Research Statement

Global-Scale Decentralized Systems with Edge-Cloud Data Management

Faisal Nawab
http://nawab.me

"*[D]ue to the often unpredictable network latency, expensive bandwidth, and privacy concerns, especially in a mobile environment, cloud computing often cannot meet the stringent requirements of applications.*"

- NSF Workshop Report on Grand Challenges in Edge Computing [9]

From Cloud Computing to Edge-Cloud Computing

Processing large quantities of data is becoming more ubiquitous and is the driving force behind the continuing success of Internet Services and Big Data applications. The way data-intensive applications are deployed has been radically transformed by the cloud computing paradigm realized through massive-scale data centers and consolidation of resources. However, the current cloud model—as a computing paradigm—is limited by many factors that might slow the progress of Internet technology. First, consolidating in data centers makes compute and storage resources far away from users. Requests to access applications are fundamentally limited by the speed of light. Second, there is a limit on how much resources can be consolidated in a data center due to factors including real estate and energy consumption. Third, placing data in consolidated data centers limits users control on their data and privacy. Fourth, consolidated data centers create a communication model where a data center is involved in all data access activity, which places a burden—and limit—on the network infrastructure.

I envision that we are now in the next phase of computing; **moving away from complete consolidation towards a semi-consolidated, semi-decentralized model that I call the *edge-cloud* model.** This looming edge-cloud model will manifest itself by pushing the compute and storage resources closer to users, including users' end devices and the network and telephony infrastructures in between. The current advances in mobility and edge computing lay the grounds for data processing and management at the edge. This is why now is the right time to study the design of data management solutions over an edge-cloud model. Such a study requires revisiting the basic principles of data management designs to address the newfound, salient challenges and take full advantage of the edge-cloud model.

Research Goal: A Data Management Infrastructure for Edge-Cloud Computing

Given the anticipated shift to edge-cloud computing, I investigate the implications of this shift on data management infrastructures. The current data management infrastructure—that supports the web, Big Data, and data science applications—is built for the cloud environment. This makes them unsuitable to be used in the edge-cloud. In fact, what is needed is a significant redesign and reconsideration of the fundamentals and principles of distributed data management. This is due to the disparities between the cloud and edge-cloud models. Unlike the cloud, the edge-cloud model of computation consists of a large number of edge nodes that are distributed around the world in addition to the cloud nodes. These edge nodes have limited performance capabilities and are prone to failures and disconnections. Edge nodes are separated by large wide-area distances—amplifying coordination costs. Furthermore, edge nodes might be run by users or third parties that are not trusted by the rest of the network. This invites a more stringent model of fault-tolerance that tolerates malicious and arbitrary behavior, referred to in the literature as byzantine failures.

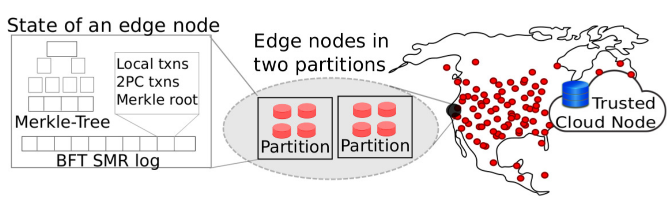

Figure 1: An example of an edge-cloud data management deployment that spans cloud and edge nodes.

To realize this vision of a data management infrastructure for edge-cloud environments, I take a systems approach, where we study the challenges and design principles in the form of building a new data management system that we call *WedgeDB* (Figure 1.) In this framework of designing WedgeDB, we tackle the various data management and systems challenges of edge-cloud systems. We begin by focusing on the fundamental distributed data management functionalities such as transaction processing and consensus. Transaction processing and consensus are the foundations of many other data management functionalities. By focusing on them, we aim to facilitate the research on the remaining challenges associated with data management systems for the edge-cloud.

