

Business Intelligence for Big Data Analytics

Tomas Ruzgas
Department of Applied Mathematics
Kaunas University of Technology
Kaunas, Lithuania

Jurgita Dabulytė-Bagdonavičienė
Department of Applied Mathematics
Kaunas University of Technology
Kaunas, Lithuania

Abstract: This article introduces methods and tools which are designed for analyzing Big Data. In the present research, the most popular software tool opportunities have been compared and the differences and advantages have been identified for Business Intelligence (BI) analytics according to the dominant market requirements of BI. The article also presents the technologies of fast calculation processing, including architecture of *in-memory* and *grid* computing.

Keywords: big data, business intelligence, grid computing

1. INTRODUCTION

Since time immemorial, mankind has been collecting and analyzing particular data. In the course of time, the necessity of fast and reliable findings has been increasing. Digital Universe Study of International market research and analysis company International Data Corporation (IDC) has revealed that the amount of created and replicated data encompassed 2.8 zettabytes in 2012. IDC predicts that digital space will have expanded to 40 zettabytes (it will be 50 times larger than it was 10 years ago) by 2020. New data is generated so quickly that a graphic data chart will represent ideal exponent. Consultation company Gartner, Inc. has reported that business increases its data from 40% to 60% per annum. This type of growth is influenced by mobile technologies and databases associated with customers and their behavior in supermarkets (such data is accumulated by trade networks). In addition to financial institutions, research data of medical and human genome is not falling behind the trend. Especially data in social networks is generated very quickly. This is the most difficult processed and unstructured multimedia data: free-form text, images, sounds and video clips. Nowadays, the data generated by devices comprises 30% of all data; therefore, it is predicted that this figure will have reached 42% by 2020. A considerable amount of data is created every day, but it is not information. In order to obtain the information from data, it is necessary to process particular data. Data Science is described as data analysis using scientific methods. Strategically important, as well as irrelevant information can be hidden in a large amount of data. The search for important information in a massive amount of data has encouraged the emergence of tools for data analysis, high quality application packages or programming tools that help to orientate in a substantial amount of information. Increase in data and information brings new requirements for information processing by computer systems.

Data mining is extraction of useful information from accumulated data. It is remarkable that technologies are able to transform factual data into useful information and knowledge, which are necessary for performance management, market analysis and the decision-making process (Han et al., 2012). Data mining is considered to be a multifaceted concept: it can be defined as identifying structures (models, connections, statistical models or templates) in databases (Fayyad et al., 1993), as well as the application of statistics for data analysis and predictive

modelling in order to discover new patterns and trends in big data sets. It may also be described as big data exploration and analysis by automated or semi-automated means with the purpose to find useful patterns and rules (Berry & Linoff, 2008).

Data mining is used for knowledge discovery in databases. During this process, new information is searched for in large amounts of data sets, that could help to gain knowledge of analyzing data and make suitable decisions (Cios et al., 2007). Data mining method helps to find rules for searching tasks and to solve problems of prediction, classification, clustering and interconnectivity; therefore, it is important to have systems, providing various methods for solving tasks of data mining (Dunham, 2002).

The main purposes of this article are to evaluate the tools for big data analytics, to conduct a comparative analysis of the most popular data mining software tools for business intelligence, to identify the differences and similarities of various opportunities and to describe the technologies of fast calculation processing.

2. BUSINESS INTELLIGENCE AND ANALYTICS

Traditional BI market share leaders are disrupted by platforms that expand access to analytics and deliver higher business value. BI leaders should track how traditionalists translate their forward-looking product investments into a renewed momentum and improved customer experience.

The BI and analytics platform market are undergoing a fundamental shift. During the past ten years, BI platform investments have largely been in IT-led consolidation and standardization projects for large-scale systems-of-record reporting. These have tended to be highly governed and centralized, where IT-authored production reports were pushed out to inform a broad array of information consumers and analysts. Now, a wider range of business users are demanding access to interactive styles of analysis and insights from advanced analytics, without requiring them to have IT or data science skills. As the demand from business users for pervasive access to data discovery capabilities is growing, IT sector wants to deliver on this requirement without sacrificing governance.

