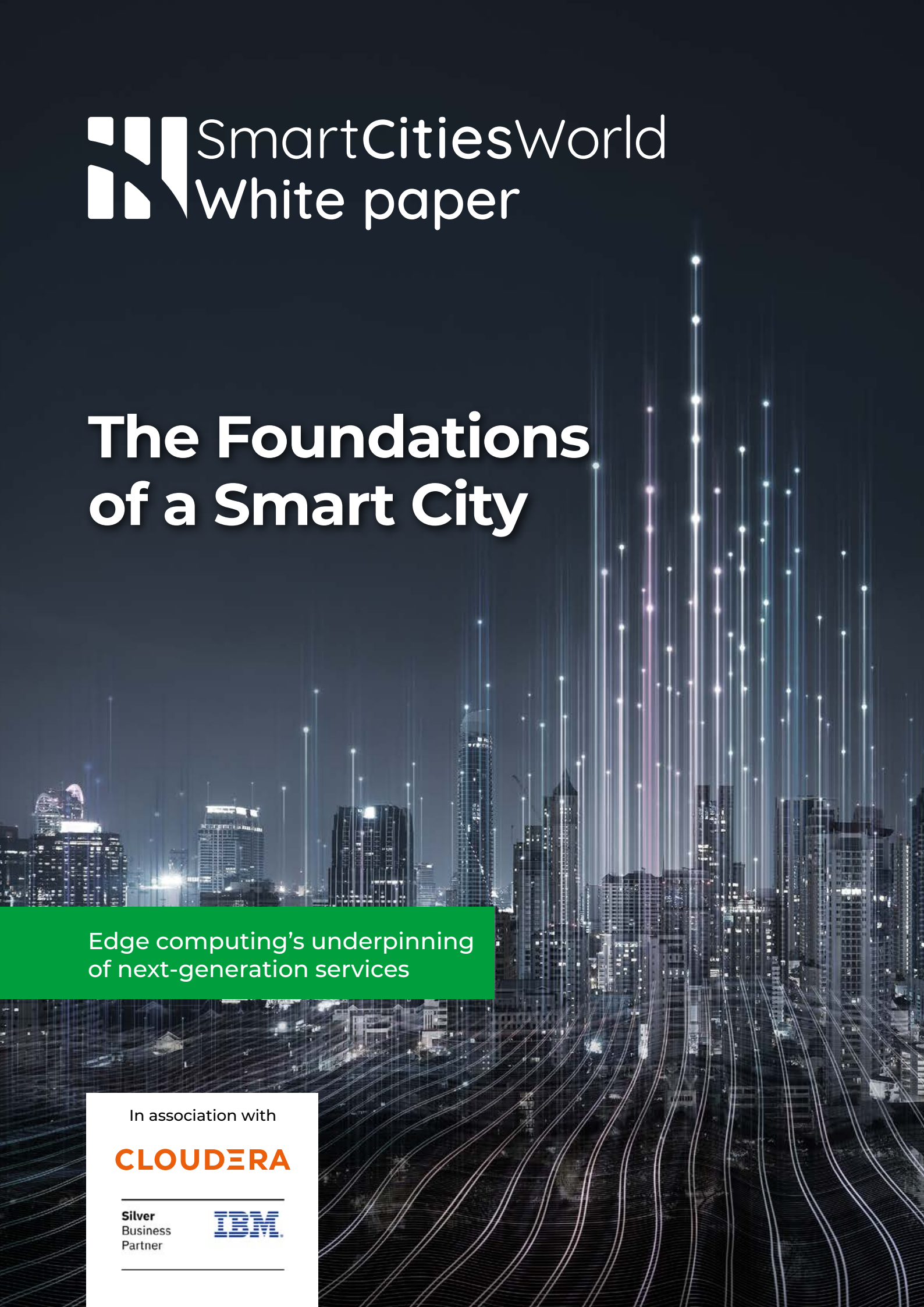
SmartCitiesWorld White paper

The Foundations of a Smart City



Edge computing's underpinning of next-generation services

In association with

CLOUDERA

Silver Business Partner

IBM

Written by

Jonathan Wright

SmartCitiesWorld White paper Reports examine an emerging or growing trend in smart cities, highlighting progress so far and future potential, as well as spotlighting case studies from cities around the world.

