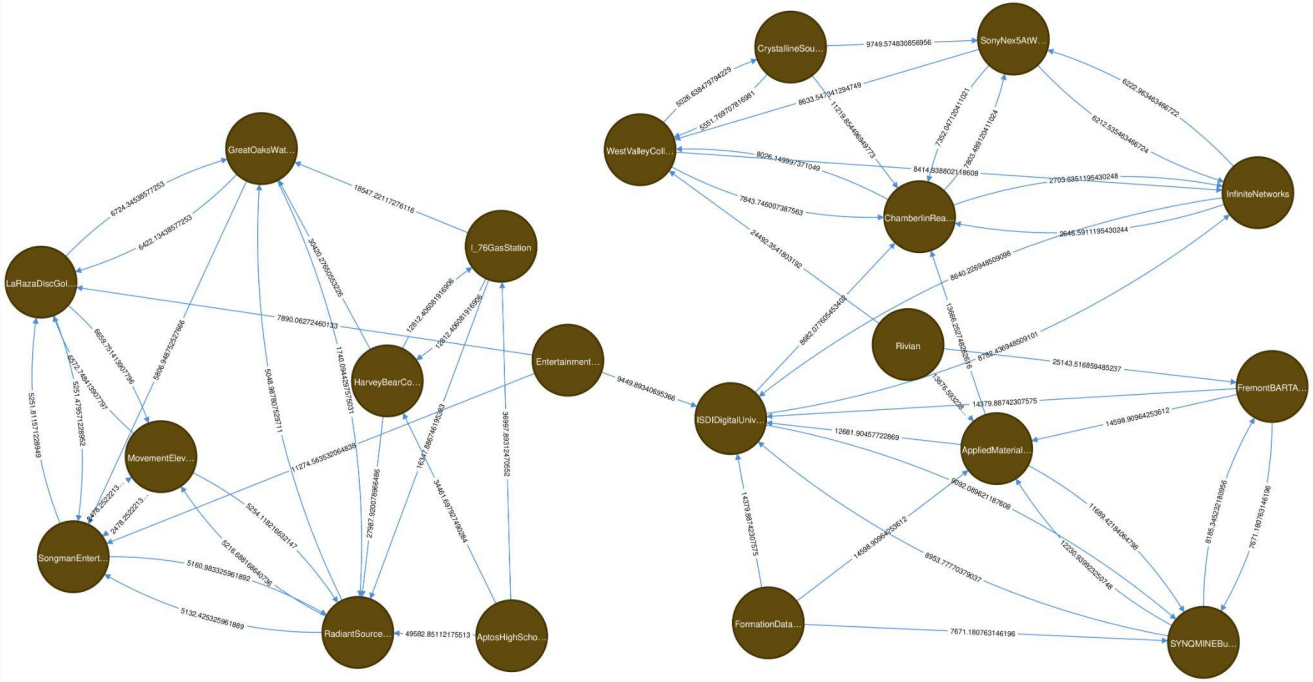


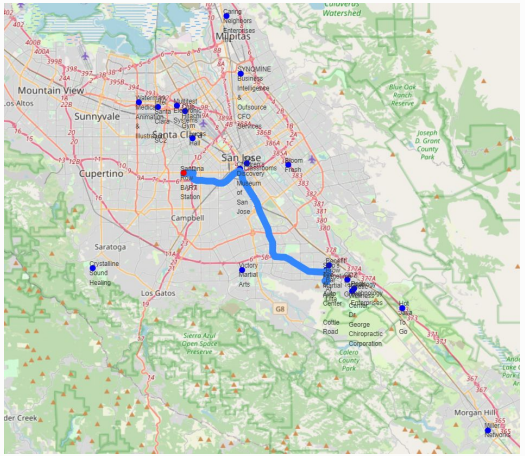
Other Neo4j graphs (Contra Costa)