While the need for system-of-record reporting to run businesses remains, there is a significant change in how

companies are satisfying these and new business-user-driven requirements. They are increasingly shifting from using the installed base, i.e. traditional and IT-centric platforms that are the enterprise standard, to more decentralized data discovery deployments that are now spreading across enterprises. There is the transition to platforms that can be rapidly implemented and can be used either by analysts and business users in order to find insights quickly, or by IT to quickly build analytics content in order to meet business requirements and to deliver more timely business benefits. Gartner estimates that more than a half of net new purchasing is data-discovery-driven (Sommer et al., 2014). This shift to a decentralized model, empowering more business users, also drives the need for a governed data discovery approach.

This is a continuation of a six-year trend, where the installed-base, IT-centric platforms are being complemented, and in 2014, they were increasingly displaced for new deployments and projects with business-user-driven data discovery and interactive analysis techniques. This is also increasing IT's concerns and requirements around governance as deployments grow. Making analytics more accessible and pervasive to a broader range of users and use cases is the primary goal of organizations, making this transition.

Traditional BI platform vendors have tried very hard to meet the needs of the current market by delivering their own business-user-driven data discovery capabilities and enticing adoption through bundling and integration with the rest of their stack. However, their offerings have been pale imitations of the successful data discovery specialists (the gold standard being Tableau) and, as a result, have had limited adoption to date. Their investments in next-generation data discovery capabilities have the potential to differentiate them and spur adoption, but these offerings are works in progress (for example, SAP Lumira and IBM Watson Analytics).

Also, in support of wider user adoption, companies and independent software vendors are increasingly embedding traditional reporting, dashboards and interactive analysis into business processes or applications. They are also incorporating more advanced and prescriptive analytics built from statistical functions and algorithms available within the BI platform into analytics applications. This will deliver insights to a broader range of analytics users that lack advanced analytics skills.

As companies implement a more decentralized and bimodal governed data discovery approach to BI, business users and analysts also demand access to self-service capabilities beyond data discovery and interactive visualization of IT-curated data sources. This includes access to sophisticated, yet business-user-accessible, data preparation tools. Business users also look for easier and faster ways to discover relevant patterns and insights in data. In response, BI and analytics vendors introduce self-service data preparation (along with a number of startups such as ClearStory Data, Paxata, Trifacta and Tamr), and smart data discovery and pattern detection capabilities (an area for startups such as BeyondCore and DataRPM) to address these emerging requirements and to create differentiation in the market. The intent is to expand the use of analytics, particularly insight from advanced analytics, to a broad range of consumers and non-traditional BI users, increasingly on mobile devices and deployed in the cloud.

Interest in cloud BI declined slightly during 2015, to 42% compared with last year's 45% — of customer survey respondents reporting they either are (28%) or are planning to

deploy (14%) BI in some form of private, public or hybrid cloud. The interest continued to lean toward private cloud and comes primarily from those lines of business (LOBs) where data for analysis is already in the cloud. As data gravity shifts to the cloud and interest in deploying BI in the cloud expands, new market entrants such as Salesforce Analytics Cloud, cloud BI startups and cloud BI offerings from on-premises vendors are emerging to meet this demand and offer more options to buyers of BI and analytics platforms. While most BI vendors now have a cloud strategy, many leaders of BI and analytics initiatives do not have a strategy on how to combine and integrate cloud services with their on-premises capabilities.

Moreover, companies are increasingly building analytics applications, leveraging a range of new multistructured data sources that are both internal and external to the enterprise and stored in the cloud and on-premises to conduct new types of analysis, such as location analytics, sentiment and graph analytics. The demand for native access to multistructured and streaming data combined with interactive visualization and exploration capabilities comes mostly from early adopters, but are becoming increasingly important platform features.

As a result of the market dynamics discussed above, for this Magic Quadrant, Gartner defines BI and analytics as a software platform that delivers 13 critical capabilities across three categories (i.e. to enable, produce and consume) in support of four use cases for BI and analytics. These capabilities support building an analytics portfolio that maps to shifting requirements from IT to the business. From delivery of insights to the analytics consumer, through an information portal often deployed centrally by IT, to an analytics workbench used by analysts requiring interactive and smart data exploration (Tapadinhas, 2014), these capabilities enable BI leaders to support a range of functions and use cases from system-of-record reporting and analytic applications to decentralized self-service data discovery. A data science lab would be an additional component of an analytics portfolio. Predictive and prescriptive analytics platform capabilities and vendors are covered in Fig. 1.



