Data Structures: A Discussion of Scalability and Storytelling

Yujun Jiang

Data Structures

Prof. Aaron Hill

December 16, 2019

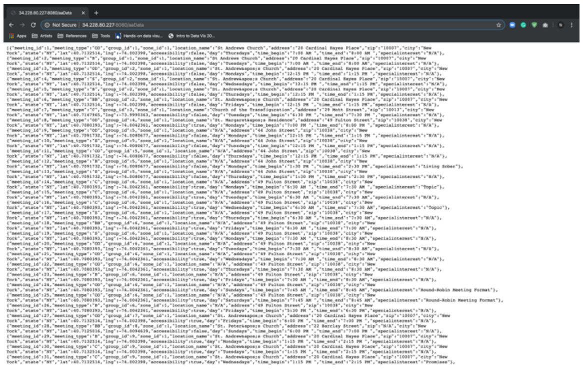
In computer science, data structure is a term about how people organize, manage, and store data to make them can be accessed or modified in a more efficient path. With the arrival of big data, people start to consider some questions such us: how to parse raw data and maintain a good performance in the whole system? Or, what is the best way to organize data with a rational assumption before we create a visualization? Therefore, I will discuss both data scalability and storytelling in here.

As we know, scalability refers to the capability of a system to handle an increasing amount of data. In the first chapter of "Designing Data-Intensive Applications", Kleppmann has asked us a question: "When you increase a load parameter and keep the system resources (CPU, memory, network bandwidth, etc.) unchanged, how is the performance of your system affected?" [1] Indeed, the answer is obvious, but I think the goal of this question is trying to guide us to find a relationship between system performance and users. Also, Kleppmann used an example of Amazon to explain how customers suffer latency when they make many purchases. Even though it only affects 1 in 1,000 requests, a huge number of users still can make complains about the delay. There are thousands of companies might have the same experience with Amazon, and it could bring various economic loses to them. Therefore, to find a solution to maintain the system performance becomes the primary task to us.

In the 2017, Microsoft has announced a new collaboration called Project Silica with Azure Datacenter. According to Jennifer Langston, "Microsoft and Warner Bros. have collaborated to successfully store and retrieve 1978 iconic 'Superman' movie on a piece of glass

[1] Martin Kleppmann. "Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems," *O'Reilly Media*, Book (2017). Page 10-18

