Hadoop World 2013 - Day 1

Many thanks to our Cloudera account team who set us up with multiple passes to the 2013 Strata + Hadoop World conference. We have a strong presence here with Jeff Woods, Keith Manthey, Jayakumar Duraisamy, Yuvaraj Sankaran, and myself making the trip. To make sure we would be *bright-eyed and bushy-tailed* for Day 1 of the conference here in New York City, we flew in on Sunday night. Today was focused on tutorials. Yuvaraj and I both attended a session in the morning that was focused on building a comprehensive solution that spanned the very common use case of a data ingestion pipeline. We leveraged HDFS, AVRO, Flume, MapReduce, HCatalog, Hive, Impala, and Oozie. Cloudera's "sugar" was their Cloudera Development Kit (CDK) which introduced a more generalized framework called Crunch that is supposed to make MapReduce programming "easier" for "business programmers". We went pretty fast today – I'll have to look at it harder before I weigh in on it. All of the CDK is available on github.

For the afternoon, I was queued up to attend the session on integrating Apache Cassandra with MapReduce, but it was overflowing at the seams. The thought of standing in the back shoulder to shoulder with 40 of my "newest friends" while trying to reproduce the labs on my laptop drove me to make an adjustment (who hasn't had this happen at a conference). A quick look at the available topics drove me to the Natural Language Processing (NLP) session instead. Very good forward looking info for us as while using these approaches could help us understand sentiment, someday we might even imagine a product that we build that leverages NLP as well data from partners we might not even know about yet.

For dinner, several of us attended Cloudera's Data Impact Awards. During the meet-n-greet, Yuvaraj and Jayakumar got to meet the legendary Doug Cutting. **Yuvaraj, don't forget to post your picture!!** Also, I have to admit that I found out that Stephen Leitner has, in fact, been pronouncing Hadoop correctly all along. *I got it straight from Doug and after all since it was his son who name that silly little stuffed elephant Hadoop in the first place, then he ought to know*. We were fortunate enough to have Jeff Hammerbacher sitting at our table. Prior to becoming a Co-Founder and Chief Scientist at Cloudera, Jeff built and lead the data team at Facebook. It was a great, and insightful, conversation and so glad for the opportunity to meet him as well as folks from other companies using Hadoop in their enterprises.

All in all, a great first day. **Looking forward to the next two!**