Strategizing Storage Deployment in Compute Centric Public Clouds: Balancing Capacity, Congestion, and Cost for Evolving Workloads

Anurag Reddy[1], Anil Naik[2], Sandeep Reddy[3]

[1]Director Capacity Engineering
Email: *anurag_reddy[at]berkeley.edu*

[2]Product Lead, Telecom & AR/VR, UC Berkeley, CA
Email: *anilgnaik9[at]gmail.com*

[3]Senior Program Manager, Consumer Technology, University of Rochester, NY
Email: *sandeepreddy4488[at]gmail.com*

Abstract: ***The paper explores the intricacies of deploying storage services in an existing compute-only environment within public clouds, considering factors such as capacity expansion, traffic congestion risks, and cost implications. The focus is on aligning data center architecture to support evolving workloads including real-time high bandwidth, low latency applications like AR/VR gaming, immersive streaming, health care (Telemedicine), military applications & monitoring systems while meeting performance, latency, reliability, and cost standards. The introduction emphasizes the importance of adapting data center architecture to seamlessly accommodate diverse workloads in public clouds. The paper outlines options for deploying storage services, aiming for reasonable and predictable lead times for capacity expansions. Key variables, including long-term forecasting, are introduced, considering high-density compute servers and forecasts for both compute and storage servers for the next year. Four design scenarios are presented, each with its benefits and risks. Design 1 proposes high storage density mixed-racks, highlighting quick deployment with no additional costs but raises concerns about potential traffic congestion. Design 2 suggests low storage density mixed-racks to handle recovery traffic but incurs incremental hardware costs and longer deployment times. Designs 3 and 4 focus on storage servers-only racks and spines, offering advantages like quick deployment and isolation from compute resources during rebuilds but come with higher failure domain risks. In conclusion, the paper underscores the need for aligning network architecture with long-term forecasts for scalability and flexibility. It emphasizes optimizing designs to balance performance, reliability, and cost considerations. The presented scenarios offer a comprehensive exploration of challenges and opportunities in deploying storage services within a compute-centric environment, providing insights for data center strategists and architects.***

Keywords: Data center architecture, compute-only environment, and storage services deployment with a focus on capacity expansion, traffic congestion risks, low latency, cost implications in public clouds

1. Introduction

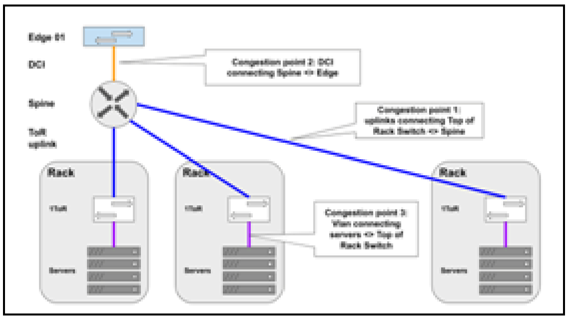
In the exploration detailed within this paper, we undertake a comprehensive analysis of diverse strategies for deploying storage services within an existing compute-only environment. The significance of this discussion lies in its dual focus:

a) Ensuring Reasonable and Predictable Lead Times for Capacity Expansions

A pivotal aspect of this examination revolves around the strategic utilization of Rack Integration for the incorporation of new racks, thereby minimizing the necessity for retrofitting existing infrastructure. This nuanced approach not only facilitates the realization of reasonable and predictable lead times for capacity expansions but also empowers Data Center Strategy teams to meticulously plan and secure the optimal layout for new data centers. This becomes particularly imperative in accommodating the growing coexistence of storage and compute within a singular spatial configuration at scale.

b) Striking a Delicate Balance in Traffic Congestion Mitigation

The paper delves into the intricacies of mitigating the risk of traffic congestion, particularly at crucial points prone to overutilization. This includes a keen examination of uplinks connecting Top of Rack Switch <> Spine, DCI connecting Spine <> Edge, and vlan connecting server <> Top of Rack Switch during the rebuild event. The analysis goes beyond identifying congestion points to encompass a thoughtful consideration of the associated cost implications linked with the adoption of various design choices in this context.