In this report, we examine the application of big data and machine learning in real world contexts.

www.smartcitiesworld.net

Introduction

Data drives our lives, how we work, how we are connected to one another and increasingly how we live. A successful smart city needs data in order to provide the services that citizens increasingly expect. Having spent years using consumer-centric digital services, they are expecting the same from their cities.

In simple terms, what cities have to contend with is lots of data from a multitude of different sensors and spanning a range of different services, from traffic to power to transport to waste management, to name a few. And they need to do so in a secure, efficient way that meets the needs of citizens as well as truly unlocks the potential of what a smart city can do.

Because as we all know, everything is connected. A burst water pipe can affect water supplies and increase congestion as traffic is slowed down. But siloed data flows mean that cities cannot see how and where one problem affects another, if they are able to spot the issues at all.

This is where big data and machine learning comes in. By running on open-source technology to avoid vendor lock-in, collecting and processing data in real time, scaling to meet this rise in data, supporting multiple data types and using AI and machine learning to predict problems before they occur, the potential of smart cities can be fully unlocked.

This white paper explores how big data and machine learning can be applied in real world contexts and how Cloudera and IBM's collaboration on technology can underpin a successful smart city strategy.

How big data and machine learning helps you capitalise on the data deluge

Even for those within the technology sector, it can be hard to comprehend the sheer volume of data we generate. According to research company IDC, the data created, captured and replicated will grow from 33 trillion gigabytes in 2018 to 175 trillion GB in 2025.

Most of it unsurprisingly comes from cities and it is only set to increase. The United Nations estimates that around two thirds of us will live in urban areas by 2050, up from one third in 1950, a year when the global population was 2.5 billion, compared to 7.9 billion today.

Data is already essential to how cities are run and can span all its constituent parts, so gathering and understanding these vast flows of information is vital. However, there's a danger of getting lost in pure numbers. What's at least as important is understanding that this data isn't necessarily organised in a way that's useful in specific scenarios or optimised for different users.

This is in part because as the number of sensors grows, they also gather information for different reasons. Consider traffic flows into London, where 22.6 billion motor vehicle miles were racked up along the city's 9,200 miles of roads in 2019. The equivalent figure for 2009, when the global financial crash led to a decline in traffic, was 18.7 billion miles.

The data gathered across 3,562 count points is rich, fascinating, and essential for urban planners and those monitoring congestion over the longer term. But what it won't do is be of much help if your job is to manage traffic flows during rush hour or react in real-time to a road traffic accident causing congestion. This is because the data is not detailed enough – video footage or emergency services estimates of how long it will take to deal with the incident will be more useful in the latter example.

> *"Data is already essential to how cities are run and can span all its constituent parts, so gathering and understanding these vast flows of information is vital."*

Then there's the issue of structured versus unstructured data. An example of the former is the traffic flow information, all neatly collated into an Excel file for sharing. Whereas unstructured data is material from the wild, gathered by Internet of Things (IoT) sensors from parking meters to traffic lights, even the interactions of apps on smartphones and more.

One is not more useful than the other; the challenge is more how to combine both data sets to find new solutions, new efficiencies and new ways to make life run more smoothly for the city's citizens – and to do this in real-time.

Supporting secure data analysis at scale

For the past four years Cloudera and IBM Power Systems have been working together to solve these problems. Their strategic partnership has meant cities can tap into technology that is open, flexible and scalable. It gives cities access to the true potential of an advanced big data and machine learning platform. Cities can tap into governance, security rich environments, advanced query and data management, all of which are outside of traditional silos. And Cloudera and IBM recognise the value of data: a 2019 Gartner survey of cross-industry CIOs revealed 92 per cent of respondents felt data was critical to improving customer experience. Given the consumer-centric demands of citizens, both companies are focused on helping cities deliver the services that matter.