Figure. 1 Magic Quadrant for Business Intelligence and Analytics. Source: Gartner

Vendors are assessed for their support of four main use cases:

- centralized BI provisioning: supports a workflow from data to IT-delivered-and-managed content;
- decentralized analytics: supports a workflow from data to self-service analytics;
- governed data discovery: supports a workflow from data to self-service analytics to systems-of-record, IT-managed content with governance, reusability and promotability;
- OEM/embedded BI: supports a workflow from data to embedded BI content in a process or application.

Vendors are also assessed according to the following 13 critical capabilities: business user data mashup and modelling, internal platform integration, BI platform administration, metadata management, cloud deployment, development and integration, free-form interactive exploration, analytic dashboards and content, IT-developed reporting and dashboards, traditional styles of analysis, mobile, collaboration and social integration and embedded BI (Sallam et al., 2015).

Fig. 1 presents a global view of Gartner's opinion of the main software vendors that should be considered by organizations, seeking to use BI and analytics platforms to develop BI applications. Buyers should evaluate vendors in all four quadrants without assuming that only the Leaders can deliver successful BI implementations. Year-over-year comparisons of vendors' positions are not particularly useful, given the market dynamics (such as emerging competitors, new product road maps and new buying centers); also, clients' concerns have changed. It is also important to avoid the natural tendency to ascribe personal definitions. For the purposes of evaluation in this Magic Quadrant, the measures are very specific and likely to be broader than the axis titles may imply at first glance.

According to the study of Gartner, Inc. (world's leading research and consultancy company of information technology), which was conducted in 2015, SAS and the Tableau were recognized as the world's greatest leaders in the field of business intelligence and analytics platforms. The results of evaluation are presented in (see Fig. 1) (Note: the best position is at the top right corner of the figure).

SAS Institute Inc. offers a vast array of integrated components within its Business Intelligence and Analytics suite that combines deep expertise in statistics and predictive modelling with innovative visualization enabled by powerful in-memory processing capabilities. SAS Visual Analytics is the flagship product in the suite for delivering interactive and self-service analytic capabilities at an enterprise level, i.e. extending the reach of SAS beyond its traditional user base of power users, data scientists and IT developers within organizations. SAS also leverages its range of platform components and expertise in various industries to offer a wide range of vertical- and domain-specific analytic applications.

SAS is again a leader this year as it continues to build momentum with SAS Visual Analytics, which was released in 2012 and has gained some traction in the market against the data discovery leaders through product differentiation and a more accessible pricing model (with a lower entry point than initially offered). SAS also continues to demonstrate very strong vision in many areas such as the expansion of both smart data discovery capabilities and embedded advanced analytics within SAS Visual Analytics, seamless navigation

between SAS Visual Analytics and SAS Visual Statistics and integration across other core analytic components of the platform in order to address enterprise requirements for governed data discovery.

- Strengths

- SAS was rated slightly higher for market understanding (by references) than the average for this Magic Quadrant; this is a composite measure combining ease of use, complexity of analysis and breadth of use. Support for complex analytic use cases is an obvious strength for SAS, but the fact that eight other vendors ranked higher for complexity of analysis may indicate that in many cases the primary product being used is Enterprise BI, which offers more traditional styles of reporting, and that penetration and adoption of Visual Analytics to address more complex use cases is a work-in-progress within SAS's BI customer base. The portfolio of products reaches a broader range of users leveraging the platform to support use cases spanning the full analytic spectrum, which is positive for SAS and a differentiator for its platform.

- The main reasons why reference customers choose SAS are functionality and product quality, which are clear strengths. SAS delivers a full range of functionality through integrated BI and analytic platform components such as SAS Visual Analytics, SAS Office Analytics and SAS BI/Enterprise BI Server (EBI) as well as complementary products used for data integration, data management, data mining and predictive modelling, all built with a focus on product quality for which SAS was rated just above the overall average.

- The SAS BI and analytics platform can be deployed to meet the needs of a diverse set of use cases, as indicated by reference organizations that ranked SAS third for frequency of deployment in both centralized and decentralized BI use cases. This diversity positions SAS favourably to differentiate itself from other vendors in the market with a platform that is able to meet both the enterprise IT needs and business self-service needs.

- Nearly 15% of survey references report using the integrated self-service data preparation capabilities offered by SAS to allow business users and analysts to access, integrate and transform data in preparation for analysis. The availability of integrated business-user data preparation capabilities is a differentiator for SAS compared with other data discovery vendors; particularly Tableau, which relies on third-party integration with vendors such as Alteryx, Paxata and Trifacta to deliver this capability to its customers.

