

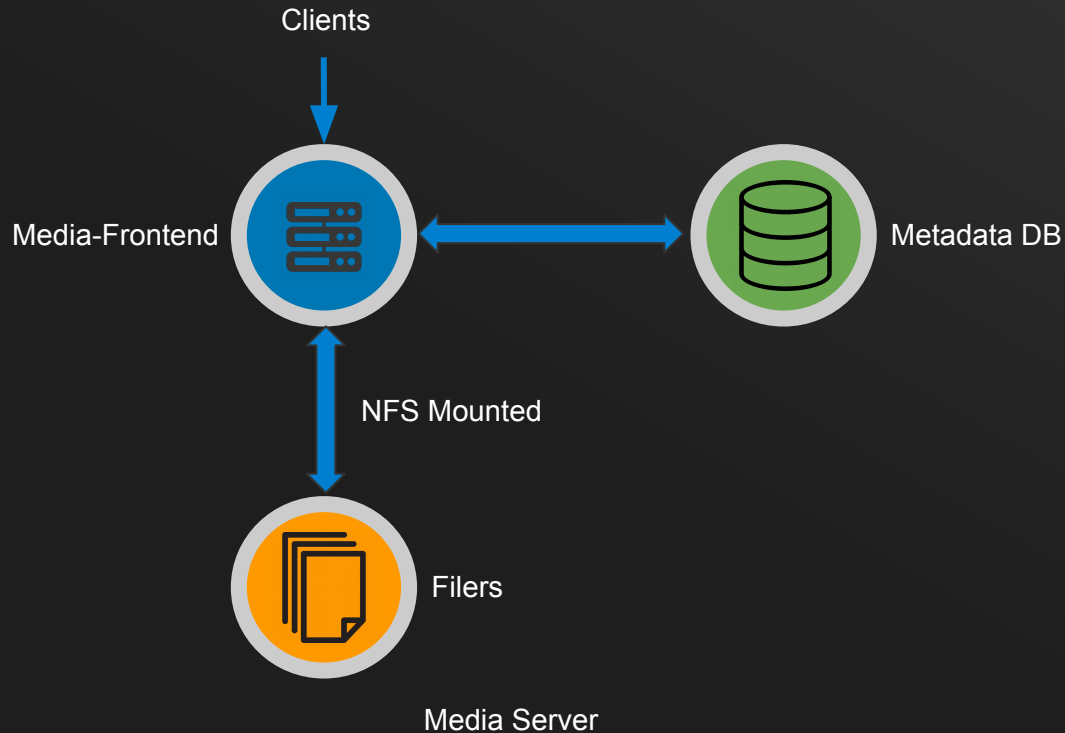


AMBRY: LinkedIn's Immutable Blob Storage System

Arjun Shenoy
Data SRE



Motivation



Issues Faced

- High MTTR
- Operational Overhead
- Scalability issue
- Not Cost effective

Motivation



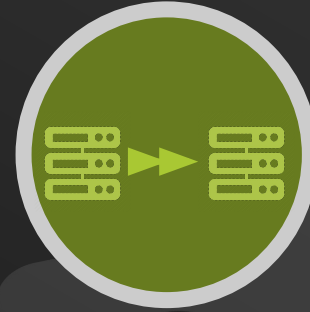
Reliability