The open-source approach at the edge does not tie a city into using a specific vendor's solutions and applications. Instead, it puts city managers in charge of their data and better able to put this data to work. It's a platform that supports cloud-based and hybrid approaches for more flexibility. And edge computing brings computation and storage nearer to where the data originally generates, giving high quality data control, quicker analysis and actions, continuous operations and lower total costs of ownership.

How does this work in practice? Essentially, it involves running an IT infrastructure that can:

- Collect, process and store data in real-time;
- Scale and cope with ebb and flow of data – and its inexorable growth as smart cities create more of it in the years ahead;
- Support multiple data types;
- Process this data so that it offers real-time operational insights;
- Enable machine learning and AI to improve outcomes based on the data by, for example, spotting and predicting anomalies or problems.

By using these technologies, data generated by an IoT sensor can tap into AI systems on the edge. It can be enriched and combined with other data, both structured and unstructured. To return to that junction where the traffic has become snarled up, it is this kind of IT infrastructure that will enable those managing rush hour to spot the problem early and perhaps to reroute other traffic in real-time – or at the very least give drivers an accurate idea of how long they are likely to be delayed, perhaps via an in-car or smartphone app.

Breaking a city out of its traditional silos is arguably the most compelling reason for using big data. Traditional legacy systems mean that data stays on premises and is locked with closed software. Different legacy systems result in different means of data storage, analytics, security and policies. Trying to marry proprietary systems together and capitalise on the amassed data is beyond most cities, especially if there is a lack of skilled staff.

The Cloudera Data Platform (CDP) removes these problems by seeing where the data lives, how it is used, by whom and for what, bringing new kinds of insights and value across a city's ecosystem. Already cities are taking advantage of Cloudera and IBM's technology and using it for the benefit of their citizens.

"Breaking a city out of its traditional silos is arguably the most compelling reason for using big data"

The reality of smart cities: Smart Dubai

Dubai's ambitions as a city are well known and that ambition has been matched by its smart city policies. The Smart Dubai government office has been charged with overseeing a smart city transformation centred on creating a seamless, safe and personalised experience for both its 3.4 million citizens and 16.7 million overnight visitors per year.

Smart Dubai's goals were ambitious from the off: to deliver 100 smart initiatives and 1,000 smart services over the course of three years. One initiative, for example, has involved building a live dashboard to monitor water and electricity consumption. This has drawn on consumption data from the Dubai Electricity and Water Authority (DEWA), GIS data and district cooling operators' datasets. This enables the analysis of water consumption at a community and individual building level, a way to improve planning in an area where water is scarce.

It has gone one step further by using DEWA's open data records to build a real-time dashboard that track flows of people moving across the city. This precious data can be used by diverse agencies such as city planners, government, real estate developers and start-ups.

Overall, more than 550 datasets have already been ingested into Smart Dubai's data programme and 115 services are being offered to citizens as a result. The plan is eventually for there to be 2,000 datasets all included.

What can other cities take from Dubai's example? Firstly, it underscores the need to take in data from the edge and then act on it in real time. As Smart Dubai's data platform, Dubai Pulse, grew in size as more data was taken in, the need for a scalable and flexible platform became clear, a key factor in Smart Dubai partnering with Cloudera.

Second is that data kept in silos cannot be used effectively. A wealth of diverse organisations, both public and private sector, would never have been able to gain essential insights if it wasn't for DEWA's open data records being made accessible via a real-time dashboard. Closed silos can mean real-time traffic data not being read by those improving vehicle and people flows or traffic pollution is not being shared with healthcare professionals.

