Social-Network-Sourced Big Data Analytics

Wei Tan • *IBM T.J. Watson Research Center*

M. Brian Blake and Iman Saleh • *University of Miami*

Schahram Dustdar • *Vienna University of Technology*

Very large datasets, also known as big data, originate from many domains. Deriving knowledge is more difficult than ever when we must do it by intricately processing this big data. Leveraging the social network paradigm could enable a level of collaboration to help solve big data processing challenges. Here, the authors explore using personal ad hoc clouds comprising individuals in social networks to address such challenges.

In recent years, the quantity of information generated by business, government, and science has increased immensely — a phenomenon known as the *data deluge*. In business, Walmart's transactional databases are estimated to contain more than 2.5 petabytes of data consisting of customer behaviors and preferences, network and device activity, and market trends data.[1] In the military, US Air Force drones collected approximately 24 years' worth of video footage from Afghanistan and Iraq in 2009.[1] In science, the Large Hadron Collider (LHC) facility at CERN produced 13 petabytes of data in 2010.[2] Moreover, sensor, social media, mobile, and location data are growing at an unprecedented rate. In parallel to this significant growth, data are also becoming increasingly interconnected. Facebook, for instance, is nearly fully connected, with 99.91 percent of individuals on the social network belonging to a single, large connected component (see http://arxiv.org/abs/1111.4503).

This astonishing growth and diversity have profoundly affected how people process and interpret new knowledge. Because most of this data both originates and resides in the Internet, one open challenge is determining how Internet computing technology should evolve to let us access, assemble, analyze, and act on big data. We believe that data are first-class citizens in the Internet landscape. The collaborative interplay between data and computation infrastructure is vital for enabling low-latency and high-throughput analytics on big data.

Advances in social networks and analytics span many Internet-based computing paradigms, including cloud and services computing.[3] Currently, most social networks connect people or groups who expose similar interests or features. In the near future, we expect that such networks will connect other entities, such as software components, Web-based services, data resources, and workflows. More importantly, the interactions among people and nonhuman artifacts have significantly enhanced data scientists' productivity. Big data analytics can accumulate the wisdom of crowds, reveal patterns, and yield best practices. For a real-world example, in recent events related to the 2013 Boston Marathon bombings, social networks of marathon participants and general high-performance computational techniques were combined to cluster and analyze large sets of candid photos and video shots — ultimately leading to the discovery of the perpetrators. This example exemplifies how cloud-oriented processing techniques can meet computational needs, while analytics are enhanced by the special expertise of social network participants.

The astonishing growth and diversity in connected data continues to profoundly affect how people make sense of this data. We can define this interplay as a virtuous circle in which

- *connected people* produce a continuous data stream that's deposited into a repository of *connected data*;
- individuals or business entities might conduct big data analytics on these *connected data* by leveraging ad hoc clouds or *connected computers*; and
- analytics on the big data from these *connected computers* generates intelligence that subsequently proliferates back to connected people.

As Figure 1 illustrates, this system is continually evolving, as is the knowledge that the interaction generates. Here, we show that the collaborative interplay of connected computers and connected people has opened new avenues with regard to how humans interpret connected data. In fact, connected data is the confluence where social networks and clouds are presented as a solution for big data analysis.

