OM^3: An Ordered Multi-level Min-Max Representation for Interactive Progressive Visualization of Time Series

– Supplementary Material –

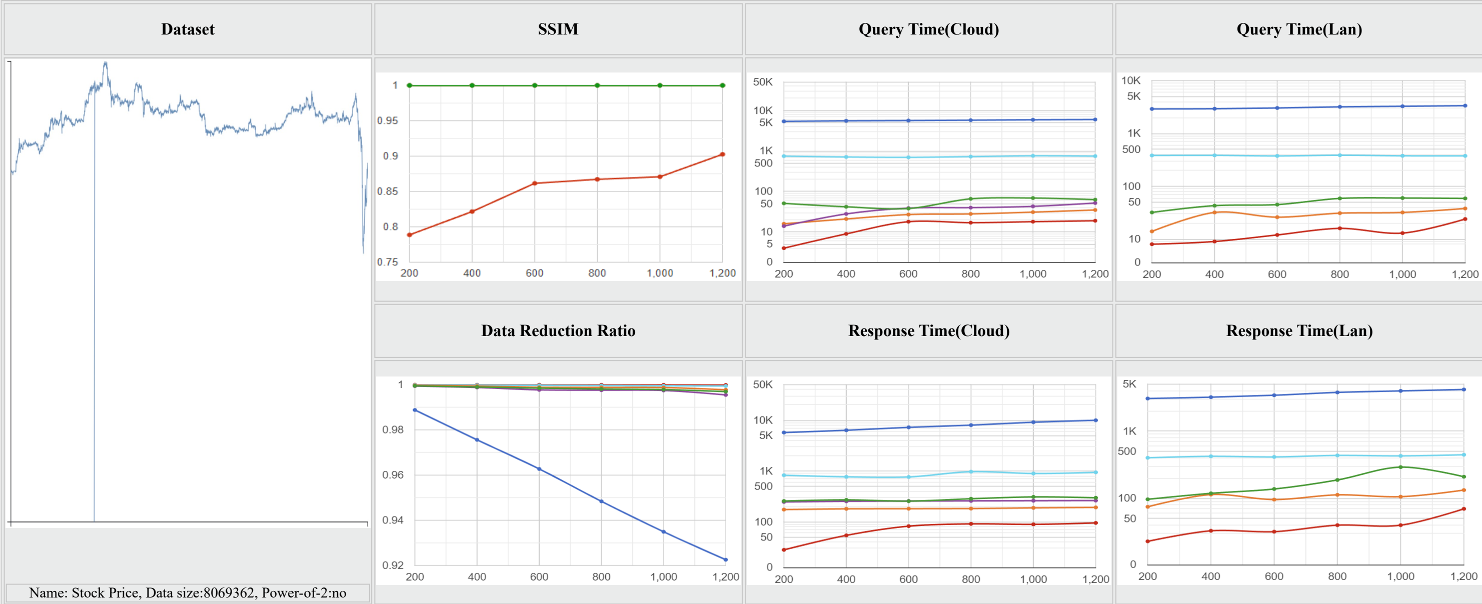
Category: Research

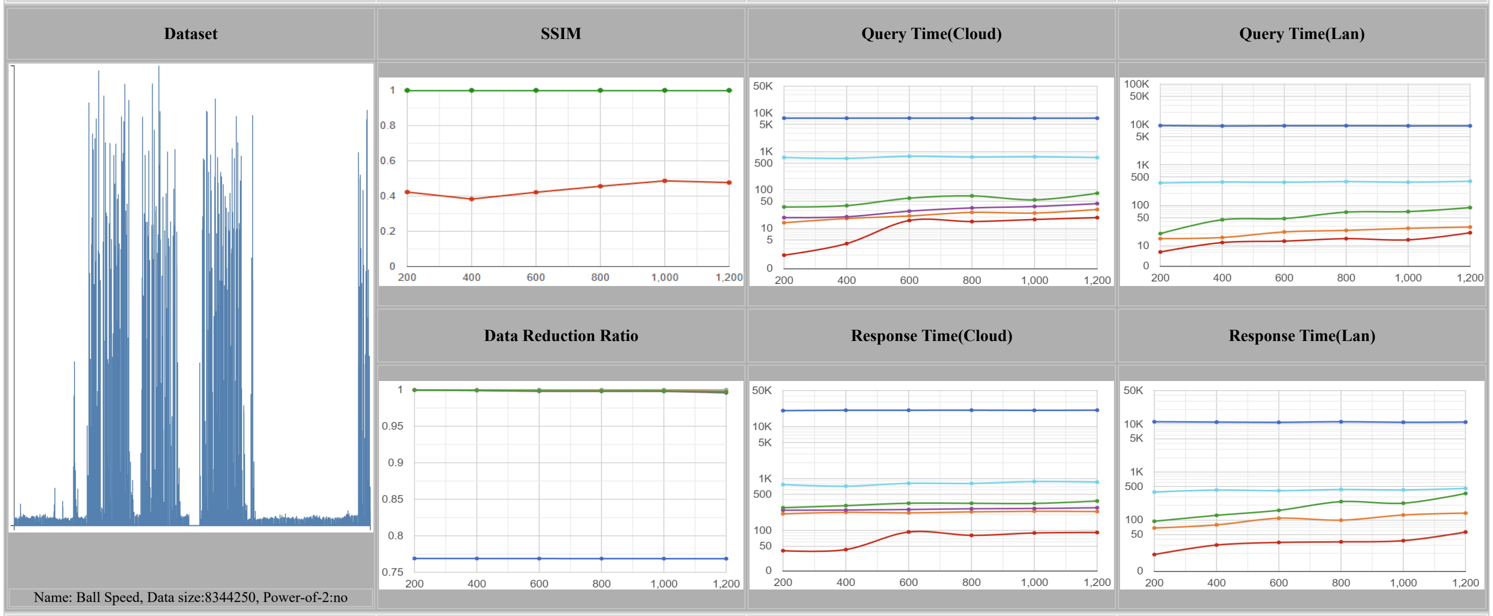
This supplemental material file provides complete experimental results for our submitted paper titled "OM^3: An Ordered Multi-level Min-Max Representation for Interactive Progressive Visualization of Time Series." In the evaluation of static visualization of our paper, three real datasets were employed in the evaluation of M4 [1] and ten synthetic datasets were generated by combining the major trend component, sine function and white noise. Here, we first provide the database query statements for generating M4 aggregations using InfluxDB and PostgreSQL as follows:

