XPRIORI

Business Intelligence and Analytics with Insight™ from Xpriori

by Dan Spaulding and Timothy Dix

Release 1.1

Xpriori, LLC
2864 S. Circle Dr.
Ste. 401
Colorado Springs, CO 80906
(719) 425-9840
www.xpriori.com

Version 1.1

Xpriori technology is protected by the following patents:
US Patent #5,742,611 (21 Apr 98)
US Patent #5,942,002 (8 Aug 99)
US Patent #6,157,617 (5 Dec 00)
US Patent #6,167,400 (26 Dec 00)
US Patent #6,324,636 (27 Nov 01)
US Patent #6,493,813 (10 Dec 02)
US Patent #6,792,428 (14 Sept 04)

Other U.S. and international patents pending.

EXECUTIVE SUMMARY

For companies that use Business and Financial Tools supported by relational databases, the day starts with hundred's of tables, thousands of fields and lots of reports to track what's going on to identify success, progress and failure. Noticeably, while reports are filled with facts of the current situation, they are lacking in the not so noticeable "why." The day can start here, but it can't end here. Insight, from Xpriori, is designed for the next step - to discover and analyze the conditions and relationships that differentiate the Wins from the Losses. Insight is equally adept for the smaller business and the larger more complex businesses.

WHAT CAN INSIGHT PROVIDE?

Most BI products use traditional Online Analytical Processing methodologies ("OLAP") to provide a multi-dimensional analysis of data. Users must construct complex data cubes (separate relational databases) geared specifically to store data gathered to answer a single query. Insight goes beyond data cube comparisons and focuses on identifying all attributes and relationships that impact the business and uses them to help define the characteristics of an item.

So, for example, instead of just showing the top sales teams and what are they selling to whom, Insight analyzes and correlates all sales teams against every collected data item and; then, orders the results by several measures of significance. It does so automatically without construction of special databases or repositories – it uses the data gathered through normal operations. Insight is not restricted to specific data but can perform the same analysis and correlation with respect to every collected data item and can do so in a semi-automated fashion.

Current approaches to BI/OLAP require the development of hypotheses about the data and then construction of data cubes to test the hypotheses. Insight lets the data speak for itself. Relationships between and among data and metadata are automatically discovered.

This approach eliminates the costly analysis phase of selecting various dimensions and attributes of the OLAP data cube in hopes of identifying a pattern or deviation from a desired outcome. For this reason, Insight enables an iterative approach to analytics. Once a relationship between data items is established, additional fields can be added to provide more detail to help answer "Why" and "Causation" concerns. This is supported by Insight without any code or configuration changes and without any database design.

The benefits to the Insight approach to BI analytics come down not only to flexibility and speed, but also to subjecting large amounts of data to multidimensional analysis which had heretofore been impossible as a practical matter. Users obtain faster knowledge creation because no data cubes have to be built. They suffer no breakdown of data cubes with complex queries on larger data sets. Users benefit from superior flexibility, because new fields are available for immediate analysis, without any code modifications.

HOW INSIGHT AND EXISTING RELATIONALLY SUPPORT PRODUCTS CAN WORK TOGETHER?

First, Insight's XML Utility is used to create comprehensive XML documents. The XML Utility uses common API's to export database table content into individual XML documents. Then, using the identified relationships, the XML Utility combines the related exported XML into a single comprehensive XML document that preserves relationships and values in a parent-child hierarchy.

For example an ORDR table, which contains exported XML order information, is combined with associated XML data exported from other application created tables, which could represent such normally recurring items such as order rows, customers, salesperson, items, invoices, and receipts. Next, Insight reads in the XML documents and creates the results. That is the entirety of the process, absolutely no programming is required.

