

Self Service BI in 10 Minutes (Client Setup)

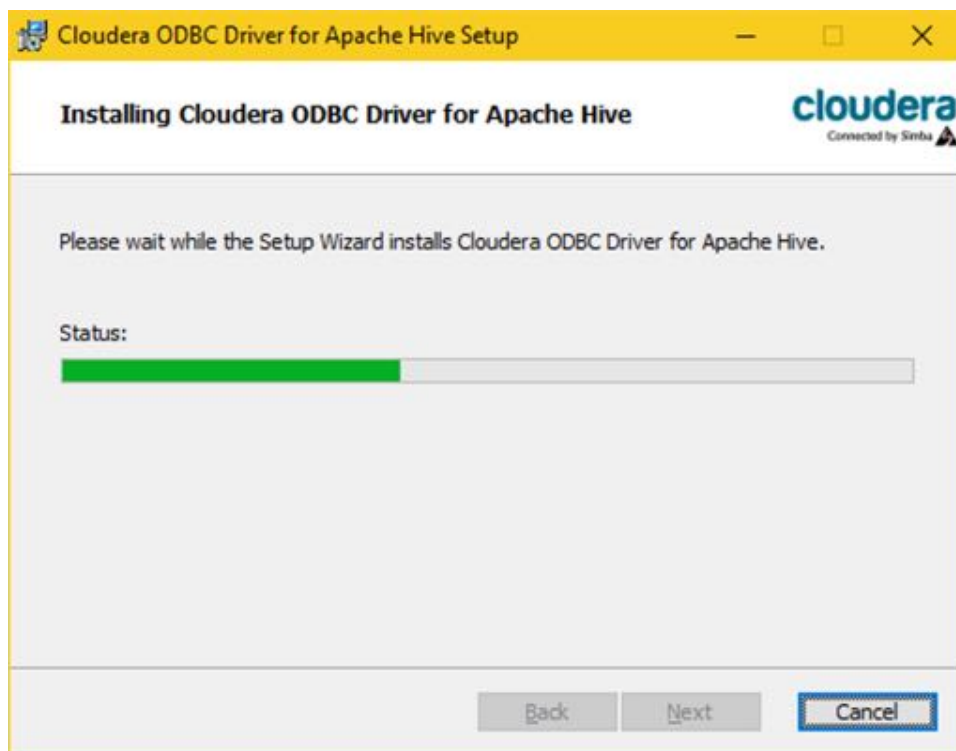
Monday, February 20, 2017 11:30 AM

Download Hive ODBC connector from

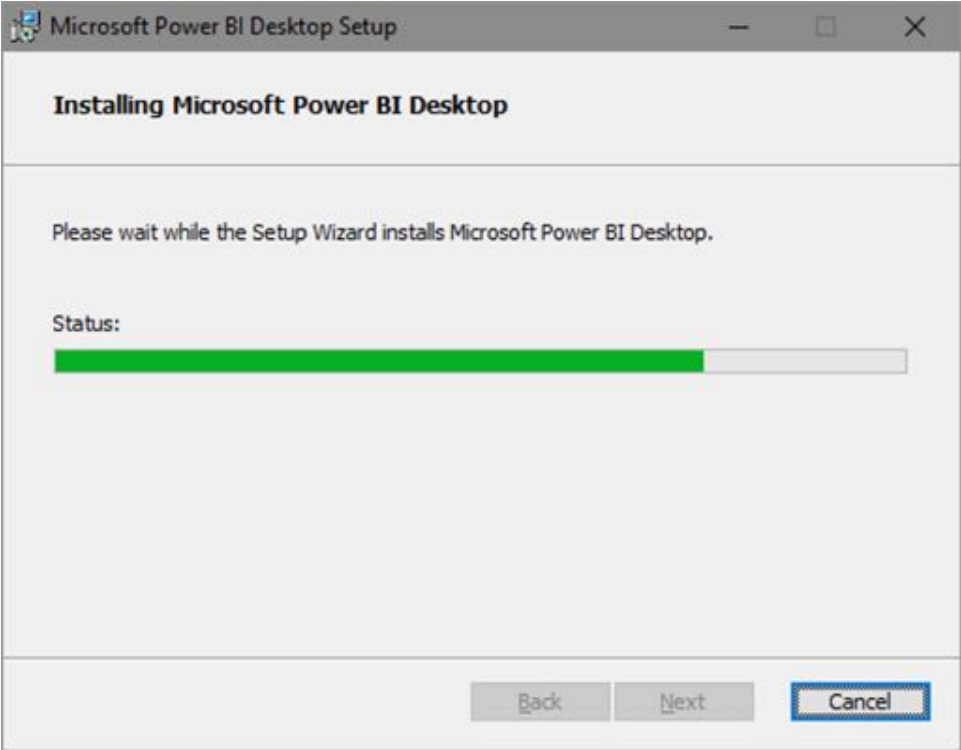
<https://www.cloudera.com/downloads/connectors/hive/odbc/2-5-17.html>.

