WHITE PAPER

ZAP-IN
SUPER FAST DATABASE TECHNOLOGY

VERSION 1.0

TURBODATA LABORATORIES INC

TABLE OF CONTENTS

0. INTRODUCTION

- Zap-In Technology is a super fast database system developed by Turbo Data Laboratories Inc. This document explains the overview, features, technologies, applications, products and company profile of Turbo Data Laboratories Inc.

1. OVERVIEW

1-1. WHAT IS ZAP-IN TECHNOLOGY

- Zap-In Technology is super fast database technology capable of providing 10 to 100 thousand times faster performance than a conventional database system.
- Processes Big Data of size 32 billion records at lightning speed.
- In addition, it offers reduced system development time functions, high performance business intelligence tools, high cost performance, high speed data indexing, client-server as well as cluster-server application models and parallelization.
- The products were installed in various companies in Japan and have received good evaluation reviews on its high performance.
- The main application areas are data processing, data analysis, procurement management and data ware houses.
- The high performance achievement and offering of advanced features are due to its unique innovative database architecture and algorithms which are designed based on *Liner Filtering Method* theory.
- Zap-In Technology has been licensed to various companies including SAP AG (Germany), NEC, Fujitsu BSC.

1-2. FEATURES OF ZAP-IN TECHNOLOGY

1. OVERWHELMINGLY FAST

- Zap-In achieves 20 or more times higher performance than Spark which is claiming to the fastest database.
- Zap-In is 10 to 1000 times faster than a conventional database system. The processing speed of 100,000 times higher is also achieved for JOIN operation. This performance difference increases with increase in volume data.
- In a real world application, data operation that took one whole night by a conventional database was completed in 1 minute by Zap-In. The performance improvement of this kind, drastically improves the efficiency of businesses. The higher performance along with the support of big data, advanced features and interactive graphical interface increases the efficiency of business in various ways.

2. LARGE SCALE BIG DATA SUPPORT

- A standalone version of Zap-In can support up to 2 billion records, while cluster version can support large scale big data of 32 billion records.
- *Proportional decrease of performance* with increase of data volume is the main advantage of Zap-In , whereas in a conventional database system the performance will *decrease exponentially with increase of data volume.*

3. REDUCED SYSTEM DEVELOPEMENT TIME

- System development time could be reduced (up to 1/10 of conventional database system) because of its efficient auto programming as well as debugging features.
- The macros generated during data operations by user interaction are converted automatically into a program. Without actually writing a program, the Program generation feature of Zap-In , shortens the system development.

In a real world application, system development that took 1 month with SQL programming language was reduced to 3 days using Zap-In.

4. HIGH LEVEL BUSINESS INTELLIGENCE TOOL

- It offers high performance business intelligence functions never before available and provides robust one stop data analysis solutions with the support of interactive batch operations, data cleaning and EXCEL like interfaces.

5. HIGH COST PERFORMANCE

- The performances that could normally be possible only with expensive hardware in a conventional database system could be achieved with small scale hardware using Zap-In. Therefore, can reduce the cost by 1/10th to 1/1000th of other conventional system.

1-3. SAMPLE APPLICATIONS

1. BIG DATA ANALYSIS SYSTEM

- The following objectives are very important for developing a database system that analyzes large volume of business data for strategic planning and business correspondence. Therefore Zap-In is an ideal system under such circumstances.
    - Handle large amount of data (aka Big Data)
    - Processes Big Data at super high speed
    - Easy-to-use、Easy-to-understand
    - Support of high performance functions
    - Flexibility of incorporating the changes in data processing

2. SUPER FAST DATABASE

- The primary advantage of Zap-In is its performance. It is an ideal database system where speed is important and cannot be substituted by other systems. It is suitable for situations where performance degradation caused primarily by volume of data.

3. DATA WARE HOUSE

- In data ware house, performance is vital for analyzing large volume of historical data. The high level functions, super fast performance, indexed architecture, handling large volume of data without the need of data mart are main merits of Zap-In for data ware house system.

4. DATA CLEANSING

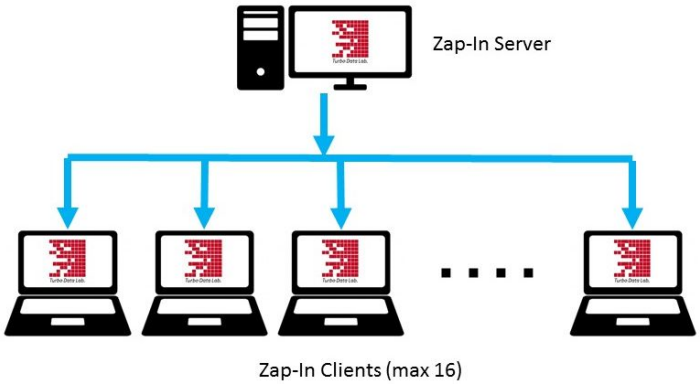
Zap-In offers high speed, powerful and advanced functions for data cleansing.

1-4. COMPARISON WITH OTHER DATABASE SYSTEMS

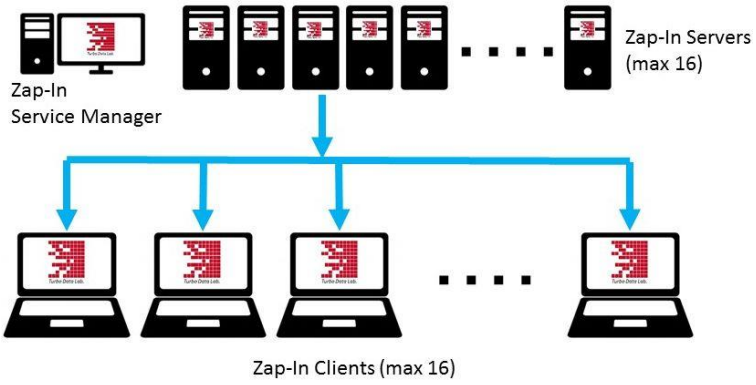
Evaluation Items	RDB(Disk based) Database	RDB(in-memory based) Database	NoSQL Database	Full Text Search Database	Spreadsheet Applications	Turbo Zap-In
Speed	1	10	50	1-10	0.1	10-1000,000
Features	Very Good	Very Good	Average	Average	Average	Good
Usability	Not Good	Not Good	Average	Average	Very Good	Good
Cost				-	Very Good	Good

