**A*DAX: Data Analytics Platform for Smart Cities**

See-Kiong Ng  
SERC Programme Director, Urban Systems Initiative  
Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

Wee Siong Ng  
Head, Data Management Laboratory  
Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

and Paul Min Chim Lim  
System Architect, A*DAX  
Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

**KEYWORDS**: big data, data exchange, urban solutions, data analytics platform

**ABSTRACT**

Cities are becoming big generators of data. In the past, decision makers of cities had to make educated guesses about their cities based on small data sets or even just their gut feel. Today, it has become possible to listen to a city’s heartbeat in the massive amounts of data that it generates—from sensors monitoring smart buildings and the environment, to positioning systems navigating moving vehicles, to netizens tweeting about the city in the social media, all in real-time. The data are no longer just accessible by the policymakers but they are also increasingly made available to the business operators and citizens of the city. This new abundance of urban data can be a game-changer for building better cities, if we can make the data work for the city using intelligent and scalable data analytics. In this paper, we will describe A*DAX, a data analytics platform for smart cities to harness the power of big data in the city into new urban solutions for the city.

**1 INTRODUCTION**

Since the industrial revolution in 19th century, technology has always been a key driver in the development of the society. The recent arrival of the Big Data era brings an unprecedented opportunity for us to listen to the city’s heartbeat closely to understand the city better. Technologies brought about by the Big Data revolution can potentially transform how cities monitor, manage, and enhance the liveability of their communities in unprecedented ways. In fact, many cities aiming to become smart are turning to data analytics and ICT innovations, developing state-of-the-art urban systems for transportation, environment, energy, etc., that use big data analytics and machine learning to extract patterns and rules from historical data to manage the future better.

While large amounts of city data are routinely collected by the public and private sectors, the data are usually kept isolated in proprietary information systems with no one having cross-domain access to these datasets, as well as the right analytics tools to generate the insights needed. In order to effectively turn the big data in the city into actionable insights, it is useful for a smart city to develop a data exchange platform designed specifically for data analytics, with the capability to bring together data of diverse natures to be managed, retrieved, integrated and analyzed. In this way, the massive volumes and variety of data generated by the city can then be fully into useful insights that lead to new urban solutions for the city.
1.1 Urban Systems Initiative

The Agency for Science, Technology and Research (A*STAR) of Singapore launched the Urban Systems Initiative to address the technological needs of the rapidly urbanising world. The Initiative comprises multiple R&D programmes addressing various smart city topics such as real-time urban sensing, city logistics, and integrated urban planning. The objective is to harness the complexity and intelligence in big data to develop solutions for real-life urban challenges. Recognising that a trans-disciplinary approach is needed, where the stages of co-definition of problem, problem solving, solution, testbed and implementation has to be a recursive feedback loop, taken into account of inputs from various stakeholders, we actively partner with public agencies and the industry to develop unique urban solutions that enhance Singapore’s competitiveness as a global model for smart cities.

To achieve the goals of the Initiative, large amounts of data collected by the public and private sectors need to be effectively managed, integrated, and analyzed. The A*STAR Data Analytics and Exchange Platform ("A*DAX") is developed to handling and merging diverse city data that can then be securely shared and efficiently accessed for use—it is the data backbone for the various solutions programs under the Urban Systems Initiative.

2 A*DAX

A*DAX is a scalable and open standards based platform designed to store, integrate and manage data for secure data exchange, analytics and visualization. The platform handles both static and real-time public and commercial data, and incorporates security and privacy features. It allows translation and integration of data into actionable insights so that citizens, businesses and public agencies can make informed decisions.

A*DAX utilises a three-tier architecture as depicted in the system diagram (Figure 1). The first tier - the client operations tier—provides open access APIs for the developers by adopting World Wide Web Consortium (W3C) open standards such as REST, SOAP, UDDI, WSDL, JSON and Open Geospatial Consortium (OGC) standards. The open access APIs will provide programmable functionalities for data querying and data manipulation operations, creating spatial visualisation, and defining data format for developers to create urban solutions. To cater for the common users, a user friendly Applications Marketplace and A*DAX portal are developed to help them easily navigate the platform for accessing, visualizing and exploring data and applications that are relevant for decision making.

The second tier, the heart of A*DAX system, is the A*DAX Fusion Middleware. It comprises of a Query Processor component that serves as a control point between client's query and data management services. It serves two purposes: i) to provide a unified query interface, where selection of query engine is transparent to users; and ii) when necessary, to re-write the query to fulfil access control policies. Unified Security is a module will provide a single access control of multiple related but independent data management services. It will ensure the integrity of access policies across all the data management services. As the data management services in A*DAX support different authentication mechanisms, the unified security module has to internally translate and store different credentials based on what is used for initial authentication. Finally, the goal of API Management module is to enable organizations to develop, deploy, manage and secure APIs.