Specifically choose the 32 bit version 2.5.0 version of the driver.

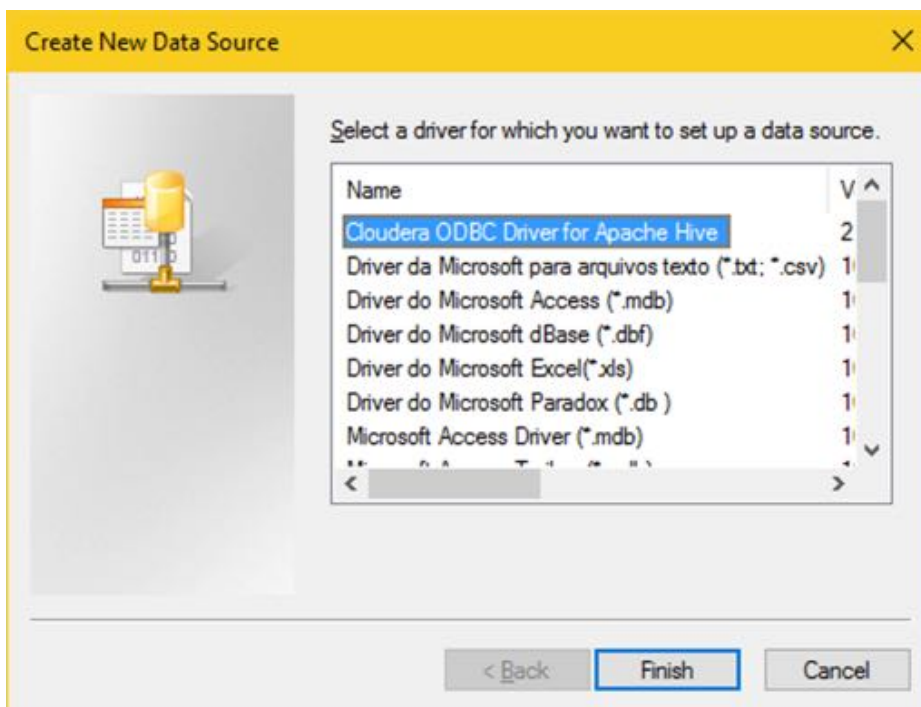
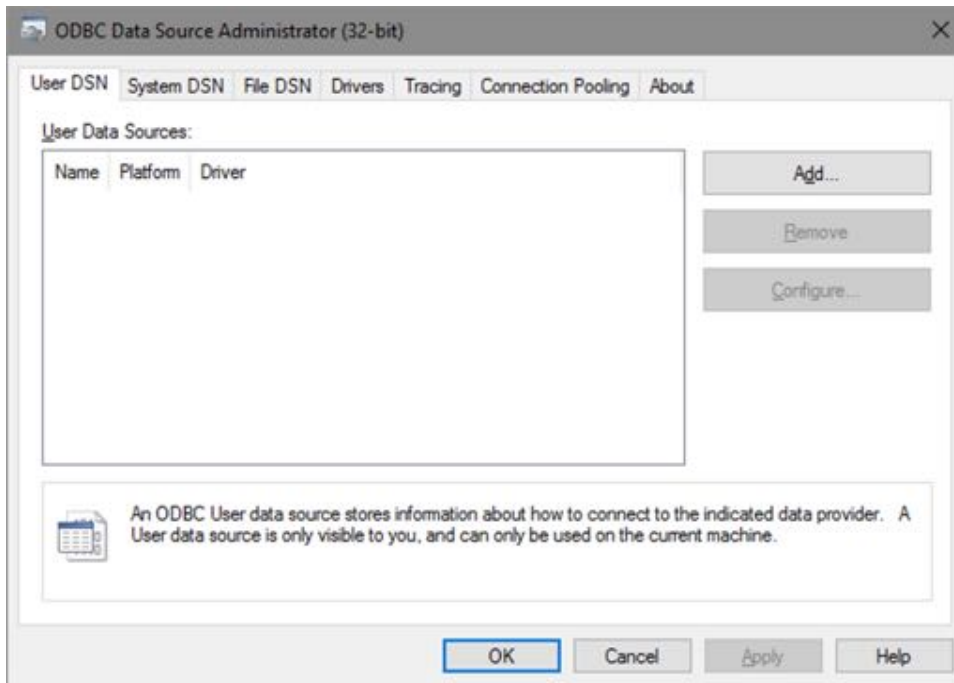
Install ODBC Connector



