Looking for Risk
Applying Graph Analytics to Risk Management
Leading Practices from YarcData
About Chartis Research

Chartis is a leading provider of research and analysis covering the global market for risk management technology. Our goal is to support enterprises seeking to optimize business performance through better risk management, corporate governance and compliance. We help clients make informed technology and business decisions by providing in-depth analysis and actionable advice on the broad spectrum of risk technology offerings. Areas of expertise include:

- Credit risk
- Operational risk and Governance, Risk and Compliance (GRC)
- Market risk
- Asset and Liability Management (ALM) and Liquidity Risk
- Financial Crime
- Insurance risk
- Regulatory requirements including Basel 2, Basel 3, Dodd-Frank and Solvency 2

Chartis is solely focused on risk technology giving it significant advantage over generic market analysts.

Chartis has brought together a leading team of analysts and advisors from the risk management and financial services industries. This team has hands-on experience of implementing and developing risk management systems and programs for Fortune 500 companies and leading consulting houses.

Chartis Research is authorized and regulated in the United Kingdom by the Financial Conduct Authority (FCA) to provide investment advice.

www.chartis-research.com

© Copyright Chartis Research Ltd 2013. All Rights Reserved.

No part of this publication may be reproduced, adapted, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Chartis Research Ltd.

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed.

Please note that the findings, conclusions and recommendations that Chartis Research delivers will be based on information gathered in good faith, whose accuracy we cannot guarantee. Chartis Research accepts no liability whatever for actions taken based on any information that may subsequently prove to be incorrect or errors in our analysis.

See Chartis “Terms of Use” on www.chartis-research.com

RiskTech100® is a Registered Trade Mark of Chartis Research Limited

RiskTech Quadrant™ is a Registered Trade Mark of Chartis Research Limited

Unauthorized use of Chartis Research’s name and trademarks is strictly prohibited and subject to legal penalties
# Table of Contents

1- Executive summary ............................................................................................................................................ 5  
2- Current Challenges in Risk Data Management ............................................................................................... 6  
3- Introduction to Graph Analytics .................................................................................................................. 11  
4- Graph Analytics as an Enabler to Improved Risk Data Management .......................................................... 15  
5- Leading Practices from YarcData ................................................................................................................ 19  
  5.1 YarcData’s offering for financial institutions ............................................................................................. 19  
6- Future outlook ................................................................................................................................................ 24  
7- How to use research and services from Chartis Research ............................................................................ 26  
8- Further reading .............................................................................................................................................. 27
List of Figures and Tables

Figure 1. Modifying existing relational database systems is not an easy process and can take weeks........... 6
Figure 2: Search-based approach to risk and compliance ............................................................................ 8
Figure 3: Bottlenecks and performance issues created by multiple schemas .............................................. 10
Figure 4: How sampling and partitioning obscures patterns in the data .................................................. 11
Figure 5: Conversion of numerous data sources into schema-free graphs .................................................. 12
Figure 6: Cyber-security visualization of a denial-of-service using graph analytics .................................. 13
Figure 7: Opportunities for graph analytics in financial services .............................................................. 15
Figure 8: Urika offers a paradigm shift in data discovery ........................................................................... 20
Figure 9: Data discovery is an iterative process. The faster each iteration completes, the sooner you can make valuable discoveries ................................................................. 21
Figure 10: Urika architecture ................................................................................................................... 22

Table 1: Search vs. discovery ..................................................................................................................... 8
Table 2: Urika and graph analytics checklist ............................................................................................. 25
1- Executive summary

In an old joke, a policeman sees a man scrabbling around near a lamp post late at night. The man explains he’s looking for his keys. The policeman helps for a while and then asks the man if he is sure this is where he dropped his keys. The man replies, “No – but the light’s much better here.”

Financial institutions also use their resources to look for risks where the light is. It’s much easier to find a “known unknown” than an “unknown unknown”. Concentrating on risks that are more easily defined and known is relatively simple and cheap, but ignores the threats that lurk in the dark.

And as we saw in the financial crisis, the risks in the dark have the potential to trigger huge systemic loss events. The financial crisis was triggered by losses that few people saw coming, simply because financial institutions weren’t looking for the interconnected risks and patterns that set off these ‘black swan’ events.

This happened because of culture (the ‘streetlight effect’ referred to above) and the limitations of technology. These two factors feed into each other. Firms focus on known unknowns because they’re easier to find and technology doesn’t get better at looking for unknown unknowns because no-one is looking for them.

Search-based methods have dominated because they provide an easy, relatively cheap way to find known risks. By contrast, discovery-based methods, aimed at discovering unknown risks, have failed to develop because until now they have been impractical and expensive. Technology systems could not examine all of the data, which prevented firms from identifying relationships between risks and discovering unknown risks.

But what if firms could leverage all of the relationships in the data from their entire portfolio? What if they could use data management systems proactively as platforms for interactive discovery? This would revolutionize the way that financial institutions manage risk, allowing them to move beyond simply reacting to known threats and enabling proactive discovery of unknown unknowns.

Graph analytics is a technology that has existed for some time, but lacked the hardware and software to make it a viable option for financial institutions to analyze their vast quantities of data. However, purpose-built graph appliances, powered by supercomputing technology, are now available and can enable financial institutions to use all their data to discover unknown risks. This technology has the potential to transform risk management.

