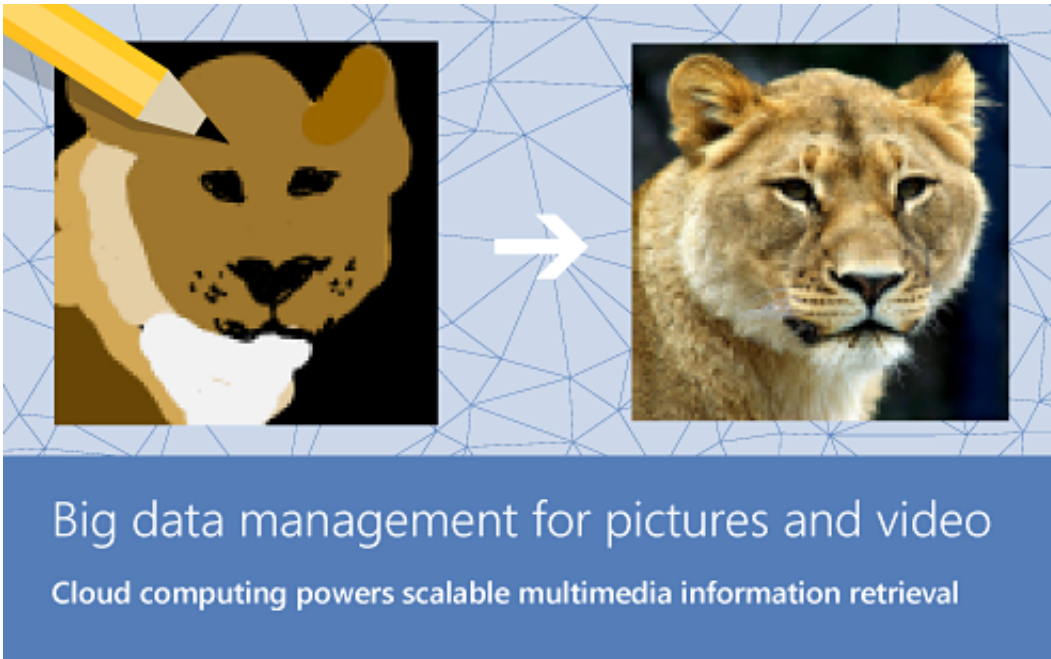


Big data tamed with the cloud

September 8 2014, by Kenji Takeda



Big data: it's the hot topic these days, promising breakthroughs in just about every field, from medicine to marketing to machine learning and more. But for many of us, the problems of managing big data hit home when we confront the welter of digital photos and videos we have recorded with our smartphones and cameras. Multiply this by the number of people doing this around the world and it is a big problem. On the surface, it does not seem like an endeavor on the order of treating cancer (more on that later), but it is a colossal headache to organize, classify, search, and retrieve our multimedia content—and designing

systems to do this at scale effectively is a huge challenge.

Thankfully, Professor Heiko Schuldt and Ivan Giangreco of the Databases and Information Systems (DBIS) Group at the University of Basel are working on a project to do just that, and a whole lot more. Their integrated system harnesses the power of the cloud, through Microsoft Azure, to understand and sort through the terabytes of data that make up multimedia content to find and return like objects.

The Basel team's system combines the power of relational databases, with the adaptability of [information retrieval](#) systems. The Basel system can handle and store any type of multimedia data, including their features. When an algorithm for feature extraction is defined, the system automatically executes the extraction, storage, and indexing of both the feature data and the object itself. This approach efficiently carries out Boolean queries as well as searches based on ranking images based on their feature similarity scores. In addition, it provides novel query paradigms and interfaces; for example, you can sketch an image or parts thereof and find images that are similar to your sketch.

It's exciting to see how this work has progressed since the Basel researchers attended our first European Microsoft Azure for Research training workshop at ETH Zurich last November. They successfully applied for an Azure Award, which got them up and running on the cloud within a few weeks. This allowed the team to quickly develop and deploy their system in a scalable way. Microsoft Azure is ideal as a fast, distributed storage and computing fabric for running the Basel team's project, whose MapReduce-style program can grow as millions of images are added to the system. By moving to the cloud, the Basel researchers have been able to develop, deploy, and demonstrate the system, testing their ideas at scale on the 14 million images that comprise the ImageNet database. They presented this work at the IEEE International Congress on Big Data ([BigData 2014](#)).

Professor Schuldt explains how Azure has helped him with his research. "In large-scale image retrieval, both effectiveness and efficiency are essential requirements. Thanks to Microsoft's support and the use of the Azure cloud, we have been able to successfully address the retrieval efficiency so that we can concentrate further on retrieval effectiveness, especially by developing novel search paradigms and user interfaces based, for instance, on gestures or sketches."

The Basel researchers are looking forward to tackling the even bigger Bing Clickture dataset, which contains 40 million images. They also plan to test the system on video content, in what they're calling the IMOTION project, which will "multiply the challenges in terms of retrieval efficiency," notes Professor Schuldt. Their next paper was presented at 37th International ACM-SIGIR Conference on Research and Development in Information Retrieval, and we're looking forward to seeing how the team continues to [push the boundaries of big data by using Microsoft Azure](#).

Now back to that earlier comment about treating cancer. Approaches similar to those used by the Basel team's project might, in fact, someday help us to better understand and treat cancer. The underlying computer science and cloud technologies could be used, for example, for managing and analyzing MRI scans of tumors.

The Basel team's project is just one example of how easy it is to get up and running on the cloud and accelerate your research—especially when by taking advantage of the Microsoft Azure for Research initiative, which offers not only training but also substantial grants of Azure storage and compute resources for qualified projects. [Read about the initiative](#) and our [requests for proposals](#). Who knows? Maybe your project will be the next big thing in [big data](#).

Provided by Microsoft

Citation: Big data tamed with the cloud (2014, September 8) retrieved 1 May 2024 from <https://phys.org/news/2014-09-big-cloud.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.