Download and install Power BI desktop from <https://www.microsoft.com/en-us/download/details.aspx?id=45331>



Create a new ODBC Connection with the Spark Thrift Server.



Configure the data source with SASL connection to the thrift host. The username is ignored mostly but necessary to establish connection.

The screenshot shows the 'Cloudera ODBC Driver for Apache Hive DSN Setup' dialog box. It contains the following fields and options:

- Data Source Name: Spark Data Lake
- Description: Spark Data Lake
- Host: 192.168.1.101
- Port: 10000
- Database: default
- Hive Server Type: Hive Server 2 (dropdown)
- Authentication section:
 - Mechanism: User Name (dropdown)
 - Realm: (empty)
 - Host FQDN: (empty)
 - Service Name: (empty)
 - User Name: user
- Advanced Options... button
- Version: v2.5.0.1001 (32 bit)
- Buttons: Test, OK, Cancel

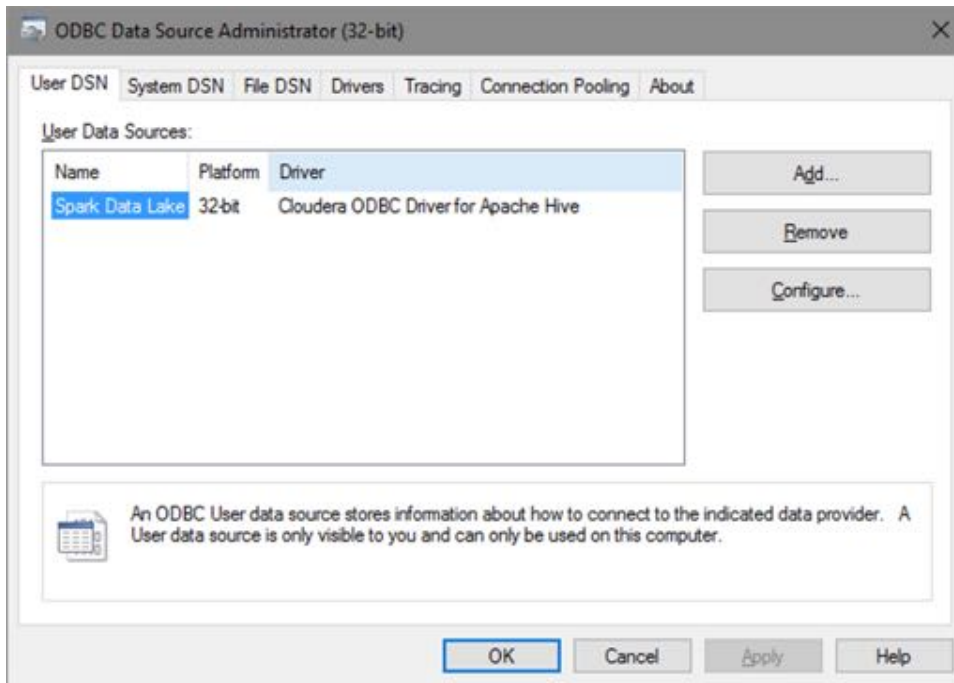
Test to make sure the driver connection is setup correctly

The screenshot shows the 'Cloudera ODBC Driver for Apache Hive Data Source Test' dialog box. It displays the following test results:

- Driver Version: V2.5.0.1001
- Running connectivity tests...
- Attempting connection
- Connection established
- Disconnecting from server
- TESTS COMPLETED SUCCESSFULLY!

