



Digging up value with big-data mining

Domenico Talia will speak about big-data mining and knowledge discovery at ISC Big Data in Heidelberg, Germany, on 1-2 October. Talia, who is a professor of computer engineering at the University of Calabria in Italy, will explain how cloud computing can offer effective support for addressing both the computational and data-storage needs of big-data mining and parallel analytics applications. In an in-depth interview, he tells iSGTW about his research in this area and argues that it's time for the public sector to invest in big-data, so as to provide better services to citizens.

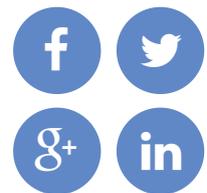
Domenico Talia will speak about big-data mining and knowledge discovery at [ISC Big Data](#) in Heidelberg, Germany, on 1-2 October. Talia, who is a professor of computer engineering at [the University of Calabria](#) in Italy, will explain how cloud computing can offer effective support for

**Posted on SEP 17
2014 7:18AM**



Andrew Purcell
European editor

Share this story



[↻ Republish](#)

Tags

analytics

big data

cloud

Heidelberg

HPC

interview

ISC

ISC Big Data

value



"Scientific sectors that fail to make full use of the huge amounts of digital data available today risk losing out on the significant opportunities that big data can offer," says Talia. Front page image shows Talia speaking at [SC13](#) in Denver, Colorado, US.

addressing both the computational and data-storage needs of big-data mining and parallel analytics applications. In an in-depth interview, he tells iSGTW about his research in this area and argues that it's time for the public sector to invest in big-data, so as to provide better services to citizens.

What is 'big data'?

Many definitions have been proposed for this term. Most of them, however, refer to the features identified by Doug Laney in [the short META Group report he wrote in 2001](#): big volume, big velocity, and big variety (even if the term big data was not actually used in that report). In more detail, big data refers to massive, heterogeneous, and often unstructured digital content that is difficult to process using traditional data-management tools and techniques.

The proliferation of data warehouses, webpages, audio and video streams, tweets, posts and blogs is generating a massive amount of complex and pervasive digital data. This trend created the big data phenomenon. However, we should recognize that data itself is not necessarily important *per se*. Data is only important if we are able to extract value from it, thus adding the fourth 'v' to big data.

So, how can value be extracted from all this data then?

Extracting useful knowledge from huge digital datasets requires smart and scalable analytics algorithms, services, programming tools, and applications. Advanced data-mining techniques and associated tools can help to extract information from big, complex datasets. This can be useful in enabling businesses and research collaborations alike to make informed decisions.

The combination of big-data analytics and knowledge-discovery techniques with scalable computing systems is an effective strategy for

producing new insights in a shorter period of time. Big data analytics use compute-intensive data-mining algorithms that require efficient high-performance processors to produce timely results.

What role does cloud computing have to play?

Among other scalable storage and computing platforms, cloud computing infrastructures serve as an effective platform for addressing both the computational and data-storage needs of big-data analytics applications.

Much big data already resides in the cloud, and this trend will increase in the future. For example, Gartner estimates that, by 2016, more than half of large companies' data will be stored in the cloud. This trend requires that clouds become the infrastructure for implementing pervasive and scalable data-analytics platforms.

Although not many cloud-based analytics platforms are available today, current research work anticipates that they will become common within a few years. Some current solutions are based on open-source systems, while others are proprietary solutions provided by big companies and start-ups.

As more such platforms emerge, researchers will port increasingly powerful data-mining programming tools and applications to the cloud, exploiting complex and flexible software models such as [MapReduce](#) and the distributed workflow paradigm. My colleagues and I from the University

of Calabria have created one such data-analysis platform, which is delivered through a [software-as-a-service \(SaaS\) model](#) and is targeted at business users. It is now being further developed through the start-up company we founded together called [DtoK Lab](#).

How, in your view, is big data changing scientific research?

Solving problems in science and engineering was the first motivation for inventing computers. Today, they are still the main areas in which innovative solutions and technologies are being developed and applied. As the data scale increases, we can address new challenges and attack ever-larger problems. New discoveries will be achieved and more accurate investigations can be carried out due to the increasingly widespread availability of large amounts of data. Scientific sectors that fail to make full use of the huge amounts of digital data available today risk losing out on the significant opportunities that big data can offer.

And how is it changing wider society, including business?

Perhaps private companies and big internet players such as Facebook, Google, and eBay are the main users and beneficiaries of big-data management and analysis. They are using it to improve their businesses and increase customer fidelity. In my view, it's now time for the public sector to invest in big-data collection and analysis, too. This could

improve the quality of life of citizens and the efficiency of public administrations.

The use of cloud for big-data management and analysis represents a good investment for governments. In Europe, for example, big-data analysis should be used for social goods, such as disease prevention and control, public security, pollution prevention, and other public services.

Finally, are there any presentations or discussions that you're particularly looking forward to attending at the ISC Big Data event?

Yes, there are many. I'm especially excited about the talks on the convergence of big-data analytics and high-performance computing systems, as well as [Dirk Slama's keynote speech addressing the impact of big-data analytics on the 'internet of things'](#). The big-data case studies session should also provide some useful insights, as should the sessions on security and privacy.

I plan to attend the ISC Cloud event immediately beforehand, too. I am really interested in some of the talks that will focus on cloud in relation to various big-data issues, systems and applications. I am sure these sessions will be helpful in gaining a more complete overview of the state of the art in the key areas of big data and cloud computing.

Join the conversation

Contribute



Do you have story ideas or something to contribute? **Let us know!**

OUR UNDERWRITERS

Thank to you our underwriters, who have supported us since the transition from International Science Grid This Week (iSGTW) into Science Node in 2015. We are incredibly grateful.

[View all underwriters](#)

CATEGORIES

Advanced computing

Research networks

Big data

Tech trends

Community building

CONTACT

Science Node

Email:

editors@sciencenode.org

Website:

sciencenode.org



Copyright © 2022 Science Node™ | [Privacy Notice](#) | [Sitemap](#)

Disclaimer: While Science Node™ does its best to provide complete and up-to-date information, it does not warrant that the information is error-free and disclaims all liability with respect to results from the use of the information.