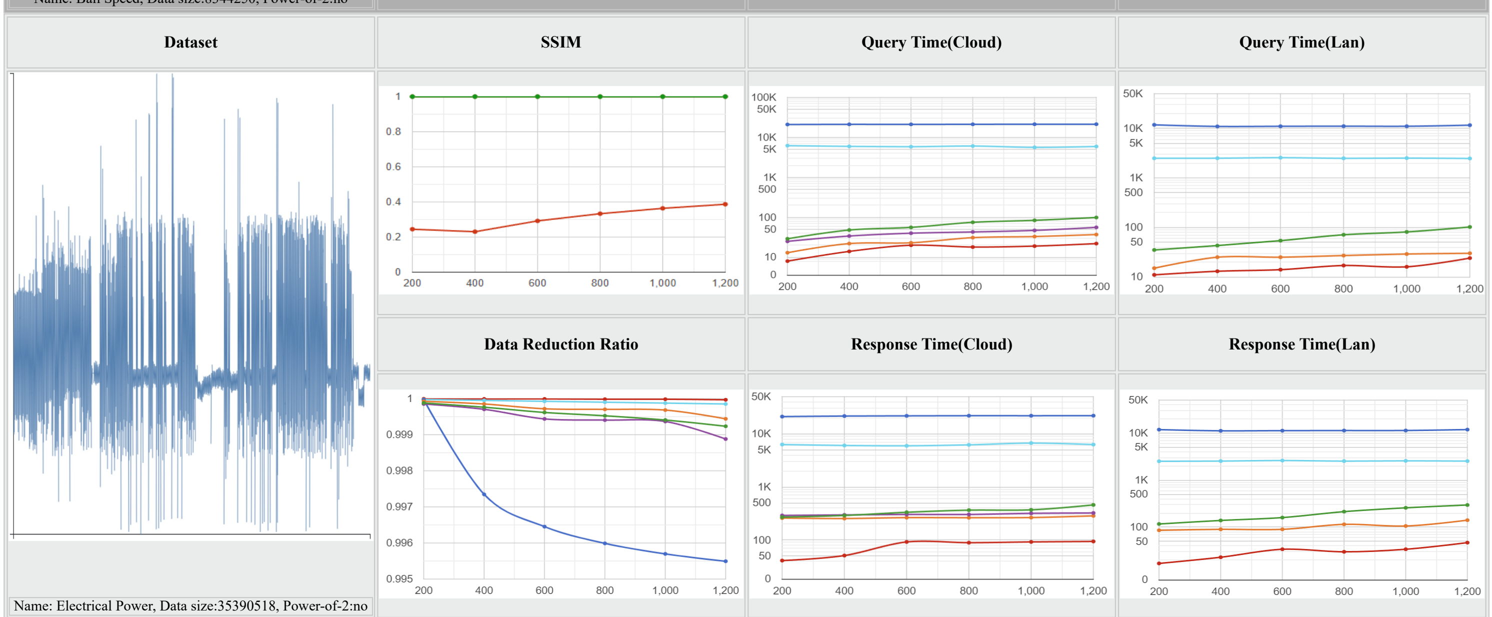
```
VMin = from(bucket: "bucket_name")
|> range(start: "t_start", stop: "t_end")
|> aggregateWindow(every: group_interval, fn: min)
VMax = from(bucket: "bucket_name")
|> range(start: "t_start", stop: "t_end")
|> aggregateWindow(every: group_interval, fn: max)
TFirst = from(bucket: "bucket_name")
|> range(start: "t_start", stop: "t_end")
|> aggregateWindow(every: group_interval, fn: first)
TLast = from(bucket: "bucket_name")
|> range(start: "t_start", stop: "t_end")
|> aggregateWindow(every: group_interval, fn: last)
```