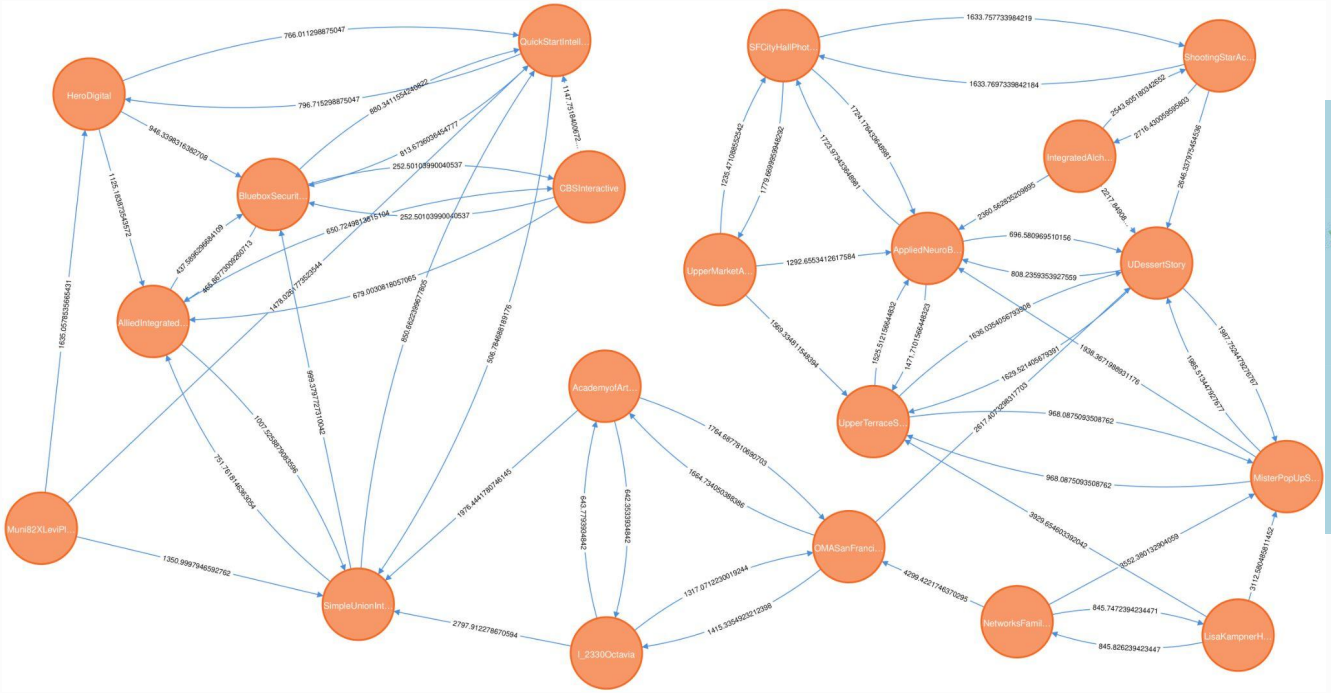


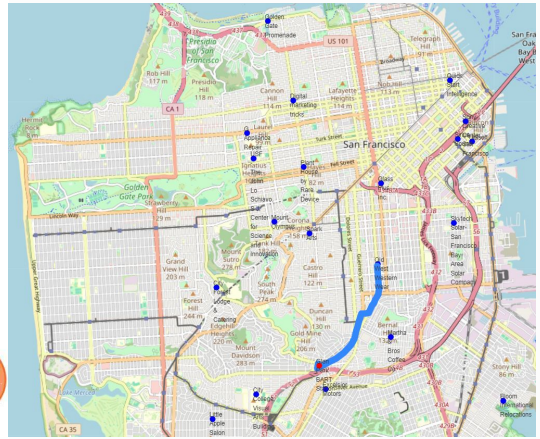
Other Neo4j graphs (Santa Clara)