roughly the size of a drink coaster, 75 by 75 by 2 millimeters thick."[2] Without a doubt, this new technology has improved the performance (specially on speed) to write and read data. Not only this, data can be saved perfectly in glass over a thousand year. Base on the research, Project Silica uses the feature of light to bring two more dimensions (polarized light and wavelength) in data saving rather than the existing 3D theory (X, Y, and Z axis) by magnetic materials. As you can see, Project Silica seems like a solution to maintain the system performance when the amount of data is increased. At this moment, it also makes another question: how can we afford a high cost of this new technology? In Kleppmann's book, he has defined two theories: move data to a more power machine (vertical scaling) and disturb the load across multiple smaller machines (horizontal scaling).[3] To be honest, either theory cannot be a perfect solution to deal with the increasing of data or load parameter and the cost of new technology. So, we have to blend both of them to create a pragmatic approach.



```
[{"meeting_id":1,"meeting_type":"OD","group_id":1,"zone_id":1,"location_name":"St Andrews Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Thursdays","time_begin":"7:00 AM ","time_end":"8:00 AM ","specialinterest":"N/A"},
{"meeting_id":2,"meeting_type":"B","group_id":1,"zone_id":1,"location_name":"St Andrews Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Tuesdays","time_begin":"7:00 AM ","time_end":"8:00 AM","specialinterest":"N/A"},
{"meeting_id":3,"meeting_type":"OD","group_id":2,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Mondays","time_begin":"12:15 PM ","time_end":"1:15 PM ","specialinterest":"N/A"},
{"meeting_id":4,"meeting_type":"S","group_id":2,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Wednesdays","time_begin":"12:15 PM ","time_end":"1:15 PM","specialinterest":"N/A"},
{"meeting_id":5,"meeting_type":"B","group_id":2,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Thursdays","time_begin":"12:15 PM ","time_end":"1:15 PM","specialinterest":"N/A"},
{"meeting_id":6,"meeting_type":"BB","group_id":2,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":false,"day":"Fridays","time_begin":"12:15 PM ","time_end":"1:15 PM","specialinterest":"N/A"},
{"meeting_id":7,"meeting_type":"C","group_id":3,"zone_id":1,"location_name":"Church of the Transfiguration","address":"29 Mott Street","zip":"10013","city":"New
York","state":"NY","lat":40.7147965,"lng":-73.9990363,"accessibility":false,"day":"Tuesdays","time_begin":"6:30 PM ","time_end":"7:30 PM ","specialinterest":"N/A"},
{"meeting_id":8,"meeting_type":"OD","group_id":4,"zone_id":1,"location_name":"St. Margaret&apos;s Residence","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Mondays","time_begin":"7:00 PM ","time_end":"8:00 PM ","specialinterest":"N/A"},
{"meeting_id":9,"meeting_type":"OD","group_id":5,"zone_id":1,"location_name":"N/A","address":"44 John Street","zip":"10038","city":"New
York","state":"NY","lat":40.7091732,"lng":-74.0080677,"accessibility":false,"day":"Mondays","time_begin":"12:15 PM ","time_end":"1:15 PM ","specialinterest":"N/A"},
{"meeting_id":10,"meeting_type":"S","group_id":5,"zone_id":1,"location_name":"N/A","address":"44 John Street","zip":"10038","city":"New
York","state":"NY","lat":40.7091732,"lng":-74.0080677,"accessibility":false,"day":"Tuesdays","time_begin":"12:15 PM ","time_end":"1:15 PM","specialinterest":"N/A"},
{"meeting_id":11,"meeting_type":"OD","group_id":5,"zone_id":1,"location_name":"N/A","address":"44 John Street","zip":"10038","city":"New
York","state":"NY","lat":40.7091732,"lng":-74.0080677,"accessibility":false,"day":"Thursdays","time_begin":"12:15 PM ","time_end":"1:15 PM","specialinterest":"N/A"},
{"meeting_id":12,"meeting_type":"B","group_id":5,"zone_id":1,"location_name":"N/A","address":"44 John Street","zip":"10038","city":"New
York","state":"NY","lat":40.7091732,"lng":-74.0080677,"accessibility":false,"day":"Mondays","time_begin":"1:30 PM ","time_end":"2:30 PM","specialinterest":"Living Sober"},
