On the Way to Fabric ...

Alexander Karl

ABOUT ME

Alexander Karl

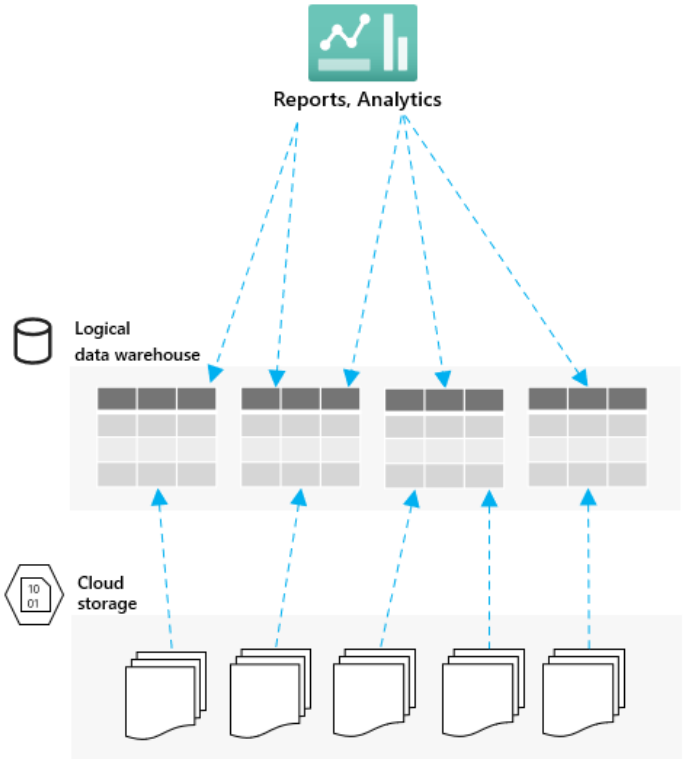
SQL & BI Developer

Wirt. Ing. (FH)