- Weaknesses

- License cost was again a concern for SAS customers in 2014 and was cited as a barrier to wider deployments by 46% of the reference organizations who responded to the survey, higher than all but one other vendor in the Magic Quadrant. It is expected that this will improve in the next year's survey as customers benefit from the fact that

SAS revamped its Visual Analytics pricing structure in September 2014 to address this concern and offer its customers a per user price point that more closely aligns with competitive data discovery products in the market. With this change, SAS has also made Visual Analytics more accessible to the SMB market with a lower point of entry, i.e. four-core server license priced at \$8,000, which can support up to five power users. Under the new pricing structure, the per-user license cost of Visual Analytics is more comparable to leading data discovery offerings, which is critical to SAS's goal of extending the reach of analytics more broadly within its customer base and to win net new customers.

- Customers reported significant difficulty in migrating to the latest release of the SAS platform components that they have deployed, as indicated by its being given the fourth-highest migration difficulty rating. While the migration difficulty rating is high (compared to other Magic Quadrant vendors included in the survey), it should be noted that the score corresponds to a rating between "straightforward" and "somewhat complex," according to the scale used in the survey. It is also likely that the complexity reported by some customers is related to platform-level migrations rather than version updates to individual products.
- Support for complex use cases is platform strength, but SAS references rate both overall ease of use and business benefits delivered as below the overall average. This could be because the adoption of Visual Analytics, while higher than other traditional market share leaders, is still early and has yet to have its full impact on the perceived ease of use; also, the most recent release of EBI, which offers usability improvements, has not yet been widely deployed. Other data discovery platforms are currently doing a better job of executing on the vision of making hard things easy and being accessible to a broader range of users, but SAS Visual Analytics is gaining awareness and traction in the market and has the potential to close the gap.

Tableau's intuitive and visual-based data discovery capabilities have transformed business users' expectations about what they can discover in data and share without extensive skills or training with a BI platform. Tableau's revenue growth during the past few years has very rapidly passed through the \$100 million, \$200 million and \$300 million revenue thresholds at an extraordinary rate compared with other software and technology companies.

Tableau has a strong position on the Ability to Execute axis of the Leaders quadrant, because of the company's successful "land and expand" strategy that has driven much of its growth momentum. Many of Gartner's BI and analytics clients are seeing Tableau usage expand in their organizations and have had to adapt their strategy. They have had to adjust to incorporate the requirements that new users/usage of Tableau bring into the existing deployment and information governance models and information infrastructures. Despite its exceptional growth, which can cause growing pains, Tableau has continued to deliver stellar customer experience and business value. It is expected that Tableau will continue

to rapidly expand its partner network and to improve international presence during the coming years.

- Strengths
 - Tableau has clearly defined the market in terms of data discovery, with a focus on "helping people see and understand their data." Currently, it is the perceived market leader with most vendors viewing Tableau as the competitor that they most want to be like and to beat. At a minimum, they want to stop the encroachment of Tableau into their customer accounts.
 - Tableau rates among the top five vendors for aggregate product score, with particular strengths in the decentralized and governed data discovery use cases. In particular, analytic dashboards, free-form exploration, business-user data mashup and cloud deployment are platform strengths. Tableau's direct query access to a broad range of SQL and MDX data sources, as well as a number of Hadoop distributions, native support for Google BigQuery, Salesforce and Google Analytics has been a strength of the platform since the product's inception and often increased its appeal to IT versus in-memory-only options. As a result, customers report having slightly below-average deployment sizes in terms of users, but among the highest data volumes (in this Magic Quadrant).
 - Tableau has managed its growth and momentum well. The company has been able to grow and scale without a significant impact on discounts extended (that is, these are very limited) or customer experience. Most technology companies struggle to manage this balance between growth and execution.
 - Tableau customers report among the highest scores in terms of breadth and ease of use along with high business benefits realized. Gartner inquiries and customer conversations reveal that Tableau users report enthusiasm for the product as a result of being able to rapidly leverage insights from Tableau that have a significant impact on their business. Customers also report faster-than-average report development times.
 - Tableau is an R&D-driven company. It continues to invest in R&D at a higher pace (in terms of revenue percentage, it was 29% in 2014) than most other BI vendors.
- Weaknesses
 - Tableau has a limited product line focused on data discovery. Organizations like buying and managing fewer software assets and vendors. At some point, many of the new generation of visualization and discovery tools that are bundled with other (competitor) applications may gain traction, particularly as they roll out smart data discovery and self-service data preparation differentiators.
 - IT-developed reports and dashboards, traditional styles of analysis, metadata management, development and integration, BI platform administration, embedded BI and collaboration are rated as weaker capabilities of the platform, making it less well suited for centralized and embedded use