Business App Table	XML Document Generated by a well known Business and Financial Management Suite

ORDR	`<?xml version="1.0" encoding="UTF-16" ?>`
	`BOM>`
	`<BO>`
	`<AdmInfo>`
	`<Object>-1</Object>`
	`</AdmInfo>`
	`<ORDR>`
	`<row>`
	`<DocEntry>1</DocEntry>`
	`<DocNum>1</DocNum>`
	`<DocType>S</DocType>`
	`<CANCELED>N</CANCELED>`
	`Etc.`

RDR1	`<?xml version="1.0" encoding="UTF-16" ?>`
	`BOM>`
	`<BO>`
	`<AdmInfo>`
	`<Object>-1</Object>`
	`</AdmInfo>`
	`< RDR1>`
	`<row>`
	`<DocEntry>1</DocEntry>`
	`<LineNum>0</LineNum>`
	`<Dscription>Consulting</Dscription>`
	`<Quantity>0.000000</Quantity>`
	`Etc.`

FIG 1.0

Insight Document	Combined XML Document Generated by XML Utility

Order	`<?xml version="1.0" encoding="UTF-16" ?>`
	`BOM>`
	`<BO>`
	`<AdmInfo>`
	`<Object>-1</Object>`
	`</AdmInfo>`
	`<ORDR>`
	`<row>`
	`<DocEntry>1</DocEntry>`
	`<DocNum>1</DocNum>`
	`<DocType>S</DocType>`
	`<CANCELED>N</CANCELED>`
	`Etc.`
	`<Order-OrderDetail>`
	`<ORDR>`
	`<row>`
	`<DocEntry>1</DocEntry>`
	`<LineNum>0</LineNum>`
	`<Dscription>Consulting</Dscription>`
	`<Quantity>0.000000</Quantity>`
	`Etc.`
	`</row>`
	`</ORDR>`
	`Additional Order Detail Records`
	`</Order-OrderDetail>`
	`</row>`
	`</ORDR>`
	`</BO>`

FIG 1.1

Insight features manual and automated drill down and drill around features that are easy to use on a point and click basis. The results of the correlations developed in the drill down and drill around work are presented to the user in a series of histograms. At all stages, the user can construct a series of query filters to further refine his/her analysis. This is all done on the fly.

