



Azure Synapse



Delta Lake - Spark, Serverless, SQL Pools & Beyond

Hiram Fleitas
Senior Customer Engineer, Data & AI

bit.ly/sql1021slides



About Me

aka.ms/hiram

- 
1. Machine Learning
 2. Synapse Analytics
 3. Power Platform
 4. DevOps

Objective

I have a big data transactional CDC (change data capture) workload in **delta** format. This is very common in Databricks, but how does it work on **Synapse**?

Why? Because you can and you'd like to host/join/index your data in a static-layer to serve n-Number of queries.

What is Azure Synapse?

- Limitless analytics service includes:
 - Data Integration
 - Enterprise data warehousing
 - Big data analytics
- Provides **unified experience** to:
 - Ingest
 - Explore
 - Prepare
 - Manage
 - Serve
 - Business Intelligence
 - Machine Learning

AZURE SYNAPSE ANALYTICS



EXPERIENCE : STUDIO.

PLATFORM:

MANAGEMENT SQL, PYTHON, .NET, JAVA, SCALA, R
SECURITY PROVISIONED, ON-DEMAND
MONITORING RUNTIMES: SQL + APACHE SPARK
METASTORE

DATA INTEGRATION

DATA LAKE STORAGE

AZURE SYNPASE

INGEST PREP TRANSFORM AND ENRICH SERVE VIZ



DELTA LAKE™

What is Synapse Spark?

- **In-memory** parallel processing framework **for big data** analytics.
- One of Microsoft's implementation of Apache Spark.
- How does it differ?
 - A config/definition of what you want a Spark Instance to look like.
 - Is an Azure Resource
 - Looks like a Spark Cluster to clients
 - Is Free (when not being used to run a workload)
 - Provides isolation and budget control
 - Can be managed from Azure Portal or Synapse Studio
 - Is securable
 - Can have additional libraries installed, different libs on different pools
 - Can be auto scaled or statically defined
 - Is NOT a cluster

What is Delta format?

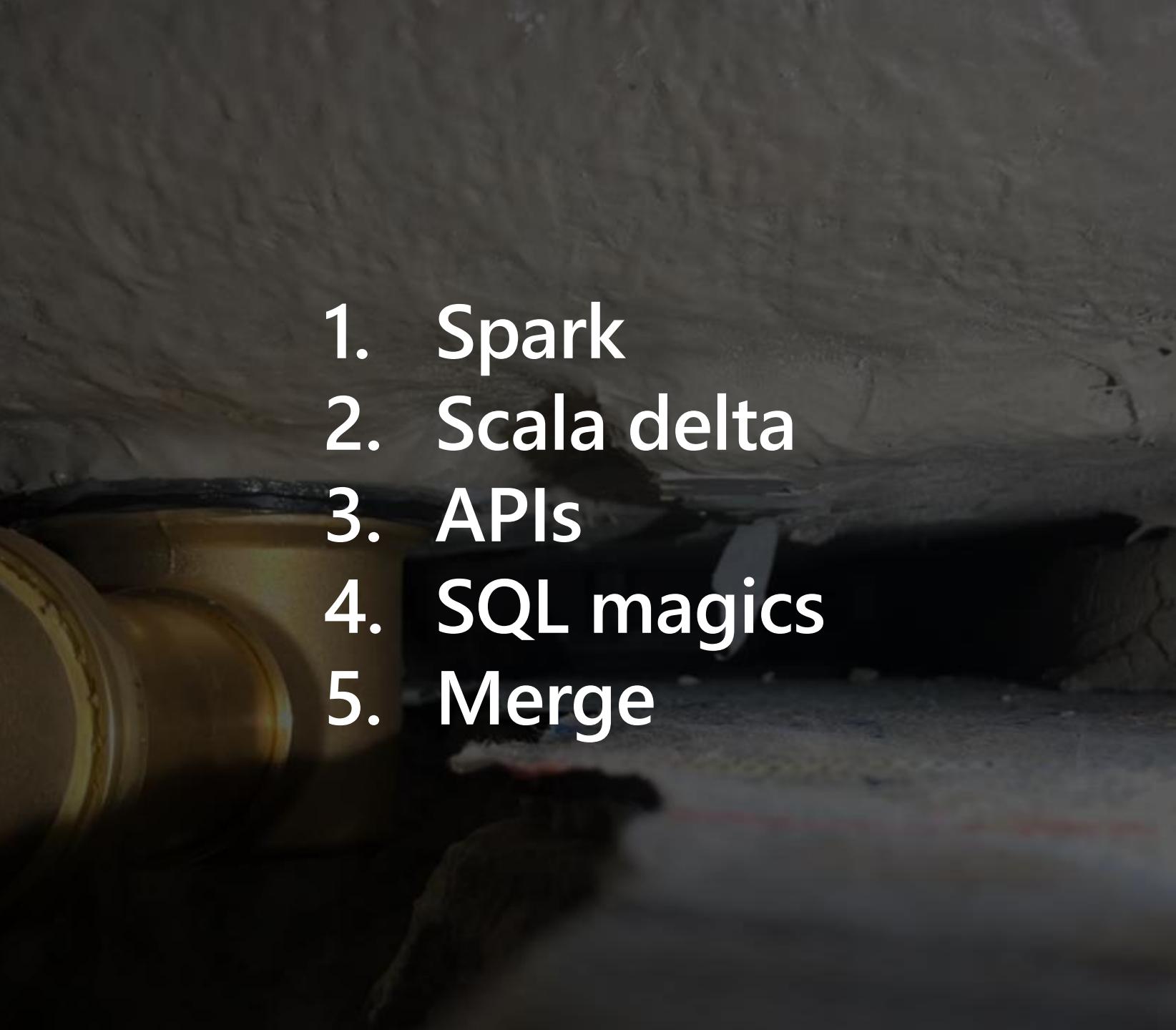
- OSS storage layer for Spark
- Provides:
 - ACID transactions (atomicity, consistency, isolation, and durability)
 - History of changes
 - Time travel in data history
 - Schema evolution
- Optimizations:
 - Z-Ordering: co-locates related info in the same set of files
 - Bin-Packaging (Compaction): OPTIMIZE command to coalesce small files into larger files.
 - VACCUM: cleanup old compacted files

Agenda

1. Intro
2. Spark
3. Serverless
4. Pipeline
5. SQL Pools
6. Resources

Demo 1

[Notebook - DeltaChanges3](#)

- 
1. Spark
 2. Scala delta
 3. APIs
 4. SQL magics
 5. Merge

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

Validate all Commit all Publish

Develop main branch DeltaChanges3

Filter resources by name

SQL scripts 24

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2
- DeltaChanges3

ReadDeltaLog

Start Delta Lake Python

Start Delta Lake Scala

Apache Spark job definit... 1

Ready

Delta Demo

1. Spark, scala delta, apis, sql magics - merge (see [this notebook](#))
2. Serverless, credentials, open query (see [sql script](#))
3. Integration pipeline, copy activity with overwrite if exists (see [pipeline](#))
4. Sql pools - merge (see [sql script](#))

```
[1] 1 spark.version
[1] ✓ 53 sec - Apache Spark session started in 52 sec 916 ms. Command executed in 511 ms by hiramfleitas on 6:06:34 PM, 10/25/21
[1] res5: String = 3.1.2.5.0-49430596
```

```
[2] ▶ 1 case class Data(key: String, value: String)
[2] 2
[2] 3 case class ChangeData(key: String, newValue: String, deleted: Boolean, time: Long) {
[2] 4   assert(newValue != null ^ deleted)
[2] 5 }
[2] ✓ 1 sec - Command executed in 1 sec 827 ms by hiramfleitas on 6:06:53 PM, 10/25/21
[2] ...
[2] defined class Data
[2] defined class ChangeData
```

+ Code + Markdown

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

Validate all Commit all Publish

Develop main branch DeltaChanges3

Filter resources by name

SQL scripts 24

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2
- DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

Ready

[2] 1 case class Data(key: String, value: String)
2
3 case class ChangeData(key: String, newValue: String, deleted: Boolean, time: Long) {
4 assert(newValue != null ^ deleted)
5 }

[2] ✓ 1 sec - Command executed in 1 sec 827 ms by hiramfleitas on 6:06:53 PM, 10/25/21

defined class Data
defined class ChangeData

[3] 1 val target = Seq(
2 Data("a", "0"),
3 Data("b", "1"),
4 Data("c", "2"),
5 Data("d", "3")
6).toDF()

[3] ✓ 2 sec - Command executed in 2 sec 749 ms by hiramfleitas on 6:07:47 PM, 10/25/21

[3] ... target: org.apache.spark.sql.DataFrame = [key: string, value: string]

+ Code + Markdown

[4] 1 sql("drop table if exists target")
2 target.write.format("delta").mode("overwrite").saveAsTable("target")

[4] ✓

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

main branch | Validate all | Commit all | Publish

Develop | DeltaChanges3

SQL scripts | Notebooks | Filter resources by name

SQL scripts: 24

Notebooks: 10

helloworld

HiSpark

Params

SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2
- DeltaChanges3

ReadDeltaLog

Start Delta Lake Python

Start Delta Lake Scala

Apache Spark job definit... | 1

Ready

[4] 1 sql("drop table if exists target")
2 target.write.format("delta").mode("overwrite").saveAsTable("target")
✓ 43 sec - Command executed in 43 sec 296 ms by hiramfleitas on 6:11:40 PM, 10/25/21

Job execution Succeeded Spark 2 executors 16 cores

| ID | Description | Status | Stages | Tasks | Submission Time | Duration |
|---------|--|-------------|--------|-----------------|----------------------|----------|
| > Job 0 | \$anonfun\$run\$2 at SynapseLoggingShim.scala:72 | ✓ Succeeded | 1/1 | 4/4 succeeded | 6:11:11 PM, 10/25/21 | 11 sec |
| > Job 1 | | ✓ Succeeded | 0/0 | 0/0 succeeded | 6:11:25 PM, 10/25/21 | 0 sec |
| > Job 2 | toString at String.java:2994 | ✓ Succeeded | 3/3 | 52/52 succeeded | 6:11:28 PM, 10/25/21 | 9 sec |

