**Motivation**

- MapReduce is standard for data-intensive computing.
  - But it is designed for a single data-center.
- Widely distributed data does not fit its design assumptions.
- Lots of network traffic

**Nebula**

- A location and context-aware Distributed Edge Cloud Infrastructure.
- Using computation and storage resources of edge volunteer and dedicated nodes.

**Nebula: Distributed Edge Cloud**

**Design Goal & Architecture**

- Voluntary compute & storage nodes can easily join, leave, and fail at any time.
- Distributed data-intensive computing.
- Location-aware.
- Sandbox for security.
- Fault tolerance.

**MapReduce on Nebula**

- Nebula MapReduce reduces network traffic.
  - The data is already dispersed on Nebula.
  - Only intermediate files need to be sent for results.
- Nebula MapReduce provides fault-tolerance and scalability.

**Evaluation**

- Nebula outperforms centralized infrastructure.
- Nebula’s location-awareness improves performance.

**Future Work**

- Explore how external data can be inserted into Nebula for aggregation and decomposition.
- Expand the range of data-intensive frameworks and applications ported to Nebula.
- Partition resources across frameworks and applications.