This document is an analysis that presents a comparison of DataStax Enterprise (DSE) and DynamoDB for a specific IoT use cases involving sensor and log data. This analysis occurred in October 2014.

As demonstrated in a recent current proof of concept, DSE provides the transactional and analytical capabilities to meet the desired use case with one integrated product stack. Understanding DSE Total Cost of Ownership (TCO) is predictable and manageable because of DSE’s licensing model combined with its linear scalability on commodity and/or cloud infrastructure.

Given the considerations of the use case, the key takeaways of how a DynamoDB solution would compare to a DSE solution are below. Our research came from online Amazon product documentation and public information such as best practices blogs. These items are described as potential Amazon solution differences compared to addressing the same use cases with the DSE stack.

**Feasibility and Complexity:**  DynamoDB internally handles data differently than DSE. For a time series data set that is expected to grow over time to a size in the 100’s of terabytes and span multiple data centers, you may want to evaluate how to engineer your solution to consider the following:

* Data hot spots around the DynamoDB cluster
* Added data model complexity and potentially increased usage costs to meet query requirements with
  + secondary indexes
  + additional tables for data aging in time series data models
* Batch data movement across Amazon services to provide analytics and other functionality (example: batch data flow from DynamoDB to Redshift)
* Batch data movement across regions in the case you need a multi-datacenter or distributed data solution (DynamoDB does not replicate in real time across regions)
* Design changes may be required as you scale in order to lower costs of storing and accessing colder data while also making hotter data readily available with low latency and highly consistent operations across regions
* Increased provisioned throughput limits per region as you scale out with approval from Amazon support

**Technology Stack:** DynamoDB and DSE are not functionally equivalent. To meet most enterprise use cases and very large data volumes you may need to employ more Amazon capabilities than just DynamoDB, including some or all of:

* DynamoDB
* Redshift
* ElastiCache
* Kinesis
* AWS Data Pipeline
* 3rd party unsupported integrations such as Spark
* Batch data movement between different Amazon services and regions
* Infrastructure: A DynamoDB based solution cannot run on-premise, on different cloud providers, or a hybrid of both

**Pricing and Total Cost of Ownership (TCO):**  Amazon’s pricing model is a complex usage-based calculation where usage is measured in Write Capacity Units and Read Capacity Units.

* As you add more volume and more complex access patterns to the base DynamoDB tables such as secondary indexes, high consistency reads, large object operations, additional tables, etc, then multiple pricing units can be consumed per operation and/or total units consumed can increase, increasing unit cost and total cost unpredictably
* As you need to add additional services (Redshift, ElastiCache, Kinesis, etc) in order to meet your analytics, multi-data center, and other use cases, additional costs for those services may be incurred over and above the base data ingestion and query in DynamoDB

**Predictable TCO:** Given the above considerations, we would recommend that you arrive at a fairly complete system design using the Amazon stack, supported by volume growth estimates, before attempting to estimate the Amazon TCO with high confidence. A functionally broad proof of concept would be a wise investment in the case where you are seriously considering an Amazon solution.

Let me reiterate that we are not DynamoDB experts. The information above is based on our basic research into publicly available information and our knowledge of your use cases.