



# Cloud 3.0

The Rise of Big Compute



# Cloud 3.0

## The Rise of Big Compute

# The next \$100B+ opportunity in cloud computing and its first killer app: Deep Learning

As we have entered into 2017, the enterprise software industry is at the inflection point for ubiquitous cloud adoption as part of the \$4 trillion dollar enterprise IT market transformation. The reasons for enterprise IT to adopt cloud are compelling, and many assume this shift has already happened. We are, however, just at the very beginning of this shift and the impact is fundamentally transforming all elements of IT. Industry experts estimate we are in fact only at about 6% enterprise cloud penetration today. <u>1</u> Most businesses have yet to realize the full value of this transformation. It is clear that adoption of cloud in core business activities by enterprise hardware and software IT players strongly correlates with improved business performance and shareholder value.



Driving market growth through embracing cloud (normalized historical 5-year market cap) 2

For instance, if we look at market cap changes over the past 5 years, companies that have fully embraced cloud (Amazon [AMZN], Adobe [ADBE], Salesforce [CRM]) have seen great success. Similarly, companies that have recently embraced cloud (Microsoft (MSFT)) are starting to see strong recent growth and those that have recently embraced cloud (Microsoft (MSFT)) are starting to see strong recent growth and those that have recently embraced cloud (Oracle [ORCL], HP [HPQ], and IBM [IBM]) are struggling to build further business value. To be sure, there are many other drivers for market cap growth in these companies, but level of cloud products, business models, and expertise has been a clear driving factor for success while cloud capability has now become table stakes for every enterprise software business.



During this multi-decade transformative shift to enterprise cloud adoption, several significant macro phases are taking place, together delivering the promise and business value of cloud in the enterprise. The first two foundational phases are SaaS (Cloud 1.0) transforming the delivery and business model, and Big Data (Cloud 2.0) unlocking the value in massive data sets. SaaS and Big Data have each already created over \$100 billion in value in the enterprise. Now, Big Compute, built on these critical foundational layers, is at an inflection point and poised to be the third.

Cloud infrastructure and Infrastructure-as-a-Service (<u>IaaS</u>) has been the critical underlying layer enabling enterprise IT cloud transformation. <u>Amazon Web Services</u>, <u>Microsoft Azure</u>, and <u>Google Cloud Platform</u> have led the way in delivering the technological and infrastructure backbone for SaaS, Big Data, and now Big Compute. As with SaaS and Big Data, Big Compute will drive technological disruption along with a new wave of category-defining companies to pursue the massive opportunity in this new breakout category. Furthermore, the existing category leaders driving billions of dollars of compute heavy workload revenue in the legacy on-premise high performance computing (<u>HPC</u>) market are facing the innovator's dilemma needing to reinvent their entire business to provide effective Big Compute solutions in the space – providing a unique opportunity for the most innovative companies to become category leaders.

Cloud infrastructure and Infrastructure-as-a-Service (<u>laaS</u>) has been the critical underlying layer enabling enterprise IT cloud transformation.



### Cloud 1.0: SaaS

## Transformation of the business model and delivery of enterprise software

SaaS-native companies such as <u>Salesforce</u>, <u>NetSuite</u>, and <u>Workday</u> have built category-defining and leading SaaS solutions for CRM, ERP, and HR, respectively, through business model innovation and frictionless software delivery. These SaaS leaders have leveraged the cloud to disrupt the legacy players historically dominating these enterprise software categories.

Legacy players have also had the opportunity to reinvent their business and implement SaaS for their product line. For example, <u>Adobe</u> reinvented their strategy and successfully transformed their organization into a SaaS leader. Even under the constraints of operating as a public company, Adobe successfully shifted from legacy on-premise software to a cloud SaaS subscription and delivery model.



Adobe's stock surge after embracing cloud SaaS model (Adobe historical stock price 2012-2017)  $\underline{3}$ 

Adobe pivoted the business model for their leading product line from a \$2,500 annual license to a \$50-100/month monthly subscription.4 The product team also transformed the delivery model of their software to a browser-based solution with continuous updates. Long story short – customers were delighted. While sacrificing meaningful short-term revenue, Adobe created significantly more future value. With the SaaS transformation, Adobe expanded their market opportunity by demonstrating broader market appeal and increased willingness to pay due to better customer value, along with the critical ability to capture that value over the long-term.