Scalability



Low MTRR



Active - Active



Low Cost



Evaluation



Inspiration



AMBRY





Handle Based



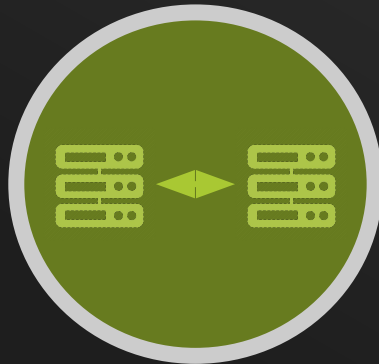
Horizontally Scalable



Petabyte Scalable



Metadata Support



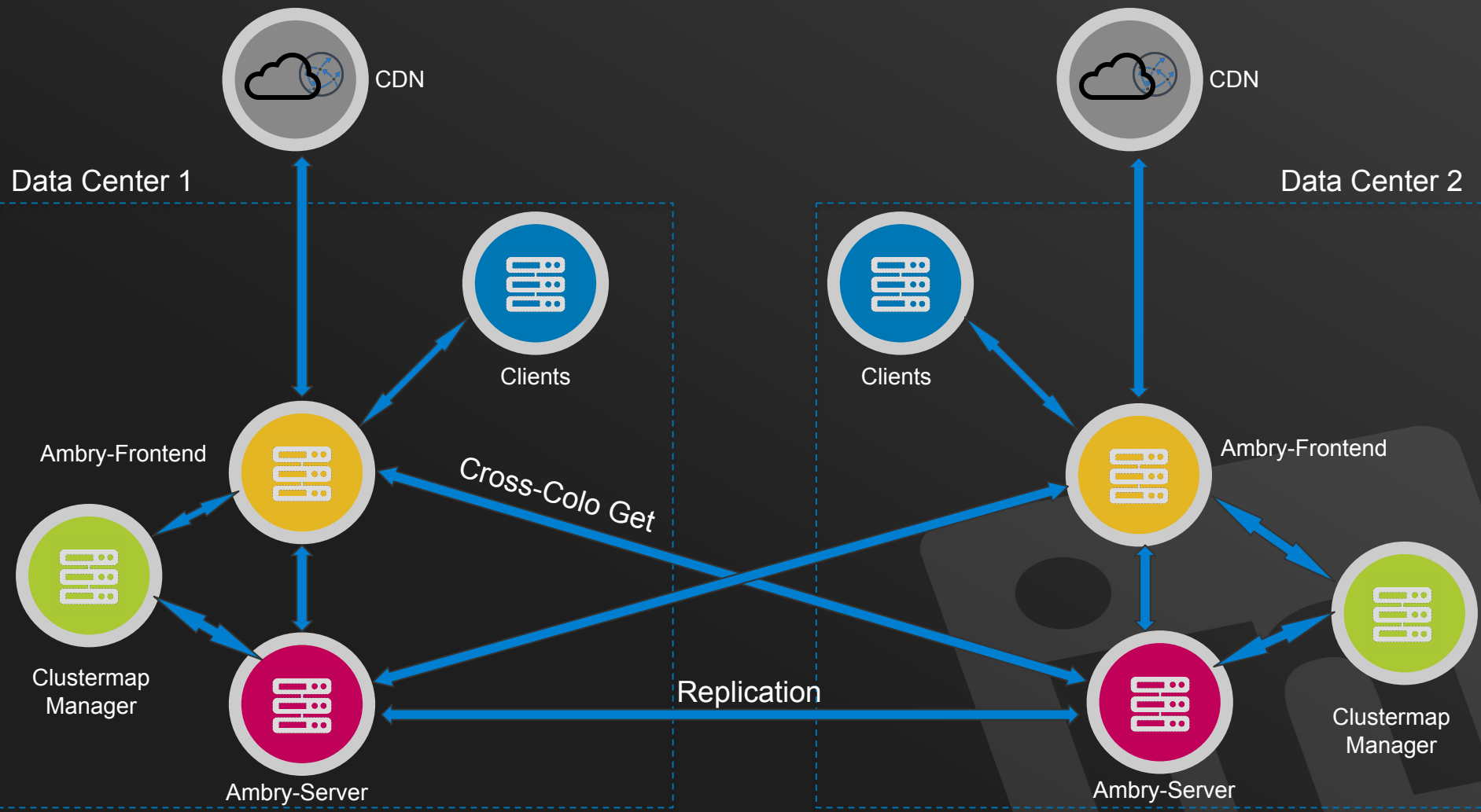
Active-Active