This report discusses why now is the right time for financial institutions to implement scalable data discovery using graph analytics and how technology advances have made these capabilities available to financial institutions and can help them meet key data management and business challenges. The report explains why financial institutions should use graph analytics to complement their current systems. It also covers best practices from YarcData, which Chartis believes is a leading platform for iterative, interactive discovery using graph analytics.
2- Current Challenges in Risk Data Management

In the past five years, the financial sector has woken up to the challenges of data. It had previously been assumed that the risk management challenge meant developing complex analytics and computational engines and that data could be easily collected and fed to risk engines. However, two major factors led to a renewed focus on data: technological advances and the financial crisis.

The march of technology: new capabilities and new challenges

Advances in electronic banking, online and mobile payments, cloud computing, and storage technology have enabled firms to store and transfer more data in more formats at greater speed. However, firms must also process more data more quickly. Unfortunately, more data does not mean more information. As well as tackling the 3 Vs (volume, variety, and velocity), firms must extract information to discover relationships buried under the ever-increasing quantities of data.

Current systems are based on searching for known patterns as shown in Figure 1. However, today analysts require dynamic abilities to look for unpredictable new patterns and relationships. Discovery starts with a hypothesis, which the analyst tries to prove or disprove. Hypothesis testing is a very iterative process. Users need to quickly add more data or change the data to support this process. Once a new hypothesis has been proven, the results can then be used to improve the search criteria.

Financial institutions want to be able to extend their search capabilities to become more responsive and adaptable to a fluid environment. Analysts need the ability to test multiple hypotheses quickly and cheaply, as many hypotheses will be false. Current systems were not designed for this flexibility and are unsuited to discovery.

Figure 1. Modifying existing relational database systems is not an easy process and can take weeks
The financial crisis – Data is not the same as information

During the financial crisis, financial institutions often had the data they needed, but could not turn it into practical information that could identify emerging risks. Individuals had a good view only of their narrowly-defined risk profile within silos or business lines. Senior management and the risk function could not see how assets, liabilities, and risks in different silos were connected. More importantly, as we saw with Lehman Brothers, AIG, Northern Rock, Greece, Ireland, etc., the compound effect of multiple risk types and unintended behavioral consequences could not be foreseen, as it was difficult to define the complex web of relationships. In a more globalized and complex financial sector, the connections were more important than individual factors, but firms were unable to discover them.

For example, Bear Stearns was undone by the rapid deterioration of the US housing market, its effect on their CDOs and CDSS, and the connection of those assets to its overall liquidity profile. The bank was unable to see the risks emerging from its investments in mortgage-backed securities and how those risks would be multiplied by a general decline in housing prices and by consequent margin calls. The failure to discover these risks and anticipate the effect of the demise of Bear Stearns’s hedge funds led to a liquidity crisis at the bank and it was taken over by JP Morgan.

Furthermore, the downturn became a crisis because of misunderstood and unknown networks. The high degree of connectivity between large banks meant that the collapse of Lehman Brothers set off a wave of losses and deteriorating market conditions. As firms lost money to bankrupt counterparties and saw the value of their portfolios plunge, they began to exit certain markets, causing credit and liquidity to dry up, deepening the crisis and putting pressure even on those banks that had little involvement in the subprime market. Both bankers and regulators were unable to identify the pattern of risks that led to the threat of unknown, systemic risk.

How to know what you don’t know

Financial institutions had fallen prey to the ‘streetlight effect’, otherwise known as observational bias, a recurring problem in scientific experiments. It refers to the problem of researchers focusing on experiments in which they can take measurements more easily and, more importantly, know what they are measuring. This limits the scope of what their experiments can find.

This can affect scientific experiments, especially medical testing, where focusing on one variable (e.g. whether a new drug cures a condition) may mean that longer term, more difficult to measure variables (e.g. elevated risk of serious conditions over time) are neglected. Similarly, economists rely on headline macro-indicators, because they cannot look at the whole economy and see how millions of investors, companies, and consumers are interacting. They are limited to measuring what they know and can observe easily.

The financial crisis showed that observational bias exists in risk management as well. Firms focused on a few key variables they could measure, which failed to alert them to real, significant threats. This demonstrated the limitations of approaches that rely on searching for individual data points (e.g. assets, risks, loss events, pre-defined risk indicators, etc.) identified in advance. This is how most risk management systems based on relational databases operate.

Successful risk management needs a dynamic and iterative approach to discover risk by testing hypotheses and complementing search-based methods with discovery methods to find unknown relationships and risks. Table 1, below, compares the different capabilities and advantages of search and discovery processes.
## Table 1: Search vs. discovery

<table>
<thead>
<tr>
<th>Search</th>
<th>Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Easily scalable</td>
<td>Difficult to scale</td>
</tr>
<tr>
<td>Limited and predictable</td>
<td>Open-ended and unpredictable</td>
</tr>
<tr>
<td>Fixed</td>
<td>Adaptable</td>
</tr>
<tr>
<td>Find known risks</td>
<td>Discover new, unknown risks</td>
</tr>
<tr>
<td>Analyze individual, isolated data</td>
<td>Analyze connected data points, patterns, and many-to-many networks</td>
</tr>
<tr>
<td>points and one-to-one relationships</td>
<td>Proactive approach to risk management</td>
</tr>
</tbody>
</table>

Discovery is more conducive to a hypothesis-based approach, but is harder to scale because partitioning data distributes connections between data points, and the resulting cross communication (between nodes) can create a performance bottleneck. Attempting to scale discovery processes using relational databases can cause significant performance constraints, slowing processing speed.

