The Evolving Cloud

Joe Weinman

JUST AS A REAL CLOUD CHANGES SHAPE AS IT IS BOTH FED AND BATTERED BY THE FORCES OF NATURE SUCH AS SUN, WIND, AND TOPOGRAPHY, cloud computing is continuously evolving due to the forces of innovative new technologies and business models and an expanding set of customer use cases and a shifting competitive landscape.

A few major trends are worth highlighting.

Dispersion and Ubiquity. The first one is the evolution of cloud and IT architecture from centralized, hyperscale data centers to a much more dispersed approach, spanning colocation and interconnection facilities and connected, highly dispersed computing and storage capabilities. This includes service provider and private nodes such as microcells that might be deployed at the edge, and “pervasive” or “ubiquitous” computing: in-device processing and storage as immense capabilities now become available at declining cost not just in smartphones but also in automobiles, video cameras, wearables, and so forth. In an executive insight roundtable I moderated recently at the Pacific Telecommunications Council, industry thought leaders from a range of firms across the IT value chain explored the drivers and implications of this new architecture. For example, IT is shifting from an internal focus to an external one, requiring a more dispersed global presence to reach global customers. The edge/fog layer can serve not only to reduce latency and backhaul costs by processing and compressing data locally, it is the source of data for centralized aggregation and processing to feed the insatiable appetite of machine learning engines. However, solving one problem can create others. Network topologies, operations, and administration may become more complex as rich business partner and computing ecosystems emerge to interconnect and manage all of these centralized and distributed elements. A condensed transcript of this discussion is in this issue in the article titled “Roundtable on Cloud, Fog, Networks and Related Technologies.”

Granularity. The second major trend is increasingly fine granularity of processing services. We have gone from mainframes to minis to PCs, of course, but more recently we have gone from virtual machines to containers to microservices and Lambda/cloud functions. Short of making a single machine instruction available for rent, it is hard to see how we can get much more granular than that. In the last issue, Disney’s Adam Eivy warned, though, that users must be wary of “per-hit” pricing: those micro-pennies per transaction can add up as your application scales, making it more cost-effective in some cases to use virtual machines or even dedicated hardware.1 Today’s firms must balance software architecture considerations including microservices and cloud functions, organization structure, economics, and scalability; a challenge that is increasingly complex due to the plethora of options.

Hybrid Multicloud. Although many cloud service providers would love for you to go “all in” to the cloud with only their services, and not their competitors’, reality—in terms of current state and trends—is much more complex. Enterprises are us-
ing a mix of private and—one or more—public clouds for infrastructure, platform, and software services. There are many reasons to use multiple clouds. At the infrastructure layer, a mix of private and public made seamless through a common stack/containers or by selecting a vendor/provider that offers consistent cloud functions in private and public flavors can optimize cost and elasticity through cloudbursts.

Using multiple public infrastructure-as-a-service clouds via a cloud broker or marketplace can leverage the best features of each, or can serve business continuity / disaster recovery needs, since cloud outages are not unheard of, and may never be completely ruled out. At the software-as-a-service layer, multiple cloud providers can be used individually for their best-in-class capabilities or in combination in an integrated workflow.

These different approaches can drive a variety of economic benefits, as I analyzed in depth in the January/February issue. However, orchestrating such workflows across multiple clouds and a hybrid cloud/edge architecture is not trivial. In this issue, Rajiv Ranjan and his colleagues address the complexities and challenges of orchestrating such workflows to service real-world big data applications in their article “Orchestrating Big Data Analysis Workflows.” The example they look at in depth is real-time flood modeling, collecting a variety of structured and unstructured data from the cloud, for example, Twitter feeds or national weather forecasts, and the edge, for example, video cameras. Edge processing—for example, filtering out irrelevant information within the video camera—can help to optimize the cost and performance of the global architecture. This data must then be fed into computationally intensive hydrodynamic models to, say, divert vehicular traffic in real time away from a flooded area. The public cloud, of course, excels at providing “near-infinite” resources on a pay-per-use basis during, say, extreme weather events. Also in this issue, David Linthicum, in “Cloud Computing Changes Data Integration Forever…What’s Needed Right Now,” overviews the general challenge of data integration across these complex new hybrid multi-cloud fog architectures, whether for use cases such as big data / analytics, hybrid cloud, centralized integration, end-user empowerment, or other drivers. David points out one of the sea changes today: the need for real-time results. Unfortunately, the greater the volume of data and the less the time available to process it, the greater the challenges around integration and elasticity.

