

Financing Lead Triggers: Empowering Sales Reps Through Knowledge Discovery and Fusion

Kareem S. Aggour
GE Global Research
One Research Circle
Niskayuna, NY 12309, USA
+1 518-387-4047
aggour@ge.com

Bethany Hoogs
GE Global Research
One Research Circle
Niskayuna, NY 12309, USA
+1 518-387-5023
hoogsbk@ge.com

ABSTRACT

Sales representatives must have access to meaningful and actionable intelligence about potential customers to be effective in their roles. Historically, GE Capital Americas sales reps identified leads by manually searching through news reports and financial statements either in print or online. Here we describe a system built to automate the collection and aggregation of information on companies, which is then mined to identify actionable sales leads. The Financing Lead Triggers system is comprised of three core components that perform information fusion, knowledge discovery and information visualization. Together these components extract raw data from disparate sources, fuse that data into information, and then automatically mine that information for actionable sales leads driven by a combination of expert-defined and statistically derived triggers. A web-based interface provides sales reps access to the company information and sales leads in a single location. The use of the Lead Triggers system has significantly improved the performance of the sales reps, providing them with actionable intelligence that has improved their productivity by 30-50%. In 2010, Lead Triggers provided leads on opportunities that represented over \$44B in new deal commitments for GE Capital.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *information filtering, retrieval models.*

H.2.8 [Database Management]: Database applications – *data mining.*

General Terms

Management, Design, Standardization

Keywords

Sales leads, lead generation, commercial finance, knowledge discovery, information fusion

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD'13, August 11–14, 2013, Chicago, Illinois, USA.

Copyright © 2013 ACM 978-1-4503-2174-7/13/08...\$15.00.

1. INTRODUCTION

GE Capital is one of the world's largest providers of commercial credit [1]. Serving over 1 million customers in over 50 countries, GE Capital provides financing in industries as diverse as healthcare, energy, aviation and semiconductor manufacturing, to name a few [2]. Across these and other industries, the majority of GE Capital's commercial loans are made to mid-sized companies across the globe. In particular, GE Capital focuses on industries in which GE also has an industrial presence, enabling GE Capital to draw from internal expertise to help companies large and small effectively manage their businesses.

GE Capital's core products involve equipment leasing and lending, financing equipment as diverse as (but not limited to) commercial aircraft, construction equipment, fleets of cars and trucks, and office printers and copiers. In addition, GE Capital provides significant corporate financing, including financing to enable business mergers and acquisitions and buyout financing.

GE Capital is divided into three main divisions, organized by geographic region: (1) the Americas, (2) Europe, the Middle East, and Africa (EMEA), and (3) Asia-Pacific. Within GE Capital Americas (GECA), the primary purpose of the sales force is to work with potential customers to understand their financing needs and discuss how GE Capital can provide that financing.

Knowing which companies are most likely to be in need of financing in the near future requires deep knowledge of each company in a sales representative's pool of prospects. However, most sales reps are expected to monitor far more prospects than they can reasonably track manually. This makes the automated mining of business data repositories, coupled with the automated delivery of actionable intelligence, invaluable to enable the reps to perform their jobs most effectively.

Information fusion and knowledge discovery capabilities operating across disparate data sets, coupled with effective information visualization, can provide valuable insights and actionable leads to sales reps to help them identify those prospects most likely to require financing. The Financing Lead Triggers system was built to address these needs.

1.1 System Overview

The Financing Lead Triggers system provides access to information aggregated from a variety of internal and external sources for use by sales reps to identify and pursue sales leads. Sales reps view the information through either a web interface or a mobile iPad application. The raw source data includes financial statements, news articles, security (e.g., stock and debt) prices and more. Once fused, sales reps use this information to educate