View in monitoring | Open Spark UI

res7: org.apache.spark.sql.DataFrame = []

+ Code | + Markdown

[5] 1 val targetDF = spark.sqlContext.sql("select * from target")
✓

[6] 1 %%sql
2 select * from target order by key
✓

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/monitoring/sparkapplication/DeltaChanges3_threedot0_1635199540?workspace=%2Fsubscriptions%2Fe4e06275-58d1-... 2

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

Apache Spark applications > Livy ID 5

DeltaChanges3_threedot0_1635199540

Completed tasks 56 of 56 Status Running Total duration 7m 52s

Cancel Refresh Spark UI

Attempts 1 of 1

All job IDs View Progress Playback 0 ms / 24 sec 973 ms

| | | |
|---|--|---|
| Job 0 | Job 1 | Job 2 |
| Tasks: 4 Duration: 11 sec 69 ms Rows: 4 Data read: 0 bytes Data written: 2.3 KB | Tasks: 0 Duration: 0 ms Rows: 0 Data read: 0 bytes Data written: 0 bytes | Tasks: 52 Duration: 8 sec 501 ms Rows: 121 Data read: 7.7 KB Data written: 6.3 KB |
| 1 Stage | 0 Stages | 3 Stages |

Diagnostics Logs Input data Output data

Driver (stderr)

Last modified 2021-10-25T22:11:49.000GMT

Length 196515

Download logs

Driver (stdout)

stdout

Latest

PreLaunch (stderr)

PreLaunch (stdout)

Livy

Search Filter errors and warnings

Showing logs since 6:13:31 PM, 10/25/2021. Load older logs

```
21/10/25 22:11:38 INFO SasClient: Response Str : {"TableType": "MANAGED", "Namespace": {"SchemaName": null, "Database": null}}
```

```
21/10/25 22:11:38 INFO HiveMetastoreClientImpl: Table default.target is found
```

```
21/10/25 22:11:38 INFO HiveMetastoreClientImpl: Start to get table default.target
```

```
21/10/25 22:11:38 INFO ConvertRequestObjectToJsonFilter: Request body json string : {"HmsRuntimeProperties": {"SpaceId": "12345678901234567890123456789012", "Region": "eastus", "StorageAccount": "synapsestorage1"}, "TableType": "MANAGED", "Namespace": {"SchemaName": null, "Database": null}, "Table": "target", "Partition": null}
```

```
21/10/25 22:11:39 INFO SasClient: Response headers : {x-ms-user-id=[e10fcc43-9abd-4aad-af2d-b6bb7d9e18eb], Server=Microsoft-HTTPAPI/2.0, Date=Wed, 27 Oct 2021 00:11:39 GMT, Content-Type=application/json, Content-Length=196515, Connection=close}
```

```
21/10/25 22:11:39 INFO SasClient: method=GET path=https://tokenservice1.eastus.azuresynapse.net:443/api/v1/proxy/1/tokenservice
```

```
21/10/25 22:11:39 INFO SasClient: Response Str : {"TableType": "MANAGED", "Namespace": {"SchemaName": null, "Database": null}, "Table": "target", "Partition": null}
```

```
21/10/25 22:11:39 INFO HiveMetastoreClientImpl: Table default.target is found
```

Stages for All Jobs

Completed Stages: 4

Completed Stages (4)

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

| Stage Id | Description | Submitted | Duration | Tasks: Succeeded/Total | Input | Output | Shuffle Read | Shuffle Write |
|----------|---|---------------------|----------|------------------------|----------|---------|--------------|---------------|
| 3 | Delta: Job group for statement 4: Compute snapshot for version: 0 toString at String.java:2994 | 2021/10/25 22:11:36 | 0.3 s | 1/1 | | | 4.3 KiB | |
| 2 | Delta: Job group for statement 4: Compute snapshot for version: 0 toString at String.java:2994 | 2021/10/25 22:11:31 | 5 s | 50/50 | | | 2.0 KiB | 4.3 KiB |
| 1 | Delta: Job group for statement 4: Compute snapshot for version: 0 toString at String.java:2994 | 2021/10/25 22:11:28 | 3 s | 1/1 | 1402.0 B | | | 2.0 KiB |
| 0 | Job group for statement 4 \$anonfun\$run\$2 at SynapseLoggingShim.scala:72 | 2021/10/25 22:11:11 | 10 s | 4/4 | | 2.3 KiB | | |

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

wplushiramsynapse - Azure Synapse | DeltaChanges3_threedit0_16351 x | +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse | Search | Help | User: hiramfleitas@microsoft.com

Develop / main branch | Validate all | Commit all | Publish

DeltaChanges3

Run all | Undo | Commit | Outline | Attach to: threedit0 | Language: Spark (Scala) | ...

Filter resources by name

SQL scripts: 24

Notebooks: 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE
- Delta
 - DeltaChanges
 - DeltaChanges2
- DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

Ready

```
1 sql("drop table if exists target")
2 target.write.format("delta").mode("overwrite").saveAsTable("target")
```

[4] ✓ 43 sec - Command executed in 43 sec 296 ms by hiramfleitas on 6:11:40 PM, 10/25/21

Job execution Succeeded Spark 2 executors 16 cores

| ID | Description | Status | Stages | Tasks | Submission Time | Duration |
|-------|--|-----------|--------|-----------------|----------------------|----------|
| Job 0 | \$anonfun\$run\$2 at SynapseLoggingShim.scala:72 | Succeeded | 1/1 | 4/4 succeeded | 6:11:11 PM, 10/25/21 | 11 sec |
| Job 1 | | Succeeded | 0/0 | 0/0 succeeded | 6:11:25 PM, 10/25/21 | 0 sec |
| Job 2 | toString at String.java:2994 | Succeeded | 3/3 | 52/52 succeeded | 6:11:28 PM, 10/25/21 | 9 sec |

res7: org.apache.spark.sql.DataFrame = []

```
1 val targetDF = spark.sqlContext.sql("select * from target")
```

[5] ✓ <1 sec - Command executed in 1 sec 43 ms by hiramfleitas on 6:16:02 PM, 10/25/21

... targetDF: org.apache.spark.sql.DataFrame = [key: string, value: string]

+ Code | + Markdown

```
1 %%sql
2 select * from target order by key
```

wplushiramsynapse - Azure Synapse | DeltaChanges3_threedit0_16351 x | +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse | Search | Help | User: hiramfleitas@microsoft.com

Develop / main branch | Validate all | Commit all | Publish

DeltaChanges3

Run all | Undo | Commit | Outline | Attach to threedit0 | Language Spark (Scala) | ...

Filter resources by name

SQL scripts 24

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2
- DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

Ready

```
targetDF: org.apache.spark.sql.DataFrame = [key: string, value: string]
```

[6] 1 %%sql
2 select * from target order by key
✓ 2 sec - Command executed in 2 sec 471 ms by hiramfleitas on 6:16:20 PM, 10/25/21

> Job execution Succeeded Spark 2 executors 16 cores

View in monitoring Open Spark UI

Table Chart Export results

| key | value |
|-----|-------|
| a | 0 |
| b | 1 |
| c | 2 |
| d | 3 |

+ Code + Markdown

1 %%sql
2 describe extended target

wplushiramsynapse - Azure Synapse | DeltaChanges3_threedot0_16351 | +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse | Search | Help | Notifications | Profile

main branch | Validate all | Commit all | Publish

Develop | Filter resources by name

SQL scripts (24) | Notebooks (10) | Delta (7) | Apache Spark job definitions (1)

DeltaChanges3

Run all | Undo | Commit | Outline | Attach to threedot0 | Language Spark (Scala) | ...

Filter resources by name

Ready

[6] 1 %sql
2 select * from target order by key
✓ 2 sec - Command executed in 2 sec 471 ms by hiramfleitas on 6:16:20 PM, 10/25/21

Job execution Succeeded Spark 2 executors 16 cores

View Table Chart Export results

Chart type: Pie chart

Key: key

Values: value

Series Group:

Aggregation: Avg

Apply Cancel

Diagram description: A pie chart showing the distribution of values across four categories labeled a, b, c, and d. Category d is the largest slice, colored blue, and is highlighted with a tooltip indicating it has a value of 50.0%. Categories a, b, and c are smaller slices colored light blue, green, and grey respectively.

| Category | Value (%) |
|----------|-----------|
| a | ~15% |
| b | ~15% |
| c | ~15% |
| d | 50.0% |

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462eb...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

main branch | Validate all | Commit all | Publish

Develop | + | <>

DeltaChanges3 | wplushiramsynapsefs

Run all | Undo | Commit | Outline | Attach to: threedot0 | Language: Spark (Scala) | ...

