

5/29

Parallel Visualization on Large Clusters using Map Reduce

说明利用分布式计算 MapReduce 进行绘制是可能的。

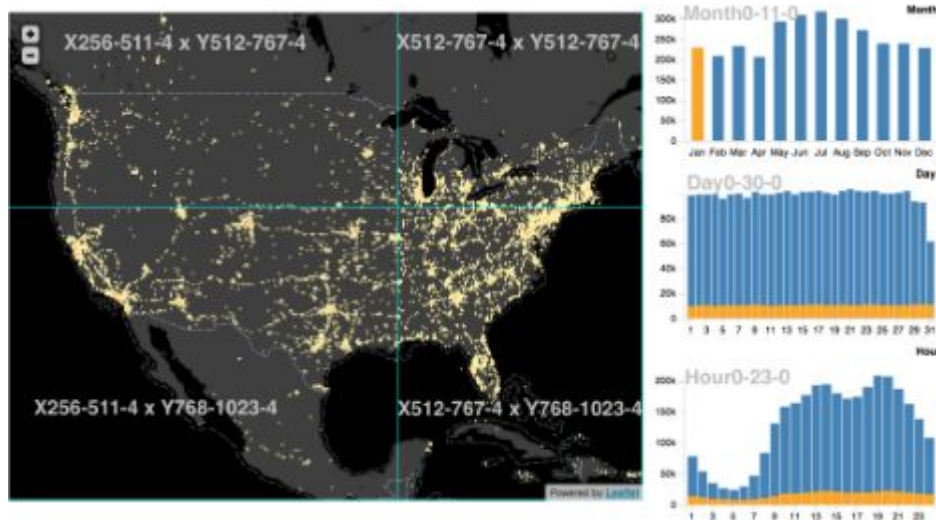
关键是如何设计将这个任务分布开来。

这篇文章用的方法似乎很 Naive ,但至少确实可行的 ,对于效率也做了很多分析。

imMens: Real-time Visual Querying of Big Data

WebGL

Integrate multivariate data tiles and parallel processing



4.5 million user checkins on Brightkite

35.6 million flight delays in the U.S. from 1989 to 2008

10K to 1B synthetic data points visualized as scatterplot matrices (SPLOM)

Weakness :

1. 受其数据结构的限制，不能多条件筛选。
2. 不能自动优化配置
3. 不能分 level 精细和过渡
4. 指支持二维的散点图，可以考虑更高维度：如平行坐标，网络（上次提到的力引导图）。

可以将其思路和算法放到集群的并行计算中。

可用的数据集

A Gallery of Large Graphs (上次提到的)

<http://yifanhu.net/GALLERY/GRAPHS/index.html>

Stanford Large Network Dataset Collection

<http://snap.stanford.edu/data/index.html>

Data Reduction Methods

- filtering

A HLBERG C., S HNEIDERMAN B.: Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In Proceedings of CHI (1994), pp. 313–317.

- sampling

D AS S ARMA A., L EE H., G ONZALEZ H., M ADHA - VAN J., H ALEVY A.: Efficient spatial sampling of large ge-

ographical tables. In Proceedings of SIGMOD (2012), ACM, pp. 193–204.

M IHALISIN T., T IMLIN J., S CHWEGLER J.: Visualizing multivariate functions, data, and distributions. IEEE Comput.

Graph. Appl. 11, 3 (May 1991), 28–35.

- aggregation

C ARR D. B., L ITTLEFIELD R. J., N ICHOLSON

W. L., L ITTLEFIELD J. S.: Scatterplot matrix techniques for large n. Journal of the American Statistical Association (1987), 424–436.

J ERDING D. F., S TASKO J. T.: The information mural: A technique for displaying and navigating large information spaces. IEEE TVCG 4, 3 (1998), 257–271.

近期需要做的：

1. 研究 Sparking Streaming 是否合适。可能要升级实验室集群上 spark 的版本。
2. 看上述对应的文章
3. 详细看 imMens 的实现。 <https://github.com/uwdata/imMens>
- 4.