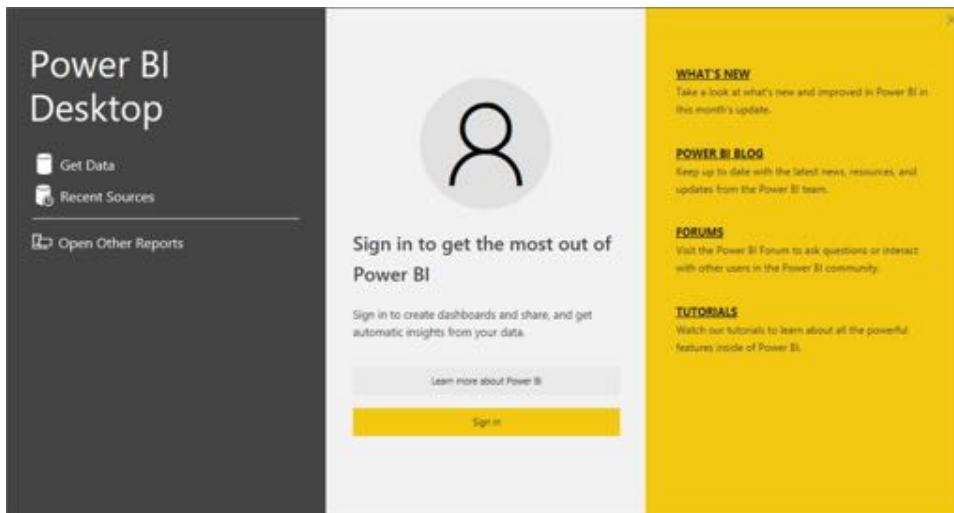
An OK button is visible at the bottom of the dialog.

Remember the name of the ODBC data source -- here "Spark Data Lake"