Filter resources by name

SQL scripts: 24

Notebooks: 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE
- Delta
 - DeltaChanges
 - DeltaChanges2
 - DeltaChanges3
 - ReadDeltaLog
 - Start Delta Lake Python
 - Start Delta Lake Scala
- Apache Spark job definit... 1

Ready [22]

1 %%sql
2 describe extended target

<1 sec - Command executed in 519 ms by hiramfleitas on 10:04:02 PM, 10/25/21

View Table Chart Export results

| col_name | data_type |
|----------|-----------|
| key | string |
| value | string |

Partitioning
Not partitioned

Detailed Table Info...

| Name | default.target |
|------------------|--|
| Location | abfss://wplushiramsynapsefs@wplushiramsynapseadls2.dfs.core.windows.net/synapse/worksaces/wplushiramsynapse/warehouse/target |
| Provider | delta |
| Table Properties | [Type=MANAGED,delta.minReaderVersion=1,delta.minWriterVersion=2] |

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/explore/linked/storageaccounts/wplushiramsynapse-WorkspaceDefaultStorage-wplushiramsynapseadlsv2%2...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

main branch / main branch Validate all Commit all Publish

Data

Workspace Linked

Filter resources by name

Azure Data Lake Storage Gen2

- wplushiramsynapse (Primary - wpl...) (Primary)
- adfstagedcommandtempdata
- adfstagedcopytempdata
- adfstagedpolybasetempdata
- dwtemp
- nyctaxistaging

Integration datasets 20

DeltaChanges3 wplushiramsynapsefs

New SQL script New data flow New integration dataset Upload Download New folder More

wplushiramsynapsefs > synapse > workspaces > wplushiramsynapse > warehouse

| Name | Last Modified | Content Type |
|-------------------|------------------------|--------------|
| lastchanges | 9/12/2021, 4:40:47 PM | Folder |
| manageddeltatable | 9/10/2021, 8:56:45 AM | Folder |
| target | 10/25/2021, 6:11:11 PM | Folder |

Select all in page
Select all cached
Size

Copy link
Rename
Manage access
Properties
Delete
Refresh

Showing 1 to 3 of 3 cached items

https://ms.web.azuresynapse.net/assets/external/storage/Scripts/0.2.19/StorageExplorer/FrameIndex.html?&id=a365TabContentDiv68219e62-61d5-47e6-aeac-3d02ad61560e&implPath=%2Fassets%2Fexternal%2Fstorage%2FStorageExplorerImpl%2F0.1.29%2F...

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1... 5

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + <>

Filter resources by name

SQL scripts 24

Notebooks 10

helloworld

HiSpark

Params

SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2

DeltaChanges3

ReadDeltaLog

Start Delta Lake Python

Start Delta Lake Scala

Apache Spark job definit... 1

DeltaChanges3 wplushiramsynapses DeltaChanges2

Run all Undo Commit Outline Attach to threedot0 ...

Ready

Shake it up

```
1 val changeDataSource = Seq(  
2   ChangeData("a", "10", deleted = false, time = 0),  
3   ChangeData("a", null, deleted = true, time = 1), // a was updated and then deleted  
4   ChangeData("b", null, deleted = true, time = 2), // b was just deleted once  
5   ChangeData("c", null, deleted = true, time = 3), // c was deleted and then updated twice  
6   ChangeData("c", "20", deleted = false, time = 4),  
7   ChangeData("c", "200", deleted = false, time = 5)  
8 ).toDF().createOrReplaceTempView("changes")
```

[30] ✓ <1 sec - Command executed in 497 ms by hiramfleitas on 10:28:26 PM, 10/25/21

changeDataSource: Unit = ()

+ Code + Markdown

SQL Example

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + <

Filter resources by name

SQL scripts 24

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE
- Delta
 - DeltaChanges
 - DeltaChanges2
 - DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

DeltaChanges3 wplushiramsynapse DeltaChanges2

Run all Undo Commit Outline Attach to threedot0 ...

Ready

SQL Example

```
1 %%sql
2 MERGE INTO target t
3 USING (
4     -- Find the latest change for each key based on the timestamp
5     SELECT key, latest.newValue as newValue, latest.deleted as deleted FROM (
6         -- Note: For nested structs, max on struct is computed as
7         -- max on first struct field, if equal fall back to second fields, and so on.
8         SELECT key, max(struct(time, newValue, deleted)) as latest FROM changes GROUP BY key
9     )
10 ) s
11 ON s.key = t.key
12 WHEN MATCHED AND s.deleted = true THEN DELETE
13 WHEN MATCHED THEN UPDATE SET key = s.key, value = s.newValue
14 WHEN NOT MATCHED AND s.deleted = false THEN INSERT (key, value) VALUES (key, newValue)
```

[31] ✓ 6 sec - Command executed in 6 sec 836 ms by hiramfleitas on 10:30:10 PM, 10/25/21

Job execution Succeeded Spark 2 executors 16 cores

No data available

View in monitoring Open Spark UI

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop DeltaChanges3 wplushiramsynapses DeltaChanges2

Run all Undo Commit Outline Attach to threedot0 ...

Filter resources by name

SQL scripts 24

Notebooks 10

helloworld

HiSpark

Params

SparkSQLCTE

Delta

DeltaChanges

DeltaChanges2

DeltaChanges3

ReadDeltaLog

Start Delta Lake Python

Start Delta Lake Scala

Apache Spark job definit... 1

Ready

```
1 val changesDF = spark.sql("select * from changes")
2 changesDF.head()
[43] ✓ <1 sec - Command executed in 498 ms by hiramfleitas on 10:32:03 PM, 10/25/21
changesDF: org.apache.spark.sql.DataFrame = [key: string, newValue: string ... 2 more fields]
res19: org.apache.spark.sql.Row = [a,10,false,0]
```

Run cell

```
1 val latestChangeForEachKey = changesDF.
2 selectExpr("key", "struct(time, newValue, deleted) as otherCols").
3 groupBy("key").
4 agg(max("otherCols").
5 as("latest")).
6 selectExpr("key", "latest.*")
[45] ✓ <1 sec - Command executed in 501 ms by hiramfleitas on 10:32:51 PM, 10/25/21
latestChangeForEachKey: org.apache.spark.sql.DataFrame = [key: string, time: bigint ... 2 more fields]
```

M ...

+ Code + Markdown

```
1 latestChangeForEachKey.show() // shows the latest change for each key
[12] ✓
```

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1... 5 5 ? hiramfleitas@microsoft.com MICROSOFT

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

main branch Validate all Commit all Publish

Develop + < Filter resources by name

SQL scripts 24

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2
- DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

DeltaChanges3 wplushiramsynapses DeltaChanges2

Run all Undo Commit Outline Attach to threedot0 ...

Ready

latestChangeForEachKey: org.apache.spark.sql.DataFrame = [key: string, time: bigint ... 2 more fields]

[46] 1 latestChangeForEachKey.show() // shows the latest change for each key
✓ <1 sec - Command executed in 1 sec 42 ms by hiramfleitas on 10:33:14 PM, 10/25/21

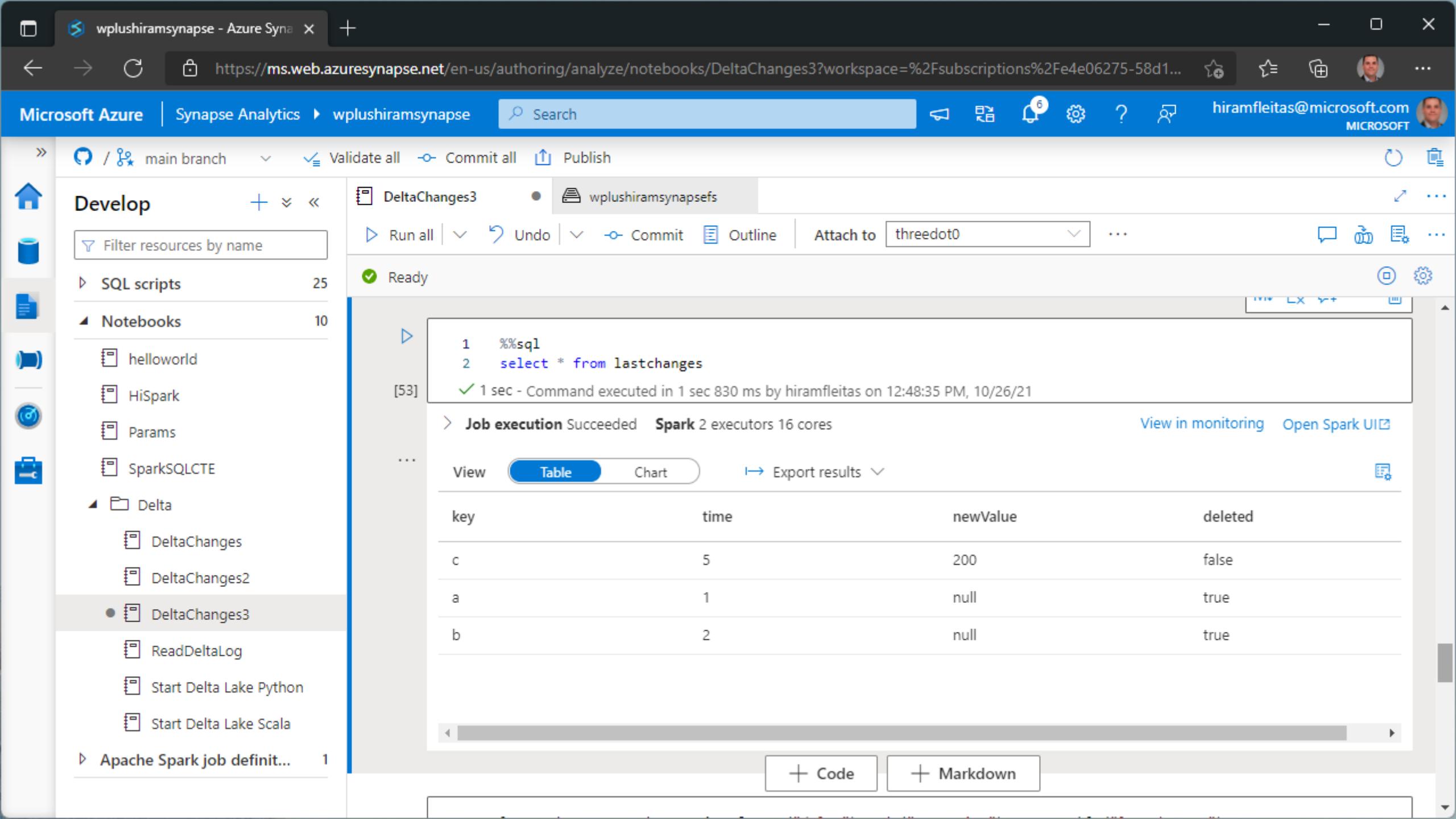
> Job execution Succeeded Spark 2 executors 16 cores

View in monitoring Open Spark UI

| key | time | newValue | deleted |
|-----|------|----------|---------|
| c | 5 | 200 | false |
| b | 2 | null | true |
| a | 1 | null | true |

+ Code + Markdown

[13] 1 latestChangeForEachKey.createOrReplaceTempView("lastchanges")



wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + <

Filter resources by name

SQL scripts 25

Notebooks 10

- helloworld
- HiSpark
- Params
- SparkSQLCTE
- Delta
 - DeltaChanges
 - DeltaChanges2
 - DeltaChanges3
- ReadDeltaLog
- Start Delta Lake Python
- Start Delta Lake Scala

Apache Spark job definit... 1

DeltaChanges3 wplushiramsynapses

Run all Undo Commit Outline Attach to threedot0 ...

