IST769 Unit G - Wide Column

## Agenda

1. Your Questions
2. Go over Problem Set
3. Unit Coursework Activities

## FRONT MATTER

Start your engines!

PS> **docker-compose down**

PS> **docker-compose up -d drill cassandra jupyter**

Get ready:

* 1. Open G-Wide-Column notebook and run the code to create the fudgemart\_order details table.
  2. Open CQL command line

PS> **docker-compose exec cassandra cqlsh**

**EVENTUAL CONSISTENCY**

**WIDE-COLUMN database**

**Cassandra:**

* What is the storage model? (relational, document, etc…) **wide column**
* What are the logical model metaphors? (tables, keys constraints, etc…)
  + Tables, keys, no constraints, no joins,
* How does it scale horizontally? CAP? (after all - its a big data course!)
  + Partition key
* Purpose of database? OLTP (CRUD) or OLAP/Analytics (CR)?\
  + Guaranteed writes and table based-analytics
* How do you design for it?
* How to connect to it from:
  + Client / Native Scripting Language
  + Drill / SQL
  + Spark / PySpark

## 1. Your Questions

* Why does Cassandra limit the usage of multi-valued output conditions, such as > (greater than) or < (less than), when querying on a secondary index compared to the primary index? Is this limitation specific to secondary indexes, or does it apply to both primary and secondary indexes both? Additionally, why is an exact match required when querying on indexed columns? Is this limitation of Cassandra
  + Since there is no master, cassandra needs to be able to determine which node has the data before it can query successfully.
* ??

## 2. Coursework Activities

1. Looking at the notebook code at the bottom: gdemo.fudgemart\_order\_details What is the:
   1. Row Key?
   2. Cluster Key?
   3. Partition Key?
2. Why is the write mode “append” and not “overwrite” change it and see what happens.
3. Write a CQL to get the customer email and name for customer ID 4.   
   Does this require ALLOW FILTERING? Why or why not?

Select customer\_id, customer\_email, customer\_name from

Fudgemart\_order\_details where customer\_id = 4

1. Write a CQL to get the names quantities and prices of items on order 10  
   Does this require ALLOW FILTERING? Why or why not?
2. Write a CQL to get the names quantities and prices of items, along with customer name and email on order 5 for customer id 4   
   Does this require ALLOW FILTERING? Why or why not?