1-5. FOUR TYPES OF SYSTEM CONFIGURATIONS

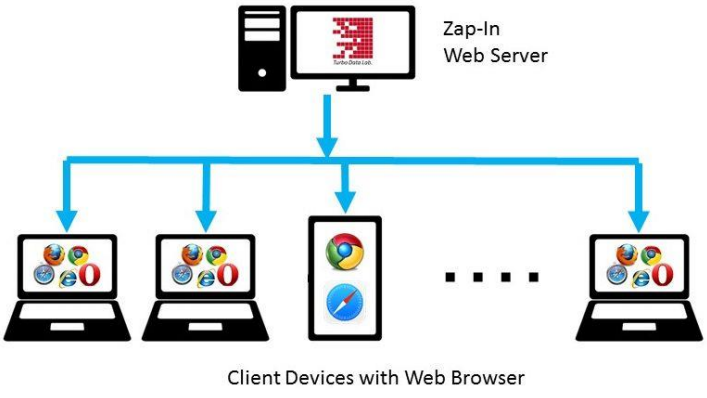
STANDALONE



CLIENT SERVER



CLUSTER・SERVER



WEB SERVICE

1-6. PATENTS

- The super fast (in-memory) database technology of Turbo Data Laboratories Inc., has been patented in Japan and Overseas.
- Number of patents:
    - The patents related to Zap-In Technology
        - Domestic (Japan) - 11
          Overseas - 26
- This technology has been licensed to following companies:
    - ・SAP AG (Germany)
      ・Fujitsu Broad Solution and Consulting, Inc
      ・NEC Corporation
      ・SEC Co. Ltd
      ・Nihon Soft KK
      ・Others

2. TECHNICAL DESCRIPTION

2-1. SUPER FAST DATA PROCESSING

- The main feature of *Zap-In technology* is overwhelmingly fast data processing.
- Super fast Data processing
    - Generally, the processing speed is 10 to 1000 times higher than conventional database system. The processing speed of 100,000 times higher is also achieved in case of JOIN operation.
    - The processing speed of 20 times or more is achieved compared to Spark which is represented as super fast database.
    - The proprietary database based on *Linear Filtering Method (LFM)* technology.
    - The super fast database operation which was not possible until now has become possible due to *Linear Filtering Method* [aka LFM] technology. Zap-In is an unique proprietary database system based on *LFM* architecture and algorithms.
- Super fast data processing possible on Big Data
    - Zap-In can handle maximum of 2 billion records in a standalone mode. The database store the data in main memory rather than hard disk (aka in-memory database) and devised with high efficient data structure. Therefore, it becomes possible to handle huge volume of data with high efficiency . The cluster configuration using 16 servers could handle maximum of 32 billion records.
- The linear relationship between volume of data and process time
    - In a conventional relational database system, by the nature of its property the process time will increase exponentially with increase in volume of data (o(n*log(n)) n-> data volume). Therefore, processing big data is time consuming operation in conventional systems. In contract, Zap-In

technology offers (o(n) n-> data volume) linear relationship between *volume of data* and *process time* and hence very efficient to process Big Data.

- Super fast data Import
    - CSV file import is extremely fast comparing to other conventional database system. Because of its unique data structure, the import or data fetch operations are effectively parallelized depending on number of (multi-core) CPUs.

2-2. COMPARISION WITH OTHER DATABASE TECHNOLOGIES (SPEED)

- Comparing the performances of Zap-In with other widely used database technologies.

Process	RDB(Disk based) Database	RDB(in-memory based) Database	NoSQL Database	Full Text Search Database	Turbo Zap-In	(Comments: Features of Zap-In)
CSV Import	1	10	50	0.05	100	
JOIN	1	10	-	-	1000-100,000	**Fast** Even with high cardinality.
SORT	1	10	-	-	100-100,000	**Fast** Even with high cardinality.
SEAERCH	1	10	100-1000 (When perfect matching)	-	10-1000	**Fast** Even with large volume.
BOM Explosion	1	10	-	-	500-700	
Categorize	1	10	-	-	10-1000,000	Particularly, **Fast O(n)**
Summary /Aggregation	1	10	-	-	1000-100,000	**Fast** Even with large volume

Calculation, Update	1	10	-	-	0.1-10,000	**Slow** when updating the records one-by-one, **Fast** when updating the records collectively
EXPORT	1	10	1 (When perfect matching data is only one)	1-10 (Fetch hit records)	100-1000	**Fast** Particularly for large hits
Full Text Search	1	10	-	100-10000	10-1000	**Fast** Even with large volume

2-3. COMPARISON WITH OTHER DATABASE TECHNOLOGIES (FEATURES)

- Comparing the features of Zap-In with other widely used database technologies.
  - **Zap-In (Super fast in-memory database)**
    - Features:

      Super fast database operation plus high performance functions similar to RDB. It also offers super fast high level full text search functions but doesn't possess transaction features.
  - **RDB(Disk based) Database**
    - Product Examples : MySQL, Oracle Database

      Features : Mainstream database system. It offers many functionalities but performs poorly with big data.
  - **RDB(In-Memory based) Database**
    - Product Examples : HANA(SAP), TimesTen (Oracle), Spark(Apache)

      Features : Faster than Disk based RDB (about 10 times).
  - **NoSQL Database**
    - Product Examples : DynamoDB(Amazon), Cassandra(Apache)

      Features : Fast but has limited functions.
  - **Full Text search Database**
    - Product Examples : Secure Enterprise Search(Oracle), Namazu

      Features : Fast but only full text search functionality.
  - **Batch Processing Database**
    - Product Example :

      Features : Programming is easy and quick but slow in performance.