Connected People: Social Networks and Big Data

Recent social networking websites such as Twitter, Facebook, LinkedIn, YouTube, and Wikipedia have not only connected large user populations but have also captured exabytes of information associated with their daily interactions. Social networking has its beginnings in the work of social scientists in the context of human social networks, mathematicians and physicists in the context of complex network theory, and, most recently, computer scientists in the examination of information or Internet-enabled social networks.[4] We can thus separate major research challenges into these areas.
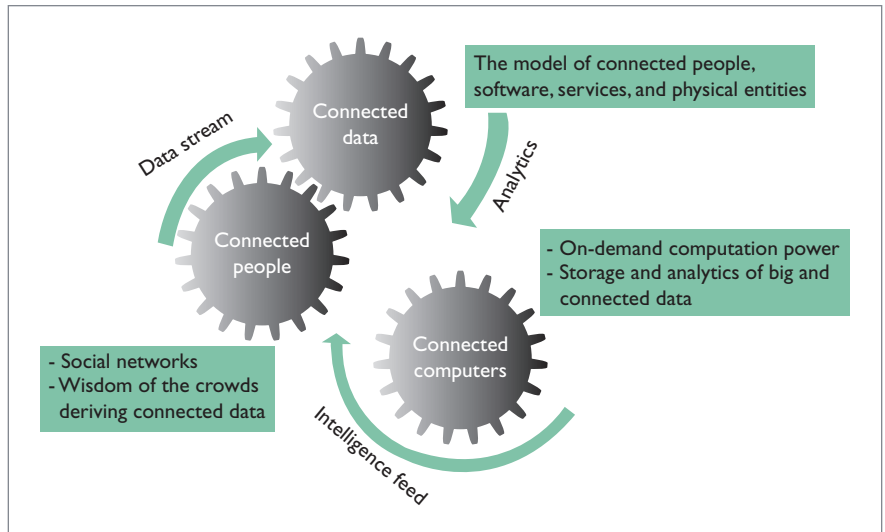

Figure 1. The virtuous circle. Connected people produce a data stream that's analyzed by connected computers, and the intelligence such an analysis generates proliferates back to connected people.

Humanistic Social Networks

Stemming back to the 1920s, social scientists have investigated interpersonal relationships as they relate to the larger network topography of societal groups of interrelated humans. These studies have attempted to systematically devise relationships' strength and have implicitly determined how trust plays into those relationships' interconnections. In managing these networks, social scientists and sociologists have employed several methods.[5] Modeling approaches include network-oriented data collection, block modeling, network-oriented data sampling, diffusion models, and models for longitudinal or emerging data. Measurements include centrality measures for groups, cross-network assessment or correspondence analysis for two-mode networks, and statistical assessment of the p^* model.

Complex Network Theory

Mathematicians and physicists perform some of the same analysis as social scientists but concentrate on the network structure's more quantitative aspects.[6] The emergence of social behavior is derived from the natural quantitative connections between nodes and links within a network. Given that network structure is irregular, complex, and dynamically evolving in time, the main focus for complex network theory is the development of principled, mathematical approaches that assess networks of millions of nodes. Furthermore, mathematicians and physicists derive insight from biological systems that form in nature. A significant vehicle for deriving these networks' behavior is the analysis of path lengths and the clustering of related path structures. Complex networks can be represented in their most fundamental forms as graphs or small-world networks, but more intricate topographies are represented as weighted, random, power-law, or spatial networks. One common approach for managing these networks that's shared with computer scientists is *spectral graph partitioning*, which determines the minimal number of edges between two sets of vertexes within a graph. *Hierarchical clustering* is an effective method for networks in which a priori knowledge of the number of communities is lacking. This approach attempts to divide nodes into clusters where the connections within the cluster are more closely

related than the connections to nodes assigned to a different cluster. Other approaches attempt to look for the largest distance between nodes until clusters are naturally formed.

Information Networks and Social Networking

Computer scientists and information engineers have combined the initial work on social and complex networks and mapped them onto networks representing information-systems-oriented environments. Many studies investigate a fundamental question: "Do online social networks resemble or behave in similar ways as people in real-world situations?" Computer scientists have employed hybrid assessment approaches similar to the traditional methods used in sociology and computational sciences. Web graph analysis, for instance, attempts to integrate the nuances of the Web when considering network analysis.