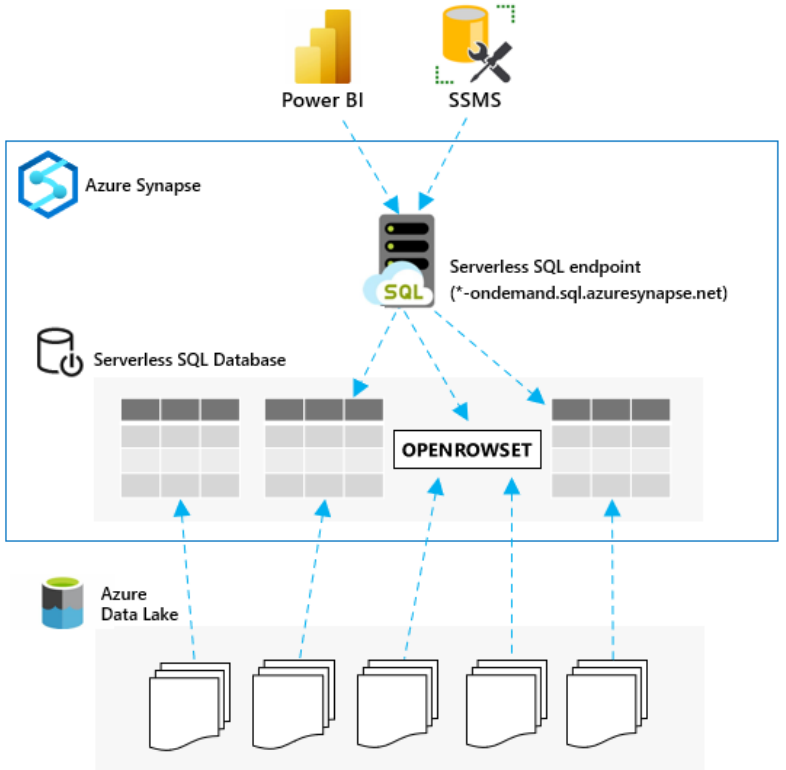
Stockstadt / PASS RG Rhein-Main

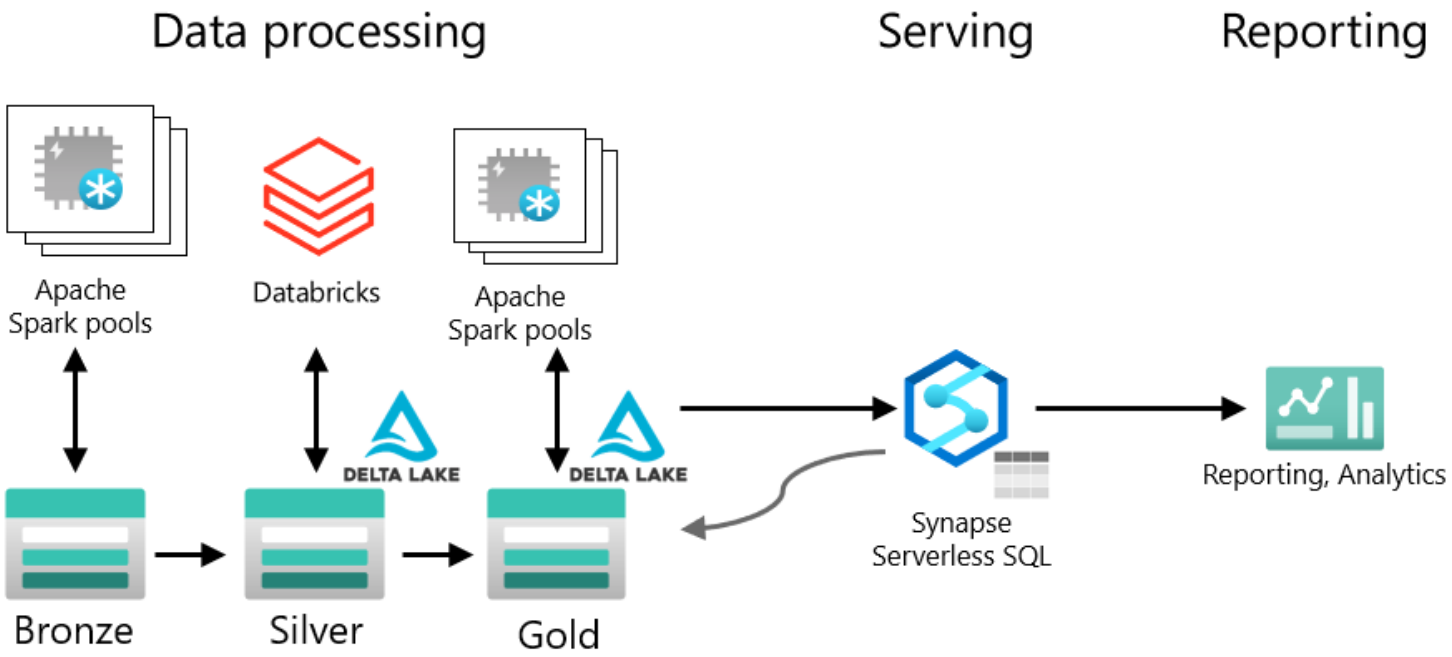
karl@net-CDE.de
& aka@sqlpass.de

Conceptual design

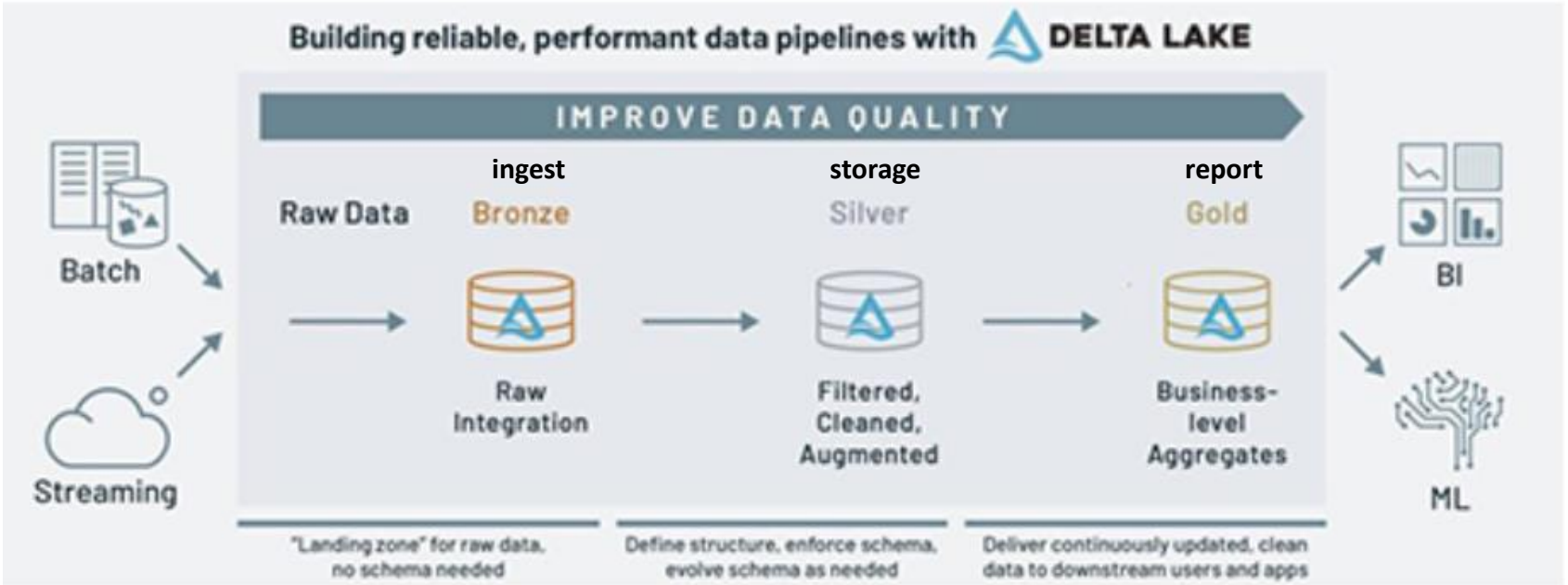


Implementation with Azure Synapse

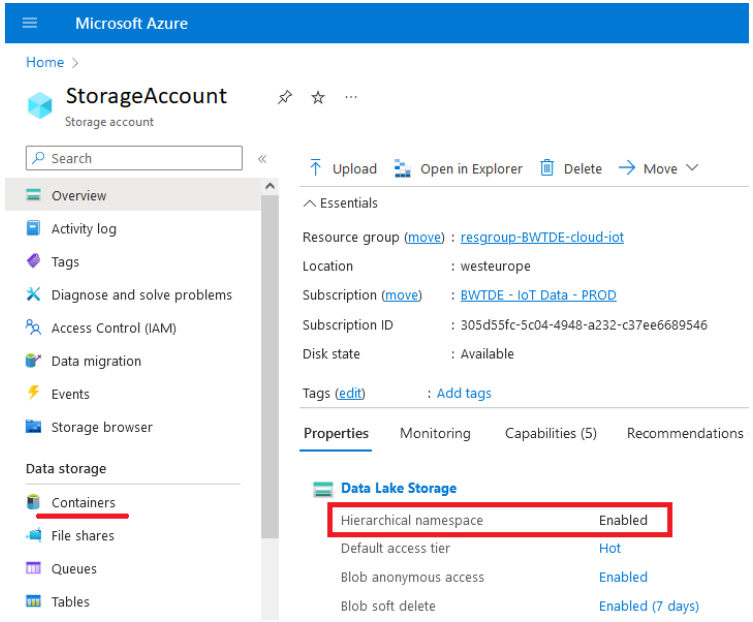


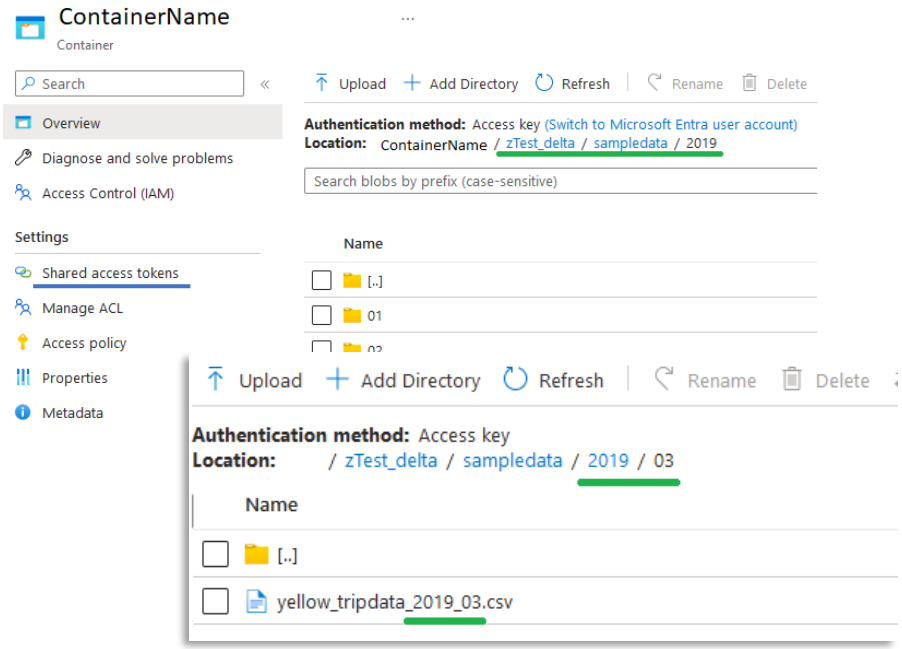

Data processing
Serving
Reporting
Apache Spark pools
Databricks
Apache Spark pools
DELTA LAKE
DELTA LAKE
Bronze
Silver
Gold
Synapse Serverless SQL
Reporting, Analytics


Building reliable, performant data pipelines with DELTA LAKE
IMPROVE DATA QUALITY
Batch
Streaming
Raw Data
Bronze
Silver
Gold
Raw Integration
Filtered, Cleaned, Augmented
Business-level Aggregates
BI
ML
"Landing zone" for raw data, no schema needed
Define structure, enforce schema, evolve schema as needed
Deliver continuously updated, clean data to downstream users and apps