My plan in the study of edge-cloud data management is to focus on two daunting challenges of edge-cloud systems that are due the edge nodes' ability to participate in the storage and computation of data:

- *(1) scalability*: in the edge-cloud, a large number of edge nodes (across wide-area links) are working together to maintain the state of the database and serve requests. All nodes must agree on the state of the database and process transactions consistently.
- *(2) edge nodes are not trusted:* edge nodes might be operated by users. There is a need to find a way to store and process data in a trusted, accountable way to the application and other edge nodes. This requires adopting a stringent fault-tolerance guarantee of tolerating arbitrary and malicious failures, *i.e.*, byzantine fault-tolerance.

PI Qualifications to Pursue The Edge-Cloud Vision (Current and Previous Research)

I believe that my background and recent work uniquely qualifies me to pursue and succeed in realizing the future of edge-cloud data management. Specifically, my expertise and work in the following areas align with the goals of edge-cloud data management:

1. **Geo-replication:** This is the area of managing data that is replicated across multiple data centers around the world. This area is relevant to edge-cloud systems as it tackles the challenges of managing data across wide-area links. This area enjoyed a lot of interest as it is one of the key ingredients to enable tolerating one of the most challenging problems in cloud computing—data center-scale outages. This area has been my specialization and dissertation topic [15]. I have published work on geo-replication [16] that span geo-replicated transaction processing [13, 17, 18, 32], foundations on the performance limits geo-replication systems [23], and communication infrastructure for geo-replication [3, 22].

   My work in this area asks the fundamental question: *what is the performance limit of access coordination in geo-replicated systems?* Answering this question started with investigations on existing replication technologies and understanding their limitations. Specifically, it turns out that the pattern of coordination that relies on requests and responses creates a performance limit on how fast coordination can be performed [13]. However, we then found that it is possible to adopt a proactive pattern of coordination that enables coordinating access between data centers in less than a round-trip time [17]. This opened the opportunity to formalize the new pattern of coordination and to derive the theoretical limits of performance in geo-replicated systems [23].

2. **Distributed transaction processing:** This is the area of managing data that is distributed across machines—which is typical in many cloud computing systems and is essential in tackling the challenges of managing data across decentralized nodes at the edge. In this line-of-work, I focus on enabling easy-to-use data access abstractions (called database transactions) to enable programmers to manage data in a consistent and efficient manner. This has been my main area of research and I have published work on transaction processing for cloud systems [2, 12, 14, 24], non-volatile memory systems [25, 26], and temporal databases [11].

3. **Systems spanning the cloud and the edge:** An important aspect of edge-cloud systems is the ability to manage and coordinate the interactions between edge and cloud nodes. I have started working on various topics in this area including a paper about the opportunities of data management on the edge [20]. This paper lays the foundation of how an edge-cloud system infrastructure looks like and discusses the research challenges that are entailed in building a data management system on that infrastructure. I also worked on observing existing basic blocks of distributed systems and redesigning them to work over an environment with both cloud and edge nodes [19]. Specifically, our proposed consensus protocol (DPaxos) redesigns the widely-used consensus protocol paxos and integrates hierarchy and locality-awareness as well as a dynamic mode of coordination to enable working on less reliable edge deployments. Lastly, I have also explored other types of workloads that are expected to be typical in edge-cloud systems such as video-based workloads and tackled the problem of managing video analytics in edge-cloud environments [10].

4. **Byzantine Fault-Tolerance:** A key ingredient in the study of edge-cloud systems in designing protocols that tolerate potentially malicious nodes at the edge. To this end, I started tackling work to build distributed systems that enable tolerating malicious nodes—known as byzantine fault-tolerance (BFT) protocols. Specifically, I target building protocols that can scale beyond the traditional BFT deployments to support large-scale deployments across wide-area latencies. Scaling to these environments is one of the main challenge that faces the adoption of BFT protocols in edge-cloud systems. For example, Blockplane [30] is a middleware design that we proposed to enable programmers to write byzantine fault-tolerant applications on geo-replicated environments. Blockplane targets environments with global-scale deployments spanning multiple data centers. It introduces the notions of hierarchy and locality awareness to BFT systems. This will enable adopting these ideas in edge-cloud systems with large-scale deployments.

