

Turbo Data Lab. (TDL) は情報量の損失も重複も一切無い形でインデックスとデータを一つに融合したデータ構造を開発した。

それにより、課題#1～#7を全て解決し、新たにビッグデータの対話型バッチ処理を可能とした。

1. ZAP-IN テクノロジ

Zap In Memory

【ビッグデータシステムに求められるもの】

1. 高速。
2. 記憶域の消費量が少ない。
3. 多様な処理ができる。

従来のインデックス技術ではこのような要請に応えることができなかった。

【従来のインデックスの課題の詳細】

従来のインデックス

代表的なものにハッシュと B-Tree がある。

課題 #1 揮発性 (すぐに破壊される)

データの更新の度、インデックスが破壊される。

課題 #3 部分集合に対して無効

インデックスは検索結果等の部分集合には使えない。

課題 #2 インデックスが記憶域を消費する

課題 #4 インデックスが低速

ハッシュは一定の計算を必要とする。

B-Tree は1つのレコードを特定するのに $O(\log(n))$ の処理を必要とする。

課題 #6 マルチコアを活用できない

課題 #5 結合インデックスの作成に時間がかかる

課題 #7 ソートとマッチングに適さない

【当社はインデックスの課題解決に成功】

課題#1～#7は、インデックスとデータが独立の存在として分離していることに起因している。

【ベンチマーク】

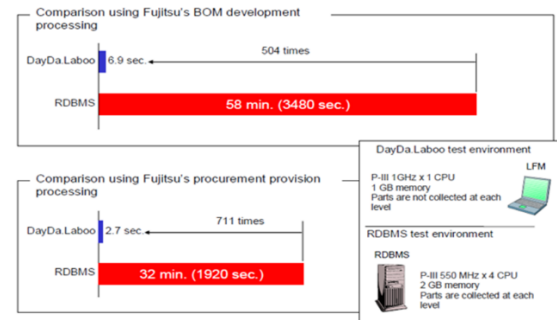


図 1.富士通でのベンチマーク @2001 x500 ~ x700

	Turbo Data Lab.	Apache	Google
	Aktblitz	Spark	BigQuery
	(mSec)	(mSec)	ratio
Bulk Import	1,803	26,925	x15
Search	19	495	x26
Join	658	234	x0.4
Summary	125	208	x2
Data Export	1,586	78,973	x50
Total	4,192	106,839	x25



図 2. Spark や BigQuery より 2.5 倍程度高速

【導入実績】

富士通 (全社集中購買システム)	2002-
大手クレジットカード会社	2003-
大手化学メーカーの生産スケジューリング	2003-
JAXA SODA システム+他	2003-2014
JAL 整備データベースシステム	2013-
上海大規模 POS 分析システム	2013-
そのほか約 200 の納入先	

【特許使用許諾先】

富士通 BSC	ZAP-IN パテント	2007-
NEC	ZAP-IN パテント	2007-
SAP	ZAP-IN パテント	2014-

1. ZAP-OVER テクノロジ

Zap over Internet

【背景】 ビッグデータだけがサーバ側処理

端末側処理の利点 (図 1 右)

これまでのインターネットのコンテンツは、html、テキスト、画像、音声、動画などのファイルであり、Web サーバから端末側ブラウザにダウンロードされ使用された。

端末側処理では、データのデコードは端末が行うので、利点①、サーバ側の負荷は小さく、投資も小さい。

また、端末側は複数のサーバを参照できるので、利点②、複数サーバのデータを融合して利用できる。

ビッグデータはサーバ側処理 (図 1 左)

ビッグデータはデータサイズが大きく、ダウンロードできない。そのためサーバ側で処理するしかなく、上記の利点①、②を享受できなかった。

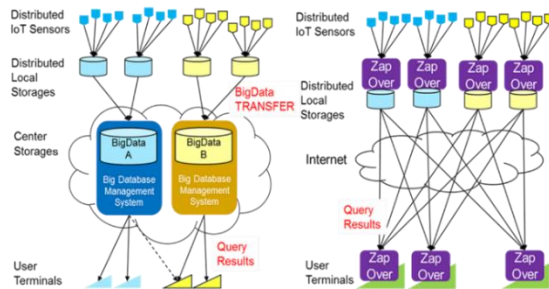


図 1. サーバ側処理 (左) と、端末側処理 (右)

【達成】 ビッグデータも端末側で高速処理

ZAP-OVER は端末からインターネットを介して世界各地のビッグデータにアクセスする技術。

- 1) 端末側から、同時に複数のビッグデータを融合 (JOINやUNION) して利用でき、上記の利点 1 および 2 を享受できる。
- 2) 全てのデータ項目が高速に検索可能

【代表的な使用方法】

1. 世界各地のデータ発生源で、インターネットに繋がった NAS にビッグデータファイルを入れる。
2. インターネットにつながった端末から、上記ビッグデータファイルを任意に選び、瞬時にマッシュアップ (UNION/JOIN)、高速な検索などの操作を端末側処理として行う。(図 2 参照)

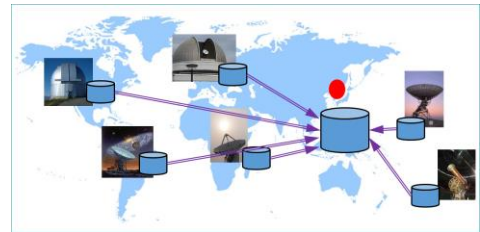


図 2. 各地のビッグデータを仮想的に統合して利用

【導入実績】

- 東京国税局 2013-
世界約 135 カ国と協力し、海外送受金データのトレースを行っている。15-20 分かっていた検索を 10 秒程度に短縮、検索スピードは約 100 倍になった。また、同時利用者数を 1-2 名から 10-20 名と 10 倍に増加させた。