

# 10x Easier Development Using Databricks Connect



---

**Martin Grund**

Principal Software Engineer, Databricks

**Vladislav Mantic Lugo**

Product Manager, Databricks

# Product safe harbor statement

This information is provided to outline Databricks' general product direction and is for informational purposes only. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

# Who prefers to develop in:

1) a Notebook?

2) an IDE?

## And why?



# IDE-BASED DEVELOPMENT W/ DATABRICKS



Why do Databricks users love developing in an IDE?

- Meet developers where they are... **muscle memory!**
- **Software engineering best practises:** debugging, code modularization, unit testing, etc.



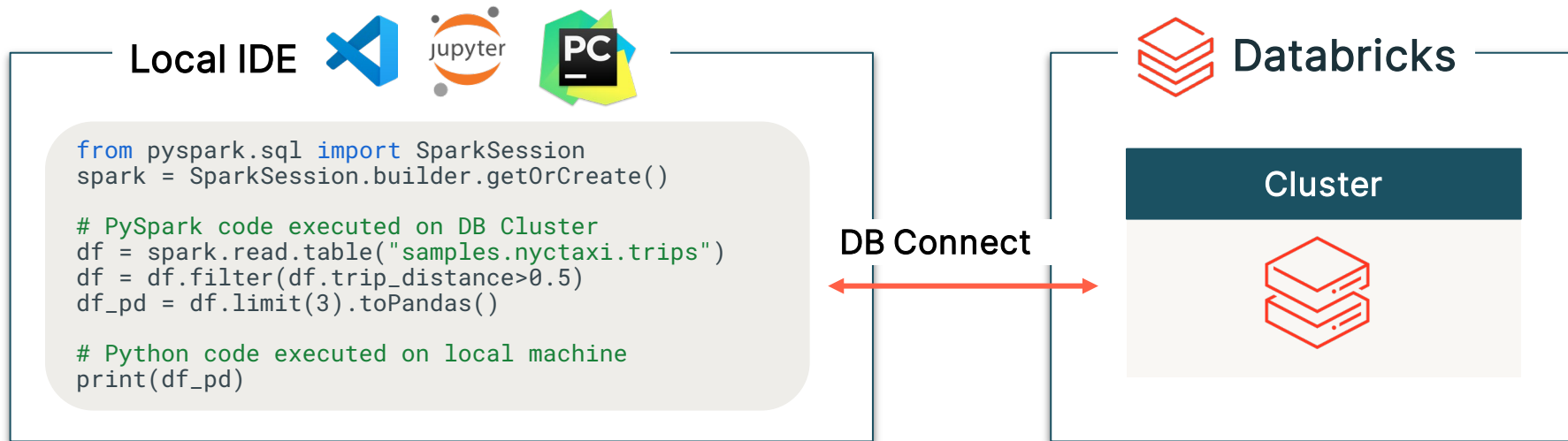
Why users love DB Connect:

- With **any IDE,**
- **Interactive debugging** of Spark code on Databricks cluster,
- High **production fidelity:** debug & run on target platform,
- **Application development:** Use DB Connect as a “Python-native JDBC driver”.

# WALK-THROUGH

# HOW TO USE DB CONNECT AND PYCHARM

# DEVELOPMENT FROM THE IDE WITH DB CONNECT



# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- 1. I create a cluster

The screenshot shows the 'New compute' configuration page for a cluster named 'My Cluster'. The page is divided into several sections, with several key options highlighted by red boxes:

- Access mode:** Set to 'Shared'.
- Performance:** Databricks runtime version is 'Runtime: 14.3 LTS (Scala 2.12, Spark 3.5.0)'. Custom Spark Version is '14.3.x-scala2.12'.
- Worker type:** Set to 'Compute Intensive Instance Pool Standard\_D16s\_v3'. Min workers is 2, Max workers is 8.
- Tags:** A section for adding tags with 'Key' and 'Value' input fields and an 'Add' button.
- Advanced options:** A section for additional configuration options.

Other visible settings include 'Using compute policy: Interactive Cluster Policy', 'Multi node' selected, 'Use Photon Acceleration' checked, and 'Terminate after 60 minutes of inactivity'.

**Summary:**

2-8 Workers	128-512 GB Memory 32-128 Cores
1 Driver	64 GB Memory, 16 Cores
Runtime	14.3.x-scala2.12

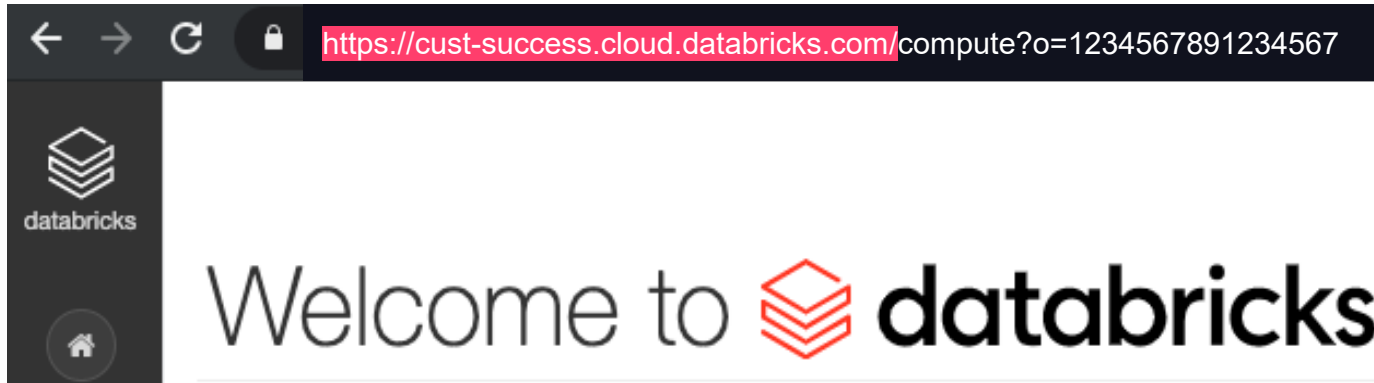
Additional tags: Unity Catalog, Photon, Standard\_D16s\_v3, 18-54 DBU/h

Buttons at the bottom: Create compute, Cancel

# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- ✓ 1. I create a cluster
- > 2. I copy my workspace URL





# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- ✓ 1. I create a cluster
- ✓ 2. I copy my workspace URL
- > 3. Install Python package. `pip install databricks-connect==14.3`

```
$ python3.10 -m venv ./venv
$ source ./venv/bin/activate
(venv)$ pip3 install --upgrade "databricks-connect==14.3.*"
```

# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- ✓ 1. I create a cluster
- ✓ 2. I copy my workspace URL
- ✓ 3. Install Python package, `pip install databricks-connect==14.3`
- > 4. Create a config profile (workspace & cluster ID, auth token)

```
$ brew install databricks
```

```
$ databricks auth login --configure-cluster --host https://cust-success.cloud.databricks.com
```

- Choose profile name: DEFAULT
- Authenticate with OAuth
- Select cluster -> cluster ID

# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- ✓ 1. I create a cluster
- ✓ 2. I copy my workspace URL
- ✓ 3. Install Python package, `pip install databricks-connect==14.3`
- ✓ 4. Create a config profile (workspace & cluster ID, auth token)

```
$ brew install databricks
```

```
$ databricks auth login --configure-cluster --host https://cust-success.cloud.databricks.com
```

- > 5. Open my IDE and I create my Spark session using a config profile

```
from databricks.connect import DatabricksSession
```

```
spark = DatabricksSession.builder.profile("DEFAULT").getOrCreate()
```

# DEVELOPMENT FROM THE IDE WITH DB CONNECT

## Install DB Connect and configure my Spark session

- ✓ 1. I create a cluster
- ✓ 2. I copy my workspace URL
- ✓ 3. Install Python package, `pip install databricks-connect==14.3`
- ✓ 4. Create a config profile (workspace & cluster ID, auth token)

```
$ brew install databricks
```

```
$ databricks auth login --configure-cluster --host https://cust-success.cloud.databricks.com
```

- ✓ 5. Open my IDE and I create my Spark session using a config profile

```
from databricks.connect import DatabricksSession
```

```
spark = DatabricksSession.builder.profile("DEFAULT").getOrCreate()
```

- > 6. Develop, debug and run my *application!*

```
df = spark.read.table("samples.nyctaxi.trips")
```

```
df.where(df.trip_distance < 10).show(10)
```

# DEVELOPMENT FROM THE IDE WITH DB CONNECT

The image shows a code editor window with a Python file named `main.py`. The code is divided into two steps:

```
12  
13 # Step 1 - Load the Bronze Tables  
14 print("Step 1 ....")  
15 table_name_bronze = "main.martingrund.taxi_demo_bronze" table_name_bronze: 'main.martingrund.taxi_demo_bronze'  
16 df = trans.load_data(spark) df: DataFrame[vendor: string, pickup_datetime: timestamp, dropoff_datetime: times...  
17 trans.save_data(df, table_name_bronze)  
18  
19 # Step 2 - Load the Silver Tables  
20 print("Step 2 ....")  
21 df = spark.table(table_name_bronze)  
22 df = trans.filter_columns(df)  
23 df = trans.transform_columns(df)  
24 df = trans.filter_invalid(df)  
25  
26 table_name_silver = "main.martingrund.taxi_demo_silver"  
27 trans.save_data(df, table_name_silver)  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000
```

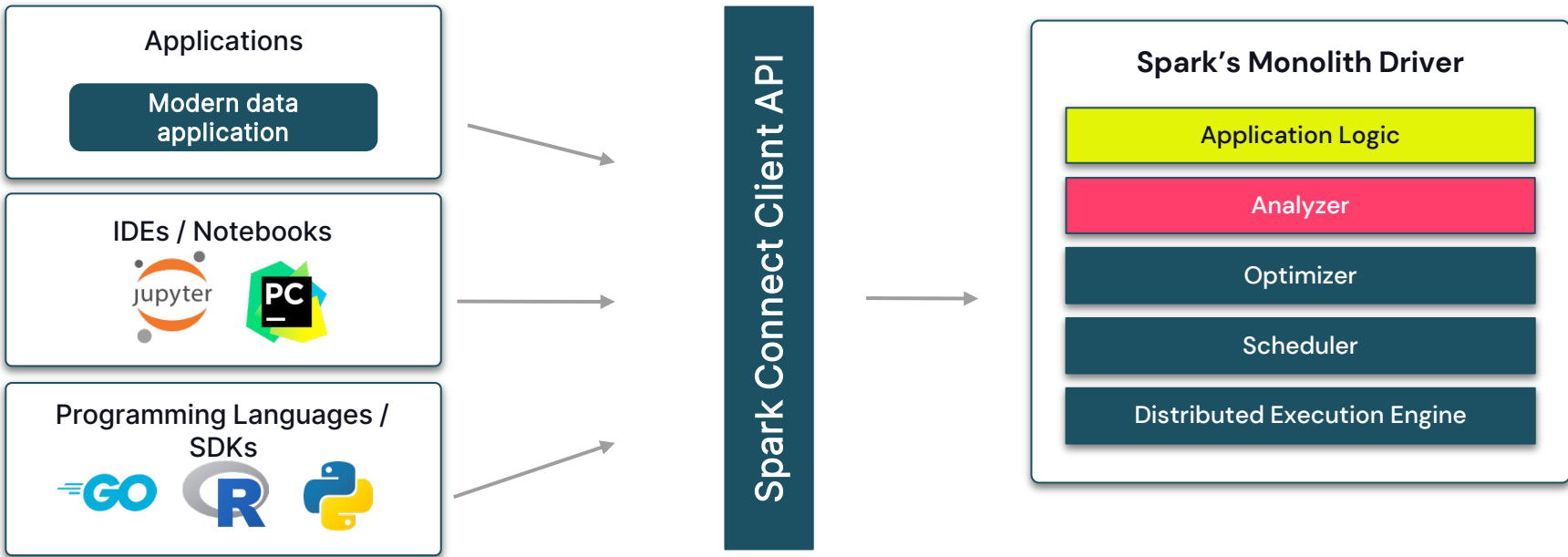
The debugger window shows the following variables:

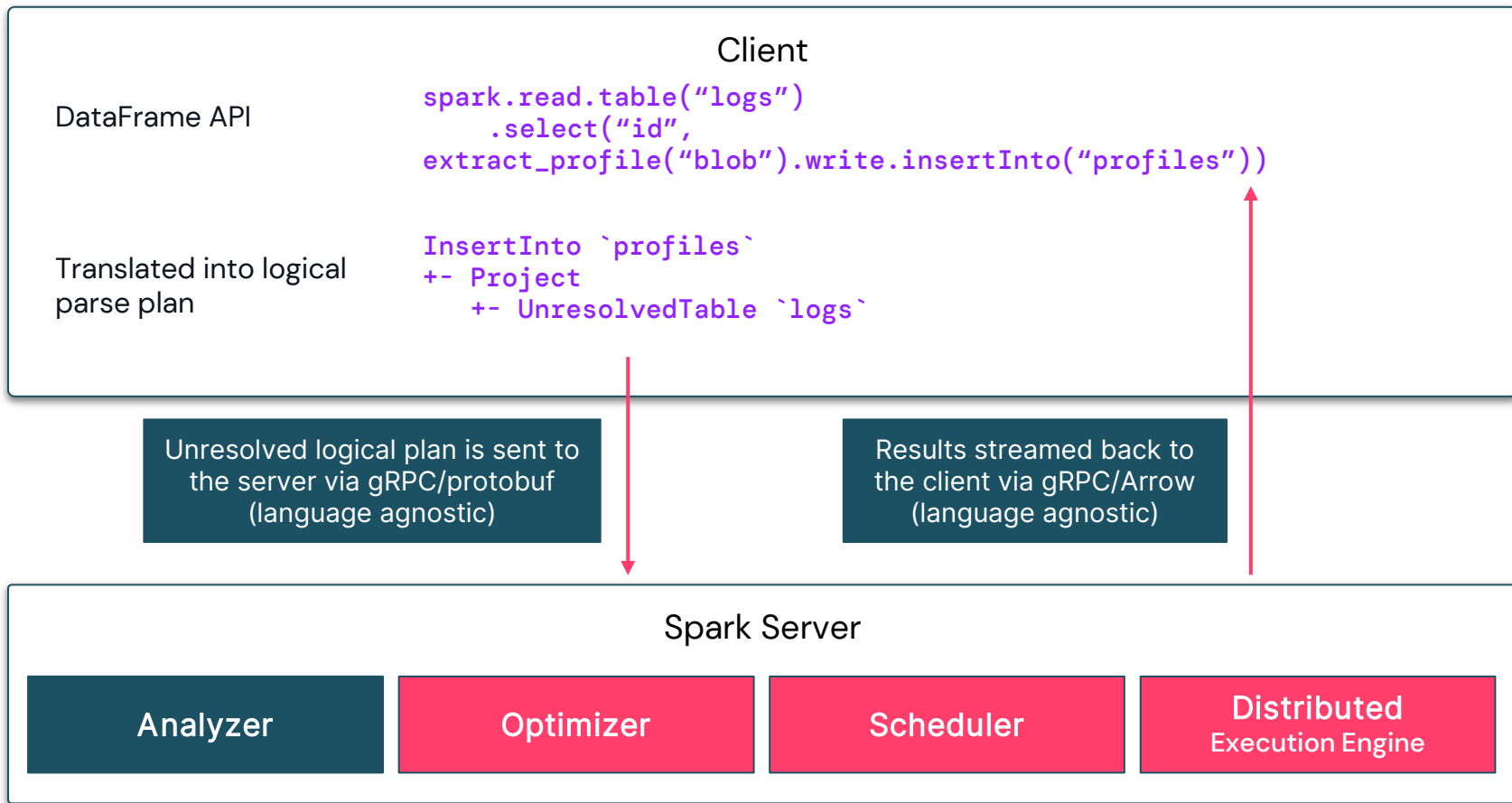
- `c = (Config) <Config: host=https://e2-dogfood.staging.cloud.databricks.com, token=***, profile=integ, auth_type=pat, c...`
- `df = (DataFrame) DataFrame[vendor: string, pickup_datetime: timestamp, dropoff_datetime: times...View as DataFrame`
- `spark = (SparkSession) <pyspark.sql.connect.session.SparkSession object at 0x108c60f70>`
- `table_name_bronze = (str) 'main.martingrund.taxi_demo_bronze'`

# BUT HOW DOES IT WORK

# SPARK CONNECT

Remote Connectivity: thin client, full power of Apache Spark

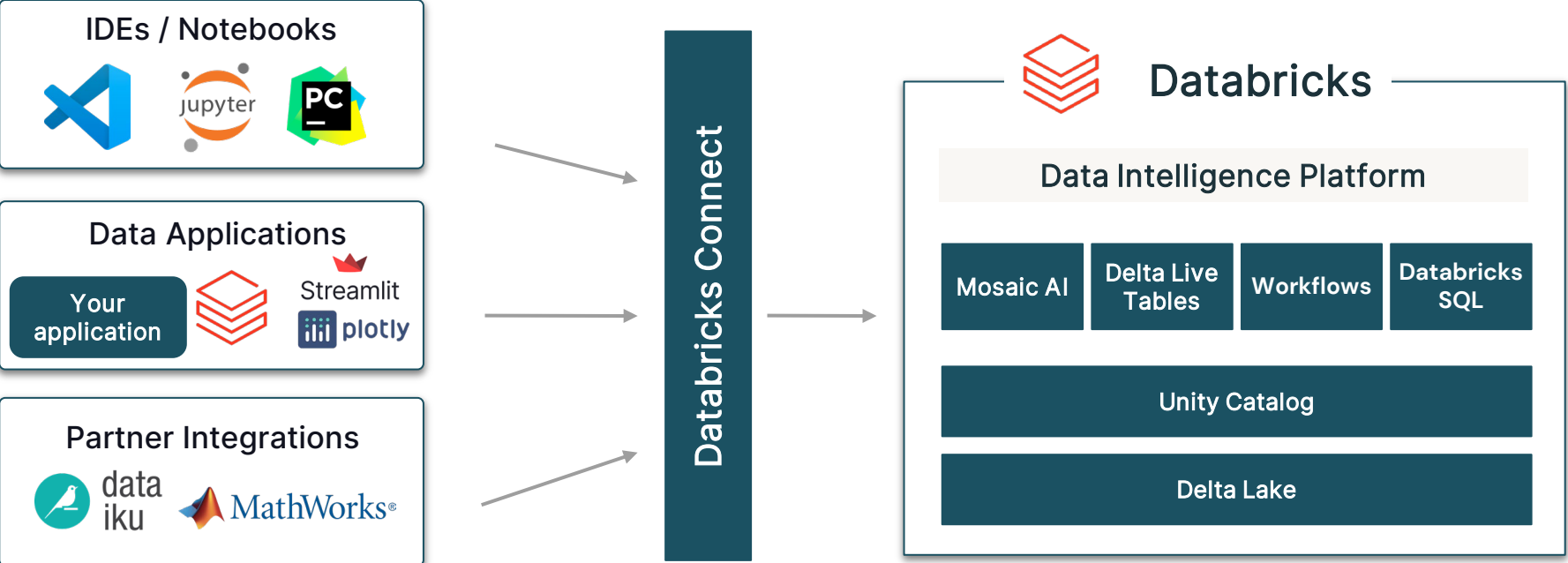






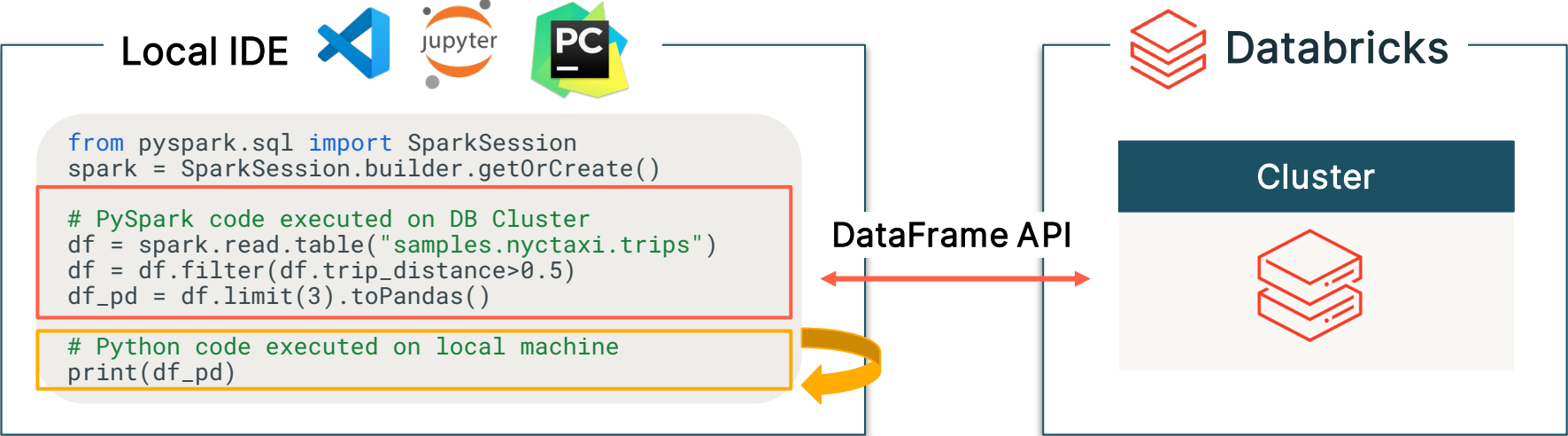
# DATABRICKS CONNECT USES SPARK CONNECT

Connect to Databricks from anywhere!