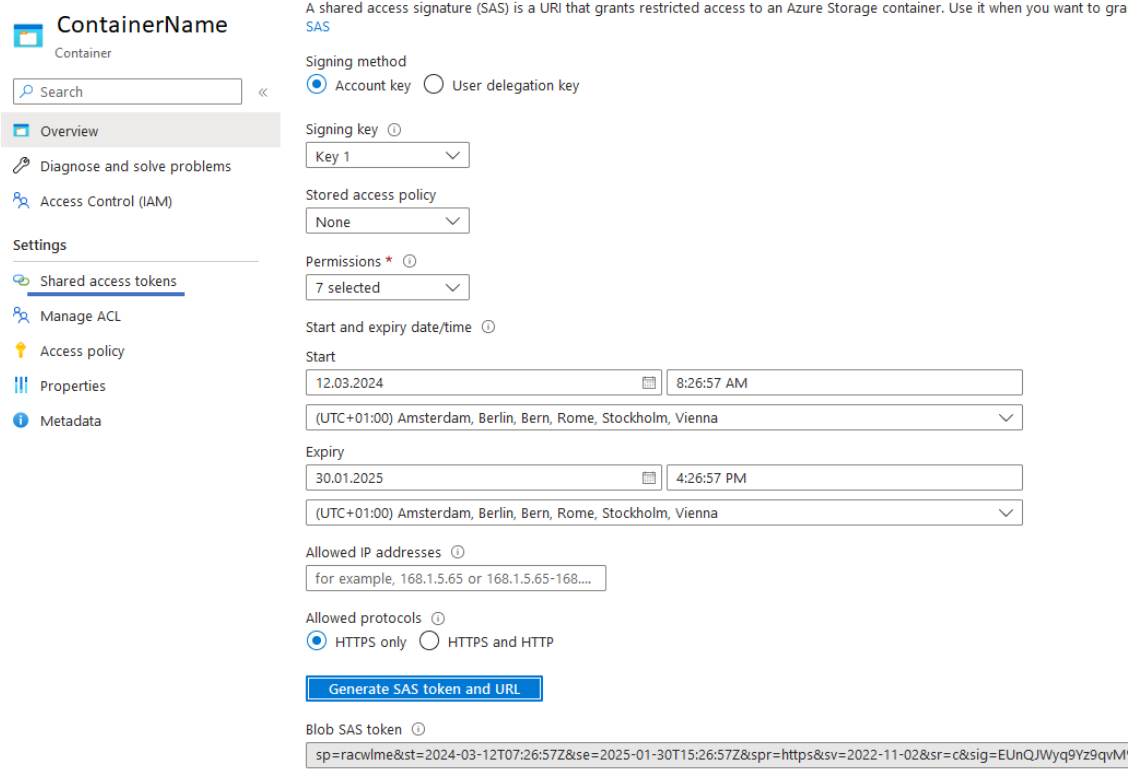

Building reliable, performant data pipelines with DELTA LAKE
IMPROVE DATA QUALITY
ingest
storage
report
Raw Data
Bronze
Silver
Gold
Batch
Streaming
Raw Integration
Filtered, Cleaned, Augmented
Business-level Aggregates
BI
ML
"Landing zone" for raw data, no schema needed
Define structure, enforce schema, evolve schema as needed
Deliver continuously updated, clean data to downstream users and apps