This paper, therefore, serves as a profound exploration of the multifaceted considerations and challenges inherent in deploying storage services within an environment predominantly oriented towards compute functionality. The discussion is underpinned by a strategic vision that seeks to harmonize the architectural decisions with the evolving landscape of data center demands, thereby providing invaluable insights for Data Center Strategy teams and architects alike.

2. Architectural Dynamics

Defining the architectural variables crucial for this study involves a meticulous consideration of the long-term forecast, a key determinant in estimating the hardware scale required for both compute and storage. This entails not only grasping the magnitude of the scale but also ensuring that the network architecture is capable of supporting dynamic changes in demand. To illustrate these considerations within our problem statement, we make certain assumptions:

a) Existing Network Configuration

The current network is characterized by high-density compute servers, with approximately 20 servers per rack adhering to a specified layout.

b) Storage Placeholder

Anticipating future demands, we designate approximately 1,000 storage servers as a placeholder for the upcoming year. Each storage server occupies 1RU (Rack Unit) and has a power consumption of 400W.

c) Next 1-Year Compute Forecast (Q2 2024 to Q4 2024)

Envisaging a growth trajectory, the forecast anticipates the addition of 10,000 compute servers within the next year. This projection translates into the need for 625 new racks or Top of Racks (ToRs), encompassing 425 new racks dedicated to scalability.

d) Edge Compute Needs

The expansion plan incorporates the deployment of 55 new Spines in 2024 which aligns is reasonable assumption of growth for public clouds, strategically aligned with the requirements of Edge Compute functionality.

1Yr Compute Forecast	10,000 Compute servers
1YR ToR / Racks	625 new racks / ToRs (incl. 425 new racks / ToRs to s)
1Yr Spine	55 new Spines

In essence, these variables, spanning from the existing network configuration to future forecasts and expansion plans, form the foundation for evaluating and optimizing the architecture to accommodate the evolving demands of both compute and storage within the data center environment.

3. Design Scenarios

a) Design 1: High-Density Storage and Compute Integration in Mixed Racks

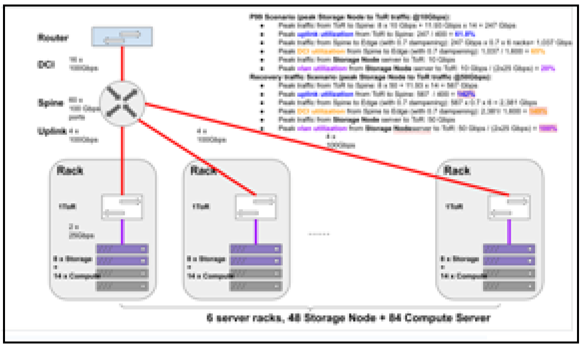
In this innovative configuration, Design 1 introduces a high-storage-density and compute-mixed rack setup, consisting of 8 storage servers and 14 compute servers per rack. This allows for a maximum deployment of 48 Storage Servers and 84 Compute Servers per spine.

Benefits:

1) *Cost Optimization:* The design offers significant cost advantages by leveraging existing deployment processes, eliminating the need for additional expenditures. The network rack initially designated for compute servers seamlessly accommodates all storage servers, minimizing costs associated with additional network equipment.
2) *Efficient Scalability:* Deployment of 1,000 storage servers requires only 125 racks (1,000 / 8 storage servers per rack) in the data center. This scalable approach perfectly aligns with plans to add 400 compute racks in the upcoming year. Importantly, this scalability is achieved without the need to retrofit existing racks, ensuring a rapid and efficient deployment process.
3) *Resource Utilization:* The design maximizes resource utilization by integrating storage and compute in a balanced manner. By avoiding the need for retrofitting, the deployment process is streamlined, leading to optimal utilization of both storage and compute resources.