{"meeting_id":13,"meeting_type":"S","group_id":5,"zone_id":1,"location_name":"N/A","address":"44 John Street","zip":"10038","city":"New
York","state":"NY","lat":40.7091732,"lng":-74.0080677,"accessibility":false,"day":"Thursdays","time_begin":"1:30 PM ","time_end":"2:30 PM","specialinterest":"N/A"},
{"meeting_id":14,"meeting_type":"C","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Mondays","time_begin":"6:30 AM ","time_end":"7:30 AM ","specialinterest":"Topic"},
{"meeting_id":15,"meeting_type":"C","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Tuesdays","time_begin":"6:30 AM ","time_end":"7:30 AM","specialinterest":"N/A"},
{"meeting_id":16,"meeting_type":"C","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Wednesdays","time_begin":"6:30 AM ","time_end":"7:30 AM","specialinterest":"Topic"},
{"meeting_id":17,"meeting_type":"S","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Thursdays","time_begin":"6:30 AM ","time_end":"7:30 AM","specialinterest":"N/A"},
{"meeting_id":18,"meeting_type":"BB","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Fridays","time_begin":"6:30 AM ","time_end":"7:30 AM","specialinterest":"N/A"},
{"meeting_id":19,"meeting_type":"S","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Mondays","time_begin":"7:30 AM ","time_end":"8:30 AM","specialinterest":"N/A"},
{"meeting_id":20,"meeting_type":"OD","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Tuesdays","time_begin":"7:30 AM ","time_end":"8:30 AM","specialinterest":"N/A"},
{"meeting_id":21,"meeting_type":"OD","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Wednesdays","time_begin":"7:30 AM ","time_end":"8:30 AM","specialinterest":"N/A"},
{"meeting_id":22,"meeting_type":"OD","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Thursdays","time_begin":"7:30 AM ","time_end":"8:30 AM","specialinterest":"N/A"},
{"meeting_id":23,"meeting_type":"B","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Fridays","time_begin":"7:30 AM ","time_end":"8:30 AM","specialinterest":"N/A"},
{"meeting_id":24,"meeting_type":"OD","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Sundays","time_begin":"7:45 AM ","time_end":"8:45 AM","specialinterest":"Round-Robin Meeting Format"},
{"meeting_id":25,"meeting_type":"OD","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Saturdays","time_begin":"7:45 AM ","time_end":"8:45 AM","specialinterest":"Round-Robin Meeting Format"},
{"meeting_id":26,"meeting_type":"B","group_id":6,"zone_id":1,"location_name":"N/A","address":"49 Fulton Street","zip":"10038","city":"New
York","state":"NY","lat":40.7080393,"lng":-74.0042361,"accessibility":true,"day":"Fridays","time_begin":"5:30 PM ","time_end":"6:30 PM","specialinterest":"N/A"},
{"meeting_id":27,"meeting_type":"OD","group_id":7,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":true,"day":"Fridays","time_begin":"6:00 PM ","time_end":"7:00 PM ","specialinterest":"N/A"},
{"meeting_id":28,"meeting_type":"BB","group_id":8,"zone_id":1,"location_name":"St. Peter&apos;s Church","address":"22 Barclay Street","zip":"N/A","city":"New
York","state":"NY","lat":40.7125016,"lng":-74.0094639,"accessibility":false,"day":"Sundays","time_begin":"6:00 PM ","time_end":"7:00 PM ","specialinterest":"N/A"},
{"meeting_id":29,"meeting_type":"B","group_id":9,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":true,"day":"Mondays","time_begin":"1:15 PM ","time_end":"2:15 PM ","specialinterest":"N/A"},
{"meeting_id":30,"meeting_type":"C","group_id":9,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":true,"day":"Tuesdays","time_begin":"1:15 PM ","time_end":"2:15 PM","specialinterest":"N/A"},
{"meeting_id":31,"meeting_type":"C","group_id":9,"zone_id":1,"location_name":"St. Andrew&apos;s Church","address":"20 Cardinal Hayes Place","zip":"10007","city":"New
York","state":"NY","lat":40.7132514,"lng":-74.002398,"accessibility":true,"day":"Wednesdays","time_begin":"1:15 PM ","time_end":"2:15 PM","specialinterest":"Promises"},
```

