MATERIŁY
XIV MIĘDZYNARODOWEJ NAUKOWI-PRAKTYCZNEJ KONFERENCJI

NAUKOWA PRZESTRZEŃ EUROPY - 2018

07 -15 kwietnia 2018 roku

Volume 10

Budownictwo i architektura
Geografia i geologia
Współczesne informacyjne technologie
Techniczne nauki
Fizyka
Chemia i chemiczne technologie

Przemyśl
Nauka i studia
2018

Adres wydawcy i redacji:
37-700 Przemyśl ,
ul. Łukasińskiego 7

Materialy XIV Miedzynarodowej naukowi-praktycznej konferencji , «Naukowa przestrzeń Europy - 2018» , Volume 10 Przemyśl: Nauka i studia -80 s.

Materialy XIV Miedzynarodowej naukowi-praktycznej konferencji , «Naukowa przestrzeń Europy - 2018» , 07 -15 kwietnia 2018 roku po sekcjach: Budownictwo i architektura. Geografia i geologia. Współczesne informacyjne technologie. Techniczne nauki. Fizyka. Chemia i chemiczne technologie.

e-mail: praha@rusnauka.com

Cena 54,90 zł (w tym VAT 23%)

ISBN 978-966-8736-05-6

Darkenbayev D.K.
Al-Farabi Kazakh National University, Almaty, Kazakhstan

INCREASING THE EFFICIENCY OF PROCESSING LARGE-SIZE DATA USING BIGSQL TECHNOLOGY

Abstract. In this article, I have considered one of the many technologies for processing large-size data. The amount of data from day to day is growing, and it can be said that there is a great deal of work for IT professionals who analyze and manage data. Many experts are eager to work with large-size data, as Apache Hadoop offers many new technologies. More specifically, programming languages require queries, management scenarios, and more. Apache Hadoop is one of the widely used platforms. The IBM Hadoop platform InfoSphere BigInsights BigSQL-SQL interface makes it easy to edit and use data.

Kew words: Apache Hadoop, data, BigSQL, obobyotka, technology, BigData, platform

Introduction. The BigSQL technology service offers SQL developers an inclusive approach to query Hadoop-driven data. It lets you create a new table for data stored in BigInsights file system, Hbase, Hive, allocated to system administrators. It also allows system administrators to populate BigSQL table data from different sources. JDBS and ODBC drivers allow the BigSQL technology to utilize existing tools that can query individual files. However BigSQL does not add Hadoop data to large individual relational databases. In this article, I have considered the basics of BigSQL, and I will clarify the various issues that arise when dealing with the database [1]. What is a BigData? Larger information we can process on even a few computers when we edit it. Consequently, BigData is large size information. At Big Date, we're able to collect large amounts of data and not just collect and process information. These methods apply to large volumes of data. Based on the Big Data definition, we can define the basic principles of working with such data: 1. Horizontal scaling. As much as you need - any system that includes large data processing needs to be expanded. The amount of iron in the cluster doubles in size and the whole process continues. 2. Tolerance to errors. The principle of horizontal scaling means that there are many

machines in the cluster. For example, in the Hadoop cluster of Yahoo there are more than 42,000 machines (this reference shows the cluster size in different organizations). This ensures that some machines are operating successfully. Methods for dealing with large volumes of data should take into account the possibility of such failures and see no significant consequences. 3. Data layout. In large systems, data is split into multiple machines. If the data is physically located on a single server and handled on another server, data transmission costs may exceed processing costs. Therefore, one of the most important principles for developing BigData solutions is the principle of data layout. If possible, we process data in a single machine that stores it [2].

BigSQL - general description of technology.

BigSQL allows IT technicians to create queries and queries with BigInsights data through SQL statements. Programmers can use standard SQL syntax and, in some cases, use Hadoop technology to compact SQL-compiled SQL capabilities. Figure 1 depicts the Big SQL Infrastructure Architecture and the BigInsights Enterprise Edition 2.1 platform.