   I am now continuing this line of work and started tackling the problem of distributed transaction processing in large-scale BFT systems. This work, which is part of WedgeDB, introduces transaction processing protocols that are capable

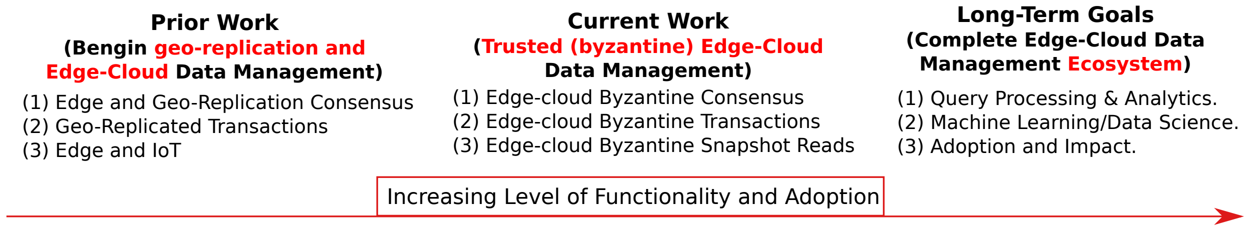

Figure 2: The vision and long-term goals of this statement build upon my prior and current work to achieve a complete data management ecosystem for the edge-cloud.

of ensuring that all participating nodes cannot lie or act maliciously to threaten the integrity of data and transactions. Additionally, we are working on efficient analytics frameworks that enables reading from multiple partitions in a consistent and trusted way even in the presence of malicious actors.

5. **Edge computing and IoT:** In edge-cloud systems, most of the nodes and interactions are going to happen at the edge of the network. To this end, I will utilize my background and work on data management infrastructures for IoT and edge applications [5, 6, 7] and applications of using more complex learning and data science models and access [8, 21]. An example of a recent work in the area is an analytical study of the characteristics and behavior of running byzantine fault-tolerant systems over IoT networks [4]. This study provides insights on the bottlenecks, challenges, and opportunities in this domain. We are currently using these insights to build new BFT protocols that target edge and IoT environments.

   More recently, we have worked in building a decentralized infrastructure to unify access to IoT data with a system called AnyLog [1]. AnyLog leverages recent research technology in decentralized coordination and blockchain, BFT systems, trusted hardware, and authenticated data structures to design a system that is capable of handling the sheer data volume and scale of a globally-scalable, unified IoT infrastructure. The main message of AnyLog is that rather than expecting that one organization is going to control the entire infrastructure of edge computing and IoT (which is unreasonable given the scale of such deployments), we envision that many organizations and control domains would collaborate to provide access to edge and IoT resources in large areas. This provides the opportunity of actually realizing emerging IoT and edge applications but also introduces the challenges of managing the infrastructure and relations across control domains. We solve this challenge via a compensation and dispute mechanisms that ensure that each party is doing what they are supposed to do and when a malicious act is performed that it is going to be detected and punished. This has been enabled by recent advances in decentralized infrastructures and blockchain-based technologies.

   Also, early in my career I have studies network designs to enable edge computing across wireless mesh networks [27, 28, 29].

Figure 2 is an illustration of how my future plans and long-term goal of a complete edge-cloud data ecosystem is positioned relative to my current and on-going work. On-going work builds to build a trusted (byzantine) edge-cloud data management system builds upon my prior work in exploring edge-cloud systems [10, 19, 20]. However, my prior work for these problems assumed a weaker (benign) fault-tolerance guarantee, where byzantine failures (malicious and arbitrary failures) are not tolerated. My ongoing work adds the functionality of tolerating byzantine failures to create a trusted edge-cloud environment. Tackling the challenges arising from considering byzantine failures is daunting and constitute the main research contribution of my ongoing work.

My long-term career goal is to build a complete ecosystem for data management and analytics for edge-cloud computing. After achieving the goals of ongoing work to build trusted (byzantine) coordination and transaction processing protocols, I plan to pursue other essential components for edge-cloud data management including query processing and analytics, and emerging data processing paradigms such as machine learning and data science. This will also enable me to pursue another long-term goal, which is the adoption and impact of the edge-cloud model in industry. This will entail building relationships with industry and demonstrating the feasibility and potential of the proposed ecosystem.