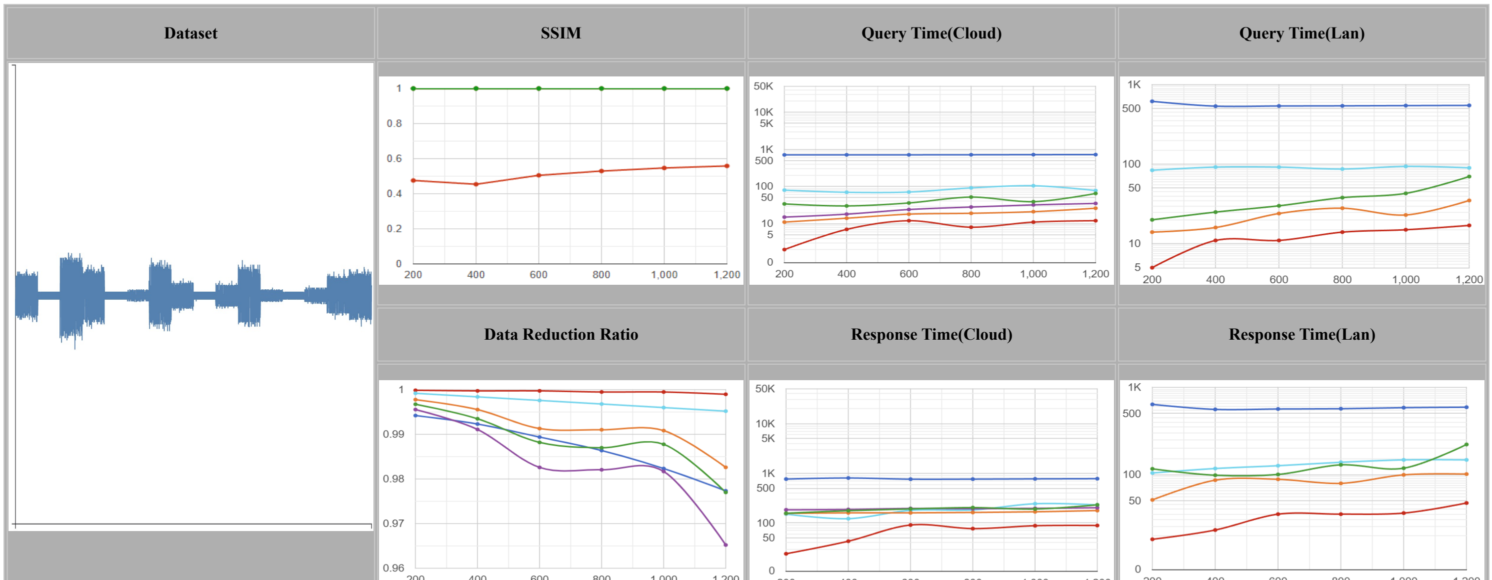
```
WITH Q AS (SELECT t,v FROM origin_table WHERE t >= t_start AND t <= t_end);
SELECT t,v FROM Q JOIN
SELECT floor(w / (t_end - t_start + 1 ) * (t - t_start) as k, --define key
min(v) as v_min, max(v) as v_max,                             --get min,max
min(t) as t_min, max(t) AS t_max                              --get 1st,last
FROM Q GROUP BY k) as QA                                      --group by k
ON k = floor(width / (t_end - tstart + 1) * (t - t_start))    --join on k
and (v = v_min OR v = v_max OR                                --&(min|max|
t = t_min OR t = t_max)                                       -- first|last)
```

Then, we provide screenshots of visualizations for all tested datasets with window width 600 and the corresponding scores on two settings (database residing either in the local area network (LAN) or on the cloud) of four metrics, which are represented by line charts, where the x-axis is the changing window width and different colors indicate different methods (■ for M4-P, ■ for M4-I, ■ for Haar wavelet [2], ■ for OM^3-NP, ■ for M^3 and ■ for OM^3). The LAN server is running on a 64-core AMD Ryzen Threadripper 3990X CPU with 4.3 GHz, 226GB RAM, and 4TB HDD with the Ubuntu 20.04.3 LTS, while the cloud server is running on an 8-core Intel Xeon Platinum 8269CY CPU with 2.5 GHz, 32GB RAM, and 250GB SSD with Ubuntu 22.04. For the evaluation of interactive visualization, we provide the complete results of data reduction ratio, query time, and response time of applying resizing, panning, zooming and hybrid interactions to two datasets (*synthetic* and *stock price*) on both the LAN and cloud settings. We do not provide SSIM [3], since all these methods are error-free. For the case study, we provide an enlarged version of the image to clearly demonstrate the process of our visual analysis.