## Data complexity and the limitations of existing technology solutions

Financial institutions have frequently identified data management systems as a key constraint on their goals and as an opportunity to improve their systems. Improved data management is a means to an end for financial institutions, who want to use sophisticated technology capabilities to meet a number of business challenges, including restoring profitability, improving risk management, and meeting regulatory requirements. Firms want solutions that will meet the challenges of data complexity, volume, and velocity and that will enable them to pivot to meet new challenges and exploit new markets quickly. Rigid technology systems can prevent firms from adding new instruments or devising new queries, as they are designed to meet defined use-cases.

## Figure 2: Search-based approach to risk and compliance

- **Load Source Data**
- **Ingest/Load/Derive Data**
- **Detect Patterns**
- **Generate Alerts**
- **Assign & Investigate**
- **Report & File**
Current risk management and compliance systems are capable of performing a number of tasks based on filtering data and carrying out pre-defined searches, as Figure 2 shows. Systems often have a broad range of well-tested policies, procedures, and applications, but have few options for discovery. They are constrained by the difficulty and expense of changing rigid data models, manually investigating across silos, testing new hypotheses, and onboarding new data. At the same time, improved what-if investigations, hypothesis testing, and the ability to augment existing systems are crucial for risk and compliance departments to meet the growing complexity and volume of regulatory requirements and criminal threats.

Breaking down risk management silos will give firms greater flexibility to pivot and will enable a consistently measured enterprise-wide view of risk. Banks will be able to draw connections between risks in different parts of the bank, which will also make regulatory calculations and submissions easier. For example, under Basel 3, being able to identify relationships between credit risk, collateral management, treasury, and ALM will give financial institutions a clearer view of their liquidity profile and will help them to optimize their distribution of highly liquid assets.

There are significant technology barriers to breaking down risk management silos. Relational database systems currently in use are unsuited to this task and their structure is resistant to these integration projects. These systems rely on knowing in advance what data items and formats would be used, so that a schema can be created and defined before being populated with data. As changing the schema is difficult and time-consuming, it is difficult to add new dimensions to existing databases.

As a result, it is often easier for users to add an additional silo (with its own, new schema) alongside the existing system than to integrate new data into the existing database, which would involve adapting or extending the structure of the database. This inflexible legacy architecture requires direct IT support and significant time and resources to modify or add to underlying data and analytics. To provide greater agility, firms avoid costly and lengthy integration processes by implementing add-ons and extra databases, at the expense of a consistent view of risk and the ability to identify patterns and networks.

These problems exist partly because relational database systems were designed 10-15 years ago to perform specific tasks in a relatively low data volume world. In particular, these systems meet the requirements for processing transactions in a single silo and producing results for a single risk type. They still perform these processes relatively well, but struggle to meet requirements for real-time, high data volume, cross-silo risk management.

Data storage is not the problem, as many firms have invested in new, advanced data warehouses to store the higher volumes of data entering their systems. The problems arise in the processing, formatting, and analysis stage of turning the data into information. As Figure 3, below, shows, these processes can create bottlenecks, slowing down the flow of data from source systems to end-users. The systems struggle to integrate large volumes of data that come in a variety of formats and schemas.

Financial institutions face major challenges around large, diverse sets of data, including integrating complex data objects (such as volatility surfaces), unstructured data (including e-mails and voice recordings), and real-time data (which includes integrating tick data with more complex static data, such as reference data). Firms often attempt to partition Big Data into different chunks (by format, by silo, by asset class, etc.). This may help firms to deal with the complexity and create data amenable to search-based methods, but it also means that relationships are missed and the information remains hidden.
Maintaining system performance is important because of the ‘need for speed’. Firms need to respond to market events and changes in the risk profile in real time. Data should not only be available to end-users as needed and on-demand, it should also be available in a practical, informative format. The data environment must also be able to adapt quickly to changing business requirements driven by markets or regulatory events. The bottlenecks and lengthy integration and aggregation processes in relational database systems make this difficult to achieve.

Current systems are square pegs for round holes – the tools designed several years ago do not meet today’s needs. Instead, they decrease performance and only add to existing complexity. To discover all patterns, firms need a unified data management environment to look at all of their data, with all of its connections. They need to be able to use data in any format and take it from anywhere to react to market events more quickly.
3- Introduction to Graph Analytics

What is graph analytics?

While graph analytics is a technology that has existed for many years, it has recently become better known. Its basic concepts will be familiar to many financial institutions. Graph analytics model complex networks of data in a graph representation of nodes and edges. The graph is not just a chart, but a fundamental data structure, in which the nodes represent entities of interest, and the edge represents the relationships between entities. The analysis of nodes and edges provides information on relationships in the data, such as network or link analysis.

Network analysis has been used by a number of financial institutions. However, while graph analytics is similar to network analysis, it is fundamentally different at a technological level. Network analysis sometimes uses graphs to model relationships, but in a simplistic and limited way, and can create performance issues. Efficient relational schemas require prior knowledge of the relationships between database fields. Many network analysis techniques rely on data sampling, instead of using full datasets. These approaches have been successful at finding and analyzing simple patterns and have reached market maturity, but are constrained by their reliance on sampling, which creates three main system performance issues:

- Partitioning graphs means cutting across connections and dividing areas of the graph. This makes it harder to analyze networks in the data and to examine connections.
- Graphs are not predictable and cannot be pre-defined; accessing graph nodes across large clusters can be unwieldy at best and does not work at scale.
- Graphs are inherently dynamic, which means that multiple, changing datasets can be loaded continuously in graph models. Updating and modifying schemas in relational models frequently introduces delays and errors.