Blob Folders

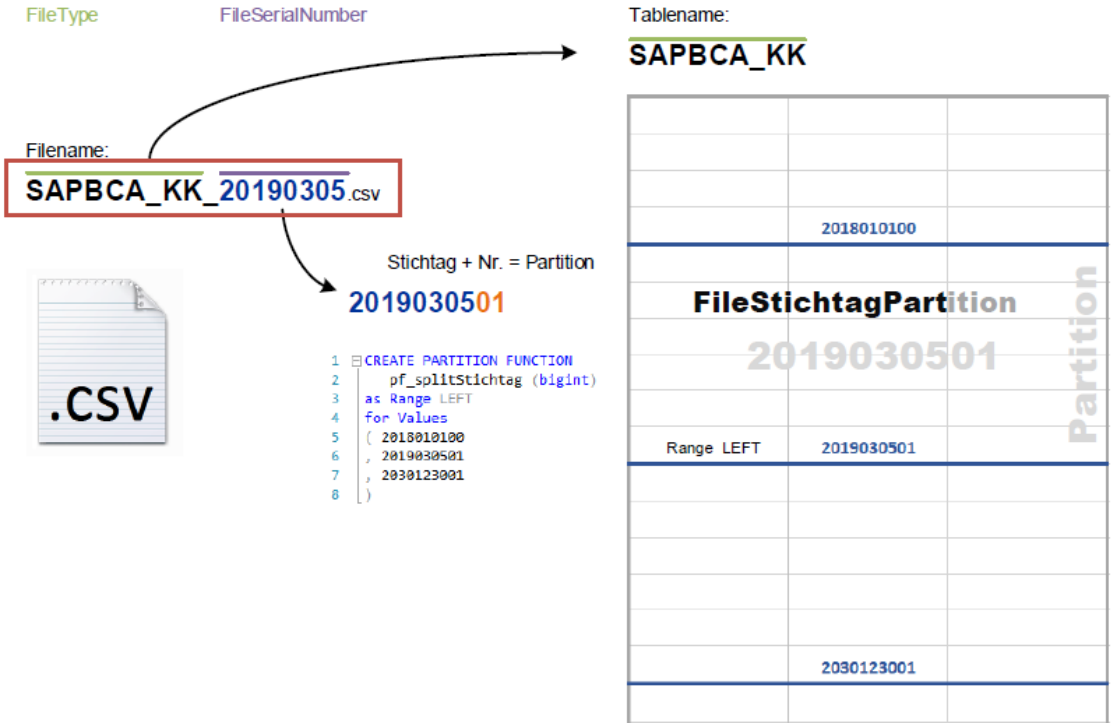




Blob Folders SAS



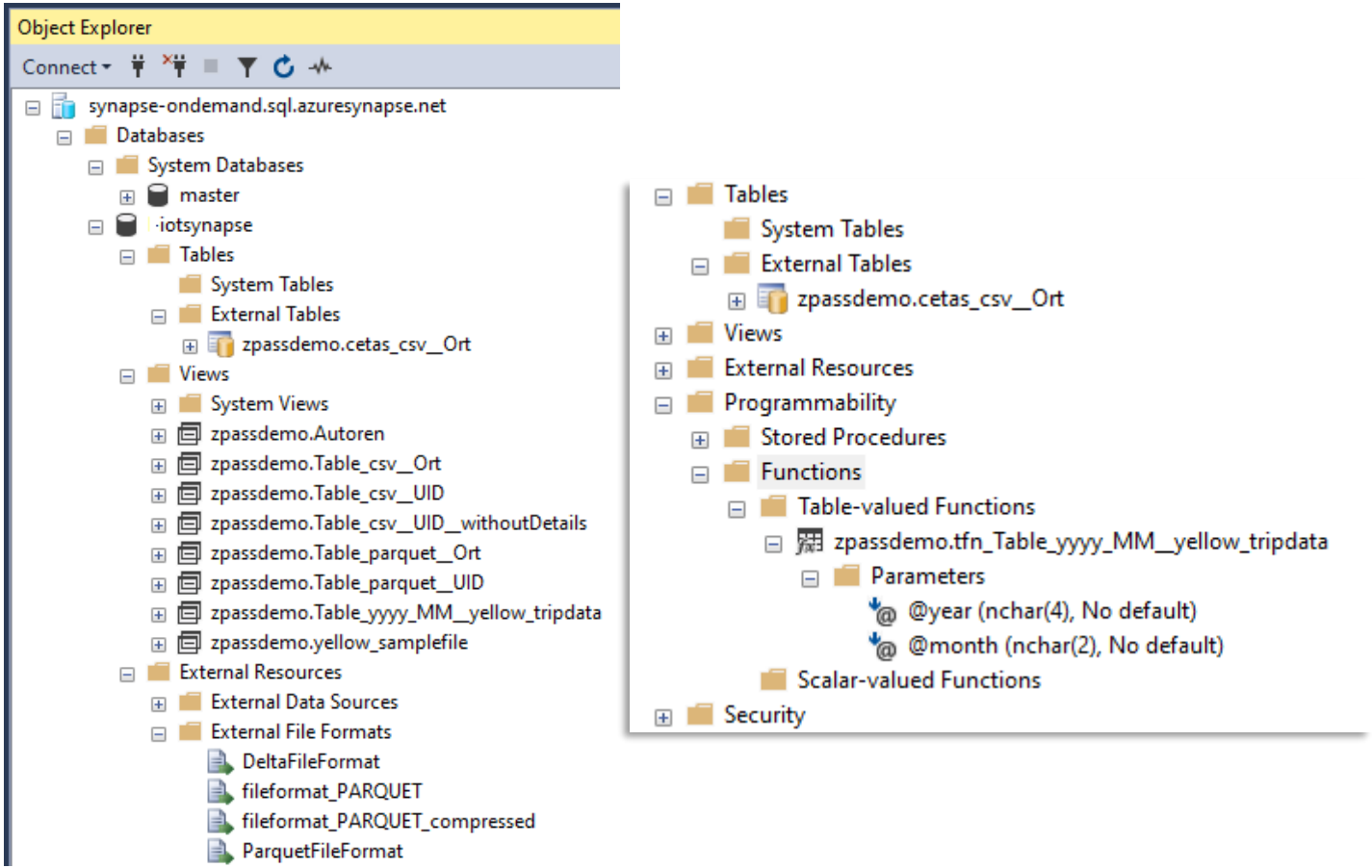
früher „Partition“ by date



DEMO

Beispiel 1

tsql_files 01 ... 04


Object Explorer
Connect
synapse-ondemand.sql.azuresynapse.net
Databases
System Databases
master
iotsynapse
Tables
System Tables
External Tables
zpassdemo.cetas_csv__Ort
Views
System Views
zpassdemo.Autoren
zpassdemo.Table_csv__Ort
zpassdemo.Table_csv__UID
zpassdemo.Table_csv__UID__withoutDetails
zpassdemo.Table_parquet__Ort
zpassdemo.Table_parquet__UID
zpassdemo.Table_yyyy_MM__yellow_tripdata
zpassdemo.yellow_samplefile
External Resources
External Data Sources
External File Formats
DeltaFileFormat
fileformat_PARQUET
fileformat_PARQUET_compressed
ParquetFileFormat
Tables
System Tables
External Tables
zpassdemo.cetas_csv__Ort
Views
External Resources
Programmability
Stored Procedures
Functions
Table-valued Functions
zpassdemo.tfn_Table_yyyy_MM__yellow_tripdata
Parameters
@year (nchar(4), No default)
@month (nchar(2), No default)
Scalar-valued Functions
Security

```sql
CREATE OR ALTER   FUNCTION [zpassdemo].[tfn_Table_yyyy_MM__yellow_tripdata]
                ( @year   nchar(4)
                , @month  nchar(2)
                )
RETURNS TABLE
AS RETURN
(
SELECT
       count(*)                       as 'cnt_Rows'
     , t.filepath(3)                  as 'blob_yyyy_MM'    ---- (3) = dritter "*"
     , t.filename()                   as 'blob_FileName'
     , t.filepath()                   as 'blob_FilePath'
FROM
    OPENROWSET( BULK '/zTest_delta/sampledata/*/*/yellow_tripdata_*.csv'
                  , DATA_SOURCE = 'eds_testfabrikpassdemo_testfabrikfilesystem'
                  , FORMAT      = 'CSV'
                  , FIELDTERMINATOR = ','
                  --, ROWTERMINATOR = '\n'
                  , PARSER_VERSION = '2.0'
                  , HEADER_ROW = TRUE
                  -- [ <reject_options> ]
                  )
                    as [t]
WHERE
       t.filepath(1) IN ( @year )
  and  t.filepath(2) IN ( @month )
GROUP  by
       t.filepath(3)
     , t.filename()
     , t.filepath()
)
--  END FUNCTION
```