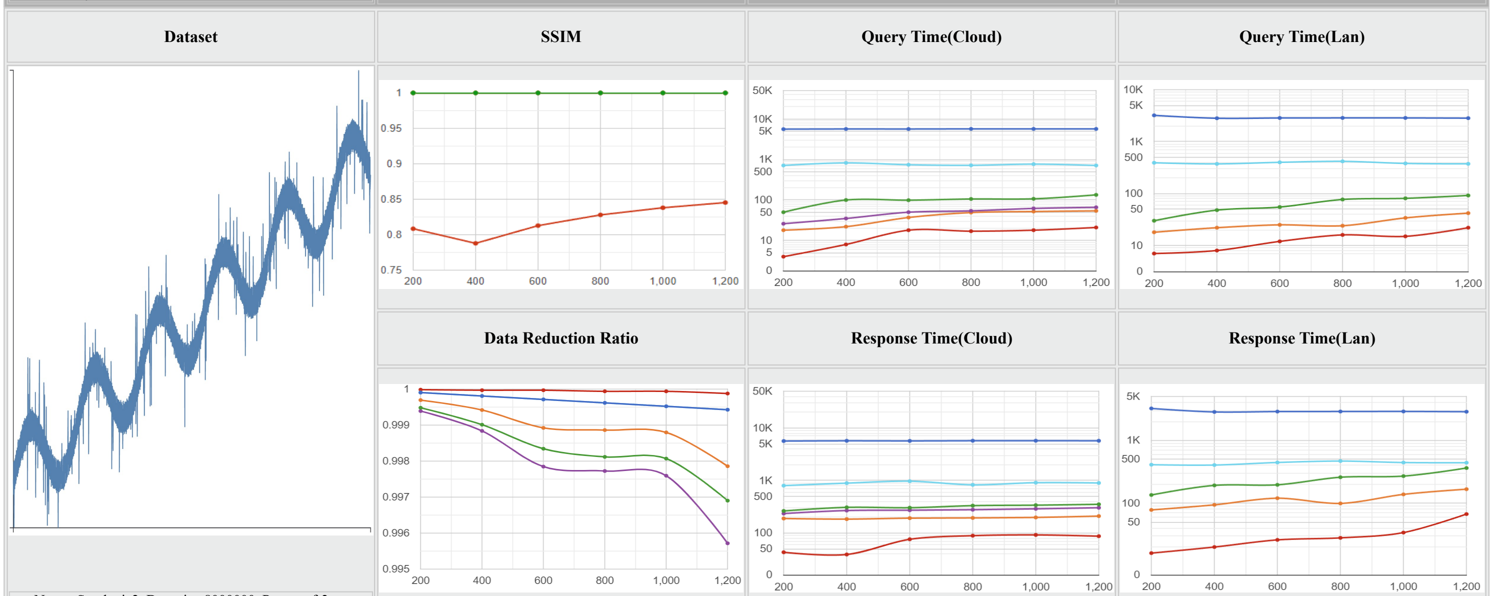

Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Stock Price, Data size:8069362, Power-of-2:no


Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Ball Speed, Data size:8344250, Power-of-2:no

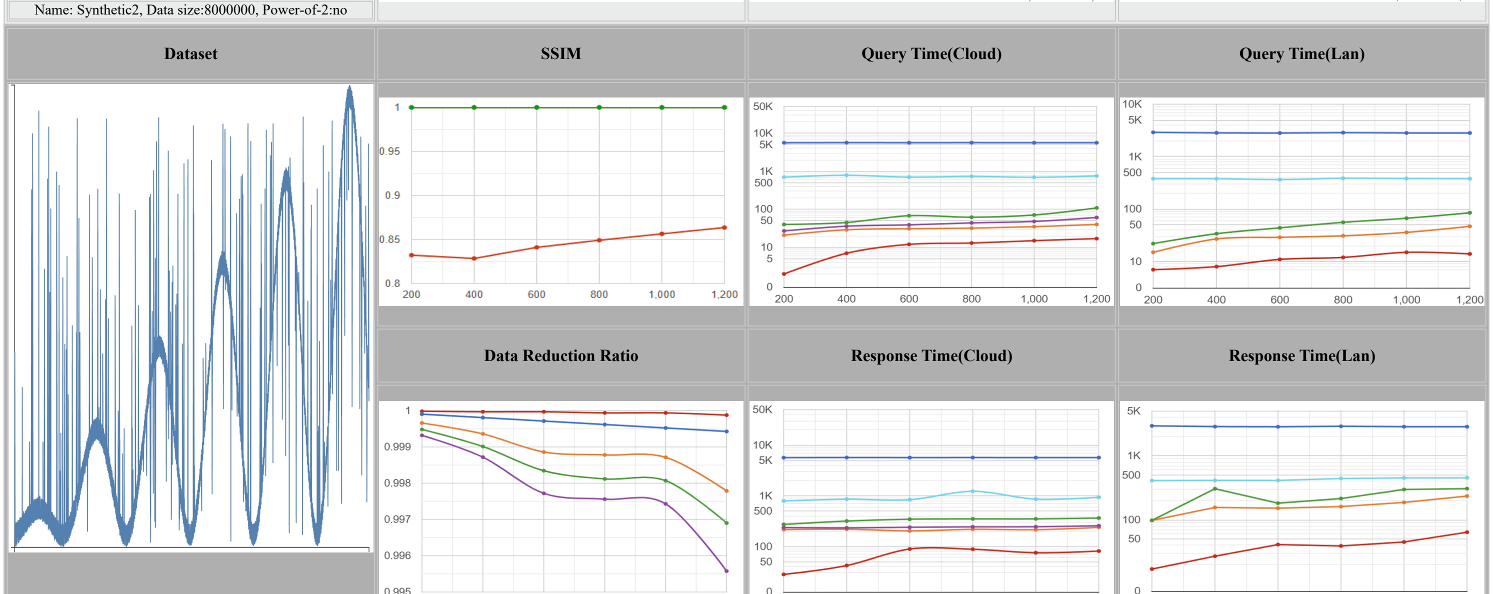

Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Electrical Power, Data size:35390518, Power-of-2:no