cases. When Tableau customers have advanced data preparation, production reporting, advanced analytics, distribution and alerting as requirements, they have to turn to third-party products and partner capabilities. This may also limit its ability for large-scale displacements, but not for large scale surrounding and marginalizing of IT and report-centric incumbents.

- Tableau is the competitive target of most other vendors in this market. It faces competitive threats from every other vendor in the market that is also focused on delivering self-service data discovery and visualization capabilities, in an attempt to slow down Tableau's momentum.
- Tableau offers limited advanced analytics capabilities. R integration has been recently added and is a major improvement for users, needing more statistical and advanced capabilities. Other vendors, such as SAS, SAP and Tibco, have more advanced native capabilities.

Tableau's enterprise features around data modelling and reuse, scalability and embeddability, which enable companies to use the platform in a more pervasive and governed way, are evolving with each release, but are still more limited than IT-centric system-of-record platforms.

3. ANALYTICS: BUSINESS VISUALIZATION

Regardless of size or industry sector, organizations collect all types and amounts of data. Unfortunately, traditional architectures and existing infrastructures are not designed to deliver the fast analytical processing needed for rapid insights. As a result, IT is swamped with constant requests for ad hoc analyses and one-off reports. Any delay can frustrate decision makers because it takes too long (or it may be impossible) to get the information needed to answer their questions quickly. Increasingly, decision makers, analysts and other business users want to share reports via email or mobile devices. To help one make sense of the growing data within organization, SAS Institute Inc. product Visual Analytics provides an interactive user experience that combines advanced data visualization, an easy-to-use interface and powerful in-memory technology. This lets a wide variety of users visually explore data, execute analytics and understand what data means. Then they can create and deliver reports wherever needed via the web, mobile devices or Microsoft Office applications.

Data visualization helps explore and make sense of data (Tagarden, 1999). Adding analytics to visualizations helps uncover insights buried in data. Analytics visualization helps discover trends within your business and the market that affect the bottom line. One can quickly recognize outliers that may affect product quality or customer churn. One can also easily recognize parameters in data that are highly correlated. Some of these correlations will be obvious, but others will not. In identifying these relationships, one is able to focus on the areas most likely to influence highest-priority goals. By combining dashboards, reporting, BI and analytics, analytic tools provide both data visualization and analytic visualization. No matter how deep one wants to dive into data, analytic tools provide the capabilities and visualization techniques to take the user there. SAS Visual Analytics lets one go directly from reporting to exploration in the same user experience. With support for data management, report

creation, collaboration through SAS Mobile BI apps and Microsoft Office integration, SAS Visual Analytics helps unlock insights and improve efficiency throughout the organization. SAS Visual Analytics reduces the number of tools that should be used and the number of systems that IT must maintain. SAS Visual Analytics combines powerful in-memory technologies with an extremely easy-to-use exploration interface and drag-and-drop analytics capabilities. No coding is required. Report creators, business analysts and even traditional consumers of BI reports can create and share visualizations to gain new insights from their data. SAS Visual Analytics is designed to handle big data, with in-memory processing designed to meet the demands of today and tomorrow. Flexible deployment options let the user easily scale system as data and analytics needs grow. SAS Visual Analytics integrates with Microsoft Office, helping share interactive and self-service reports directly within familiar Microsoft Office applications. These are more than static reports. SAS Visual Analytics allows to build reports that enable collaborative and engaging discussions that can drive deeper insights and better decisions.

The SAS LASR Analytic Server is the in-memory analytics engine for SAS Visual Analytics. In-memory analytics allows quickly determine relationships across hundreds of parameters in billions of rows of data. After all, speed and accuracy are critical to effective analytics. With social media data and freeform text documents becoming part of data ecosystem, the question is often "What valuable information is in all this data?" Data from the social media world, including Twitter streams, Google Analytics and Facebook, as well as call center logs, online comments and other text-based documents can be analyzed to determine much more than the frequency of common terms and phrases. The sentiment around topics, terms and entire text documents can also be determined. Through the combination of text sentiment analysis and data visualization techniques, documents can be filtered by topic and sentiment; therefore, areas that need attention may be isolated.