One could argue that Dubai has a competitive edge; being one of the fastest growing cities in the Arabian Peninsula means that it is relatively easy to build in smart city technology from the off in new developments. But cities around the world are increasingly focused on retrofitting long-established legacy systems in addition to building new ones. By 2025, it is estimated that smart cities will create business opportunities worth $2.46 trillion as they look to build collaborative and data driven infrastructure.

> *"Dubai underscores the need to take in data from the edge and then act on it in real time."*

This is why we are seeing a suite of diverse and ambitious smart city projects in some of Europe's most historical cities. Cologne, for example, has a 'climate street' intended to act as a living lab for energy-saving techniques and green transport initiatives. Berlin Smart City is using technology to make Germany's capital more sustainable. Vienna is pioneering district cooling, using some of the same technology involved in district heating networks, which uses heat from industry and waste incineration. In Zurich, city employees can access a 3D map that helps reduce the need for on-site inspections of public spaces or for dedicated spatial knowledge to collect the accurate coordinates of objects. The city can now plan and maintain public spaces, or grant permits remotely.

Common goals

This is just a flavour of some real-life examples across Austria, Germany and Switzerland but as these initiatives get rolled out, themes and common ideas have begun to recur. In order to imagine how even the most venerable and longest-established world cities will become smart within the next five years, it's instructive to focus on a few of the areas where technologies and techniques are being most widely adopted.

Traffic management: we have seen how most city administrations have huge amounts of sensor data to indicate traffic levels and traffic flows. Less happily, many cities will rapidly become gridlocked if heavily used parts of the system become congested. Tackling the issues requires acting on real-time information so that, for example, traffic lights can be used to help regulate the flow of traffic around the city. This kind of approach has been used for many years, but it becomes far more sophisticated within a smart city where AI analysis of edge data means it's not just possible to react quickly to problems, but to predict them.

Parking: where should city authorities provide parking and when? Conversely, where and when should city authorities discourage parking? The use of big data can help, again by tracking traffic flows and measuring which parking areas are most used and when. Artificial intelligence again can come into play by predicting which parking areas are likely to be full and giving drivers real-time updates as to which alternatives to use.

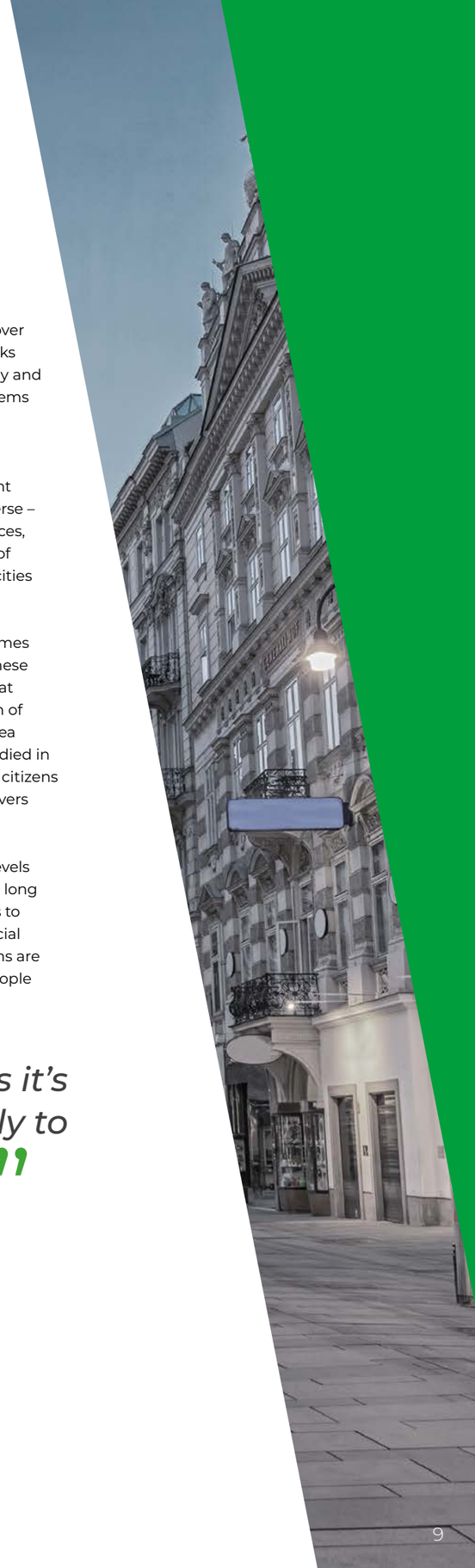
Public Transport: Last year the European Union's spending watchdog found commuters across the continent are continuing to choose cars over the likes of buses and trains, an indication that public transport networks still have considerable room for improvement. For cities to run smoothly and ambitious climate targets to be met, efficient mass transportation systems need to be an integral part of city life.

