

ETL & REPORTING CAPABILITIES

At BitWise we understand that a company's profitability hinges on its ability to act swiftly and make sound business decisions, based on a complete and accurate single view of customers, suppliers, and operations. However, the critical information required to gain this 360° view and make these key decisions is scattered throughout the enterprise, across multiple applications, departments and divisions. And while each source system contains pieces to the puzzle of enterprise profitability, each is an island of raw data with a gulf of incompatibility separating them.

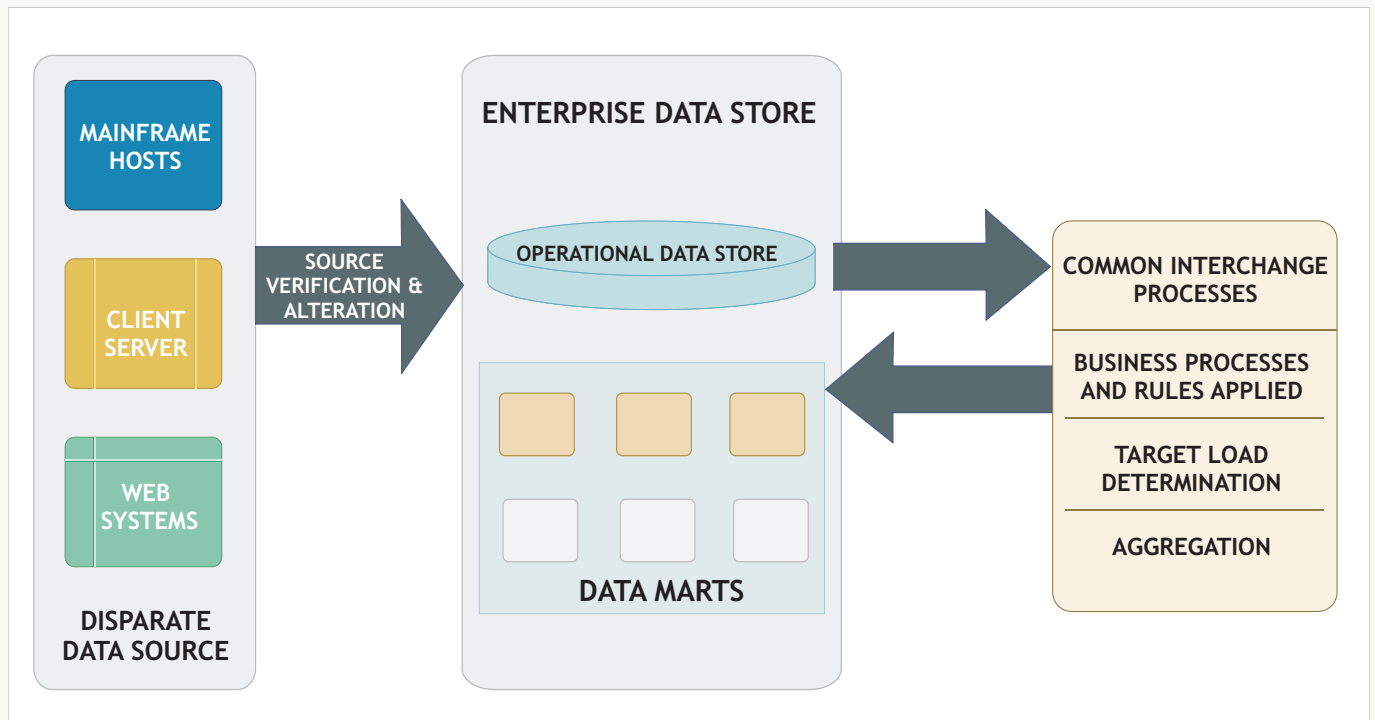
This paper describes some of the ETL processes BitWise has implemented to bridge these islands of raw source data to create Information using approaches that eliminate proliferating, distinguish when applying source-specific and non source-specific business rules and phased transformations.

MERCHANT DECISION SUPPORT SYSTEM

Our Client is one of the largest Credit Card companies in the USA, both as an acquirer of card transactions as well as a card issuer. Its customer portfolio has about 50 million Card Members and more than 4 million Merchants across the USA.

