

Adaptive Data Splitting and Column Chunk Ordering in Wide Tables

First Author #¹, Second Author *², Third Author #³

DEKE Key Lab (MOE), Renmin University of China, Beijing, China

¹first.author@first-third.edu

³third.author@first-third.edu

Abstract—**«TODO write later.»**

I. INTRODUCTION

«TODO Write later. 1) What is wide table and where it is used; 2) Why do we use wide tables instead of normal tables; 3) What are the problems we try to solve in this paper; 4) Challenges; 5) Our achievements and contributions»

II. PRELIMINARIES

In this section, we review today's HDFS-based column stores in which wide tables are stored, and how queries are executed on top of these wide tables [1].

A. Disk Seek Cost

B. Wide Table Layout in HDFS

«TODO 1) columnar format in HDFS; 2) wide tables are store in these formats; 3) disk seek cost problem.»

C. Query Execution on Wide Tables

«TODO 1) one task per split / row group; 2) task schedule and initialization cost; 3) adaptive split size problem.»

III. QUERY-WISE DATA SPLITTING

IV. COLUMN CHUNK ORDERING

V. IMPLEMENTATION

VI. PERFORMANCE EVALUATION

VII. RELATED WORK

VIII. CONCLUSION

REFERENCES

[1] (2007) The IEEE website. [Online]. Available: <http://www.ieee.org/>