Risks: Traffic Congestion Dynamics: Despite the efficient design, there is a potential risk of traffic congestion during rebuild events. Under normal circumstances, ToR uplink and DCI utilization remain below 100%. However, during recovery with peak storage node to ToR traffic reaching 50Gbps (assuming peak NIC capacity), there is a possibility of surpassing these thresholds. This scenario could lead to congestion across ToR uplinks, DCI, and VLAN, necessitating careful consideration and mitigation strategies during the planning phase. A detailed analysis of traffic patterns and potential congestion points is essential to address this risk effectively.

This approach presents an innovative solution with high-density storage and compute integration, optimizing costs and facilitating efficient scalability without the need for retrofitting. While the design maximizes resource utilization, careful consideration and mitigation strategies are essential to address potential traffic congestion risks during rebuild events. Overall, a balanced approach and thorough planning are crucial to harness the benefits of this configuration effectively.



Paper ID: SR24209005017 DOI: https://dx.doi.org/10.21275/SR24209005017

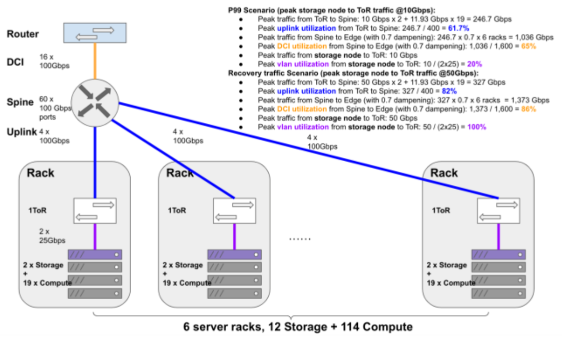
b) *Design 2: Low Storage Density Mixed-Rack Configuration*

This approach adopts a low storage density mixed-rack configuration, featuring 2 storage nodes and 19 compute nodes per rack, allowing for a maximum deployment of 12 storage nodes and 114 compute nodes per spine. This design emphasizes recovery traffic management but introduces additional considerations in terms of incremental hardware costs and deployment strategies.

Benefits: Recovery Traffic Handling: Design 2 ensures efficient management of recovery traffic. During the rebuild event at 50 Gbps, both Top-of-Rack (ToR) uplink utilization and Data Center Interconnect (DCI) utilization maintain a safety margin, minimizing the risk of congestion.

Risks:

1) *Incremental Hardware Costs:* The implementation of Design 2 incurs incremental hardware costs in the range of millions of dollars. This is attributed to the need for more network equipment and the impact on Colo (Colocation) Space & Power fixed costs, which can increase by 20% to 30%. The less dense new racks or longer deployment time contribute to these elevated costs.
2) *Exceeding Planned Rack Additions:* Deploying 1,000 storage servers under this design necessitates 500 racks (1,000 / 2 storage nodes per rack), exceeding the originally planned 1-year new rack additions for compute (400 racks). To address this challenge, two potential strategies are proposed:
3) *Adding New Racks Without Retrofitting:* This strategy involves adding approximately 100 new racks, requiring 100 new Top of Racks (ToRs), 100 Management Switches, 100 racks, 200 PDUs, 17 Spines, and 17 1.6 T Data Center Interconnects (DCIs) in conjunction with the new racks. While this approach is CapEx heavy, it provides a solution without retrofitting.
4) *Retrofitting Existing Racks:* Alternatively, retrofitting approximately 100 existing compute racks is proposed. This approach, though resource-intensive in terms of Smart Hands hours, extends the deployment lead time. It offers a different cost profile compared to adding new racks.



Design 2 presents a nuanced solution balancing recovery traffic considerations with the challenges of incremental hardware costs and strategic deployment choices. The choice between adding new racks and retrofitting existing ones involves trade-offs in terms of capital expenditure, deployment timelines, and resource utilization. The careful evaluation of trade-offs between hardware costs and deployment strategies underscores the need for a holistic approach, where optimal decision-making aligns with both cost-effectiveness and operational efficiency.

c) *Design 3: Storage Servers-only Rack Configuration*

Design 3 introduces a specialized configuration featuring Storage Servers-only racks, comprising 20 Storage Servers per Storage Servers rack, with 2 Top-of-Rack (ToR) units per rack and 10 Storage Servers under each ToR. Additionally, Compute Servers racks consist of 20 compute servers. This design allows for 1 Storage Servers rack and 5 compute racks per spine, accommodating 20 Storage Servers and 100 compute servers per spine.

