

Understanding the CCSQ Data & Analytics Environment

1. **From a high level view, how will the new notebook alternative tool affect the architecture?**

A - The notebook tool would be replacement for Zeppelin, and that would be the tool that folks can do their development in Python, Java, and Scala to run their to run the Spark code. Then additionally, the decision on the new platform would essentially replace the current CDR Hive clusters that we have, instead of operating on a single large cluster that all groups are connecting to. Folks would have access to personal compute clusters in order to isolate their jobs from from one another, and be able to access resources that can auto scale versus the static cluster that we have today.

2. **In the current use case, what is the file storage used with Apache spark?**

A - Spark is accessing data via hive and the data is stored in S3 buckets.