

Big Data in the Public Sector

Turning Bytes into Valuable Information for Society



Target Audience

Business and IT decision makers in the Public Sector are looking for practical use cases of how they can start reaping the promised benefits of Big Data – today. Away from the marketing hype, what is possible now? What are others doing? How do you make it work?

Table of Contents

Big Data in the Public Sector: The Big Picture	3
Use Case #1: Virtualise heterogeneous public data sets to guide decision making	4
Use Case #2: Shape new policies with smarter information about citizens and businesses	6
Use Case #3: Publish and secure Public Sector data sets	8
Use Case #4: Real-time visibility on public service operations, to optimise processes and detect anomalies	10
Use Case #5: Analytics for the business of IT	12
Big Data Use Cases and Technologies for Public Sector (Infographic)	14
Next Steps	16

Authors

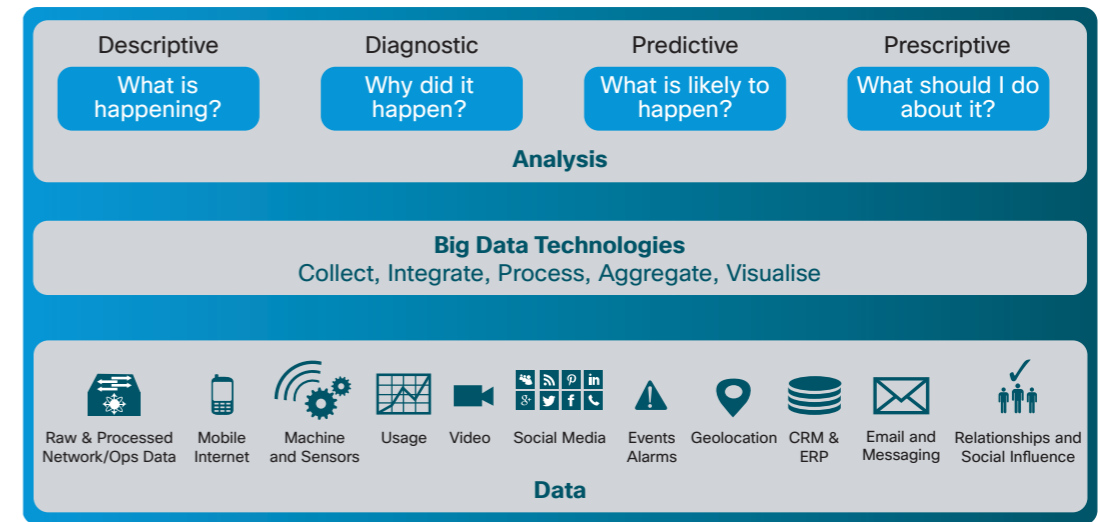
The co-authors of this report include Big Data specialists and experts on government: Alexander Timmerman, Frank de Groot, Greg Page, Holger Mueller, Nick Salt, Nicola Villa, Patrick Bikar, Paul Davies, Roberto Moral and Vernon Thaver. If you have any questions or comments, please contact pbikar@cisco.com.

Big Data in Public Sector: The Big Picture

New economic analysis by Cisco (<http://internetofeverything.cisco.com/>) reveals that the Internet of Everything is a \$4.6 trillion opportunity for global Public Sector organisations (e.g. Government, Healthcare, Education) over the next decade. This is the result of cost savings, increased productivity, new revenues and the enhanced citizen experiences it will bring.

Capturing this value relies on Public Sector organisations:

1. Prioritising the key use cases – gaining insight through descriptive, diagnostic, predictive or prescriptive analysis.
2. Deploying the right technology – to collect, integrate, process, aggregate and visualise information.
3. Connecting the previously unconnected – from sensors and databases to usage patterns and social media.



Many Public Sector organisations have announced their “Big Data strategies” but then failed to deliver the results... why?

Typically, there are three reasons:

1. The focus is put on the technology to deploy, which is often the wrong one, rather than key use cases.
2. The right technologies are deployed for each use case, but there is no common platform architecture. As each solution is in its own silo, it’s not scalable and is costly.
3. Data sources are not standardised or interoperable so it’s not possible to correlate data sets and generate new perspectives.

The four Vs of Big Data are well known: Volume, Velocity, Variety and Veracity. But this paper focuses on the fifth, most important V: Value – the value that Public Sector organisations can create for society.

In the following pages, you will discover:

- Some of the most relevant use cases for the Public Sector
- How to build a future-proof platform that will evolve with your needs
- Why you should consider Cisco for your Big Data projects

Use Case #1:

Virtualise heterogeneous public data sets to guide decision making

While Big Data platforms offer a new way of storing and exploiting large data volumes, the danger is that they become another data silo. In fact, Gartner estimates that 90% of the information assets from Big Data analytics will be siloed by 2017.

The success of evidence-based policy making depends on the quality and variety of the information available. In theory, the more data sources, the better data-driven policy outcomes. But there are hurdles:

- There is too much data to use effectively
- Data is spread too widely and hidden away
- Formats are not easily shared
- Security policies range from public domain to highly sensitive, secret or privacy impacting

So how do you start to action the following?

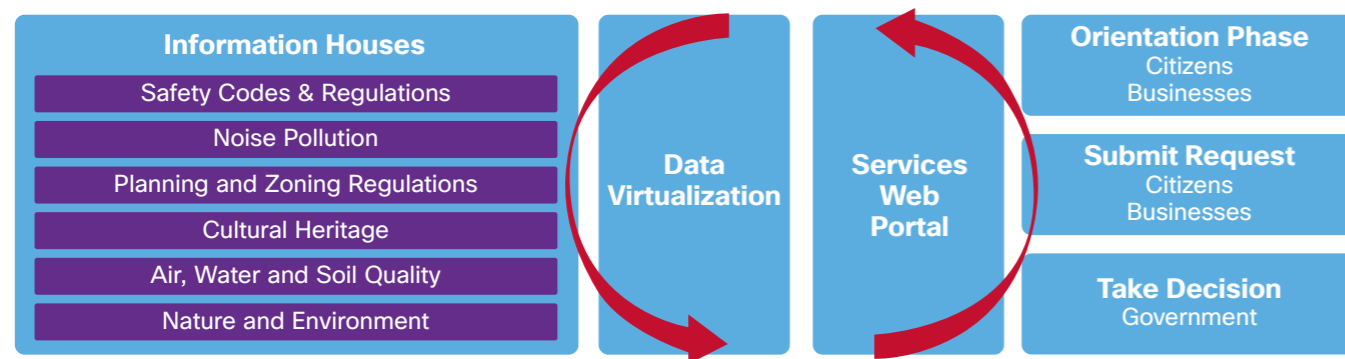
- Access all regional data stores when transitioning to a centralised national police force
- Investigate the pathways to health and wellbeing, education, disability, child abuse and neglect, and juvenile delinquency