But how do you ensure these systems, which are expensive to run and maintain, are working optimally? Again, it's necessary to collate different data sets in real-time to keep people moving. These people can be diverse – the number of passengers on different trains and buses, delays in services, connections between different services made and missed, the impact of events such as concerts and football games – but the more efficiently cities can do this, the more smoothly people can get around.

Water levels: many of our cities are built on rivers. London has the Thames running through its centre, Paris the Seine, Berlin the Spree. Many of these cities' citizens, whether they realise it or not, live with the constant threat of flooding. In 2007, the Thames Barrier, designed to stop the floodplain of London being deluged by high tides, protected the city from a North Sea storm surge comparable to that of 1953, when more than 2,500 people died in the Netherlands, Belgium and eastern England. There is also the water citizens don't see, which often runs underground as tributaries of these main rivers have become buried by urban development.

In addition, sewage systems come under heavy pressure when water levels are high because waste water levels rise too, one reason why it takes so long to clean up houses and business premises after flooding. Using sensors to monitor water levels and other factors that may affect them, plus artificial intelligence to analyse the data enables cities to predict where problems are likely to occur and, in the most extreme situations, to arrange to get people to safety.

> *"AI analysis of edge data means it's not just possible to react quickly to problems, but to predict them"*



"By 2025, it is estimated that smart cities will create business opportunities worth $2.46 trillion"

Waste disposal: the amount of waste city life creates is extraordinary and the figures are typically higher in the developed world because there is greater consumption per capita. The European Union's most recent figures found that 5.2 tonnes of waste were generated per inhabitant in 2018. That amounts to the equivalent of 14 kilograms per day.

Disposing of this waste efficiently improves the quality of life of citizens and not just because waste is unsightly. Picking up and then disposing of waste requires man hours, and potentially creates pollution and congestion because it all needs to be carried away. Smart bins, which have sensors in them to monitor how full they are and the weight of the rubbish they contain, have sometimes been portrayed as something of a novelty. In truth, when you can analyse where and when waste is likely to build up, you can also begin to dispose of it more efficiently so that garbage trucks are not unnecessarily trundling through the city.

What underpins the use cases above is that successful networks all depend on data and they all interrelate. Efficient waste management will be achieved if trucks know which roads to avoid at certain times. Sophisticated public transport networks can operate if demand can be accurately tracked and predicted. The only means of doing this in the way that citizens expect is tapping into what big data and artificial intelligence offers.

The city of the future, fuelled by data and run by AI

Smart cities may revolutionise our ways of living and working but they will be a constant evolution. There's no end point to becoming a smart city. Where we are now represents just an early stage in an ongoing process. To understand why, it helps to return to issues around working with data.