2-4. REASON FOR HIGH PERFORMANCE IN ZAP-IN TECHNOLOGY

ZAP-IN TECHNOLOGY - REVOLUTIONIZED IN-MEMORY DATABASE

- An in-memory database is a database system that keeps the data entirely in main memory instead of hard disk. Many companies are offering in-memory database systems. The existing database systems could achieve up to 10 times performance improvement by keeping the data entirely in main memory without changing the data structures or algorithms. However, this alone does not guarantee the effective use of main memory for high speed data operation.

- Turbo Data Laboratory has designed a database architecture and algorithms from the ground up so to perform the data operations with high performance in main memory. The data structures and algorithms devised using its proprietary *Linear Filtering method* (*) technology, brings 10 times to 10,000 times performance improvement over existing databases. In some cases, the performance improvement of 100,000 times also achievable. In addition, it also could handle (because of its data structure) maximum of 2 billion records (Cluster version : 16 billion records).

LINEAR FILTERING METHOD (LFM)

- Turbo Data Laboratories has invented its own Linear Filtering Method (aka LFM) to further enhance the efficiency of in-memory data processing. Zap-In Technology is an innovative in-memory database based on data structure and algorithms based on LFM. The LFM and associated algorithms have been patented worldwide.

EFFICIENTLY UTILIZING MULTI CORE CPU SYSTEMS

- Most computers of today are equipped with 4 to 16 CPUs. The performance of database operations could be improved if we could effectively utilize multiple CPUs. To utilize the power of multiple CPUs, the algorithms of database should be designed for parallel execution. The algorithms of

conventional database systems are generally difficult to parallelize and hence it could not effectively utilize the multi-core's advantage. In contrast, by nature of structure and algorithms of LFM, almost all of data processing functions could be parallelized and thus improves the performance. In Zap-In not only database operations but also data import/export operations are parallelized to effectively utilize multi-core systems.

INDEXES IN ALL FIELDS

- Designing efficient indexes is paramount for achieving good performance in any system development. In a conventional database system, designing indexes is time-consuming operation to build efficient high performance system. The fields which are to be indexed should be chosen carefully for high performance. Even after having gone through most careful design process, there are cases to redo whole process again from scratch because of not achieving expected performance or infusing additional new requirements.

  In Zap-In Technology by the nature of LFM architecture, indexes are attached to all fields and hence separate operation to add indexes is not required. Since Indexes are available for all fields, database operations on any field is very fast as well as time-consuming indexes design phase is unnecessary during the system development process. Therefore, development could be carried out in shortest time. Any new requirements during the later part of system development could also be easily accommodated as re-index process is not required.

ALGORITHM

- Zap-In has realized super fast data processing using Linear Filtering Method (LFM) technology. In order to materialize this theory into a software product; various algorithms, peripheral technologies, software

technologies have been developed. Those techniques and algorithms along with LFM have been patented worldwide.

2-5. PERFORMANCE BENCHMARK

2-5-1. COMPARISION BETWEEN SPARK AND ZAP-IN TECHNOLOGY

OVERVIEW

- Performance was compared with Spark, a high speed data processing database system popular for handling large volumes of data.
- Zap-In was 181 times (in some cases) faster when performing import, search, join, aggregation and export operations individually. The total performance improvement of 25 times higher was recorded in benchmarks. The performance difference will increase proportionate with volume of data.

ENVIRONMENT

- - CPU ----------

  Intel(R) Xeon(R) CPU E5-2403 v2 @ 1.80GHz

  cpu MHz : 1201.148

  cache size : 10240 KB

  fpu : yes

  fpu_exception : yes

- - Memory ----------

  MemTotal: 32726272 kB

  MemFree: 28515008 kB

  MemAvailable: 30948488 kB

  Buffers: 884 kB

  Cached: 2568612 kB

  SwapCached: 0 kB

- - OS ----------

  CentOS Linux release 7.1.1503 (Core)

DATA

- ・A salesman_master.csv 1,000 records
  ・B products_master.csv 100 records
  ・C saleslog.csv 6,070,000 recodes
  ・D export.csv 73,000 records

TASKS

- 1. Data Import
  2. Search
  3. Join
  4. Aggregate/Summary
  5. Data Export

Process		Zap-In	Spark	Performance (ratio)
Data Import	1. A Data	13 ms	2,352 ms	x181
	1. B Data	2 ms	150 ms	x75
	1. C Data	1,788 ms	24,424 ms	x14
Data Processing	2. Search	19 ms	495 ms	x26
	3. Join	658 ms	234 ms	x0.4 (*)
	4. Aggregate	125 ms	208 ms	x1.7
Data Export	5. D Data	1,586 ms	78,973 ms	x50 (*)
Total Time		4,192 ms	106,839 ms	x25

2-5-2. PERFORMANCE COMPARISON BETWEEN BIGQUERY AND ZAP-IN TECHNOLOGY

OVERVIEW

- Performance was compared with BigQuery - a popular cloud based massively large data analysis system.
- Zap-In was 826 times(in some cases) faster when performing import, search, join, aggregation and export operations individually. The total performance improvement of 27 times higher was recorded in the benchmark. The performance difference will increase proportionate with volume of data.