Last but not least, in the third tier, various data management services deliver data storage and query processing capabilities needed for smart city applications:

- **Data Warehouse**—Traditional data management system in which data are stored in tables (i.e., row based) and the relationships amongst the data are also stored in tables;
• GIS System—Spatial data store that supports for geographic objects such as geometry types for points, line strings, polygons, multi-points, multi-line-strings, multi-polygons and geometry collections;
• Complex Event Processing—A system that processes a constant data (or events) stream, or a concept in which the content of a database is continuously changing over time;
• Triple Store—A purpose-built database for the storage and retrieval of triples, a triple being a data entity composed of subject-predicate-object. Unlike a relational database, a triple store is optimized for the storage and retrieval of triples. In addition to queries, triples can usually be imported/exported using Resource Description Framework (RDF) and other formats; and
• Big Data Cluster that enables the distributed processing of large data sets across clusters of commodity servers.

![Figure 1. A*DAX System Components](image)

1.1 System Architecture

A smart city is continually evolving as new requirements become evident and new technologies become available to be adopted. As such, the architecture of a data analytics platform for smart cities should be designed with scalability and flexibility in mind. A*DAX is built upon scalable infrastructure with a target towards share-nothing cluster with federation capabilities. Middleware components such as Geographic Information System (GIS), data analytics toolkit, visualization libraries and external data ingestion modules are designed to be pluggable and integrated as needed. It is also built upon open standards, so that these components are replaceable with other products with equivalent functions that support the same open standards. In this way, cities that adopt A*DAX will have the choice of using the components that suit their requirements and budget, with the flexibility of continually evolving the platform’s architecture together with the growth of the smart city, without costly vendor lock-ins.
Figure 2 above shows the reference system architecture for A*DAX. As A*DAX increases in complexity and scope, this reference architecture will serve as a tool to provide direction, architecture baseline and architecture blueprint to guide and constrain the development of new architecture versions, solutions and system extensions. The common architecture vision and lexicon will result in streamlining of requirements collection, system development and life-cycle support of A*DAX. Future A*DAX adopters can use the reference architecture as a starting point to implement their own system and specialize it solve specific problems. To achieve interoperability between many different and evolving systems, the Reference Architecture will capture, share and encourage the adherence to common standards, specifications and patterns. For more detailed technical descriptions of the A*DAX architecture, please see (Amudha et al. 2014).

3 Data Analytics in Smart Cities: Urban Mobility Use Cases

As breakthroughs in technologies such as internet-of-things and big data enable us to sense and make sense of the city in real-time, a data analytics platform such as A*DAX is useful for smart cities to turn the big data into actionable insights to benefit the decision makers as well as businesses and people. In this section, we present some use cases in urban mobility to demonstrate the usefulness of big data analytics for smart cities.

3.1 Urban Mobility Analytics

Understanding the urban dynamics of the city is critical for urban planning as well as discovering business opportunities and solutions to meet urban challenges. Traditional methods of conducting surveys to gather data are expensive, time-consuming and provide data with limited reliability. The new abundance of big data in the city about transportation, from various sources such as public transport smart cards, GPS on taxi, and even mobile phones, can provide new opportunities to reveal quantifiable patterns to better inform us about the trends of urban mobility and usage of city facilities on the ground.

3.1.1 Public Transport Smart Card Data

Smart cards are increasingly being used worldwide in public transportation for transit fare payments. In Singapore, the EZ-Link card is used by the public commuters for public transportation services such as the MRT\textsuperscript{1} and the public buses. The trip fare is calculated based on the travelling distance, and

\textsuperscript{1} Mass Rapid Transit (MRT) is the subway system in Singapore.
automatically deducted from a commuter’s EZ-Link card when he/she taps out at a bus stop or an MRT station. For special groups of people, such as students and senior citizens, there are also concession EZ-Link cards which offer discounted fares. The dataset is structured and each record follows a predefined schema as shown in Table 1.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card_Number_E</td>
<td>Card ID for this ride</td>
</tr>
<tr>
<td>Transport_Mode</td>
<td>BUS, LRT, or MRT</td>
</tr>
<tr>
<td>Entry_Date</td>
<td>Date when ride started</td>
</tr>
<tr>
<td>Entry_Time</td>
<td>Time when ride started</td>
</tr>
<tr>
<td>Exit_Date</td>
<td>Date when ride ended</td>
</tr>
<tr>
<td>Exit_Time</td>
<td>Time when ride ended</td>
</tr>
<tr>
<td>Payment_Mode</td>
<td>Method of payment</td>
</tr>
<tr>
<td>Origin_Location_ID</td>
<td>Starting location of the ride</td>
</tr>
<tr>
<td>Destination_Location_ID</td>
<td>Ending location of the ride</td>
</tr>
</tbody>
</table>