Large Blob Support

ARCHITECTURE





COMPONENTS





Ambry-Frontend



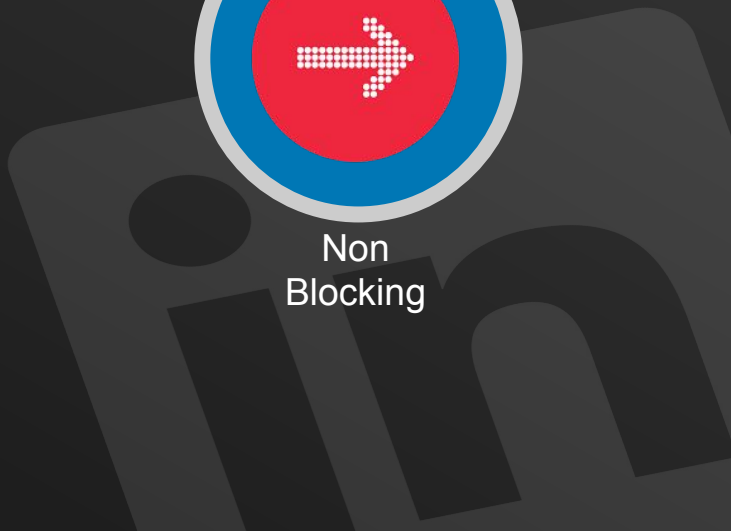
Coordination



Security



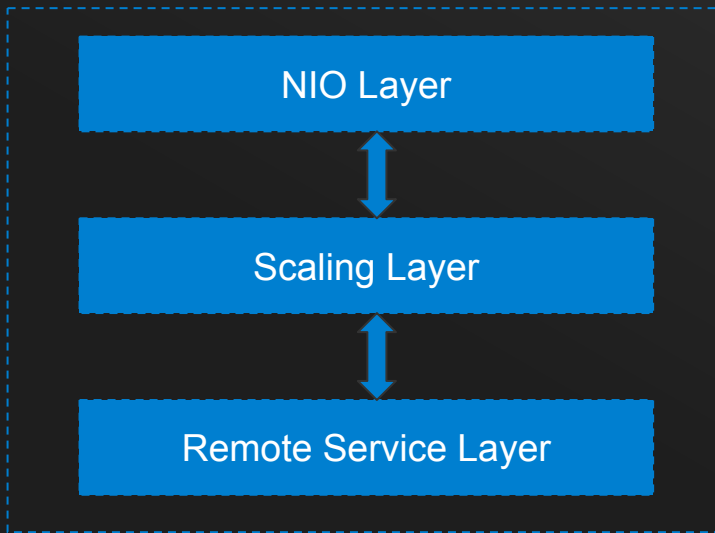
Non
Blocking



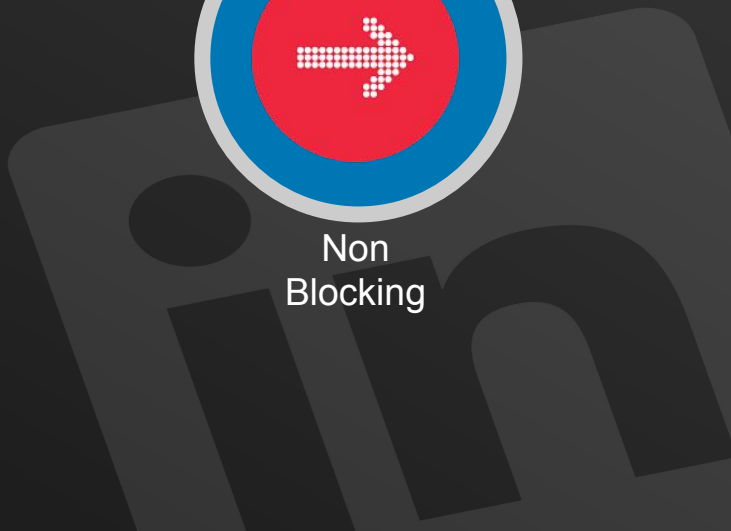


Ambry-Front End

Ambry-frontend (Non-Blocking Model)