With web-based exploratory analysis and other easy-to-use features, even users without analytical expertise can use predictive analytics to gain precise insights (Matthew et al., 2006). Nontechnical users can create and change queries simply by selecting items from a sidebar or dynamically filtering and grouping data items. Autocharting selects the visualization that best suits the type of data chosen. "What does it mean" pop-up boxes provide explanations of analytical techniques, helping everyone understand the data and what the analysis means. Analytically savvy users can use visualization techniques to spot trends and derive deep intelligence quickly and easily. This eliminates much of the everyday trial-and-error process currently used to identify areas that need further analysis.

How do customers navigate website of organisation? What is the customer journey through organisation support structure? The data accumulated from operational systems provides information to paint a clear picture of how transactions move within those systems. Path analysis with SAS Visual Analytics allows to see those flow patterns and recognize trends, such as where customers enter the website, where they navigate and where they exit. With SAS Visual Analytics, successful flow patterns and isolate flows that failed to deliver the desired action can be identified. This level of analytics visualization provides decision makers with the information required to pinpoint opportunities for improvement. Analytic features are tailored for ease of use; therefore, everyone can

create analytic visualizations on their own without learning new skills or engaging IT. Self-service autoloading allows the users to load their own data from Excel spreadsheets and other sources for analysis.

Growing volumes and varieties of big data make it difficult to visualize and understand valuable relationships in data and obtain the analytically based answers, which require to take the best actions. Traditional IT infrastructures are just not designed for rapid and iterative analytical processing and on-the-fly changes to predictive models. It is hard for statisticians, data scientists and business analysts to build the number of models that are needed. They cannot easily experiment with segments or groups, or quickly refine their models to find the best one. SAS Visual Statistics solves these issues. As an add-on to SAS Visual Analytics, it combines interactive data exploration and discovery with the ability to easily build and adjust huge numbers of predictive models. It is really very easy as no coding is required. The in-memory engine reads data into memory once, putting an end to constant and expensive data shuffling.

SAS Visual Statistics provides an interactive, intuitive, drag-and-drop, web-browser interface for creating descriptive and predictive models on data of any size rapidly. It takes advantage of LASR Analytic Server to persist and analyze data in memory and deliver near instantaneous results. When combined with SAS Visual Analytics, it provides a fast and single environment for interactive data exploration and model development. SAS Visual Statistics is designed for statisticians, data scientists and business analysts who want to visually and instantly interact with and analyze complex data. The easy-to-use, drag-and-drop interface provides nonprogramming access to powerful SAS statistical modeling and machine-learning techniques. These techniques are used to predict outcomes that result in better and more targeted actions.

SAS Visual Statistics is an add-on to SAS Visual Analytics Explorer. The common SAS Visual Analytics Explorer environment provides interactive data exploration and analytical modeling capabilities. It can quickly identify predictive drivers among multiple exploratory variables, and

interactively discover outliers and data discrepancies. Then, this information may be used to populate interactive environment for sophisticated predictive modelling. The web browser interface makes it a simple drag-and-drop process to create powerful descriptive and predictive models. Multiple users can easily collaborate to build and refine the best models. Interactive processing is very fast; thus, users can quickly and easily experiment with different techniques.

4. GRID: FASTER PROCESSING

These days, IT budgets are typically limited in most organizations, which makes meeting the computing demands of today's business environment a constant challenge. Buying the latest and greatest servers (i.e., scaling up) to meet peak-demand computing loads is one solution, but it can be both costly and inefficient. Organizations' use of business analytics grows, as well as the need for a flexible IT infrastructure that can scale cost-effectively while meeting peak demands and managing growing and increasingly diverse user workloads. Grid enables organizations to create a managed, shared grid computing environment for processing large volumes of data and analytic programmes. The solution provides critical capabilities for meeting an organization's business analytics needs, including workload balancing, job prioritization, high availability, parallel processing, resource assignment and monitoring.