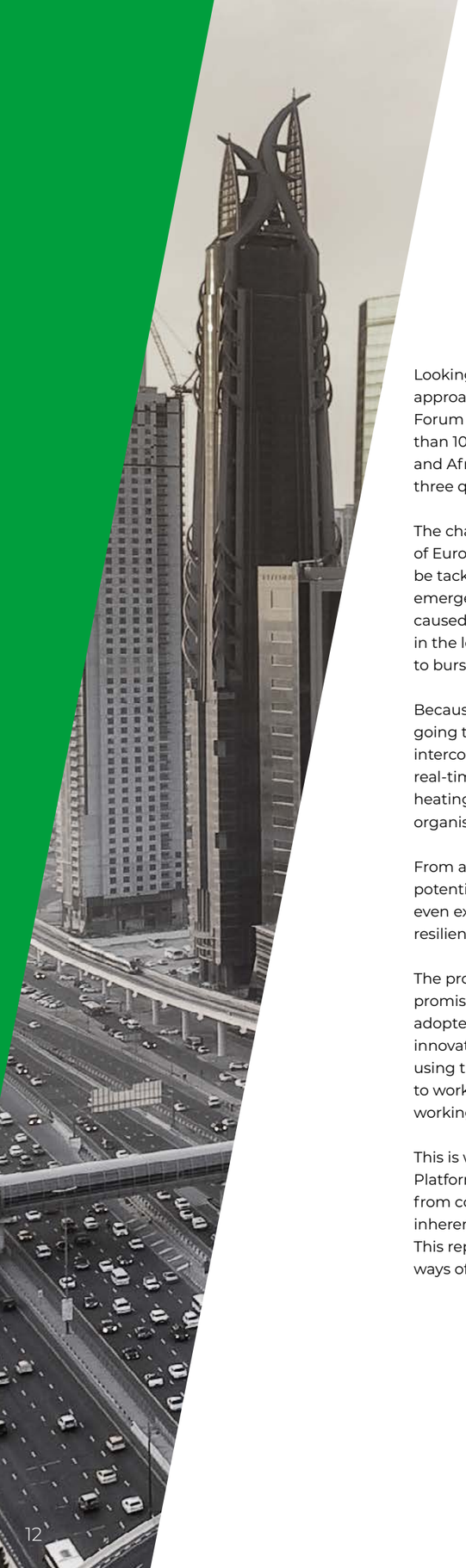
Until now, as we have seen, one of the main problems for city administrators when it comes to working within data has been how to take in unstructured data from the edge and use this information in new and innovative ways in real-time. Given the inefficiencies of legacy hardware and software, it just hasn't been possible. Data sets have solved problems within limited boundaries: traffic data helps with traffic flows and waste sensors help those in charge of the city to clear up more efficiently. But looking ahead to a near-future where the data becomes richer and it's freer, there's no need to stick with an inherently siloed approach.

In the immediate term, edge computing and AI can help cities treat data in a truly open manner, drawing connections between different data sets and dealing with issues before they occur. But that's only step one. Human ingenuity being what it is, people will quickly start using this functionality in wholly unexpected ways. The granular information we have about the city will be scaled up to produce a far more detailed picture of the city. This process will in turn produce new insights and, potentially, new ways of doing things. In 10 years' time, using real-time traffic data to route vehicles around a city might seem old hat.

Faster, higher, stronger

It's not just the data that's driving this process. It's the advent of high-speed internet too. Ten times faster than 4G, 5G is designed to work at average speeds of around 150-200Mbps, with peak speeds of greater than 1Gbps. As access to superfast broadband becomes the norm, people will be more likely to use online services. 5G-enabled devices, especially our smartphones, will also become access points to the smart city. This is already happening in Dubai where citizens can access 115 city services via the DubaiNow application. Services for citizens will be fueled by the data supplied from citizens.

"Smart cities may revolutionise our ways of living and working but they will be a constant evolution"



Looking further into the future, we are going to need these kinds of new approaches as cities get bigger and more complex. The World Economic Forum predicts 10 cities will achieve megacity status, with populations of more than 10 million people, over the next nine years, bringing the total to 43. Asia and Africa will dominate the megacity landscape, accounting for just over three quarters of this new total.