SQL

PASS DEUTSCHLAND e.V. Microsoft Data Platform Community

```sql
CREATE OR ALTER   FUNCTION [zpassdemo].[tfn_Table_yyyy_MM__yellow_tripdata]
                ( @year   nchar(4)
                , @month  nchar(2)
                )
RETURNS TABLE
AS RETURN
(
SELECT
       count(*)                     as 'cnt_Rows'
     , t.filepath(3)                as 'blob_yyyy_MM'   ---- (3) = dritter "*"
     , t.filename()                 as 'blob_FileName'
     , t.filepath()                 as 'blob_FilePath'
FROM
    OPENROWSET( BULK '/zTest_delta/sampledata/*/*/yellow_tripdata_*.csv'
                  , DATA_SOURCE = 'eds_testfabrikpassdemo_testfabrikfilesystem'
                  , FORMAT      = 'CSV'
                  , FIELDTERMINATOR = ','
                  --  ROWTERMINATOR = '\n'
      , t.filename()
      , t.filepath()
)
--  END FUNCTION
```

```sql
SELECT *
FROM   [zpassdemo].[tfn_Table_yyyy_MM__yellow_tripdata] ( '2019' , '08'  )
GO
```

100 %

Results | Messages

	cnt_Rows	blob_yyyy_MM	blob_FileName	blob_FilePath
1	6073357	2019_08	yellow_tripdata_2019_08.csv	/zTest_delta/sampledata/2019/08/yellow_tripdata_2019_08.csv

DEMO

Beispiel 2

tsql_admin files

CETAS **C**reate**E**xternal**T**able**A**s**S**elect

?? Warum kein SELECT ... INTO

CETAS CreateExternalTableAsSelect

```
1  CREATE EXTERNAL TABLE zpassdemo.cetas_csv__Ort
2  WITH
3      ( LOCATION    = '/report/2024/03/12/cetas_csv__Ort'
4      , DATA_SOURCE = eds_testfabrikpassdemo_testfabrikfilesystem
5      , FILE_FORMAT = PARQUET
6      )
7  AS
8  SELECT *
9  FROM   zpassdemo.Table_csv__Ort;
10
```

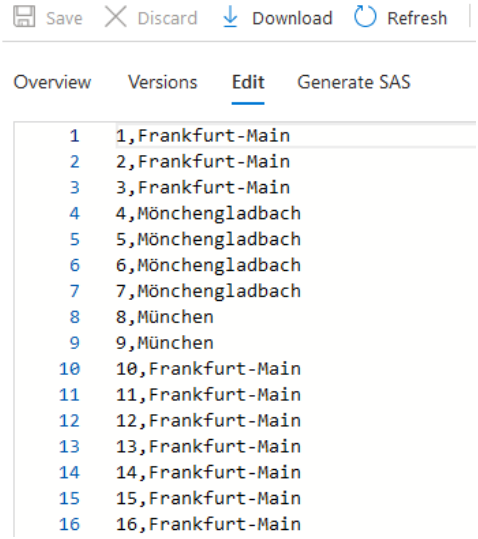
```
CREATE OR ALTER PROC [dbo].[usp_cetas_from_VIEW]
              @currDatetime smalldatetime
            , @viewName     sysname   -- = 'zpassdemo.Table_csv__Ort'
AS
BEGIN
----
DECLARE @StorageAccount  sysname  = 'testfabrikpassdemo'
      , @Container       sysname  = 'testfabrikfilesystem'

DECLARE @sqlCmd nvarchar(4000)
        --SELECT  @currDatetime  -- sys.time_zone_info --> UTC
----
SET @sqlCmd = '
IF EXISTS (select * from sys.objects where object_id = OBJECT_ID(N''' + Replace( @viewName, 'Table', 'cetas' ) + ''') )
                                          DROP EXTERNAL TABLE  ' + Replace( @viewName, 'Table', 'cetas' ) + ';
' --
--select @sqlCmd
EXEC sp_executesql @sqlCmd

SET @sqlCmd = N'
CREATE EXTERNAL TABLE ' + Replace( @viewName, 'Table', 'cetas' ) +'
WITH
(
  LOCATION    = ''/report/' + Format( @currDatetime, 'yyyy/MM/dd/HHmm__' ) + Replace( Replace( @viewName, 'Table', 'cetas'
                                                                                              , 'zpassdemo.', '' )
                                                                                +'.parquet''  -- plus appropriate fileExtension
, DATA_SOURCE = eds_' + @StorageAccount + '_' + @Container + '
, FILE_FORMAT = ParquetFileFormat
)
AS
SELECT *
FROM   ' + @viewName + ';
' --
--select @sqlCmd
EXEC sp_executesql @sqlCmd

END  -- end PROC
GO
```