Grid gives IT greater flexibility to meet service level commitments by easily reassigning computing resources to meet peak workloads or changing business demands (Smith et al., 2002). The solution provides a central point of control for administering policies, programmes, queues and job prioritization across multiple types of users and applications to achieve business goals under a given set of constraints. Having multiple servers in a grid computing environment enables jobs to run on the best available resource. If a server fails, its jobs can be transitioned seamlessly to another server, providing high availability. In addition, IT staff can perform maintenance on specific servers without interrupting analytics jobs, as well as introduce additional computing resources without disrupting the business. Multiprocessing capabilities let divide individual jobs into subtasks that are run in parallel

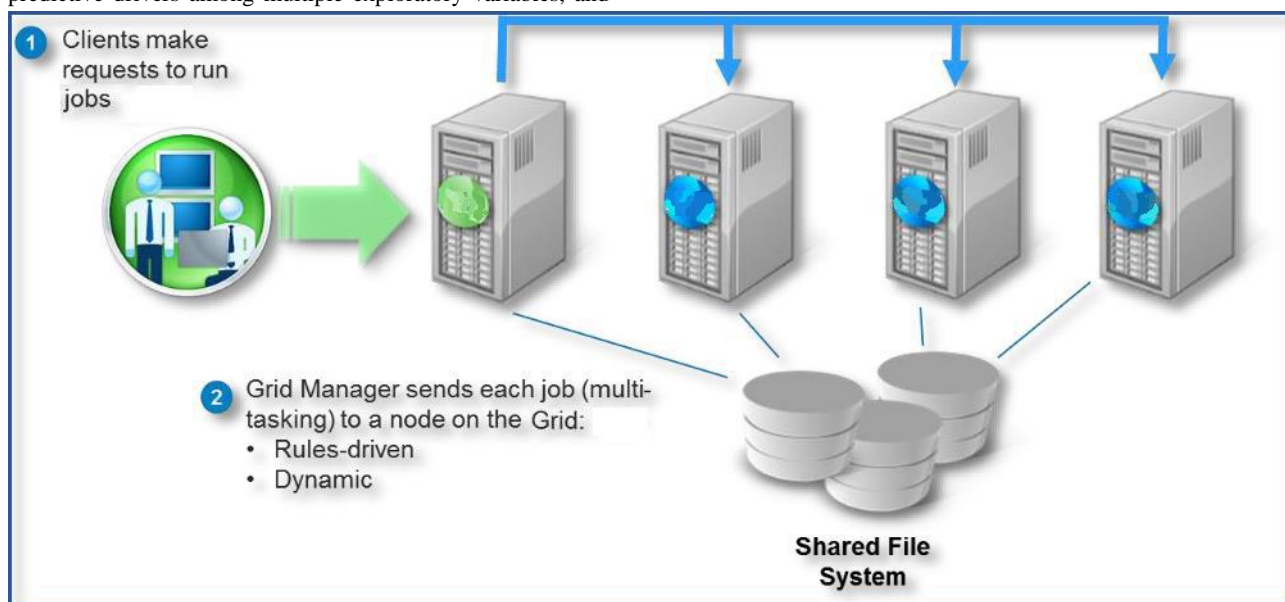


Figure 2. Grid Computing Architecture.

on the best available hardware resource. The programmes best-suited for parallel processing are those with large data sets and long run times, as well as those with replicate runs of independent tasks running against large data sets. Processing data integration, reporting and analytical jobs accelerate decision making across the enterprise. Grid lets fully utilize all available computing resources now and cost-effectively scale out as needed, adding capacity in single-processing units to keep IT spending in check (Joseph, 2004). As it can add low-cost commodity hardware resources incrementally, there is no need to size today's environment.

SAS Grid Manager's patented technology uses industry-leading grid computing middleware from Platform Computing to get maximum availability from business analytics environment. The solution gives a competitive advantage by enabling to balance user and application workloads among available computing resources; consequently, it is possible to obtain results much more quickly. IT can add computing resources in the form of lower-cost commodity hardware incrementally, eliminating the need to size today's environment for tomorrow's demands.

SAS data integration and analytical products are automatically tailored for parallel processing in a grid computing environment. To achieve maximum processing efficiency with minimum user intervention, these programs detect the grid environment at the time of execution. The grid-enabled logic, that is produced, can be saved as stored processes for the use by other reporting clients to generate results for more users as cost-effectively as possible. Other SAS solutions, including SAS Enterprise Guide and SAS Risk Dimensions, can automatically submit jobs to a grid of shared computing resources. All programmes can take advantage of grid computing environment with the addition of programming syntax and a structure that allows the submission of entire programmes to the grid or the parallel execution of programme steps (subtasks).