There are tools that can do graph analytics and produce graphical results for large graphs of millions of nodes, but not without a considerable reduction in speed. Fortunately, the barriers to using graph analytics on a large scale and at speed have begun to fall.

Figure 4: How sampling and partitioning obscures patterns in the data

The diagram demonstrates how sampling data prevents firms from analyzing the whole graph. Each of the red boxes represents a sample, containing some of the data. However, as the diagram shows, this sampling partitions the data, cutting across connections - preventing effective analysis of the graph.
Graph analytics can now operate in large, shared in-memory data management environments to process high volumes of data swiftly. This demonstrates the difference between graph analytics and graph databases. Graph analytics entails analyzing a graph of connected data points; a graph database entails storing data as a fundamental graphical data structure.

These tools have been designed to meet a data management and analysis problem; they are not designed for a specific area of financial services and therefore can model any kind of network. They provide a number of methods to analyze data and the relationships in the data. They enable ranking of entities within the data and ranking of relationships, as well as clustering analysis and matching analysis, and can provide users with a web of contextual information.

More importantly, graph analytics allows end-users to use the data more interactively. Graph analytics systems are both more iterative and exploratory. Graph databases are schema-free (i.e. relationships between nodes are individually represented, not pre-determined by an overall structure), making them more flexible. It is easier to add data and hierarchies and to eliminate or reduce silos. Figure 5, below, illustrates that some graph analytics applications can take data in any format and convert it into a schema-free graph database environment. The graph can be built up in an iterative fashion, allowing users to use their discoveries to direct their graphs.

**Figure 5: Conversion of numerous data sources into schema-free graphs**

This process means that graph analytics programs can provide users with an IT platform for creative discovery and an exploratory process of investigation into the data. A large, shared, in-memory environment means users do not face performance delays due to graph traversal or missing information due to sampling. The schema-free structures can use data from anywhere, making graph analytics systems more adaptable to changing requirements, and do so more quickly at lower cost.
How does graph analytics differ from conventional systems?

Graph analytics differs from traditional data management systems because:

- It is oriented towards relationships and connections and is designed to model and analyze many-to-many relationships.
- It can direct queries towards relationships and connections (a particularly complex and data-processing intensive process), so firms can discover and explore networks.
- Queries can be pattern-based.
- A pattern can be easily visualized and extended through the data by modeling it as a graph.
- Users can move from a reactive to a proactive risk management approach.

Figure 6, below, shows the illustration of complex patterns in a denial-of-service for a cyber-security application. Each node represents an IP address used to launch a DNS attack and uses the clusters to identify the IP addresses of those coordinating attacks.

**Figure 6: Cyber-security visualization of a denial-of-service using graph analytics**
What are the practical applications of graph analytics?

The potential of graph analytics for data discovery has already been recognized in a number of industries that face significant challenges around big data. The most publically visible use of graph analytics has been in social media, notably through the introduction of Facebook Graph Search. Graph analytics can assist many organizations dealing with large and complex datasets, including doctors modeling mortality rates and biological networks, pharmaceutical companies modeling relationships between different proteins, and law enforcement agencies mapping criminal trends and terrorist networks.
4- Graph Analytics as an Enabler to Risk Management in Financial Services

Graph analytics offers a range of capabilities to its users in any industry, but for financial institutions, it offers a paradigm shift in the way that they analyze risk data. It will enable firms to investigate their data in greater depth and in ways that previously had been impossible. Graph analytics now offers financial institutions the opportunity to:

• Address the limitations of sampling.
• Break down organizational silos and look at relationships across the enterprise.
• Move from search-based to discovery-based risk management, allowing financial institutions to investigate risk proactively and uncover new risks.

As Figure 7 shows, it will help firms to meet a number of existing and emerging challenges. In particular, graph analytics will enable financial institutions to expand their conception of how they can manage risk and will enable a more proactive approach. Financial institutions will be able to consider risk as a product of patterns and networks in different departments in their organization.

**Figure 7: Opportunities for graph analytics in financial services**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Current</th>
<th>Emerging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk - Market/Credit</td>
<td>Departmental computations &amp; aggregations</td>
<td>Risk as a function of the connections between networks of counterparties</td>
</tr>
<tr>
<td></td>
<td>Fragmented view of operations risk and controls</td>
<td>Holistic, connected view of causes, events and impact</td>
</tr>
<tr>
<td></td>
<td>Prescriptive recognition of known behaviors</td>
<td>Proactive discovery of new patterns of suspicious behavior</td>
</tr>
<tr>
<td></td>
<td>Profitability, retention and marketing</td>
<td>Follow connections across customer social networks to drive revenue</td>
</tr>
<tr>
<td></td>
<td>Silo'd examination of network incidents</td>
<td>Find hidden threats through a connected view of network activity</td>
</tr>
<tr>
<td>Risk - Operational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Insight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyber Security</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The collapse of the hedge fund Long Term Capital Management (LTCM) is a classic example of the danger of failing to find connected risks or discover new risks – and the opportunities provided by graph analytics. LTCM famously employed Nobel Prize winners Myron Scholes and Robert Merton and believed that it had adequately mitigated all of its risk through its risk managers’ models. However, these models were designed to measure known, individual risks.

