Nov. 2016

http://turbo-data.co.jp/en/

1. ZAP-IN Technology

Zap In Memory

【What's required for Big Data Systems】

#1. Fastness
#2. Low storage consumption
#3. Available for many kinds of operations

But existing indexes cannot endure them.

【Problems of Big Data's Index】

Existing Indexes
Hash and B-Tree are widely used.

Problem #1. Indexes are Volatile
Every time data updated, indexes are destroyed.

Problem #2. Impossible to handle subsets
Hash and B-Tree indexes don't applicable to subsets.

Problem #3. Indexes consume memory area

Problem #4. Slowness
Hash needs calculation time.
B-Tree needs O(log(n))'s time to specify a record.

Problem #5. Difficulty in utilizing multi core

Problem #6. Takes much time for creating index combination

Problem #7. Inefficient in Sort or Matching

【We achieved to solve them】

➔ **Solution of these problems**

ZAP-IN's index is derived from its data structure: combination of Ordered-Set, Value-No and Value-List.

Because that index is derived from data itself, that index is not volatile (problem 1). And it can solve problem 2 ~ 7 also.

Every table's every field is given efficient index always.

So BigData's interactive batch processing becomes possible.

【Benchmarks】

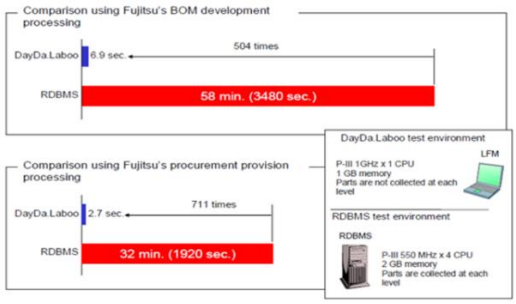

Figure 1. Benchmark at Fujitsu @2001 x500~x700

	Turbo Data Lab. Aktblitz	Apache Spark		Google BigQuery	
	(mSec)	(mSec)	ratio	(mSec)	ratio
Bulk Import	1,803	26,925	x15	73,100	x41
Search	19	495	x26	15,700	x826
Join	658	234	x0.4	10,200	x16
Summary	125	208	x2	3,000	x24
Data Export	1,586	78,973	x50	12,000	x8
Total	4,192	106,839	x25	114,000	x27

Figure 2. Aktblitz runs x25 faster than Spark

【Used at】

Fujitsu's central procurement system 2002-
Major credit card company 2003-
Major chemical industry company's production scheduling 2003-
JAXA (Japan Aerospace Exploration Agency) 2003-2014
JAL (Japan Air Line) maintenance parts management system 2013-
Shanghai big scale POS data analysis system 2013-
And about 200 copies are running until now

【Patent licensees】

Fujitsu BSC	ZAP-IN patents	2007-
NEC	ZAP-IN patents	2007-
SAP	ZAP-IN patents	2014-

Nov. 2016

http://turbo-data.co.jp/en/

1. ZAP-OVER Technology

Zap over Internet

【Existing Trouble】 BigData cannot be processed at the client side.

The merits of processing on client side

Usual internet contents except for BigData are almost always files (html, text, picture, and movie, etc.).

They are downloaded from web servers to clients and decoded by clients.

Then these case's merits are as follows.

Merit 1: server's load becomes small and investments to servers become small.

And clients can look up many servers at the same time,

Merit 2: Mash up of many server's contents is possible.

……

But BigData is too huge to download, so processing should be done on server side, then said merit 1 and 2 are lost inevitably.

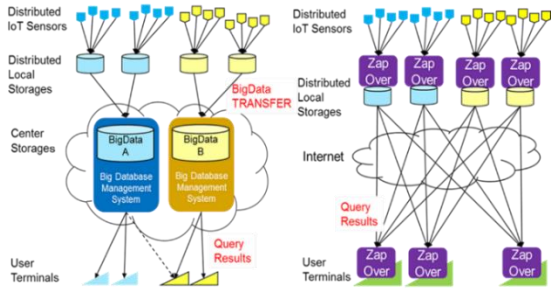

Fig 1. Processing in server (left) and clients (right)

【Accomplishment】

Client side processing of BigData via internet

ZAP-OVER is the technology which enables clients to access simultaneously to many distributed BigData services scattered all over the world via Internet.

Client side can make union and join using plural BigData files on different servers all over the world.

And then Every field of the BigData has efficient built-in index.

【Typical Usage】

At data generation source, make BigData file from original data and put that BigData file on NAS connected to Internet.

From client's site, clients access multiple BigData files stored in NASs simultaneously and mash-up these files and search them using built-in indexes.

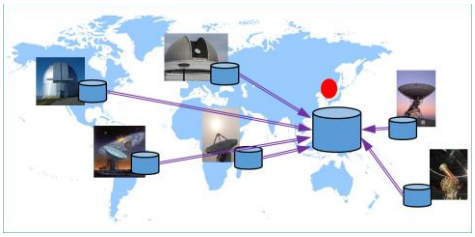

Fig 2. Immediate virtual synthesis of BigData across the earth.

【Used at】

Tokyo National Taxation Bureau 2013-
Cooperating with about 135 countries, it traces traffic of international illegal money exchange.
Before ZAP-Over, it took about 15 ~ 20 min. to get search result for 1 step.
Zap-Over reduced search time down to about 10 sec. (Search speed x100)
And it raised simultaneous user from 1~2 to 10~20.