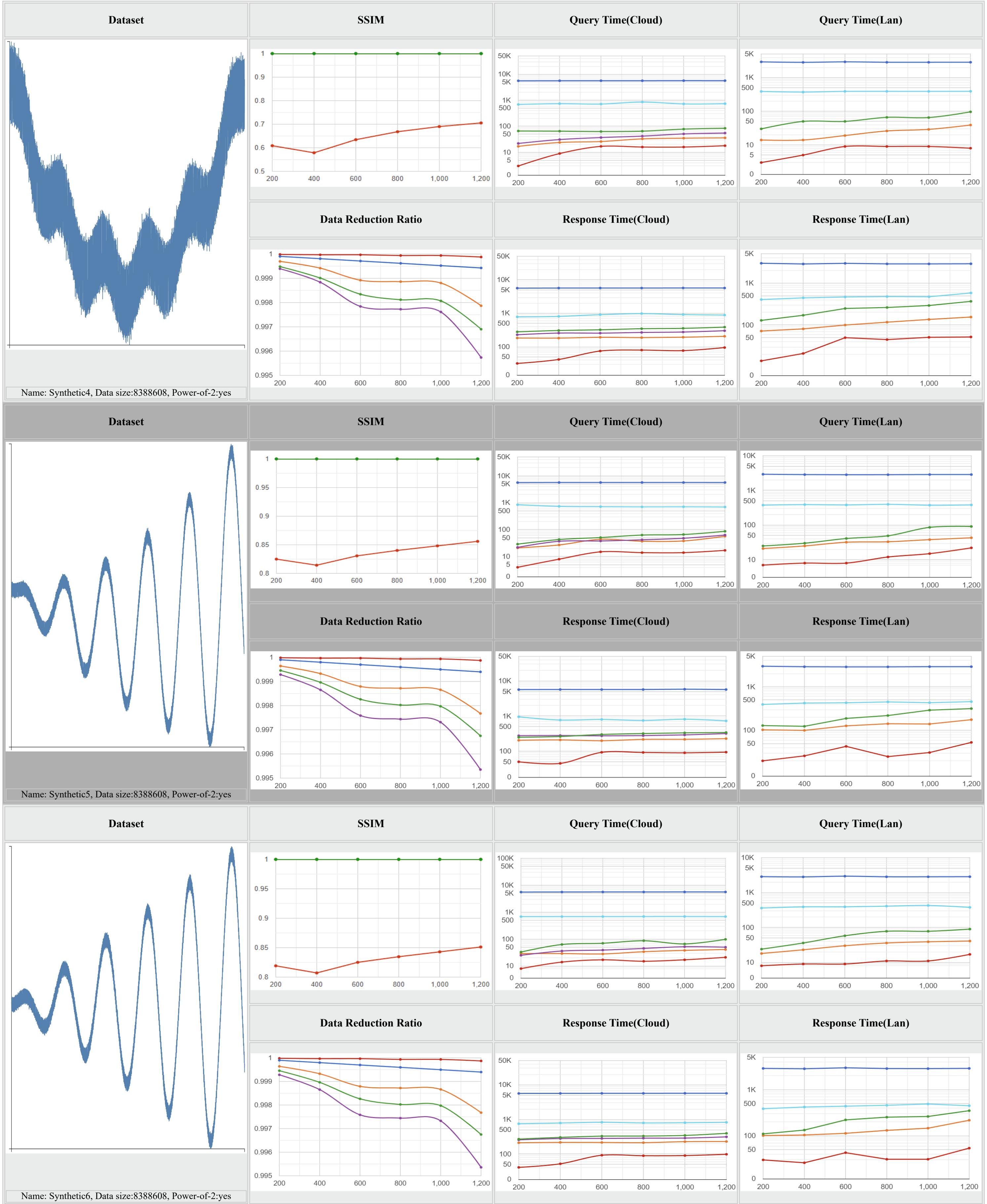
Name: Synthetic1, Data size:8000000, Power-of-2:no