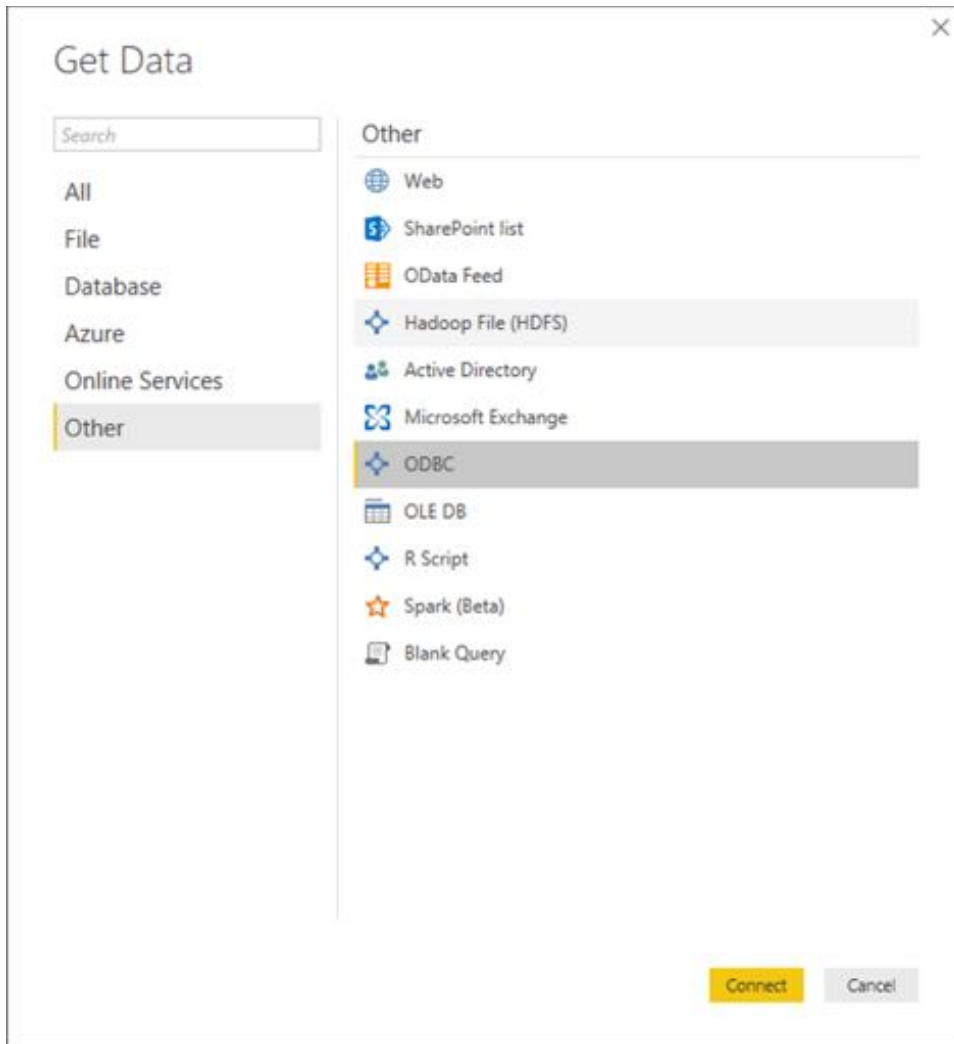


Build the BI.

Open Power BI. See <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-getting-started/> for more details



Choose to Get Data... and Select the ODBC source.



Specifically choose the Data Lake DSN you created in the ODBC Control Panel.



Since we already know the table name ("sales_sample") in the data lake, we might as well choose to bring the data a priori into the PowerBI desktop. Use the SQL statement to select the data

From ODBC

Data source name (DSN)
 Spark Data Lake

Advanced options
 Connection string (non-credential properties) (optional) ⓘ

SQL statement (optional)
 select * from sales_sample

OK Cancel

Verify the data integrity.

ODBC (dsn=Spark Data Lake select * from sales_sample)

Row_ID	Order_ID	Order_Date	Order_Priority	Order_Quantity	Sales	Discount	Ship_Mode	Profit	Unit_Price	Shipping
3817	27201	1.34317E+18	High	45	3500.1	0.01	Delivery Truck	-2661.318	70.89	
3897	27808	1.26446E+18	Low	9	196.12	0.01	Express Air	4.41	20.89	
4173	29573	1.32667E+18	Critical	43	792.21	0.05	Regular Air	42.24	18.97	
4203	29862	1.30844E+18	Medium	34	214.64	0.08	Regular Air	-76.88	6.48	
4276	30433	1.27604E+18	Medium	17	253.38	0	Express Air	-35.78	13.9	
4284	30532	1.32227E+18	High	28	431.37	0.03	Regular Air	-28.6695	15.22	
4285	30532	1.32227E+18	High	23	262.31	0.04	Express Air	27.55	11.09	
4286	30532	1.32227E+18	High	45	514.03	0.05	Regular Air	82.19	11.34	
4287	30532	1.32227E+18	High	9	1696.7	0.04	Delivery Truck	-82.81	180.98	
4297	30597	1.34412E+18	High	18	93.02	0.07	Regular Air	20.01	5.08	
4298	30597	1.34412E+18	High	42	639.84	0.03	Regular Air	-95.76	15.14	
4349	30581	1.29773E+18	Low	29	7837.44	0.1	Delivery Truck	1817.76	279.81	
4350	30981	1.29773E+18	Low	22	3653.22	0.09	Delivery Truck	-116.02	180.98	
4455	31751	1.30637E+18	High	31	413.12	0.1	Express Air	-10.2235	14.27	
4473	31873	1.25764E+18	Low	2	141.59	0.05	Regular Air	-91.09	70.97	
4474	31873	1.25764E+18	Low	14	2465.5015	0.06	Regular Air	13.014	205.99	
4475	31873	1.25764E+18	Low	50	8558.4715	0.06	Regular Air	2539.458	205.99	
4476	31873	1.25764E+18	Low	42	236.89	0	Regular Air	-125.29	5.28	
4492	32000	1.34749E+18	Low	49	162.16	0.05	Regular Air	24.95	3.29	
4493	32000	1.34749E+18	Low	26	15168.82	0.02	Delivery Truck	-1096.776	550.98	

Load Edit Cancel

Compute any transformations needed from within the PowerBI platform.

