

Literature Review

Dylan Davis
Oregon State University
Corvallis, Oregon

November 20, 2017

Abstract

This document is a review of several relevant research sources that pertain to the HP Big Data Compression research project. In this review, two different areas of research will be addressed, with each having explanations of several different research papers or other documents that contribute to that research. These sources will include research white papers as well as articles by experts in the research field this project is in.

1 Oracle Advanced Compression

1.1 Overview

Oracle advanced compression is database table compression algorithm that is similar to Oracle's basic compression algorithm but with more features included. This algorithm initially compresses tables using the same method as basic compression. However, when normal data manipulation operations are performed on a table that is compressed with advanced compression, new data is also compressed, maintaining the tables compression ratio. When these operations are performed on tables compressed with basic compression, new data is not compressed and the table begins to take up more storage space. Below, three useful sources of information to be studied when researching Oracle advanced compression are discussed.

1.2 Oracle Database - Compression Best Practices

This document from Oracle describes the use of their various compression algorithms in Oracle Database release 12.2.0.1. It describes the proper syntax for the SQL statements to create tables using each type of Oracle compression, including advanced compression, as well as how to change existing uncompressed tables into ones compressed using any of these algorithms. It also goes into detail about the strengths and limitations of each compression algorithm, including compression ratios and how well the table stays compressed after several data manipulation operations, and under what circumstances each algorithm should be used. The document does not go into detail about how each algorithm actually compresses the data. [6]

1.3 Oracle Advanced Compression with Oracle Database 12c Release 2

This white paper by Oracle goes into more detail on Oracle advanced compression specifically. It describes the benefits of advanced compression over basic table compression for data tables that will often have data updated or new data inserted. It also gives a general description of how the compression algorithm actually compresses the data and keeps it compressed over the course of many data manipulation operations. Besides just advanced row compression, this document also describes other compression options included with Oracle Database. These include compression for backup data, network compression, both basic and advanced index compression, and data pump compression. Finally, the document also describes several tools Oracle provides which help with managing compression, such as Heat Map, which tracks activity levels of data within the database, and Automatic Data Optimization, which can use the data from Heat Map to choose the best compression practices for a given set of data. [5]

1.4 Master Note for OLTP Compression

This document is another release from Oracle describing Oracle advanced compression. It provides further information on how advanced compression compresses data and its uses and limitations. It also provides some examples of tests done on advanced compression and their results, as well as useful statistics. Finally, it provides information on possible bugs with advanced compression which was not included in the documents described earlier. [7]

2 Oracle Database Block Structure

2.1 Overview

Oracle Database stores data in data blocks, usually 8 kilobytes in size. These blocks, and the trace file dumps used to examine them, have a specific structure and various components and headers. These components are especially important for compressed data blocks. Understanding how these data blocks work and being able to study them is vital when studying how Oracle Database compression algorithms work. Four sources of information on accessing and examining Oracle Database data blocks are described below.

2.2 Compression in Oracle - Part 1: Basic Table Compression

This article is a detailed description of how the Oracle basic table compression algorithm compresses data. In it, the author also tests the block counts of a table when uncompressed and compares that to the block count when the table is created using compression and when it is changed from uncompressed to compressed. More important, though, is the description of Oracle Database data blocks and how to study the block structure within a trace file containing a data block dump. This is used to describe how the Oracle compression algorithm works, but it is also

a very useful tool for learning about how data blocks are structured. The article explains how to tell if a data block is compressed, what order the columns of the table have been put in within the block, how to read the token table used to represent duplicate data within the block with a single identifier, and how to tell which rows of data contain which tokens. With the information provided in this article, an Oracle Database data block can be easily decoded and examined to see exactly how Oracle is representing the data within the block. [3]

2.3 Dumping Oracle: Version 12.2.0.1

This online source describes how to dump Oracle Database data blocks so that their contents can be examined in a trace file. For all the steps in this process, the SQL code is given. First, the SQL statements for creating an example table are given. Once this table is created and committed, the SQL statements to display the datafile and block number for this table are given. Next, the code is provided that uses the numbers retrieved in the last step to dump the contents of the data block to a trace file. Finally, an explanation of where to find the correct trace file within Oracle is given. Example contents of a data block dump within a trace file are given to show the structure of the data block. Without the information provided by this source, it would be nearly impossible to study the contents and structure of Oracle Database data blocks. [1]

2.4 Data block - Oracle FAQ

This article from Oracle FAQ's again describes the process of acquiring an Oracle Database data block's datafile and block numbers and dumping the contents of the block to a trace file but in a more condensed manner. In addition, it also describes how to dump the contents of multiple data blocks at one time, which is very useful when studying the differences between different blocks from the same table. Next, it provides a line by line description of an example data block dump, explaining how the rows and columns of the table are represented within the trace file. The article also gives some general information on data blocks and extents, or continuous sets of data blocks, and shows where overall information about blocks and extents in a given table can be found. [2]

2.5 Disassembling the Oracle Datablock

This paper describes the bbed, or block browser and editor, tool provided by Oracle Database and how it can be used to examine data blocks. The document warns that bbed is a dangerous tool because it can alter and corrupt the data within the data blocks of any Oracle database. It first provides detailed information on how to setup and start the bbed tool from within Oracle. Next, it gives an overview of the various available commands that can be used with this tool and what each does. Some examples of these are set dba, set filename, and set blocksize which set the current data block, current filename, and size of the block respectively. Finally, the paper provides several worked examples to show how to properly use the bbed tool to perform tasks such as changing data, recovering deleted rows, or uncorrupting a block. [4]

3 References

- [1]D. Morgan, "Dumping Oracle: Version 12.2.0.1", Morganslibrary.org, 2017. [Online]. Available: https://www.morganslibrary.org/reference/dump_oracle.html. [Accessed : 21 – Nov – 2017].
- [2]"Data block - Oracle FAQ", Orafaq.com, 2009. [Online]. Available: http://www.orafaq.com/wiki/Data_block. [Accessed : 21 – Nov – 2017].
- [3]J. Lewis, "Compression in Oracle Part 1: Basic Table Compression - Simple Talk", red-gate.com, 2013. [Online]. Available: <https://www.red-gate.com/simple-talk/sql/oracle/compression-oracle-basic-table-compression/>. [Accessed: 21- Nov- 2017].
- [4]G. Thornton, Disassembling the Oracle Data Block. OraFAQ, 2005.
- [5]Oracle Advanced Compression with Oracle Database 12c Release 2 (white paper). Oracle, 2017.
- [6]C. Pedregal, Oracle Database - Compression Best Practices. Oracle, 2015.

[7]L. Varady, "Master Note for OLTP Compression", Blogs.oracle.com, 2010. [Online]. Available: <https://blogs.oracle.com/db/master-note-for-oltp-compression>. [Accessed: 21- Nov- 2017].