Visual analysis of such spatio-temporal data can reveal interesting mobility patterns. For example, Figure 3 depicts the morning commutes of adult residents living in different regions in Singapore. One can visually observe that residents living in the northern region do travel longer distance to get to work, whereas the residents living in the eastern region of the city tends to live near where they work.

![Figure 3. Different Morning Commute Patterns of Residents](image)

The ability to identify the travel profiles of groups such as families and tourists is very useful for the government and the businesses such as childcare centres, restaurants, hotels and advertising companies, to improve their existing services and create new business opportunities. Figure 4 shows the travel patterns of families with kids on weekends that we were able to extract from the smart card data to evaluate whether various regions serve the recreational needs of young families well. This shows that with the use of advance data mining and machine learning algorithms, it is possible to extract meaningful information from the public transport smart card data. Our recent work (Xue et al. 2014) show the design of an advanced data analytics algorithm for extracting travel profiles of specific population subgroup of interest such as tourists.

The insights extracted from public transportation data are useful for urban planners to design the cities better. However, to fully understand how the underlying complex interactions of land use, demographics, job types, and transport networks lead to different commute patterns, it will be necessary to collect, share, integrate, and analyze all the relevant data about the various aspects of the city affecting urban mobility. A*DAX is designed to be such an integrated data analytics platform for understanding smart cities.
3.1.2 Taxicab GPS Data

In addition to public transportation, commercial transportation services such as taxicabs are also an indispensable transportation alternative for the people in the city. In many cities where private cars are no longer viable due to the high-density setup, taxicabs are often used as a substitute for private vehicles by passengers.

Taxicabs today are usually equipped with GPS devices, generating large scale spatio-temporal data that includes their location and status (e.g. “available”, “on-call”, or “occupied”). Figure 5 below shows the GPS trajectories of taxicabs who are currently available (“supply”), and the probabilities of customer pickups at various locations in the city (“demands”) based on historical data. We can apply data mining and machine learning algorithms to analyze the historical data of taxi trips and identify patterns of taxi demand in various regions of the city. We have also recently developed sophisticated machine learning algorithms that can be applied on the GPS data to perform real-time predictions of the trajectories dynamically (Zhou, Tung, Wu & Ng 2013). These patterns and predictive models can then be used to build location-aware taxi demand prediction models that can help a taxicab company dispatch its taxi fleet more efficiently to pick up potential passengers in the city.

In fact, taxicab passengers in the city often find it difficult to find a taxi at the places where they are waiting although there could be free taxis around just a few blocks away. This is compounded by the problem of taxicab drivers not knowing where they can find their next passenger. These happen because passengers and taxicabs are not at the right place at the right time due to the lack of real-time information. With data analytics and data streaming via a platform such as A*DAX, the passengers and drivers can be directed to the right places, thereby optimizing the effectiveness of the taxicab transportation system in the smart city.
3.1.3 Mobile Phone Location Data

Mobile phone locations are routinely collected by telecommunication operators for network management purposes. Location traces extracted from mobile phone usage data (e.g. call detail records) are emerging as an important data source for large-scale urban sensing for mobility research. Compared to travel survey data and even the public transport smart card data mentioned above, mobile phone location data are much more fine-grained and pervasive, and covers practically all possible transportation modes used by the commuters. Recent works on the mobile phone data have showed that it is indeed possible to extract (Manoranjan et al. 2014, Wu et al. 2014) that are important insights for urban planners to plan and design our cities better.

3.2 Integrated Analytics for Urban Mobility

Urban transportation is an excellent example of a complex urban system that is made up of many interacting subsystems. People in the city commute by using public transportation systems (e.g. buses, subways), taxicabs provided by possibly multiple private operators, as well as private cars. In order to piece together a complete picture of urban mobility in a city, it is necessary to integrate cross-domain transportation datasets from multiple public, private, and personal sources.