Row_ID	Order_ID	Order_Date	Order_Priority	Order_Quantity	Sales	Discount	Ship_Mode	Profit	Unit_Price	Shipping_Cost	Customer_Name	Province
415	2104	20180101000000000000	Standard	21	116.16	0.01	Regular Air	149.78	6.49	8.89	Tim Inzer	New Scotia
538	5980	20180112000000000000	Not Specified	27	277.92	0.01	Regular Air	-52.78	6.49	8	Dean Henry	New Scotia
5392	22182	20180101000000000000	Not Specified	33	328.83	0.04	Regular Air	-278.85	6.49	8.34	Scott Stearns	New Scotia
6870	48983	20180101000000000000	Not Specified	33	82.87	0.07	Regular Air	-51.14	6.49	8.94	David Fleming	New Scotia
6966	49594	20180101000000000000	Low	33	227.87	0.02	Regular Air	126.03	6.49	8.4	Sarah Leung	New Scotia
4402	52192	20180101000000000000	High	23	128.96	0.2	Regular Air	-42.81	6.49	8.28	George Debraze	New Scotia
470	478	20180101000000000000	Critical	29	188.38	0.06	Regular Air	-119.32	6.49	7.49	Michelle Tran	Quebec
278	1892	20180101000000000000	Low	9	64.29	0.01	Regular Air	-18.83	6.49	8.4	Jan Kamboonsa	Ontario
3387	23919	20180101000000000000	Medium	18	122.02	0.04	Regular Air	-121	6.49	10.89	Pete Talarico	Ontario
3388	57893	20180101000000000000	High	28	183.38	0.04	Regular Air	-49.34	6.49	8.4	Steven Beale	Ontario
5742	46380	20180101000000000000	Not Specified	41	288.52	0.01	Regular Air	175.29	6.49	7.92	Jack Lee	Ontario
4864	48902	20180101000000000000	Medium	7	33.29	0.02	Regular Air	-28.89	6.49	7.69	Bergamot-Patterson	Ontario
889	48111	20180101000000000000	Not Specified	41	277.91	0.02	Regular Air	189.39	6.49	8.97	Walter Prince	Ontario
5217	23189	20180101000000000000	Critical	27	272.7	0.07	Regular Air	-43.97	6.49	8.11	Van Brown	Ontario
5218	23190	20180101000000000000	Not Specified	89	228.42	0.01	Regular Air	120.39	6.49	8.32	Steph-Anne	Quebec
5274	23396	20180101000000000000	Medium	19	221.2	0.06	Regular Air	-75.44	6.49	7.97	Cass Adams	Quebec
4512	52191	20180101000000000000	Medium	3	29.81	0.02	Regular Air	-33.88	6.49	8.22	Vivik Lunde-wan	Quebec
3507	38619	20180101000000000000	High	2	21.02	0.01	Regular Air	-13.12	6.49	8.23	Tom Sherry	Quebec
9429	45676	20180101000000000000	Low	2	26.34	0.06	Regular Air	-13.32	6.49	7.86	Betsy Frank	Quebec
6882	49018	20180101000000000000	High	40	368.97	0.01	Regular Air	208.38	6.49	8.4	Cathy Huang	Quebec
5281	23192	20180101000000000000	Critical	36	233.38	0.2	Regular Air	180.17	6.49	8.79	Alan Donnelly	Quebec
5149	23188	20180101000000000000	Critical	19	127.48	0.01	Regular Air	-41.89	6.49	8.84	Tracy Foster	Quebec
4023	47129	20180101000000000000	Medium	29	290.7	0.02	Regular Air	147.72	6.49	8.17	Alan Daykin	Quebec
3869	56382	20180101000000000000	Not Specified	34	223.76	0.01	Regular Air	128.69	6.49	7.92	Julia Liu	Quebec
8342	58619	20180101000000000000	Not Specified	31	221.4	0.08	Regular Air	-101.48	6.49	8.4	Walter	Quebec
112	768	20180101000000000000	Not Specified	22	289.91	0.08	Regular Air	99.88	6.49	8.14	Caroline	Quebec

Now that the canvas is primed with data, drag-n-drop various visualizations, measures and filters for rendering a compelling BI dashboard.

Explore Rich Interactive Visualizations