Ready

View Table Chart Export results

Chart type: Pie chart

Key: key

Values: time

Series Group:

Aggregation: Avg

Apply Cancel

time

The chart displays the following data:

| Category | Value |
|----------|-------|
| a | ~15% |
| b | ~20% |
| c | 62.5% |

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e06275-58d1...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + < DeltaChanges3 wplushiramsynapses

Filter resources by name

SQL scripts 25

Notebooks 10

helloworld

HiSpark

Params

SparkSQLCTE

Delta

- DeltaChanges
- DeltaChanges2

DeltaChanges3

ReadDeltaLog

Start Delta Lake Python

Start Delta Lake Scala

Apache Spark job definit... 1

Run all Undo Commit Outline Attach to threedot0 ...

Ready

time

[68]

1 latestChangeForEachKey.write.format("delta").mode("overwrite").saveAsTable("lastchanges")

✓ 8 sec - Command executed in 8 sec 560 ms by hiramfleitas on 12:53:33 PM, 10/26/21

> Job execution Succeeded Spark 2 executors 16 cores

View in monitoring Open Spark UI

+ Code + Markdown

SCALA Merge example (Skipped)

```
1 // deltaTable.as("t")
2 // .merge(
3 //   latestChangeForEachKey.as("s"),
4 //   "s.key = t.key")
5 //   .whenMatchedUpdateAll()
```

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/explore/workspace/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e0...      

Microsoft Azure | Synapse Analytics > wplushiramsynapse  Search     ?  hiramfleitas@microsoft.com

main branch   Publish

Data   DeltaChanges3  wplushiramsynapses  SQL script 1 
Workspace  Linked     Attach to  threedot0 
 Filter resources by name  
Databases  3
wplussynapsedw (SQL) 
Tables 
dbo.Character 
dbo.dateformats 
dbo.dateformats2 
dbo.devs 
dbo.Episode 
dbo.heapTable 
dbo.latestonly 
dbo.Letters 
dbo_LOCATIONS 
dbo_newtable 
dbo_parks 

```
1 import com.microsoft.spark.sqlanalytics.utils.Constants
2 import org.apache.spark.sql.SqlAnalyticsConnector._
3
4 val lastchangesDF = spark.sqlContext.sql("select * from lastchanges")
5
6 lastchangesDF.write.option(Constants.
7 SERVER, "wplushiramsynapse.sql.azure-synapse.net").
8 synapsesql("wplussynapsedw.dbo.lastchanges", Constants.INTERNAL)
[*] 5 sec - Running
```

Job execution In progress Spark 2 executors 16 cores  

+ Code + Markdown

Resources

- <https://docs.databricks.com/static/notebooks/merge-in-cdc.html>
- <https://techcommunity.microsoft.com/t5/azure-synapse-analytics/query-delta-lake-files-using-t-sql-language-in-azure-synapse/ba-p/2388398>
- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-delta-lake-format>

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/explore/workspace/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e0...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Data DeltaChanges3 wplushiramsynapses SQL script 1

Run all Undo Commit Outline Attach to threedot0

Ready

Filter resources by name

Databases 3

wplussynapsedw (SQL)

Tables

dbo.Character

New SQL script

dbo.dateformats Refresh

dbo.dateformats2

dbo.devs

dbo.Episode

dbo.heapTable

dbo.latestonly

dbo.Letters

dbo.Locations

dbo.newtable

dbo.parks

1 import com.microsoft.spark.sqlanalytics.utils.Constants
2 import org.apache.spark.sql.SqlAnalyticsConnector._
3
> val lastchangesDF = spark.sqlContext.sql("select * from lastchanges")

> lastchangesDF.write.option(Constants.
. SERVER, "wplushiramsynapse.sql.azure-synapse.net").
8 synapsesql("wplussynapsedw.dbo.lastchanges", Constants.INTERNAL)
✓ 18 sec - Command executed in 18 sec 768 ms by hiramfleitas on 1:06:54 PM, 10/26/21
[70]
[70] > Job execution Succeeded Spark 2 executors 16 cores
View in monitoring Open Spark UI
...
import com.microsoft.spark.sqlanalytics.utils.Constants
import org.apache.spark.sql.SqlAnalyticsConnector._
lastchangesDF: org.apache.spark.sql.DataFrame = [key: string, time: bigint ... 2 more fields]

+ Code + Markdown

Resources

- https://docs.databricks.com/_static/notebooks/merge-in-cdc.html
- <https://techcommunity.microsoft.com/t5/azure-synapse-analytics/query-delta-lake-files-using-t-sql-language-in-azure-synapse/ba->

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/explore/workspace/notebooks/DeltaChanges3?workspace=%2Fsubscriptions%2Fe4e0...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Data DeltaChanges3 wplushiramsynapses SQL script 1

Run all Undo Commit Outline Attach to threedot0

Ready

Filter resources by name

dbo.Character
dbo.dateformats
dbo.dateformats2
dbo.devs
dbo.Episode
dbo.heapTable
dbo.lastchanges
dbo.latestonly
dbo.Letters

ML

```
1 import com.microsoft.spark.sqlanalytics.utils.Constants
2 import org.apache.spark.sql.SqlAnalyticsConnector._
3
4 val lastchangesDF = spark.sqlContext.sql("select * from lastchanges")
5
6 lastchangesDF.write.option(Constants.
7 SERVER, "wplushiramsynapse.sql.azure-synapse.net").
8 synapsesql("wplussynapsedw.dbo.lastchanges", Constants.INTERNAL)
✓ 18 sec - Command executed in 18 sec 768 ms by hiramfleitas on 1:06:54 PM, 10/26/21
```

[70] Job execution Succeeded Spark 2 executors 16 cores

View in monitoring Open Spark UI

```
import com.microsoft.spark.sqlanalytics.utils.Constants
import org.apache.spark.sql.SqlAnalyticsConnector._
lastchangesDF: org.apache.spark.sql.DataFrame = [key: string, time: bigint ... 2 more fields]
```

+ Code + Markdown

Resources

- https://docs.databricks.com/_static/notebooks/merge-in-cdc.html
- <https://techcommunity.microsoft.com/t5/azure-synapse-analytics/query-delta-lake-files-using-t-sql-language-in-azure-synapse/ba...>

Demo 1 - Resources

1. github.com/hfleitas/synapseDelta/blob/main/notebook/DeltaChanges3.ipynb
2. docs.databricks.com/_static/notebooks/merge-in-cdc.html
3. techcommunity.microsoft.com/t5/azure-synapse-analytics/query-delta-lake-files-using-t-sql-language-in-azure-synapse/ba-p/2388398
4. docs.microsoft.com/azure/synapse-analytics/sql/query-delta-lake-format
5. databricks.com/blog/2019/03/19/efficient-upserts-into-data-lakes-databricks-delta.html
6. docs.microsoft.com/sql/t-sql/statements/merge-transact-sql?view=azure-sqldw-latest
7. databricks.com/blog/2019/02/04/introducing-delta-time-travel-for-large-scale-data-lakes.html
8. docs.microsoft.com/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-develop-ctas
9. mssqltips.com/sqlservertip/6282/azure-data-factory-multiple-file-load-example-part-2
10. docs.microsoft.com/azure/synapse-analytics/spark/synapse-spark-sql-pool-import-export#write-api-1

Demo 2

[SQL Script - DeltaSrvlessViews](#)

- 
1. Credentials
 2. Openrowset

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/DeltaSrvlessViews?workspace=%2Fsubscriptions%2Fe4e06275-58d...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + <

Filter resources by name

SQL scripts 26

- CopyIntoDevs
- Create external table with SQL
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult

Delta

- DeltaDropTarget
- DeltaSQLPoolMerge
- DeltaSrvlessViews
- MegreLatest
- Reset
- ShakeNBake

DeltaChanges3 wplushiramsynapsefs DeltaSrvlessViews

Run Undo Committed Query plan Connect to Built-in ...

```
1 create credential [https://wplushiramsynapseadls2.dfs.core.windows.net/wplushiramsynapsefs]
2 with
3     identity = 'Managed Identity'
4 go
5
6 select * from sys.credentials;
7 go
8
```

Results Messages

View Table Chart Export results

Search

| credential_id | name | credential_ident... | create_date | modify_date | target_type | target_id |
|---------------|----------------------------------|---------------------|-------------------|-------------------|-------------|-----------|
| 65548 | https://sqlvas6gnicddvebxy.bl... | SHARED ACCESS ... | 2021-10-25T16:... | 2021-10-25T16:... | (NULL) | (NULL) |
| 65550 | https://wplushiramsynapseadl... | Managed Identity | 2021-10-27T15:... | 2021-10-27T15:... | (NULL) | (NULL) |

00:00:00 Query executed successfully.

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/DeltaSrvlessViews?workspace=%2Fsubscriptions%2Fe4e06275-58d...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

Develop / main branch Validate all Commit all Publish

DeltaChanges3 wplushiramsynapses DeltaSrvlessViews

Run Undo Commit Query plan Connect to Built-in ...

