THE JOURNEY TO COMPOSABLE INFRASTRUCTURE

CISCO TAKES AN EVOLUTIONARY APPROACH WITH UCS M-SERIES & C3260 COMBINED WITH UCS MANAGEMENT FRAMEWORK

EXECUTIVE SUMMARY

Today's digital economy is driving IT as a focal point for competitive differentiation. But siloed, traditional IT architectures often are not equipped to provide the flexibility that today's business-critical applications require. In addition, many IT organizations are transitioning to a DevOps methodology for application development which involves continuous integration, automation, and measurement throughout the process. DevOps requires an agile infrastructure to keep up with the rapidly evolving application demands and often uses an "infrastructure as code" approach to optimize allocation of resources. To address these needs, a new category of solutions is emerging called composable infrastructure.

Moor Insights & Strategy defines composable infrastructure as fluid pools of resources that can be configured dynamically through software and the application of policy to optimize application performance and drive efficient use of infrastructure. Composable infrastructure has the potential to offer the best application performance possible, reduce underutilization / overprovisioning of infrastructure, and create a more agile, cost-effective method to provision / re-provision applications. However, composable infrastructure requires ongoing dynamic configuration and reconfiguration of a large number of infrastructure assets—which has the potential to increase IT complexity.

The portion of the market that requires truly composable infrastructure solutions is still in its infancy. There are a number of infrastructure vendors coming to market with a variety of approaches. It is important for IT organizations who are evaluating composable infrastructure solutions to look at all aspects of a solution to make sure it can scale as business requirements evolve. Four key tenets to consider include…

1. Disaggregated programmable infrastructure resources
2. Composition and orchestration capabilities
3. Robust management and automation framework
4. An application-centric solution approach

Copyright ©2015 Moor Insights & Strategy

Cisco is taking a pragmatic approach to composable infrastructure with the introduction of their latest UCS M-Series and C3260 products which include disaggregated compute, network, and storage subsystems. The “secret sauce” to Cisco’s composable infrastructure will happen via Cisco UCS Manager which handles the composition, orchestration, automation, and management of resource pools driven by a user-defined set of application requirements. IT organizations evaluating composable infrastructure solutions should add Cisco to their shortlist of vendors for consideration.

New Business Models Require More Agile IT

Major technology drivers such as Big Data, the Internet of Things, mobility, and cloud-based services are shaping the IT landscape of the future. These megatrends are driving line of business (LOB) executives to deliver new products and services more quickly, improve operational efficiencies, and increase bottom line results. To address the changing environment driven by these trends, IT is now a business differentiator that rapidly delivers new applications to help the business achieve their goals.

IT infrastructure models that are rigid or inflexible will not meet the needs of a growing business today. Each environment is unique and has a specific set of requirements. In an economy where IT is core to the business, an application-centric development model is required to deploy resources that can adjust quickly to changes in end user demands, account for traffic spikes, or rapidly deliver new services. In addition, many organizations are adopting DevOps as a software development methodology for these next-generation applications. DevOps involves continuous integration, automation, and measurement, and it requires an agile infrastructure to keep up with the rapidly evolving environments of these applications. Provisioning and re-provisioning of resources must happen “on demand”, in a matter of seconds / minutes, not days / weeks.

All of these business challenges and application demands have the potential to greatly increase IT complexity. But IT organizations are demanding infrastructure that is more flexible, efficient, and easier to manage to solve today’s problems.

Composable Infrastructure for Changing Business Needs

Many IT organizations today are constrained by complex, siloed infrastructure that was originally designed to help reduce costs for traditional applications and architectures. This infrastructure is not equipped to drive the degree of flexibility and efficiency required for next-generation applications that are central to business success.