Non
Blocking





Ambry-Server

Responsible

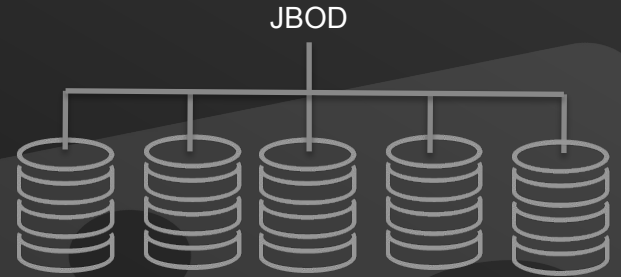


Storage



Replication

Built on





Clustermap-Manager



- HardwareLayout file
- Node State
- Disk Capacity and State

```
{
  "clusterName": "TestCluster",
  "version": 1,
  "datacenters": [
    {
      "dataNodes": [
        {
          "disks": [
            {
              "capacityInBytes": 21474836480,
              "hardwareState": "AVAILABLE",
              "mountPath": "/tmp"
            }
          ],
          "hardwareState": "AVAILABLE",
          "hostname": "*****",
          "port": 15088
        }
      ],
      "name": "Data-Center"
    }
  ]
}
```

- PartitionLayout file
- Mapping to nodes
- Partition State

```
{
  "clusterName": "TestCluster",
  "version": 1,
  "partitions": [
    {
      "id": 0,
      "partitionState": "READ_WRITE",
      "replicaCapacityInBytes": 10737418240,
      "replicas": [
        {
          "hostname": "*****",
          "mountPath": "/tmp/x001/ambrydata/",
          "port": 15088
        }
      ]
    }
  ]
}
```

