opportunities to create faster, cheaper and better solutions for big data analytics than traditional approaches can solve."

Process challenges

A major challenge in this context is how to analyse. Shilpa Lawande from Vertica explained that "It can take significant exploration to find the right model for analysis, and the ability to iterate very quickly and 'fail fast' through many (possible throw away) models – at scale – is critical."

According to Laura Haas from IBM Research, process challenges in regard to deriving insights include:

- Capturing data
- Aligning data from different sources (e.g., resolving when two objects are the same)
- Transforming the data into a form suitable for analysis
- Modelling it, whether mathematically, or through some form of simulation
- Understanding the output, visualizing and sharing the results, considering how to display complex analytics on a mobile device.

Management challenges

The main management challenges are related to data privacy, security, governance, and ethical issues.

The main management related challenges are ensuring that data is used correctly, which means abiding by its intended uses and relevant laws, tracking how the data is used, transformed and derived, as well as managing its lifecycle.

According to Michael Blaha, "Many data warehouses contain sensitive data such as personal data. There are legal and ethical concerns with accessing such data. So the data must be secured and access controlled as well as logged for audits".

Conclusion

Big data has definitely made it beyond the buzzword status. It is key for innovation and has a high potential for value creation. There are huge opportunities, for example concerning healthcare, location related data, retail, manufacturing, or social data. There are also challenges, for example concerning data volume, data quality, data capturing, and data management, such as privacy, security or governance.

I would like to conclude this article mentioning an interesting opportunity for big data:

"As more data become less costly and technology breaks barriers to acquisition and analysis, the opportunity to deliver actionable information for civic purposed grows. This might be termed the 'common good' challenge for big data." (Jake Porway, DataKind)

Further information

- "Big Data for Good", Roger Barca, Laura Haas, Alon Halevy, Paul Miller, Roberto V. Zicari. ODBMS Industry Watch June 5, 2012, http://www.odbms.org/blog/2012/06/big-data-for-good/
- Managing Big Data. An interview with David Gorbet, ODBMs Industry Watch, July 2, 2012. ODBMS Industry Watch, http://www.odbms.org/blog/2012/07/managing-big-data-an-interview-with-david-gorbet/
- McKinsey Global Institute (MGI), Big Data: The next frontier for innovation, competition, and productivity, Report, June, 2012, http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.html

Big data analytics for Smart Grid Cities

Stamatis Karnouskos
SAP Research
stamatis.karnouskos@sap.com

The emerging Smart Grid Cities constitute complex ecosystems of heterogeneous entities that densely interact. The huge amount of data generated by the billions of envisioned interconnected devices will need to be analysed in a timely manner and is expected to provide new insights. Advanced business services will take advantage of the near real-time information flows among all participants and provide new innovative functionalities that will impact the way we design, develop and operate complex infrastructures as well as the associated provision of applications.

Big data infrastructures

In an era where advanced networking capabilities are built in everyday appliances, any device that can communicate its energy production or consumption can be considered as a smart meter. However, in conjunction with such energy measurements, additional information with respect to the device's status, its processes as well as other task- or device-specific information could be shared. This transforms smart cities into a big data infrastructure that provides detailed information at multiple layers (device, network, application, etc.) and where services and applications can now tap into and enhance their own capabilities.

As we can see in the figure (next page), devices, systems, and services will take part in "flat" and information-driven interactions; the latter will be done not only in a peer-to-peer method, but also assisted via advanced cloud services that will make the benefits of big data analytics available to all stakeholders. As an example, the IMC-AESOP project (www.imc-aesop.eu) is designing cloud-based SCADA/DCS systems that are in the heart of modern industrial infrastructures, and which rely on a mix of on-device and in-cloud services to do monitoring and management of the real-world processes.

Big data analytics

The huge amount of data acquired will be amalgamated in the cloud, enriched with context-specific and system-wide aspects as well as business relevant information, and enable us through analytics to better understand the physical world, its processes, the impact on the business side and eventually take more informed decisions. Although big data existence and analytics don't guarantee better decisions, potential new insights that may be acquired may materialize to more effective problem tackling and business advantages.