ENVIRONMENT

BIGQUERY ENVIRONMENT

- Google Cloud

ZAP-IN ENVIRONMENT

- - CPU ----------

  Intel(R) Xeon(R) CPU E5-2403 v2 @ 1.80GHz
  CPU Clock : 1201.148 MHz
  cache size : 10240 KB
  fpu : yes
  fpu_exception : yes

- - Memory ----------

  MemTotal: 32726272 kB
  MemFree: 28515008 kB
  MemAvailable: 30948488 kB
  Buffers: 884 kB
  Cached: 2568612 kB
  SwapCached: 0 kB

- - OS ----------

  CentOS Linux release 7.1.1503 (Core)

DATA

- ・A salesman_master.csv 1,000 records
  ・B products_master.csv 100 records
  ・C saleslog.csv 6,070,000 recodes
  ・D export.csv 73,000 records

TASKS

- 1. Data Import
  2. Search
  3. Join
  4. Aggregate/Summary
  5. Data Export

Process		Zap-In	BigQuery	Performance (ratio)
Data Import	1. A Data	13 ms	1,000 ms	x77
	1. B Data	2 ms	1,000 ms	x500
	1. C Data	1,788 ms	71,100 ms	X40
Data Process	2. Search	19 ms	15,700 ms	X826
	3. Join	658 ms	10,200 ms	x15
	4. Aggregate	125 ms	3,000 ms	X24
Data Export	5. D Data	1,586 ms	12,000 ms	X8
Total Time		4,192 ms	114,000 ms	X27

2-5-3. PERFORMANCE MEASUREMENT OF ZAP-IN TECHNOLOGY

OVERVIEW

- Measured the performance of Zap-In technology against category/grouping operation.
- Categorizing 100 million records into 10 groups has taken as low as 187 milliseconds.

ENVIRONMENT

- - CPU ----------

  AMD Phenom2 CPU x4 925(4core) @ 2.80GHz

  CPU Clock: 1201.148 MHz

  L1 : 2MB

  L2 : 6MB

- - Memory ----------

  Memory Total: 8 MB (1333MHz)

- - OS ----------

  Windows 7 Ultimate 64bit

DATA

- Data Records 100 million records
- Number of Fields 8

PROCESS RESULT

- Categorize into 10 groups

RESULT

- PROCESSING TIME 187 ms (535 million updates/second)

2-5-4. PERFORMANCE MEASUREMENT OF ZAP-IN TECHNOLOGY

OVERVIEW

- Measured the performances of Zap-In Technology against search, sort, aggregate and update operations.
- Sorting a string field, having 100 million records has taken as low as 5,659 milliseconds.

ENVIRONMENT

- – CPU ————————

  Intel Xeon x2 (2core) @ 3.16GHz

  L1 : 64kB

  L2 : 1MB/1core

- – Memory ————————

  32 MB (1333MHz)

- – OS ————————-

  Windows Server 2008 Standard

DATA

- Data records 100 million records
- Number of Fields 8 Fields
- Field 1: Integer, 100 million variations, Sequential

  Field 2: Integer, 10,000 variations, Random

  Field 3: Integer, 100 variations, Random

  Field 4: Integer, 100 variations, Random

  Field 5: String, 100 million variations, Random

  Field 6: String, 100 variations, Random

  Field 7: Double, 100 million variations, Random

  Field 8: Numeric(38 digit), 100 million variations, Random

TASK

- Process 1. Search : String (Unique and random) .
  Process 2. Sort : Fields 5 (Unique and random strings) .
  Process 3. Summary : 1 Dimension (100 variations), 1 Measure (Unique and random Double value).
  Process 4. Update : Overwrite 1 million cells.

RESULT

Data Process (100 million records)	Zap-In
1. Search 10 hits	0 ms
1. Search 10,000 hits	0 ms
1. Search 1 million hits	7 ms
2. Sort	5,659 ms
3. Summary	9,312 ms
4. Update/Overwrite 1 million cells	10,784 ms

2-6. AUTO PROGRAMMING FEATURES

BUILDING THE SYSTEM IN A VERY SHORT TIME

- Zap-In technology offers auto programming using its macro recording functionality. The spreadsheet like mouse operations performed on database are simultaneously recorded and then recoded operations are converted in to Python program. Thus, able to build a system without actually writing the program. Because of Python being a procedural language, it becomes possible to simply convert the operations in to program.

- In addition, the results generated from the database processing could be verified (correct or not) instantly using graphical display. Thus effectively shortening debug time and improve efficiency. In a case study, system construction that generally take 1 month had been shortened to 3 days using Zap-In Technology .

ADOPTING PYTHON

- In RDMS, SQL is commonly used program for building a system. SQL is non-procedural language, therefore writing a program without using loops is very difficult. In addition, as the variables and functions are not exposed outside, debugging the system is also cumbersome.

- In Zap-In Technology, above mentioned issues are resolved by adapting the Python (which is a procedural language). Python is a convenient, reliable and high level programming language incorporated with rich library functions. The Python scripts could also be executed using job schedulers.

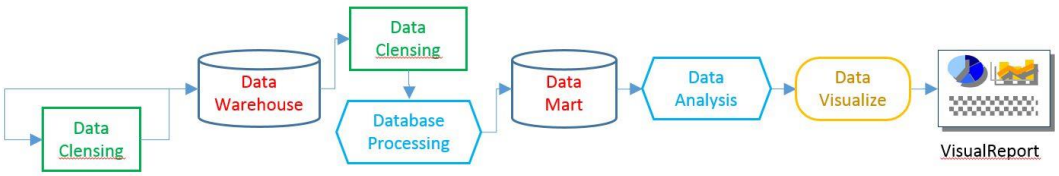
2-7. HIGH PERFORMANCE BUSINESS INTELLIGENCE TOOLS

2-7-1. WHAT IS BUSINESS INTELLIGENCE

- Business Intelligence is a process of analyzing data for helping the corporations to make actionable business decision. It helps the decision making process or grasp the situation by collecting the valuable information using operations such as search, multi-dimensional analysis (also called Data Mining) and present them in a user understandable graphical format.

- To retrieve valuable information using search or analysis, the operations should be performed with repeated trails. Therefore system usability, support of multiple functions and processing speed are very vital.