Benefits:

1) *Recovery Traffic Handling:* Design 3 ensures robust recovery traffic management, maintaining safety margins in both ToR uplink utilization and Data Center Interconnect (DCI) utilization even during rebuild events at 50 Gbps.
2) *Quick Deployment with Rack Integration:* The configuration enables quick deployment through Rack Integration, eliminating the need for retrofitting for approximately 1,000 Storage Servers. This streamlines the deployment process, enhancing efficiency.
3) *Minimal Impact on Cage Space & Power:* Design 3 introduces no significant changes to cage space and power purchase, ensuring that the existing infrastructure can accommodate the new configuration without major modifications.

Risks: Higher Failure Domain: One notable risk associated with Design 3 is a higher failure domain compared to the other scenarios. The specialized focus on Storage Servers-only racks may introduce a larger failure domain, necessitating careful consideration of redundancy and failover mechanisms.

This design presents a tailored solution emphasizing efficient recovery traffic handling and quick deployment through Rack Integration. While offering benefits in terms of streamlined deployment and minimal impact on existing infrastructure, the higher failure domain introduces a risk that requires thoughtful risk mitigation strategies.

d) *Design 4: Storage-Only Rack and Spine Configuration*

Design 4 introduces a specialized Storage-Only configuration with 28 servers per storage rack and 6 storage-only racks per spine, accommodating a total of 168 Storage Servers per spine.

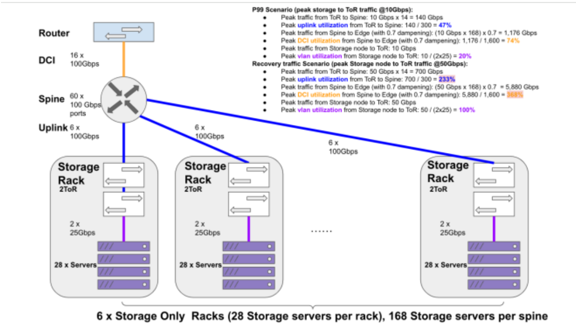
Benefits:

1) *'Plug-n-Play' Supply Allocation:* Design 4 facilitates a seamless 'plug-n-play' approach for supply allocation, streamlining the process of integrating and allocating resources.
2) *Isolation from Compute Resources:* The design ensures isolation from compute resources, preventing over-consumption during rebuild events. This isolation enhances the efficiency of storage operations without impacting compute functionality.

3) *Leveraging Rack Integration for Quick Deployment:* Similar to other designs, Design 4 leverages Rack Integration as a 'block' for quick deployment, eliminating the need for retrofitting and expediting the deployment process.
4) *Minimal Impact on Cage Space & Power Purchase:* Design 4 introduces no significant changes to cage space and power purchase, maintaining compatibility with existing infrastructure.

Risks:

1) *Higher Failure Domain:* Similar to other scenarios, Design 4 presents a risk of a higher failure domain. The specialized focus on Storage-Only racks may increase the impact of potential failures, necessitating robust redundancy measures.
2) *Demand Forecast Accuracy:* The design may require more accurate demand forecasting at a cluster level, especially considering the cluster size, which now encompasses at least 3 spines worth of servers (168 Storage x 3 spines, 504 Storage machines, ~121 PB based on current servers). Accurate demand forecasting becomes critical to align resources with actual storage needs.



Design 4 offers a tailored solution with 'plug-n-play' supply allocation, isolation from compute resources, and rapid deployment through Rack Integration. Despite the benefits, the higher failure domain and the need for accurate demand forecasting underscore the importance of meticulous planning and risk mitigation strategies for successful implementation.