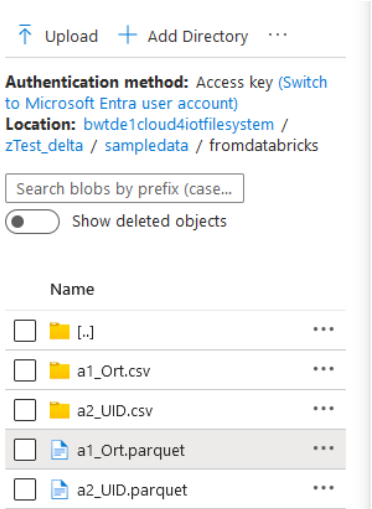
DEMO

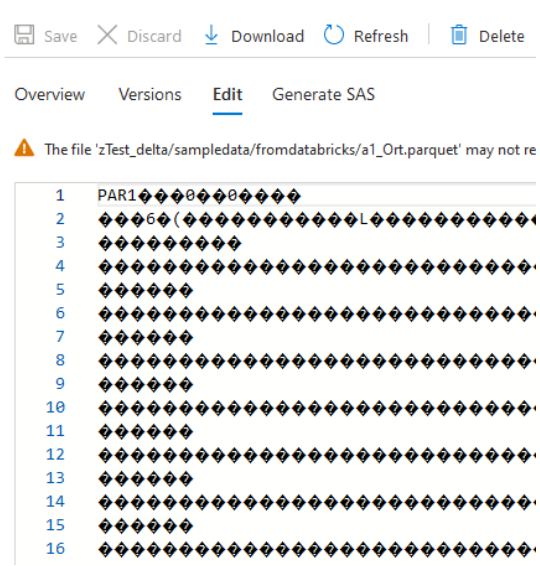
Beispiel 3

tsql_cetas files

.csv vs .parquet







.csv vs .parquet

Save Discard Download Refresh

Overview Versions Edit Generate SAS

```
1  1,Frankfurt-Main
2  2,Frankfurt-Main
3  3,Frankfurt-Main
4  4,Mönchengladbach
5  5,Mönchengladbach
6  6,Mönchengladbach
7  7,Mönchengladbach
8  8,München
9  9,München
10 10,Frankfurt-Main
11 11,Frankfurt-Main
12 12,Frankfurt-Main
13 13,Frankfurt-Main
14 14,Frankfurt-Main
15 15,Frankfurt-Main
16 16,Frankfurt-Main
```

Block blob 1.9 MiB

Upload Add Directory ···

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: bwtde1cloud4iotfilesystem / zTest_delta / sampledata / fromdatabricks

Search blobs by prefix (case...