- For such an analysis - consolidating/correcting the original data (also called data cleansing), storing the consolidated data into a large database (also called data warehouse) and then slicing the larger database into smaller easy-to-use databases (also called data mart) are involved. Then finally perform the data analysis using variety of techniques.



2-7-2. BUILDING THE DATA ANALYSIS SYSTEM (AS MENTIONED ABOVE) CAN BE ACHIEVED USING FOLLOWING METHODS

- **Using Conventional Database System**

  Data analysis is fast because of supporting multiple high level functions. However, an expert is required to write data analysis program .

- **Using Business Intelligence Tools (BI Tools)**

  Data analysis can be performed using graphical operations. However, volume

of data that could be handled is limited and also performance, analytical functionalities are not sufficient.

- **Using Spread Sheet Software (such as EXCEL)**

  Data analysis can be performed easily. However, volume of data that could be handled is limited and also performance, analytical functionalities are not sufficient.

- **Using Zap-In Technology**

  Data analysis can be performed at very high speed using rich functions. The analysis can be achieved using EXCEL like operations. The data mart is unnecessary even for handling huge data. It offers powerful data cleansing functions.

2-7-3. COMPARISION OF BI FUNCTIONS

Process	Conventional database	BI Tools	Spread Sheet Software (such as EXCEL)	Zap-In Technology
Data Cleansing	Yes	No	No	Powerful
Data Analysis	Powerful	Cannot cope with large varieties of values	Impossible for operations such as JOIN	Powerful
Operation	Using SQL (Depends on Expert)	Easy for Graphical Representation	Easy for Tabular Representation	Easy for Tabular Representation
Programming Process	SQL Program (Depends on Expert)	Unnecessary (or depends on Expert)	Unnecessary (Possible to certain extent)	Auto Programming(Expert is not required)
Process Speed	High Speed	Low Speed	Low Speed	Super High Speed
Huge Data	Manageable	Cannot Manage	Cannot Manage	Manageable
Data Mart	Necessary to handle huge data	Necessary	Necessary	Not required even for huge data
visualization and graphics	Using Visualization tool	Powerful	Powerful	Connect with EXCEL for visualization

2-7-4. BUSINESS INTELLIGENCE FUNCTIONALITIES IN ZAP-IN TECHNOLOGY

Zap-In Technology offers excellent Business Intelligence functionalities.

- **High Business efficiency using One-Step Process**

  The operational flow – *data cleansing, data analysis and visualization of analysis report* are achieved smoothly.

  Data cleansing is essential for data analysis but most of the BI tools lacks effective data cleansing features. Zap-In Technology offers many high-speed data cleansing functions. As it can handle big data with high speed, building d*ata marts* became unnecessary. The excel integration feature allows to visualize the analysis output of Zap-In using rich functionalities of EXCEL.

- **High business efficiency due to Interactive data processing on Big Data**

  Generally, conventional BI tools take longer time (difficult to be called interactive) to process big data and hence reduce the business efficiency of the enterprise. In contract, The BI functionalities of Zap-In can perform the data operations interactively at super high speed (even on big data) and thus significantly improve the business efficiency.

- **Rich data handling functions, yet High Speed**

  The data conversion operations (such as unit conversion, category assignment, etc.,) are generally considered to be slow or inefficient in BI tools. Although, BI tools are popular for its interactive graphical visualization, it is unrealistic to plot customer registration information even for a database having 10,000 variations, the system becomes unresponsive for any click operations inside the graph.

  In Zap-In, data transformations using fact and dimension tables could be achieved instantly. Excel-like calculations, categorization and data conversion could also be performed at lightning speed. Zap-In technology product (Aktblitz Ⅲ) offers tabular display and provides easy-to-use graphical interface for users to perform the operations at high speed.

2-8. HIGH COST PERFORMANCE

Similar performance could be achieved with small scale hardware. Therefore, can reduce the cost by 1/10th to 1/1000th of the other conventional systems.

2-9. TECHNICAL DETAILS

2-9-1, ABOUT LINEAR FILTERING METHOD

1. OVERVIEW OF LINEAR FILTERING METHOD

LFM(Linear Filtering Method) is an innovative database technology developed by Turbo Data Laboratories, Inc. The number of Japan and Overseas patents have been acquired on this technology.

Originally, the followings are expected from a database system to manage big data.

- Securely update the data
- Handle huge volume of data
- High speed data processing
- Suitable for Data analytics
- Compute complex calculation using batch operation
- Suitable for distributed environment

Thus LFM technology is well suited to serve the pressing needs of the database system.

2. FEATUES OF LINEAR FILTERING METHOD

Linear filtering method offers following advantages.

1. Computational Complexity O(n)

The computational complexity , for almost all the algorithms could be reduced to O(n) (where *n* is number of records). Since the computational complexity of conventional database system is O(n*log(n)), improved performance could be achieved in the rate of order. This is one of the very important feature when dealing with big data.

2.High speed indexes and high performance functions

The indexes (creating the indexes for selected fields) are key to the high performance as well as efficiency of the database. In LM technology, various issues related to indexes have been resolved, accomplished indexes on fields even under repeated data processing operations. Demonstrated an overwhelming performance during multi-stage BOM process.

3.Parallel Processing

The parallelization is possible almost on all data base functions. Thus taking the advantage of multi core/multi CPU systems (which became very common in recent times) for performance improvement.

2-9-2. INDEXES IN LINEAR FILTERING METHOD

1. OVERVIEW OF DATABASE INDEX

Database index is the common and most efficient method of increasing the performance of the database processing. The *B-tree* or *Hash-index* are the most common type of indexes.

Adding indexes to the fields, significantly improve performance of data operations such as search. However, updating or modifying the data on field alter indexes and hence leads to performance degradation. Therefore, it is a common practice to build indexes only for few fields.