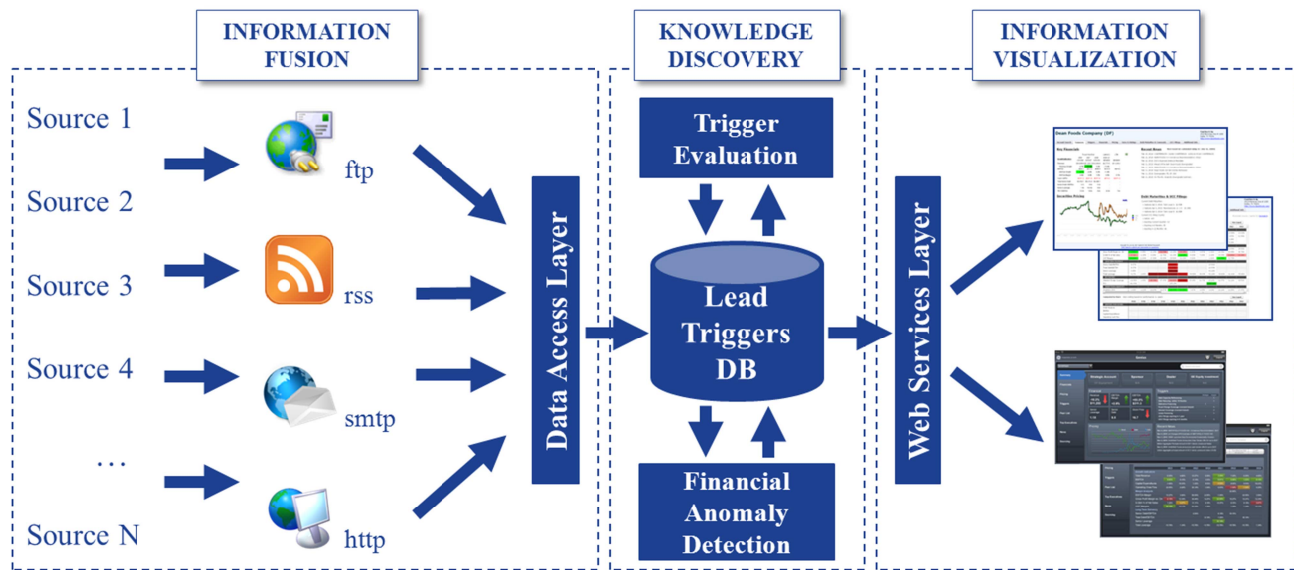


Figure 1: Financing Lead Triggers conceptual system architecture

themselves about the performance of potential customers and to find indicators that a company may be in the market for a financing product. The system also proactively identifies triggers that indicate that a representative should pursue a company as a potential lead.

1.2 System Architecture

The architecture and key components of Lead Triggers are shown in Figure 1. Lead Triggers is integrated with a number of internal and 3rd party data provider systems. Data is pulled from these providers through a diverse set of data acquisition modules and then aggregated into a centralized Lead Triggers database. Algorithms are then applied to detect anomalous financial values and generate sales leads driven by a combination of expert-defined and statistically derived triggers. This aggregated and enriched content feeds the Lead Triggers user interfaces. Important features of the system include the ability to collect and fuse data from different sources, the ability to mine for actionable sales leads, and the ability to visualize the aggregated information and sales leads in a rich, web-based user interface or iPad application.

The system was first built in mid-2009, and was launched in production at GE Capital in November of 2009. Over the past few years the system has undergone a number of enhancements and has had a significant impact on the sales force, both in terms of increased productivity and identification of a significant volume of new sales opportunities.

This paper is organized as follows. Section 2 summarizes prior art in the domain of knowledge discovery and information fusion for automated sales lead generation. Section 3 describes how Lead Triggers performs information fusion across the different data providers. Section 4 then details how the financing lead triggers were developed and operate. Section 5 describes the user interface and how the information and leads are visualized. Section 6 then describes how the system was developed and currently operates. Section 7 reviews the impact the system has had on GE Capital Americas, and finally, Section 8 provides conclusions about the system.

2. PRIOR ART

A number of commercial systems exist that are designed to mine content with the objective of discovering sales leads for companies. However, these systems are largely focused on identifying consumer sales leads, and often focus exclusively on mining publicly available information [3][4].

OneSource's iSell application is quite similar to Lead Triggers in that it is focused on identifying commercial sales leads [5]. iSell is offered as a SaaS (Software as a Service) solution that allows sales reps to define a target prospect pool and then provides them with real-time intelligence on any companies that fall within that pool. The most interesting aspect of iSell is its claim to be able to aggregate data from over 50 different sources, and then to cross-check the information it has pulled to identify the most reliable information across all of the sources. One area where Lead Triggers appears to outperform iSell is the definition of sales triggers. iSell incorporates over 20 triggers, but they appear to be exclusively event driven (e.g., executive management changes) whereas Lead Triggers also incorporates triggers based on changes in financial performance. Lead Triggers is less focused on event-driven triggers, and so in this respect iSell and Lead Triggers could be quite complementary. Even with this limitation, iSell appears to be a very powerful system.

Salesforce's Data.com also focuses on gathering company information and goes a step further by integrating contact information for the key individuals within those companies [6]. The data is integrated directly into Salesforce.com's existing Customer Relationship Management system, giving users the ability to perform the same territory planning and lead identification they normally do within Salesforce, but augmented with more information. With a focus on data cleaning and data quality, Data.com primarily addresses only the first challenge covered in Lead Triggers – the access and fusion of information to drive lead generation.

Salesgenie from Infogroup describes capabilities to access in-depth profiles of prospects (either individuals or businesses), with the ability to define a profile of preferred prospects [7]. Salesgenie then identifies other companies or individuals that match the

profile. The system also assists in the development of email, direct mail or calling sales campaigns. The in-depth profile aspect of Salesgenie is conceptually similar to the information fusion and visualization aspects of Lead Triggers, but Salesgenie appears more focused on identifying a broad list of prospects as opposed to generating a short list of companies to call.

While research publications in this application domain appear limited, Ramakrishnan et al describe a system called ETAP (Electronic Trigger Alert Program) built to mine web pages for ‘trigger events’ [4]. The team focused on collecting text from the web, splitting those text documents into ‘snippets’, and then annotating and classifying those snippets as indicating a sales lead or not. The trigger events defined in ETAP were all expert driven, and utilized publicly available content from online news sites and the like to analyze and extract instances of the trigger events. While the system is conceptually similar to Lead Triggers, it focuses on entirely different data sources (publicly available data) and different data types (text exclusively). The triggers are also expert driven, whereas Lead Triggers includes both expert-defined and statistically derived triggers.

In addition to the commercial products and research cited, a few patents also exist in this space. In U.S. Patent 5,901,287 “Information Aggregation and Synthesization System,” Bull et al describe a generic system for collecting, aggregating and displaying information from disparate online data sources [8]. While the patent begins by describing a generic system, it then goes into some detail describing a specific instance of a sales lead generation system. Their system provides travel information online, tracking user search activity and analyzing that activity to send recommendations to travel agents based on the user’s presumed travel interests. While this is a consumer-focused rather than commercial application, it does match some of the key concepts embodied in the Lead Triggers system from a knowledge discovery and lead distribution perspective.