**Security and Privacy.** Security and privacy are complex, nuanced topics, especially where the cloud is concerned. Public cloud services offer numerous advantages in security, such as greater ability to withstand the kinds of large-scale DDoS attacks such as we recently saw with the estimated 1.2 Terabit per second Dyn attack, which greatly surpassed the attack bandwidth of the previously largest attack, and which will no doubt be eclipsed by the next large attack. In addition, distributed architectures such as a cloud can enable secure object storage through various combinations of object fragmentation or sharding, replication, and encryption, an approach dating back decades to projects such as Berkeley’s Oceanstore. On the other hand, any shared and/or connected infrastructure such as the cloud inherently has vulnerabilities that a dedicated, never-connected architecture does not.

Raymond Choo and his colleagues explore the vulnerabilities—and opportunities—of an underexplored area of security and privacy—cloud forensics—in “Evidence and Forensics in the Cloud: Challenges and Future Research Directions.” This fits with the cloud-edge and hybrid multicloud trends and the complex big data challenges explored by David Linthicum and Rajiv Ranjan. Fragments of data left on cloud client devices such as smartphones, cloud storage services, or edge devices may enable law enforcement to investigate and solve physical or cyber crimes that involve the use of cloud. There are many nuances, here, however, beyond technology. For example, cloud-based evidence may have been tampered with by a corrupt insider. There are also legal, regulatory, and jurisdictional concerns, and different jurisdictions won’t always agree where data sovereignty conflicts with prosecutorial needs, much less corporate objectives for cloud providers and device manufacturers and users, especially those who have committed crimes.

**Cost to Value.** Cloud computing was originally viewed by many as a cost take-out play. Moving from costly, inefficient enterprise data centers to a provider exhibiting “economies of scale” could reduce costs. I explored the elements of this argument that are valid and those that are fallacious in my book Cloudonomics, but suffice it to say here that the realization
As Prof. Keith Jeffery has argued, this has implications spanning organizational management structures, communication through supply chains and product distribution to market, customer service, and real-time command and control environments and sense and respond loops. This shift to value is a good thing, because as Ignacio Llorente points out in “The Limits to Cloud Price Reduction,” a comprehensive analysis and projection of cloud price trends, it is overly simplistic to think that public or private cloud prices will reflect Moore’s Law-type exponential price drops, because CPU costs are an increasingly small fraction of the total cost of delivering cloud services. Other elements, such as physical facilities and power costs, are not dropping at Moore’s Law rates; in fact, they are often rising, and, therefore, perversely, the lower the cost of computing hardware goes, the less important the cost of computing hardware is, in terms of aggregate unit cost.

Innovation. Speaking of innovation, we are experiencing something unprecedented in the history of high tech thanks to the cloud. Cloud service providers are exposing not just basic infrastructure and services for the classic use cases of test/dev, load testing, and production, but also the most advanced technologies on the leading edge of innovation either for free or at bite-size pricing. Rather than, say, buying your own5 leading-edge high-performance or quantum computer for tens of millions of dollars, you can acquire such functionality in the cloud via an API for a song. The same with blockchain / distributed ledger capabilities, and various cognitive functions such as AI platforms, speech-to-text, sentiment analysis, facial recognition, image processing and analysis, and the like.

In short, cloud computing has become a catalyst for a plethora of new solutions, new challenges, and new innovation, with every day bringing new developments.***

References

JOE WEINMAN is a frequent global keynoter and author of Cloudonomics and Digital Disciplines. He also serves on the advisory boards of several technology companies. Weinman has a BS in computer science from Cornell University and an MS in computer science from the University of Wisconsin-Madison. He has completed executive education at the International Institute for Management Development in Lausanne. Weinman has been awarded 22 patents. Contact him at joeweinman@gmail.com

MAZIN YOUSIF is the editor in chief of IEEE Cloud Computing. He’s the chief technology officer and vice president of architecture for the Royal Dutch Shell Global account at T-Systems International. He has a PhD in computer engineering from Pennsylvania State University. Contact him at mazin@computer.org