Filter resources by name

SQL scripts 25

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult
- Delta
 - DeltaDropTarget
 - DeltaSQLPoolMerge
- DeltaSrvlessViews
- MegreLatest
- Reset
- ShakeNBake

10 drop view if exists vDeltaTarget;
11 go
12 create view vDeltaTarget
13 as
14 SELECT [key], [value]
15 FROM OPENROWSET(
... BULK 'abfss://wplushiramsynapsefs@wplushiramsynapseadlsv2.dfs.core.windows.net/synapse/workspaces/wplushiramsynaps...
16 ... FORMAT = 'delta') as rows
17 go
18
19 drop view if exists vDeltaLastChanges;
20 go
21 create view vDeltaLastChanges
22 as
23 SELECT [key], [time], [newvalue], [deleted]
24 FROM OPENROWSET(
... BULK 'abfss://wplushiramsynapsefs@wplushiramsynapseadlsv2.dfs.core.windows.net/synapse/workspaces/wplushiramsynaps...
25 ... FORMAT = 'delta') as rows
26 go
27
28 select * from vDeltaTarget
29 select * from vDeltaLastChanges
30
31

Results Messages

00:00:18 Query executed successfully.

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/DeltaSrvlessViews?workspace=%2Fsubscriptions%2Fe4e06275-58d...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + < Filter resources by name

SQL scripts 25

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult
- Delta
 - DeltaDropTarget
 - DeltaSQLPoolMerge
 - DeltaSrvlessViews
 - MegreLatest
 - Reset
 - ShakeNBake
- M01L05Lab01

DeltaChanges3 wplushiramsynapses DeltaSrvlessViews

Run Undo Commit Query plan Connect to Built-in ...

Results Messages

Select Query 0 View Table Chart Export results

Search

| key | value |
|-----|-------|
| c | 200 |
| d | 3 |

Note: changes were merged to target (delta table) in previous demo.

00:00:18 Query executed successfully.

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/DeltaSrvlessViews?workspace=%2Fsubscriptions%2Fe4e06275-58d...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + < Filter resources by name

SQL scripts 25

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult
- Delta
 - DeltaDropTarget
 - DeltaSQLPoolMerge
 - DeltaSrvlessViews
 - MegreLatest
 - Reset
 - ShakeNBake
- M01L05Lab01

DeltaChanges3 wplushiramsynapses DeltaSrvlessViews

Run Undo Commit Query plan Connect to Built-in ...

Results Messages

Select Query 1 View Table Chart Export results

Search

| key | time | newvalue | deleted |
|-----|------|----------|---------|
| a | 1 | (NULL) | True |
| c | 5 | 200 | False |
| b | 2 | (NULL) | True |

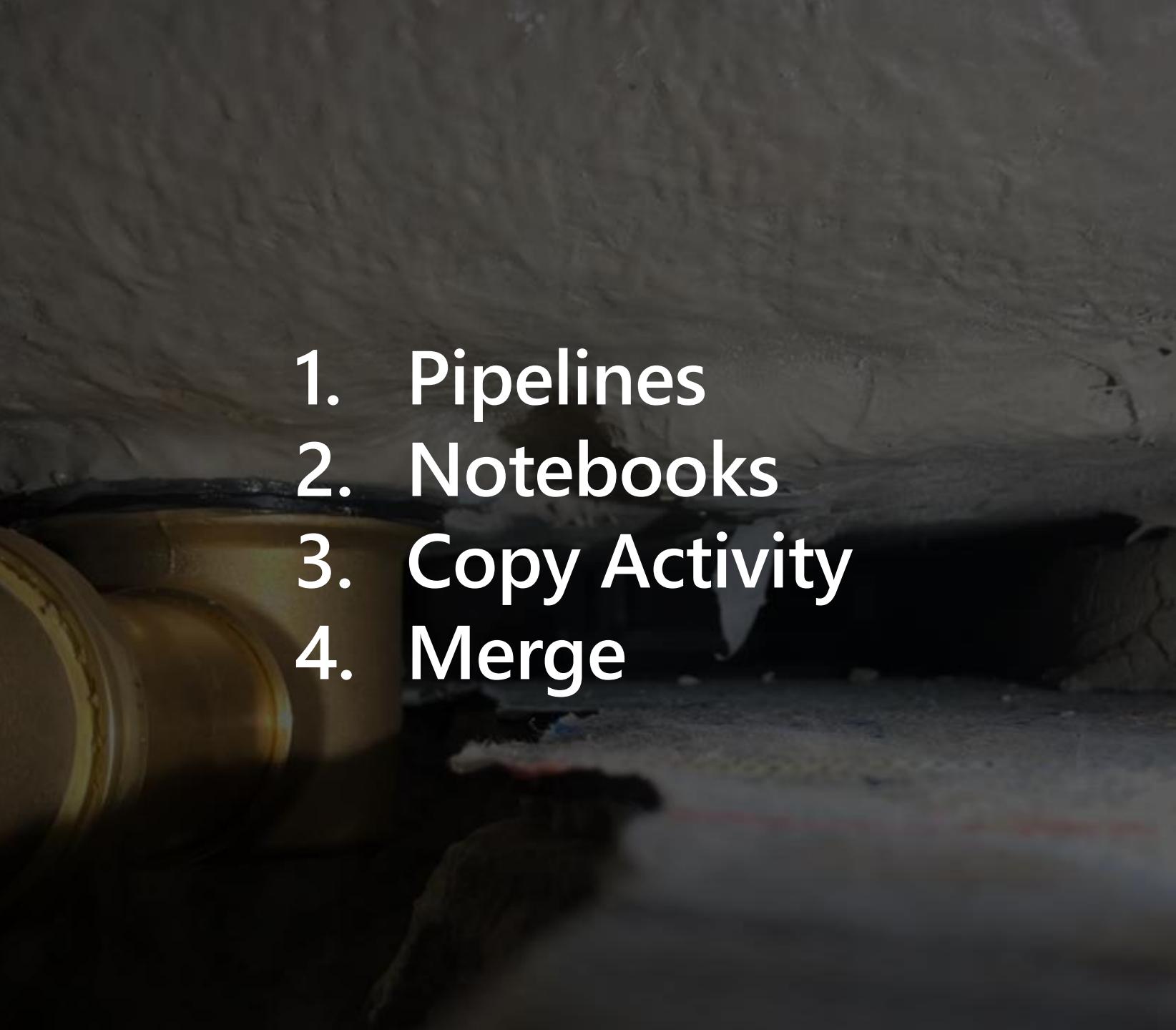
00:00:18 Query executed successfully.

Demo 2 - Resources

1. github.com/hfleitas/synapseDelta/blob/main/sqlscript/DeltaSrvlessViews.sql
2. docs.microsoft.com/azure/synapse-analytics/sql/overview-features#data-formats

Demo 3

Pipeline: DeltaChanges-Synapse

- 
1. Pipelines
 2. Notebooks
 3. Copy Activity
 4. Merge

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/orchestrate/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e062... hiramfleitas@microsoft.com MICROSOFT

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

main branch Validate all Commit all Publish

Integrate + <

Filter resources by name

Pipelines 13

- ADForIntDW
- CrimeData-ADB
- DevOpsCICD
- Params
- Primer
- sparkjob
- wait

Delta

- DeltaChanges-Synapse
- ReadDeltaLog-ADB
- ReadDeltaLog-Synapse

M01L05Lab01

Activities

Search activities

- Synapse
- Move & transform
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning

DeltaChanges3 wplushiramsynapsefs DeltaSrvlessViews DeltaChanges-Synapse

Committed Save as template Validate Debug Trigger (1)

Notebook DeltaChanges

Copy data LastChanges

Copy data Target

Stored procedure MergeLatest

General Settings User properties

Notebook * DeltaChanges Open New

Base parameters

```
graph LR; Note[Notebook DeltaChanges] --> C1[Copy data LastChanges]; Note --> C2[Copy data Target]; C1 --> SP[Stored procedure MergeLatest]; C2 --> SP
```

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1-... 9

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

DeltaChanges3 DeltaChanges-Syna... DeltaChanges DeltaSrvlessViews

Run all Undo Commit Outline Attach to XXL Language Spark (Scala) Variables ...

Not started

Table of contents

- 0. Create classes
- 1. Add sequence of changes
- 2. Read changes into DF
- 3. Expression for only latest changes
- 4. Temp view of latest changes
- 5. Save delta table

0. Create classes

```
1 case class Data(key: String, value: String)
2
3 case class ChangeData(key: String, newValue: String, deleted: Boolean, time: Long) {
4   assert(newValue != null ^ deleted)
5 }
```

[1] ✓

1. Add sequence of changes

```
1 val changeDataSource = Seq(
2   ChangeData("a", "10", deleted = false, time = 0),
3   ChangeData("a", null, deleted = true, time = 1), // a was updated and then deleted
4   ChangeData("b", null, deleted = true, time = 2)) // b was just deleted once
```

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1-...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search ? hiramfleitas@microsoft.com MICROSOFT

main branch Publish

DeltaChanges3 DeltaChanges-Syna... DeltaChanges DeltaSvlessViews

Run all Undo Commit Outline Attach to XXL Language Spark (Scala) Variables ...