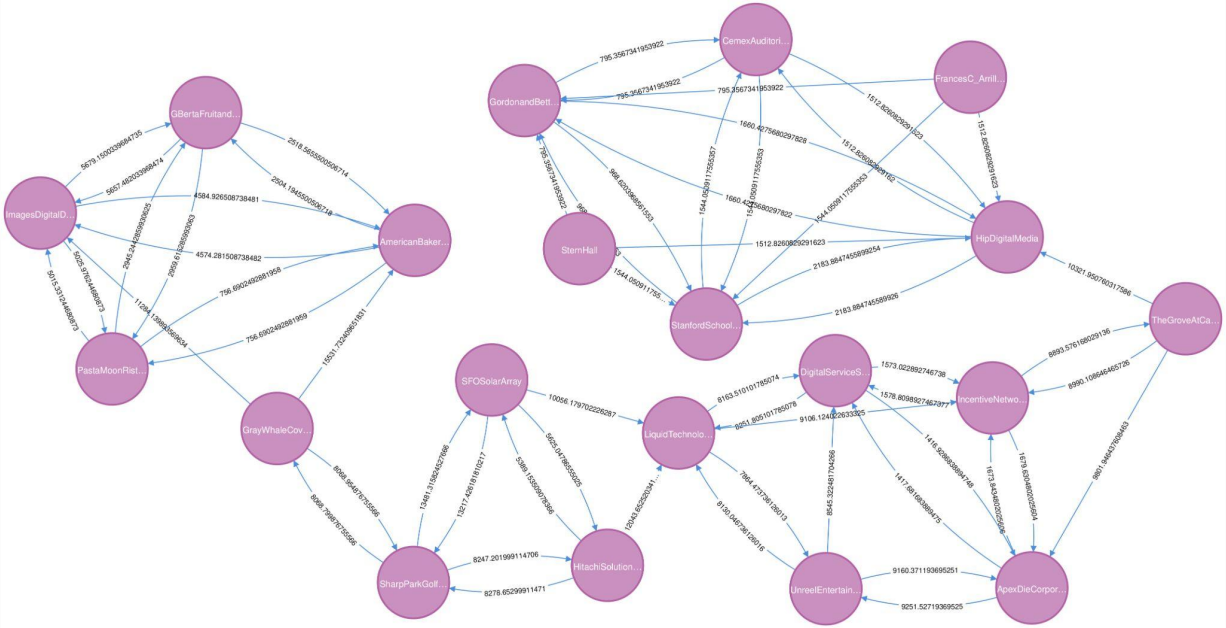


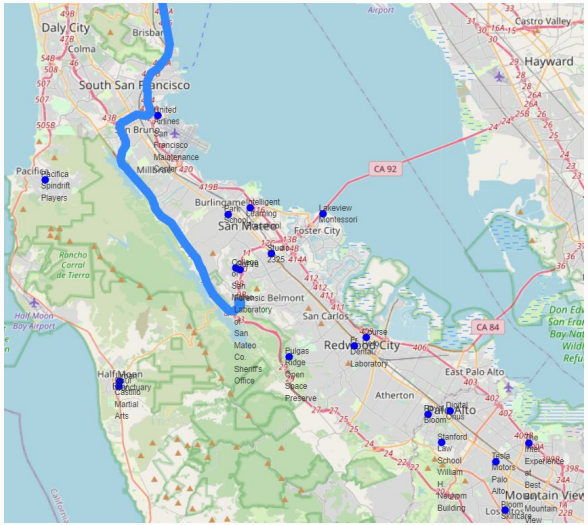
Other Neo4j Graphs (San Francisco)





Other Neo4j Graphs (San Mateo)





What do we use MongoDB for?

MongoDB

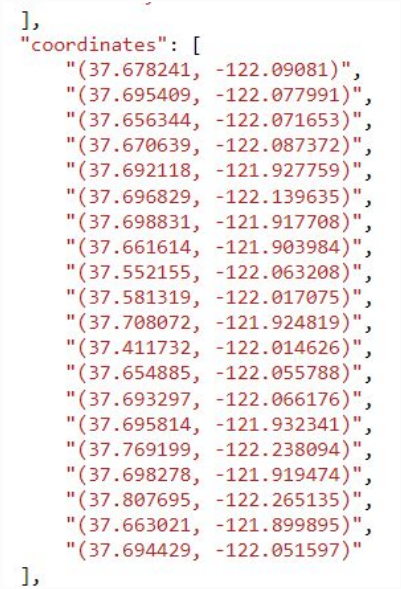
- Can handle both structured and unstructured data.
- Stores and returns data in JSON like formats.
- Useful for running analytics from multiple POVs effectively.
- Scales up better than SQL for very large datasets.
- Able to access certain subset (query) of data quickly
- Low latency refresh rate (ACID vs flexibility)

Why a relational database is not a good fit for this business example:

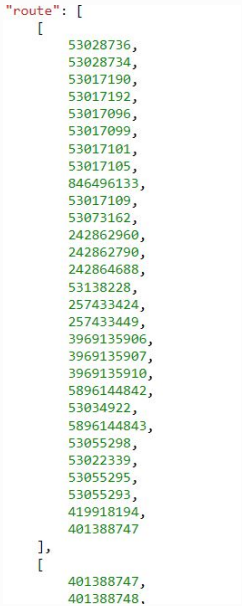
- Lists of varying route node lengths aren't natural fits for SQL structure
- Better for handling and scaling up for large amounts of end-of-day data