- Gain a single patient-centric view of health data, not only within a hospital but across the entire healthcare chain
- Provide roadside sensor information to interested parties developing technology for aided driving/self-driving cars

Data Virtualization (DV) provides a unique solution by allowing access to all data, without replication or consolidation.

For example, the Dutch government has a 10-year programme to digitise the rules and regulations surrounding planning, zoning and environmental permits. The programme's goals include:

- Better and faster decisions
- A 100% digital workflow
- A single view of information for citizens, government and businesses
- Sharing accurate and trustworthy data, owned by the responsible departments



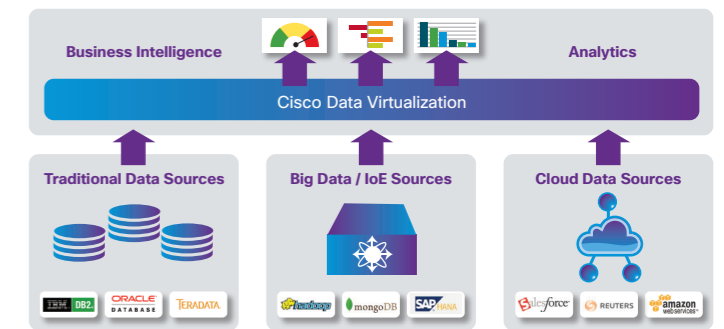
Imagine now that you would like to open a restaurant:

- Data Virtualization would enable a unified self-service portal to find information on the local demographics, laws, regulations and required permits, and manage the application process
- The government uses the same portal and information to decide on the project and manage the legal steps
- The required data is not in a central data warehouse but gathered, managed and owned by a distributed set of 'Information Houses'

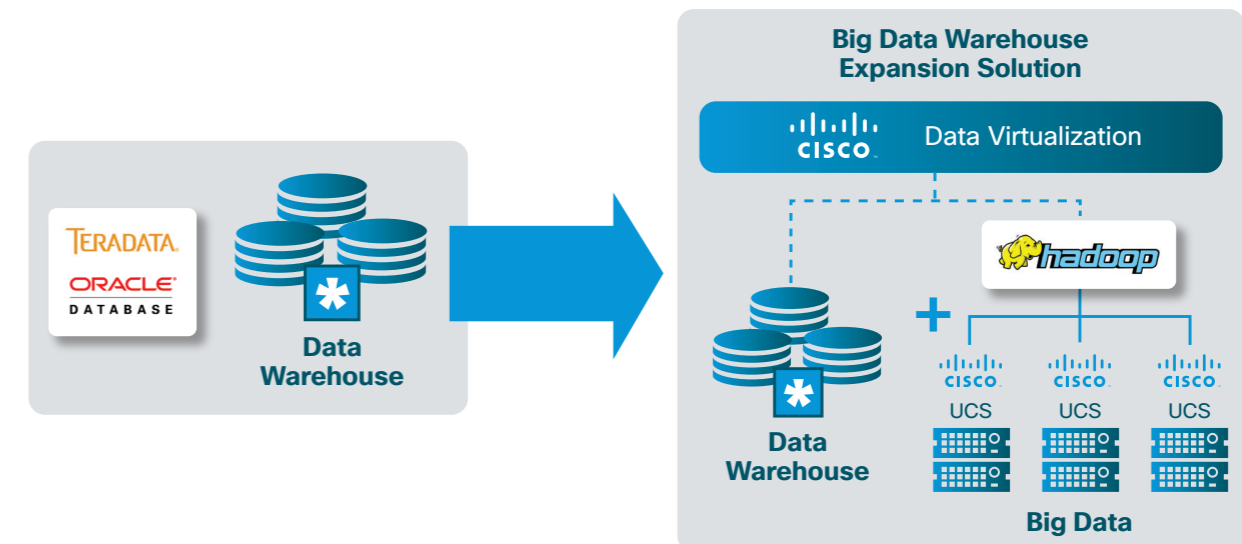
Spotlight on Cisco's Big Data Portfolio: Data Virtualization

The **Cisco Data Virtualization (DV) Suite** abstracts and consolidates multiple different sources to give decision-makers a unified, business-friendly view of the data. One great benefit is its non-invasiveness: no change is needed on existing database systems. You can also bypass the long delays of data replication and physical consolidation traditionally required.

Part of the Cisco DV Suite, **Cisco Business Directory (BD)** is an intuitive web-based interface for business users to easily access and use data. BD provides a self-service directory empowering users to search, browse, and comment on all available data, and categorise large, diverse data sets. The data can be queried by users' preferred analytics or business intelligence (BI) tools.



Based on Cisco DV, **Cisco Big Data Warehouse Expansion (BDWE)** dramatically reduces the costs of the Enterprise Data Warehouse (up to \$100k/TB/year) by offloading infrequently used data to a low-cost Hadoop cluster (typically \$1k/TB/year).



Many Public Sector organisations have turned to Enterprise Data Warehouses (EDW) to centralise their data sets. But the rate at which new data is produced (both structured and unstructured) makes them cost prohibitive. It is also very difficult to build a clear policy on what data needs to be kept 'warm' and what can be archived.

With BDWE, Public Sector decision makers are not only able to access the current data and recent history, but also the extended historical data that is typically archived and not readily accessible. DV acts as a 'virtual database' to access data regardless of whether it is in the original warehouse or the new Hadoop data store.