When Russia defaulted on its debt in 1998, it set in train a series of events that revealed LTCM’s exposure to unknown correlations and risks. LTCM was unaware that its attempt to diversify its holdings had been undermined by the strong correlations between credit trades in different geographic markets, which meant it faced a much greater risk than anticipated. Its exposure was increased by the complex web of swaps it had built up to hedge against risk. These swaps were, however, also linked to each another and to LTCM’s credit trades, some of which had been used as collateral for the swaps, causing further margin calls, losses, and a reduction in liquidity.

Finally, the strength of the relationships between LTCM and investment banks had been underestimated. These banks belatedly began to understand the extent of their exposure to LTCM and the systemic risk. This led to a bail-out for LTCM from Wall Street. If LTCM had had the technology to enable discovery, it could have found these unknown risks and altered its strategies or increased its capital stocks and/or liquidity buffers to mitigate them.

This example helps to demonstrate that at both micro and macro levels, risk management is a network problem and therefore needs a network solution. Network problems are often best represented and analyzed by graph-based methodologies and technologies.

**A step change in hardware capabilities**

The shift in available capabilities does not come simply from better modeling and discovery. Hardware plays a key role. Commodity hardware systems, such as those based on the popular Intel x86 architecture, can model graphs in-memory, but do not scale beyond more than four terabytes of RAM based on currently available systems. As a result, users are forced to partition the graphs and cannot efficiently extend them. These systems also lack the parallel processing power and bandwidth to enable real-time discovery in all of a firm’s data.

As the volume of data has grown and the need for speed has risen, the ability of current systems to leverage graphs has diminished, but dedicated graph analytics platforms can enable users to discover and extend patterns across business lines and silos in large datasets. A dedicated graph analytics platform with large shared in-memory capability can reduce these performance problems. The technology barriers that previously made graph analytics less attractive and too expensive have begun to come down. Until now, financial institutions have known what they needed to identify relationships and networks, and have wanted to be able to carry out these kinds of operations, but lacked an efficient platform to do so.

**New hardware capabilities**

Modeling graphs in a large, globally shared in-memory environment means that as queries don’t need to traverse the data, there are no networking performance bottlenecks, allowing users to follow graphs through the environment to make discoveries. The schema-free environment will make adding data easier and improve performance. This functionality will help firms to use graph analytics at enterprise scale and in real-time.
What specific functionalities are available?

The use of a data discovery appliance for graph analytics may be more significant for risk management and financial services than in any other sector. It offers financial institutions the opportunity to solve problems that have the following characteristics:

- **High volumes of data and the need to look at all of the data at once** – A simple data model and support for multiple data types makes this possible. This means that graph analytics will, for example, be useful for enterprise risk management (ERM). Risk managers can use all of the data to discover interconnected risks and build up an enterprise risk profile, including additional risks created by networks, rather than just aggregating different silos.

- **Frequent updates and new data source additions** – Graph analytics will permit this because additions can be made quickly without mapping or formatting. Users can augment data and definitions, add new instruments, new transactions, and new assets on the fly, and provide incremental, real-time updates to risk measurements. For example, financial institutions could add new transactions, perform a stress test quickly, and examine the results to see if new risks had been created.

- **High connectivity/correlations between different entities in the data** – This includes any problems that cross silos and include a large number of ERM calculations, such as searching for correlations between different assets or connections between credit risks in different business units.

- **Requirement to test multiple hypotheses quickly and cheaply** – For example, stress testing and scenario analysis could be expanded and accelerated through the use of graph analytics.

- **Requirements for high performance** – This would mean any problem in which risk managers needed to analyze a high volume of data quickly. For example, it would include carrying out integrated stress-testing in real time and examining the results.

Chartis believes that, in the short to medium term, graph analytics is most likely to be adopted in areas such as operational risk, financial crime, and enterprise risk, and when carrying out stress testing and scenario analysis for financial institutions. Financial institutions will be able to use graph techniques such as path analysis (to determine the shortest non-obvious path between two entities), centrality analysis (to identify who/what is at the center of a network), and cluster analysis. For these risk areas, relationships and networks are crucial to discovering risks, especially those that would otherwise be difficult to detect.

**Graph analytics for financial crime risk management and compliance** – Examining networks can help firms to identify suspicious activity more easily and create more accurate thresholds for investigation. Using graph analytics to examine payment networks or networks of potential fraudsters can reveal the existence of false identities (e.g. by examining people connected to a single address) and fraud rings. It can also help link numerous red flags that, while not enough to trigger action by themselves, when brought together provide a clear indication of fraud, corruption, or money-laundering. Patterns of suspicious payments or transactions will also be easier to explore through graph analytics.

**Graph analytics for operational risk management** – Graph analytics will be able to uncover unknown risks, which are a particular problem in operational risk because there is less historical data. For example, it can be difficult to estimate the potential risk of exposure to cyber-attacks because successful attacks are relatively rare and the nature of the threat is constantly changing. Hackers often use multi-step attacks to compromise and disable critical systems and steal data. Counting and assessing individual vulnerabilities is not enough because of the rapidly evolving and multi-vector threats, which create new vulnerabilities and multiply their risks. Graph analysis can examine the relationships between vulnerabilities and between different machines in the network, allowing the firm to build up an attack graph that will help it to discover how a multi-step attack might take place. Graph analytics will allow this process to be automated, fast, and repeatable, giving firms a clearer view of security risk, as part of operational risk.
**Graph analytics for enterprise risk management** – ERM will be a prime candidate for graph analytics, as it attempts to build an overall risk profile for the firm. Having the ability to examine the relationships between different assets, liabilities, and risks will allow risk managers to provide senior management with a more accurate view of the firm’s risk profile. Graph analytics will also be more likely to reveal significant, unknown risks caused by relationships in the organization.