Merchant Decision Support System (MDSS) is one of the many systems designed and developed by BitWise to help our Client as an organization to take efficient decisions for ongoing and future activities to improve business with merchants. The system helps in monitoring the relationship progress with the merchants based on the information derived from data. Detailed information about the merchant is also used in identifying fraud and the level of risk associated with the merchant.

ETL ARCHITECTURE



PERFORMANCE CONSIDERATIONS

Database

- Database design was based on star schema so as to have the minimum traversal through tables and improve query performance.
- The output tables were partitioned so that the data could be loaded and fetched in parallel.
- Indexing the tables helped in scanning, sorting and join operations performed during the ETL process.
- Stored procedures were executed separately to perform a batch of data base operations.
- Since the row count to be loaded into a single table was in millions to maintain the atomicity, a replica of the original table is loaded first in the process. Auditing is done using this replica table to make sure data is consistent with the source. This replica is merged with the master table where the history is maintained. As this is a logical move provided inherently by RDBMS, it doesn't require any separate process to be invoked and is faster then the actual loading process.

Data Handling

- During the ETL process, huge files were processed multi-way to speed up the execution resulting in efficient throughput.
- I/Os were reduced by performing in-memory caching of records for processing similar group of records.
- Using checkpoints helped in eliminating redundant processing of data in case of abnormal termination of process.

CHALLENGES

Some of the key technical challenges BitWise overcame during the development of this system are described below:

- **Source Data Complexity:** Some data sources were more complex than others. Examples of complex source data were multiple record types, bit fields, COBOL OCCURS clauses and packed decimal fields. This kind of data had to be taken care of in the ETL solution since they did not exist in the target data structures

TRANSFORMATION PROCESS

A staged transformation approach was used to optimize data acquisition and transformation from source systems. This approach consisted of five stages, which provided an adaptable transformation process:

- **Source Verification** - In this stage data from the disparate sources is accessed, extracted and verified. Here we also built a temporal view of the data at the time of extraction. The source extract built in this stage was included in backup process of the entire batch cycle for reload purposes and for audit purposes during testing. Audit files provided from the source system, were compared against the extract files to verify items such as row counts, byte counts, amount totals and hash sums
- **Source Alteration** - This stage comprised of a variety of transformations unique to each of the disparate sources depending on the business requirements. Also integration of data from multiple source systems, Cleansing, Standardization, application of source specific business logic and conversions unique to the source systems was done here. The transformed integrated data was then loaded into Operational Data Store.
- **Common Interchange** - The non-source specific business rules and enterprise definitions were applied here on the data loaded in the ODS in the previous stage.
- **Target Load Determination** - At this stage final formatting of data was done to make it ready for loading into the target tables (Data Mart). Here identification and segregation of rows was done for insert versus update.
- **Aggregation** - In this stage aggregate tables were created which improved query performance against the Data Mart.

This approach provided a modular and adaptable means to efficiently load the Data Mart.