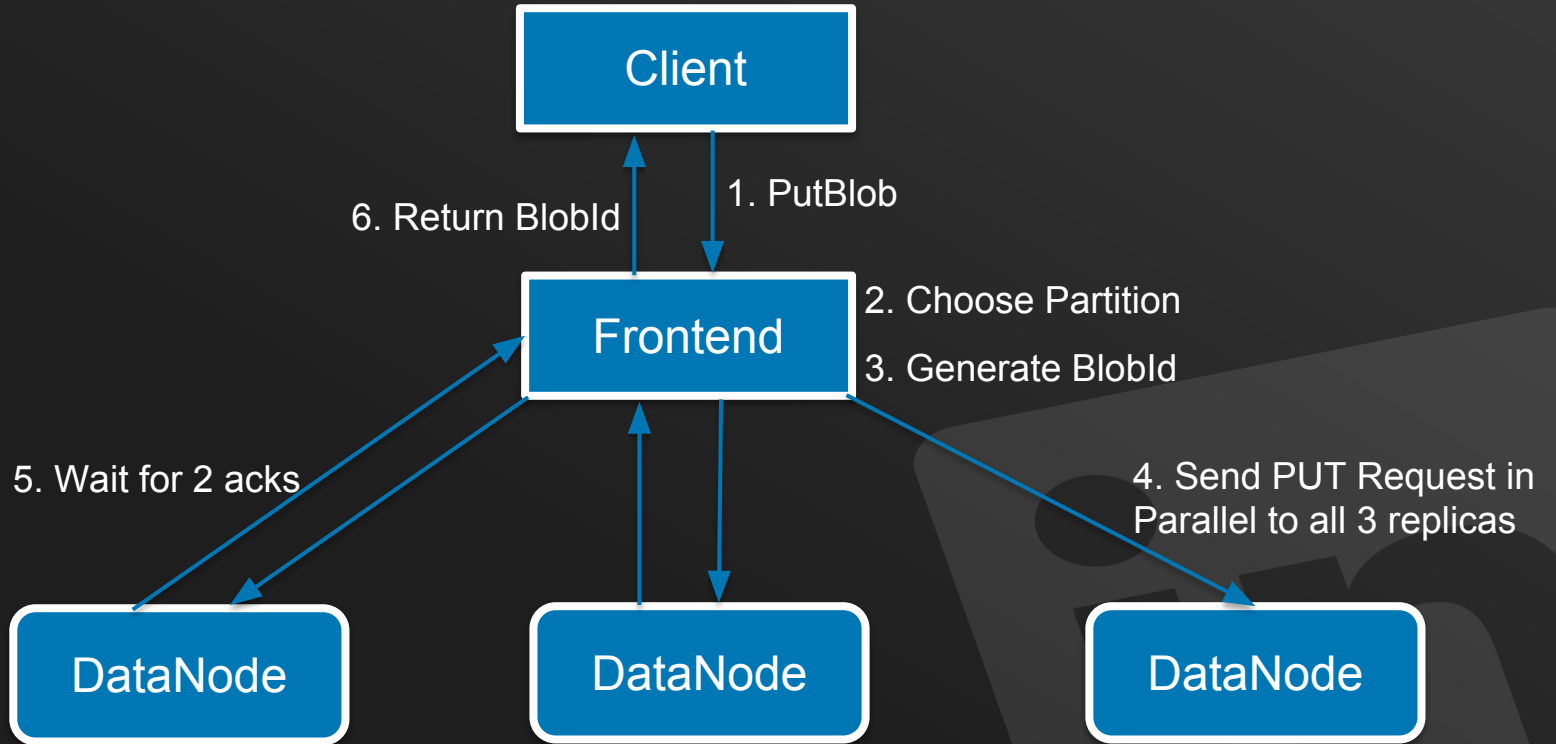

HOW DO THEY INTERACT?



PutBlob WorkFlow

Sample Blob-Id:

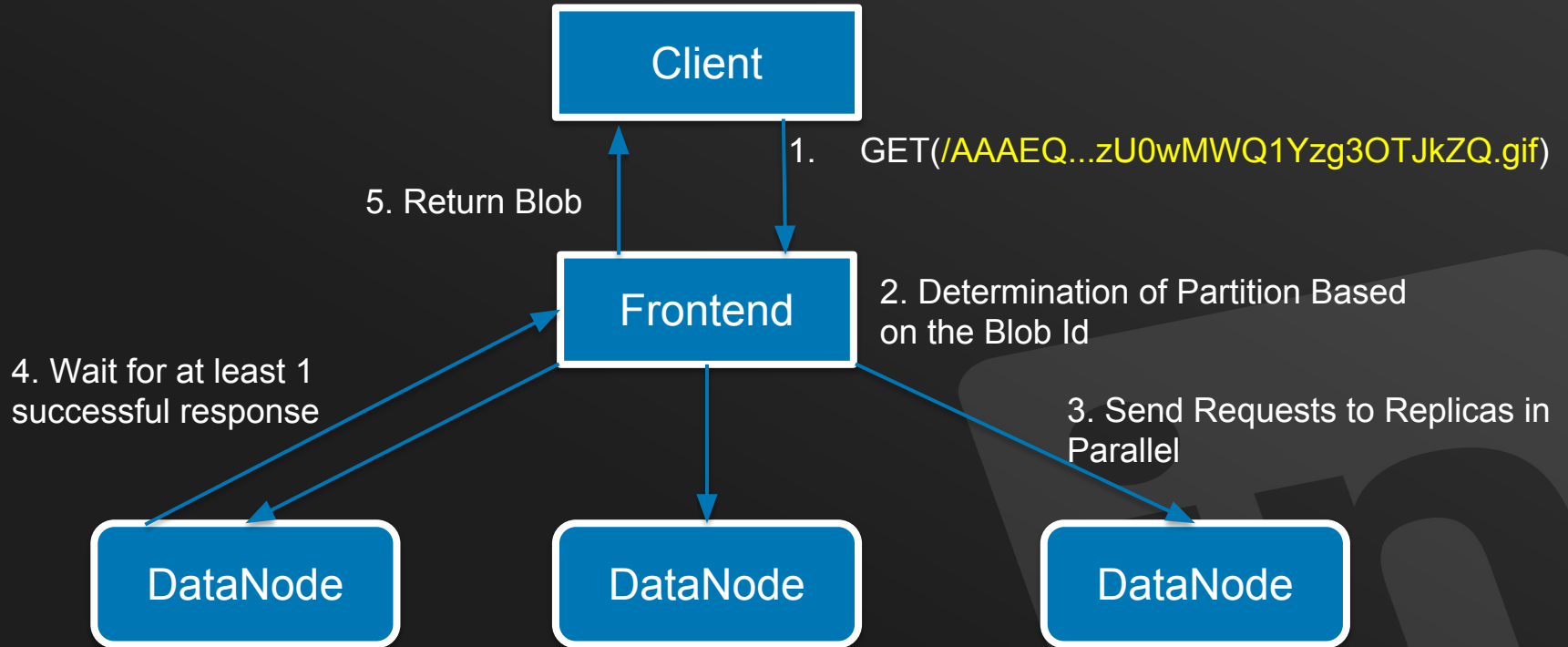
/AAEAAQAAAAAAAAADFAAAAJDMYyWZiOTJmLTBkNDYtNDQyNS1iYzU0LWEwMWQ1Yzg3OTJkZQ.gif