Show deleted objects

Name
[..]
a1_Ort.csv
a2_UID.csv
a1_Ort.parquet
a2_UID.parquet

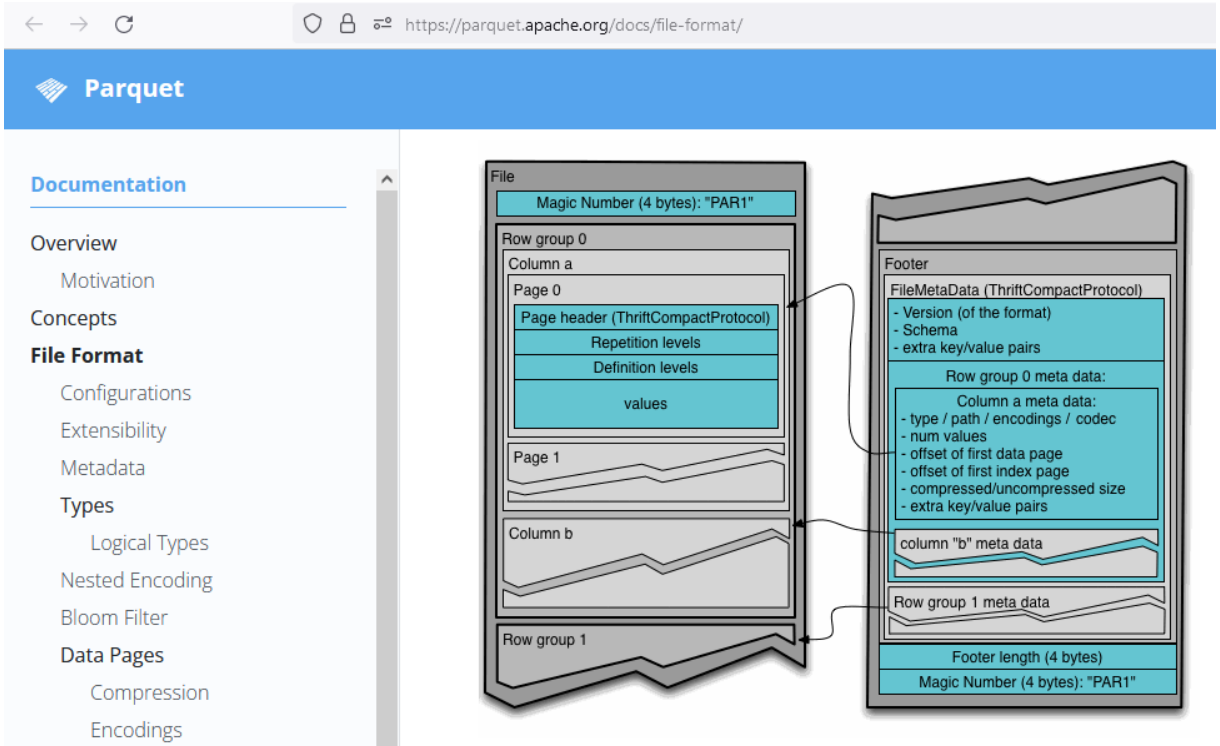
Save Discard Download Refresh Delete

Overview Versions Edit Generate SAS

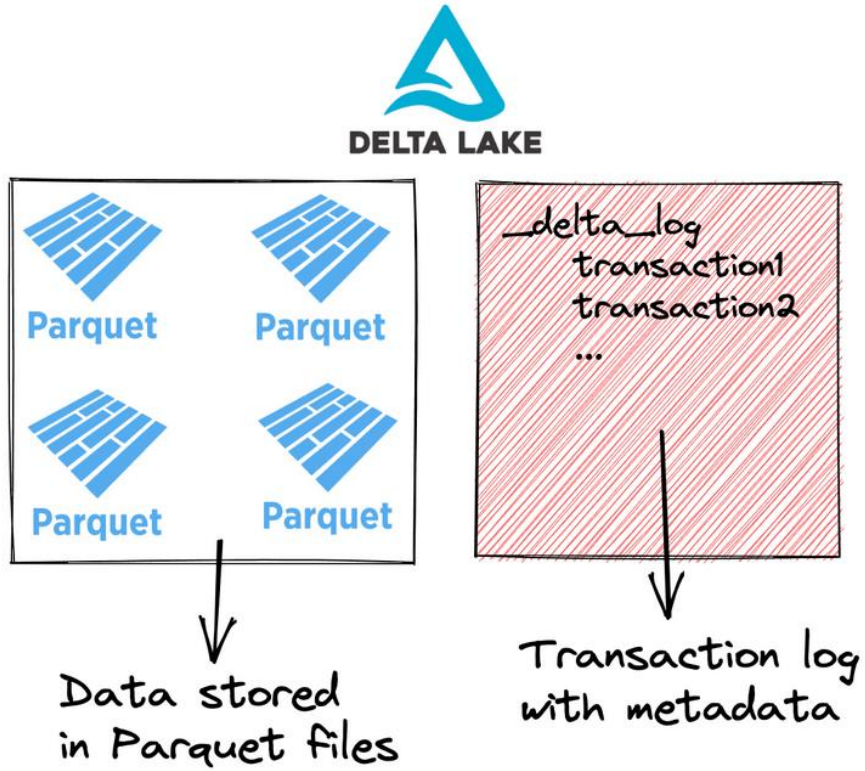
⚠ The file 'zTest_delta/sampledata/fromdatabricks/a1_Ort.parquet' may not re

```
1  PAR1▯▯▯0▯▯▯0▯▯▯▯
2  ▯▯▯6▯(▯▯▯▯▯▯▯▯▯▯▯L▯▯▯▯▯▯▯▯▯▯▯
3  ▯▯▯▯▯▯▯▯▯▯
4  ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
5  ▯▯▯▯▯▯▯
6  ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
7  ▯▯▯▯▯▯▯
8  ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
9  ▯▯▯▯▯▯▯
10 ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
11 ▯▯▯▯▯▯▯
12 ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
13 ▯▯▯▯▯▯▯
14 ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
15 ▯▯▯▯▯▯▯
16 ▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
```

Block blob 391.48 KiB



https://parquet.apache.org/docs/file-format/



https://delta.io/blog/delta-lake-vs-parquet-comparison/

SQL

PASS DEUTSCHLAND e.V.
Microsoft Data Platform Community

DEMO

Beispiel 4

.parquet vs. .delta-Files

LINKS

- https://github.com/sql-fabrik/AzureSynapse_gettingStarted
- >> README.md

- https://www.databricks.com/de/glossary/medallion-architecture
- https://learn.microsoft.com/de-de/azure/synapse-analytics/sql/develop-tables-cetas
- https://parquet.apache.org/docs/file-format/
- https://www.databricks.com/de/glossary/what-is-parquet
- https://delta.io/blog/delta-lake-vs-parquet-comparison/

THANK YOU TO OUR SPONSORS

PASS DEUTSCHLAND e.V.
Microsoft Data Platform Community

pmOne

SOLARWINDS

smart casual datadesign

Mainzer Datenfabrik

b.telligent
smart data. smart decisions.

HMS analytical software

ORAYLIS
turn your data into value

TIMEXTENDER

Microsoft

ITGAIN

SOLISYON
www.solisyon.de

GDS BUSINESS INTELLIGENCE GmbH

Thank you very much for your attention.

Vielen Dank für Eure Aufmerksamkeit.