Not started

Table of contents

- 0. Create classes
- 1. Add sequence of changes
- 2. Read changes into DF
- 3. Expression for only latest changes
- 4. Temp view of latest changes
- 5. Save delta table

1. Add sequence of changes

+ Code + Markdown

```
1 val changeDataSource = Seq(  
2   ChangeData("a", "10", deleted = false, time = 0),  
3   ChangeData("a", null, deleted = true, time = 1), // a was updated and then deleted  
4   ChangeData("b", null, deleted = true, time = 2), // b was just deleted once  
5   ChangeData("c", null, deleted = true, time = 3), // c was deleted and then updated twice  
6   ChangeData("c", "20", deleted = false, time = 4),  
7   ChangeData("c", "200", deleted = false, time = 5)  
8 ).toDF().createOrReplaceTempView("changes")  
9  
10 // read from last checkpoint  
11  
12 // val changeDataSource = Seq(  
13 //   ChangeData("d", "82", deleted = false, time = 6),  
14 //   ChangeData("e", "19", deleted = false, time = 7),  
15 //   ChangeData("f", "12", deleted = false, time = 8),  
16 //   ChangeData("g", null, deleted = true, time = 9),  
17 //   ChangeData("d", "21", deleted = false, time = 10)  
18 // ).toDF().createOrReplaceTempView("changes")
```

[2]

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1-... hiramfleitas@microsoft.com MICROSOFT

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

main branch Validate all Commit all Publish

DeltaChanges-Syna... DeltaChanges

Run all Undo Committing... Outline Attach to XXL Language Spark (Scala) Variables ...

Not started

Table of contents

- 0. Create the class
- 1. Add sequence of changes
- 2. Read changes into DF
- 3. Expression for only latest changes
- 4. Temp view of latest changes
- 5. Save delta table

2. Read changes into DF

+ Code + Markdown

```
[3] 1 val changesDF = spark.sql("select * from changes") ✓
```

3. Expression for only latest changes

```
[5] 1 val latestChangeForEachKey = changesDF.selectExpr("key", "struct(time, newValue, deleted) as otherC") ✓
```

4. Temp view of latest changes

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1-...     

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search    ?  hiramfleitas@microsoft.com MICROSOFT

main branch  Validate all  Commit all  Publish

DeltaChanges-Syna... DeltaChanges  Run all  Undo  Committing...  Outline  Attach to XXL Language Spark (Scala) Variables ...

Ready

Table of contents 

- 0. Create the class
- 1. Add sequence of changes
- 2. Read changes into DF
- 3. Expression for only latest changes
- 4. Temp view of latest changes
- 5. Save delta table

3. Expression for only latest changes

+ Code  + Markdown 

```
1 val latestChangeForEachKey = changesDF.selectExpr("key", "struct(time, newValue, deleted) as otherC")
2 latestChangeForEachKey.show()
```

[] Press shift + enter to execute cells

4. Temp view of latest changes

```
1 latestChangeForEachKey.createOrReplaceTempView("lastchanges")
```

[] Press shift + enter to execute cells

5. Save delta table

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/notebooks/DeltaChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1-... 6

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

DeltaChanges-Syna... DeltaChanges

Run all Undo Committing... Outline Attach to XXL Language Spark (Scala) Variables ...

Ready

Table of contents

- 0. Create the class
- 1. Add sequence of changes
- 2. Read changes into DF
- 3. Expression for only latest changes
- 4. Temp view of latest changes
- 5. Save delta table

5. Save delta table

+ Code + Markdown

```
1 latestChangeForEachKey.write.format("delta").mode("overwrite").saveAsTable("lastchanges")
```

Press shift + enter to execute cells

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/analyze/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e06275-...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

DeltaChanges3 | DeltaChanges-Synapse | DeltaChanges

Activities | Committed | Save as template | Validate | Validate copy runtime | Debug | Trigger (1)

Search activities

Synapse | Move & transform | Azure Data Explorer | Azure Function | Batch Service | Databricks | Data Lake Analytics | General | HDInsight | Iteration & conditionals | Machine Learning

Copy data | LastChanges | Stored procedure | MergeLatest

Notebook | DeltaChanges

General | Source | Sink | Mapping | Settings | User properties

Source dataset * vDeltaLastChanges | Open | New | Preview data | Learn more

Use query | Table | Query | Stored procedure

Query timeout (minutes) 120

Isolation level None

Partition option None | Physical partitions of table | Dynamic range

Please preview data to validate the partition settings are correct before you trigger a run or publish the pipeline.

Additional columns | New

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/dataset/vDeltaLastChanges?workspace=%2Fsubscriptions%2Fe4e06275-58d1...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

DeltaChanges3 | DeltaChanges-Syna... | DeltaChanges | vDeltaLastChanges

Committed

Azure SQL Database
vDeltaLastChanges

Connection Schema Parameters

Linked service * SQLOnDemand Test connection Edit New Learn more

Integration runtime * AutoResolveIntegrationRuntime Edit

Table dbo.vDeltaLastChanges Refresh Preview data Edit

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/dataset/vDeltaLastChanges?workspace=%2Fsubscriptions%2Fe4e06275...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch / main branch Validate all Commit all Publish

DeltaChanges3 DeltaChanges-Syna... DeltaChanges Committed

Azure SQL Database vDeltaLastChanges

Connection Schema Parameters

Linked service * SQLOnDemand Test connection

Integration runtime * AutoResolveIntegrationRuntime Edit

Table dbo.vDeltaLastChanges Refresh

Edit linked service (Azure SQL Database)

To avoid publishing immediately to workspace, please use Azure Key Vault to retrieve secrets securely. Learn more [here](#)

Name * SQLOnDemand

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Connection string Azure Key Vault

Account selection method Enter manually

From Azure subscription Enter manually

Fully qualified domain name * wplushiramsynapse-on-demand.sql.azuresynapse.net

Database name * hiram

Apply Cancel Test connection

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e0...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

DeltaChanges3 | DeltaChanges-Synapse | DeltaChanges

Activities | Committed | Save as template | Validate | Validate copy runtime | Debug | Trigger (1)

General | Source | Sink | Mapping | Settings | User properties

Sink dataset *: LoaderXLLLastChanges

Copy method: PolyBase

Allow PolyBase:

Reject type: Value

Reject value: 0

Use type default:

Table option: Auto create table

Pre-copy script: exec DeltaDropLastChanges

Write batch timeout

Write batch size

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/explore/workspace/sqlscripts/SQL%20script%202?workspace=%2Fsubscriptions%2Fe...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Data DeltaChanges-Syna... DeltaChanges SQL script 2

Run Undo Commit Query plan Connect to wplussynapsedw ...

Filter resources by name

Databases 3

- wplussynapsedw (SQL)
 - Tables
 - External tables
 - External resources
 - Views
- Programmability
 - Stored procedures
 - System stored procedures
 - dbo.DeltaDropLastChan...
 - dbo.DeltaDropTarget
 - dbo.GetRowCounts
 - dbo.MegreLatest
 - dbo.QueryActivity

Properties

General Related (0)

Name * SQL script 2

Description

Type .sql script

Size 190 bytes

Results settings per query ⓘ

First 5000 rows (default)

All rows

CREATE

DROP

DROP and CREATE

ALTER

Execute

New SQL script >

Add to pipeline >

Refresh

```
1 SET ANSI_NULLS ON
2 GO
3 SET QUOTED_IDENTIFIER ON
4 GO
5
6 CREATE PROC [dbo].[DeltaDropLastChanges] AS
7     if object_id('LastChanges') is not null
8         begin
9             drop table LastChanges
10        end
11 GO
```

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynthesize.net/en-us/authoring/analyze/dataset/LoaderXLLastChanges?workspace=%2Fsubscriptions%2Fe4e06...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

DeltaChanges3 | DeltaChanges-Syna... | DeltaChanges | LoaderXLLastChanges

Committed

Azure Synapse Analytics
LoaderXLLastChanges

Connection Schema Parameters

Linked service * wplushiramsynapseLoaderXL Test connection Edit New Learn more

Integration runtime * AutoResolveIntegrationRuntime Edit

Table dbo.lastchanges Refresh Preview data Edit

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/dataset/LoaderXLLastChanges?workspace=%2Fsubscriptions%2Fe4e06...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch / DeltaChanges3 DeltaChanges-Syna... DeltaChanges Committed

Azure Synapse Analytics
LoaderXLLastChanges

Connection Schema Parameters

Linked service * wplushiramsynapseLoaderXL Test connection

Integration runtime * AutoResolveIntegrationRuntime Edit

Table dbo.lastchanges Refresh

Edit linked service (Azure Synapse Analytics)

Description

Connect via integration runtime *

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Fully qualified domain name *

Database name *

Authentication type *

User name *

Password Azure Key Vault

Apply Cancel

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e0...

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

DeltaChanges3 DeltaChanges-Synapse DeltaChanges

Activities Committed Save as template Validate Validate copy runtime Debug Trigger (1)

General Source Sink Mapping Settings User properties

Import schemas Preview source New mapping Clear Reset Delete

| Source | Type | Destination | Type | |
|----------|---------|-------------|----------|---------|
| key | String | → | key | String |
| time | Int64 | → | time | Int64 |
| newValue | String | → | newValue | String |
| deleted | Boolean | → | deleted | Boolean |

Synapse Move & transform Azure Data Explorer Azure Function Batch Service Databricks Data Lake Analytics General HDInsight Iteration & conditionals Machine Learning

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e0...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

DeltaChanges3 | DeltaChanges-Synapse | DeltaChanges

Committed | Save as template | Validate | Validate copy runtime | Debug | Trigger (1)

Activities | Search activities

Synapse | Move & transform | Azure Data Explorer | Azure Function | Batch Service | Databricks | Data Lake Analytics | General | HDInsight | Iteration & conditionals | Machine Learning

General | Source | Sink | Mapping | Settings | User properties

You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Data integration unit: Auto | Edit

Degree of copy parallelism: | Edit

Fault tolerance:

Enable logging:

Enable staging:

Staging settings:

Staging account linked service: * [wplushiramsynapse-WorkspaceDefa...](#) | Test connection | Edit | New

Integration runtime: * [AutoResolveIntegrationRuntime](#) | Edit

Storage Path: [wplushiramsynapsefs](#) | Browse

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/orchestrate/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e062... ⚡ 🔍 📈 🌐 🏷️

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

DeltaChanges-Synapse DeltaChanges

Activities Committed Save as template Validate Debug Trigger (1)

Notebook DeltaChanges

Copy data LastChanges

Stored procedure MergeLatest

General Settings User properties

To reference SQL pool, use the SQL pool stored procedure instead.