## Cloud 2.0: Big Data

MGI has compiled a heat map of

the value potential of using big data

## Scaling and connectivity enablement for the full-stack data layer

As cloud infrastructure matured to support the rapid growth of SaaS companies, Infrastructure-as-a-Service (IaaS) provided a fundamental layer of scalable and ubiquitous storage and databases. This broad infrastructure capability enabled dramatic innovation on the data layer and is driving the <u>Big Data</u> transformation.

Leaders in the Big Data category such as <u>Splunk</u>, <u>Cloudera</u>, and <u>Palantir</u> have transformed our ability to extract insights from the past by scaling, manipulating, and connecting enterprise data sets. At a basic level, Splunk provides the ability to drive insights from large and disparate enterprise application logs, Cloudera took this to the next level with the ability to deploy distributed storage and processing clusters for the enterprise, and finally Palantir has developed proprietary stacks to connect siloed and unstructured data sets in the public and private sectors. Fundamentally, Big Data is a horizontal capability that can be applied across the enterprise to distill insights from existing data sets.

Top quintile

2nd quintile

(highest potential)

4th quintile

Bottom quintile

(lowest potential)



1 See appendix for detailed definitions and metrics used for each of the criteria. SOURCE: McKinsey Global Institute analysis

Heat map showing Big Data opportunities across all major industries (For a more detailed analysis of Big Data, the comprehensive McKinsey Global Institute publication is an excellent resource: "Big data: The next frontier for innovation, competition, and productivity")<u>5</u>



## **Cloud 3.0: Big Compute**

# Scaling and workload enablement for the full-stack compute layer

Big Compute is the next transformational shift for enterprise cloud computing. Just like Big Data removed constraints on data and transformed major enterprise software categories, Big Compute eliminates constraints on compute hardware and provides the ability to scale computational workloads seamlessly on workload-optimized infrastructure configurations without sacrificing performance.

#### "Supercomputer in your pocket" fallacy

<u>Moore's Law</u> and IaaS cloud compute have resulted in the perception that "everyone has a supercomputer in their pocket" and ubiquitous compute power is universally accessible by all. However, while compute capabilities have been growing at an exponential rate and have become much more accessible, computing requirements for applications have also grown dramatically to meet the insatiable demand for complex, next-generation algorithms. In fact, the iPhone "supercomputer in your pocket" is really a supercomputer from 25+ years ago,<u>6</u> and in that same time the computing requirements for the software ecosystem have grown at a rate on par with or greater than Moore's Law.

## Many critical enterprise problems and innovations are still compute-bound

Despite popular belief, many of the world's most challenging software problems are still compute-bound. Since the dawn of enterprise software, compute limitations have highly constrained both the applications and scope of the work due to the multi-million dollar expense and expertise required for implementation of supercomputers and high performance computing systems in the enterprise.

High performance computing is used heavily in most major Fortune 500 companies, including verticals such as aerospace, automotive, energy, life sciences, and semiconductors. In almost all of these verticals, companies are faced with significant constraints in compute capacity. For example, in aerospace design, large-scale problems have complex computational fluid dynamics (<u>CFD</u>) algorithms that can take days or weeks to solve on large high performance computing clusters. In automotive crash testing simulation analysis, scaling finite element analysis (<u>FEA</u>) physics to millions of components and interactions creates a massive computational problem. And in life sciences, the massive computational requirements of molecular dynamics (<u>MD</u>) simulation restricts computational drug discovery, one of the largest opportunities for disruption for pharmaceutical companies.

Despite popular belief, many of the world's most challenging software problems are still compute-bound.