# HOW DATABRICKS CONNECT WORKS

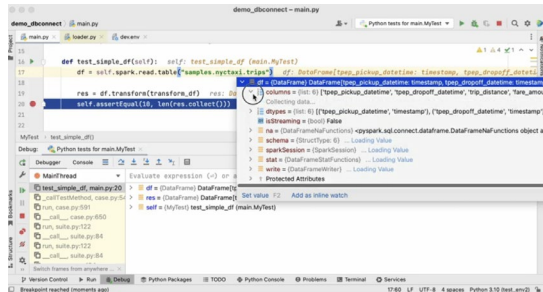
## Build on Spark Connect



# USE DATABRICKS FROM ANYWHERE!

It's more than debugging: Remote Connectivity to Databricks

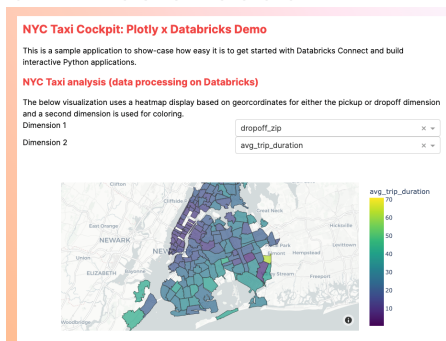
Interactively develop & debug from IDE



Dataframe API (Python, Scala, R)  
UDFs, Streaming

[ETL github example](#)

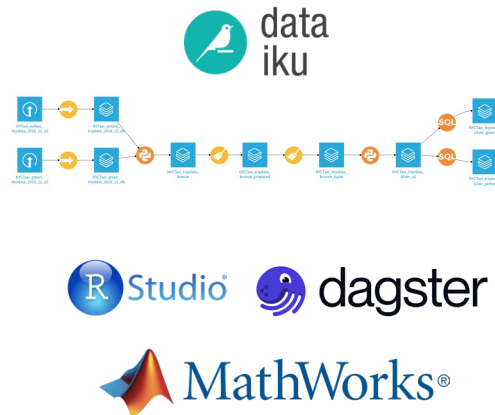
Interactive Data Apps with a few lines of code!



Databricks Lighthouse APPs!

[Plotly app github example](#)

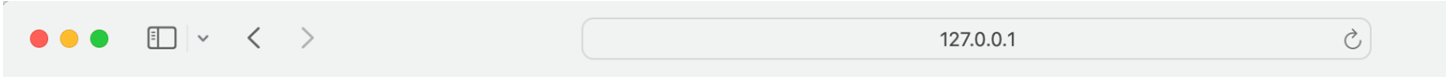
Partner integrations made easy!



[Dataiku Demo](#)



# YOUR APPS WITH DB CONNECT



	index	tpep_pickup_datetime	tpep_dropoff_datetime	trip_distance	fare_amount	pickup_zip	dropoff_zip
0	0	2016-02-16 22:40:45	2016-02-16 22:59:25	5.35	18.5	10003	11238
1	1	2016-02-05 16:06:44	2016-02-05 16:26:03	6.50	21.5	10282	10001
2	2	2016-02-08 07:39:25	2016-02-08 07:44:14	0.90	5.5	10119	10003
3	3	2016-02-29 22:25:33	2016-02-29 22:38:09	3.50	13.5	10001	11222
4	4	2016-02-03 17:21:02	2016-02-03 17:23:24	0.30	3.5	10028	10028
5	5	2016-02-10 00:47:44	2016-02-10 00:53:04	0.00	5.0	10038	10005
6	6	2016-02-19 03:24:25	2016-02-19 03:44:56	6.57	21.5	10001	11377
7	7	2016-02-02 14:05:23	2016-02-02 14:23:07	1.08	11.5	10103	10167
8	8	2016-02-20 15:42:20	2016-02-20 15:50:40	0.80	7.0	10003	10011
9	9	2016-02-14 16:19:53	2016-02-14 16:32:10	1.30	9.0	10199	10020
10	10	2016-02-16 21:01:22	2016-02-16 21:11:29	2.74	11.0	10002	11211
11	11	2016-02-22 20:13:11	2016-02-22 20:27:35	2.09	11.5	10012	10009
12	12	2016-02-12 11:35:59	2016-02-12 11:50:53	2.80	13.5	10103	10016
13	13	2016-02-27 07:50:33	2016-02-27 07:56:38	1.00	6.0	10022	10110
14	14	2016-02-28 12:04:09	2016-02-28 12:07:01	0.70	4.5	10065	10021
15	16	2016-02-11 12:44:24	2016-02-11 13:07:43	3.80	18.0	10021	10027
16	17	2016-02-28 18:25:18	2016-02-28 18:31:08	0.98	6.0	10119	10009
17	18	2016-02-13 20:00:15	2016-02-13 20:17:20	2.83	13.5	10003	10282
18	19	2016-02-21 01:07:04	2016-02-21 01:09:43	0.60	4.0	10018	10018
19	20	2016-02-27 15:56:56	2016-02-27 16:04:51	1.53	8.0	10023	10025
20	21	2016-02-18 07:40:28	2016-02-18 07:49:00	0.93	7.0	10020	10167



# YOUR APPS WITH DB CONNECT

## Avoid Impedance Mismatch

```
@app.route("/")
def main():
    table = "samples.nyctaxi.trips"
    df = spark.table(table)

    if request.args.get('bounds'):
        filter = request.args.get('bounds')
        df = df.where(df.trip_distance < filter)

    if request.args.get('limit'):
        df = df.limit(int(request.args.get("limit")))
    else:
        df = df.limit(100)

    return str(df.toPandas().to_html())
```

```
@app.route("/")
def main():
    con = sqlite3.connect("data.db")
    query = "select * from trips"

    bounds = None; has_filter = False; params = []
    if request.args.get('bounds'):
        has_filter = True
        query += " where trip_distance < :1"
        params.append(request.args.get('bounds'))

    limit = 100
    if request.args.get('limit'):
        limit = int(request.args.get("limit"))

    query += " limit :2"
    params.append(limit)
    pdf = pd.read_sql(query, con, params=params)
    return str(pdf.to_html())
```

# YOUR APPS WITH DB CONNECT

String concatenation!