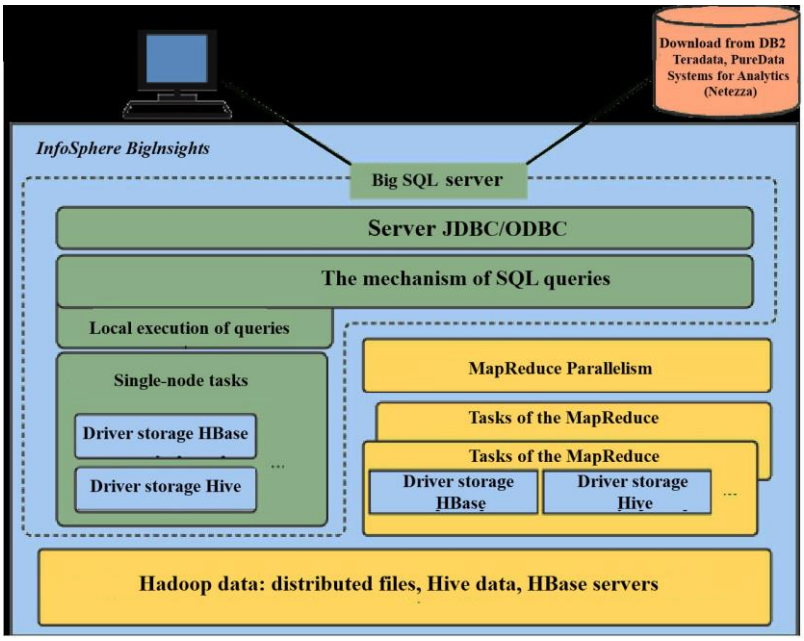

Figure 1. BigSQL architecture.

BigSQL enables users of Linux, Windows to complete JDBC and ODBC access. In addition, BigSQL can read data from some local or distributed BigInsights file system directly from a relational database management system. BigInsights EE 2.1 file system can be used to improve the Hadoop Distributed File System with the File Placement Optimizer optimizer. The SQL queries mechanism is convenient for joining, joining, grouping, converting to a shared table, and converting

conventional SQL. We also access data identified by queries through configuration settings and optimization tips. Depending on the nature of the request, size of the data depends on the other factors of the BigSQL, it can refer to the Hadoop MapReduce infrastructure for parallel processing or perform the query on a personal node on the local BigSQL server [3].

Principles of working with BigSQL

BigInsights has tools and interfaces similar to BigSQL's interface with the relational database management tools. As previously mentioned, BigSQL offers developers 32, 64-bit, Java, C, C ++ languages to support JDBC and ODBC drivers. These drivers support high capabilities. BigInsights's Eclipse plugin also allows Java developers to test and query BigSQL application requests. Figure 2 shows the plugin, including the JDBC driver's BigSQL server, and the test result [4].

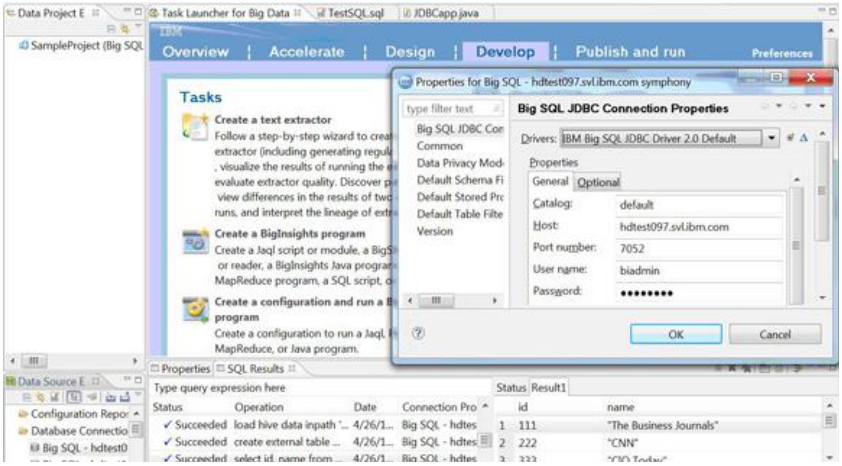

Figure 2. BigInsights Eclipse plugin fully supports BigSQL.

Big SQL offers the BigInsights command line interface and the Web interface for an interactive calling request. These tools are used to perform scenarios for creating prototypes. Big SQL can be configured for external devices that support JDBC and ODBC data. For example, CognosBusiness Intelligence uses the JDBC interface to create and analyze functionality for BigSQL data requests. Like other BigInsights components, BigSQL is a service that sends system administrators as much as possible from the Web Console or command line [5].

Uploading a table and loading data

When working with BigSQL, we know that it is necessary to create tables and fill them with data. This applies to GREAT TABLE and LOAD operators. Table creation syntax and data loading seem to be normal, but have minor problems. This

means that Hadoop is used for a certain type of technology [6]. Let's look at Listing 1 (table 1 in the table is a shortcut to the GOSALES storage database.)