MongoDB for Storing Points, Truck Routes, and Item Orders

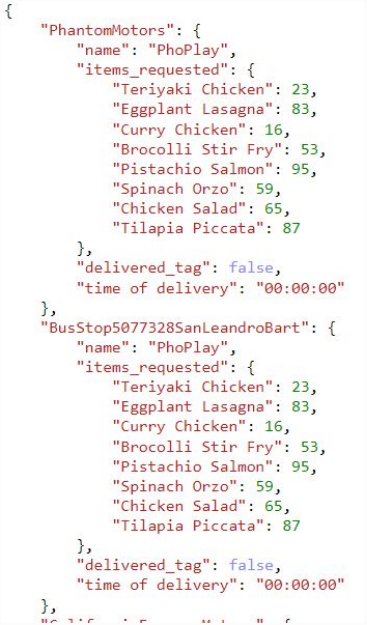
```
{
    "Alameda": {
        "nodes": [
            "PhantomMotors",
            "BusStop5077328SanLeandroBart",
            "CaliforniaExpressMotors",
            "HaywardBARTGarage",
            "MensWearhouse",
            "RedApplePizza",
            "Up2SpeedSportsTherapy",
            "SimbolMaterials",
            "Logitech",
            "MyUnionCityPlumberHero",
            "SpeeDeeOilChangeandTuneUp",
            "EnvieInteractive",
            "UniversityUnion",
            "SpeedeeOilChangeTuneUp",
            "SafeAmericaCreditUnion",
            "TintNSound",
            "DesignEnergyGroup",
            "RandallE_Strauss",
            "EmcorIntegratedSolutions",
            "PhoPlay"
        ],
```



```
],
"coordinates": [
    "(37.678241, -122.09081)",
    "(37.695409, -122.077991)",
    "(37.656344, -122.071653)",
    "(37.670639, -122.087372)",
    "(37.692118, -121.927759)",
    "(37.696829, -122.139635)",
    "(37.698831, -121.917708)",
    "(37.661614, -121.903984)",
    "(37.552155, -122.063208)",
    "(37.581319, -122.017075)",
    "(37.708072, -121.924819)",
    "(37.411732, -122.014626)",
    "(37.654885, -122.055788)",
    "(37.693297, -122.066176)",
    "(37.695814, -121.932341)",
    "(37.769199, -122.238094)",
    "(37.698278, -121.919474)",
    "(37.807695, -122.265135)",
    "(37.663021, -121.899895)",
    "(37.694429, -122.051597)"
],
```



```
"route": [
    [
        53028736,
        53028734,
        53017190,
        53017192,
        53017096,
        53017099,
        53017101,
        53017105,
        846496133,
        53017109,
        53073162,
        242862960,
        242862790,
        242864688,
        53138228,
        257433424,
        257433449,
        3969135906,
        3969135907,
        3969135910,
        5896144842,
        53034922,
        5896144843,
        53055298,
        53022339,
        53055295,
        53055293,
        419918194,
        401388747
    ],
    [
        401388747,
        401388748,
```



```
{
    "PhantomMotors": {
        "name": "PhoPlay",
        "items_requested": {
            "Teriyaki Chicken": 23,
            "Eggplant Lasagna": 83,
            "Curry Chicken": 16,
            "Brocolli Stir Fry": 53,
            "Pistachio Salmon": 95,
            "Spinach Orzo": 59,
            "Chicken Salad": 65,
            "Tilapia Piccata": 87
        },
        "delivered_tag": false,
        "time of delivery": "00:00:00"
    },
    "BusStop5077328SanLeandroBart": {
        "name": "PhoPlay",
        "items_requested": {
            "Teriyaki Chicken": 23,
            "Eggplant Lasagna": 83,
            "Curry Chicken": 16,
            "Brocolli Stir Fry": 53,
            "Pistachio Salmon": 95,
            "Spinach Orzo": 59,
            "Chicken Salad": 65,
            "Tilapia Piccata": 87
        },
        "delivered_tag": false,
        "time of delivery": "00:00:00"
    },
```

