

Kinesis Consumers Types

Shared (Classic) Fan-out Consumer - pull

- Low number of consuming applications
- Read throughput: 2 MB/sec per shard across all consumers
- Max. 5 GetRecords API calls/sec
- Latency ~200 ms
- Minimize cost (\$)
- Consumers poll data from Kinesis using GetRecords API call
- Returns up to 10 MB (then throttle for 5 seconds) or up to 10000 records

Enhanced Fan-out Consumer - push

- Multiple consuming applications for the same stream
- 2 MB/sec per consumer per shard
- Latency ~70 ms
- Higher costs (\$\$\$)
- Kinesis pushes data to consumers over HTTP/2 (SubscribeToShard API)
- Soft limit of 5 consumer applications (KCL) per data stream (default)