

Improving High Performance Data Analytics Platforms & Smart Systems: Resource Management and Middleware

Research by
Shikharesh Majumdar and team



IMPROVING HIGH PERFORMANCE DATA ANALYTICS PLATFORMS & SMART SYSTEMS: RESOURCE MANAGEMENT AND MIDDLEWARE

Our rapidly increasing production of data is straining computer infrastructures to unprecedented levels. As researchers develop much-needed coping mechanisms, **Dr Shikharesh Majumdar** at Carleton University focuses on two important solutions to the issue: creating techniques for efficiently handling the analysis of large amounts of data and developing middleware platforms that unify geographically scattered computer and storage resources.

According to IBM, the amount of data we generate every day now exceeds a staggering [2.5 quintillion bytes](#). As this figure increases unrelentingly, the current infrastructures in place to deal with this information are becoming increasingly strained. The world of technology is now faced with key challenges in making sense of the sheer amount of data and extracting useful information from it. As Dr Shikharesh Majumdar of Carleton University explains, these challenges stem from what he refers to as the well-known ‘3V characteristics’ of Big Data.

The first ‘V’ refers to the ‘volume’ of data, which describes the extremely large size of the information being produced. The second ‘V’ stands for ‘velocity’, which is important when describing flows of information, in contexts ranging from Twitter feeds, to sensor data streams in sensor-based smart systems that include buildings, bridges and machinery. Finally, ‘variety’ refers to the different types of data that are produced, in forms including text, images, and numbers.

In his research, Dr Majumdar aims to address these challenges in two critically important areas of data handling. Both areas have involved developing new types of ‘middleware’ – a type of software that bridges the

gap between an operating system and its applications – for effective resource management. One area involves creating resource management techniques that allow large-scale data processing platforms to operate at a high performance. The other focuses on middleware that unifies various resources scattered over multiple locations, for managing sensor-based smart systems.

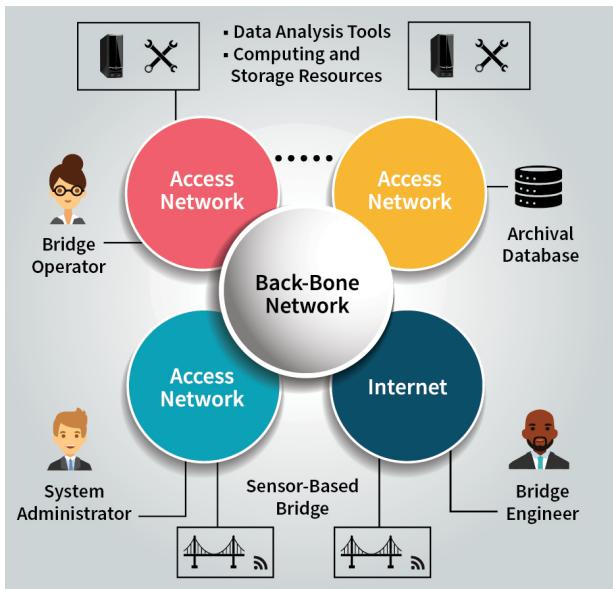
Over the last few years, Dr Majumdar and his colleagues have addressed a variety of the issues involved in each area. Research in the first has led to a novel resource management algorithm, which resulted in an article in *IEEE Transactions on Parallel and Distributed Systems* [2], one of the premier journals in the field, and another paper that received the best paper award at a reputed IEEE International Conference. Dr Majumdar and his team’s pioneering R&D work in the second area has produced a platform for research collaboration among researchers of smart facility management, described by the Canadian Network for the Advancement of Research, Industry and Education (CANARIE) as: ‘first of its kind, it allows geographically dispersed researchers to share data-analysis tools, sensor data, and expertise to manage smart facilities’.

Managing Resources on Big Data Processing Platforms

Today, many data storage and processing resources are provided by cloud and cluster-based platforms, which link many systems together to share information. These systems are fundamentally important to many different users, from businesses and financial institutions to researchers and engineers, who will often generate large amounts of data.

Among the achievements of Dr Majumdar and his colleagues in this field is in resource management for batch data analytics platforms, which allows for techniques that allocate computer resources to specific tasks and schedule the orders and times at which the tasks are carried out – to the convenience of the user of the system. The team has also devised resource management techniques for streaming data analytics, which can take in flows of information from sources including Twitter feeds and financial transactions. This was achieved through a variety of studies that Dr Majumdar has carried out in recent years.