Designing indexes are very important task as it directly impact the performance of database system. Most importantly, *which fields to be indexed* should be carefully considered for balancing the performance of the system. Any changes later in the design is very complex, time consuming and requires long term redevelopment.

2. INDEXING IN LINEAR FILTERING METHOD

The *Linear Filtering Method* technology used in Zap-In/Zap-Over has natural way of incorporating indexes in data structure and hence all the fields are build inherently with indexes. Therefore, data processing can be performed at high speed on any field. Since, a separate *indexes designing phase* as required in conventional database is unnecessary, system development is very efficient and simple.

In addition, the new fields generated from the result of data processing results are also having naturally build indexes. Therefore, high speed data processing is also possible on generated fields. Devising re-index or index is not required at any point of time in LFM.

3. COMPARISION OF EACH METHOD

The database *processing speed* comparison has been carried out using following 3 methods:

- Hash Index method

- B-Tree Index method
- *Linear Filtering Method* (LFM)

In each of comparison, Linear Filtering Method performs better than other methods and found that LFM has good advantage over other methods when performing on large volumes of data. In addition, LFM has performed well on multi-core CPU systems because of its multi-thread support.

Comparison of Index methods

Index Method	Hash	B-Tree	Liner Filter Method (p is number of threads)
Indexed Fields	Primary Key Only	Few Fields	All Fields
Index Creation Time	Not Small	Long	First time Only
Required Memory	O(n)	O(n*log(n))	None
SEARCH	o(1) Exact match, 1 item only	o(m*log(n)) *m*-hit counts	During Default O(n/p) During High Speed O(log(n)/p)
SEARCH (Inside Subset) *s* - size of subset	O(s) Exact match Only	O(s)+O(m*log(n))	O(s/p)
SORT	Not applicable	O(n*log(n))	o(n/p)

SORT (Inside Subset) s - size of subset	Not applicable	O(n*log(n))	o(n/p)
UNION n1, n2 are size of tables	Not applicable	Not applicable	o((n1+n2)/p)
UNION (Subset) s1, s2 are Subset sizes	Not applicable	Not applicable	o((s1+s2)/p)
JOIN n1, n2 are Table sizes	O(n1) (SEARCH JOIN)	O(n1*log(n2)) (SEARCH JOIN)	o((n1+n2)/p) (SORT JOIN)
JOIN (Subset) s1, s2 are Subset sizes	Not applicable	Not applicable	o((s1+s2)/p) (SORT JOIN)
Summary	Not applicable	Not applicable	o(n/p)
Summary (Subset)	Not applicable	Not applicable	o(n/p)
Updating Index	Sequential	Difficult	Not required

3. APPLICATIONS : EXAMPLES OF SYSTEM DEVELOPMENT

3-1. A TYPICAL EXAMPLE OF SYSTEM DEVELOPMENT

- Zap-In offers super fast data processing functionalities by itself or integrating with other systems.

  This section introduces typical examples of *system development* that employs super fast, high level functions and special features of Zap-In.

1. SUPER FAST DATABASE FOR DATA ANALYSIS

- The process that could take one whole day by a conventional database system could be finished in few minutes by Zap-In. Similarly, few minutes operation could be completed in few seconds.

  The shortening of process time could impact the business efficiency in the following ways:

  - Monthly reports [due to time constraint] can be changed to daily reports, that will revolutionize the business analysis and management decision making processes.
  - Reducing the waiting time (from minutes to seconds) for analytical output could drastically improve the operational efficiency of data analyst.

2. BIG DATA SYSTEM

- In a conventional database system, the performance of speed will drastically decrease for large volume of data. However, as processing speed is already high in Zap-In, it becomes possible to build *big database system* which was traditionally given up by other conventional system due to performance bottlenecks.

3. DATA CLEANSING APPLICATION

- Data cleansing is a process of detecting, normalizing and correcting or removing the duplications, errors, inaccurate notations etc., to improve the quality of database. This process is very essential for any database system.
- Zap-In offers super fast powerful data cleansing functions. The data cleansing could be performed either with Interactive operation (confirm and correct the data in real time) or with automated auto-checking/ cleansing batch operation.

4. DATA WARE HOUSE APPLICATION

- Data ware house is a system that analyze large volumes of historical data to guide decision making processes. As a storage place for organized data, it is expected to support variety of data processing requests on large volumes of data.
- Comparison of Zap-In and conventional Data ware house system
  - **Conventional- Data ware house system**

Hardware :	Build with dedicated System having many modules(CPU, Storage)
Cost :	Very Expensive
Performance of Complex Operations :	Low Performance (Because of data referencing across many modules)

  - **Zap-In - Data ware house system**

Hardware :	Possible with Generic PC Server
Cost :	Less Expensive (1/10th of conventional data ware house)
Performance of Complex Operations :	High Performance

5. DATA ANALYSIS APPLICATION

- Zap-In is a useful application for analyzing Big Data.

- In *Spreadsheet applications*, manipulations on tabular data are easier to perform. However, it cannot handle large volume of data and as well cannot support complicated operations such as JOIN. The *BI tools* are powerful in plotting graphs, however it does not have high performance data processing functions. In case of c*onventional database* application, experts are required to write program to handle data processing.

  In contrast, Zap-In offers high level manipulation functions for tabular data, support of large volumes of data, super fast data processing, auto programming features and support of big data. The Zap-In can also be connected to spread sheet applications such as excel to visualize and plot output results in graphs or charts.

6. DATABASE SERVER FOR WEB SERVICE APPLICATION

- The web service applications such as e-commerce are required to handle large volumes of product data. The processing time, turn-around time, high performance functions are key to any web service. Offering high level search *operations as a web service* is challenging task, customers are often required to wait long to get response

- By adopting Zap-In (in server side database), web services could utilize its super high speed performance and high level functionalities.