Fig. 1

In my recent project, I have analyzed and reorganized the data of alcoholic anonymous meetings in Manhattan. As my assumption for the users, I was focusing on two key values: time

[2] Jennifer Langston. "Project Silica: Proof of Concept Stores Warner Bros. 'Superman' Movie on Quartz Glass," *Microsoft*, Innovation Stories (November 4, 2019). https://news.microsoft.com/innovation-stories/ignite-project-silica-superman/

[3] Martin Kleppmann. "Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems," *O'Reilly Media*, Book (2017). Page 10-18

and location. The goal is to help people locate a best meeting efficiently with the fast-paced life in New York (Fig. 1). The whole concept targets to make a data of storytelling which are connect with people, try to convey one idea, keep it simple, and explore the thing I know best. As you can see, all the values are grouped by different categories to avoid redundant query. The location (same with time) has been parsed and reorganized into location_name, address, zip, city, state, lat (latitude), and lng (longitude). As the result, the users can find a matched meeting base on these values and the qualifiers that they know.

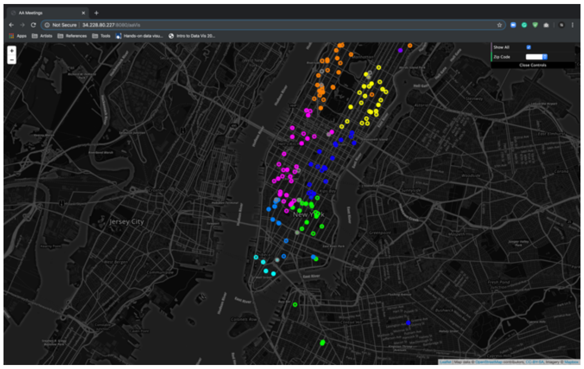


Fig. 2

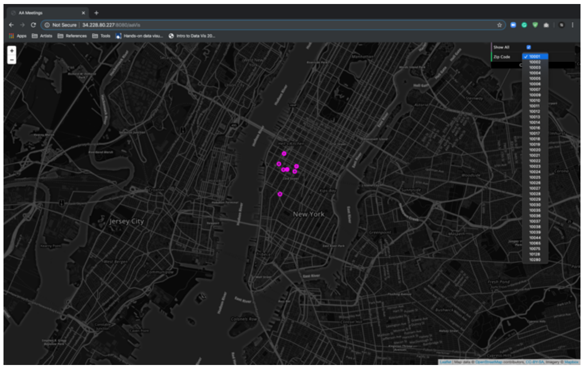


Fig. 3

Last summer, I went to a conference hosted by UXPA in Boston for some inspiration of the user experience in today’s life. I still remember there was a panel about how to use data visualization to build a connection between scientist, designer, and user. As Thomas Watkins mentions during the conference, the best option to find a connection is to adhere the idea of storytelling in the font-end (e.g. user interface). For example, I organized all the meetings by zip code, and each color represents several zip codes that have been grouped into area (Fig. 2 and 3). The reason is I want to users can find nearby meetings in the same area because the traffic is unpredictable in Manhattan. Meanwhile, the zip code can assist users to contact each meeting location by mail regarding to there is no phone number in the raw data. In the discussion of aesthetics by Robert Simmon, “Follow good design practice as well as good visualization practice when developing imagery. In addition to color, consider the other aspects of design: typography, line shape, alignment, etc. Be aware of the media you're designing for. It may be trite, but a good visualization is better than the sum of its parts."[4] I think Simmon brings an important point between atheistic and functionality in data visualization. As a graphic designer, I totally understand a visualization is not only about to create a beautiful image, and my mission is to make people can understand the meaning (or tell a story) behind it through the visual elements. To realize this goal, I have to explore a standard meaning of each line, shape, and color with the reference of cultural differences or human perception. On another hand, a design of the clean and reasonable visualization always can deliver information to users more accurate rather than to spend a lot of time on some fancy but unnecessary visual elements.

As an overview of data visualization, it is not just to covert data from values to a colorful diagram, and we need to consider each part that could affect user experience. In the design system, either scalability or storytelling plays its own role in the process to guarantee users can have the best experience. Given the above, my duty is to become a storyteller to make data be accessible and understandable to everyone.

[4] Robert Simmon. "Subtleties of Color," *NASA*, Blog (August 5, 2013). https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/05/subtleties-of-color-part-1-of-6/

Bibliography

Kleppmann, Martin. “Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems,” *O’Reilly Media*, Book (2017). Page 10-18

Langston, Jennifer. “Project Silica: Proof of Concept Stores Warner Bros. ‘Superman’ Movie on Quartz Glass,” *Microsoft*, Innovation Stories (November 4, 2019). https://news.microsoft.com/innovation-stories/ignite-project-silica-superman/

Robert Simmon. “Subtleties of Color,” *NASA*, Blog (August 5, 2013). https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/05/subtleties-of-color-part-1-of-6/