In a similar vein, in U.S. Patent 5,930,764 “Sales and Marketing Support System Using a Customer Information Database,” Melchione et al describe a system that focuses on aggregating information from both internal and external sources to generate a customer information database used by financial institutions selling products to end consumers [9]. This customer database is used to generate and distribute sales leads to branches of the financial institution, which then use the leads to direct targeted marketing campaigns around the financial products. This patent is conceptually similar to the one cited above, both focusing on collecting and aggregating information on consumers to develop sales leads, but for very different products and domains. Both are only tangentially related to Lead Triggers in that they share the goal of aggregating information from disparate sources to generate leads.

3. INFORMATION FUSION

The information fusion component of Lead Triggers collects and integrates data on hundreds of thousands of companies in GE Capital Americas’ target industries. It aggregates company data from four external data sources and two internal sources.

3.1 Data Acquisition

The data source access methods and update frequencies vary depending on the data type and the frequency of updates generated by the sources, as shown in Table 1.

Table 1: Data types, access methods and update frequencies

| Data Type | Access Method | Update Frequency |
|-----------------------------|---------------|------------------|
| Public company financials | FTP | nightly |
| Private company financials | RSS (HTTP) | weekly |
| Stock prices | RSS (HTTP) | nightly |
| Debt prices | FTP | nightly |
| Credit Default Swap prices | FTP | nightly |
| Secondary loan pricing | SMTP | nightly |
| Credit Ratings | JDBC | weekly |
| News Articles | RSS (HTTP) | every 3 days |
| Debt Maturities & Covenants | SMTP | monthly |
| UCC Filings | JDBC | weekly |

To manage the variety of delivery methods and timing of the updates, Lead Triggers includes a suite of data acquisition modules that retrieve data from each source separately. Each provider has a distinct data structure and method of integration, requiring the development of a distinct data acquisition module for each source. The FTP feeds are processed by Bash shell scripts using Oracle SQL and SQLLDR (SQL Loader). The other feeds are processed by Java programs. Additional data feeds can be easily incorporated by adding new modules to this suite. A cron (UNIX scheduler) process automates the execution of the import processes at predefined intervals, defined in the ‘update frequency’ column of Table 1.

3.2 Data Aggregation

Data from the internal and external feeds are fused in an Oracle database, as shown in Figure 1. The database contains separate tables for each of the data types shown in Table 1. Since each data source has a different unique identifier representing each company, a critical part of the integration is a master table that maps a company’s internal GE Capital unique identifier to the unique identifiers for that same company across each of the different data sources. This master company table is the result of several manual and semi-automated processes, outside the scope of the Lead Triggers system, that match company identifiers across the different input sources.

The master company table is not static. Several internal sources of company records are merged to create this table, resulting in many duplicate records and occasional mismatches of company identifiers across different systems. Before the table could be used, it had to be analyzed extensively to identify inconsistencies and likely errors. Based on feedback from the Lead Triggers development team, the master company table was cleaned and further enhanced to include references between master records. These internal links were used when it was suspected that multiple records referred to the same corporate entity.

As it is cleaned and updated over time, the external source identifiers associated with a specific company may change. As a result, we chose to keep the unique identifiers from each external data source as the foreign keys that link the different data tables to the master company table. This allows the various identifiers to change in the master company table without forcing an update to all of the linked data tables.

In addition to retrieving and storing raw data from the various feeds, the data acquisition processes also calculate and store summary and trend information. Key ratios, year-over-year comparisons, and Modified Z-scores (defined by Senturk et al) are calculated from the raw financial data [10]. The Modified Z-scores (described in detail in Section 4) identify statistically significant variations in the financials of a company and can be effective indicators of financing needs. Counts of active, new and expiring Uniform Commercial Code (UCC) transactions are calculated from raw UCC data [11], and debt covenant breaches are identified by comparing debt covenants to current financial values. UCCs are records of existing financing deals along with their intended durations, and debt covenants provide information on the terms of those deals, all of which can be used to determine the timing and conditions in which new financing may be needed.

Both the raw data and summary information are stored in the Oracle database, which is the primary data source for the knowledge discovery and information visualization components of Lead Triggers.

3.3 Information Fusion Key Challenges

The key challenges faced during the design and development of the information fusion component came primarily from the many disparate data providers that were integrated into the system. Each provider had a distinct data structure and method of integration, requiring the development of many different data acquisition modules. Another challenge came from the use of the master company table. This table was the result of merging several internal sources of company records and contained many duplicate records and occasional mismatches of company identifiers across different systems. Before the table could be used, it had to be analyzed extensively to identify inconsistencies and likely errors in order to be sufficiently cleaned.

Next we describe the triggers that were defined to mine the fused information for actionable sales leads.

4. SALES LEAD DETECTION

A collection of over 50 expert-defined and statistically derived triggers were designed and integrated into the system to detect actionable leads for the sales reps. These lead triggers identify significant events or changes that indicate a company is likely to need financing at some point in the next 6 to 12 months.

The triggers are automatically pushed via email to the sales reps linked to the companies that fired the triggers, delivering targeted intelligence directly to the relevant sales reps so they can be kept up-to-date on developments occurring at the companies in their prospect portfolio. Some of the triggers are generic and apply across the GE Capital financing spectrum, while others are focused on specific product lines and business verticals.