3-2. CASE STUDIES

1. NATIONWIDE FRANCISE COMPANY (A COMPANY)

SALES, INVENTORY MANAGEMENT SYSTEM

- System Overview
    - This system quickly grasp changes in item, arrange the items according to demand or popularity and initiate the storage of items in each branch. In addition, disposing surplus items and dynamically determining prices of items at the time of sale. Thus this system helps to keep suitable items of optimum quantity at each stores.
- Shortened the monthly batch processing by several 10 times
    - A monthly process that took 20 days was shortened to 6 hours.
    - Reduced monthly processing time enables to correspond other issues promptly.
    - The decision making speed of each person-in charge has been significantly improved.
    - Failing to observe or overlooking early procurement of necessary items for peak season have been avoided.
    - It became possible to dispose unsold goods that have been occupied in the shelf at an early stage.
- Substantially reduced development time and cost
    - The system was developed in a short time with no errors by employing interactive development tools and using actual data.

      (Due to Zap-In automatic programming feature)
    - Using existing hardware, deploy similar batch operations effectively.

2. CHEMICAL INDUSTRIAL PRODUCTS MANUFACTURING COMPANY (B COMPANY)

MANUFACTURING COST MANAGEMENT SYSTEM FOR 4 FACTORIES

- System Overview
    - In a cost management process, BOM (Bill of Materials) calculation for a chemical component is very complicated and time consuming during manufacturing activities of chemical industrial product. Therefore, only rough calculations were possible until now.
- Achieved very precise calculation
    - Zap-In performed the calculations hundreds of times faster and hence very precise calculations could be accomplished at a short time.
      As a result,
      ・Over production components that existed during the rough calculation were suppressed to near Zero.
      ・Able to achieve precise cost calculations by refining the calculation of sub components. The near perfect cost calculation enabled customer to quote most accurately than their competitors and handled orders without any deficit.
- Reduced batch processing time helps to transform processes from monthly to daily
    - In a monthly process, the old system used to take 4 days to perform rough calculation. This process time has been reduced to 4 hours even with complete calculation. Therefore once in a month aggregation process has been changed to daily aggregation process .
    - By understanding operational status of factory up to the previous day, it becomes possible to incorporate every day plan modifications and decision-makings without having to wait until the end of the month.
    - Because of shortened process time, it became possible to handle urgent orders from customers .

- Because of shortened calculation time, it became possible to reflect the various cost changes into latest cost calculation.

- Information utilization
  - The *"contingency that is too late to handle"* had been occurred many times with monthly data aggregation processes, however those situations can be dealt easily with daily processes.
  - By understanding detailed product cost, it became possible to provide meticulous support and thus greatly improved the business efficiency.

- Substantially reduced development time and cost
  - The system was developed in a short time with no errors by employing interactive development tools and using actual data.

    (Due to Zap-In automatic programming feature)
  - Migration from legacy system to open system became possible and thus reduced the costs of both hardware and software.

3. NON LIFE INSURANCE COMPANY(C COMPANY)

CUSTOMER MANAGEMENT SYSTEM

- System Overview
  - In insurance systems, apart from having very large volumes of data, the customer information also changes very often. So, master data could not function properly. In such situation - the popular method to remember customer information (deposit, payment, etc.,) before and after the transactions is by storing the new information along with data. By doing so, the increase in fields and records are inevitable and analyzing such large tables are not an easy task.

In Zap-In, indexes are naturally embedded in all fields , therefore high speed search along with inherently fast aggregation/summary functions helps to discover information which was not possible until now.

- Information utilization
  - There were occasions where information processing could take several days because of some ad-hoc requests. However, because of high efficiency, Zap-In could handle any ad-hoc requests and provides the result with high accuracy.
  - Thus, improved the speed and increased the efficiency of customer decision making process.
- Substantially reduced development time and cost
  - Reduced development time and cost by employing interactive development process on real data.
    (Due to Zap-In automatic programming feature)
  - The information *whether wanted or not* can be verified on working spot, therefore error-free processing result can be achieved without much redoing.
  - Suppressed the new investments by off-loading online batch processes from host system.

4. GENERAL TECHNOLOGY TRADING COMPANY (D COMPANY)

FINANCIAL MANAGEMENT SYSTEM

- System Overview
  - Detecting the inconsistencies are very critical for huge transactions. Therefore, functions such as search, aggregation, matching, categorization, etc., should be performed at high speed to detect inconsistencies in the transactions.

- Reduced the cost by abolishing special purpose machines.
- Migration from special purpose machine to open system became possible and thus reduced the cost of hardware, software, system development, operational maintenance, etc.,
- Migrating the system operation from *fixed operation* to *flexible operation* helped to cope up quick and instant changes.
- Able to respond quickly for changing business needs and thus improved the decision making process.

- Substantially reduced development time and cost
  - Migrating the old system to zap-In by employing interactive development process on real data without writing any program .

    (Due to Zap-In automatic programming feature)
  - Using interactive development on real data, system migration was achieved smoothly without any errors.

5. MAIL MARKETING COMPANY(E COMPANY)

CUSTOMER MANAGEMENT SYSTEM

- System Overview
  - For mail marketing companies, *hit rates* of direct mail impact the success and failure of their business. In their analysis not only aggregation but also functions such as join, matching, search and categorization are also needed to perform simulation. The efficient analysis is possible only with high performing functions.
  - Improved the sales by increasing hit rates of direct mail.
  - The persons-in-charge were able to add new features to system by utilizing their work experience and business skills.

    (Due to Zap-In automatic programming feature)

- The business advantage was achieved using database as a information utilization tool.
- The direct mail hit rate improved 3 to 5 times.

- Substantially reduced development time and cost
  - Comparing to conventional system, able to develop a system with significantly at low cost and in short time.