Social Networks as Big Data

Understanding social networks evolves into a big data problem when business, management, or information systems specialists hope to predict behavior to ultimately enhance marketing, sales, and online commerce. Many social networking sites have between 10 and 200 million users, so data sampling is central to most studies. Although significantly time-consuming, gaining insight from the entire dataset might provide the most optimal solutions. Big data is usually characterized by the "three Vs" — that is, volume, velocity, and variety.[7] In terms of volume, at the end of 2011, Facebook had 721 million individuals and 68.7 billion friendship edges (see http://arxiv.org/abs/1111.4503). In terms of velocity, Twitter and Facebook respectively generate 7 Tbytes and 10 Tbytes of data daily. These data also need to be processed at the speed of thought. For example, on 11 November 2012, a sales event at TaoBao, the largest online shopping marketplace in China, generated 100 million transactions and reached a peak transaction rate of 205,000 per minute (see http://tech.sina.com.cn/i/2012-11-12/00207788375.shtml). In terms of variety, data today come from various sources, ranging from surveillance videos, to satellite images, to mobile tweets, to sensors and meters in the power grid.

Connected Computers: Advances in Scale-Out Systems

Given the astonishing amount of data being produced and the need to store and process them economically, organizations are widely adopting scale-out rather than scale-up systems to acquire and interpret data. Key features of the scale-out pattern include commodity server clusters, share-nothing architecture (no shared memory, storage, and so on), a TCP/IP network connection, and a parallel programming framework such as MapReduce. Cloud computing, which offers scale-out and on-demand computing resources in a pay-per-use manner, is an ideal technology to enable big data for mainstream uses. For example, Netflix stores movies and TV shows, and Dropbox stores customers' files, both in Amazon's Simple Storage Service (S3). Yelp not only uses Amazon's storage but also Amazon Elastic MapReduce to power its user-behavior analytics. Microsoft Windows Azure and IBM SmartCloud Enterprise+ offer similar functions. Startup companies such as Cloudera, Hortonworks, and MapR Technologies are building value-added software and solutions on top of the Apache Hadoop ecosystem.

In recent years, scale-out data stores, popularly referred as NoSQL systems,[8] are rapidly gaining popularity as a potential solution to support Internet-scale applications. These stores include commercial systems such as Amazon's DynamoDB, Google's BigTable, and Yahoo's PNUTS, as well as open source ones such as Cassandra, HBase, and MongoDB. These stores usually provide limited APIs (create, read, update, and delete operations) compared to relational databases, and focus on scalability and elasticity on commodity hardware. Such platforms are particularly attractive for applications that perform relatively simple operations while needing low-latency guarantees as they scale to large sizes. NoSQL stores offer flexible schema and elasticity to overcome relational databases' limitations. However, in doing so, they trade off full ACID guarantees. Clearly, several challenges exist for computational systems that process big data.

Data Models and High-Level Abstraction

Relational models and SQL provide an abstraction layer between the database's physical layer and the application layer. This feature lets users specify a query in a language-dependent and declarative manner, while a query engine schedules and optimizes its execution. No similar solution exists for big data analysis. Instead, NoSQL data stores offer various forms of data structures — such as document, graph, row-column, and key-value pair — that are directly exposed to users. So, users must understand data's physical organization and employ vendor-specific APIs to manipulate these data. Current state of the art attempts to devise a SQL layer on top of NoSQL, but without an abstract data model, this effort is ad hoc and limited to the underlying technology.

Incremental Processing and Approximate Result

Volume and velocity impose contradictory requirements on big data systems. A large volume of data is injected into such a system at a high speed, while analysis and interpretation must occur at the same pace. In traditional business intelligence (BI) analytics,[9]

transactional data is processed initially on an *online transaction processing* (OLTP) system before flowing through an *extract, transform, load* (ETL) process in a batch mode. Eventually, data are loaded into an *online analytical processing* (OLAP) data warehouse, where they're analyzed to provide strategic insights. This OLTP-ETL-OLAP approach trades timeliness for accuracy, given that a long delay occurs between when data becomes available and insight generation.