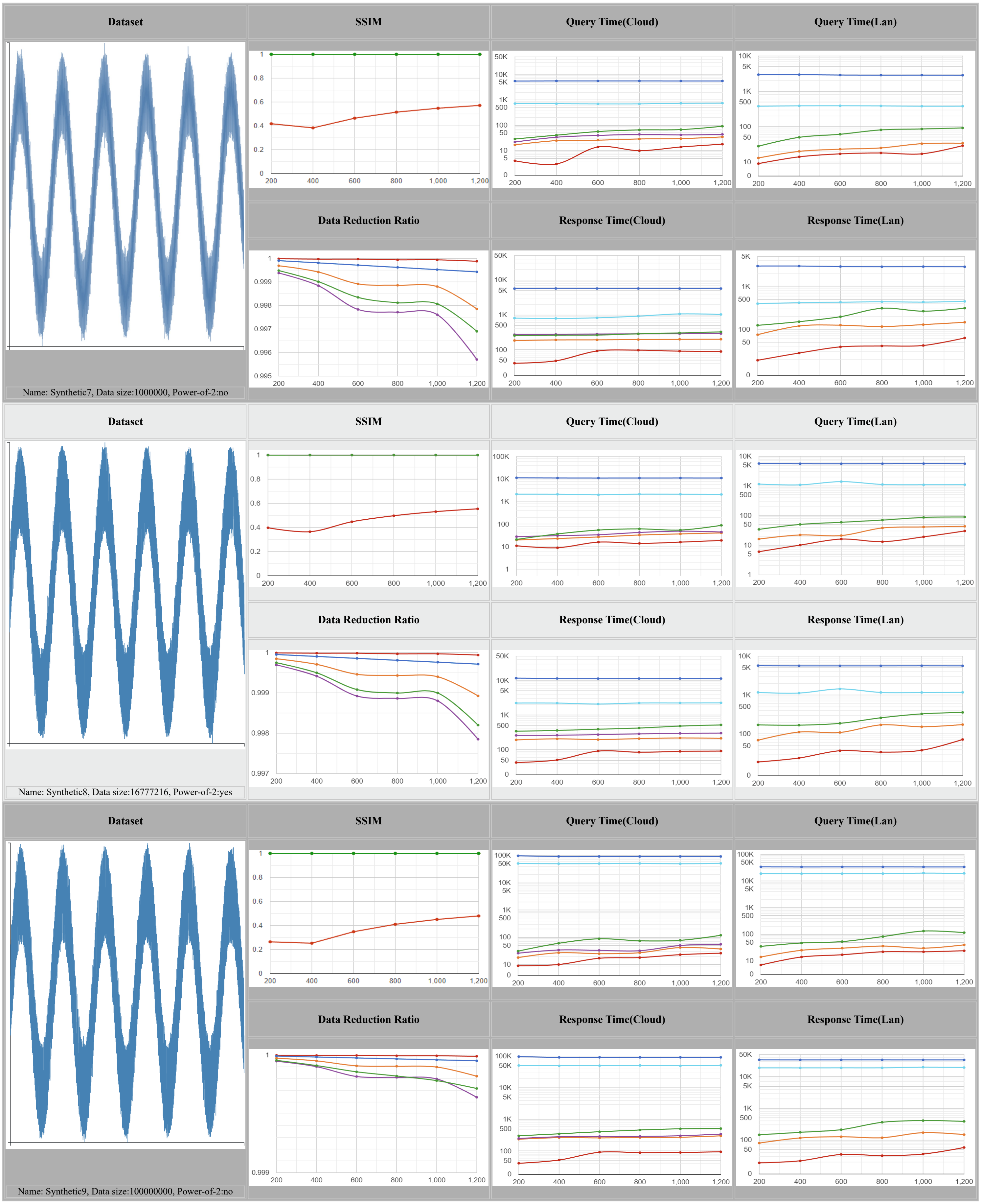


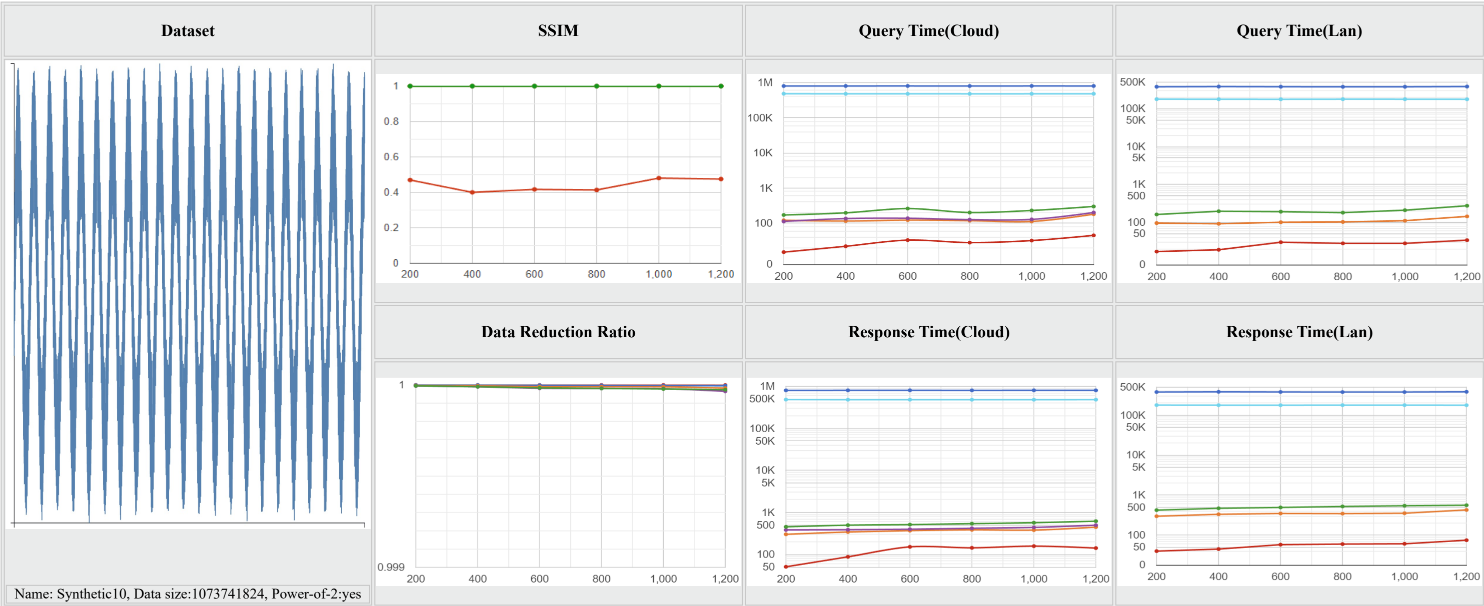
Name: Synthetic2, Data size:8000000, Power-of-2:no



Name: Synthetic3, Data size:8388608, Power-of-2:yes


Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic4, Data size:8388608, Power-of-2:yes
Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic5, Data size:8388608, Power-of-2:yes
Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic6, Data size:8388608, Power-of-2:yes


Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic7, Data size:1000000, Power-of-2:no
Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic8, Data size:16777216, Power-of-2:yes
Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic9, Data size:100000000, Power-of-2:no


Dataset
SSIM
Query Time(Cloud)
Query Time(Lan)
Data Reduction Ratio
Response Time(Cloud)
Response Time(Lan)
Name: Synthetic10, Data size:1073741824, Power-of-2:yes