Cloud computing innovations such as SaaS, Big Data, and IaaS are the building blocks that enable the Big Compute category. A comprehensive Big Compute stack now enables frictionless scaling, application-centric compute hardware specialization, and performance-optimized workloads in a seamless way for both software developers and end-users. Specifically, Big Compute transforms a broad set of full-stack software services on top of specialty hardware into a software-defined layer, which enables programmatic high performance computing capabilities at your fingertips, or more likely, as back-end function evaluations part of software you touch every day. This approach flips the legacy approach of a one-size-fits-all compute cluster for a broad set of applications upside down, into an a-la-carte solution where the compute workloads are profiled and deployed onto an optimal specialized hardware solution for tailored to the workload. Big Compute turns a homogenous cluster full of performance and cost trade-offs, into a heterogenous solution with no compromises.

#### Hardware fragmentation and specialization enabling Big Compute

Big Compute is a horizontal technology that enables enterprises to scale complex algorithms and workloads on flexible, specialized compute infrastructure. Specialized compute infrastructure is a critical element for success, since there is a multiple magnitude performance gap between commodity infrastructure and specialty servers. Many components of the servers can be specialized, such as processors, networking, memory, storage, etc. Big Compute is democratizing the access to this level of specialty hardware from a software-defined point of view, thus enabling the ability to spin up a \$10 million high performance computing cluster for just a few thousands of dollars per hour and integrate this capability into any enterprise software application while, most critically, still meeting the cutting-edge performance expectations for high-end applications.

Within the data center there are many areas of tuning and optimization from the physical racks, to the networking, and down to the specific server components. For example, there is already a broad spectrum of processor capabilities that is fragmenting at an increasing rate. The specialized capabilities of these processors is critical for specific Big Compute workloads that need to be mapped to the right architectures, depending on their algorithms and implementation.



Example mapping of increasing fragmentation in the spectrum of processor capabilities



Big Compute enables users, software applications, and algorithms to seamlessly take advantage of these capabilities without the need to understand or deploy the middleware and libraries required. This accelerates the adoption cycle and deployment of specialized hardware capabilities in IaaS and data centers, but more importantly enables a dramatic new category of Big Compute enterprise software applications and tools.

Fundamentally, Big Compute provides the layer of software-defined computing that allows a more application-centric compute approach in the enterprise, bringing seamless optimization and scalability of the stack, while ultimately democratizing the Big Compute capability to developers and end-users. This capability is resulting in a dramatic transformation within the enterprise, from multi-disciplinary teams of highly-trained experts in hardware, middleware, and the domain-specific software algorithms supported by large investments in computing resources, to a seamless service which allows the average enterprise user to focus on the application layer while relying on a highly complex but seamless Big Compute layer for execution. This transformation allows organizations to move up a level of abstraction, from formulating workloads to meet the underlying hardware capabilities, to solving the higher level problems without any compute constraints.

The rise of Big Compute is not only a fundamental new capability for all enterprise software applications, but it provides the opportunity to shift from backward-looking analysis (as is typical for Big Data), to forward-looking analytics and simulations to drive predictive insights and predictions of the future states. In fact, Big Compute can provide simulations that generate data sets which leverage Big Data tools to evaluate the potential quality of a future state a simulation generated. An asynchronous process will work for the largest problems, but this will move toward real-time where the Big Compute stack can process faster than the required response time. For example, in Deep Learning training (asynchronous) and inference (real-time), both running on a Big Compute stack of capabilities local on the edge, or centralized in the cloud if latencies allow.

#### Deep Learning, the first Big Compute killer app

The first real breakout in Big Compute has been the category of <u>deep</u> <u>learning</u> software. Deep learning is a subset of machine learning algorithms that requires massively parallel computations and is thus a great fit for GPU hardware, historically designed for parallel graphics thread processing. Deep learning algorithms have been around for decades, but their recent success is attributable to highly specialized GPU hardware tailored for this workload along with the ability to spin up a turnkey full-stack solution on which the software algorithms operate without needing to be an expert on the implementation, parallelization, and orchestration of the algorithms on the specialty hardware.

The first real breakout in Big Compute has been the category of deep learning software.