Collaborations

Work to realize this vision of global-scale decentralized applications via edge-cloud systems requires expertise in many fields as well as physical infrastructure to enable experiments and evaluations on real world scenarios. For these reasons, I have invested in collaborating with colleagues with similar interests from UC Davis (Prof. Mohammad Sadoghi), UC Irvine (Prof. Sharad Mehrotra), University of Maryland (Prof. Daniel Abadi) and KAUST (Prof. Basem Shihada) as well as industry collaborators, most notably a collaboration with AnyLog—a startup that aims to build a decentralized data processing infrastructure for IoT. These collaboration has resulted in a number of recent publications [1, 4, 30, 31] and ongoing projects that are related to the topic of this research statement.

References

[1] D. J. Abadi, O. Arden, F. Nawab, and M. Shadmon. Anylog: a grand unification of the internet of things. In *CIDR*, 2020.

[2] D. Agrawal, A. El Abbadi, V. Arora, C. Budak, T. Georgiou, H. A. Mahmoud, F. Nawab, C. Sahin, and S. Wang. Mind your ps and vs: A perspective on the challenges of big data management and privacy concerns. In *2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015, Jeju, South Korea, February 9-11, 2015*, pages 1–6, 2015.

[3] D. Agrawal, A. El Abbadi, H. A. Mahmoud, F. Nawab, and K. Salem. Managing geo-replicated data in multi-datacenters. In *Databases in Networked Information Systems - 8th International Workshop, DNIS 2013, Aizu-Wakamatsu, Japan, March 25-27, 2013. Proceedings*, pages 23–43, 2013.

[4] M. Alaslani, F. Nawab, and B. Shihada. Blockchain in iot systems: End-to-end delay evaluation. *IEEE Internet of Things Journal*, 2019.

[5] V. Arora, F. Nawab, D. Agrawal, and A. El Abbadi. Multi-representation based data processing architecture for iot applications. In *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, GA, USA, June 5-8, 2017*, pages 2234–2239, 2017.

[6] V. Arora, F. Nawab, D. Agrawal, and A. El Abbadi. Typhon: Consistency semantics for multi-representation data processing. In *2017 IEEE 10th International Conference on Cloud Computing (CLOUD), Honolulu, HI, USA, June 25-30, 2017*, pages 648–655, 2017.

[7] V. Arora, F. Nawab, D. Agrawal, and A. El Abbadi. Janus: A hybrid scalable multi-representation cloud datastore. *IEEE Trans. Knowl. Data Eng.*, 30(4):689–702, 2018.

[8] C. Biafore and F. Nawab. Graph summarization for geo-correlated trends detection in social networks. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2247–2248, 2016.

[9] M. Chiang and W. Shi. NSF workshop report on grand challenges in edge computing. *NSF Workshop Report*, 2016.

[10] P. Grulich and F. Nawab. Collaborative edge and cloud neural networks for real-time video processing. *PVLDB*, 11(12):2046–2049, 2018.

[11] D. B. Lomet and F. Nawab. High performance temporal indexing on modern hardware. In *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*, pages 1203–1214, 2015.

[12] H. A. Mahmoud, V. Arora, F. Nawab, D. Agrawal, and A. El Abbadi. Maat: Effective and scalable coordination of distributed transactions in the cloud. *PVLDB*, 7(5):329–340, 2014.

[13] H. A. Mahmoud, F. Nawab, A. Pucher, D. Agrawal, and A. El Abbadi. Low-latency multi-datacenter databases using replicated commit. *PVLDB*, 6(9):661–672, 2013.

[14] S. Maiyya, F. Nawab, D. Agrawal, and A. El Abbadi. Unifying consensus and atomic commitment for effective cloud data management. *PVLDB*, 12(5):611–623, 2019.