For more information:

www.cisco.com/go/datavirtualization

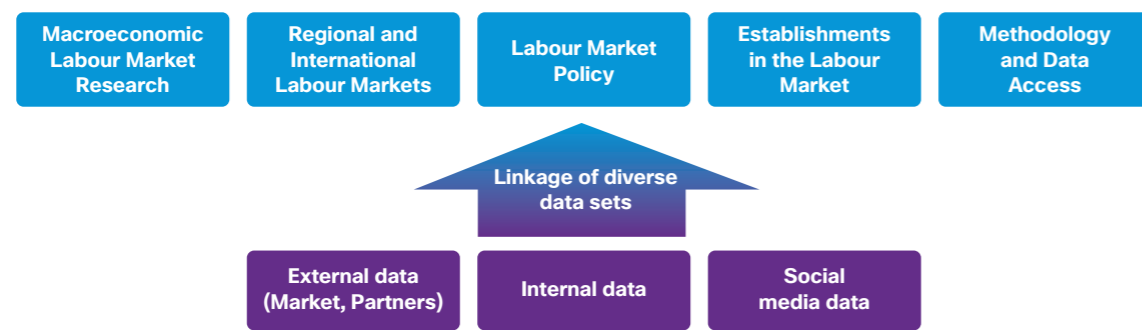
Use Case #2:

Shape new policies with smarter information about citizens and businesses

Germany's Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB) is a special office of the Federal Employment Agency (Bundesagentur für Arbeit, BA). IAB's research on the labour market is designed to help political leaders make well-informed decisions and optimise outcomes, e.g. finding jobs for the unemployed.

Economists, sociologists and researchers create a comprehensive information base, feeding into IAB's key research areas, shown below.

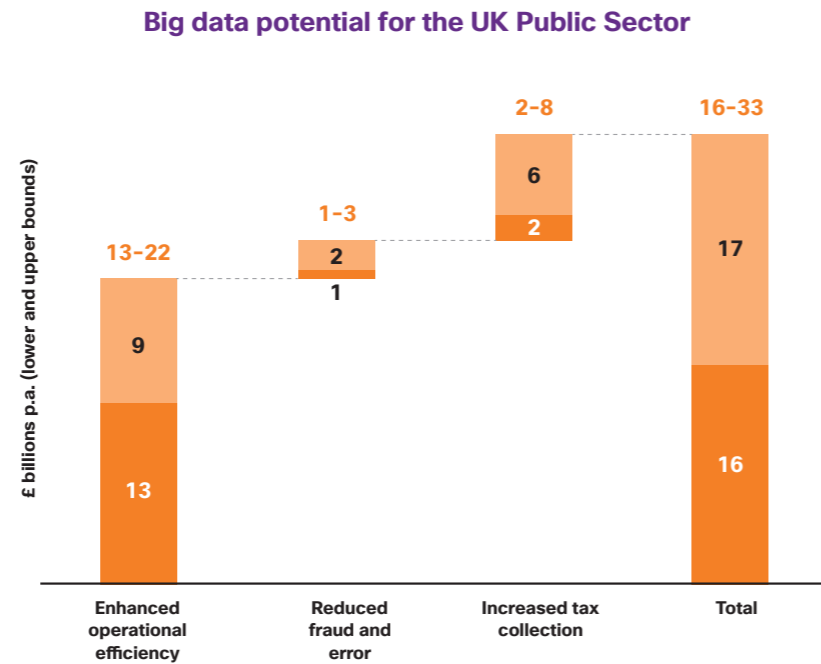
IAB's Big Data capabilities produce and analyse data that helps the agency offer more effective advice. For example, by accurately analysing the outcomes of placement programmes, they can identify those that are less efficient and improve or eliminate them.



In the UK, a recent study (policyexchange.org.uk, The Big Data Opportunity) detailed how Big Data can make public services more responsive. This could help:

- Ensure households receive the benefits and other support they are entitled to, but fail to claim
- Respond to individual public enquiries faster and with more useful information
- Detect public concerns and priorities at an early stage and adjust policy accordingly

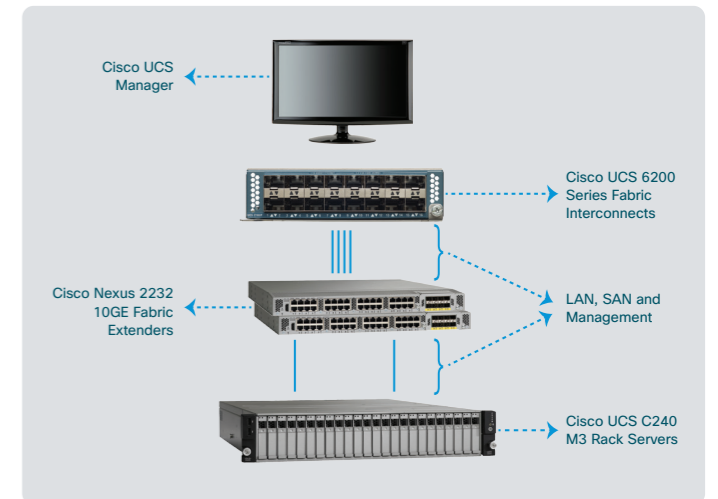
In addition, by making smarter decisions about how departments are organised and what work gets prioritised, the direct cost of government operations can be reduced.



Spotlight on Cisco's Portfolio: Common Platform Architecture for Big Data

Cisco's Common Platform Architecture (CPA) for Big Data provides a highly scalable platform that can be optimised and easily scaled for any size of Hadoop cluster and compute-intensive applications. It comes with pre-validated configurations for leading Hadoop and NoSQL distributions, including Actian, Cloudera, DataStax, Elastic Search, HortonWorks, MapR, MarkLogic, MongoDB, Pivotal, Platfora, SAP, SAS, Splunk, and others.

Built on Cisco's vision of shared infrastructure and unified management for enterprise applications, CPA has been widely adopted in government, education and healthcare. This platform delivers scale without complexity through automated server deployment, using service profiles, which improves IT efficiency and lowers TCO relative to legacy infrastructure.