In some big data applications, such as financial fraud detection and market promotion, long delays aren't tolerable. A newly emerged paradigm called *stream computing* enables continuous queries over streaming data such as social media feeds and call data records. Stream computing opens a gateway to real-time analytics, but a few challenges remain. One is the interplay between building the batch mode model and sensing the real-time streams. On one hand, the accumulated historical data in the data warehouse can help information specialists build a statistical model to guide stream processing – for example, decide which features to observe and help set the reacting threshold. On the other hand, the newly arrived data from the stream system should be leveraged to tune the model to reflect the recent trends. An incremental data processing and model-tuning mechanism is vital to this interplay.

With respect to the volume-velocity challenges, another perspective is to provide approximate, just-in-time results to queries, or prioritize different queries by allocating a varying amount of resources.[10] As such, different data consistency levels are possible in which queries can be either accurate but slow or best-effort but fast.

NoSQL, Scalable SQL, and NewSQL

To address the big data challenge, NoSQL proponents limit ACID constraints, provide fully scalable solutions with preliminary database features, and then slowly add back the relational database management system (RDBMS) features such as index and transaction support. We can observe this trend in Google's BigTable to Spanner evolution.

On the other end of the spectrum, the RDBMS community is rethinking its systems' design and is attempting to scale them in a share-nothing environment. These approaches add the ability to autopartition and autoscale data while offering more options for trading off consistency for performance. Moreover, other NewSQL[11] projects seek to modernize the RDBMS architecture to provide the same scalable performance of NoSQL while preserving the ACID guarantees of a traditional, single-node database system.

Connected Data: New Challenges for Clouds and Social Networks

Research has shown that users primarily employ social networking sites to articulate and make visible their existing social networks.[12,13] In other words, users on these sites aren't usually trying to connect with strangers but are primarily communicating with people who are already part of their direct or extended social network. This observation implies that a level of trust already exists between social network users, and that these users share at least one aspect of their lives: career, hobbies, political views, and so on. We envision that these characteristics are vital to enabling interesting opportunities, including establishing security policies that leverage existing trust relationships, promoting data and resource sharing within networks of people with similar interests, and optimizing data analytics by leveraging the fact that people in the same network potentially share the same interests and will thus submit similar queries. Finally, we propose leveraging the wisdom of socially connected individuals to build and maintain service reputation systems. Clouds comprising social network connections open numerous research opportunities.

Resource Sharing

Social networking on the cloud could enable resource sharing based on the social relationship between users. This would potentially build on technologies such as volunteer computing, which is a distributed computing model in which connected users donate computing resources to a project. Storage@home[14] and Boinc[15] are two examples. In these cases, the computing resources are owned by individuals and can be shared in return for access to other resources. This could potentially change the cloud's economics and raises questions related to reliability and quality-of-service (QoS) guarantees. Again, we can leverage the social aspect to build reputation for users and establish their corresponding resource reliability.

Locality of Reference in the Cloud

The cloud's big data aspect constitutes a challenge for both efficient data analysis and mining. From a performance perspective, the cloud's social aspect can be leveraged to compute, cache and share the analytics results within a circle of connected users. These users are potentially interested in the same patterns, so computations would exhibit high locality of reference, which can help to optimize performance.