In the case below, the analysis is that of data from a public database concerning complaints about automobiles. Note that all of the data elements – tags or metadata—are presented in the upper left box of the UI. All of the data associated with those elements are automatically associated and correlated with the results expressed in the histogram to the right. The filter or query is reflected in the lower left box.
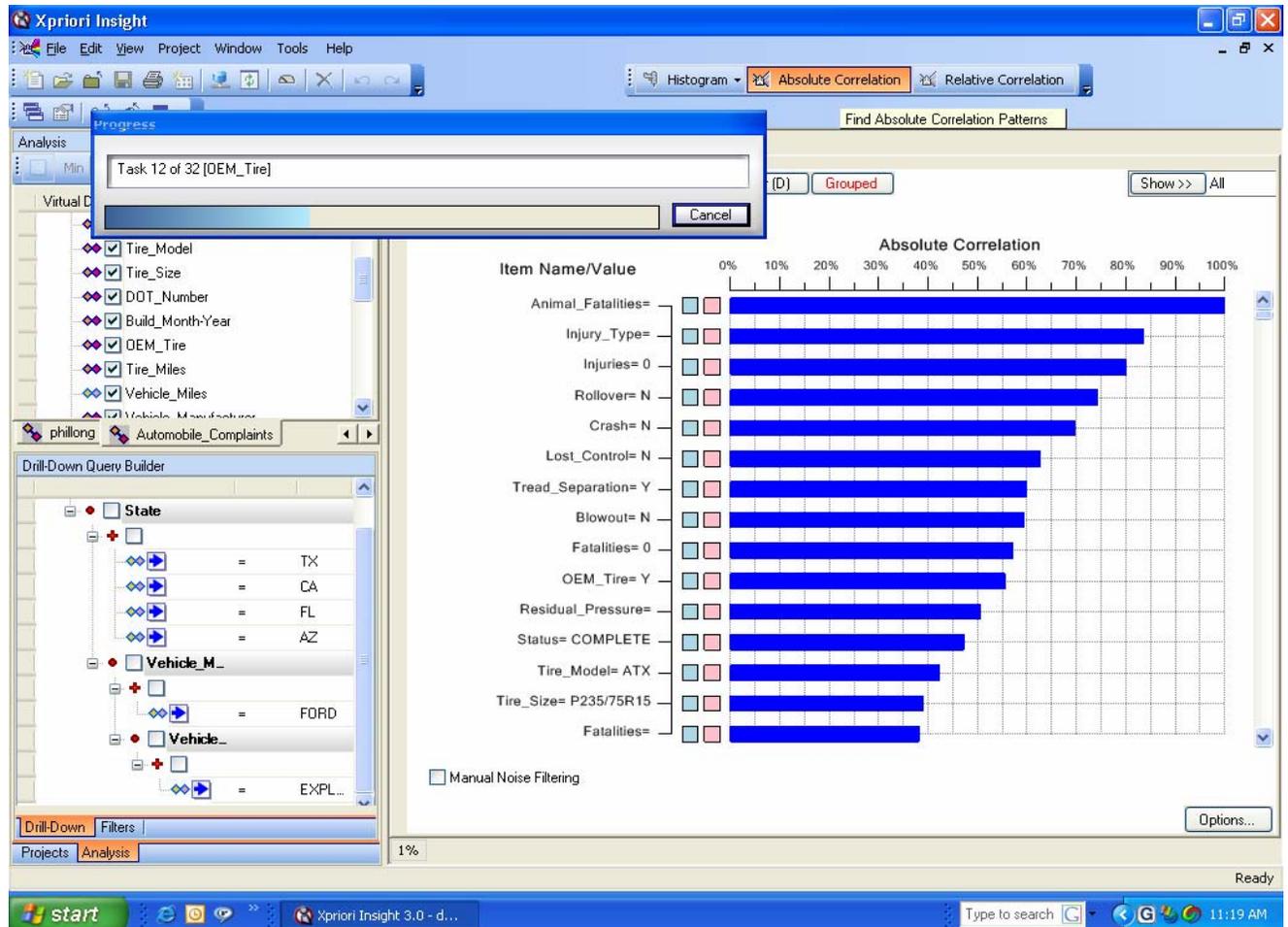

Fig 2.0

There is no database programming required and a business user - not necessarily an IT staff person - can readily do the analysis with a modicum of training. Business intelligence professionals will be able to use Insight for semi-structured decisions and analysis without the complex and cumbersome procedures typically required in OLAP. The data can be aggregated and stored in Insight with scheduled or real time export from the data store that supports normal business operations.

Once a filter is built - or any number of them - the supporting query can examine new data as received. This is done through use of gauges or reports or even the development of Expert Systems. There is an "alerts mechanism" - send an email, ring a bell etc., - to alert users to the