Some key considerations have been described below

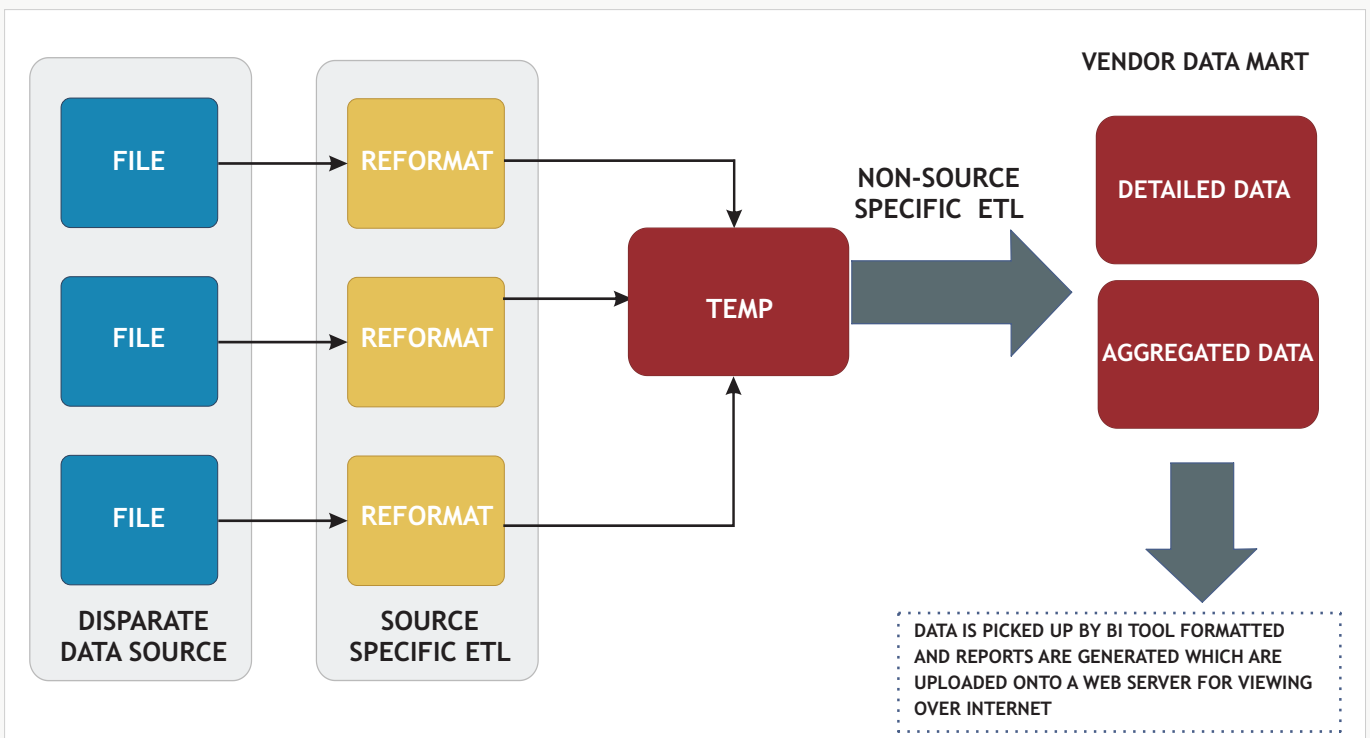
- **Interoperability between Platforms:** Data was being transmitted using a secure file transfer process. The data coming from the Mainframes was in EBCDIC format while from the other sources was in ASCII format. This requirement had to be catered to during the design of the system.
- **Scheduling:** Loading the Operational Data Store involved many source files originating from systems running at different times during the day. Here BitWise used advanced scheduling techniques and options available with Tivoli workload scheduler to meet these complex demands.

VENDOR STATEMENT SYSTEM

One of our clients is one of United States of America's top media companies, operating businesses in broadcasting and publishing. Our client operates 14 leading daily newspapers including a wide range of targeted publications.

For the sale and distribution of their daily newspapers they employ a number of vendors. Each vendor is allotted a certain number of newspapers and areas to distribute. Depending on the distribution volume they are billed on a monthly basis.

ETL ARCHITECTURE



TRANSFORMATION PROCESS

This ETL process revamped the existing Vendor Statement Billing Application to make it more maintainable, modularized, user friendly, cost effective and accessible.

The ETL process that was designed outlined the necessary actions for mapping specific elements to successfully integrate disparate data sources. Source specific ETL processes were developed to transform and decode the data from a format recognizable only to the source system into a standardized format.

This design externalized the Business rules from the repetitive ETL processes being carried out on the entire source data. The business rules were made to be a part of the common interchange ETL process. Any changes to the business logic would mean change only to this common interchange module applying the business rules.

This design also eliminated the need for intermediate files and instead the data from source tables and files were extracted and loaded by source specific ETL routines into temporary tables that were dropped after each successful run. The temporary tables added value to the system by allowing faster data access using index, keys and cursors thereby significantly reducing the run times of the common interchange module.

Next a Non-Source specific ETL (common interchange) process was run on the standardized and cleansed data from this temporary table where in business rules were applied and finally loaded into the data mart with detailed as well as aggregated information.