Copyright ©2015 Moor Insights & Strategy

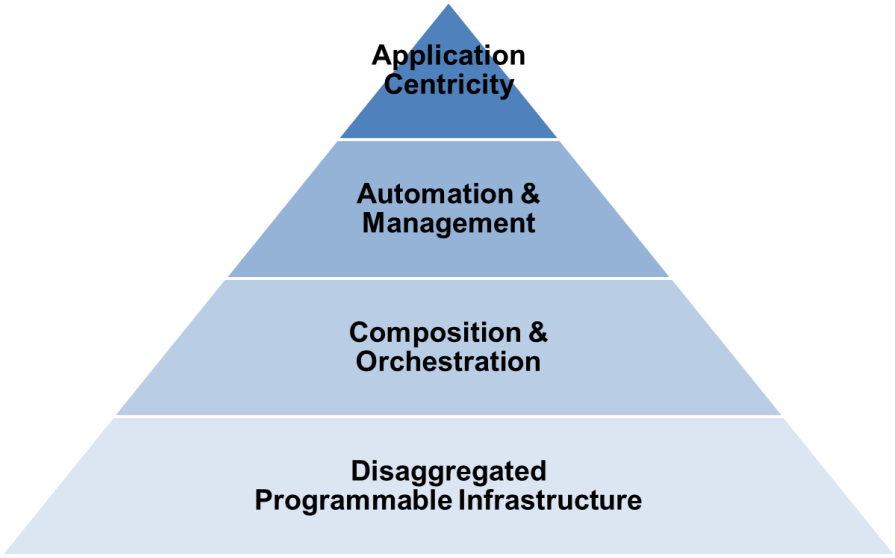
A new category of datacenter infrastructure solutions called ““composable infrastructure” is emerging to address the gap in capabilities from traditional servers. **Composable infrastructure** is designed to provide **fluid pools of resources** that can be **configured dynamically through software and the application of policy** to optimize application performance and drive efficient use of infrastructure**.**

KEY REQUIREMENTS FOR COMPOSABLE INFRASTRUCTURE

Composable infrastructure offers the potential to provide the best application performance possible, reduce underutilization / overprovisioning of infrastructure, and create a more agile, cost-effective method to provision / re-provision applications. However, composable infrastructure requires ongoing, dynamic configuration and reconfiguration of a large number of infrastructure assets—which has the potential to increase IT complexity. A robust systems management framework is required to help address this potential complexity.

Figure 1 describes the key attributes IT organizations should consider when evaluating composable infrastructure solutions. While companies can begin to plan for and implement the different attributes across their infrastructure, starting at the base and working upwards will yield the most efficient results.

FIGURE 1: HIERARCHY FOR COMPOSABLE INFRASTRUCTURE



Source: Moor Insights & Strategy

Copyright ©2015 Moor Insights & Strategy

- **Disaggregated Programmable Infrastructure**: Infrastructure resources (such as power and cooling, network I/O, storage, and compute) that are disaggregated from the other subsystems in the server can be optimally scaled and allocated to meet the needs of a given application. Disaggregated resources allow for efficiency in scale and give IT organizations the ability to upgrade each subsystem as new technologies become available without having to replace the entire framework. Programmability of the infrastructure with a common API across subsystems using an object model is a critical element that makes it possible for the layers above to compose, allocate, and manage these resources.
- **Composition & Orchestration**: Robust composition and orchestration policies are critical to take advantage of disaggregated resources. Composition and orchestration allow higher-level application frameworks to manage pools of hardware resources dynamically without interrupting ongoing service operations. Composition and orchestration also enable different devices and applications to work together seamlessly in a coordinated way. Composition occurs at the subsystem level through the use of profiles, whereas orchestration occurs across infrastructure domains. Composition and orchestration capabilities must enable consistent application of all policies across infrastructure resources.
- **Management & Automation**: A comprehensive management framework is required to automate discovery of resources in the infrastructure pool and implement policies to standardize the provisioning of those resources. In addition, a consistent API with broad support for other systems management tools is required to provide ease of integration and to extend monitoring, provisioning, automation, and orchestration across the datacenter.
- **Application-Centricity**: Through an application-centric approach, IT can define workflow, resource requirements, and policies in advance to drive “infrastructure as code” and keep applications optimized dynamically. Condition-based actions and action synchronization can eliminate guesswork and manual processes—preventing operational slowdowns and preventing increases in long-term operational costs. When application conditions drive actions, platforms can run at peak performance, and applications / services can be delivered quickly and more securely.