## Avoid Impedance Mismatch

```
@app.route("/")
def main():
    table = "samples.nyctaxi.trips"
    df = spark.table(table)

    if request.args.get('bounds'):
        filter = request.args.get('bounds')
        df = df.where(df.trip_distance < filter)

    if request.args.get('limit'):
        df = df.limit(int(request.args.get("limit")))
    else:
        df = df.limit(100)

    return str(df.toPandas().to_html())
```

Significant whitespace!

```
@app.route("/")
def main():
    con = sqlite3.connect("data.db")
    query = "select * from trips"

    bounds = None; has_filter = False; params = []
    if request.args.get('bounds'):
        has_filter = True
        query += " where trip_distance < :1"
        params.append(request.args.get('bounds'))

    limit = 100
    if request.args.get('limit'):
        limit = int(request.args.get("limit"))

    query += " limit :2"
    params.append(limit)
    pdf = pd.read_sql(query, con, params=params)
    return str(pdf.to_html())
```

# YOUR APPS WITH DB CONNECT

## Debugging!

```
@app.route("/")
def main():
    table = "samples.nyctaxi.trips"
    df = spark.table(table)

    if request.args.get('bounds'):
        filter = request.args.get('bounds')
        df = df.where(df.trip_distance < filter)

    if request.args.get('limit'):
        df = df.limit(int(request.args.get("limit")))
    else:
        df = df.limit(100)

    return str(df.toPandas().to_html())
```

```
@app.route("/")
def main():
    con = sqlite3.connect("data.db")
    query = "select * from trips"

    bounds = None; has_filter = False; params = []
    if request.args.get('bounds'):
        has_filter = True
        query += " where trip_distance < :1"
        params.append(request.args.get('bounds'))

    limit = 100
    if request.args.get('limit'):
        limit = int(request.args.get("limit"))

    query += " limit :2"
    params.append(limit)
    pdf = pd.read_sql(query, con, params=params)
    return str(pdf.to_html())
```

Project

- dais-demo ~/Development/dais-
  - data.db
  - dbconnect.py
  - sqlbased.py
- External Libraries
- Scratches and Consoles

```

9
10 @app.route("/")
11 def main():
12     table = "samples.nyctaxi.trips" table: 'samples.nyctaxi.trips'
13     df = spark.table(table) df: ['tpep_pickup_datetime', 'tpep_dropoff_datetime', 'trip_distance', 'fare_amount'
14
15     if request.args.get('bounds'):
16         filter = request.args.get('bounds')
17         df = df.where(df.trip_distance < filter)
18
19     if request.args.get('limit'):
20         df = df.limit(int(request.args.get("limit")))
21     else:
22         df = df.limit(100)
23
24     return str(df.toPandas().to_html())
25
26
27
28
main() > if request.args.get('bounds')

```

Debug Flask (dbconnect.py)

Debugger Console

- Thread-40 (pr...quest\_thread)
  - main, dbconnect.py:16
  - dispatch\_request, app.py:865
  - full\_dispatch\_request, app.py:880
  - wsgi\_app, app.py:1473
  - \_\_call\_\_, app.py:1498
  - execute, serving.py:324
  - run\_wsgi, serving.py:363
  - handle\_one\_request, server.py:424

Switch frames from anywhere in the ID...

Evaluate expression (e) or add a watch (w)

```

> df = {DataFrame} ['tpep_pickup_datetime', 'tpep_dropoff_datetime', 'trip_distance', 'fare_amount', 'pickup_zip', 'dropoff_zip'] ...View as DataFrame
table = {str} 'samples.nyctaxi.trips'

```



**HOW CAN I BE  
EVEN MORE  
EFFICIENT?**

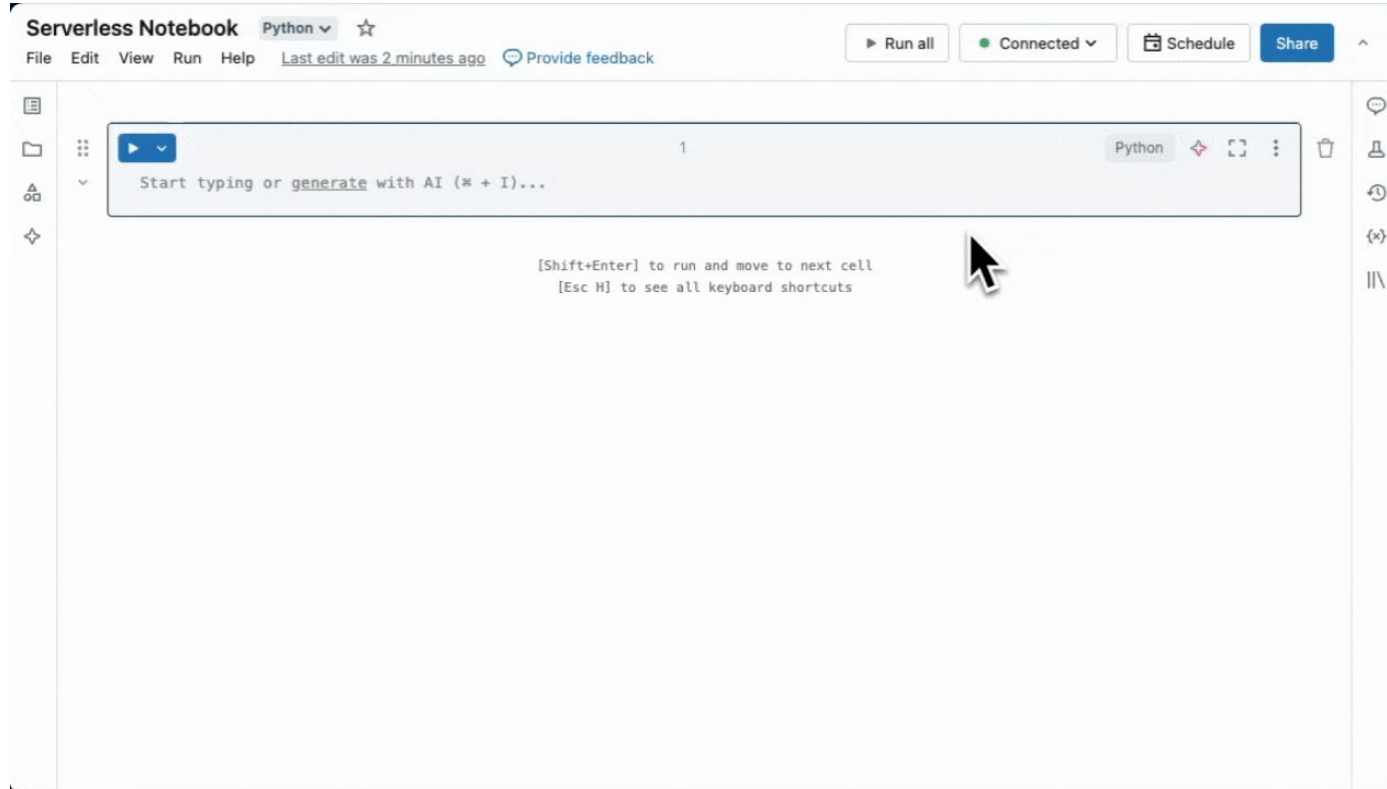
What if you never  
had to wait for the  
cluster to startup?

# Serverless Compute

# SERVERLESS NOTEBOOKS

Public Preview

So simple, it's hard to show!