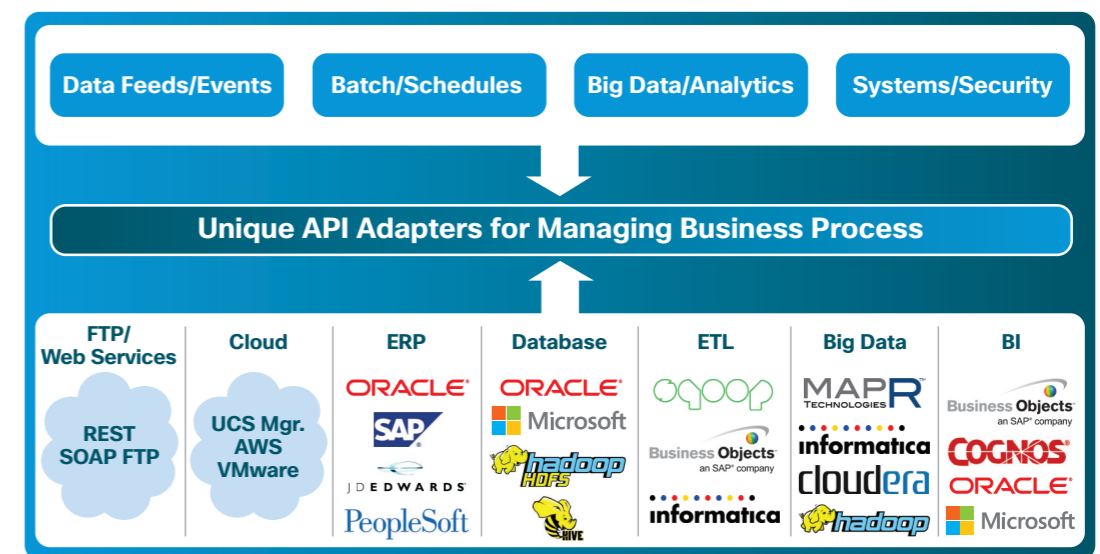
Cisco Tidal Enterprise Scheduler

To keep up with growing demand for the exploration of large data sets, organisations need an easy-to-use workload automation solution.

Cisco Tidal Enterprise Scheduler (TES) fits the bill. An end-to-end workload automation solution, TES offers built-in adapters to Hadoop, Sqoop, and Hive as well as leading enterprise resource planning (ERP), database, data warehouse, data integration and business intelligence applications.

Along with complex, time-based batch scheduling, Cisco TES automates workloads by initiating event-based actions. An event, such as the creation of a new customer record, may trigger an action to move a set of records into a data warehouse. Events can also include running self-service, ad hoc reports for end users or watching FTP folders for changes.

TES runs many thousands of jobs and events a day from a single, centralised server, lowering the burden of managing a distributed environment.



Use Case #3: Publish and secure Public Sector data sets

Open data sets include: Economic Activities, Public Administration, Culture, Education, Tourism, Health, Water and Air quality, Hospitals and Health Centre locations, Disease Statistics, Transport and Mobility, Work and Employment, Weather, Justice, etc.

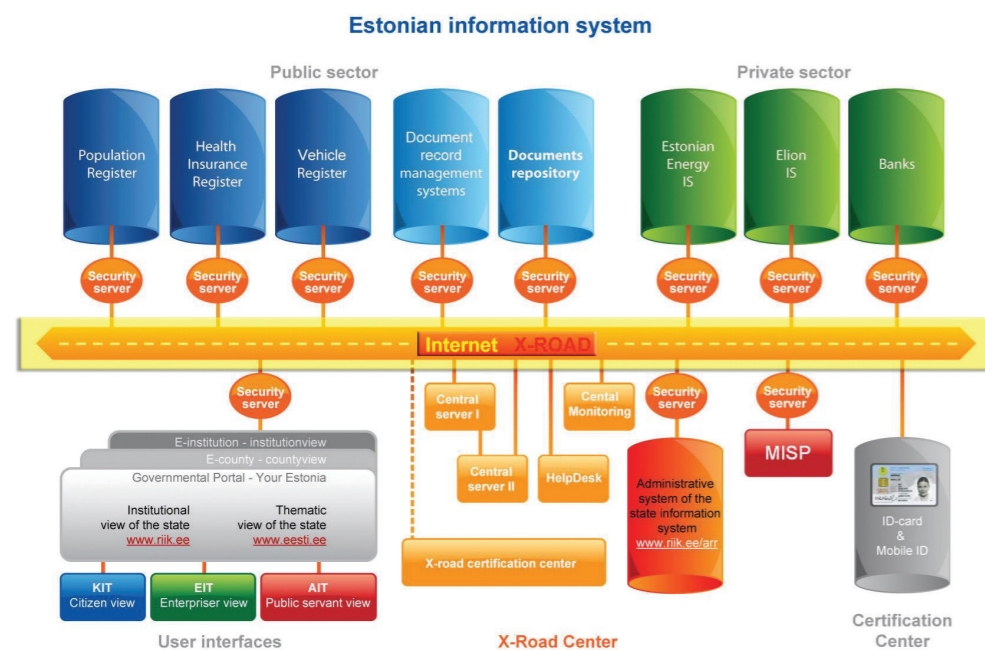
Opening the Public Sector information vault can bring significant benefits:

- Create value and wealth through new data services. For example, the UK Government's platform for third party applications: <http://data.gov.uk/apps>
- Generate transparency, efficiency and trust by evaluating public management

- Facilitate interoperability between administrations
- Complement Public/Private Sector efforts to boost skill sets

Government must standardise the open data publication process, while minimising human intervention. Formats should move from simple (e.g. XLS, CSV, MDB) to advanced (e.g. RDF/XML + APIs) or specialised (Geo Data – ODS, SHP, WMS).

Estonia is widely recognised as a leader in government e-services. The backbone of e-Estonia is X-Road, which allows the nation's various e-services databases, both Public and Private Sector, to connect and operate in harmony.



