Press Release Addendum - Detailed Questions and Answers about DataStax Enterprise 6 New Features

1. Can you go into more detail on DSE 6 new features above and beyond operational simplicity and doubled performance.

Operational simplicity and doubled performance are the two big impact capabilities we are bringing to DSE 6 (see more technical information on these features in the response to Question 2). But there are several other features important to our enterprise customers. These additional features include:

- Advanced Apache Spark™ communication layer: delivers a number of improvements for analytical queries, including a 3x query performance increase.
- DataStax Bulk Loader: allows you to rapidly move data into and out of DSE up to 4x faster than other data loading utilities.
- AlwaysOn SQL: brings nonstop reliability to your existing investment in SQL tools and expertise for BI, ETL, and application data pipelines.
- Apache Spark Structured Streaming: enables simple, efficient, and robust streaming of data from Apache Kafka, file systems, or other sources.
- DataStax Studio 6 Notebook support for Spark SQL: a visual and intelligent query-builder interface for writing Spark SQL queries and reviewing and sharing the results.
- DSE Graph enhancements:
  - Advanced performance
  - Smart analytics query routing
  - Advanced schema management
  - Batch Fluent API Gremlin support
  - Support for TinkerPop 3.3.0
  - User experience enhancements to Studio
- Advanced Security additions (already has identity management, transparent data encryption, auditing, etc.)
  - Private Schemas: administrators now have more control over schema visibility. Administrators can control whether or not a user can see certain schema definitions, which can be especially helpful in securing multi-tenant applications. Private Schemas supports the principle of least privileges, which is key for meeting many security compliance standards.
  - Auditing by Role: We've enhanced auditing with the ability to audit changes and user activity by role. Traditionally, auditing in DSE was controlled by which respective database object you wanted to keep track of. Having role-based auditing greatly reduces the audit trail, since most administrators want to keep track of human activity rather than machine-generated activity.
  - Separation of Duties: There are many cases where administrators need full control of the database but should not have access or visibility to the data itself. For example, imagine a doctor or nurse who requires access to sensitive medical data. In this case, the administrator would still have the correct privileges to grant access to the doctor or nurse but the administrator would not be able to access the data.

2. You claim "half the latency for superior application responsiveness and twice the throughput and user capacity" compared with Apache Cassandra. Can you elaborate?

Apache Cassandra uses a traditional staged event-driven architecture (SEDA). With the SEDA architecture, Apache Cassandra assigns thread pools to events or tasks and connects them via a message service. This architecture also uses multiple threads per task, meaning that threads need to be coordinated. Additionally, events in this architecture are synchronous, which can cause contention and slow downs. Because of this, adding CPU cores eventually sees diminishing returns. DSE 6 utilizes a new coordination-free design — a thread-per-core architecture that yields incredible performance gains. Each node in a cluster owns part of the token range; that's not new. What's new is that a respective node's token range is divided evenly among CPU threads: one per CPU core to be exact. A respective

thread is now responsible for incoming writes for the part of the token range it owns, and any available thread can be used to handle read requests. This means that evenly distributed data in the cluster results in evenly distributed CPU utilization on a server. This architecture also means that very little coordination is needed between threads, which ensures that a CPU core can be used to its fullest capabilities. Since a single thread owns the writes for its respective token range, what about contention? In DSE 6, we've moved reads, writes, and other tasks from synchronous to asynchronous operations. This allows us to eliminate thread contention and always keep threads working. Combined with the thread-per-core architecture, this allows us to scale performance linearly as we scale the number of CPU cores. This is extremely important since multi-socket motherboards and high-core-count cloud instances have become the standard today.

Besides ingesting and serving data faster with thread-per-core, we've also made improvements to the storage engine that improve latency and optimize compaction, which can also be a bottleneck for write-heavy workloads.

3. What is DSE NodeSync and how does this ease operational management?

Operational simplicity is the biggest improvement here and very important to our enterprise customers. Apache Cassandra is the leading database technology at scale because it's masterless architecture ensures continuous availability great performance no matter where the millions of users are around the globe. (And we doubled that performance.)

But you must understand the Cassandra technology well to run it effectively. It's an open source technology, not an enterprise product. DSE 6, on the other hand, is designed to be operationally very simple to manage. That's achieved with three key features:
- **DSE NodeSync:** removes manual "repair" maintenance operations by keeping data synchronized in DSE clusters in a transparent and continuous fashion. NodeSync frees resources to allow you to focus on the core business, and by making applications run more consistently, it also improves planning predictability.
- **DSE Upgrade Service**: effortlessly handles patch upgrades by automatically notifying you that an upgrade is available, downloading the software you need, applying it to a cluster with zero downtime, and freeing you to do other things.
- **DSE TrafficControl:** delivers advanced resiliency that ensures DSE nodes stay online under extreme workloads.

That means that teams can focus on innovation and customer success instead of operations. And, ultimately, that translates to easy, enterprise-wide adoption for operational data management at scale. Something that's been elusive to digital transformation initiatives. Until now.

###