Linked service * wplushiramsynapseLoaderXL Test connection Edit New

Integration runtime * AutoResolveIntegrationRuntime Edit

Stored procedure name * [dbo].[MegreLatest]

Edit

Stored procedure parameters

Import New

```
graph LR; DeltaChanges[DeltaChanges] --> CopyData[Copy data]; CopyData --> MergeLatest[Stored procedure MergeLatest];
```

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/MegreLatest?workspace=%2Fsubscriptions%2Fe4e06275-58d1-408...      

Microsoft Azure | Synapse Analytics > wplushiramsynapse  Search     ?  hiramfleitas@microsoft.com MICROSOFT

main branch   Publish

Develop    

SQL scripts 25 

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult

Delta  

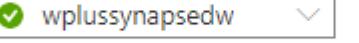
- DeltaDropTarget
- DeltaSQLPoolMerge
- DeltaSrvlessViews

MegreLatest  

Reset  

ShakeNBake  

M01L05Lab01  

DeltaChanges-Syna... DeltaChanges MegreLatest     Connect to  

```
1 if object_id('MegreLatest') is not null drop proc [dbo].[MegreLatest]
2 go
3
4 create proc [dbo].[MegreLatest]
5 as
6
7     merge into [ReportTarget] as t
8     using (select * from [lastchanges]) as lc ([key], [time], [newvalue], [deleted])
9     on lc.[key] = t.[key]
10    when matched and lc.[deleted] = 1
11        then delete
12    when matched
13        then update set t.[key] = lc.[key], t.[value] = lc.[newValue]
14    when not matched and lc.[deleted] = 0
15        then insert ([key], [value]) values(lc.[key], lc.[newValue]);
16
17 go
```

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/authoring/analyze/sqlscripts/ShakeNBake?workspace=%2Fsubscriptions%2Fe4e06275-58d1-408...    

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search       hiramfleitas@microsoft.com MICROSOFT

Develop / main branch   Publish

DeltaChanges-Syna... DeltaChanges MegreLatest ShakeNBake

Run Undo Commit Query plan Connect to  wplussynapsedw ... 

1 if object_id('ShakeNBake') is not null DROP PROCEDURE [dbo].[ShakeNBake]
2 GO
3
4 CREATE PROC [dbo].[ShakeNBake] AS
5 -- rename object target2hash TO ReportTarget;
6 --ctas to merge
7 if object_id('target2hash') is not null drop table [target2hash];
8 create table [target2hash]
9 with (
10 distribution = hash([key]),
11 clustered columnstore index
12) as select * from [target];
13 --merge
14 merge into [target2hash] as t
15 using (select * from [lastchanges]) as lc ([key], [time], [newvalue], [deleted])
16 on lc.[key] = t.[key]
17 when matched and lc.[deleted] = 1
| then delete
when matched
| then update set t.[key] = lc.[key], t.[value] = lc.[newValue]
18 when not matched and lc.[deleted] = 0
| then insert ([key], [value]) values(lc.[key], lc.[newValue]);
19 --rename
20 if object_id('ReportTarget') is not null drop table [ReportTarget];
21 rename object target2hash TO ReportTarget;
22
23
24
25
26 GO

DO NOT USE
Requires reloading target.

wplushiramsynapse - Azure Synapse

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/Reset?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

Develop + <

Filter resources by name

SQL scripts 25

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult

Delta

- DeltaDropTarget
- DeltaSQLPoolMerge
- DeltaSrvlessViews
- MegreLatest

Reset

DeltaChanges-Syna... DeltaChanges Reset MegreLatest

Run Undo Commit Query plan Connect to wplussynapsedw ...

```
15 insert [target] values('c','200')
16 go
17
18 --reset lastchanges
19 select * from lastchanges;
20 delete from [lastchanges];
21 insert [lastchanges] values ('a','1',NULL,1)
22 insert [lastchanges] values ('c','5','200',0)
23 insert [lastchanges] values ('b','2',NULL,1)
24 go
25
26 --reset reporttarget
27 select * from [reporttarget];
28 delete from [reporttarget];
29 go
30 insert [reporttarget].values('a','0')
31 insert [reporttarget].values('b','1')
32 insert [reporttarget].values('c','2')
33 insert [reporttarget].values('d','3')
34 go...
35
36 exec mergelatest
```

Results Messages

00:00:01 Query executed successfully.

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/authoring/orchestrate/pipeline/DeltaChanges-Synapse?workspace=%2Fsubscriptions%2Fe4e062... ⚡ 🔍 🌐 🏷️ 🎯

Microsoft Azure | Synapse Analytics > wplushiramsynapse Search

hiramfleitas@microsoft.com MICROSOFT

main branch Validate all Commit all Publish

DeltaChanges-Synapse DeltaChanges Reset MergeLatest

Activities Committed Save as template Validate Debug Trigger (1)

Trigger now

Copy data LastChanges MergeLatest

Notebook DeltaChanges

General Settings User properties

To reference SQL pool, use the SQL pool stored procedure instead.

Linked service * wplushiramsynapseLoaderXL Test connection Edit New

Integration runtime * AutoResolveIntegrationRuntime Edit

Stored procedure name * [dbo].[MergeLatest]

Edit

Stored procedure parameters

Import New

The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. A pipeline named 'DeltaChanges-Synapse' is displayed. The 'MergeLatest' activity is selected, showing its configuration details. The 'Settings' tab is active, containing fields for 'Linked service' (set to 'wplushiramsynapseLoaderXL'), 'Integration runtime' (set to 'AutoResolveIntegrationRuntime'), and 'Stored procedure name' (set to '[dbo].[MergeLatest]'). Below these, there is a section for 'Stored procedure parameters' with 'Import' and 'New' buttons. On the left, the 'Activities' pane lists various integration types like Synapse, Move & transform, and Azure Function. The main workspace shows a flow from a 'LastChanges' activity to a 'MergeLatest' activity, with a tooltip indicating a 'Copy data' operation between them. The top navigation bar includes 'Validate all', 'Commit all', and 'Publish' buttons.

wplushiramsynapse - Azure Synapse +

https://ms.web.azure-synapse.net/en-us/monitoring/pipelineruns?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462e...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

hiramfleitas@microsoft.com MICROSOFT

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Search by run ID or name Local time : Last 24 hours Pipeline name : All Status : All Runs : Latest runs Copy filters

Triggered by : All Add filter

Showing 1 - 2 items

| <input type="checkbox"/> Pipeline name | Run start ↑ | Run end | Duration | Triggered by | Status |
|---|-----------------------|----------------------|----------|----------------|-------------|
| DeltaChanges-Syn... <input type="checkbox"/> <input type="button"/> | 10/27/21, 12:06:52 AM | -- | 00:00:21 | Manual trigger | In progress |
| wait <input type="checkbox"/> | 10/26/21, 5:01:50 AM | 10/26/21, 5:01:54 AM | 00:00:03 | Manual trigger | Succeeded |

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/monitoring/pipelineruns/c0b70414-16b2-4fb5-bb35-292c8cdbe3f3?workspace=%2Fsubscriptions%2F... hiramfleitas@microsoft.com MICROSOFT

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

All pipeline runs > DeltaChanges-Synapse - Activity runs

DeltaChanges-Synapse

List Gantt

Rerun Rerun from activity Rerun from failed activity Refresh Update pipeline

```
graph LR; DeltaChanges[Notebook  
DeltaChanges] --> CopyData1["Copy data  
LastChanges"]; CopyData1 --> MergeLatest[Stored procedure  
MergeLatest];
```

Activity runs

Pipeline run ID c0b70414-16b2-4fb5-bb35-292c8cdbe3f3

All status

Showing 1 - 1 of 1 items

| Activity name | Activity type | Run start ↑ | Duration | Status | Error | Log | Integration ru |
|---------------|---------------|-----------------------|----------|-------------|-------|-----|----------------|
| DeltaChanges | Notebook | 10/27/21, 12:07:06 AM | 00:01:00 | In progress | | | |

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/monitoring/sparkapplication?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

hiramfleitas@microsoft.com MICROSOFT

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

Apache Spark applications

All Spark session Batch job Refresh Edit columns

Filter by keyword Local time : Last 24 hours Status : All Add filter

Showing 1 - 3 of 3 items

| Application name | Submitter | Submit time | Status | Pool | Type |
|---------------------------|------------------------------|-----------------------|---------------------------|-----------|-----------|
| DeltaChanges_XXL_163... | 6c183999-a58e-4295-bd6c-ac01 | 10/27/21, 12:07:07 AM | Submitting | XXL | Spark ses |
| DeltaChanges3_threedot... | hiramfleitas@microsoft.com | 10/26/21, 3:02:53 PM | Stopped (session timed ou | threedot0 | Spark ses |
| DeltaChanges3_threedot... | hiramfleitas@microsoft.com | 10/26/21, 12:42:59 PM | Stopped (session timed ou | threedot0 | Spark ses |

wplushiramsynapse - Azure Synapse +

https://ms.web.azuresynapse.net/en-us/monitoring/sparkapplication/DeltaChanges_XXL_1635307627?workspace=%2Fsubscriptions%2F...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

hiramfleitas@microsoft.com MICROSOFT

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

Apache Spark applications > Livy ID 20

DeltaChanges_XXL_1635307627

Completed tasks 206 of 206 Status Running Total duration 7m 34s

Cancel Refresh Spark UI

Attempts 1 of 1

All job IDs View Progress Playback 0 ms / 3 sec 611 ms

Job 0 Tasks: 7 Duration: 2 sec 635 ms Rows: 6 Data read: 0 bytes Data written: 531 bytes 2 Stages