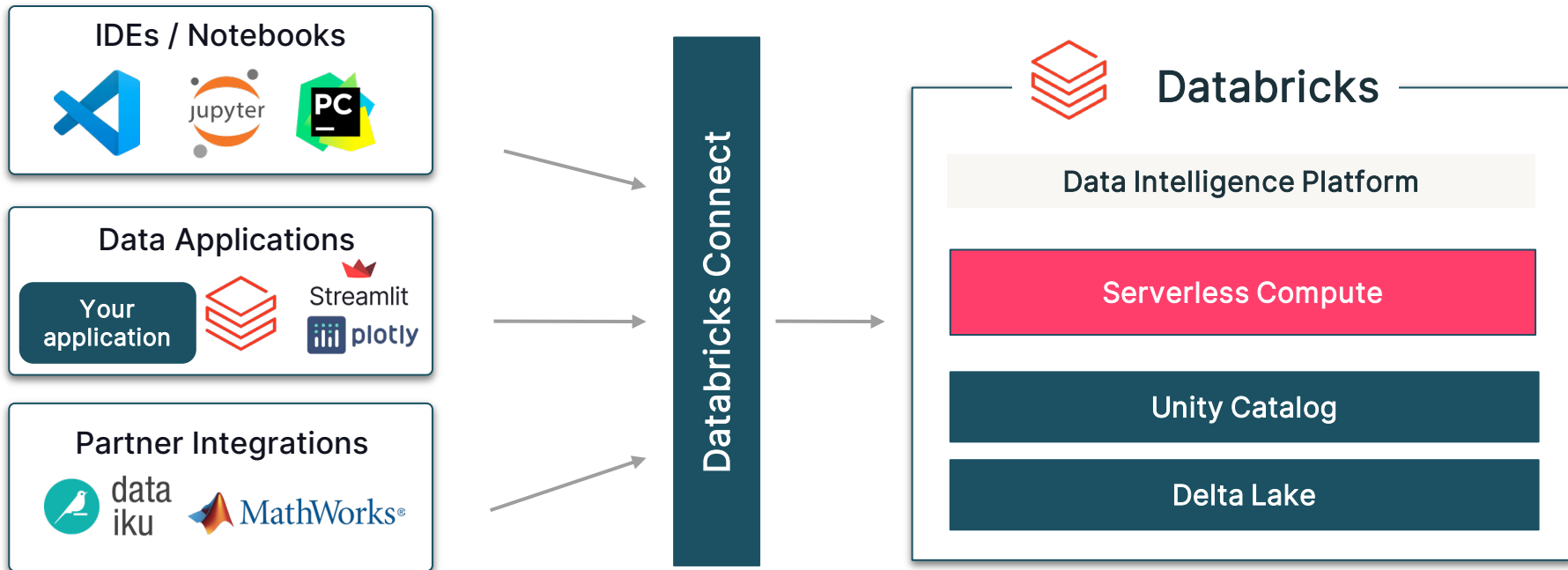
Imagine you connect  
from the IDE and it  
just runs...

# INTRODUCING

# SERVERLESS DATABRICKS CONNECT

# RUN ON SERVERLESS COMPUTE FROM ANYWHERE WITH DB CONNECT!

Public Preview  
(Soon)



# SERVERLESS DB CONNECT

- Serverless enabled in your workspace
- Use your existing code
- Use your existing profiles
- Simply use “serverless()”

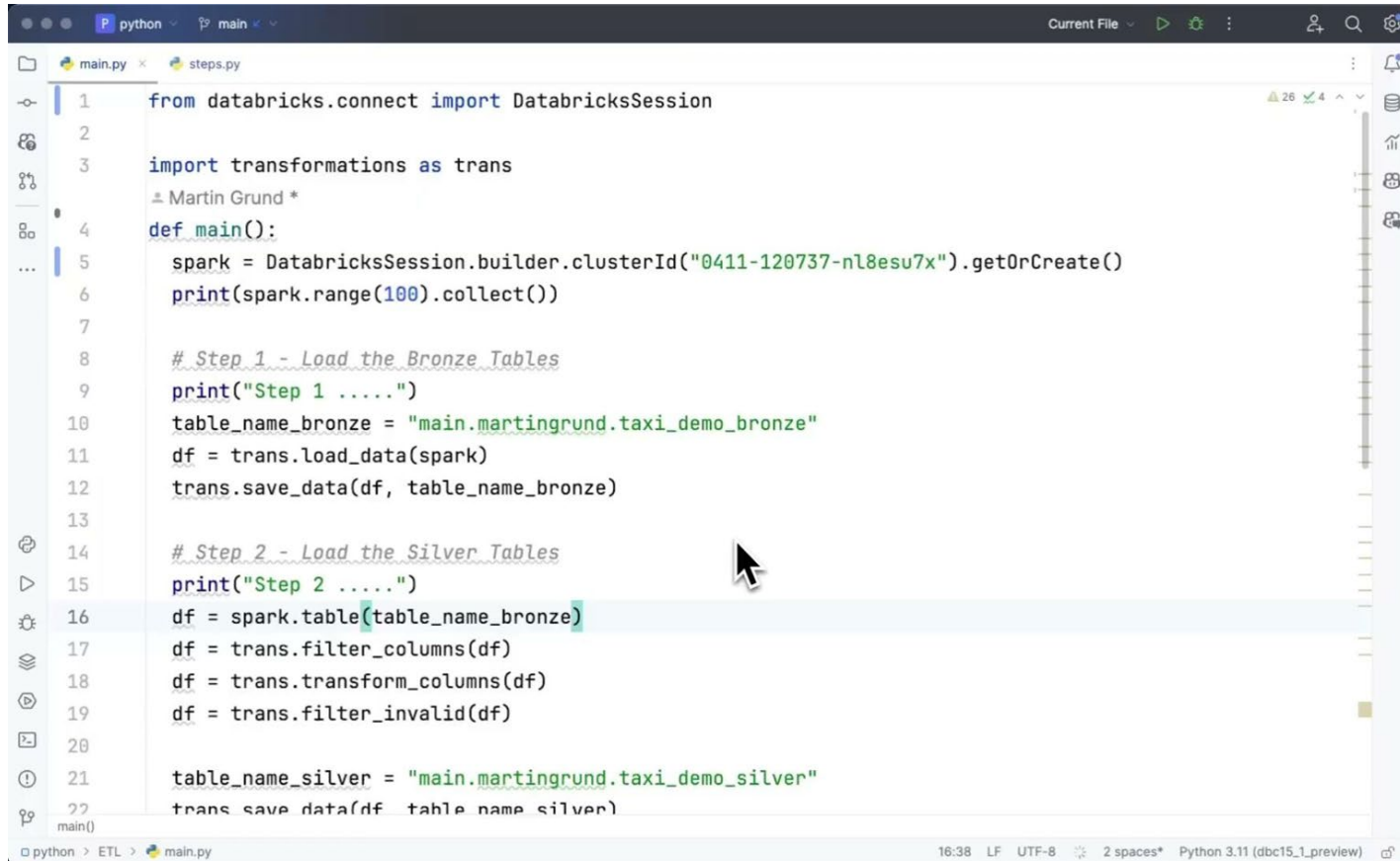
Local IDE



```
from databricks.connect import DatabricksSession  
  
spark = DatabricksSession.builder.serverless().getOrCreate()  
  
df = spark.read.table("samples.nyctaxi.trips")  
df.count()
```