**Graph analytics for stress testing and scenario analysis** – Impact analysis is often described in terms of a network effect, which is best modeled as a graph of nodes and edges. The computational power needed to carry out stress testing on an integrated basis is beyond the traditional database technologies and infrastructures. However, graph analytics will allow firms to carry out impact analysis as a fast and repeatable process. This will enable firms to run multiple stress tests and scenarios to identify unknown risks and hidden patterns and networks in their data.

**Graph analytics for customer intelligence** – Banks can identify cross-selling and up-selling opportunities by examining networks in their existing customer base. Looking at communities of individuals can give banks a clearer perspective of their customers’ needs and wants. Improved customer intelligence can then inform credit risk and financial crime risk systems.

These are areas in which financial institutions need tools to help them to deal with an increased amount of data, particularly unstructured data. The high performance of purpose-built graph analytics platforms means that networks with a high number of data points and many-to-many relationships can now be investigated effectively. Firms can utilize purpose-built graph analytics platforms that allow them to explore hundreds of terabytes of data at high speed.

**Complementary, not a substitute**

Graph analytics can add value to existing systems and architectures, rather than being a like-for-like substitute. Firstly, it will allow them to focus on the tasks they are designed for, including monitoring and searching for risks, making these processes faster. Secondly, the results of graph analysis can be used to update traditional systems. For example, if graph analytics reveals a new kind of fraudulent activity, it can be added to the search parameters so the new fraud can be detected by established transaction monitoring systems. Improved knowledge of the causes of operational failures could help firms develop more accurate and responsive Key Risk Indicators (KRI). Crucially, graph analytics is relatively accessible to non-experts. Interactivity and visualization are key aspects of graph analytics that allow business users to investigate the data and discover new connections.
5- Leading Practices from YarcData

In the opinion of Chartis, YarcData’s Urika is currently one of the most sophisticated data discovery platforms for graph analytics available today. Launched at the start of 2012, YarcData is a new Cray company, set up to provide a purpose-built appliance for graph analytics to the Big Data market. Cray is well-known for its high performance super computers, and YarcData focuses on delivering scalable graph analytics for enterprises to gain business insight by discovering unknown relationships and non-obvious patterns in big data. YarcData is based in the San Francisco bay area.

YarcData has leveraged its Cray background in supercomputing to create a graph analytics appliance based on innovative, powerful technology, which has been proven in production use at some of the largest and most complex organizations. Urika is designed to be an enterprise-ready, real-time solution for financial institutions. It is a flexible platform with a range of applications and can solve problems across a number of risk types. Urika can be used to detect new fraud patterns, understand or model systemic/counterparty risk, identify potential insider trading relationships, and improve cyber-security. It also has applicability in market-sensing and customer-profiling solutions.

5.1 YarcData’s offering for financial institutions

Urika is a graph analytics appliance that is purpose-built for discovery in big data and is well-suited for tackling a number of challenges facing financial institutions. It offers:

- **A platform for discovery** – Urika enables users to interact with the data to discover patterns and networks and to test hypotheses quickly and cheaply. The graph database can run multiple queries at once to enable quick, effective discovery.
- **The ability to look at all the data** – Urika allows financial institutions to look at all of their data using a shared, in-memory environment with up to 512 terabytes of RAM. This means the data does not have to be partitioned and firms can analyze links across different entities and business lines. This enables users to discover non-obvious patterns and risks.
- **The ability to add new data** – Urika uses a schema-free graph database to take data from anywhere and quickly add new and diverse data sources, without needing time-consuming processes to map or format the data.
- **Scalability** – Urika is purpose-built to be able to process high volumes of data. It can scale to very high data volumes to meet the needs of financial institutions to investigate patterns and risks in large datasets.

The Urika software stack was crafted with several goals in mind:

- Create a standards-based appliance to support enterprise graph analytics workloads.
- Facilitate migration of existing graph workloads onto the Urika appliance.
- Allow users to dynamically update the graph database with new incoming data.
- Enable ad-hoc queries and pattern based searches of the dynamic graph database.
- Provide a software layer to enable organizations to manage the appliance like any other LINUX environment.
- Provide real-time performance and scalability.

What differentiates Urika from the many graph databases available today is its ability to enable data discovery at scale and on an interactive basis. As shown in Figure 8, Urika redefines how financial institutions can manage risk by finding new relationships in their existing data. In addition, Urika’s graph database facilitates adding new data sources, thereby enhancing the quality of existing datasets.
Typically, an analyst may start their typical day with a hunch about their data and how it might be enhanced, say by adding demographic data to validate a hypothesis. Adding a new data source to Urika is fairly straightforward and takes a fraction of the time it does to modify an existing data warehouse. The analyst can now begin to explore how this additional data might expose new relationships and patterns that were non-obvious before. This can either prove or disprove the initial notions that the analyst had harboured and can force a new path of exploration, but the key point here is that Urika can significantly accelerate the pace of data discovery, by enabling many such iterations to occur within the span of a typical business day.

**Figure 8: Urika offers a paradigm shift in data discovery**

As shown in Figure 9 data discovery is inherently an iterative process that involves trying out various hypotheses. Analysts sequentially test several hypotheses prior to arriving at a valid one. It is a trial-and-error process and Urika can speed this up by validating a hypothesis quickly, in seconds and minutes, instead of the hours and days it can take with traditional tools when working with complete datasets. The faster each iteration completes, the sooner a discovery can be made and refined.
Figure 9: Data discovery is an iterative process. The faster each iteration completes, the sooner you can make valuable discoveries.