As an example, in the smart grid, analytics empower scenarios of grid infrastructure optimi-

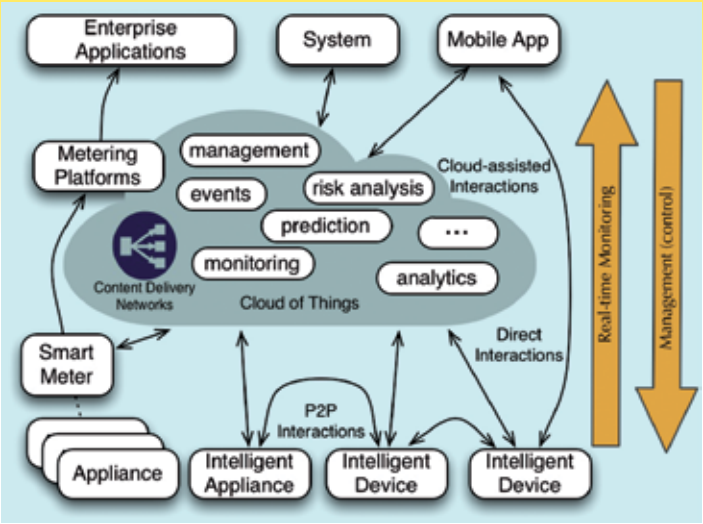

Figure: Value-added services empowered by big data analytics in the Cloud of Things

zation, energy management scenarios with participation of residential prosumers (energy producers and consumers), energy trading, and better planning of energy infrastructure in cities. Big data analytics is seen also as the key to understanding complex systems of systems, such as the emerging smart cities. The SmartKYE project (www.smartkye.eu), for example, aims at enabling municipalities to better understand and manage aspects of a smart city via a business cockpit. The SmartKYE solution is based on analytics that can be done on massive data obtained from energy management systems and other infrastructure and business systems.

Challenges

There are significant challenges when big data analytics in such complex cyber-physical infrastructures are considered. For instance, designing software solutions for such infrastructures, and analysing the impact, e.g. of malfunctions, at system-wide level can be assisted by big data analytics. A new generation of data exploration tools as well as sophisticated algorithms considering context-specific information at several levels for very large-scale systems will need to be designed, developed, and piloted. A major challenge is extracting and understanding the business-relevant information under temporal constraints and being able to effectively build in solutions that utilize the monitor-analyse-decide-manage approach for a multitude of domains. The high heterogeneity of systems and models, the varying quality of data and associated information as well as the uncertainties and complex system-wide interactions will need to be investigated to identify business opportunities and realize a business benefit. Considering also that in smart cities much of this data will be directly attributed to their citizens, data lifecycle management approaches will need to be investigated.

Conclusion

Massive data generated by billions of devices, users and their interactions will need to be acquired and analysed in specific contexts, such as the smart city, in order to better understand and manage the increasing complexity. This implies a paradigm shift on how we design, realize and operate software services, applications and systems. New insights based on big data analytics may lead to better decision-making and business advantages.

Further information:

- IMC-AESOP project – www.imc-aesop.eu
- SmartKYE project – www.smartkye.eu

The cost of big data – An inconvenient truth

Anastasius Gavras
Eurescom
gavras@eurescom.eu

There is a widespread expectation that big data offers tremendous opportunities. The potential economic value to be generated from the vast amount of data is given in tens or even hundreds of billions of euros per year. However, the cost of big data is rarely mentioned.

In my non-exhaustive research I hardly found figures explaining the long-term cost of big data. Yet, there must be a price tag attached to the investments needed for handling big data, including operational expenses for maintaining the data and making them available in the long term.

At the EIT Foundation Annual Innovation Forum that took place end of March 2013 in Brussels, Commissioner Neelie Kroes said that every two days we create as much information as was created from the dawn of civilisation to 2003. In addition she claimed that big data is growing by 40% per year, a figure that is hard to correlate to the previous statement and in fact even harder to believe if we accept that Moore's law will likely continue to be valid until at least 2020, implying that most likely the growth of big data is higher.

In 2011, data management provider EMC claimed that the world's data volume is growing faster than Moore's law, and that 1.8 trillion gigabytes would be created and replicated in 2011.

Expectations

The traditional big stakeholders in very large database management systems see a great opportunity and a new market potential. However, a large share of the value to be created will come from new types of data use which are unprecedented. There are tremendous investments to be made in storage, computation and transmission capacity; and there are undoubtedly costs for keeping the systems running, including energy costs.

High expectations are put in non-ICT sectors that so far, although users of ICT, had little exposure to big data because it was not easy and it was costly due to highly specialised solutions. Sectors often mentioned include energy, environment, agriculture, medicine, government and many others. The sectors today are either in-