Privacy-Preserving Data Analytics

On the other hand, privacy-preserving statistical techniques, such as differential privacy, can be employed in conjunction with social links to maximize query result accuracy without revealing private data. Privacy levels and accuracy can be defined differently within a social setting. For example, privacy constraints can be
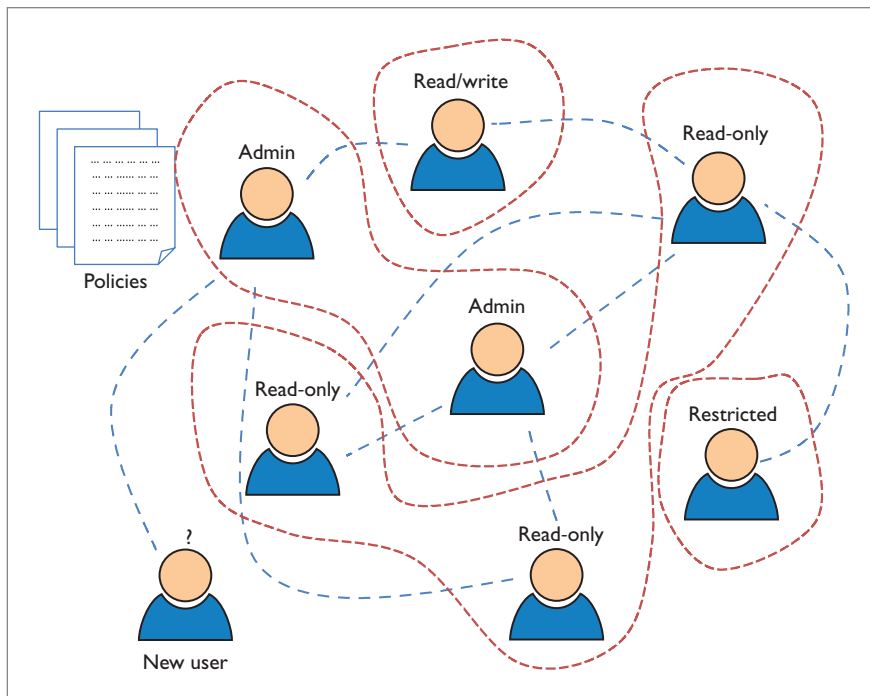
Figure 2. Overlaying the social graph with security groups, roles, and policies. Based on their social links, new users can be automatically classified into groups as they join the network.

relaxed depending on the number of links between sets of users in a social graph. Differential privacy techniques must also be refined to deal with incremental data that has social annotations.

Cross-Domain Data Analytics

Aggregating data from multiple social networks enables data analytics that correlate the datasets' various networks. Given that social networking vocabulary varies from one network to another, we anticipate the need for cross-domain vocabulary mapping as a data preprocessing step. For example, the Twitter glossary defines terms such as "followers" and "tweet." Facebook defines terms such as "friends" and "status." Google Plus uses "circles" and "hangout." To perform cross-domain data analytics, we must develop and maintain a common ontology that will capture the differences and similarities in terminologies and define relationships between terms within and across the network.

Socializing Access Control Policies

Security is a major concern that we must address when coupling social networks with the cloud. User groups, roles, and access control policies must be in place to govern users' access to cloud resources. To facilitate this process, we could leverage social relationships to build an evolving access control system that self-adapts to the addition, deletion, and update in users and their relationships. Some work has proposed semantically annotating these relationships and using semantically described rules to infer relationships between users and resources.[16–18] These relationships can then help to establish trust and form the basis of access control policies. Because cloud resources are largely dynamic, self-adapting policy rules are needed to determine users' access rights as new resources become available and new users connect to the social network. These rules can use just-in-time data classification schemes to infer access rules for new data items as they're digitally born

within the cloud. As Figure 2 shows, the outcome is a social graph overlaid with security groups and policies; based on their social links, new users can be automatically classified into groups as they join the network.

Service Reputation Frameworks

Cloud computing reaches its potential when software is implemented as services that can be mixed and matched over the cloud to address users' requirements. Automatic service discovery and composition can occur based on services' reputation. A service reputation can be built from users' feedback and by auditing a service invocation and execution. The service reputation is hence a function of both the QoS a service delivers, measured over the historical execution log, and the explicit community's feedback.

Some generic frameworks propose incorporating service reputation as a selection criterion when composing services.[19] Incorporating the social dimension can largely enrich these frameworks. Consider a travel reservation website that composes and invokes different services to find the best deals on air tickets. By binding this functionality to a social network, not only can we effectively build a service reputation by incorporating community wisdom, but a consensus for evaluating services will exist among users because they're potentially of the same mindset. For example, some communities would appreciate price over the length of a flight, others a service's response time over result quality. Consequently, the reputation value calculated within social settings is a more accurate measure of satisfaction within a user community.

