**Review on**

**“Mesa: Geo-Replicated, Near Real-Time, Scalable Data Warehousing”**

**by group 6**

**Introduction**

This paper focuses on a new data-warehouse system called Mesa, which is invented and used by Google for Google’s Internet advertising business. Mesa is geo-replicated across multiple data-centers and could handle petabytes of data and serve billions of queries at low latency. It introduces the structure, functionality and production metrics of Mesa. This topic is very important and useful because it solves the problem of extensive increasing demand of advertisements on google’s platform. Below I will discuss my comments and suggestions, which hopefully can help the author to improve the paper.

**Merits**

The paper introduces very comprehensive about some important aspects of Mesa system. In Introduction proposes authors seven requirements, which this data warehousing system must satisfy. Then in the following description introduce authors the mechanisms about how are those requirements actualized. Explanations for the mechanisms are easier to understand by using of comparisons with older system such as BigTable, Colossus.

**Critique**

1 There is one small mistake in figure 2 about how two updates for table B are added to generate figure 1 table B. There is a AdvertiserId 1 from US on 2014/01/01 in figure 1 table B with clicks 5 and cost 3. Then in figure 2 (d) update version 1 for table B changes the date to 2013/01/01.

2 Some arguments are too abstract to be understood. For example in section 4.2 describe authors that in order to parallel worker operation, must sample a Mesa worker every s-th row key. Then definite the maximum number of input rows per partition is at most n/p+(m-1)s. Here is only qualitative analysis, which comes to conclusion that larger s is, is the total number of samples smaller. Yet, quantitative analysis of the relationship between s and n/p is absent. I mean s should be large set but exactly which value or formula should s satisfy is not clear.

3 Some figures couldn't match corresponding functional descriptions. In section 6.1 introduce authors update processing metrics of Mesa. The first figure describes the number of row updates for one data source over a seven day period. According to the description reads Mesa 30-60 megabytes of compressed data per second and adds about 3000 thousand new rows. Those two metrics are not illustrated in the first figure.

**Conclusion**

Overall, I think that this paper about Mesa is important and well founded. Although there are some small flaws in the paper, counts this paper still very worthful because of innovation and future foreseeability. The authors team have devoted a great deal of time in the research and development of this system and therefore accumulated a lot of experiences, which they are also mentioned in section 5. I hope that these comments could help to improve this paper.