Many of the large server infrastructure providers have developed plans to bring composable infrastructure products to market over the next several years. Implementation plans, vendor approaches, and production timelines vary widely across vendors. Composable platforms should not be confused with converged systems which are defined as pre-integrated configurations that combine server, storage, and

Copyright ©2015 Moor Insights & Strategy

networking with a unified management framework. In many cases, capabilities are added on top of a converged system infrastructure to create a composable platform, but not all converged systems have composable infrastructure capabilities.

In June 2015, HP Enterprise laid out their multiyear plan called Project Synergy to build the ecosystem and groundwork for a composable infrastructure solution in the future including an open API and partner / developer program.

Intel has developed the Intel Rack Scale Architecture (RSA) which is focused on optimizing fluid pools of resources within a rack. Intel is focused on enabling partners to bring to market solutions based on RSA with software, hardware, and management reference architectures (currently version 1.0 available) and partner development support. Intel's implementation timeline includes shared power, cooling, and rack management today; rack-level fabric and shared boot / standard management by the end of 2015; and pooled compute, pooled storage, and pooled memory optimizations in future RSA revisions. Intel partners developing platforms based on RSA include Dell, Ericsson, Quanta, and others.

In 2015, Cisco introduced its first set of Cisco UCS composable infrastructure solutions. Cisco's approach to **composable** infrastructure builds on their **converged** infrastructure portfolio. It does so by taking advantage of the combined compute, network, and storage resources, by decomposing the infrastructure resources, and by leveraging the UCS management framework. The underlying hardware used in the new UCS composable infrastructure platforms is an extension of the abstraction that UCS management provides.

Cisco's Approach to Composable Infrastructure

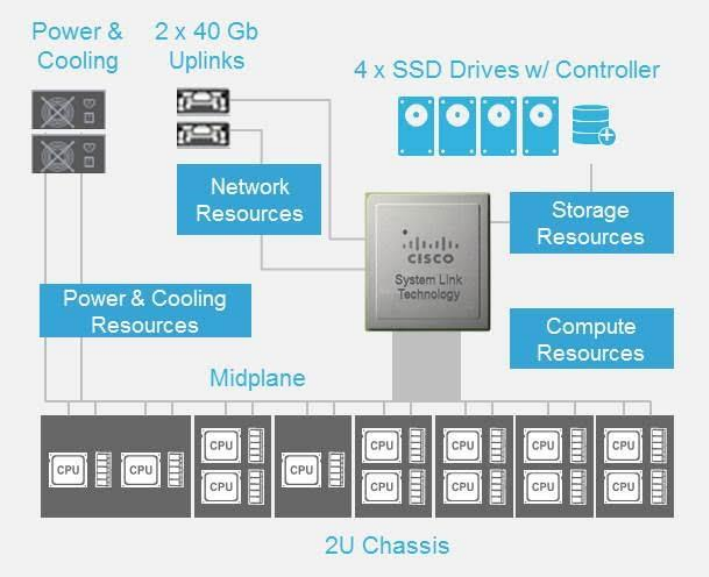
Cisco currently has two primary offerings in their composable infrastructure portfolio to meet the needs of a variety of modern applications. The **UCS M-Series** product is designed for compute-intensive workloads such as scale-out applications, grid, EDA, online gaming, genomic applications, web serving, memcached, and MaaS (metal as a service). The recently announced **UCS C3260** is targeted at data-centric workloads such as Big Data analytics (MapR, Cloudera, *etc*.), content delivery, Microsoft Storage Spaces, and software-defined storage environments (CEPH, Scality, *etc*.). Each workload varies significantly in terms of specific resource requirements which makes composable infrastructure a good potential fit to drive intelligent resource allocation.