Classification for Social Networks

The success of Facebook and LinkedIn demonstrates that the Web's power can not only foster but can also capitalize on a social network. Such

networks, both for the general public and specifically for the scientific community, are changing user communication and practices. We classify all social networks using two criteria: level of generality and ability to execute.[20] In the level of generality dimension, we distinguish a social network for general and specific purposes. In the ability to execute dimension, we distinguish informative and executable (that is, able to run computation) social networks. We show this classification in light of scientific networks, but it applies to nonscientific ones as well.

Informative vs. Executable

When considering the overlap of social networking techniques and commodity or cloud computation, a distinct difference exists between the system being informative or being executable.

General-purpose social networking sites have aspects of both:

- *Informative*. General-purpose social networks such as Facebook and LinkedIn have been harnessed to cultivate communication and collaboration.[2] For example, major scientific associations such as the American Association for the Advancement of Science (AAAS) and the IEEE have set up groups on both Facebook and LinkedIn. In these major community groups and many smaller ones, members can share research progress, search for jobs, and seek collaborations.
- *Executable*. Besides these informative social networks, many websites provide open and collaborative platforms to search for executable mashups, Web services, and so on. This category includes ProgrammableWeb (www.programmableweb.com), an online community for Web APIs and mashups, and Amazon Elastic Compute Cloud (EC2; http://aws.amazon.com/ec2).

Figure 3. Social networks for scientists. Each network is positioned based on its relative level of generality and its ability to execute. (Some online services included in this figure, such as Amazon EC2, Globus Online, Galaxy, and caGrid, are arguably social networks by themselves. However, we list them here because they all provide an open collaborative environment that's very close to a social network and can rapidly evolve toward that direction.)

Research-oriented social networks tend to be naturally integrated with informativeness and execution capabilities:

- *Informative*. Various social networking sites exist for general academia, such as CiteULike (www.citeulike.org) and Nature Network (http://network.nature.com). These websites are based on author-publication-citation networks and can be used to identify connections among authors, publications, and research topics. Sites also exist for specific communities, such as life scientists (http://prometeonetwork.com) and doctors (www.doctors.net.uk).
- *Informative-executable*. Many sites go beyond just bringing people together. Rather, they enable researchers to share data and

protocols that describe methodologies for conducting experiments and obtaining data. OpenWetWare (http://openwetware.org) is such an example for biology.
- *Executable*. Some research-specific social networks are computation-oriented — that is, they facilitate the sharing of executable computational components. For example, myExperiment (www.myExperiment.org) offers a curated registry of scientific workflows and a platform on which to execute them; nanoHub[21] provides a nanotechnology research gateway hosting not only user groups and tutorials, but also simulation tools.
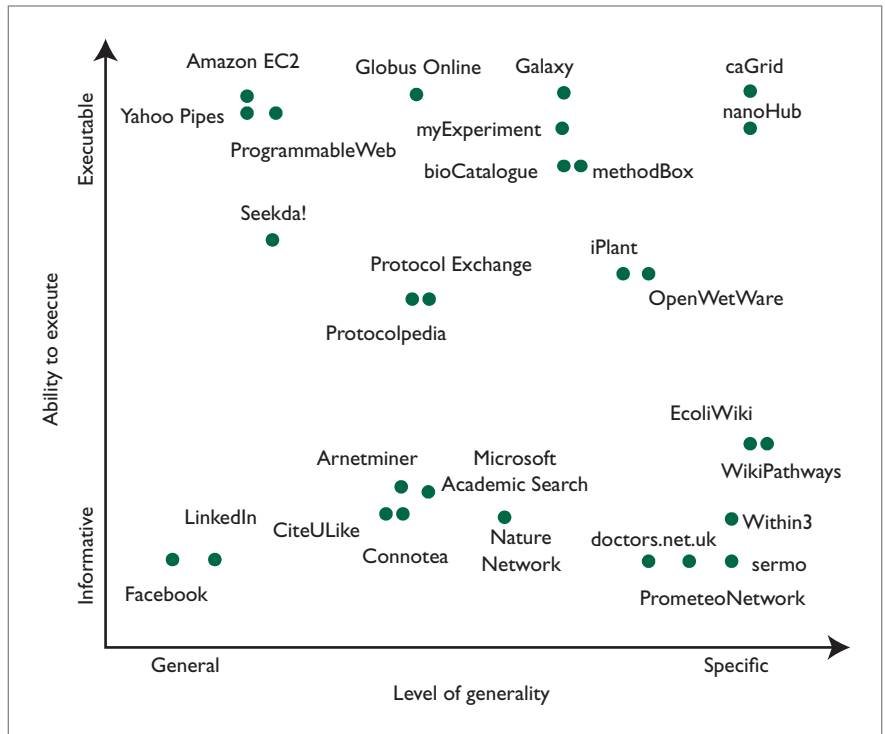
Figure 3 lists social networks for scientists. Each one is positioned based on its relative level of generality (the *x*-axis) and ability to execute (the

