



Big Data - No Longer The Preserve of Big Business



There is no longer any need for companies to miss out on the benefits of big data, because the vast web search and analysis capabilities of Hadoop are now accessible to everyone. The key is that no-compromise bare metal processing capabilities are now increasingly being used to deliver them, making big data analytics viable and cost-effective for even the smallest start-up.

Big data, big advantage

One of the reasons 'big data' has attracted so much hype is because it presents one of the most striking examples of the gulf between the workplace and life outside. In the wider world, knowledge discovery is a rich and immediate experience thanks to the sophistication of Internet search. By contrast most businesses remain swamped with data they can't readily combine, mine or interpret. This is inhibiting their ability to innovate and compete.

Recognising the opportunity, cloud providers are clamouring to offer Hadoop on a pay-as-you-go basis so that companies don't have to invest in huge, expensive data centres if they want to do more with their data.

Hadoop is seen as the panacea to data blindness because it allows companies to delve deeper and discover more. It does this by applying supercomputer/parallel processing-style capabilities to everyday business data. Hadoop is a collection of tools, amongst which a distributed filesystem (HDFS), used to split vast amounts of data into manageable chunks across multiple servers, so it can be analysed quickly. At its core is MapReduce technology which allows related data to be split up, analysed and brought back together again.

Although other manifestations of the principle exist, Hadoop is the most mature in an immature market.

Who needs Hadoop?

Making Hadoop facilities available in the cloud opens the door to any business of any size to begin mining its data and experimenting with it to drive commercial advantage.

At the top end of the scale, it offers retailers, telecoms companies or oil exploration organisations a cost-efficient means of drilling down into the minute, most granular detail of their data - to arrive at new insights and discover untapped opportunities. At a more transformational level (if the business model is right) cloud-based Hadoop brings sophisticated business analytics to the masses, allowing even the smallest start-up to begin harnessing big data – with potentially disruptive consequences.

To be truly responsive, companies today need to be able to track trends on social media for example. While large organisations are still trying to work out what to do with Twitter and Facebook, enterprising start-ups are using cloud-based Hadoop to cut to the chase – identifying less obvious trends and distilling from these the elements of a new product or service which fills a gap in the market.

Would-be ‘big data’ consultants or systems integrators, meanwhile, can use cloud-based Hadoop to inexpensively power up and test new services, developing a proof of concept to show to clients – ie. without the risk of investing in their own big-data infrastructure. For providers of cloud migration and cloud integration services, it offers the chance to embrace big data as a new revenue stream.

Protecting performance

Although cloud-based Hadoop services have been available before now, typically these have been dependent on server virtualisation, the building blocks of most cloud services. One of the highest-profile examples is Amazon’s Elastic MapReduce. The virtual-machine model of cloud delivery is at odds with the way Hadoop works, however.

Hadoop was designed to run in large physical data centres, on dedicated physical machines. A whole wiki page on [Apache Hadoop](#) has been devoted to the implications of running Hadoop in a virtualised environment. It advises that where organisations are forced to choose between *virtual* Hadoop or *no* Hadoop, they need to be prepared for reduced performance - and compensate by allocating more virtual machines (incurring greater cost). Performance is likely to be variable too, storage costs are likely to escalate, and data integrity and security could be at risk of compromise if not adequately provided for.

This is where bare-metal cloud services come in. These provide a third alternative in the form of a dedicated high-performance server environment (using no virtualisation, therefore no performance-hampering hypervisor) accessed on demand via the web. For big-data applications, where the need for reliable, uncompromised performance is paramount, bare-metal cloud services have huge appeal because they promise to deliver results faster and more economically.

Fostering ambition

Buying a dedicated Hadoop infrastructure outright isn’t practical or affordable for most companies; it would cost hundreds of thousands of pounds just to deploy a farm of 20 servers; then there’s the cost of acquiring the right technical skills, supporting the systems and keeping the technology

current. Given that a lot of companies are still at an experimental stage with big data and don't have a clear enough idea at the outset of what they will do with the technology, this kind of investment is never going to get the green light in any case.

It is no coincidence that Gartner estimates that the global cloud computing market will have grown by 18.5% in 2013 to £87 billion; cloud-based infrastructure (IaaS) services alone are experiencing 42% annual growth. Forrester meanwhile has noted that one of bare metal's biggest benefits is its support for more complex applications - those that couldn't previously be moved to the cloud because of their incompatibility with server virtualisation.

Now, rather than compromise with a 'make do' virtualised Hadoop scenario in the cloud, companies with pressing big-data needs have the option to buy affordable, fit-for-purpose bare-metal services - allowing them to power up a dedicated infrastructure as needed, meaning they can get started with big data in a sustainable, low-risk way, gradually adding servers as experimentation pays off.

www.bigstep.com