Copyright ©2015 Moor Insights & Strategy

DISAGGREGATED INFRASTRUCTURE: UCS M-SERIES & C3260

Figure 2 provides an overview of the disaggregated programmable infrastructure in the Cisco UCS M-Series. Each M-Series chassis includes up to 16 independent servers including Intel Xeon E3 v3 processor and Intel Xeon E5 v3 processor configurations with a variety of core-count and memory footprint options. The compute cartridges are connected via an in-chassis PCIe midplane to the Cisco VIC 1300 Series with Cisco System Link Technology which extends the PCIe Fabric within the chassis to provide access to the local shared I/O resources. Via the Cisco System Link, each server cartridge has access to a pool of 2x 40 Gb network resources and a pool of 4x SSD storage resources which can be distributed as needed and scaled independently to the servers within the chassis. These shared I/O resources, along with shared power and cooling, spread subsystem costs across all 16 servers in the chassis and have the potential to improve system efficiencies.

FIGURE 2: CISCO UCS M-SERIES DISAGGREGATED INFRASTRUCTURE



Source: Cisco

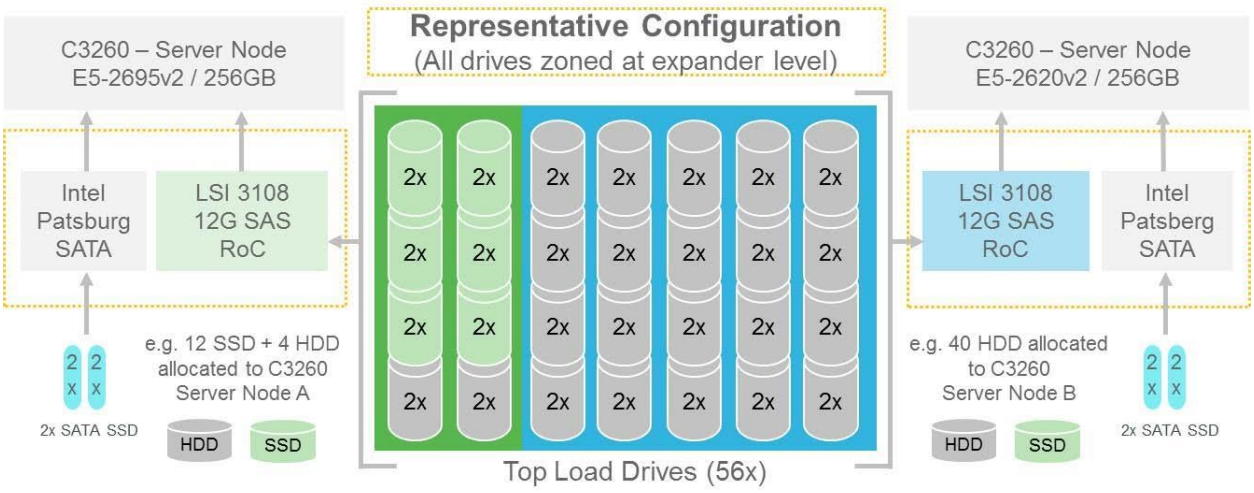
The Cisco 1300 Series VIC with System Link Technology is the core technology that provides flexible resource sharing and configuration for the Cisco UCS M-Series. System Link Technology presents the vNIC (virtual network interfaces) and the sNIC (virtual storage controller) to the operating system as a dedicated PCIe device for that server. In addition to presenting these PCIe devices to the operating system, the System Link Technology provides a method for mapping the vNIC to a specific uplink

Copyright ©2015 Moor Insights & Strategy

port from the chassis. For the sNIC, the System Link Technology provides a mapping of virtual drive resources on a chassis storage controller drive group to a specific server as a local resource. The modularity of UCS M-Series provides the same potential for system efficiency and utilization gains through its ability to share subsystem resources. In addition, this modularity gives IT organizations the ability to upgrade the different subsystems within the server on separate lifecycles, providing the potential for a more rapid adoption rate of new compute components as well as investment protection for longer-lived components.

The Cisco UCS C3260 is a capacity-intensive architecture with up to 56x drives (supports both SSD and HDD options) and 2x server nodes per chassis. UCS C3260 uses up to two system I/O controllers to create pools of storage which can be allocated dynamically via the management controller to each server node. The flexibility to define and allocate storage pools for each server node is critical for storage-intensive workloads, as the specific capacity and technology requirements (SSD vs. HDD) vary greatly across workloads and storage tiers. Policy-driven local storage allocation is a differentiated capability for Cisco compared to many of the competitive composable infrastructure offerings.