The loaded data mart is then used up by a BI tool which generates the required reports/statements. These reports/statements are transferred using secure file transfer to a web-reporting server for viewing over the intranet.

PERFORMANCE CONSIDERATIONS

- Indexed temporary tables helped in faster, easy access since it helped in scanning, sorting and join operations performed during the ETL process.
- Since the need for granular data was more than summarized data it meant that the tables would contain redundant data, which in turn meant more storage. Also as the query performance was not a priority as reporting process was to be an overnight batch process the target data mart design was modeled on the Snowflake schema as opposed to the Star Schema which ensured optimal space usage.

- Audit and Checkpoints at every stage of the batch cycle were implemented. With a backup process in place to counter contingencies
- When new sources are added only source-specific actions need to be implemented
- Modularization of the process meant easy maintainability giving room for enhancements at a later stage.

CHALLENGES

As the existing system was to be redesigned the system consisted of two parts.

- The conversion and migration involved design development and execution of a one time ETL process to migrate the existing data to suit the target system. The considerations for this one time process though of low priority owing to it's use and throw nature were:
 - » Overall time taken for the conversion and migration process
 - » Script performance of conversion and migration process
 - » Code documentation
- **Disk Space:** Since the reports were to be transmitted to a web server and the business requirement was to archive reports of past one year with access to last six months reports, this requirement potentially meant lot of disk space and additionally there was some "hidden" disk space to be catered to for staging areas. To counter this BitWise developed a compressing/decompressing utility which was plugged into the system which now required 45% of the originally conceived disk space.

STANDARD PROFITABILITY FORMULA

Standard Profitability Formula (SPF) is used to produce a complete monthly Profit and Loss for each account. This helps to identify, track and analyze company wide programs and policies in terms of Profitability.

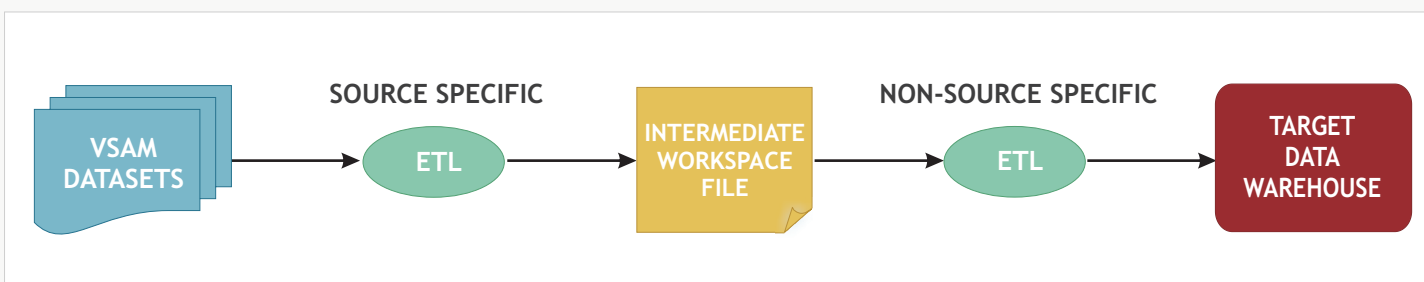
The objectives of the System were multi-fold:

- Track performance of various programs and policy decisions in terms of profitability
- Identify accounts that client wants to reward and modify relationship accordingly

SYSTEM
INTEGRATION
& TEST

- Improve accuracy of assumptions used in financial models
- Use as source of information for predictive modeling to optimize profitability

ETL ARCHITECTURE



TRANSFORMATION PROCESS

The upstream subsystems produce a daily file having account specific information like finance charges, expenses, net revenue and many such fields.

This daily file is accepted by the SPF system running on AIX on a monthly basis. Before the daily file is actually processed by the core SPF Business Modules, there is a need to perform a Source-Specific ETL process so that the mainframes based VSAM files can be reformatted into SPF specific format.