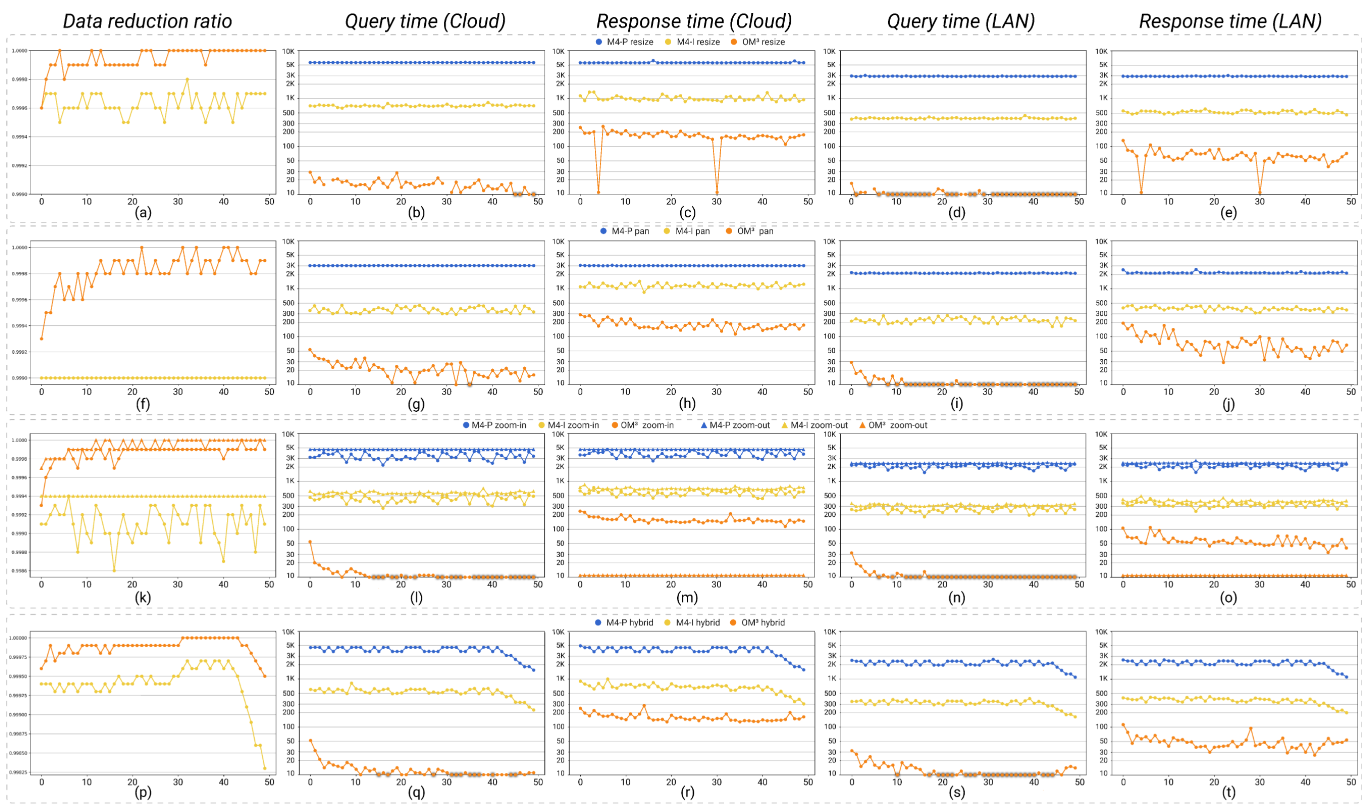


Fig. 1: The results of interactive visualization evaluation of the synthetic dataset. The curves show how the data reduction ratio, query time, and response time vary within 50 trials of applying four types of interactions: resizing (a,b,c,d,e), panning (f,g,h,i,j), zooming (k,l,m,n,o), and hybrid of resizing, panning, and zooming (p,q,r,s,t) with two settings, times are on a logarithmic scale. For values out of the plot range (see, e.g., (d)), we indicate them by a dark transparent shadow. (a,f,k,p) The data reduction ratios of M4-I and M4-P overlap and are lower than OM^3 in general. The query time of the cloud setting (b,g,l,q) is longer than the one of LAN setting (d,i,n,s) since the hardware of LAN server is better. For OM^3, the response times of the cloud setting (c,h,m,r) and the LAN setting (e,j,o,t) are less than 300 ms and 200 ms, respectively.