In a 2017 study, Dr Majumdar and his colleagues developed techniques to effectively allocate and schedule tasks on cloud and cluster platforms by efficiently harnessing the resources of the systems. To gauge the effectiveness



The diagram (based on [3]) shows the middleware-based unification of geographically dispersed resources, which can be shared by various users through a two-level communication system comprising local access networks that are connected to the main back-bone network.

of their techniques, the team paid attention to maximising the quality of service of the systems, based on the requirements of users. To do this, Dr Majumdar's team created an algorithm that effectively performed resource management on systems that processed open streams of batch data analytics jobs, in which the large data file was split into smaller chunks each of which was processed by a set of concurrent tasks in the job. For each task, the algorithm set an earliest acceptable start time, a required execution time, and specified which particular tasks should be started after another has finished. These times were ultimately derived based on what the user had specified as a deadline for completion through a service level agreement. After analysing the performance of the algorithm, Dr Majumdar's team found that compared to a leading previously developed technique, 63% fewer overall tasks missed their deadlines when using their algorithm.

In the same year, Dr Majumdar led a further study that explored how energy consumption could be reduced in cloud and cluster-based systems. Unsurprisingly, processing huge amounts of data requires systems to consume huge amounts of energy, accounting for a large fraction of their maintenance costs, and making significant contributions to greenhouse gas emissions. To address this pressing issue, Dr Majumdar's team developed their algorithm further to take the energy consumption required for each smaller task into account. This time, the algorithm could match tasks together to be performed at the same time, and adjust the CPU operating frequency in such a way that energy consumption was reduced while the quality of service requirements of users could still be met. After testing the updated algorithm in a range of scenarios, the team achieved a reduction in energy consumption of up to 45% for a simulated cloud system and workload – potentially a hugely significant step towards energy-efficient data centres.

Again in 2017, Dr Majumdar and his colleagues started analysing how streaming data analytics that concerns extracting information from continual streams of data could be performed more efficiently. Particularly for social media companies such as Twitter, information needs to be processed and scheduled in particular orders, based on whether one piece of information has priority over another. Multiple levels of priority are often needed as well in the context of smart systems generating sensor data streams that need to be processed in real time. To address the issue, the team developed two scheduling techniques. These could assign higher priority to information to be scheduled depending on whether the system is 'static' (where the priorities are unlikely to change), or 'dynamic' (where priorities can continually change). After prototype implementation, the team demonstrated how their proposed scheduling techniques can be highly effective.

Platform for Managing Smart Systems

As well as working with processing platforms, Dr Majumdar and his team are dedicated to analysing unified systems of many computer systems, scattered over many geographical locations. This is often required in the management of sensor-based smart systems. Such systems leverage the 'Internet of Things' (IoT) technology for communication among a large number of independent objects, which are often found in sensor-based bridges, buildings, machinery and patient monitoring systems.

One problem currently facing IoT technology-based smart systems is the difficulty of basing the smart system component being monitored in one location while having the tools and computing systems required for the analysis of its data in other places. Through several recent studies, Dr Majumdar and his colleagues have created middleware to act as a 'glue' that connects various system components, allowing a variety of data sources and tools to become available on demand. The advances will allow authenticated users of smart systems to analyse data and manage devices from anywhere in the world.

In 2015/2016, a team led by Dr Majumdar described how a cloud-based system could be used to effectively manage large IoT-based smart systems. Despite these systems having widely varied management needs, all smart systems share the same basic operations: monitoring the state and health of their infrastructure and analysing and making decisions on their future states and maintenance. Smart systems require various different resources to carry out these tasks, including computers for data analysis, storage for sensor data and maintenance history, and software for analysing sensor data. Dr Majumdar's team realised that clouds can help to manage such complex smart systems by unifying dispersed resources required for managing a smart facility. In two case studies, they showed that cloud-based systems can greatly improve operation in smart systems of geographically-scattered resources.



The figure (based on [1]) presents the architecture for a remote patient monitoring system. Wearable Health Sensors (WHS) send sensor readings to the mobile device, which uses the embedded Complex Event Processing (CEP) Engine to detect complex events and send them to the remote IoT Hospital Server (IHS) for further processing at the backend and notification.

Dr Majumdar and his colleagues also explored how cloud-based platforms could be used in research collaboration. In a 2015 study, the team proposed a system where privately-owned software and hardware resources are linked together into a unified platform, available for use by different groups of researchers working in the same field. Named the ‘Research Platform for Smart Facilities Management’ (RP-SMARF), the system would allow researchers to carry out tasks using resources and datasets that would have previously been unavailable to them.