In the amount of time it takes to explore one hypothesis, we can now explore thousands of hypotheses, massively improving our success rate.

Dr. Ilya Shmulevich, Professor, Institute for Systems Biology.
Graph analytics platforms are meant to augment an institution’s risk management capabilities. When a new discovery is made with Urika, this new information can be incorporated into the institution’s existing operational systems to in effect productionize the learning gained from discovery. In this way, each tool is best leveraged for its intended purpose.

Figure 10 shows one scenario how Urika can be incorporated into existing IT infrastructures that include data warehouses and other big data appliances, as well as Hadoop clusters. Urika can accept data straight from core systems, or from existing relational databases and data marts with minimal processing. Adding additional data sources is also relatively simple and straightforward. Furthermore, Urika has been designed to easily integrate with popular applications and visualization tools that analysts are already familiar with.

Figure 10: Urika architecture

Urika’s graph database is based on Apache’s open-source Jena software combined with YarcData’s own graph database optimizations. Urika uses the industry standard RDF (Resource Description Framework) format for storing data, and SPARQL (SPARQL Protocol and RDF Query Language). RDF represents a relationship (an edge in a graph) using a triple (<subject>, <predicate>, <object>), where each member of the triple is identified individually. Data from different sources can be added by simply combining the sets of RDF triples, without schema issues, making the database quite flexible and enabling users to quickly and easily add new data. SPARQL is a query language that allows users to create unambiguous queries comprised of triple patterns, conjunctions, disjunctions, and optional patterns. SPARQL is considered one of the key technologies that supports true discovery.

Urika also supports the use of inferencing to automatically discover new relationships in the graph database by applying logical rules of deduction on the data. This allows for:

- Reconciliation – The use of RDF triples makes it easy to combine data from different sources.
- Inference of new relationships using deduction.
- Chaining – A rule can add a relationship to the database that can trigger one or more other rules to add other relationships.
Case Study 1 – Emerging risks

YarcData is helping outfit the “Emerging Risks” department of a financial institution with the technology it needs to exploit the data in the firm’s existing technology systems. The role of this department in the firm is to take a discovery-based approach to uncovering potential risks and patterns. Currently, the department carries out this process manually. The department has no responsibility within the organization relating to specific risk types (e.g. credit risk, operational risk, financial crime etc.) and also has no designated systems of its own.

To discover emerging risks, analysts formulate their query and extract the data from the firm’s core systems by issuing a data source request to the IT department. This process takes several weeks, if not months. The analysts are often asked what format the data needs to be in, but often don’t know what would be appropriate.

YarcData’s technology capabilities improve performance in the following ways:

- **Data management** – YarcData’s solution will allow the department to take data from across the organization in any format to enable it to find risks throughout the organization and patterns between risks.
- **Find unknown patterns** – The department will use Urika to look for evidence that entities might be using the firm to launder funds or to conduct other criminal activities. Using the graph analytics capabilities, they aim to identify users who might be human traffickers (or engaged in other criminal activities) by examining ATM withdrawal patterns and comparing them to known trafficking routes. This will allow them to uncover a previously unknown pattern and identify an emerging risk.
- **Strengthening existing systems** – Using discoveries made in the data, the Emerging Risks department can send notifications to other departments (e.g. fraud, operational risk, or AML) to update their queries and controls to make search-based efforts smarter.
- **Efficiency** – The firm aims to move from a slow, manual and resource-heavy risk discovery process to a much quicker, automated one that produces better results, allowing the firm to do more with less.

Case Study 2 – Improving Know-Your-Customer processes

YarcData can help banks identify new requirements for its Know Your Customer (KYC) processes and can help to improve these processes. Quite often banks struggle to process and effectively use a large amount of incoming data. Banks want to understand what attributes are important when processing new customers and transactions.

Most banks already filter data using official watch-lists and are aware that some illicit transactions are eluding the filters, but rarely know how to find them efficiently. The bank’s attempts to find these false negatives and refine their filters were hampered by the size of data, the size of lists, and its existing technology system’s inability to scale. As a result, its attempts to discover patterns were largely manual and very resource-intensive.

Using Urika, a bank can look at all of its data and leverage Urika’s discovery capabilities to identify connections between individuals, to connect data from multiple lists, and to discover patterns of individuals and transactions. Urika offers the ability to investigate the data on a large scale and allows it to bring in new data on-the-fly, which is a key differentiator.

A bank can discover new patterns quickly and efficiently by looking at all of its data and improve its filtering and detection processes. Urika’s speed and ability allows banks to use an iterative process to test hypotheses quickly and cheaply. It could improve efficiency and redeploy resources by reducing the amount of manual work needed to discover patterns and improve the effectiveness of KYC procedures.
6- Future outlook

Graph analytics platforms provide financial institutions with the opportunity to solve many of the challenges they currently face. Many firms may nonetheless be skeptical about implementing graph analytics, as they have seen new technologies that promise to revolutionize their approach to risk data that ultimately disappoint. New technologies have, in the past, either become outdated themselves, or have simply added to the complexity. Many firms question whether graph analytics will be any different.

For financial services the trends of higher data volumes, variety, and velocity look set to continue and compliance requirements will continue to grow. The marketplace is unlikely to reduce its demand for speed or the level of complexity in a more globalized financial world. Firms will continue to need tools that can identify hidden risks and discover information from a growing mass of data by separating signals from the noise.