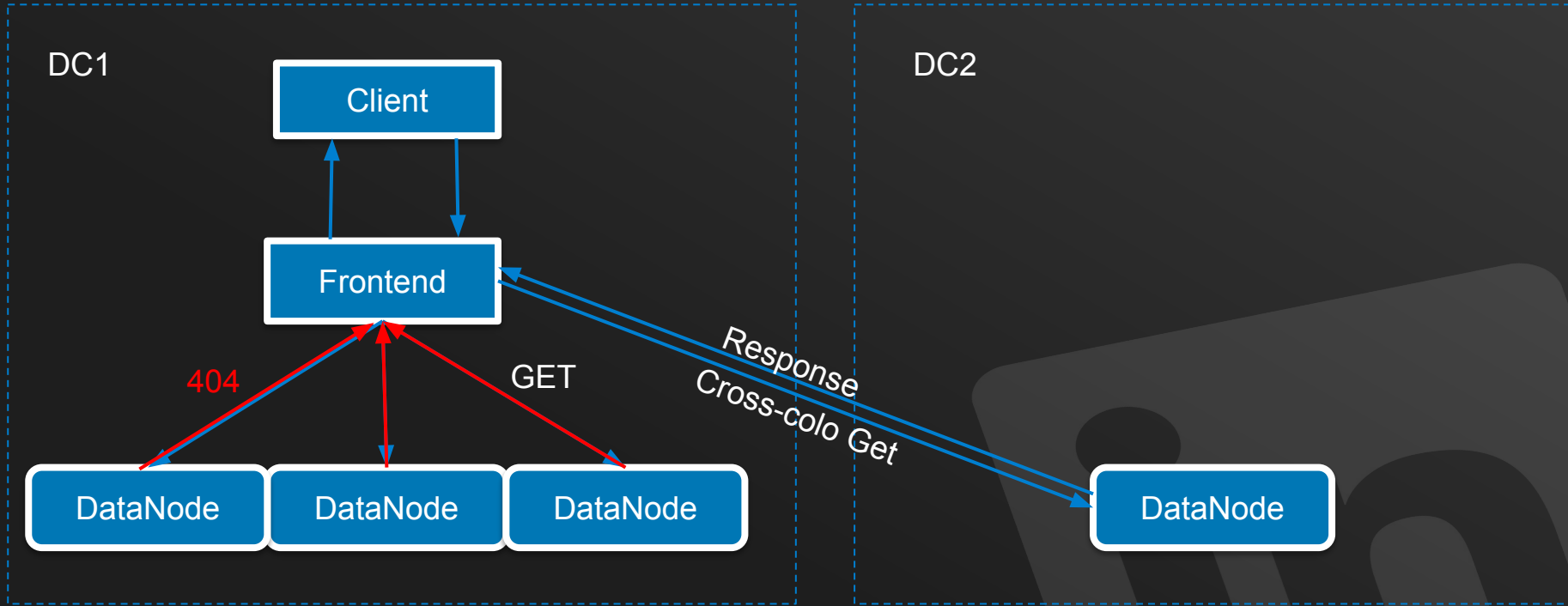
PutBlob Statistics

- Average QPS: **370**
- Latencies(95th Percentile):
 - SmallBlob (<100KB): **<10ms**
 - MediumBlob(100KB - 4MB): **<10ms** (Goes up to **20ms** extreme cases)
 - LargeBlob (>4MB): **<50ms** (Goes up to **80ms** extreme cases)