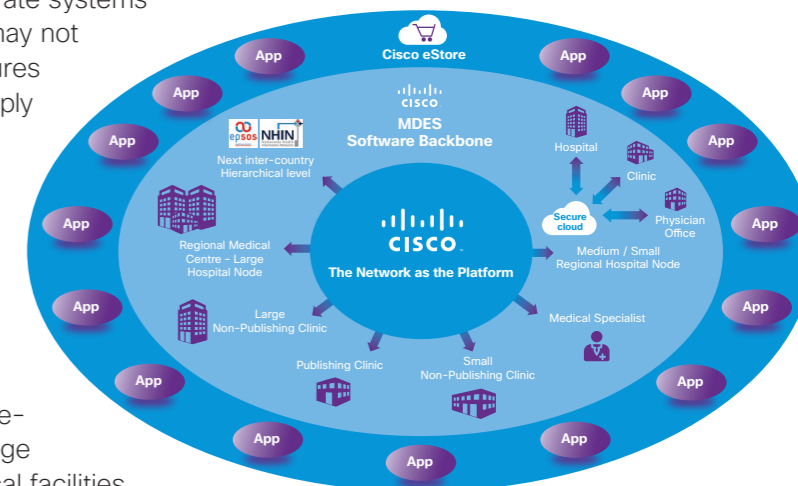
One key element is that databases are decentralised: there is no single controller, every government agency or business can choose the product that's right for them – so services can be added as they're ready.

Designed with growth in mind, X-Road can scale:

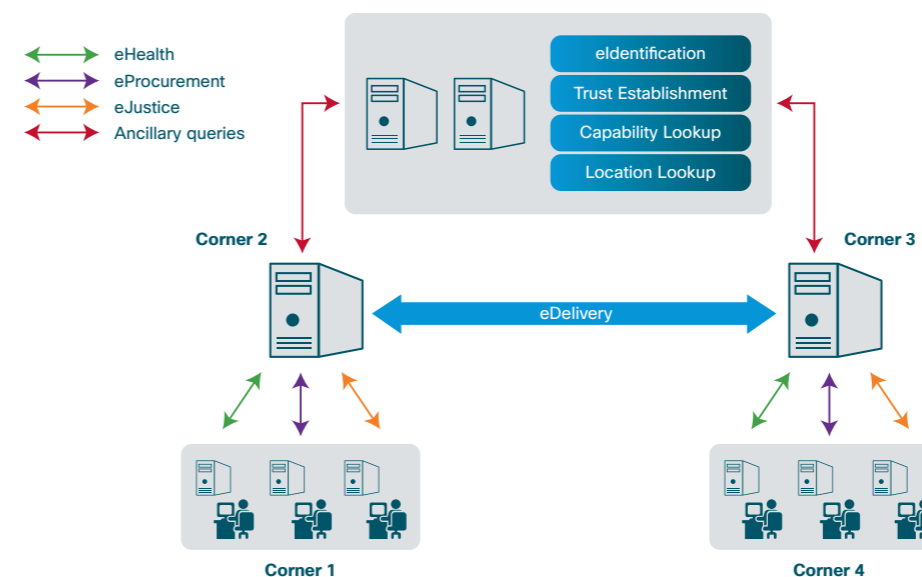
- Over **287 million queries** in 2013
- Over **170 databases** offer their services
- Over **2,000 services** are used
- Over **900 organisations** use it daily
- More than **50% of the inhabitants of Estonia** use X-Road, through the information portal eesti.ee, for e-services like: registering a new residence; checking/updating personal data; submitting taxes; declaring a new-born child; etc.

Spotlight on Cisco's Portfolio: Medical Data Exchange Solution (MDES)

Today, most patient information is stored in disparate systems across healthcare offices. As a result, clinicians may not have full information about patients, redo procedures unnecessarily or expert systems lack access to amply structured, semantically enriched data to further analyse and process. This interoperability challenge is solved with Cisco MDES, a unique combination of Cisco Network/Computing and Tiani-Spirit innovative identity management and communication facility. MDES recognises information coming from known disparate and highly distributed data sources, then conditions and transforms it into human-readable and machine-understandable information according to knowledge representation standards in real time. The technical facilities are based on IHE methodology, an international initiative to improve the way healthcare systems share information by harmonising existing standards through profiling.



The MDES solution integrated **27 Hospitals in Lower Austria** and their systems including: Patient Index, Document Management and Sharing via IHE XDS, workflow handling (e.g. tele-consultation), Identity Management, DICOM, Visualisation, etc.



e-SENS,* a large-scale EU funded project (2013-2016) to enable cross-border and cross domain pilots for consolidating the exchange infrastructures of eHealth, eProcurement, eJustice, eTendering, and Maritime Surveillance (planned), has a similar framework.

The project's target architecture enables the connection, in a trusted environment, of existing national transport solutions (Corner 1 and Corner 4) via single points of contacts or access points (Corner 2 and 3).

Fraunhofer FOKUS together with the Austrian Ministry of Health contribute to the definition of the target architecture for both security and eDelivery. Principally, the team coordinates the definition and facilitates the implementation of pilot strategies, plans and assets. In this case Cisco MDES modular and pluggable architecture is ready to enable the use of additional and sector-specific standards as sender, receiver, and even as a consolidator to enrich and safeguard the semantic expressiveness of medical data as well as an aggregator for subsequent analysis facilities.

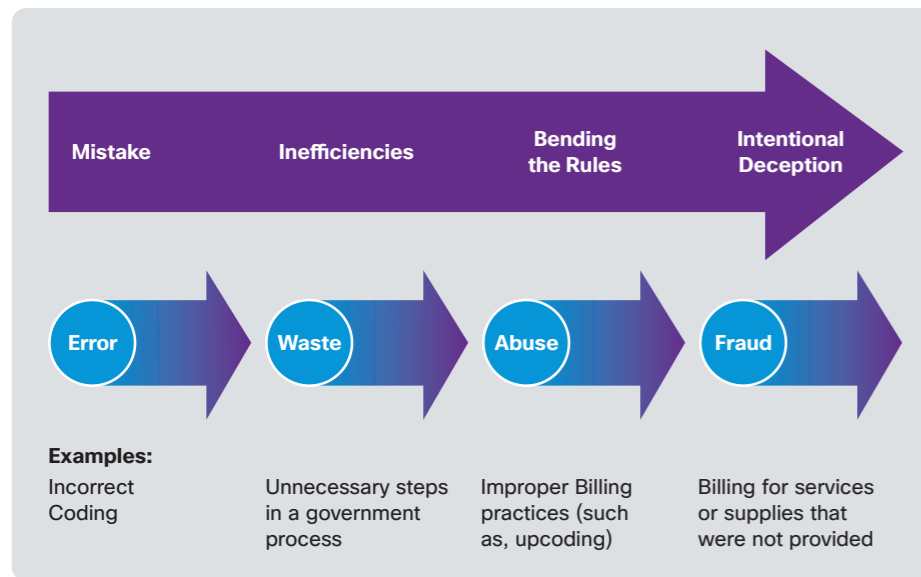
Use Case #4:

Real-time visibility on public service operations, to optimise processes and detect anomalies

By analysing the live data of government, it's possible to see how departments are performing, and find ways to reduce errors, waste, abuse and fraud as well as increase efficiency.

A new generation of analytics technology is helping managers uncover past and current anomalies, while also detecting trends that highlight potential risks.

Turning high volumes of raw data into valuable information with this type of predictive analytics is especially powerful when combined with in-memory database technology. This brings large sets of information into high-speed memory - avoiding the processing delays common when data is stored outside the database.



As City of Boston demonstrates every day with its longstanding performance management programme named Boston About Results (BAR), the technology to detect anomalies can also optimise processes. BAR's interactive web-based platform lets city officials and Bostonians alike know what agencies are doing, how well they are doing it, and where they can be improved.

The BAR programme provides the information technology to help city agencies set strategic goals and measure key performance indicators (KPIs) in a consistent way.

The public has 24x7 access, via the city's website and from mobile devices, to a performance scorecard detailing Boston's progress. This visibility helps ensure public transparency, greater accountability, and a more efficient use of city resources.

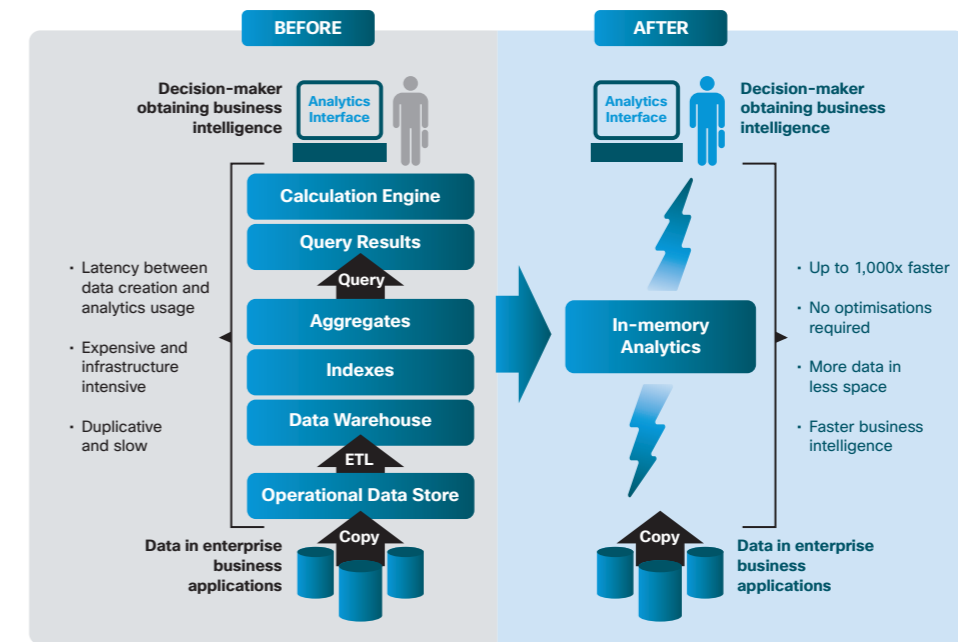
For example, data analytics is helping the city identify and respond to properties where there are persistent criminal activities or violations of the building codes. The city can also predict where problem properties are likely to be and act, even before complaints are made.

Key Benefits

- >2,000** KPIs tracked on a monthly or quarterly basis
- 3 million** Visitors to city websites Q2 of 2013
- 1,200** Unique visits to the public scorecard each month
- 45** Departments all using a common performance tool
- 16** Departments conveying performance through an interactive public scorecard, and more being added

Spotlight on Cisco's Big Data Portfolio: SAP HANA on Cisco UCS

SAP HANA is the next generation of SAP's in-memory analytics, which gives organisations the instant insight to react quickly to changing conditions. Business users can immediately access, model and analyse all their transactional and analytical data in real time, from virtually any data source.



Cisco offers an SAP HANA integrated system built on the Cisco Unified Computing System (UCS) along with the associated Cisco networking technologies. The benefits this offers, according to IDC (<http://cs.co/SAP-HANA-on-Cisco>), include:

ROI Summary for Cisco UCS as a Platform for SAP HANA and other SAP Mission Critical Applications

- 5 Year ROI: **368%**
- 5 Year Total Business Benefit: **\$4.79M**
- Payback Period: **10 MONTHS**

Key Performance Improvements realised from customers who deployed Cisco's UCS Platform

- Reduction in Staff Time Needed for Server Management: **68%**
- Reduction in Staff Time Needed for Server Deployment: **84%**
- Reduction in Productive Employee Time Lost due to Hardware Downtime and Degradation: **96%**

Business Value Benefits - Average Annual Benefits per 100 Users

- Business Productivity: **\$3,108**
- Risk Mitigation/User Productivity: **\$2,580**
- Infrastructure Cost Reduction: **\$701**
- IT Staff Productivity: **\$3,172**

Use Case #5: Analytics for the business of IT

In today's rapidly evolving and complex world, unnecessary inefficiencies and risks keep emerging in the operational IT environments. Did you know that:

Only 10% of compute power is used today to run IT processing workloads (NY Times)

70% of storage remains unused every year in IT departments (Forrester)

Costs per unplanned outages in the datacentre grew 32% between 2010 and 2013 (Ponemon)

