Loading and Analyzing IBM dashDB data in IBM Analytics for Apache Spark

Step	Audio Script	What's happening on the screen
1	This video will show you how to create a connection to dashDB data in IBM Analytics for Apache Spark. You'll then see how to use a Scala notebook for easy access to interactive visualizations and analysis of the dashDB data using Spark SQL.	Welcome screen displays with copyright statement and link to text version.
2	Let's begin in IBM Analytics for Apache Spark. First, you'll need to create a data connection to your dashDB instance on Bluemix. Provide a name and description for the connection. You can connect to a Bluemix service or an external service. In this case, select IBM Bluemix, then select the dashDB instance and database from the lists before completing the connection.	1. The IBM Analytics for Apache Spark dashboard displays. 2. Click Data. 3. Click Create A Connection. 4. Type a name and description for the connection. 5. Select the service type which is IBM Bluemix in this case. 6. From the Instance drop-down box, select the dashDB instance. 7. Select the database which is BLUDB in this case. 8. Click Create Connection.
3	Next, on the Analytics tab, open the existing Spark instance. On this screen, you can create new notebooks using either the scala or python kernel. In this case, we'll create a new scala notebook.	1. Click the Analytics tab. 2. Click the existing instance. 3. Click New Notebook. 4. Type ScalaSparkdashDBNotebook for the notebook name. 5. (Optional) Type a description. 6. Click Create Notebook.
4	First, add dashDB as a data source to this notebook.	1. On the right side of the screen, click Data Source. 2. Click Add Source. 3. Select dashDB, and click Add Data Source. 4. Highlight dashDB in the Data Source side panel.
5	The notebook is in the main part of the screen, and that's where you'll add a number of commands to load and analyze the data. In this example, you'll see how to load a table from a dashDB database and retrieve the information for a specific sales associate.	Zoom in on the notebook area.
6	This first command contains SQLContext which is the entry point into all functionality in Spark SQL and is necessary to execute SQL queries.	1. Paste the first statement: val sqlContext = new org.apache.spark.sql.SQLContext(sc) 2. Click Run.
7	This second command loads the sales table from the dashDB account and assigns it to the dashdata variable. The url and account information comes directly from the dashDB data source.	1. Paste the second statement: val dashdataDF = sqlContext.load("jdbc", Map("url" -> "jdbc:db2://bluemix05.bluforcloud.com:50000/BLUDB:user=<account>;password=<password>;", "dbtable" -> "dash013382.SALES")) 2. Click Insert to code. Highlight the URL and credentials. 3. Delete the inserted code.

Step	Audio Script	What's happening on the screen
8	This third command takes the dashdata and registers it as a table called salesdata.	1. Paste the third statement: dashdata.registerTempTable("salesdata") 2. Click Run.
9	This next command lets you take a look at the schema.	1. Paste the fourth statement: dashdata.printSchema 2. Click Run. 3. The schema displays.
10	Next, the dashdata.collect command shows the data collected for each sales associate.	1. Paste the fifth statement: dashdata.collect 2. Click Run. 3. The data displays.
11	Once the data is registered as a table, you can use SQL to process the data. This next line returns the SALESMAN_ID data from the salesdata table.	1. Paste the sixth statements: val results = sqlContext.sql("SELECT SALESMAN_ID from salesdata") results.collect 2. Click Run. 3. The SALESMAN_ID data displays.
12	Or you could retrieve the information for a specific sales person by ID.	1. Paste the seventh statements: val salesman = sqlContext.sql("select * from salesdata where SALESMAN_ID='NC100 '") salesman.collect 2. Click Run. 3. The information for the sales person with ID NC100 displays.
13	You can load data from other Bluemix services, such as IBM Cloudant. We hope this video has shown you how easy it is to load and analyze dashDB data in IBM Analytics for Apache Spark. Find more videos in the Spark Learning Center at http://developer.ibm.com/clouddataservices/spark.	Closing animation displays with video title, copyright statement, IBM Analytics for Apache Spark logo, and a link to the Spark Learning Center.