happening of future events. This push analytical function can monitor deltas and a dashboard can be created to present an unlimited number of gauges.
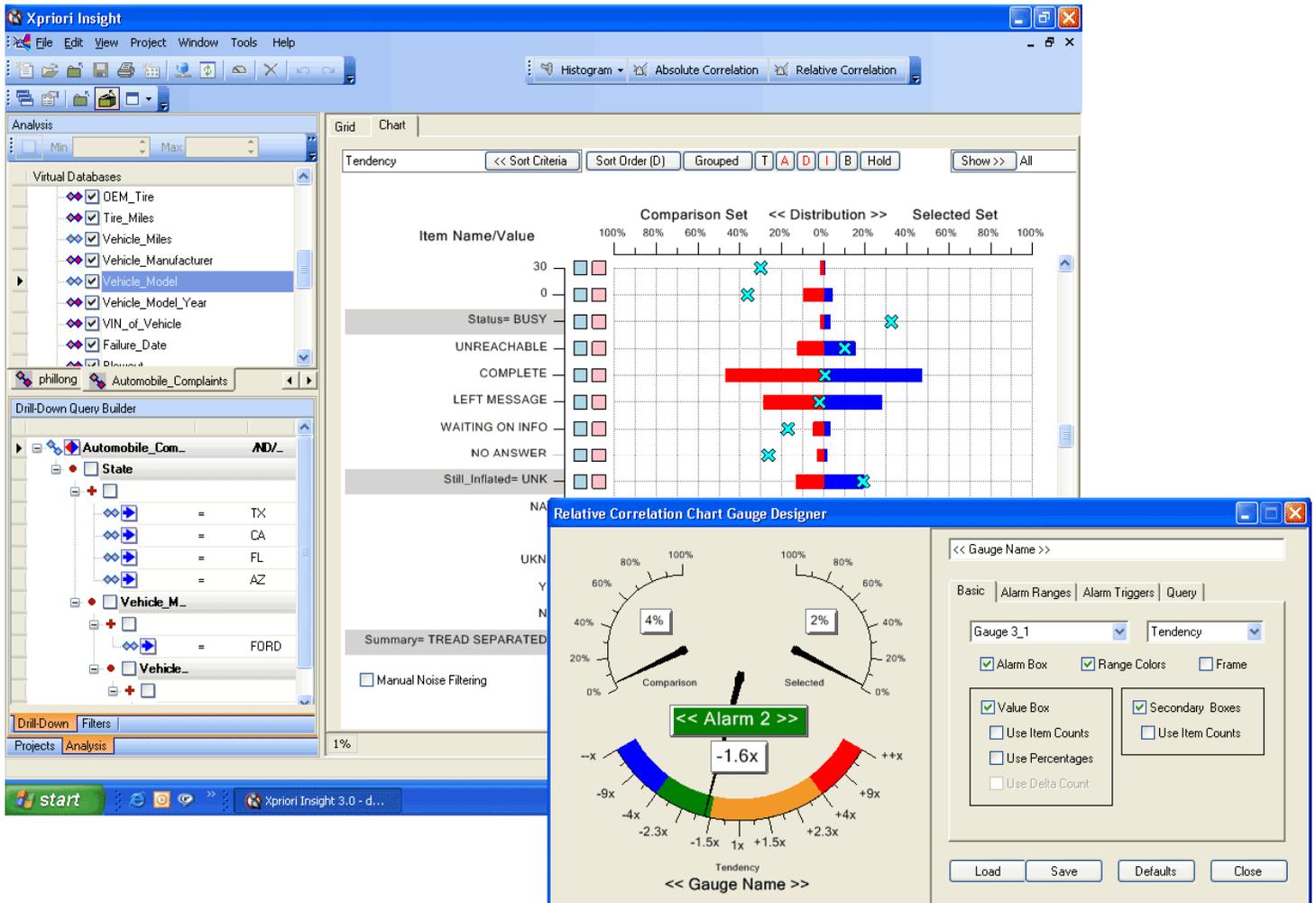

FIG 3.0

Insight provides a zero programming solution and more intimate engagement with enterprise information for business managers without requiring substantial IT involvement. **It can operate on existing data generated from normal business operations without any further database programming being required.** Users can add third party data - such as from surveys - for comparison too and without any database design or programming.

INSIGHT AUTOMATES THE CREATION OF ADVANCED ANALYTICAL ARCHITECTURES FOR LARGER MORE COMPLEX BUSINESSES

Insight provides competitive and operational analytics capabilities that are currently not available to even much larger organizations - those who likely have spent significant amounts of money on database programming and application development. In short, users may not have to "graduate" to new business process management systems to obtain more sophisticated business analytics. They will be able to gain knowledge and compete using advanced analytics with the data that their existing systems collect and in the formats that it provides. Further, business consultants can use Insight to help their customer users to achieve competitive analytics usually reserved for the Fortune 500. The availability of Insight enables both the SMB and large multinational businesses in like fashion.