Deep learning tools are now recognized as a must-have horizontal solution for the enterprise. These tools provide analytics, insights, and intelligence capabilities that were previously unattainable without massive investments and expertise across the stack, but are now universally accessible with a Big Compute stack of specialized hardware, middleware, and integrated algorithms. These machine learning algorithms are now universally accessible on the latest generation of specialized scalable GPUs and can quickly be implemented as a solution to complement existing software capabilities with algorithmic intelligence across the enterprise, leveling up internal capabilities. As the leader in the GPU market, <u>NVIDIA</u> has been "selling pickaxes during a gold rush" and seen tremendous growth and adoption fueled by the Big Compute layer making these specialized hardware capabilities quickly and easily accessible to anyone looking to use deep learning tools.



Dramatic growth and success of NVIDIA, leaders in GPU chipsets (NVIDIA Historical Stock Price 2012 -2017)  $\underline{Z}$ 

Google (<u>Alphabet</u>) has famously seen deep learning as such a fundamental capability to remain the leader in search. It has been considered so instrumental to their research and development efforts that they have implemented the <u>TensorFlow</u> algorithm broadly across their broad portfolio of endeavors unrelated to search. So instrumental, that they have even developed their own custom Tensor Processing Unit (TPU) chip to accelerate the performance to levels not attainable with standard chipsets. This clearly illustrates the critical importance of specialized hardware in the Big Compute stack for optimal algorithmic performance to drive new capabilities and innovation.



#### Growing Use of Deep Learning at Google



Growing use of Deep Learning at Google 8

At the start of 2017, deep learning is likely at the peak of its hype cycle, but there is clear fundamental long-term value being created. At a minimum, deep learning is creating value as a back-end process to optimize highly complex situations in which the datasets are too large for deterministic algorithms. For instance, it is likely that Google recouped their \$500 million acquisition of DeepMind through a single implementation of their machine learning tools to improve their data center cooling.<sup>9</sup> It is also likely that deep learning will provide new capabilities and accelerate further innovation in autonomous driving, drug discovery, medical imaging, robotics, and many other exciting areas. Deep learning is just a single software application category powered and enabled by the Big Compute stack.

#### The emerging Big Compute stack

The Big Compute stack is the critical enabling layer to bring Big Compute capabilities to any software application and handle the management overhead complexity to seamlessly execute workloads on the broad selection of specialized infrastructure. There are entirely new application categories that have yet to be unlocked with the Big Compute stack, and analogous to deep learning, have algorithms which have historically been compute-bound requiring specialized hardware, complex middleware, and comprehensive tuning and optimization of the stack to become commercially viable to the enterprise.

Big Compute has the potential to bring a magnitude larger fundamental impact across the entire economic landscape beyond that which Big Data already has. Not only will Big Compute bring the computation and algorithmic capabilities to amplify the power and impact of the Big Data analytics tools, but Big Compute will also drive an entirely new category of capabilities that were not possible before.



The success of deep learning as the first Big Compute killer app signals an inflection toward exponentially greater use of large-scale, specialized computing. As with Big Data, Big Compute stack capabilities will be a must-have in the enterprise.

Here at Rescale, we believe Big Compute is the foundation upon which new future innovations will be built – driven by the algorithms and software that have historically been highly constrained by deep expertise and significant upfront investment, but have now been unleashed with the Big Compute stack. We are building the universal platform to bring Big Compute capabilities to the enterprise – a critical capability we believe will be necessary to drive most major new innovations in the leading Fortune 500 enterprises across industry verticals. From space exploration to electric vehicles to computational drug discovery, we are working with the leaders of industries which are implementing the Big Compute stack as a competitive advantage in their respective verticals.

The rise of Big Compute will unlock a broad spectrum of innovation and incredible business value in the enterprise previously gated by specialized hardware capability, accessibility, and scalability. While late adopters of Big Compute capabilities will face tougher competition for undifferentiated capabilities, early leaders are finding the unique opportunity to leverage the existing internal expertise and core competencies of their organization and supercharge these with Big Compute capabilities to create the future innovations that will transform our world.

The rise of Big Compute will unlock a broad spectrum of innovation and incredible business value in the enterprise previously gated by specialized hardware capability, accessibility, and scalability.

Author:

Joris Poort CEO, Rescale Inc.