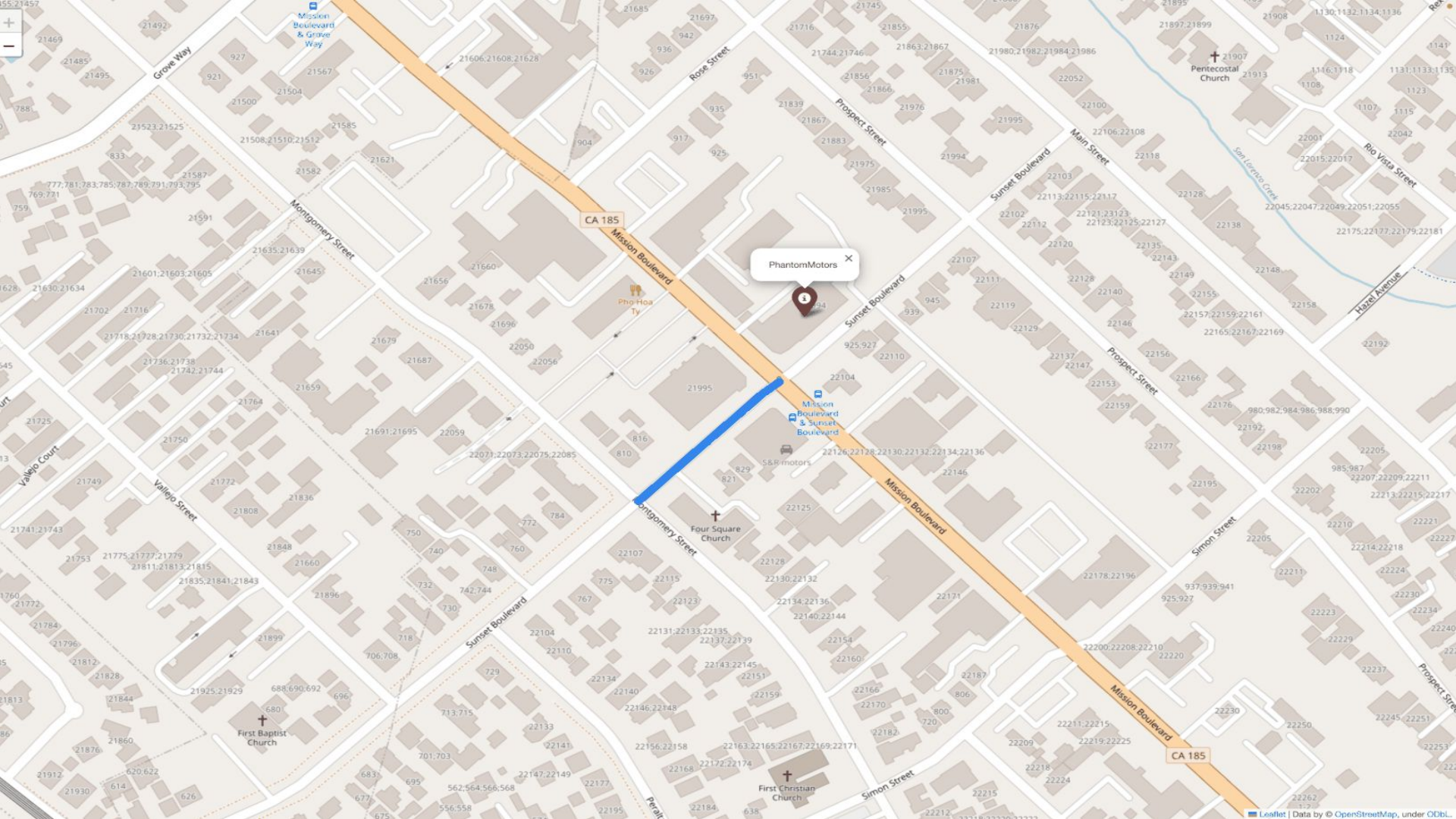
What do we use Redis for?

- Redis is a data storage solution optimized for real-time data
- Why redis over a relational database?
    - For real time updates like delivery and location updates, Redis excels due to its speed and ability to handle high-throughput data.
- Examples:
    - Delivery Updates: Instantly track and update delivery statuses.
    - Location Updates: Real time monitoring of users locations

redis

Redis for Simulated Real Time Updates

- Our System simulates real-time updates with:
    - Realistic variation of wait times
    - Tracking of items being offloaded at each store
- How it works
    - Delivery information is pushed to redis
    - Updates are instantly reflected in our simulation


PhantomMotors
CA 185
Mission Boulevard
Sunset Boulevard
Prospect Street
Rose Street
Main Street
Grove Way
Montgomery Street
Vallejo Court
Vallejo Street
Simon Street
Hazel Avenue
Rio Vista Street
Pho Hoa Ty
Four Square Church
First Baptist Church
First Christian Church
Pentecostal Church
S&R motors
Leaflet | Data by © OpenStreetMap, under ODbL.

Thanks for your time!

Questions?