4.1 Trigger Discovery Process

The lead triggers are considered a trade secret and thus we can describe only a few simple examples and will focus more on the knowledge discovery process that led to their creation.

Perhaps the simplest expert-defined trigger is one that notifies sales reps when a company experiences a ratings change, either in a positive or a negative direction. Ratings changes may indicate a financing need, since ratings are largely driven by the financial performance of a company. Both positive and negative ratings changes may indicate financing needs. Strong companies often utilize financing to grow their operations and struggling

companies often pursue financing in an attempt to restructure and right themselves.

The statistical triggers focus primarily on the financial performance of the companies and look for patterns across changes in their financials. For example, a combination of a decrease in cash flow from operations combined with an increase in capital expenditures may indicate a company is investing for long-term growth but could be short on cash at present, and therefore could benefit from a cash infusion.

4.1.1 Dataset Selection

The financial metrics were first analyzed to determine their completeness within the dataset. Only metrics with valid values more than 80% of the time for all public companies for the most recent 40 quarters (10 years) were analyzed, down-selecting the population to approximately 80 unique financial metrics (this included metrics such as total revenue, net income, and long-term debt, to name a few).

This ensured that there was a sufficiently complete dataset, reducing the likelihood that meaningless coincidences would be found from metrics that were rarely populated.

From this dataset of well-populated raw financial metrics, derived values were then generated to normalize the metrics for data mining.

4.1.2 Creating Derived Normalized Values

Financial metrics are challenging to process directly due to the wide distribution of the raw values. For example, while some companies may measure their revenue in the thousands or millions, often large public companies measure theirs in the billions, and so on. Therefore, the Modified Z-Score calculation developed and patented by Senturk et al [10] was used to generate derived, normalized values from the raw financial data.

Modified Z-scores identify anomalies in a company's financials by looking for significant deviations using two orthogonal comparisons: (1) comparing a company's financials to its own past performance, and (2) comparing a company's financials to a set of its closest peers/competitors. The Modified Z-score is calculated through the same method as a traditional Z-score with one key difference. In a traditional Z-score the mean and standard deviation are calculated for all of the points in a population. The Z-score is then calculated as the number of standard deviations a specific value is from the mean of the population to which it belongs. In the Modified Z-score calculation, the target point is not used in the calculation of the population's mean and standard deviation. The mean and standard deviation are calculated for the remaining points, and then the Z-score is calculated (as normal) as the number of standard deviations the target is from the mean.

The team that developed the concept concluded that the Modified Z-score is more resilient to small datasets. Whereas a traditional Z-score requires a population of at least 20 values to be statistically significant, a Modified Z-score has proven to be effective with populations as small as 5 values.

Because Z-scores are directional, the Modified Z-score is further normalized by flipping the sign for metrics where smaller values are considered to be better (such as total debt). Thus, positive values are always considered to be better than negative values. Interestingly, Modified Z-score values can be too positive – values that appear too good to be true are considered suspect and may be signs of anomalies in financial statements.

The Modified Z-scores are extremely valuable for analyzing financial data in that they normalize the financial metrics to enable direct comparisons between diverse companies, large and

small. Modified Z-scores allow users to identify statistically significant variations in a company’s financials when compared to that company’s historical performance or to the performance of its peers.

4.1.3 Trigger Discovery

Statistical models were used to analyze the Modified Z-score metrics across time to find patterns indicative of a financing need. Figure 2 provides a visual representation of a representative pattern that a trigger might look for.

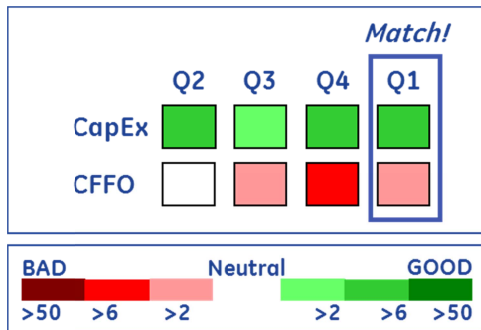


Figure 2: Representative Modified Z-score pattern. In this example, Capital Expenditures (CapEx) is statistically significantly higher than the past for four consecutive quarters, and Cash Flow From Operations (CFFO) is statistically significantly lower than the past for three consecutive quarters. In this example, four quarters of CapEx growth overlapping with three quarters of CFFO decline matches our target pattern.

The models were built using supervised learning, utilizing a historical dataset of positively-labeled (financing was required) and negatively-labeled (financing was not required) cases. The positive cases were built using the financing start dates and financial products found in the debt maturity dataset described in Section 3.2.

The normalized dataset was first explored using descriptive and inferential statistics to analyze the distributions of the various Modified Z-scores of the financial metrics. This analysis provided an awareness of which individual metrics appeared to diverge the most between the positive and negative financing cases.

Decision trees were then used to develop the actual statistical triggers. Specifically, decision trees were used to identify combinations of metrics and thresholds over time that define patterns common across the positive cases but not present in the negative cases. Those patterns were specifically designed to trigger 6-12 months prior to the financing start date for the positive cases. Figure 2 provides an example of one such pattern. In Figure 2, the CapEx Compared to Past Modified Z-score is greater than two for four consecutive quarters and the CFFO Compared to Past Modified Z-score is less than negative two for the most recent three overlapping consecutive quarters.

