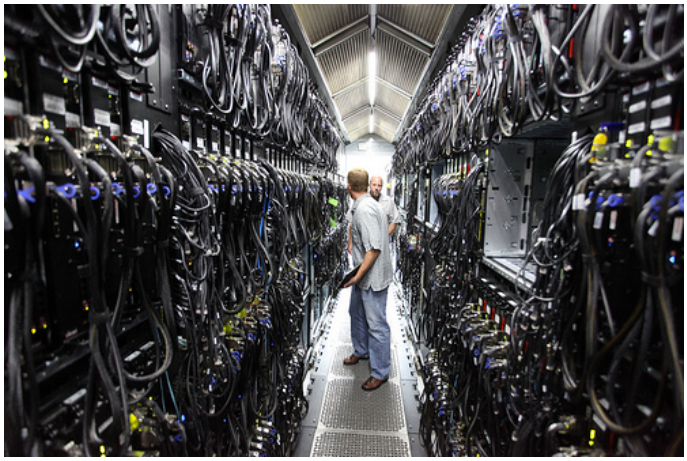
Making mega data centers more durable

Researchers in China have come up with a blueprint for restructuring mega data centres to boost performance by reducing the impact of hardware failures



A shipping container packed with computers. Image courtesy Robert Scoble, Flickr, CC-BY 2.0.

Posted on AUG 22 2012 8:32AM

Andrew Purcell
European editor

Share this story

Behind every cloud is a mega data center. And, as the popularity of cloud-computing grows, so does

the need for these centers to become ever-larger and ever-more efficient. Facebook, Google, Twitter, Amazon, Apple, and Microsoft all rely heavily on mega data centers to provide their services to growing numbers of users around the world. Thus, it is of paramount importance to these companies and others like them, to ensure that their centers are operating as efficiently as possible at all times. Key to achieving this, of course, is minimizing the detrimental effects localized hardware failures have on the overall capacity of mega data center networks. Now, research from the Chinese National University of Defense Technology (NUDT) has provided a significant step forward in solving this problem.

A team of scientists based at the National Key Laboratory of Parallel and Distributed Processing at the NUDT has come up with a way to restructure mega data center networks to make them significantly more fault tolerant, allowing distributed software applications to maintain performance even in the case of multiple hardware failures. Their research is published in the journal *Science China Information Sciences*.

Lead author Feng Huang explains that as cloud providers have seen the amount of IT they manage grow, they have resorted to packing computers into shipping containers and linking these together. Typically, a standard 20 or 40 foot shipping container is equipped with 1200-2500 servers, with the number of servers in a container fixed during its lifetime. Such containerization lowers the total

cost of ownership for companies and allows operators to manage the mega data center using what they term a "service-free" model, which basically means a container as a whole is never repaired during its deployment lifespan (usually around 3-5 years). Provided the performance of the entire container meets an engineered minimum criterion, there is no continuous component repair. While popular with companies such as Google and Amazon, Huang argues that using containers in this way to rapidly building up mega data centers with a modular structure can leave companies vulnerable, since any crimp in the inter-container networking performance can have a huge effect on facilities: the fact that network performance degrades faster than computation or storage capacity is this system's ultimate weakness, since this causes the container's overall performance to decrease below the threshold criterion and end its lifespan prematurely.

To tackle this problem, Huang and his team have come up with a new way of structuring these 'inter-container' networks, which they call SCautz. The key to the SCautz method is that it allows servers to carry out many of the typical functions of network switches, thus leaving the actual switches to focus on inter-container data transfer. The full logical structure of this hybrid approach can be seen in the diagram below:

In tests against BCube, a Microsoft-led experimental network architecture for modular data centers, SCautz performed well. While BCube

just had the edge in terms of performance when both systems were running at 100 per cent, SCautz was able to route around hardware failures much more quickly - when a server fails somewhere in the system, SCautz simply finds a peer server in the same cluster to bypass the failed one. In cases where 10 per cent of network hardware failed, SCautz throughput dropped by just seven per cent, whereas BCube throughput dropped by as much as 15.3 per cent. Equally, in cases where 20 per cent of network hardware failed, Scautz throughput dropped by just 14 per cent, compared to 25 per cent for BCube. Thus, in the case of SCautz, its ability to route around failed hardware meant that network performance actually degraded by less than the total amount of hardware which became unavailable in each instance.

This increased resilience potentially gives operators of mega data centers far greater flexibility in responding to hardware crises. However, this isn't the only advantage SCautz offers over other typical architectures for modular mega data center networks. SCautz's ability to run in different modes - with switches on or off - means that it can easily handle sudden increases in network flows effectively without lowering the quality of bandwidth-intensive applications. However, perhaps most importantly, because SCautz requires far fewer switches than other architectures, it is generally a much cheaper solution. According to the theoretical analysis conducted by the researchers involved, a typical SCautz-based container with 1280 servers only needs 160

commercial off-the-shelf switches. Of course, as impressive as this all sounds, the next stage is for SCautz to be implemented and examined in a larger production data center. Should these further trials prove equally successful, then cloud-computing providers may start to get really excited.

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.o
Website:
sciencenode.org

GIVE NOW