1st listing. Creating a BigSQL table and entering data in an online file.

```
create table mygosales.product_brand_lookup
 (product_brand_code int,
product_brand_en varchar(180)
)row format delimited fields terminated by'\t';
load hive data local inpath '/home/user1/data/product.tsv'
overwrite into table mygosales.product_brand_lookup;
```

The Great table operator creates a Hive table with two columns. The first one is intended for writing a brand digital ID of the product, and the other one is for branding in English. The last line of this operator indicates that data is stored in tabular form with a string. It's worth noting that BigSQL has Hive's shared directories, so when we talk about the Hive chart, we're talking about the BigSQL table. 1st listing shows one way to save BigSQL tables. For example, you can create a table that stores data separately from the Hive table, which is stored in a separate file system. That way, when you delete a table from BigSQL, only metadata will be deleted, not physical data. As the default storage manager, when using the Hbase statement in the Great table statement, it is best to specify the keyword between the SQL columns and Hbase columns and the closest columns. If you're not familiar with Hive and HBase, you can get acquainted with the website on the Internet. Let's consider Load operator in Listing 1. Here we provide a complete way to the local file system that we want to include in the table. According to the definition of the table, each entry must include this field, separated by the symbol \ t in this file [7]. The Overwrite statement affects the content of the table that contains data for Big SQL. However, this does not reload data entry in the first listing listed in the BigInsights table on a separate file system. To do so, we need to create an external table using the Location statement and indicate where the existing data is. Also, there are several versions of data loading that can affect the database structure and affect the way the application works. For example, Hadoop systems are often used to store data types of different types like JSON. Contributors, that is, direct Hive developers use this service to read data for specific serializers / deserializers. Therefore, Big SQL fully supports specific SerDes functionality and affects the name of the SerDes class when creating a table. Big SQL supports multiple types of relational databases management systems, but some of them do not support data type Hive and Hbase. The user-defined real boundaries are consistent with relational databases and do not match those versions. The limitations of reference

uniqueness should be at the application orientation level. When using the GRANT and REVOKE operators together with the Hadoop standard commands, they limit the ability of data. They grant access to Hive data in the file system. This means that we have to think about the right at the table level, not at the level and column level. What does this mean for relational database developers or system administrators? If you want to create a simple test database on Hadoop, it's very efficient and convenient with Big SQL. In addition, BigInsights Quick Edition contains a step-by-step scenario for creating GOSALES where the database stores it. However, the creation of a BigInsights database requires a full understanding of Hadoop technology, and the design of the platform as appropriate [8].

Conclusion

If you have at least one of the BigSQL tables in your hand, you can query it using a SQL syntax. To design data, BigSQL can use SELECT to combine data into merge and sort data. It also provides a complete overview of internal queries that begin with WITH operators, and generalized table formulas. There are dozens of combined functions, including Hive-identified [9]. If needed, program developers can limit the number of lines that can be returned to the particular query. Big SQL technology IBMInfoSphere BigInsights adds the standard query interface to the Hadoop platform. However, Big SQL does not replace BigInsights with a relational database. It provides usual approaches to SQL users to the rapidly developing large-scale data processing [10]. In the article, I also highlighted the difference between the Big SQL core function, BigInsights, and the traditional relational databases management system. To learn more about the technology, you can download BigInsights Quick Start Edition and follow the detailed instructions for InfoSphere BigInsights to develop large SQL queries by analyzing large volumes of data.

References

[1] Tom White. Hadoop: The Definitive Guide, 3rd Edition. O'Reilly Media, 2012, 688 p.

[2] Darkenbayev D. K. Big Date. Ulken kolemdi derektermen zhumys isteu kagidalary

KazUPU khabarshysy, ISSN 17-28-7901, Almaty 2017., 204-207 p.

[3] Mark A. Beyer, Douglas Laney. The Importance of «Big Data»: A Definition.

http://www.gartner.com/DisplayDocument?id=2057415, 21 June 2012.

[4] Carlo Strozzi. NoSQL: A Relational Database Management

System.http://www.strozzi.it/cgi-bin/CSA/tw7/I/en_US/nosql/Home.20 p.

[5] Apache Hadoop. http://hadoop.apache.org/

[6] Apache CouchDB. http://couchdb.apache.org/

[7] Konstantin Shvachko, Hairong Kuang, Sanjai Radia, Robert Chansler. The Hadoop

Distributed File System. MSST '10 Proceedings of the 2010 IEEE 26th Symposium on

Mass Storage Systems and Technologies (MSST), 2010, 1-10 p.

[8] Ajay Anand. Hadoop Sorts a Petabyte in 16.25 Hours and a Terabyte in 62 Seconds.

http://developer.yahoo.com/blogs/hadoop/posts/2009/05/hadoop_sorts_a_petabyte_in_162/,200.

[9] Apache Hive. http://hive.apache.org/

[10] Apache Hbase. http://hbase.apache.org/

CONTENTS

TECHNICZNE NAUKI

Hutnictwo

Branżowa budowa maszyn

Transport

Energetyka

Электротехника і radioelektronika

Górska sprawa

WSPÓŁCZESNE INFORMACYJNE TECHNOLOGIE

Komputerowa inżynieria

Obliczeniowa technika i programowanie

BUDOWNICTWO I ARCHITEKTURA

CHEMIA I CHEMICZNE TECHNOLOGIE

GEOGRAFIA I GEOLOGIA

FIZYKA

234562

235490

235483

235380

235539

234460

234461

235386

235383

235472

234778

235461

235447

235477

235364

235333

235229

235218