The challenges of dealing with this kind of growth, or even the steadier growth of European cities, are huge. Moreover, these are challenges that need to be tackled alongside environmental problems. In February 2021, a state of emergency was declared in Texas in the United States, after an Arctic blast caused outages in ill-equipped power networks, unsafe traffic networks, chaos in the local water system and further disruption as thawing temperatures led to burst pipes.

Because of global warming, these kinds of extreme weather events are only going to become more commonplace around the world. It also underlines the interconnectedness of cities and why non-siloed data networks operating in real-time are vital. In the case of Texas, the Arctic blast affected road, water and heating networks, among others. A city needs to be treated as an interlinked organism and can only do so in a limited manner today.

From a different angle, the Covid-19 pandemic has shown us that we are potentially vulnerable to a health emergency from a virus few of us knew even existed as 2019 ticked over into 2020. We need our cities to be robust and resilient, able to help us cope with what lies ahead.

The prospect of using data in new ways within smart cities holds out the promise of helping us to deal with these kinds of scenarios. But as early adopters – and most cities have enthusiasts for using digital technologies in innovative ways, people who drive change – find new and creative ways of using the data, one key lesson of the present is that they will need to be able to work flexibly and to experiment rather than be locked into specific ways of working, even specific applications.

This is where a flexible, connected data platform such as the Cloudera Data Platform (CDP) has huge advantages. As does the flexibility that comes from controlling, analysing and experimenting with data where it's located – inherent in Cloudera & IBM's Private, Hybrid-Cloud and Multi-Cloud approach. This represents an approach that's the opposite of getting locked into specific ways of doing things.

"Smart cities are not just better for city administrators; they are also better places to live for their citizens"

Apply this kind of thinking to dealing with the recent pandemic and you begin to get an idea of where we are going. Think of the questions we had to ask for the first time. Even as cities went into lockdown, we knew there would be more people working from home, but how many? Conversely, how many people would still need to get around by public transport? How would so many people working from home impact on the demand for power and water? What kinds of services would health providers be called upon to deliver to patients being struck by a deadly condition we didn't fully understand and whose behaviour was changing in ways it was difficult to predict?

We answered these questions as we went along, but next time smarter cities may be able to answer them more quickly, and better understand how one answer impacts on different areas of life. We will do this by combining data in new ways, again by breaking data out of silos to create. Think of it as having the tools to deal with transport rather than roads, taxis, subways, buses and trains, cycle lanes and pedestrians.

In this way, smart cities are not just better for city administrators, those who want to make their cities run more efficiently, they are also better places to live for their citizens. The smart city is often portrayed as somewhere that's hyperconnected, exciting, where life moves fast, and there's certainly truth in this picture. But another way to think of smart cities is as places where it will be easier to do things because people will be at the centre of how the city works.

Conclusion

As citizens become more aware of the power of technology and what they can do thanks to the digital innovations they have seen over the past 20 plus years, so do their expectations of what they expect from their cities. Their personal lives and their work lives have changed for the better because of digital technology; why shouldn't where they live change accordingly?

Big data and machine learning is a technology that unlocks truly digital centric services. As we are able to use more data sets and to use machine learning in innovative ways, we are finding new ways to make our cities better places to live and work. This is especially true when those in charge of these kinds of initiatives build out new ways of doing things that keep the citizen at the centre of their thinking.

But what truly excites is the scalable nature of a flexible big data platform. As more and more of a city becomes connected and the data becomes richer and more varied, so too do the services. We don't know the services that big data and machine learning will unlock in five or 10 years' time, perhaps even two. This is what technology should do – inspire and excite us and deliver services we previously never thought possible.

About Cloudera & IBM Power Systems
Cloudera and IBM Power Systems bring advanced data and analytics solutions to more organizations to promote a digitally connected future with the enterprise data cloud. Together we offer cutting- edge solutions through open-source innovation for any data, anywhere, from the Edge to AI.