The team used decision trees primarily because of their transparency. The statistically derived triggers were evaluated both for accuracy and for their comprehensibility by Capital marketing experts. Patterns that did not make sense to the marketing experts were discarded, since it was felt that the sales reps would not take action in response to triggers that could not be explained logically from a business perspective. Therefore, it was critical that the triggers be developed with a transparent technique and decision trees were chosen as being particularly

straightforward to interpret. After discarding the least comprehensible triggers, the remaining triggers were ranked based on their goodness of fit and given an associated priority. Priority 1 triggers had accuracies greater than 95%, while Priority 2 triggers had accuracies between 88% and 95%. Priority 3 triggers had accuracies between 80% and 88%, and Priority 4 triggers had accuracies below 80%.

The expert-defined triggers were learned through a process of eliciting knowledge about the key events that experienced sales reps look for when evaluating companies. A group of expert sales reps was identified based on years of experience and consistently successful track record. This group was then interviewed to learn their preferred indicators of financing needs. Candidate triggers were defined for those indicators that were corroborated across several experts. The candidate expert-defined triggers were then evaluated for accuracy. The result of this process is a knowledge base of successful predictors used by expert sales reps that can now be applied consistently by sales reps regardless of their level of experience.

4.1.4 Trigger Implementation

Each of the 50+ expert and statistical triggers are implemented as Oracle SQL queries that may join as many as 5 or 6 tables at a time to integrate all of the different data sources that may feed a single trigger. These trigger queries are inserted as records in a table that defines the trigger name, category and query. This table is iterated over on a nightly basis by a trigger evaluation application written in Java and initiated via cron, and each query is evaluated one-by-one. A separate table stores the results of the triggers that fired. The system is designed to refrain from inserting duplicate records if the same trigger fires night after night for the same company on the same data.

In theory, triggers can be defined solely for the purpose of feeding other triggers, rather than being sent directly to the sales reps. In other words, it is possible to build hierarchies of triggers, with super-triggers looking for situations where multiple lower-level triggers fired in unison. In practice, however, all of the triggers implemented to date are considered worthy of notifying a sales rep.

4.2 Sales Lead Detection Key Challenges

Many challenges were faced during the design and development of the lead triggers. For the expert triggers, sales reps had to be surveyed to understand what events would typically warrant a call to a prospect. These indicators then had to be reviewed to identify the triggers that appeared relevant to more than one sales rep, and then those triggers had to be designed and implemented as queries for mining the data.

The statistical triggers required the development of a large dataset of quarterly Modified Z-scores for nearly one hundred variables as the inputs, and financing start dates and financial products as the outputs. Once the derived dataset was assembled and normalized, patterns had to be mined and validated.

The final challenge was implementing the triggers in an environment that would enable them to be evaluated on a nightly basis without having the same trigger stored (or worse, distributed to the sales reps) repeatedly.

5. INFORMATION VISUALIZATION

Visualization of the fused information and sales leads is supported through two interfaces—a web-based application and an iPad application. Both applications provide graphical user

interfaces for rich user experiences, enabling the viewing of summary and detailed company information and sales leads.

5.1 Web Application

The web application was developed using Adobe Flex, HTML, Java, and JavaScript, and is comprised of a tabbed interface beginning with a summary tab followed by several tabs for drilling into the details about a company. The views use a variety of tables, charts and other graphical methods to highlight important information and enable a quick understanding of the content.

The Summary tab, shown in Figure 3, provides an overview of the company information, including quadrants for key financial values, a securities (stock, debt and credit default swap) pricing chart, recent news and borrowing activity. The financial section displays both raw and derived financial values with color-coded highlights that indicate significant trends based on the Modified Z-score values.



Figure 3: Company Summary Tab, displaying high-level information about each company. The summary is divided into four quadrants, showing summaries of the most critical information—key financial values, recent news and credit rating changes, a chart of securities prices, and debt maturities & UCC filing counts.

The Triggers tab displays both the expert-defined and statistically derived triggers that fired over the last three months (current month and two prior), as shown in Figure 4, providing both current and trend information for identifying new sales leads.

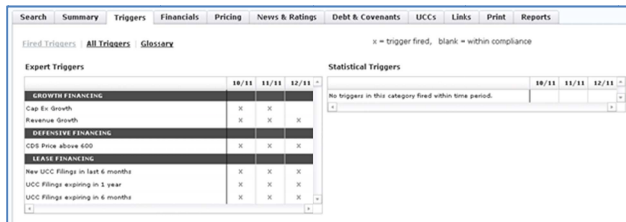


Figure 4: Triggers Tab, displaying the triggers that fired for the selected company over the current and prior two months.

The Financials tab contains several sub-tabs. The sub-tabs provide diverse views of the financial data customized for certain functions, such as risk analysis or financial statement views such as income statements, balance sheets and cash flow statements. Another Financials sub-tab, shown in Figure 5, displays

comparative information based on the Modified Z-scores in a heatmap-style visualization [12]. This view highlights financial conditions that may indicate a financing need, or, conversely, increased risk that contraindicates financing.



Figure 5: Financials Heatmap Sub-Tab, displaying key quarterly financial values over the prior five years. Red and green color-codings in the cells indicate statistically significant variations in a negative or positive direction, respectively.

The Debts & Covenants tab provides a summary of active debts and their covenants, with any breached covenants highlighted in red, as shown in Figure 6.