Figure 4. A word cloud for recent cloud computing and big data research. Resource management and performance issues are gaining the research community's attention.

y-axis). To understand how big data research is overlapping with cloud computing research, Figure 4 shows a word cloud generated from more than 60 recent research papers on cloud computing and big data in the last two years. Based on the frequency of words, we can see that resource management and performance issues are gaining the community's attention. Technologies such as MapReduce and Hadoop are becoming the leading examples in this field. Research has also started addressing energy issues related to the cloud. Interestingly, social and mobile domains aren't gaining the expected attention despite the popularity of social networking and mobile devices.

With beginnings in social science, mathematics, physics, and now computer science, social interactions among humans have been widely investigated. However, the vast amount of data available in digital form, coupled with larger, well-organized groups of users, facilitate a significant enhancement in collective human intelligence and knowledge derived from collective data. We can summarize this as the overlap of social networks for big data analysis. This area presents a wealth of new research opportunities for engineers and scientists.

Engineers will need to introduce new distributed data analysis frameworks in which users have access to subsets of the "big data" datasets as well as situational awareness into global processing. This framework should enable engineers to share computational resources while leveraging them on desktops, servers, and mobile phones. Big data analysis over clouds can't be done by trial and error, but rather will require just-in-time assessments. Consequently, the operational research community must investigate new simulation techniques for predictive decision support when deciding when or if to initiate a new analysis. Data will no longer reside in standard relational databases, but in more distributed data stores spanning users of a larger network. As such, new comprehensive cross-network, cross-cloud data models must be developed that are designed to optimize performance based on the distribution of information and users. Finally,

conventional security and access control systems, such as the active directory, are based on the tree-structured organization of users. In a socially connected world, however, these policies must leverage interconnected, graph-based social relationships. A need will exist for highly self-configurable security policies to protect users' security and privacy while also preserving privacy embedded within the data. These and other techniques will significantly enhance and extend the information age. ⌷

References

1. "The Data Deluge: Businesses, Governments and Society Are Only Starting to Tap Its Vast Potential," *The Economist*, 25 Feb. 2010; www.economist.com/opinion/displaystory.cfm?story_id=15579717.
2. V. Gewin, "The New Networking Nexus," *Nature*, vol. 451, no. 7181, 2008, pp. 1024–1025.
3. Y. Wei and M.B. Blake, "Service-Oriented Computing and Cloud Computing: Challenges and Opportunities," *IEEE Internet Computing*, vol. 14, no. 6, 2010, pp. 72–76.
4. A. Mislove et al., "Measurement and Analysis of Online Social Networks," *Proc. 7th ACM SIGCOMM Conf. Internet Measurement*, ACM, 2007, pp. 29–42.
5. P.J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*, Cambridge Univ. Press, 2005.
6. S. Boccaletti et al., "Complex Networks: Structure and Dynamics," *Physics Reports*, Feb. 2006, pp. 175–308.
7. I.Z. Paul, C. Eaton, and P. Zikopoulos, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw Hill Professional, 2011.
8. M. Stonebraker et al., "MapReduce and Parallel DBMSs: Friends or Foes?" *Comm. ACM*, vol. 53, no. 1, 2010, pp. 64–71.
9. S. Chaudhuri, U. Dayal, and V. Narasayya, "An Overview of Business Intelligence Technology," *Comm. ACM*, vol. 54, no. 8, Aug. 2011, pp. 88–98.
10. S. Chaudhuri, "What Next? A Half-Dozen Data Management Research Goals for Big Data and the Cloud," *Proc. 31st Symp. Principles of Database Systems*, ACM, 2012, pp. 1–4.