GetBlob Workflow

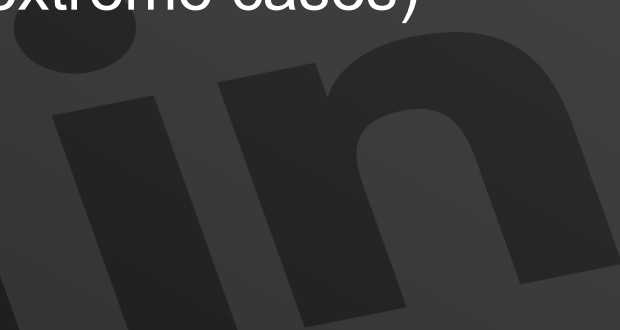


GetBlob Workflow

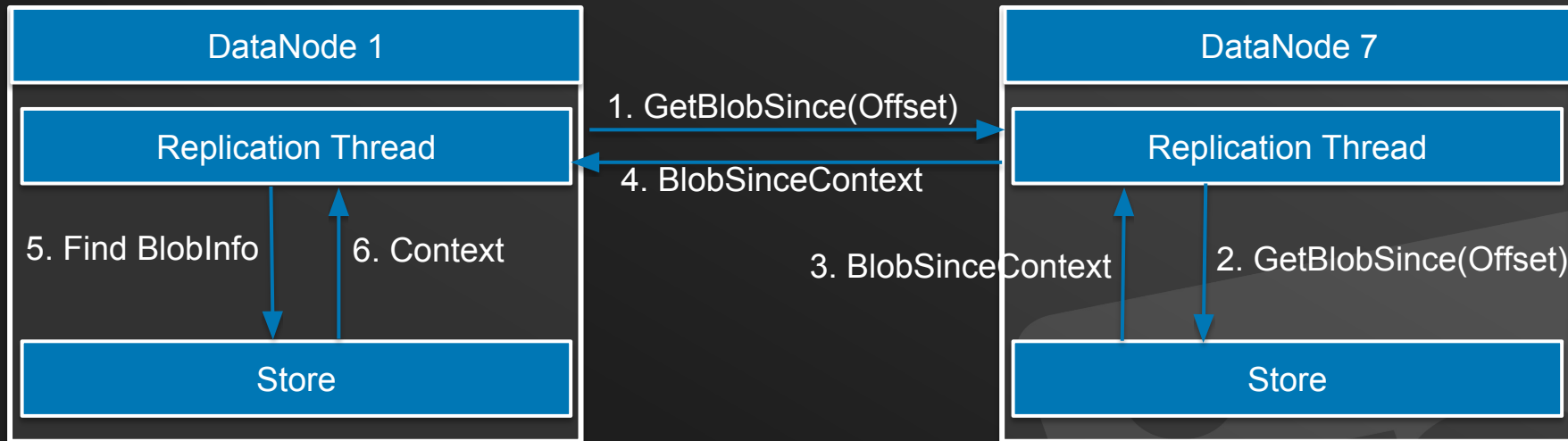


GetBlob Statistics

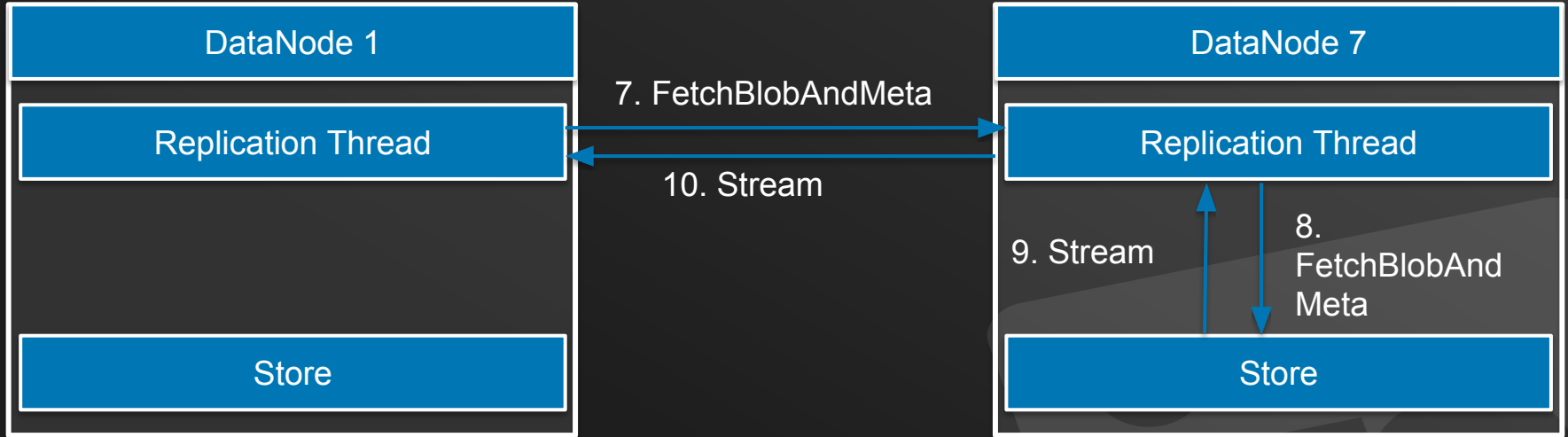
- Average QPS: 24K
- Latencies(95th Percentile):
 - SmallBlob: <15ms
 - MediumBlob: <75ms (Goes up to 700ms extreme cases)
 - LargeBlob: <200ms (Goes up to 1.5s extreme cases)



Replication Workflow



Replication Workflow



- Inter DC Replication Time: **20-100ms**
- Intra DC Replication Time: **<10ms**

SOME STATISTICS



Total Capacity: **900TB**

Serving: **300TB**

Growth Rate: **1160GB** per day

Use-cases: Images, PDFs, Static files, Videos so on...



We're Open Source!

Git Link: github.com/linkedin/ambry/wiki

Blog Post:

engineering.linkedin.com/blog/2016/05/introducing-and-open-sourcing-ambry---linkedins-new-distributed-

Thank You!!



DISCLAIMER

The information in this presentation was compiled from sources believed to be reliable for informational purposes only. It does not constitute legal or professional advice. The views and opinions expressed in this presentation are those of the authors and do not necessarily represent official policy or position of LinkedIn

All product and company names are trademarks™ or registered® trademarks of their respective holders.