A wide variety of SAS jobs can be scheduled across grid environments for optimal resource utilization and faster processing. Individual jobs can be divided into subtasks that are then executed in parallel to accelerate processing and increase workload throughput. In today's international organizations, nightly batch-processing windows no longer exist. As a result, data is available 24/7 and can be quickly loaded and analyzed.

5. CONCLUSIONS

The need for platforms to scale and perform for larger amounts of diverse data will also continue to dominate BI market requirements. At the same time, the ability to bridge decentralized business-user-led analytics deployments with those centralized to serve the enterprise will be a crucial ongoing challenge for IT and BI vendors. With the added complexities introduced by new data sources (such as the cloud, real-time streaming events and sensors and multistructured data) and new types of analysis (such as link/network and sentiment analysis, and new algorithms for machine learning), new challenges and opportunities will emerge to integrate, govern and leverage these new sources to build business value. Leaders of BI initiatives will be under pressure to identify and optimize these opportunities and to deliver results faster than ever before.

In-memory analytical processing build models faster (Zaharia et al., 2012). With the LASR Analytic Server, there is no need to write data to disk or perform data shuffling. SAS Visual

Statistics loads all data into memory once and interacts with the data without reloading it each time when a new task is performed. This means the impact of changes to models (e.g., adding new variables or removing outliers) is instantly visible. Because it is designed for concurrent processing, many users can create and run complex models simultaneously. Data and analytic workloads are performed in a distributed form across multiple server nodes, and are multithreaded on each node for blazingly fast speeds.

Because SAS has made grid computing an automatic capability within multiple applications, processing times are greatly reduced. As a result, one can integrate, cleanse and analyze larger volumes of data more quickly.

6. REFERENCES

- [1] Berry M. J. A. and Linoff G. S. 2008. *Mastering Data Mining: The Art and Science of Customer Relationship Management*, Wiley, p. 512.
- [2] Cios K.J., Pedrycz W., Swiniarski R.W., and Kurgan L. 2007. *Data Mining: A Knowledge Discovery Approach*, Springer, p. 606.
- [3] Dunham M.H. 2002 *Data Mining: Introductory and Advanced Topics*, Pearson, p. 315.
- [4] Fayyad U., Chaudhuri S., Bradley P. 1993. *Data Mining and its Role in Database Systems*, vol. 5, no. 6, 914–925.
- [5] Han J., Kamber M., and Pei J. 2012. *Data Mining: Concepts and Techniques – 3rd edition*, Elsevier, p. 740.
- [6] Joseph J. 2004. *Evolution of Grid Computing Architecture and Grid Adoption Models*, IBM System Journal, vol. 43, iss. 4, 624-645.
- [7] Matthew K.O.L., Christy M.K.C, Kai H.L., Choon L. 2006. *Understanding Customer Knowledge Sharing in Web-based Discussion Boards: An Exploratory Study*, Internet Research, vol. 16 iss. 3, 289 – 303.
- [8] Sallman R. L., Hostmann B., Schlegel K., Tapadinhas J., Parenteau J., and Oestreich T. W. 2015. *Magic Quadrant for Business Intelligence and Analytics Platforms*.
- [9] Smith J., Gounaris A., Watson P., Paton N.W., Fernandes A.A.A., Sakellariou R. 2002. *Distributed Query Processing on the Grid*, Springer.
- [10] Sommer D., Buytendijk F., Schlegel K. 2014. *Market Trends: Business Intelligence Tipping Points Herald a New Era of Analytics*.
- [11] Tagarden D.P. 1999. *Business information visualization*, Communications of the AIS, vol. 1, iss. 1, article 4.
- [12] Tapadinhas J. 2014. *How to Architect the BI and Analytics Platform*.
- [13] Zaharia M., Chowdhury M., Das T., Dave A., Ma J., McCauley M., Franklin M.J., Shenker S., Stoica I. 2012. *Resilient Distributed Datasets: A Fault-tolerant Abstraction for In-memory Cluster Computing*, Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, April 25-27
- [14] Gartner Inc. <<http://www.gartner.com>>
- [15] International Data Corporation <<https://www.idc.com>>
- [16] SAS Institute Inc. <<http://www.sas.com>>

[17] Tableau <<http://www.tableau.com>>