This Source-Specific ETL process performs following activities:

- Conversion from EBCDIC to ASCII format
- Reformat the EDS file into SPF specific format
- Identify records based on transaction types and mark them for specific processing so as to handle large volumes of data
- Data validation and integrity
- Discard bad records from entering into the system

- Merge daily file into one consolidated monthly compressed multi-way file
- Perform an audit on the records processed to maintain accountability

The output of the Source Specific ETL process is a merged file. This file serves as an Intermediate Workspace for a Non-Source Specific ETL (common interchange process) where in the core SPF business rules are applied so as to perform various transformations like filter, route and aggregations and load the output into Oracle based target Data Warehouse.

The Non-Source Specific ETL performs following activities:

- Identify accounts that would qualify for that month's SPF processing.
- Perform account reconciliation process on qualified accounts. Here, multiple transactions for the same account are reconciled into a single transaction for that account.
- Perform business rules so as to create company wide profitability statistical data as well as account based granular data.

PERFORMANCE CONSIDERATIONS

- Indexing the files helped in improved performance operations during the ETL process.
- Multi-way/Parallel batch processing of file helped in reducing the execution time of the process.
- Dealing with large volume of data (5 to 8 million records) requires discarding of unwanted data from the source before processing. This reduced the I/O operations and the overall execution time within acceptable limits.

CHALLENGES

- **Data Dependencies:** SPF receives a feed of about 30 different VSAM files from upstream systems. There is a lot of Data Dependency which needs to be satisfied in order to load the target data warehouse. As Dependencies tend to reduce parallel loading operations, especially if data is from different sources, which are on a different business cycle, It was decided to first load an intermediate work file thereby resolving all the data dependencies, then apply common interchange (non source ETL) to the intermediate work file and load the target data warehouse. This ensured that data dependencies which

tend to make the load processes more complex, encourage bottlenecks and make support more difficult is bypassed.

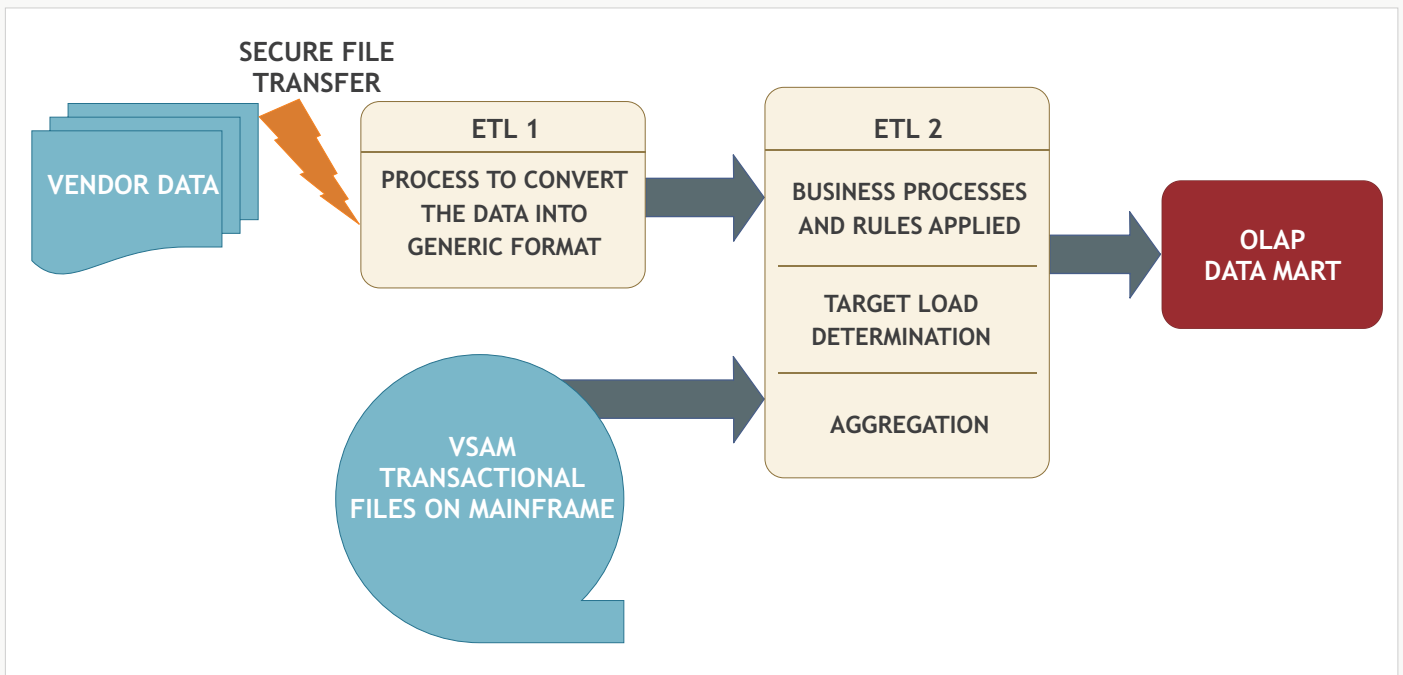
- **Meta Data:** Technical Meta data describes not only the structure and format of the source and target data sources, but also the mapping and transformation rules between them. The system was designed in a manner where the Meta data is visible, accessible and usable to both programs and people.