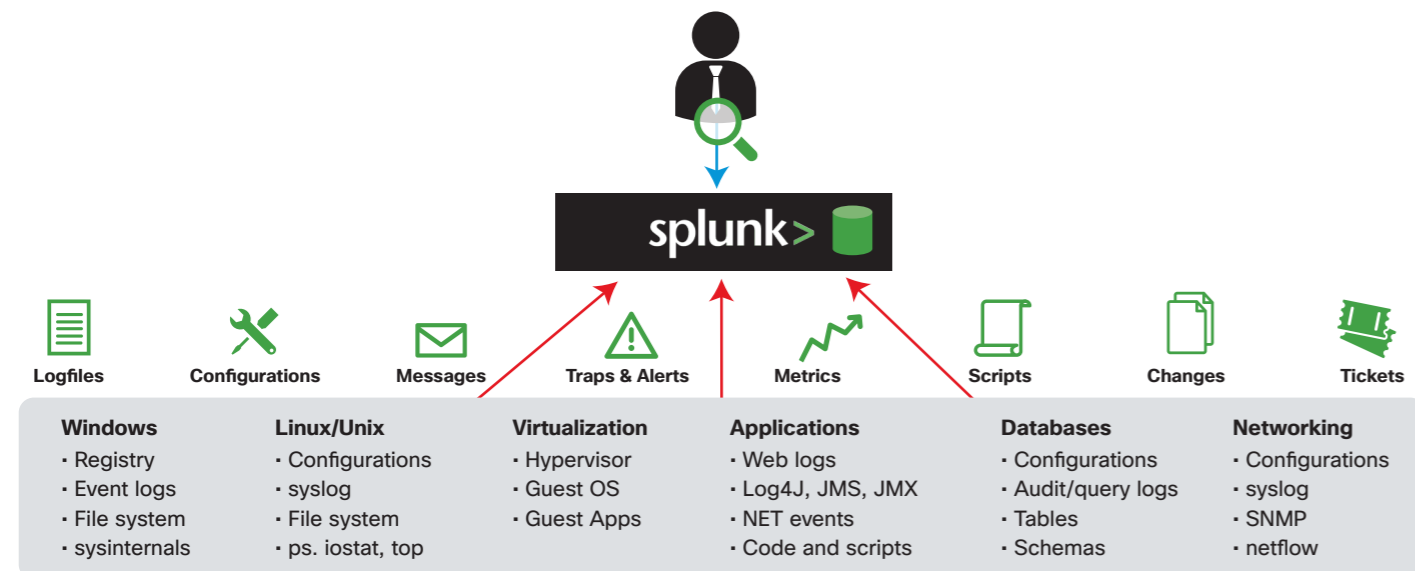
50% of the energy power in a datacentre is wasted (Wired)

CIOs are therefore starting to focus on how to improve IT management by using operational data to inform day-to-day decisions. Issues include:

What are the main problem areas in IT now and in the next 3-6 months?	What is the utilisation and performance of collaboration software?
What IT support skill sets will I need?	What are the process bottlenecks?
What capacity (CPU, storage, bandwidth) will I need to plan for?	Which departments are consuming the most CPU and storage?
How have recent organisational changes impacted core network performance?	Do I need to modify my chargeable accounting model?
How to enhance service management and enhance the IT user experience?	What IT support staff do I need to resolve issues more quickly and within budget?
What are my security vulnerabilities?	Is someone trying to hack into my IT?

But today's datacentres have become very complex. Siloed and interconnected technologies, working in an environment without boundaries, mean it is nearly impossible to find the root cause of problems or proactively identify and prevent outages.

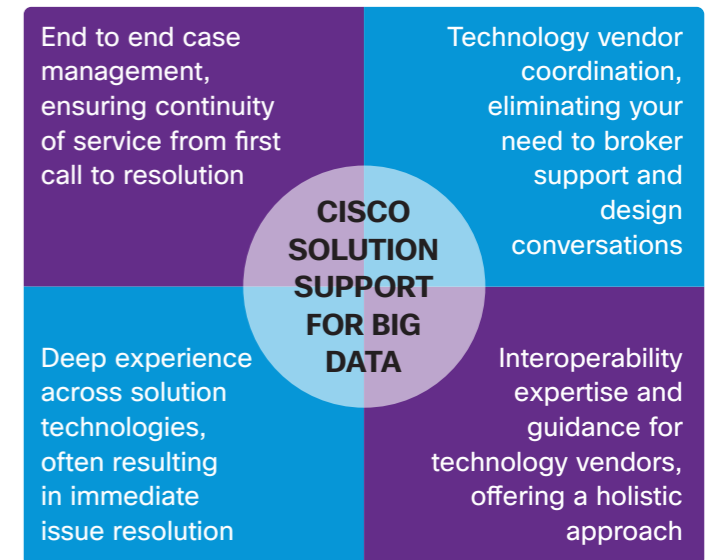
Splunk is a widely-recognised solution. It's a highly versatile and scalable engine for the machine data generated by the IT infrastructure. It collects, indexes and harnesses live data generated from virtually any source, format or location. Splunk lets administrators search and report, in real time, on every single element of the IT environment.



Spotlight on Cisco's Portfolio: Cisco Solution Support for Big Data

Implementing Big Data technology is a challenge for any business: analytics solutions always involve multiple technology vendors. Supporting your new solution can be disruptive and time consuming. While traditional product support ensures that each technology is up-to-date and operating correctly, you also need a solution-level perspective to expertly and quickly resolve system-wide issues. Here's where Cisco Solution Support Services step in.

Solution Support for Big Data features Cisco experts, specially trained in Big Data architectures based on Cisco Validated Designs. It fast-tracks technical support between Cisco and global solution vendors – using an industry-standard framework, recognised code of conduct and established processes to resolve issues across all components in your architecture. Our solution-focused engineers will work to quickly isolate and resolve your problem, regardless of which technology is involved.