6. TELECOMMUNICATION EQUIPMENT MANUFACTURING AND SALES COMPANY (F COMPANY)

PROCUREMENT MANAGEMENT SYSTEM

- System Overview
  - The procurement systems in largest manufacturing industries, in most cases, have been managed separately at each factories. If you could centralize this operation from headquarters, we could achieve the cost performance of 10% as stated below:
    - Components inventory can be reduced.
    - By consolidating components, procurement can be greatly reduced.
    - Selecting cost effective components while maintaining the quality.
    - Information utilization cost can be improved.
  - With the addition of user specific new functions, able to achieve the typical information utilization.
    (Due to Zap-In automatic programming feature)
  - Migration to open system became possible and thus reduced hardware, software, operational and outsourcing costs.
  - The paperless model is achieved due to latest and quality information.
    (Report output was reduced by 80%)

- Quick and accurate decision making became possible by leveraging the detailed procurement information from components changes, price negotiation etc.,

4. PRODUCT INTRODUCTION

4-1. ZAP-IN SERIES AKTBLITZ Ⅲ PRODUCT OVERVIEW

Operating Environment

OS: Windows 7/8/10, Linux
Disk: 100GB or more recommended

4-1-1. AKTBLITZ Ⅲ STANDALONE EDITION

- Equipped with Powerful Graphical User Interfaces.
- Type S :
  - OS: 32bit/64bit
    RAM : Up to 2GB (Generally up to 10 million records)
    Threads : Up to 4
    JOB Count : Up to 16
- Type L :
  - OS: 64bit
    Data Upper Limit: 2 Billion records
    Threads : Up to 16
    JOB Count : Up to 999

4-1-2. AKTBLITZ Ⅲ CLIENT SERVER EDITION

- Distributed Network Architecture wherein Server is responsible for performing data processing and Client is responsible for presenting the data in usable form through Graphical User Interface (Multiple).
- Type S : (Server Side)

- OS: 32bit/64bit
  RAM : Up to 2GB (Generally up to 10 million records)
  Threads : Up to 4
  JOB Count : Up to 16

- Type L : (Server Side)

  - OS: 64bit
    Data Upper Limit: 2 Billion records
    Threads : Up to 16
    JOB Count : Up to 999

4-1-3. AKTBLITZ Ⅲ CLUSTER SERVER EDITION

- In Client-Server version, maximum of 16 server units can run in parallel.

- Type L : (For 16-Server machines)

  - OS: 64bit
    Data Upper Limit: 2 Billion records
    Threads : Up to 16
    JOB Count : Up to 999

4-2-4. AKTBLITZ Ⅲ WEB SERVICE EDITION

- The Server configured as a web service for Client-Server System. It can be accessed through web browsers from Client PC, Smart Phone etc.,

- Type S : (Server Side)

  - OS: 32bit/64bit
    RAM : Up to 2GB (Generally up to 10 million records)
    Threads : up to 4
    JOB Count : up to 16

- Type L : (Server Side)

- OS: 64bit
  Data Upper Limit: 2 Billion records
  Thread : Up to 16
  JOB Count : Up to 999

- Client Terminal
  - PC equipped with Google Chrome, Microsoft IE etc., and mobile terminals such as Tablets.

4-1-5. AKTBLITZ Ⅲ ENGINE INTEGRATED MODULE VERSION

- This module integrates AktblitzⅢ functionalities inside the customer's software system/application. It offers powerful APIs for integration.
- Type S :
  - OS: 32bit/64bit
    RAM : Up to 2GB (Generally up to 10 million records)
    Thread : Up to 4
    JOB Count : Up to 16
- Type L :
  - OS: 64bit
    Data Upper Limit: 2 Billion records
    Thread : Up to 16
    JOB Count : Up to 999

4-2. ZAP-IN SERIES AKTBLITZ Ⅳ (TENTATIVE NAME)

- The product that performs super fast data processing on records larger than 2 billion, an upper limit of AktblitzⅢ. This product is under planning stage.

5. COMPANY PROFILE

- Business Philosophy
    - Mission
        - Introduce ground-breaking innovations in database technologies that could broadly assist worldwide growth of Big Data.
    - Vision
        - We will continue to develop ground-breaking innovative Big Data technologies.
        - With the power of ground-breaking innovative technology, we strive to reach worldwide market by creating products that are helpful to the society.
- Company Name
    - Turbo Data Laboratories, Inc.
- Home Page
    - http://turbo-data.co.jp
- Head Office
    - Herios kannai Building 4F,
      3-21-2 Motohama cho,
      Naka-ku Yokohama, Kanagawa Prefecture 231-0004
      Telephone : 045-222-8826
- Nagareyama office
    - ESS Bldg., 61,
      Maehirai, Nagareyama-shi,Chiba, 270-0152
      Telephone: 047-151-0398 FAX: 047-115-8512

- Corporate Officers

  Representative Director & President: Shinji Furusho
  Development Board Director : Jahir Hussain
  Outside Board Director : Ogasawara kuniyoshi
  Outside Board Director : Kawakami Tatsu

- Equity Capital
  - 200 million yens
- Establishment
  - 28th Aug 2000

5-2. SALES PARTNERS

- IT Value Associates, Ltd. http://www.itvalue.jp/
- Web I Laboratories, Inc. https://www.webi.co.jp/
- Esperant System Co., Ltd. https://www.ess-g.com/
- Cross Cat Co., Ltd. http://www.xcat.co.jp/
- Seeds Co.Ltd. http://www.kkseeds.com/
- Systems Engineering Consultants Co., Ltd. http://www.sec.co.jp/
- Needs Ltd. http://www.needs-inc.net/
- Nisseicom, Ltd. http://www.nisseicom.co.jp/
- Nippon Computer Kaihatsu Ltd. http://www.nck-tky.co.jp/
- Nippon Software Knowledge Corp. http://www.nihonsoft.co.jp
- HIMACS, Ltd. http://www.himacs.jp/
- BaySoft Ltd http://baysoft21.co.jp/