11. M. Stonebraker, "New Opportunities for New SQL," *Comm. ACM*, vol. 55, no. 11, 2012, pp. 10–11.

12. N.B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *J. Computer-Mediated Communication*, vol. 13, no. 1, 2007, pp. 210–230.

13. C. Haythornthwaite, "Social Networks and Internet Connectivity Effects," *Information, Communication & Society*, vol. 8, no. 2, 2005, pp. 125–147.

14. A.L. Beberg and V.S. Pande, "Storage@ home: Petascale Distributed Storage," *Proc. Parallel and Distributed Processing Symp.*, IEEE CS, 2007, pp. 1–6.

15. D.P. Anderson, "Boinc: A System for Public-Resource Computing and Storage," *Proc. 5th IEEE/ACM Int'l Workshop Grid Computing*, IEEE CS, 2004, pp. 4–10.

16. B. Carminati et al., "A Semantic Web Based Framework for Social Network Access Control," *Proc. 14th ACM Symp. Access Control Models and Technologies*, ACM, 2009, pp. 177–186.

17. B. Ali, W. Villegas, and M. Maheswaran, "A Trust Based Approach for Protecting User Data in Social Networks," *Proc. 2007 Conf. Center for Advanced Studies on Collaborative Research*, IBM, 2007, pp. 288–293.

18. B. Carminati, E. Ferrari, and A. Perego, "Enforcing Access Control in Web-Based Social Networks," *ACM Trans. Information Systems Security*, vol. 13, no. 1, 2009, pp. 6:1–6:38.

19. E.M. Maximilien and M.P. Singh, "Conceptual Model of Web Service Reputation," *SIGMOD Record*, vol. 31, no. 4, 2002, pp. 36–41.

20. W. Tan and M.C. Zhou, *Business and Scientific Workflows: A Web Service-Oriented Approach*, Wiley-IEEE Press, 2013.

21. G. Klimeck et al., "nanoHUB.org: Advancing Education and Research in Nanotechnology," *Computing in Science & Eng.*, vol. 10, no. 5, 2008, pp. 17–23.

Wei Tan is a research staff member at IBM T.J. Watson Research Center. His research interests include big data, cloud computing, service-oriented architecture, business and scientific workflows, and Petri nets. Tan has a PhD in automation engineering from Tsinghua University, China. Contact him at wtan@us.ibm.com.

M. Brian Blake is a professor of computer science and concurrent professor of electrical and computer engineering, and human genetics at the University of Miami. His research interests include service-oriented computing, workflow systems, and software engineering. Blake has a PhD in information and software engineering from George Mason University. He's a senior member of IEEE and an ACM Distinguished Scientist. Contact him at m.brian.blake@miami.edu.

Iman Saleh is an assistant scientist at the University of Miami. Her research interests include data modeling, Web services, formal methods, big data, and cryptography. Saleh has a PhD in software engineering from Virginia Tech. She's a member of ACM, IEEE, and the Upsilon Pi Epsilon Honor Society for Computer Science at Virginia Tech. Contact her at iman@miami.edu.

Schahram Dustdar is a full professor of computer science and head of the Distributed Systems Group, Institute of Information Systems, at the Vienna University of Technology. His research interests include service-oriented architectures and computing, cloud and elastic computing, complex and adaptive systems, and context-aware computing. Dustdar is an ACM Distinguished Scientist and IBM Faculty Award recipient. Contact him at dustdar@dsg.tuwien.ac.at.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*