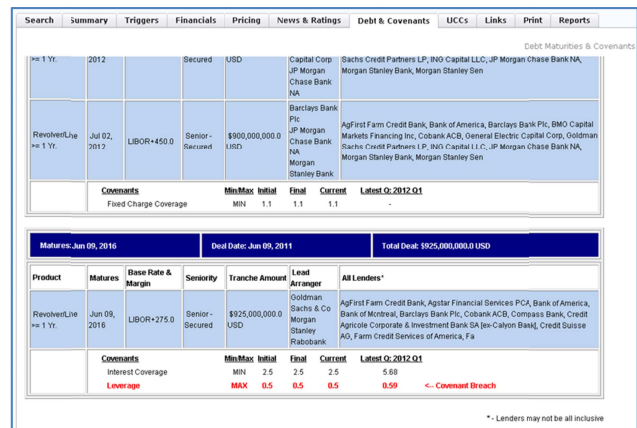


Figure 6: Debt & Covenants Tab, displaying key information on existing company debts including deal start and end dates, primary lenders, and key covenants.

Other tabs provide additional company information including:

- Charts of stock, debt, credit default swap (CDS) and secondary loan price history,
- Recent news about the company and credit ratings over time,
- UCC counts and filings, and
- Links to additional company and industry information available through 3rd party sources.

Two final tabs provide printing and reporting interfaces. The print tab formats the company information from all of the previous tabs into a single report that can be exported to a PDF file. The reporting tab allows users to query for companies in a

specific region or industry and produce a report of the matching companies with key information such as their recent financials, current debts, recent security prices, credit rating and UCC counts.

5.2 iPad Application

The iPad application provides a similar set of views, optimized for mobile devices. This iPad app was specifically highlighted by Apple CEO Tim Cook during a presentation in September 2012 when he first unveiled the new iPhone 5 [13].

Access to the fused company information and triggers is accomplished through a suite of RESTful JSON web services that retrieve and structure the data from the centralized Lead Triggers database. Additionally, the iPad application augments the Lead Triggers information with market and industry information, proving additional context around the company view.

Figure 7 shows the iPad application's financials heatmap view. Like the web application heatmap view in Figure 5, this shows key financials with color-coded highlights that indicate significant trends based on the Modified Z-scores. The triggers view, with a pop-up showing UCC details, is shown in Figure 8.



Figure 7: iPad Financials Heatmap View, displaying key quarterly financial values over the prior five years. Red and green color-codings in the cells indicate statistically significant variations in a negative or positive direction, respectively.



Figure 8: iPad Triggers View, displaying triggers that fired for the selected company over the current and prior months.

5.3 System Access

Lead Triggers provides multiple methods of access. Sales reps can access the web or iPad applications and use search functionality to find companies in either system. In the web application, a search tab allows for the searching of companies by name or by one of several identifiers, such as DUNS (Dun and Bradstreet unique identifier), master table identifier, or internal customer relationship management (CRM) system identifier.

Links to the web application are embedded directly in GECA's CRM system, with the company identifier as a parameter in the URL so that sales reps can easily navigate from the company in the CRM system directly to the summary view of the same company in Lead Triggers.

Links are also embedded in weekly Lead Triggers emails that are automatically pushed to the sales representatives. These emails notify the sales reps of triggers that have recently fired for the companies in the sales rep's unique portfolio. They provide proactive and timely communication informing the sales rep when salient changes occur that may create an opportunity with one of their accounts. An example email is shown in Figure 9. The email includes the names of the companies and the list of triggers that have fired for each company, with a link the sales rep can click to navigate directly to the company summary page in Lead Triggers. Navigating to the Lead Triggers system from the emails is the most frequently used method to connect.

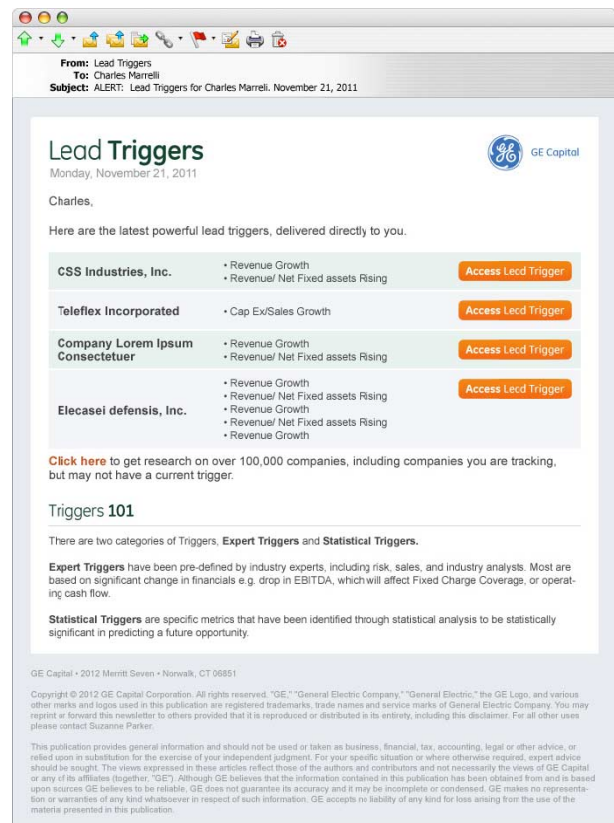


Figure 9: Lead Triggers Email listing companies in the sales rep's portfolio that fired triggers in the past week. The email includes the specific triggers that fired for each company, and links to each company's Lead Triggers summary web page.

5.4 Information Visualization Key Challenges

The key challenges faced during the design and development of the information visualization components were driven largely by the need to present the company information and sales leads in a coherent format that was meaningful and actionable for the sales reps. Information layouts were designed and selected based on how that information would be consumed by the sales reps, with data formatted in ways that most suited the data type. Financial data was shown primarily in tables, stock and securities prices were plotted in graphs, and so on.

