Case ID:M14-075P^

Published: 2/26/2020

Inventors

Lei Ying

Weina Wang

Matthew Barnard

Contact

Shen Yan
shen.yan@skysonginnovations.com

Peer-to-Peer Architecture for Processing Big Data

Big data is collected and analyzed by organizations everywhere to help make more accurate predictions and decisions. However, big data is useless without sufficient infrastructure for accessing, storing, and transferring the increasingly massive data sets. Current architectures rely on a centralized job scheduler and resource allocator, where a few designated master nodes assign storage roles to the many remaining slave nodes within a network. Failure at one or more of these master nodes increases the workload of the remaining nodes, often bottlenecking network bandwidth which causes a degradation or interruption of service. Centralized systems have limited scalability as they are unable to expand beyond the capacity of the master nodes, and their architectures require continuous monitoring and configuration to ensure hierarchy is preserved and to accommodate for unexpected workload. These constraints tax the efficiency of a data network and contribute to unnecessary hardware and administrative costs.

Researchers at ASU have developed peer-to-peer architecture for processing big data that provides totally decentralized, redundant, fault-tolerant data storage and map task scheduling with high throughput and data locality. Nodes within the system are equal in their assumption of master and slave roles with regards to data storage, job delegation, and reporting. This facilitates rapid failover in the event of single or massive node loss. While data or current tasks may be lost, the availability of the system remains high as workload is spread evenly across remaining nodes. The architecture establishes a network that is infinitely scalable, where newly introduced nodes are welcomed as if they were already a part of the network.

Potential Applications

- Big Data Processing
- Data Assembly
- Dynamic Cloud Computing
- Process Control Analytics

Benefits and Advantages

- Economical
  - Eliminates administrative costs for reconfiguring master and slave nodes.
  - Reduces need for expensive, high-bandwidth failsafe hardware.
- Practical – Self-organizing and readily adaptable to different network sizes (and network size changes).
- Reliable
  - Facilitates rapid failover in the event of single or massive node loss.
  - Balances processing within a network so that no node is overworked.

For more information about the inventor(s) and their research, please see

[Dr. Lei Ying's directory webpage](#)