# PYCHARM WITH SERVERLESS DB CONNECT



```
1 from databricks.connect import DatabricksSession
2
3 import transformations as trans
4 # Martin Grund *
5 def main():
6     spark = DatabricksSession.builder.clusterId("0411-120737-nl8esu7x").getOrCreate()
7     print(spark.range(100).collect())
8
9     # Step 1 - Load the Bronze Tables
10    print("Step 1 ....." )
11    table_name_bronze = "main.martingrund.taxi_demo_bronze"
12    df = trans.load_data(spark)
13    trans.save_data(df, table_name_bronze)
14
15    # Step 2 - Load the Silver Tables
16    print("Step 2 ....." )
17    df = spark.table(table_name_bronze)
18    df = trans.filter_columns(df)
19    df = trans.transform_columns(df)
20    df = trans.filter_invalid(df)
21
22    table_name_silver = "main.martingrund.taxi_demo_silver"
23    trans.save_data(df, table_name_silver)
24
25 main()
```

The screenshot shows the PyCharm IDE interface. The top bar indicates the current file is 'main.py' and the Python version is 3.11 (dbc15\_1\_preview). The code editor displays a Python script with the following content:

- Line 1: `from databricks.connect import DatabricksSession`
- Line 2: Blank line
- Line 3: `import transformations as trans`
- Line 4: `# Martin Grund *`
- Line 5: `def main():`
- Line 6:  `spark = DatabricksSession.builder.clusterId("0411-120737-nl8esu7x").getOrCreate()`
- Line 7:  `print(spark.range(100).collect())`
- Line 8: Blank line
- Line 9:  `# Step 1 - Load the Bronze Tables`
- Line 10:  `print("Step 1 ....." )`
- Line 11:  `table_name_bronze = "main.martingrund.taxi_demo_bronze"`
- Line 12:  `df = trans.load_data(spark)`
- Line 13:  `trans.save_data(df, table_name_bronze)`
- Line 14: Blank line
- Line 15:  `# Step 2 - Load the Silver Tables`
- Line 16:  `print("Step 2 ....." )`
- Line 17:  `df = spark.table(table_name_bronze)`
- Line 18:  `df = trans.filter_columns(df)`
- Line 19:  `df = trans.transform_columns(df)`
- Line 20:  `df = trans.filter_invalid(df)`
- Line 21: Blank line
- Line 22:  `table_name_silver = "main.martingrund.taxi_demo_silver"`
- Line 23:  `trans.save_data(df, table_name_silver)`
- Line 24: Blank line
- Line 25: `main()`

The status bar at the bottom shows the current file is 'main.py', the encoding is UTF-8, the indentation is 2 spaces, and the Python version is 3.11 (dbc15\_1\_preview).

# BENEFITS OF SERVERLESS DB CONNECT



## Easy setup

- No need to create a cluster
- No cluster Id



## Efficient everyday use

- Instant access with no idle time
- Dynamically adjust to workload and data size