The financial data proved particularly challenging, both from a design layout and implementation perspective. The final layouts were chosen because they represented how the financial data is spread and analyzed by the GECA risk organization, which has the final say in approving or rejecting potential deals. Presenting company financials to the sales reps in a layout consistent with the way that the risk organization analyzes the data gives the sales force the ability to better identify and address potential areas of concern that might reduce the likelihood of a deal being approved.

From an implementation perspective, the financial spreads were quite complex and required large tables to visualize the data. While the selection of Flex facilitated generating the tabbed interface and graphs, it is not well suited for displaying large volumes of textual or numerical data in complex tabular formats. As a result, IFrames were used to embed standard HTML pages into the Flex user interface [14].

6. DEVELOPMENT AND OPERATIONS

Lead Triggers was designed and developed by a four-person team at GE Global Research, GE's industrial research laboratory headquartered in Niskayuna, NY. The team included a project leader/software developer, a user interface developer, a back-end infrastructure developer and a database expert. The design and development team worked closely with a Capital Marketing team responsible for gathering requirements from the sales force, negotiating contracts with the third party data vendors, and coordinating the sales rep training. Development began in August of 2009, and the first generation of the system was launched in production 14 weeks later, in late November 2009.

At the time of the production launch, the system was hosted at GE Global Research and managed by the original development team. To facilitate system usage, the Capital Marketing team facilitated four training sessions attended by over 200 sales reps. After the launch and throughout 2010, feedback was gathered from these sales reps, leading to the design and implementation of a number of enhancements including the development of the reporting interface and the integration of additional data sources. In 2009 a total of \$175K was spent by GE Capital to develop and launch the first generation of Lead Triggers, and in 2010 an additional \$200K was spent to further enhance the system.

In early 2011, Lead Triggers was transitioned from GE Global Research to its permanent production environment, managed and maintained by GE Capital's IT organization. To facilitate the transition and ongoing maintenance, the development team authored several documents including guides on the software infrastructure design, database design, user interface design and a standard operating procedure (fault management) guide. The Capital IT team is now completely responsible for the application.

Lead Triggers runs on two commodity servers. One Linux server hosts the Oracle database, and a second Linux server hosts the application, from the data acquisition and information fusion components through to the web interface.

6.1 System Usage Tracking

As part of managing Lead Triggers, the maintenance team monitors system usage, including the number of users, number of companies accessed, and tab views. The tab view analysis, in particular, has given GECA valuable insight into what types of data the sales force finds most useful when learning about a company. Table 2 displays the tab names with the percentage of sessions in which a user viewed those tabs throughout 2010. In this analysis, a session is defined as a request for company data by a user, and a tab view is defined as one or more clicks on the specified tab during the session. Multiple clicks on a given tab during a session count as a single tab view in this analysis.

Table 2: Percent of tab visits for all users

| Tab Label (Data Type) | Percent of Time Tab is Visited |
|--------------------------|-----------------------------------|
| Summary | 100% |
| Triggers | 60% |
| Financials | 56% |
| Debts & Covenants | 53% |
| News & Ratings | 45% |
| Pricing | 43% |
| UCCs | 41% |
| Links | 37% |

The Summary is at 100% because users start on this tab. The details tabs are accessed less frequently for a couple reasons. First, the Summary tab includes information from many of the details tabs, so those tabs only need to be accessed when the sales rep wishes to dive deeper into the data. Second, certain tabs will be empty for certain companies. For example, private companies will rarely have security pricing data.

The Triggers tab, which provides the actionable sales leads, is visited 60% of the time, and is the most visited after the Summary. The Triggers tab is one of two ways that sales reps can review the triggers that fired for a company, since trigger results are also sent directly to them via email. The Financials and Debts & Covenants are the next most-visited tabs, both in the 50%+ range. This is not surprising given that the financials, debts and covenants provide the most quantifiable information.

The News & Ratings, Pricing and UCCs are the next most visited tabs, all in the lower-to-mid 40% range. These tabs provide valuable information, but aren't always as easily translatable into actionable leads, so it is not surprising that these tabs are visited less often than the rest of the content.

7. BUSINESS IMPACT

With the Lead Triggers system, GECA sales reps no longer have to manually scour the web or printed resources to look for meaningful intelligence on prospects. Relevant company information, plus specific and actionable sales leads, are now automatically discovered and made available to the reps through a single system. An evaluation at the end of 2010 revealed that Lead Triggers had been accessed by over 275 sales representatives who had performed over 2,000 company searches, and searching is not even the primary method by which companies are accessed. The most common point of access to a company is by clicking the links embedded in the lead triggers emails.

According to Richard James, former Commercial Initiatives Leader at GECA, "Lead Triggers has become an integral part of the originators' lives, especially in Corporate Finance. Originators' productivity has been improved by 30-50% in terms of preparing for calls and meetings. In addition, Lead Triggers provides a reason to touch the prospects and customers on relevant issues." By providing the sales reps with insights at their fingertips (as opposed to spending hours or more searching for new information), the sales reps have become more productive, making more calls armed with more information. Further, pushing actionable leads brings greater consistency across the sales force, arming even the least experienced reps with meaningful intelligence and increasing the entire sales force's effectiveness.