The YarcData Urika appliance shows the value that a data discovery graph analytics platform can provide for financial institutions. Looking at all of their data in unison in addition to being able to take data from a multitude of sources facilitates more discovery-based processes, and improves a firm’s ability to find unknown patterns and emerging risks. Advances in technology have enabled the use of a shared, in-memory environment that allows firms to use graph analytics without worrying about the performance issues that have dogged previous attempts to implement discovery-based platforms.

Firms that adopt graph analytics now can reap considerable benefits. Graph analytics offers a competitive advantage by enhancing the risk data analysis that financial institutions already carry out. At a time when risk departments find their risk management abilities stymied by regulatory burdens, firms may welcome the emergence of a new tool that enhances their flexibility and allows them to uncover unknown patterns ahead of their competitors.

Graph analytics also offers firms the ability to carry out genuinely effective enterprise risk management. Enterprise risk management is ultimately a network problem, as it is based on the links between the different risks across a financial institution. It therefore requires a solution that can identify and analyze networks and graph analytics is one approach that can effectively do this.

Graph analytics makes an important shift from reactive to proactive risk management. By becoming risk intelligent, firms gain major competitive advantage discovering what was previously unknown. To truly achieve business intelligence, organizations must recognize their potential for technological change and enthusiastically rise up to meet it.
Table 2: Urika and graph analytics checklist

<table>
<thead>
<tr>
<th>Is Urika right for you? Has your organization been faced with any of these data analytics challenges?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do some of your analytics involve looking for patterns of relationships in the data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you currently have a desire to look for “unknown unknowns” or relationships in your data that you suspect are there, but are unable to define explicitly?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you find that the analysis you need to do frequently involves changing data schema because you need to add new relationships or new data sources?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you having performance challenges because the data schema never seems to be right for the queries you want to run today?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you find it difficult to predict what sort of a data model or a schema you will need for your analytics?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it difficult to know all of the questions to ask, for a given subject area?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are there diverse data requirements from multiple sources – structured, unstructured or semi-structured?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it difficult to predict what data is needed?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would interactive data exploration and near real-time response on 10’s of TB of data allow your data scientists to find patterns that are otherwise lost to sampling?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On new management requests, would iterative hypothesis testing with quick cycle time lead to higher quality and confidence in your results?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there a need for both forward and backward looking analytics, and prescriptive analytics within the organization?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you ever wanted to explore complex queries that involve multiple database joins or other complex operations without having to re-architect your data warehouse?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you discovered all the low-hanging fruit and now need to explore complex relationships in your data e.g. Connect-the-dots type problems?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>You think advanced analytic techniques like community detection, path analysis, clustering and influencer networks would improve monetization but haven’t had an easy way to test this out?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you answered Yes to any of these questions, then your organization may benefit from real-time data discovery on your entire data that only Urika™ can offer.
7- How to use research and services from Chartis research

In addition to our flagship industry reports, Chartis also offers customized information and consulting services. Our in-depth knowledge of the risk technology market and best-practice allows us to provide high quality and cost-effective advice to our clients. If you found this report informative and useful, you may be interested in the following services from Chartis.

For risk technology buyers

If you are purchasing risk management software, Chartis’s vendor selection service is designed to help you find the most appropriate risk technology solution for your needs.

We monitor the market to identify the strengths and weaknesses of the different risk technology solutions, and track the post-sales performance of companies selling and implementing these systems. Our market intelligence includes key decision criteria such as TCO (total cost of ownership) comparisons and customer satisfaction ratings.

Our research and advisory services cover a range of risk and compliance management topics such as credit risk, market risk, operational risk, GRC, financial crime, liquidity risk, asset and liability management, collateral management, regulatory compliance, risk data aggregation, risk analytics and risk BI.

Our vendor selection services include:

• Buy vs. Build decision support
• Business and functional requirements gathering
• Identification of suitable risk and compliance implementation partners
• Review of vendor proposals
• Assessment of vendor presentations and demonstrations
• Definition and execution of Proof-of-Concept (PoC) projects
• Due diligence activities

For risk technology vendors

Strategy

Chartis can provide specific strategy advice for risk technology vendors and innovators, with a special focus on growth strategy, product direction, go-to-market plans, and more. Some of our specific offerings include:

• Market analysis, including market segmentation, market demands, buyer needs, and competitive forces
• Strategy sessions focused on aligning product and company direction based upon analyst data, research, and market intelligence
• Advice on go-to-market positioning, messaging, and lead generation
• Advice on pricing strategy, alliance strategy, and licensing/pricing models

Thought Leadership

Risk technology vendors can also engage Chartis to provide thought leadership on industry trends in the form of in-person speeches and webinars, as well as custom research and thought-leadership reports. Target audiences and objectives range from internal teams to customer and user conferences. Some recent examples include:

• Participation on a “Panel of Experts” at global user conference for leading ERM (Enterprise Risk Management) software vendor.
• Custom research and thought-leadership paper on Basel 3 and implications for risk technology
• Webinar on Financial Crime Risk Management
• Internal education of sales team on key regulatory and business trends and engaging C-level decision makers

Visit www.chartis-research.com for more information.
8- Further reading

- Anti-Money Laundering Solutions 2013
- Operational Risk Management Systems for Financial Services 2013
- Enterprise GRC Solutions 2012
- Collaborative Risk Management
- Basel 3 Technology Solutions 2012: Horses for Courses
- Global Risk IT Expenditure 2011
- RiskTech100® 2012

For all of these reports see: www.chartis-research.com