COMPETITIVE MEDIA RESEARCH

One of our clients is a leading global media company that has operating businesses in broadcasting and publishing too. Our client operates 14 leading daily newspapers including a wide range of targeted publications.

This system was designed to convert the raw data into statistical information and then into reports. These reports were to help the senior management in strategy and planning for increasing market share.

ETL ARCHITECTURE



TRANSFORMATION PROCESS

The Client receives data from an external media intelligence data collecting agency. This data is received monthly in file format through secure file transfer process and contains details of advertisements published in the entire United States by all publishers, across different media. Our Client maintains its own database with which this data can be compared to determine the share of our client against its competitors.

The process begins with extracting the data from external agency's server running on Microsoft Windows Advance Server 2000. The data is transformed so that it can be mapped with client's data that is in DB2 on IBM Mainframe. After applying a chain of transformations and mappings, the summarized data is ready for reporting and loaded into client's Oracle database maintained on Sun Blade Server.

Since the Data Marts were being developed for an OLAP system the ETL process was designed accordingly. OLAP systems use data warehouses/marts that are primarily used for analyzing the transactional data. These are decision support systems (DSS) providing the management with useful and critical information about the business.

While designing the ETL process for this OLAP system following were the considerations:

- ETL process complexity.
- Performance of ETL
- Frequency of execution of ETL process
- Code documentation

PERFORMANCE CONSIDERATIONS

Database

- Various reports were generated using this data mart, which were used by the senior management for decision-making. Keeping in mind the expected quick results and response time Star Schema was best suited for designing the Data Mart.

- Indexing the tables helped in scanning, sorting and join operations performed during the ETL process.
- Batch processing instead of record-by-record processing of data helped in reducing the execution time of the process.

Data Handling:

- Reducing Data Flow: When dealing with 5 to 8 million rows, reducing a single byte from a record helps in improving the performance.
- Eliminating I/O: In memory caching of records by processing similar group of records together helped in reducing I/O operations.

CHALLENGES

- **Similarity of Source and Target Data Structures.** The source systems are owned by two different entities running on disparate software technology platforms. One of the source data coming into the system is a third party vendor providing the captured market intelligence data. And the second is a legacy system running on Mainframe consisting of the daily transactional data. The system needed to match the data from these two systems, apply business rules and load the target data mart which would be used by the reporting system to report the competitive analysis for the Senior Management of the organization.
 - » Since the two sources and the target data structures were not similar, the load processes typically had to parse the records, transform values, validate values, substitute code values and generate keys, etc.
 - » Since the ETL process had complex algorithms and the system was a typical OLAP system used by the senior management, the daily end of day batch ETL process had to complete within the stipulated window. generate keys, etc.
 - » The system had to be scalable to meet the growing volume of data and at the same time maintain it's level of efficiencies that it has been built for to handle current data volumes

CONTACTS ADHOC RETENTION AND TELEMARKETING

We are currently helping a large financial organization in Account Management in Data Mining, Analytics and reporting initiatives in various business areas. These are ongoing request based engagements.

Following is an overview of some of the applications that we have worked on

CUSTOMER CAMPAIGNS

BitWise was involved in the development of Customer Campaign Management which is an automated iterative process which solicits account members for different offers based on offer history in the past.