Job 1 Tasks: 10 Duration: 196 ms Rows: 0 Data read: 0 bytes Data written: 0 bytes 2 Stages

Job 2 Tasks: 26 Duration: 121 ms Rows: 0 Data read: 0 bytes Data written: 0 bytes 2 Stages

Job 3 Tasks: 106 Duration: 375 ms Rows: 3 Data read: 274 bytes Data written: 0 bytes 2 Stages

Job 4 Tasks: 81 Duration: 238 ms Rows: 3 Data read: 257 bytes Data written: 0 bytes 2 Stages

Diagnostics Logs Input data Output data

Driver (stderr)

Last modified 2021-10-27T04:14:41.000GMT Length 132838

Download logs

Driver (stderr)

Driver (stdout)

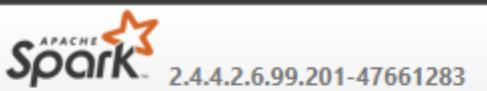
stdout

Latest

Search Filter errors and warnings

Showing logs since 12:14:51 AM, 10/27/2021. Load older logs

21/10/27 04:14:50 INFO BlockManagerInfo: Added broadcast_8_piece0 in memory on 6f4df57ad80c4c
21/10/27 04:14:50 INFO BlockManagerInfo: Added broadcast_8_piece0 in memory on 6f4df57ad80c4c



Jobs

Stage

Storage

Environment

Executors

Graph

Diagnosi

50

DeltaChanges_XXL_1635307627
application UI

Spark Jobs [\(?\)](#)

User: trusted-service-user

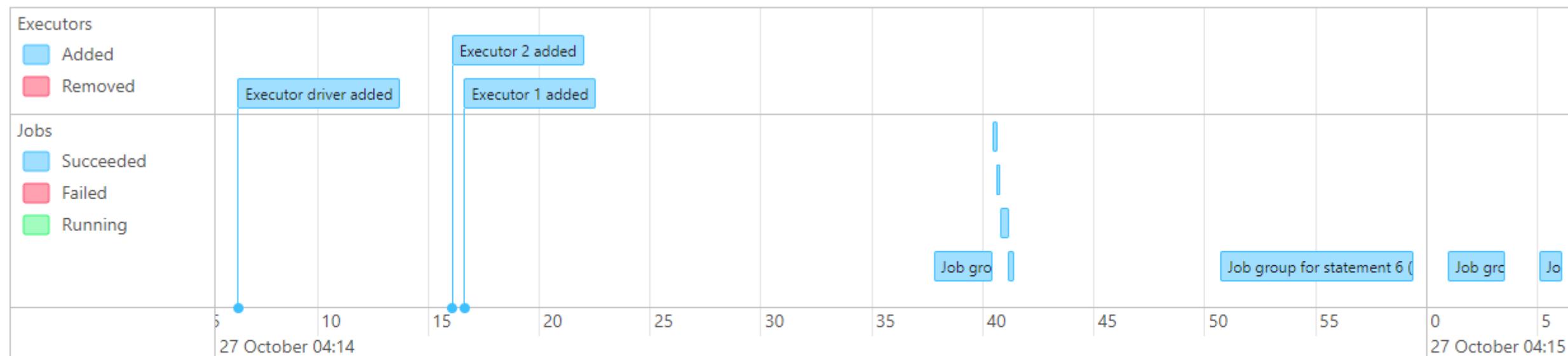
Total Uptime: 1.1 min

Scheduling Mode: FIFO

Completed Jobs: 8

► Event Timeline

Enable zooming



▼ Completed Jobs (8)

wplushiramsynapse - Azure Synapse DeltaChanges_XXL_1635307627 - +

https://ms.web.azure-synapse.net/en-us/monitoring/pipelineruns?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b-be12462e...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

hiramfleitas@microsoft.com MICROSOFT

Analytics pools

SQL pools

Apache Spark pools

Activities

SQL requests

Apache Spark applications

Data flow debug

Integration

Pipeline runs

Trigger runs

Integration runtimes

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Search by run ID or name Local time : Last 24 hours Pipeline name : All Status : All Runs : Latest runs Copy filters

Triggered by : All Add filter

Showing 1 - 2 items

| <input type="checkbox"/> Pipeline name | Run start ↑ | Run end | Duration | Triggered by | Status |
|--|-----------------------|-----------------------|----------|----------------|------------------------|
| DeltaChanges-Synapse | 10/27/21, 12:06:52 AM | 10/27/21, 12:15:57 AM | 00:09:05 | Manual trigger | ✓ Succeed |
| wait | 10/26/21, 5:01:50 AM | 10/26/21, 5:01:54 AM | 00:00:03 | Manual trigger | ✓ Succeed |

wplushiramsynapse - Azure Synapse | DeltaChanges_XXL_1635307627 | +

https://ms.web.azure-synapse.net/en-us/authoring/analyze/sqlscripts/Reset?workspace=%2Fsubscriptions%2Fe4e06275-58d1-4081-8f1b...

Microsoft Azure | Synapse Analytics > wplushiramsynapse | Search | Notifications | Help | hiramfleitas@microsoft.com | MICROSOFT

Main Branch / main branch | Validate all | Commit all | Publish

Develop | + | <<

Filter resources by name

SQL scripts (25)

- CopyIntoDevs
- Dateformats
- ExternalStuff
- PDWBackups
- ServerlessOpenRowsetDevs
- WordCountResult

Delta

- DeltaDropTarget
- DeltaSQLPoolMerge
- DeltaSrvlessViews
- MegreLatest
- Reset
- ShakeNBake

M01L05Lab01

DeltaChanges-Syna... | DeltaChanges | Reset | MegreLatest

Run | Undo | Committed | Query plan | Connect to: wplussynapsedw | ...

--reset reporttarget
select * from [reporttarget];

Results | Messages

View | Table | Chart | Export results

Search

| key | value |
|-----|-------|
| d | 3 |
| c | 200 |

00:00:00 Query executed successfully.

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/explore/workspace/notebooks/Notebook%201?workspace=%2Fsubscriptions%2Fe4e...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

Data

Workspace | Linked

Filter resources by name

Databases

- wplussynapsedw (SQL)
- hiram (SQL)

default (Spark)

- Tables
 - externaldeltatable
 - lastchanges
 - manageddeltatable
 - target
- Views

DeltaChanges-Syna... | DeltaSrvlessViews | Reset | Notebook 1

Run all | Undo | Commit | Outline | Attach to XXL | ...

Ready

[2]

```
1 %%pyspark
2 df = spark.sql("SELECT * FROM `default`.`target`")
3 df.show(10)
```

✓ 4 sec - Command executed in 3 sec 932 ms by hiramfleitas on 12:47:51 AM, 10/27/21

Job execution Succeeded Spark 2 executors 128 cores

View in monitoring Open Spark UI

New SQL script >

New notebook > Load to DataFrame

Machine Learning >

Refresh

+ Code + Markdown

| key | value |
|-----|-------|
| c | 200 |
| d | 3 |

wplushiramsynapse - Azure Synapse

https://ms.web.azuresynapse.net/en-us/authoring/explore/workspace/notebooks/Notebook%201?workspace=%2Fsubscriptions%2Fe4e...

Microsoft Azure | Synapse Analytics > wplushiramsynapse

Search

hiramfleitas@microsoft.com MICROSOFT

main branch | Validate all | Commit all | Publish

Data

Workspace | Linked

Filter resources by name

Databases

- wplussynapsedw (SQL)
- hiram (SQL)

default (Spark)

Tables

- externaldeltatable
- lastchanges
- manageddeltatable
- target

Views

DeltaChanges-Syna... DeltaSrvlessViews Reset Notebook 1

Run all | Undo | Commit | Outline | Attach to XXL ...

Ready

Run cell

1 %%sql
2 SELECT * FROM default.target

[3] ✓ 1 sec - Command executed in 1 sec 802 ms by hiramfleitas on 12:49:35 AM, 10/27/21

Job execution Succeeded Spark 2 executors 128 cores

View in monitoring Open Spark UI

key value

c 200

d 3

Code Markdown

This screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, there's a sidebar with navigation icons and a tree view of databases (wplussynapsedw, hiram), default (Spark) tables (externaldeltatable, lastchanges, manageddeltatable, target), and views. The main area displays a notebook cell with the following content:

```
1 %%sql
2 SELECT * FROM default.target
```

The cell status is "Ready". It shows a success message: "✓ 1 sec - Command executed in 1 sec 802 ms by hiramfleitas on 12:49:35 AM, 10/27/21". Below the cell, it says "Job execution Succeeded" with "Spark 2 executors 128 cores". The results are presented in a table:

| key | value |
|-----|-------|
| c | 200 |
| d | 3 |

At the bottom, there are buttons for "+ Code" and "+ Markdown".

Summary

- 1. Intro**
 - What is Azure Synapse
 - What is Synapse Spark
 - What is Delta format
- 2. Spark**
 - Scala delta
 - APIs
 - SQL magics & Merge
- 3. Serverless**
 - Credential
 - Openrowset
- 4. Pipelines**
 - Activities
 - Monitoring

Next Steps

1. github.com/Azure/Test-Drive-Azure-Synapse-with-a-1-click-POC
2. Home > Knowledge center > Browse Gallery > Notebooks > Delta
3. github.com/hfleitas/synapsedelta
4. dba2o.wordpress.com/2021/10/25/synapse-spark-getting-started
5. techcommunity.microsoft.com > Blogs > Azure Synapse Analytics
6. docs.microsoft.com/azure/synapse-analytics/spark/apache-spark-overview
7. docs.microsoft.com/azure/synapse-analytics/spark/apache-spark-what-is-delta-lake
8. docs.microsoft.com/azure/synapse-analytics/spark/synapse-spark-sql-pool-import-export

Eval

bit.ly/sql1021

Your feedback is appreciated.



Thank you

Hiram Fleitas
aka.ms/hiram