FIGURE 3: CISCO UCS C3260 DISAGGREGATED INFRASTRUCTURE



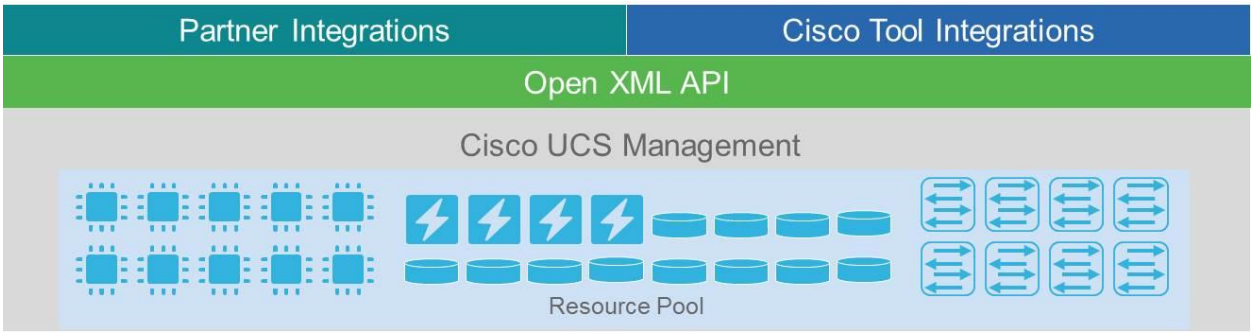
Source: Cisco

Copyright ©2015 Moor Insights & Strategy

CISCO UCS MANAGEMENT SOFTWARE: "THE SECRET SAUCE" FOR COMPOSABLE INFRASTRUCTURE

Programmable disaggregated infrastructure is only one piece of the puzzle to enable a composable infrastructure solution. Cisco's composable infrastructure capabilities are made possible through enhancements to their Cisco UCS management infrastructure—the management framework used for more than 6 years across the UCS product portfolio. Cisco UCS Manager allows IT organizations to catalog the infrastructure assets for composability, program each infrastructure element, manage each element at a granular level, and develop domains / subdomains to provide logical groupings of systems with similar configuration characteristics. Cisco UCS Central Software can provide centralized coordination for up to 6,000 servers and tens-of-thousands of infrastructure elements. Figure 4 illustrates UCS management's role in a composable infrastructure environment.

FIGURE 4: CISCO UCS MANAGEMENT & COMPOSABLE INFRASTRUCTURE



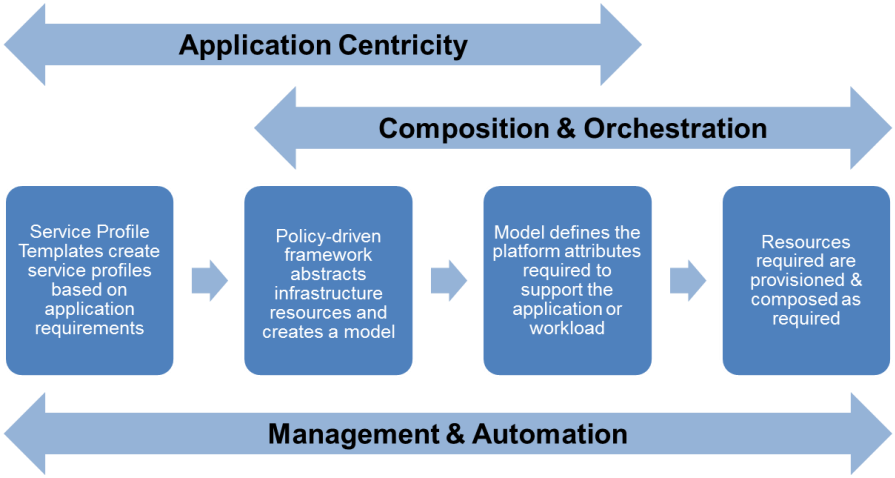
Source: Cisco

Based on an application-driven set of criteria, Cisco UCS Manager[1] coordinates the best platform and optimized block of resources to support each application and workload. Figure 5 describes the process that allows the UCS management software to "direct" the applications to the appropriate compute resources.