Index	Design 1 (High Storage Servers density mixed-rack)	Design 2 (Low Storage Servers density mixed-rack)	Design 3 (Storage Servers only rack)	Design 4 (Storage Servers only spine)
2024 Storage Servers + Compute Servers Pickup	1,000 Storage Servers + 8,149 Compute Servers	1,000 Storage Servers + 8,149 Compute Servers	1,000 Storage Servers + 8,149 Compute Servers	1,000 Storage Servers + 8,149 Compute Servers
Rack density	8 Storage + 14 Compute	2 Storage + 19 Compute	20 storage per Storage only rack, 20 compute servers	28 Storage nodes per rack
Spine density	48 Storage + 84 Compute	12 Storage + 114 Compute	20 Storage + 100 Compute	168 Storage only
Traffic Congestion Risks				
ToR Uplink Utilization @10Gbps	62%	62%	17%	47%
DCI Utilization @10Gbps	65%	65%	61%	74%
ToR Uplink Utilization @50Gbps (rebuild event)	142%	82%	100%	233%
DCI Utilization @50Gbps (rebuilt event)	149%	86%	96%	368%
Velocity / Costs Considerations				
# of racks triggered by 1,000 Storage servers	125	500	50	36
Original # of new compute racks 1 yr forecast	400	400	400	400
# of existing Compute racks needing retrofitting	0	100	0	0
Additional Network Equipment CapEx	$0	$$Millions	$0	$0
Max Storage Servers servers our racks can house				
Max Storage Servers in sites by Q4 24'	12,000 (1,500 racks * 8 per rack)	3,000 (1,500 racks * 2 per rack)	4,120 i.e. 20 Storage Servers per spine * (151 existing spine + 55 new spine)	-

4. Conclusion

This paper delves into the intricate challenges and considerations associated with deploying storage services within an environment predominantly focused on compute functionality in public clouds. The presented design scenarios, ranging from high-density mixed-racks to specialized storage-only configurations, offer a nuanced exploration of the trade-offs between cost optimization, efficient scalability, and the mitigation of traffic congestion risks. The emphasis on Rack Integration, quick deployment strategies, and the delicate balance between hardware costs and deployment timelines underscore the need for a holistic approach in aligning network architecture with long-term forecasts for scalability and flexibility. Each design presents unique benefits and risks, requiring careful evaluation based on the specific requirements and priorities of the data center. Ultimately, this paper provides valuable insights for Data Center Strategy teams and architects, guiding them in navigating the complexities of deploying storage services while adapting to evolving workloads. The multidimensional considerations explored here pave the way for informed decision-making in optimizing designs that align with the evolving landscape of data center demands.

References

[1] A. Johnson and B. Williams, "Optimizing Data Center Architecture for Scalability and Efficiency," IEEE Trans. Comput., vol. 10, no. 3, pp. 201-215, Mar. 2019.
[2] C. Brown and D. Miller, "Towards a Compute-Only Environment: Challenges and Opportunities," IEEE Cloud Comput., vol. 7, no. 1, pp. 45-58, Jan. 2021.
[3] X. Chen and Y. Wang, "Efficient Strategies for Storage Services Deployment in Cloud Environments," IEEE Trans. Cloud Comput., vol. 12, no. 2, pp. 89-104, Feb. 2022.
[4] M. Davis and E. Smith, "Strategies for Dynamic Capacity Expansion in Cloud Infrastructures," IEEE Internet Comput., vol. 15, no. 4, pp. 76-89, Apr. 2018.
[5] K. Lee and S. Kim, "Mitigating Traffic Congestion Risks in Cloud Networks," IEEE Trans. Netw. Serv. Manag., vol. 9, no. 3, pp. 301-315, Sep. 2020.
[6] R. Patel and S. Gupta, "Cost Implications of Elastic Cloud Computing Services," IEEE Cloud Comput., vol. 11, no. 4, pp. 123-137, Dec. 2019.