The system generates the files for all regular, recurring mail campaigns. This process identifies which accounts are eligible for each of the mail campaigns being generated. It also segments the eligible accounts based on the targeting strategy for each respective campaign and creates the files to be printed and mailed.

PROMOTIONAL OFFERS

The purpose of this initiative was to re-establish the ability to offer a promotion for skipping a payment. It was offered to select card members who could skip the monthly minimum payment due while not affecting the credit history but continue to accrue the finance charges.

Skip-A-Pay was introduced with improved profitability and reduced risk as in the past; the variables used to determine who would receive an offer for Skip-a-Pay did not predict customers who were inconsistent in their payment habits. The new model for Skip-a-Pay has stronger prognostic indicators in order to improve the profitability of the process and reduce the risk.

Test and control parameters were designed, developed, and implemented by BitWise in order to validate the option. It was designed to generate incremental finance charges, but not compromise any other revenue streams.

PRODUCT CROSS SELLING

The Client offers Life and Credit Insurance products which were to be offered to the existing customer base. Client did not have customer contact data available to perform Customer Contact Strategy (CCS) related research and analysis. Client also wanted to use such data to eliminate overlap within its outbound telemarketing efforts.

In order for client to build models and perform analysis on the life and credit products BitWise designed and setup a process which supplied the CCS data on a daily extract. The new extract included the call data only.

CARD USAGE PATTERNS USING AD-HOC LIST PULLS—EXAMPLES

- Pull out the list of card members who shopped at client list of merchants published in the previous mail promotion and also segregate them in various card types
- Convenience Revolver Trigger Initiative. This list pull was for the project to offer \$10 Cash back bonus with \$100 purchase as a loyalty mailing to Best Customers. This project was expected to stimulate retail activity and usage.

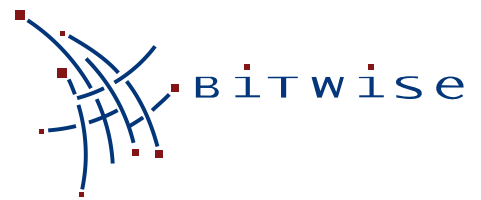
CREDIT LINE INCREASE

Client wanted to judgmentally review accounts for a credit line increases. This process allowed client to leverage additional information from the credit bureau and customer to set the appropriate credit line. In addition client was able to gather data for potential use in future strategy development. As part of this process, enrollment department would attempt to contact those customers that were telemarketing eligible. In order to achieve this BitWise identified the telemarketing eligible accounts prior to providing the file.

TELEMARKETING LIST PULLS AND QUALITY ASSURANCE

The Telemarketing(TM) team of client needs Ad-hoc lists to be produced. These lists are generated using the existing Enterprise Data Stores.

Telemarketing List Management team creates outbound campaigns, each of these campaigns create a Dialer file for TM to contact a card member. Process of creating campaigns, start with a request and requirements from the business sponsor.



BitWise is involved as the Telemarketing-QA (TMQA) team which Quality Assured these list-pulls by running parallel processes and comparing the lists with actual lists generated and provided critical inputs for process improvements

In addition to above BitWise is also involved in

- Profiling and Segmentation
- Delinquency/Charge Off forecasting
- Recovery/Collection Performance trends

BitWise is using Ab Initio for the above engagements across Databases such as DB2/UDB, Oracle, Teradata and Legacy File Systems on various operating systems.



CONTACT INFORMATION

BitWise Inc.
1515 Woodfield Rd.
Suite 930
Schaumburg, IL 60173

Phone : 847-969-1544
Fax : 847-969-1500
Email : info@bitwiseusa.com

BitWise Australia Pty Ltd.
Level 39, 2 Park Street
Sydney,
NSW 2000

Phone : 61 2 9004 7887
Fax : 1300 790 860
Email : info@bitwiseaustralia.com

BitWise Solutions Pvt. Ltd.
BitWise World
Off Int'l Convention Centre
Senapati Bapat Road
Pune - 411016 - INDIA

Phone : 91 20 40102000
Fax : 91 20 40102010
Email : info@bitwiseindia.com

SUPPORTING PARTNERSHIPS