Understanding the need to obtain and integrate data from multiple data owners, we have initiated a project called Singapore-In-Motion (SIM) to form a data consortium that facilitates collaboration between government agencies and industry partners. The common goal of the SIM data consortium is to study urban mobility through a public-private partnership for data exchange (Yu, Yee, Ng, Lim & Ng 2014). The SIM as an experimental platform to facilitate multi-party, multi-objective data-sharing as a collaborative exercise between public and private data owners through a trusted third-party data analytics platform such as A*DAX to bring together the various kinds of urban data to derive useful insight that can then be shared amongst the data contributors. In this way, new and useful insights which cannot be obtained individually by each of the data owners, can now be extracted to benefit all. The resulting insights can be utilized by government agencies for city planning and by the industry partners to discover new business opportunities.

While A*DAX is designed to provide the platform (Lim, Ng, Ng, Wu & Quek 2014) to deploy the latest data analytics technologies that address the privacy and security concerns of the data owners, the bigger challenge is to develop a viable and sustainable business model for which the various data owners will be motivated to contribute their data. This is especially challenging for the topic of urban mobility, for some of the data owners who own useful data for understanding urban mobility are not particularly interested in transportation (e.g. telecommunication companies). As such, we have to look beyond public-private partnerships and seek ways to engage the people who are the original data generators. It will be important to develop viable participatory sensing models for smart cities so that data can also be contributed directly by individuals into A*DAX for data analytics.

4 SMART CITY TESTBEDDING

It is useful to think of the development of a smart city as an R&D (research and development) undertaking. As such, it is important to provide testbedding opportunities for the technologies and applications to find out what works and what don’t on the ground, and to identify unexpected challenges when technologies and new concepts meet the city and its people. The Jurong Lake District (JLD) Testbed Project is one of Singapore’s Whole of Government (WOG) approaches to pilot the concept of a mixed-use urban precinct that is sustainable, smart and connected. The vision is to develop JLD to be a leading model for Singapore where pervasive ICT connectivity, real time large
data collection infrastructure and data management platform can be provided as “utilities” for smart cities, similar to power and water utilities.

In the JLD Testbed, A*DAX serves as the experimental innovation platform for smart city applications. A*DAX will be deployed as the core data management platform for the city that ingests and shares data collected from a shared sensor and camera network. Figure 6 below shows the interactive map of sensors and cameras deployed in JLD. Participating developers from public agencies, industries and research institutions will leverage on the RESTful A*DAX API to develop their data-driven applications for pilot projects and trials in JLD. The adoption of innovations from such smart city models and testbeds will translate into improved liveability for residents, and more efficient operations and informed planning for agencies and organisations.

![Figure 6. A*DAX's Interactive Map for Sensor Nodes for the JLD Testbed](image)

4 CONCLUDING REMARKS

Cities are becoming big generators of data. In the past, decision makers of cities had to make educated guesses about their cities based on small data sets or even just their gut feel. Today, it has become possible to listen closely to a city’s heartbeat in the massive amounts of data that it generates. Data can indeed become the new electricity to power the intelligence and innovations for a smart city.

However, there are still the following key hurdles to be overcome:

a) Data Silos
Although large amounts of social, economic, geographic, business, and urban data are routinely collected by the public and private sector, the data are usually kept isolated in various proprietary systems and formats. However, oftentimes, cross-domain access to multiple datasets is needed to perform holistic data analytics. In addition, merely putting datasets together does not automatically yield new correlations. This often requires developing advanced algorithms to mine the data to discover hidden linkages among disparate datasets.

b) Security & Business Secrecy Concerns
Security and business secrecy concerns are the main show-stopper for data owners to share data with each other. Without a proper framework to address privacy concerns, liability and regulatory issues, data owners are not comfortable with sharing their data. New privacy-preservation
techniques such as (Wu et al. 2013) and (Wu, Xiang, Ng and Wu 2014) needs to be developed through R&D to unlock the potential of big data for the smart cities.

c) Accessibility of Data Analytics and Visualization Tools
Large datasets and complex data analysis require powerful computational resources for data integration, analytics and visualization. Currently such powerful tools and the specialized expertise needed to use the tools are not easily available to the decision makers who need them. In order to effectively harvest actionable insights from the vast and diverse data generated by the city, it is important to develop accessible data analytics and visualization tools in order for data to become as easy to utilize as electricity.

A*DAX is a data analytics and exchange platform designed for smart cities to harness the power of big data in the city into new urban solutions for the city. It is a work-in-progress, as the smart city is always evolving, and there are many technical and non-technical challenges that we will have to overcome. As such, the architecture of A*DAX is designed to serve as an experimental platform to testbed emerging technologies and new business models for smart city applications. With data and technology being at the core of smart cities, it is our hope that A*DAX can be our vehicle to take us into the future of smart cities, where public agencies, businesses and people can make informed decisions to respond to dynamic conditions in the city intelligently through real-time sensing and data analytics.

REFERENCES