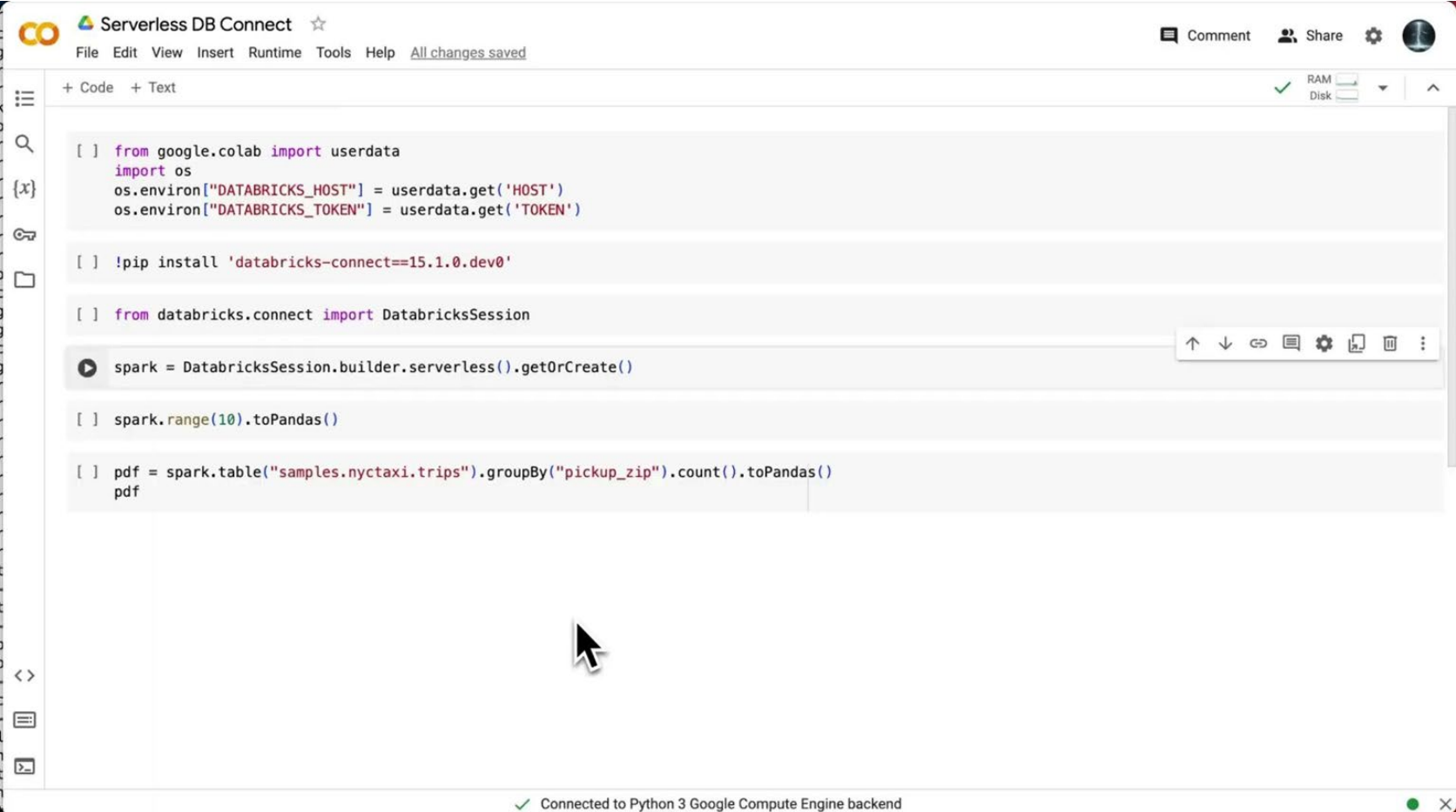


## Pay only for what you use

- Same cost as for

**USE DB CONNECT FROM  
REALLY EVERYWHERE**

# Google Collab □ Databricks Connect



```
Serverless DB Connect ☆
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[ ] from google.colab import userdata
import os
os.environ["DATABRICKS_HOST"] = userdata.get('HOST')
os.environ["DATABRICKS_TOKEN"] = userdata.get('TOKEN')

[ ] !pip install 'databricks-connect==15.1.0.dev0'

[ ] from databricks.connect import DatabricksSession

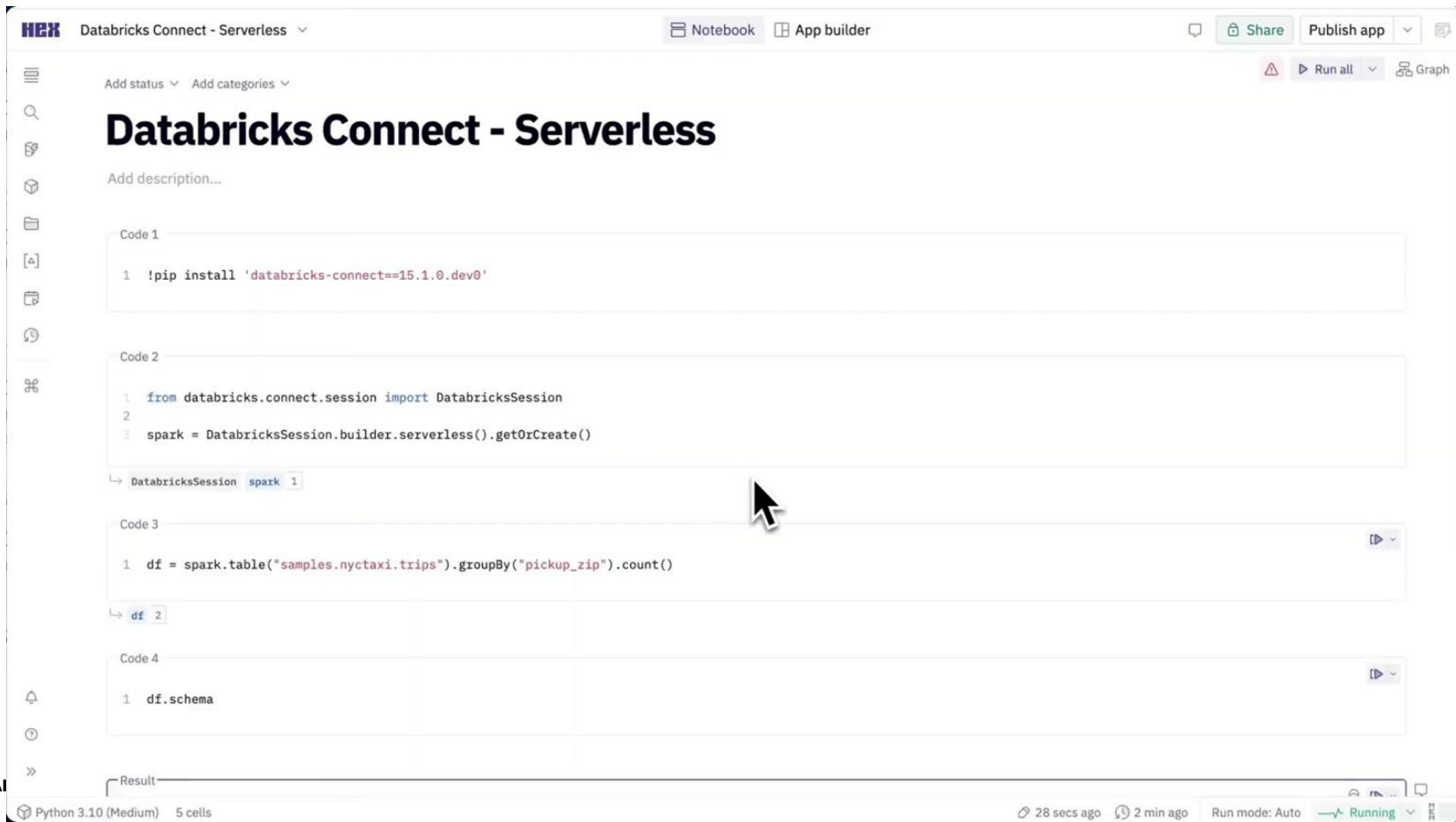
[ ] spark = DatabricksSession.builder.serverless().getOrCreate()

[ ] spark.range(10).toPandas()

[ ] pdf = spark.table("samples.nyctaxi.trips").groupBy("pickup_zip").count().toPandas()
pdf
```

Connected to Python 3 Google Compute Engine backend

# Hex Databricks Connect



The screenshot shows a Databricks Hex notebook interface. At the top, the title is "Databricks Connect - Serverless" with a dropdown arrow. To the right, there are buttons for "Notebook" and "App builder", and a "Share" button with a lock icon. Further right, there is a "Publish app" button with a dropdown arrow. Below the title, there are "Add status" and "Add categories" dropdowns. The main content area has a large heading "Databricks Connect - Serverless" and a placeholder "Add description...". There are four code cells:

- Code 1:** `!pip install 'databricks-connect==15.1.0.dev0'`
- Code 2:** `from databricks.connect.session import DatabricksSession  
spark = DatabricksSession.builder.serverless().getOrCreate()`
- Code 3:** `df = spark.table("samples.nyctaxi.trips").groupBy("pickup_zip").count()`
- Code 4:** `df.schema`

Below the code cells, there are labels for the variables: "DatabricksSession spark 1" and "df 2". A mouse cursor is pointing at the "spark 1" label. At the bottom, there is a "Result" section and a status bar showing "Python 3.10 (Medium) 5 cells", "28 secs ago", "2 min ago", "Run mode: Auto", and "Running".

# SUMMARY



# BUILD ON YOUR LAKEHOUSE FAST AND EFFICIENTLY

## Databricks Connect + Serverless



**DB Connect:** Full power of Spark from anywhere

- Governance with Unity Catalog
- Authentication (OAuth, Azure AD, PAT)



**Serverless:** Fast and efficient

- No Cluster, No Cluster ID
- Immediate response & Scalability



**Databricks**

Data Intelligence Platform

Mosaic AI

Delta Live  
Tables

Workflows

Databricks  
SQL

Unity Catalog

Delta Lake

# SERVERLESS DATABRICKS CONNECT

## Call to Action

Try DB Connect **git examples**:

- Develop ETL pipeline with DB Connect
- Build your first APP with DB Connect



Sign up to **Serverless DB Connect Preview -> form** in QR code





# Questions?



# Learn more at the summit!



Mobile App



## Tells us what you think

- We kindly request your valuable feedback on this session.
- Please take a moment to rate and share your thoughts about it.
- You can conveniently provide your feedback and rating through the Mobile App



## What to do next?

- Discover more related sessions in the mobile app!
- Visit the Demo Booth: Experience innovation firsthand!
- More Activities: Engage and connect further at the Databricks Zone!



## Get trained and certified

Visit the Learning Hub Experience at Moscone West, 2nd Floor

- Take complimentary certification at the event; come by the Certified Lounge
- Visit our Databricks Learning website for more training, courses and workshops!  
[databricks.com/learn](https://databricks.com/learn)