[1] Composable Infrastructure capabilities via UCS Manager are currently supported on the Cisco UCS M-Series. The Cisco UCS C3260 will be supported in a future release of UCS Manager scheduled for CY 2016.

Copyright ©2015 Moor Insights & Strategy

FIGURE 5: CISCO UCS MANAGEMENT PROCESS FLOW FOR COMPOSABLE INFRASTRUCTURE



Source: Moor Insights & Strategy

Cisco leverages an API framework to enable IT organizations to use their existing tools as well as new cloud-native tools as they transition to DevOps methodologies and implement "infrastructure as code". Ecosystem partnerships and enablement includes Microsoft System Center, VMWare (vRealize and vCenter), BMC, HPE, IBM, Openstack, and others. Cisco does not yet support tools like Ansible, Chef, Puppet, or Salt directly through UCS management, but Cisco does support these tools through Cisco ACI, UCS Director, or via other third party partner tools.

POTENTIAL DIRECTIONS FOR CISCO COMPOSABLE INFRASTRUCTURE

Cisco is leveraging their Cisco UCS management software as the foundation for composable infrastructure. Since all of the Cisco UCS products use Cisco UCS management software, MI&S expects Cisco will expand its composable infrastructure product portfolio by designing new UCS products with disaggregated resources to take advantage of these proven management capabilities.

Cisco's composable infrastructure framework uses a technology-agnostic approach to policy management. MI&S believes as more components in Cisco's infrastructure become disaggregated in future server solutions (remote storage resources, memory, *etc*.) these resources could be added to the pools of infrastructure resources and intelligently allocated based on application needs.

Copyright ©2015 Moor Insights & Strategy

CALL TO ACTION

Moor Insights & Strategy believes that IT is on the cusp of a major datacenter architecture transition. This transition is driven by 24 × 7 global business reach, dramatically increased use of Big Data analytics, and pushing sensors and intelligence into our physical world in the form of the Internet of Things. It is impossible to predict exact technology directions even in a three-to-five year timeframe, but improved efficiencies driven by the public cloud are helping form a high-level framework for the future of IT operations.

Composable infrastructure has emerged as a category of solutions to meet the needs of next-generation applications. Key potential benefits include increased application performance and greater provisioning agility. However, not all environments are well-suited to composable infrastructure today. Using composable infrastructure for traditional IT environments could result in increased complexity without the benefit of lower costs or greater efficiencies.

IT organizations that have next-generation applications with dynamic resource needs that are core to their business success—such as Big Data, software-defined storage, and cloud-based services—should consider evaluating composable infrastructure solutions as a potential fit for their environments. IT organizations must have a detailed understanding of the performance characteristics of their applications to ensure the resources allocated are right-sized appropriately. While many vendors claim to have composable infrastructure strategies and solutions, it is important to consider each vendor's capabilities and roadmaps across all layers in the composable infrastructure hierarchy (see Figure 1) when comparing solutions. Choosing the right management software, and not just the hardware roadmap, is essential to this decision, because it will establish the framework for automation, orchestration, and operations management for many years.

Cisco's composable infrastructure strategy addresses the layers in the composable infrastructure hierarchy with a combination of the UCS M-Series and C3260 products and robust capabilities of UCS management software. IT organizations evaluating composable infrastructure solutions should add Cisco to their shortlist of vendors for consideration.

Copyright ©2015 Moor Insights & Strategy

IMPORTANT INFORMATION ABOUT THIS PAPER

AUTHOR

Gina Longoria, Senior Analyst at Moor Insights & Strategy

REVIEW / PUBLISHER

Patrick Moorhead, President and Principal Analyst at Moor Insights & Strategy

EDITOR / DESIGN

Scott McCutcheon, Director of Research at Moor Insights & Strategy

INQUIRIES

Please contact us here if you would like to discuss this report, and Moor Insights & Strategy will promptly respond.

CITATIONS

This note or paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING

This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

DISCLOSURES

Cisco is a research client of Moor Insights & Strategy, and this paper was commissioned by Cisco. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.

©2015 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.