Ask yourself:

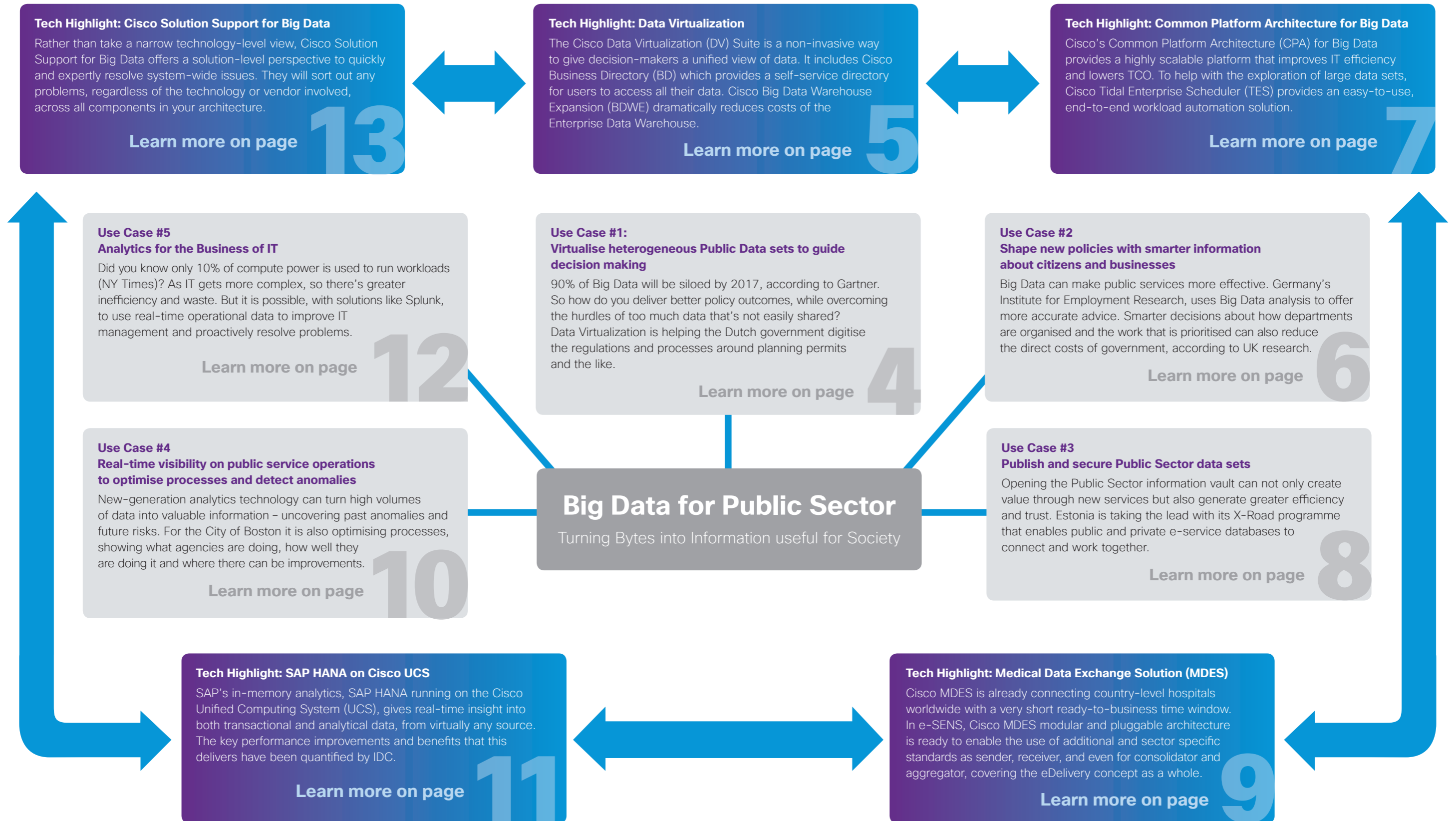
- Do you have the staff and resources to determine what might be happening, if there's a problem?
- Do you have time to coordinate multiple vendors to get your issue resolved – while keeping an eye on interoperability?
- Could you afford downtime, losing access to your mission-critical data, while you are juggling vendors to resolve your issue?

With Solution Support, you hand over critical processes such as case management and vendor coordination to Cisco, freeing your IT team and resources for more strategic initiatives.

Cisco experts will support you through the lifecycle of your Big Data project: Preparing, Planning, Design, Implementation, Operation and Optimisation.



Big Data Use Cases and Technologies Highlights for Public Sector



Next Steps

Are you serious about reaping the benefits of Big Data?

If so, you should nominate a Chief Data Officer to own the Big Data strategy and put in place the policies, practices and organisational model to foster innovation, collaboration and practice sharing.

Here are some key lessons from the past and current Big Data projects with our Public Sector customers.

Start small...

- Focus on the most business-relevant use case(s)
- Deploy the right software with the right hardware
- Build trust and credibility with quality and governance

But think BIG!

- Expect a huge growth of data Volume, Variety and Velocity
- Architectural approach is key to long-term success
- Build a platform: new Big Data projects will emerge, requiring a different mix of technologies

And grow your capabilities

- Leverage existing skill sets, e.g. translate SQL skills to Hadoop
- Build a Common Platform Architecture for all your present and future Big Data projects
- Avoid operational issues with a sound design phase

Cisco brings together Big Data and analytics in a way no other company can. We connect more people, processes, data, and things than anyone else. We bring data access and analytics to the edge of the network to provide near-real-time information and insight. You can use these insights to change behaviour, capture opportunity, respond to threats, and improve your operations.

Cisco has a broader strategy.

www.cisco.com/go/bigdata

This document was co-authored by a multi-national team of cloud experts and Cisco solution architects responsible for Public Sector agencies in EMEA: Alexander Stoklasa, Atanu Roy, Beat Baumberger, Chris Blenkhorn, Frank De Groot, Greg Page, Holger Müller, Jerome Paquay, John Johnson, Luca Lironi, Luca Relandini, Patrick Bikar, Shawn Dawson Troutt, Vernon Thaver, Viktor Hagen.



Americas Headquarters

Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)

Asia Pacific Headquarters

Cisco Systems, Inc.
108 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters

Cisco Systems, International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www.europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

© 2014 Cisco Systems, Inc. and/or its affiliates. All rights reserved. Cisco, the Cisco logo, and Cisco Systems are trademarks or registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries. All other trademarks mentioned in this document are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company.