With the emphasis being placed on analytics to meet competition and to improve performance, it is useful to understand where Xpriori Insight fits. In enterprise information analytics, the Insight analytics model bears some resemblance to traditional OLAP (on-line analytical processing) and DM (data mining). However, Insight provides significant advantages over the OLAP and DM paradigms. Data mining and OLAP are non-equivalent, complimentary technologies. Insight uniquely implements some necessary OLAP and DM features and renders others unnecessary. Industry efforts to integrate OLAP and DM indicate a trend in the information analysis industry toward the Xpriori Insight model.

OLAP
OLAP belongs to a sub-set of decision support tools used to discover patterns and relationships in a data set. The user forms a hypothesis about a relationship and verifies it with a series of queries against the data. In other words, the OLAP analyst generates a series of hypothetical patterns and relationships and uses queries against the database to verify them or disprove them. As such, OLAP analysis is a deductive process. An OLAP database is often synonymous with a data warehouse and typically constructed using star/snowflake schemas, or data cubes.

DATA MINING
Data mining is an inductive process that uses the data, itself, to uncover patterns. Data mining relies heavily on the fields of artificial intelligence (AI) and statistics for pattern recognition and classification. Techniques include algorithms, such as neural networks, decision trees, and discriminant analysis. Traditional statistical techniques involve formal algorithmic modeling based on hypotheses. Data mining commonly includes predictive models using statistical techniques, such as linear regression. A data-mining tool might access an OLAP/relational database, or support various file formats, for data retrieval.

FEATURE COMPARISON
The following table presents a capability/feature comparison of Insight, OLAP, DM and traditional statistical analysis. Subsequent paragraphs provide further explanation.

Feature	Insight	OLAP	Data Mining	Traditional Statistical Analytics
Identify patterns natively, i.e. within DB context	X			
Enable real-time/dynamic changes to data sets	X	X		
Implement dynamic schema modifications	X			
Quickly analyze n-dimensional hierarchic data sets (n-dimensional analysis)	X			
Rapidly discover inherent limitations of data sets	X			
Discover all existing correlations within a data set	X			
Quickly develop multiple hypothesis	X			
Extract insights without hypothetical prejudice	X			
Use to support traditional statistical analysis	X	X	X	X
Quickly deploy business intelligence environments	X			
Identify & monitor KPI's from live data sets	X			

PATTERN PROCESSING

Traditional OLAP and data mining methods use separate approaches to finding patterns and relationships within data sets. OLAP methods involve hypothetical query construction and data manipulation (data cubes). Data mining methods use database queries to retrieve data before applying a variety of statistical and artificial intelligence algorithms. Insight realizes a superior advantage over traditional methods by leveraging its patented pattern processing technology, called DPP (Digital Pattern Processing). The native pattern-processing capability of the Insight engine virtually eliminates any need for data cubes and external statistical analysis for identifying patterns and relationships.

REAL-TIME/DYNAMIC UPDATES

While many OLAP databases support real-time and dynamic updates to data sets, doing so can be difficult and costly. An OLAP database may rely heavily on caching in order to improve query performance. In a real-time environment with low-latency updates, the performance of an OLAP database will suffer due to frequent cache invalidation. Insight users will suffer none of this frustration. Although Insight implements a cache, the performance of its core architecture is not affected in the same manner. In fact, the cache remains valid during and after updates and does not require reload, thereby contributing to significantly better performance.

DYNAMIC SCHEMA MODIFICATION

In data cube design, a star schema defines measures and dimensions that describe which data to present and how to present it. During the design process, the analyst must consider questions that ask why certain data should be included and why it requires a specific presentation. Resulting schemas are primarily incapable of supporting dynamic change.

Insight neither relies on nor requires explicit schemas to understand the structure of stored data. Instead, Insight recognizes implicit schemas from XML context information. XML modifies context information on the fly without re-design. Storing a new document with additional XML tag hierarchy or inserting new context into an existing document automatically updates the implied schema.

Insight supports dynamic schema changes to data since it also does not rely on an explicit schema. Insight inspects the database and learns its structure through contextual information inherent in the stored XML. Therefore, Insight can adapt to frequent schema changes.