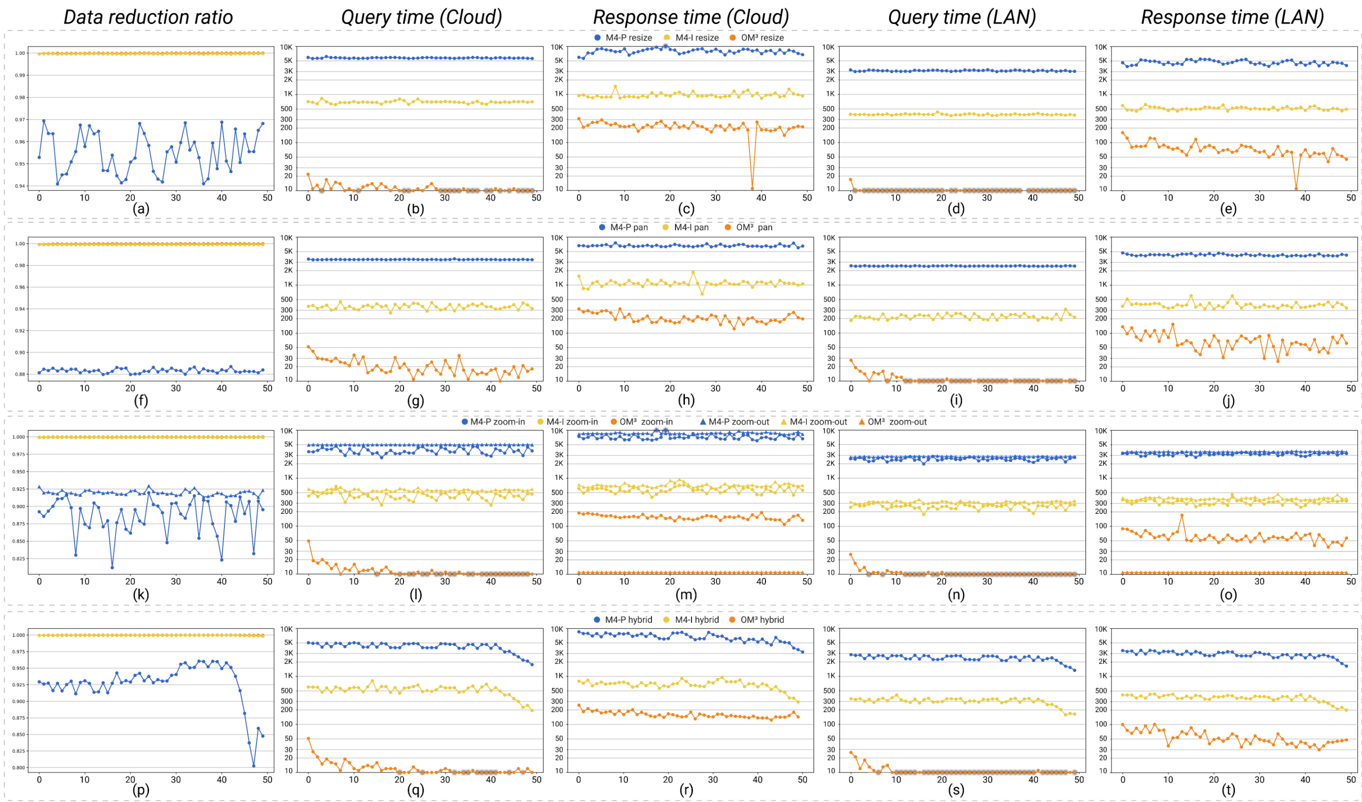


Fig. 2: The results of interactive visualization evaluation of the stock price dataset. The curves show how the data reduction ratio, query time, and response time vary within 50 trials of applying four types of interactions: resizing (a,b,c,d,e), panning (f,g,h,i,j), zooming (k,l,m,n,o), and hybrid of resizing, panning, and zooming (p,q,r,s,t) with two settings, times are on a logarithmic scale. For values out of the plot range (see, e.g., (d)), we indicate them by a dark transparent shadow. When the . (a,f,k,p) The data reduction ratio of M4-P is much lower than M4-I and OM^3 because M4-P will find duplicates of minima and maxima in each pixel column from the server. The query time of the cloud setting (b,g,l,q) is longer than the one of LAN setting (d,i,n,s) since the hardware of LAN server is better. The response times of the cloud setting (c,h,m,r) and the LAN setting (e,j,o,t) show OM^3 is significantly faster than M4 methods.

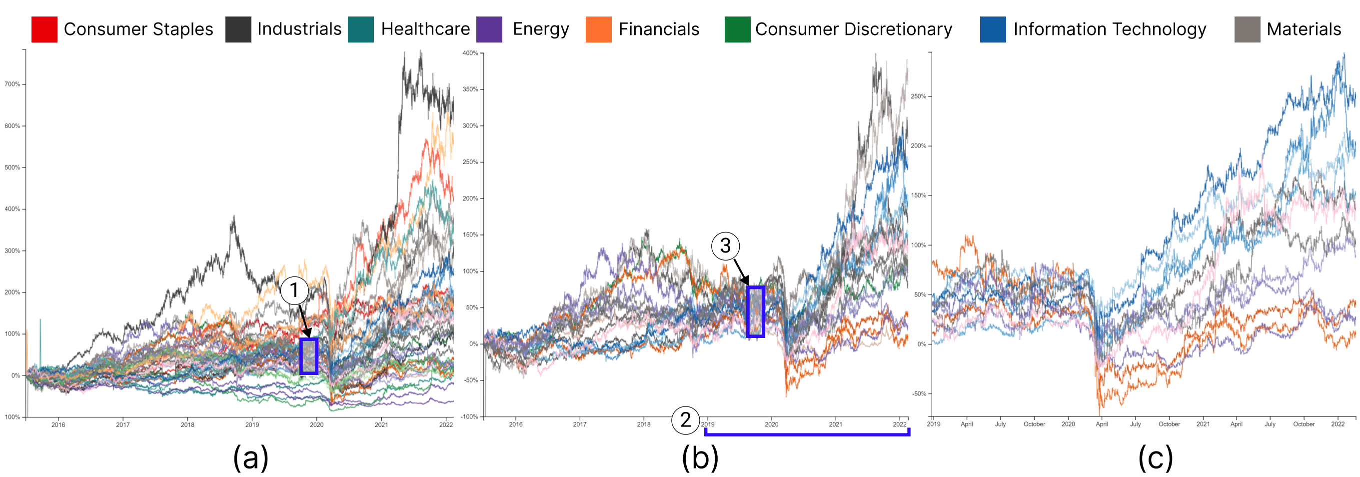


Fig. 3: The enlarged version of the image of the case study. (a) A line visualization of 44 stock prices, exhibiting heavy visual clutter. Hence, we put the timebox marked by the blue box to filter out the stocks with different behaviors and produce the visualization shown in (b). (c) Further, we zoom into a specific time interval between 2019 and 2022 and then use a timebox to further filter the stocks.

References

[1] U. Jugel, Z. Jerzak, G. Hackenbroich, and V. Markl. M4: A visualization-oriented time series data aggregation. vol. 7, p. 797–808. VLDB Endowment, jun 2014. doi: 10.14778/2732951.2732953

[2] D. F. Walnut. *An introduction to wavelet analysis*. Springer Science & Business Media, 2002.

[3] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861