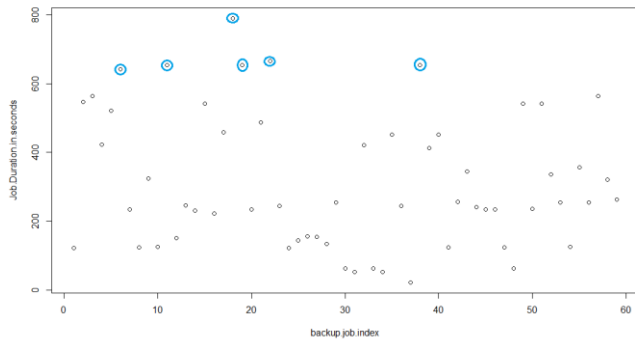


(b)



(c)

Figure 2: Top 6 outliers of (a) data transferred (b) number of backup files, and (c) job duration during for a backup client.

References

- [1] J. Kelly. (2013). *Taming Big Data* [Online]. Available: <http://wikibon.org/blog/taming-big-data/>
- [2] J. H. Howard *et al.*, "Scale and performance in a distributed file system," *ACM Trans. Comput. Syst.*, vol. 6, no. 1, pp. 51_81, 1988.
- [3] R. Cattell, "Scalable SQL and NoSQL data stores," *SIGMOD Rec.*, vol. 39, no. 4, pp. 12_27, 2011.
- [4] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107_113, 2008.
- [5] T. White, *Hadoop: The Definitive Guide*. Sebastopol, CA, USA: O'Reilly Media, 2012.
- [6] IBM 2012, What is big data: Bring big data to the enterprise, <http://www-01.ibm.com/software/data/bigdata/>, IBM.
- [7] Michel F. 2012, How many photos are uploaded to Flickr? <http://www.flickr.com/photos/franckmichel/6855169886/>.
- [8] T. Rabl, M. Sadoghi, H.-A. Jacobsen, S. Gomez-Villamor, V. Muntès-Mulero, and S. Mankovskii, "Solving big data challenges for enterprise application performance management," *VLDB*, vol. 5, 2012.
- [9] RevolutionAnalytics, <https://github.com/RevolutionAnalytics/RHadoop/wiki>.
- [10] P Anchalina, Kaushik Roy The k-Nearest Neighbor Algorithm Using MapReduce Paradigm