Perhaps the main appeal of RP-SMARF is the ability of researchers to access data generated by experiments at any location around the world. Through a sophisticated authorisation framework, resource owners would be able to precisely control the availability of particular elements to other users of the system, allowing for a secure, trustworthy way of sharing information about experiments. Dr Majumdar’s team believes that RP-SMARF will unify researchers around the world, significantly increasing their productivity. In addition to facilitating research collaboration, RP-SMARF-like systems could be used to monitor and manage smart facilities and extend

the lifespan of public infrastructure including buildings, machinery and renewable energy sources such as wind turbines and hydroelectric dams. The system would allow this by helping engineers to collaborate and share streams of data from the smart system with one another over many locations. With this greater ease of communication between engineers, the inspection, maintenance and repair processes of public infrastructures could be made far more efficient.

In their latest studies, Dr Majumdar and his team have introduced an architectural framework for performing complex event processing for smart systems. A complex event is a combination of multiple raw events, each of which may correspond to the respective sensor data crossing a pre-determined threshold value. The team proposed a smartphone-based, remote patient monitoring system that uses data from sensors attached to patients’ bodies to detect complex events that may indicate impending health problems. Currently, mobile devices that forward sensor data streams to hospital servers need to remain connected to the server and increase the consumption of the overall network when large amounts of data are transferred. However, the

smartphone-based system can process complex events on the device itself and is therefore able to generate local alarms for the patient in the event that the mobile network becomes temporarily unavailable, disconnecting the user from the hospital server. The researchers demonstrated the viability of their approach through a proof-of-concept prototype built from a Google Pixel smartphone, and open source software. Analysis of the device’s performance provided new insights into how patient monitoring systems can be scaled, and into the relationship between the complexity of the system and its performance.

The research performed by Dr Majumdar and his colleagues was carried out in the laboratories of the *Real Time and Distributed Systems* (RADS) Research Centre in Carleton University. With internationally recognised researchers and talented graduate students, the centre is a seat of world class research in real-time and distributed computing systems. Dr Majumdar and his team are currently engaged in further research in each of the two themes: resource management on big data processing platforms and platforms for managing smart systems.



Meet the researcher

Dr Shikharesh Majumdar

Department of Systems & Computer Engineering
Carleton University
Ottawa, Ontario
Canada

Dr Shikharesh Majumdar is internationally known for his research in resource management and middleware for high performance parallel and distributed systems. He is a Full Professor and the Director of the Real Time and Distributed Systems Research Centre at the Department of Systems and Computer Engineering in Carleton University, Ottawa. He was awarded his PhD in Computational Science from the University of Saskatchewan in 1988. Dr Majumdar's research interests are in the areas of cloud and grid computing, smart systems, operating systems and performance engineering. He has received a number of awards for his research and services to the professional community, including the IEEE's Best Paper Award in 2017, the Glory of India Award and Recognition of Service awards from ACM and IEEE. He is a member of ACM, is a senior member of IEEE and has provided lectures in various countries as a Distinguished Visitor for The IEEE Computer Society (1998–2001).

CONTACT

E: majumdar@sce.carleton.ca
W: <http://www.sce.carleton.ca/faculty/majumdar.html>

KEY COLLABORATORS

Professor David Lau, Professor Jie Liu, Professor Marc St-Hilaire, Norman Lim and Amarjit Dhillon, Carleton University
Dr Blswajit Nandy, Solana Networks
Peter Ashwood-Smith, Huawei
Ali El-Haraki, Telus
Dr Nishith Goel, Cistel Tech

FUNDING

Natural Sciences and Engineering Research Council (NSERC) of Canada
Ontario Centres of Excellence (OCE)
CANARIE
Huawei Canada, Telus and Cistel Tech

ACADEMIC SPONSORS

Department of Systems and Computer Engineering, Carleton University – a research-intensive department home to a dynamic and innovative team of active faculty members, instructors, and undergraduate and graduate students.

Faculty of Engineering and Design, Carlton University – one of the nation's leading institutions for teaching and research of engineering, architecture, industrial design, and information technology.

REFERENCES

1. A Dhillon, S Majumdar, M St.-Hilaire, A El-Haraki, A Mobile Complex Event Processing System for Remote Patient Monitoring, Proc. IEEE International Congress on Internet of Things (ICIOT), San Francisco, July 2018.
2. N Lim, S Majumdar, P Ashwood-Smith, MRCP-RM: A Technique for Resource Allocation and Scheduling of MapReduce Jobs with Deadlines, IEEE Transactions on Parallel and Distributed Systems, 2017, 28, 1375–1389.
3. S Majumdar, Cloud-Based Smart Facilities Management, Book Chapter in Internet of Things: Principles and Paradigms (Eds: Buyya, Desjardins), Elsevier, 2016, 319–339.