[15] F. Nawab. *Global-Scale Data Management with Strong Consistency Guarantees*. PhD thesis, University of California, Santa Barbara, USA, 2017.

[16] F. Nawab. Geo-scale transaction processing. In *Encyclopedia of Big Data Technologies.* 2019.

[17] F. Nawab, D. Agrawal, and A. El Abbadi. Message futures: Fast commitment of transactions in multi-datacenter environments. In *CIDR 2013, Sixth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 6-9, 2013, Online Proceedings*, 2013.

[18] F. Nawab, D. Agrawal, and A. El Abbadi. The challenges of global-scale data management. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2223–2227, 2016.

[19] F. Nawab, D. Agrawal, and A. El Abbadi. Dpaxos: Managing data closer to users for low-latency and mobile applications. In *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, pages 1221–1236, 2018.

[20] F. Nawab, D. Agrawal, and A. El Abbadi. Nomadic datacenters at the network edge: Data management challenges for the cloud with mobile infrastructure. In *Proceedings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018.*, pages 497–500, 2018.

[21] F. Nawab, D. Agrawal, A. El Abbadi, and S. Chawla. COP: planning conflicts for faster parallel transactional machine learning. In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT 2017, Venice, Italy, March 21-24, 2017.*, pages 132–143, 2017.

[22] F. Nawab, V. Arora, D. Agrawal, and A. El Abbadi. Chariots: A scalable shared log for data management in multi-datacenter cloud environments. In *Proceedings of the 18th International Conference on Extending Database Technology, EDBT 2015, Brussels, Belgium, March 23-27, 2015.*, pages 13–24, 2015.

[23] F. Nawab, V. Arora, D. Agrawal, and A. El Abbadi. Minimizing commit latency of transactions in geo-replicated data stores. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 1279–1294, 2015.

[24] F. Nawab, V. Arora, V. Zakhary, D. Agrawal, and A. El Abbadi. A system infrastructure for strongly consistent transactions on globally-replicated data. *IEEE Data Eng. Bull.*, 40(4):3–14, 2017.

[25] F. Nawab, D. R. Chakrabarti, T. Kelly, and C. B. M. III. Procrastination beats prevention: Timely sufficient persistence for efficient crash resilience. In *Proceedings of the 18th International Conference on Extending Database Technology, EDBT 2015, Brussels, Belgium, March 23-27, 2015.*, pages 689–694, 2015.

[26] F. Nawab, J. Izraelevitz, T. Kelly, C. B. M. III, D. R. Chakrabarti, and M. L. Scott. Dalí: A periodically persistent hash map. In *31st International Symposium on Distributed Computing, DISC 2017, October 16-20, 2017, Vienna, Austria*, pages 37:1–37:16, 2017.

[27] F. Nawab, K. Jamshaid, B. Shihada, and P. Ho. TMAC: timestamp-ordered MAC for CSMA/CA wireless mesh networks. In *Proceedings of 20th International Conference on Computer Communications and Networks, ICCCN 2011, Maui, Hawaii, USA, July 31 - August 4, 2011*, pages 1–6, 2011.

[28] F. Nawab, K. Jamshaid, B. Shihada, and P. Ho. Mac-layer protocol for TCP fairness in wireless mesh networks. In *2012 1st IEEE International Conference on Communications in China (ICCC), Beijing, China, August 15-17, 2012*, pages 448–453, 2012.

[29] F. Nawab, K. Jamshaid, B. Shihada, and P. Ho. Fair packet scheduling in wireless mesh networks. *Ad Hoc Networks*, 13:414–427, 2014.

[30] F. Nawab and M. Sadoghi. Blockplane: A global-scale byzantizing middleware. In *35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019*, pages 124–135, 2019.

[31] F. Nawab, M. Sadoghi, and A. Singh. Trusted wide-area and edge systems using hierarchical permissioned blockchain. In *1st Workshop on Blockchain and Distributed Ledger (BCDL)*, 2019.

[32] V. Zakhary, F. Nawab, D. Agrawal, and A. El Abbadi. DB-Risk: The Game of Global Database Placement. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2185–2188, 2016.