The qualitative benefits of the increased sales force productivity and effectiveness just described are validated through quantified benefits that were calculated for the first three quarters of 2010. Lead Triggers fired triggers for 59% of the deals that were approved from Q1 through Q3 of 2010, representing 78% of the total volume of business approved for that period. Specifically, this amount represented over \$346B in new finance opportunities that Lead Triggers helped to identify.

By the end of 2010, within the GE Capital Americas pipeline, 61% of the volume had triggers fire, representing \$44B in commitments. Simply put, 59% of the deals approved in the first three quarters of 2010 and 61% of the pipeline of deals in progress had Lead Triggers identifying leads pointing to those deals. While one should not conclude that the sales reps used Lead Triggers exclusively to find each and every one of these deals, there is no doubt among the GECA senior leadership that Lead Triggers provides substantial value to their business.

8. CONCLUSIONS

Lead Triggers is a complete, end-to-end knowledge discovery system that integrates information fusion, knowledge discovery, and information visualization capabilities to address the needs of GE Capital Americas sales reps. No longer burdened by having to manually search through news reports and financial statements to gather information about the prospects in their portfolio, sales reps are now automatically emailed sales leads as they are discovered. Lead Triggers provides a comprehensive view of companies and enables deep-dive analyses into a wide range of information that has been automatically collected, integrated, and mined for actionable intelligence. The system contains multiple access points including a web interface, iPad application, and email notifications of fired triggers to facilitate the sales reps' activities.

The sales leads are generated through a set of over 50 expert-defined and statistically derived triggers. The expert triggers were defined through a set of interviews and corroborations between experienced sales reps, and analyzed for their historical accuracy. The statistical triggers were developed through a knowledge discovery process that focused on financial metrics. The process began by studying the completeness of the dataset to target the most populated metrics, and then involved generating a derived and normalized dataset using Modified Z-score calculations. Finally, the process involved using decision trees to identify combinations of metrics and Z-score cutoffs to identify patterns over time that act as positive indicators of future financing events.

Lead Triggers has been in production since late 2009 and has had a quantifiable impact, improving sales rep productivity by 30-50%. The accuracy and relevance of the triggers has been validated through the analysis of approved deals. In 2010 alone, Lead Triggers provided leads on opportunities that represented over \$44B in new deal commitments for GE Capital Americas.

9. ACKNOWLEDGMENTS

The authors would like to thank Trish Anderson, Magda Kotek, Richard James, Deniz Senturk-Doganaksoy, Richard Aldrich, Craig Reynolds, Cary Correia, and Charles Norton from GE Capital Marketing. We would also like to thank John Interrante and Michael Woellmer from GE Global Research for their contributions.

We would like to acknowledge Sean Morton and Murali Narasimman from GE's Corporate Initiatives Group for building the iPad application. In addition, we would like to recognize John Dressel and the Capital IT team for supporting Lead Triggers in Capital's production environment.

Finally, the authors would like to thank David Spangler, Terry Suppers, Lee Cooper, Janet Barnett and Mike Clark for their support of the project and team.

10. REFERENCES

- [1] GE Capital, <http://www.gecapital.com/>.
- [2] GE Capital Fact Sheet, http://www.gecapital.com/en/pdf/GE_Capital_Fact_Sheet.pdf.
- [3] Davidson Marketing Group, Sales RockStar <http://www.mysalesrockstar.com/>.
- [4] Ramakrishnan, G., Joshi, S., Negi, S., Krishnapuram, R., Balakrishnan, S. 2006. Automatic Sales Lead Generation from Web Data, *Proceedings of the 22nd International Conference on Data Engineering* (Atlanta, GA, USA).
- [5] OneSource iSell, <http://www.onesource.com/isell.aspx>.
- [6] Salesforce Data.com, <https://www.salesforce.com/data/whatisdata/>.
- [7] Infogroup Salesgenie, <http://www.salesgenie.com/>.
- [8] Bull, D.S., Carr, Jr., R.N., Offutt, Jr., J.R., 1999. U.S. Patent 5,901,287, Information Aggregation and Synthesization System. Issued to The Sabre Group Inc., Fort Worth, TX (May 4, 1999).
- [9] Melchione, A., Martinez, R., Seifert, E., Hirsch, M. 1999. U.S. Patent 5,930,764, Sales and Marketing Support System Using a Customer Information Database. Issued to Citibank, New York, NY (July 27, 1999).
- [10] Senturk, D., LaComb, C., Doganaksoy, M., Neagu, R. 2004. Financial Anomaly Detection: a Six Sigma Approach to Detecting Misleading Financials and Financial Decline. *ASA Joint Statistical Meetings* (Toronto, Canada).
- [11] Uniform Commercial Code, J. Michael Goodson Law Library, Duke University School of Law, <http://www.law.duke.edu/lib/researchguides/pdf/ucc.pdf>.
- [12] LaComb, C.A., Hoogs, B.K., Miele, J.P., Senturk-Doganaksoy, D., Neagu, R.E., Bufi, C.N., Moitra, A., Deitsch, A.I., Arthur, R.B. 2006. U.S. Patent Application 11/028,685, Methods and Systems for Visualizing Financial Anomalies. Filed by the General Electric Company, Niskayuna, NY.
- [13] GE Reports, There's An App for That: GE's Genius App Makes a Cameo in iPhone 5 Roll Out, <http://www.gereports.com/theres-an-app-for-that/>.
- [14] Rutherford, A. Flex: Towards a Better IFrame HTML Component, <http://ccgi.arutherford.plus.com/blog/wordpress/2007/09/16/flex-towards-a-better-iframe-component/>.