N-DIMENSIONAL DATA SET ANALYSIS

While OLAP databases enable analysis of n-dimensional data sets, the time associated with hypotheses development and data cube design results in a high cost compared to Insight. Insight supports dynamic and hierarchic heterogeneity without schema design, or re-designs. Insight adapts to the XML structure based on its inherent context regardless of the existence of heterogeneous structures within a given data type. In other words, Insight views heterogeneous data sets of a given type in a manner similar to the abstraction concept of object oriented software design. This enables Insight to quickly analyze n-dimensional data sets without any need for data cube design and de-normalization, as with OLAP.

INHERENT LIMITATION DISCOVERY

When applying traditional analytical methods commonly used in OLAP and DM, the focus tends to be on answering questions formed from preconceived notions about the expected answers. Significant percentages of analysis resources may be spent making decisions about which

questions to ask and creating a corresponding analysis model. Using Insight, analyses quickly converge toward the most significant correlations that exist within a data set. Insight correlations point to the important questions with relatively little effort from the user. Due to the high performance of Insight and rapid question/answer cycles, Insight analyses can quickly uncover deficiencies in data sets.

Analysis of a survey database might indicate that the data set chosen for the survey is incorrect or inadequate. Users realize the benefits of inherent limitation discovery at any stage in such a survey. For example, a behavioral health organization in Arizona used Insight to analyze the YRBS (Youth Risk Behavior Survey) raw data set from the CDC (Centers for Disease Control). Within minutes of the analysis, they realized that the data set was deficient because it did not include certain protective social factors that might help indicate causality of certain conditions. A previous attempt to analyze a summarized version of the same data set using traditional statistical analysis required nearly a man month of human resources and failed to uncover the deficiency.

Likewise, Analysis of data might cause users to establish new fields and fact gathering strategies – all of which can be done without significant involvement of IT.

CORRELATION DISCOVERY
Insight effectively enables the discovery of virtually all correlations that exist within a data set in a relatively short period-of-time. In the OLAP and DM paradigm, correlation analysis moves at the pace of hypothesis development and data cube design.

HYPOTHESIS DEVELOPMENT AND PREJUDICE
In traditional analytics, the requirement for hypotheses effectively contributes to delayed discovery and preclusion of potential correlations by virtue of choosing a hypothesis. A hypothesis provides an analytical focus and direction for the analyst. By nature, a hypothetical focus creates limitations that would preclude alternate views of the information under analysis. Often, in spite of subconscious or cognitive dissonance concerning a chosen hypothesis, the analyst must proceed to the proof or rejection of the hypothesis.

The absence of a requirement for pre-developed hypotheses allows the Insight user the freedom to explore a data set without limit. Focus and hypotheses emerge as output of the analysis instead of being requisite input. In instances where traditional statistical methods are needed, Insight may be used to drive hypothesis development and significantly reduce the total time and cost of analysis.

CONCLUSIONS
Xpriori Insight makes available next generation competitive analytics to users today. Deployed with existing business applications, it brings OLAP style analytics to all aspects of the enterprise. Drill down and around in the existing information and without the high end IT costs normally associated with analytics.

Let the data and not thoughts or hypotheses about the data lead the way. Use Insight to analyze all the available data collected to determine what data relationships are relevant. Discover what the data can tell you first; make changes and adjustments as you go forward and do it on the fly.

Build business rules around the facts that you discover and not about what you guess might be there. Monitor queries in real or near real time as datasets change; sound alerts; do follow on analysis at anytime; and where appropriate, use the findings to assist in the creation of expert systems and appropriate rules. Support decisions with today's facts.

Identify the quality or important data and map out what is poor quality or unimportant. Add customized data values without any additional effort; expand analysis to third party information coupled with existing data to produce integrated results.

Get all of this without any design or development of specialized databases or OLAP cubes.

Competitors will increasingly rely on analytics for strategic and tactical decisions and Insight provides a key to maintaining a competitive position.

# # #