

Strange New Worlds

from SQL to Lakehouse and beyond

Claudia Huber

Do you use cloud services for data processing? – Raise you hands!



SQL KONFERENZ 2024

Not in the cloud (yet)? – Raise you hands!



SQL KONFERENZ 2024

Do you use cloud services? – A few numbers from 2023:

- A few numbers from 2023:
  - 46,5% of Germany's companies use cloud technology
  - 11,1% plan to use the cloud
  - 18,2% discuss about it

Source: https://www.ifo.de/pressemitteilung/2023-07-12/unternehmen-deutschland-cloud-computing



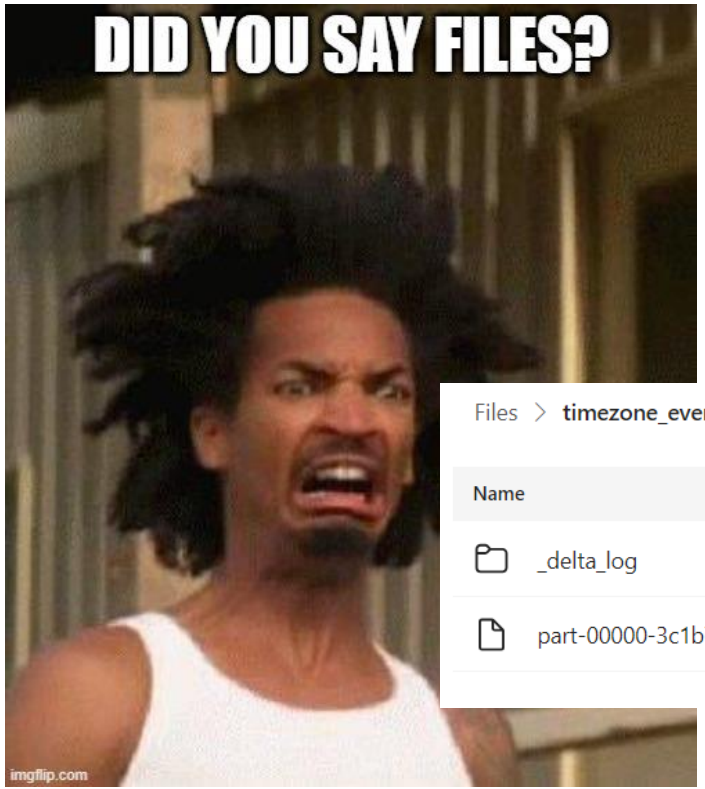


Let‘s start exploring the strangeness of the Lakehouse …

SQL KONFERENZ 2024

Everything is based on files!

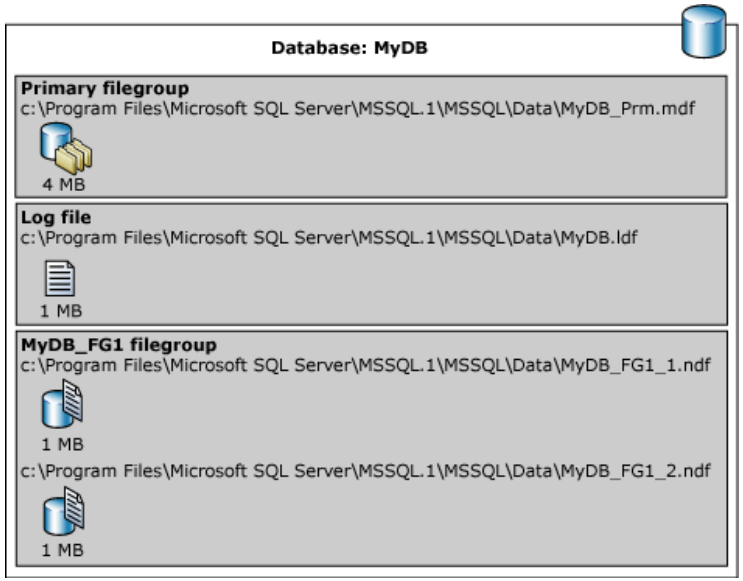




Everything is based on Files

… Delta ***Tables*** are just parquet file with meta data information!

SQL KONFERENZ 2024



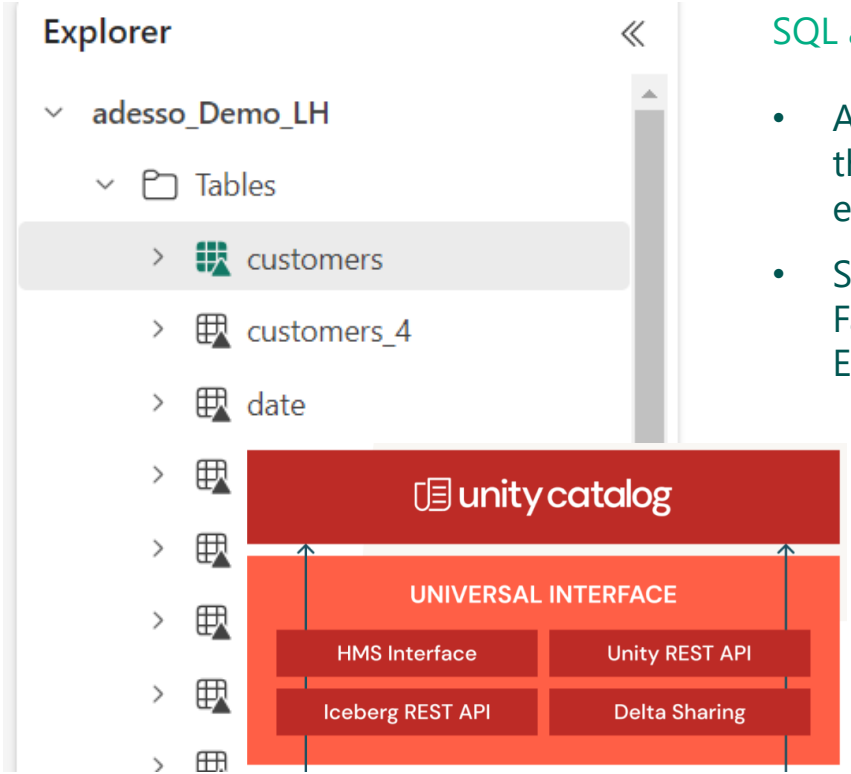
SQL Server uses files as well!

- … proprietary formats.
- We can now use open source formats instead!

SQL KONFERENZ 2024

„Database(s)“ and SQL





SQL and more

- Allow us to work with all those files like in the „old“ relational days organized in tables etc.
- Services like Azure Databricks and Microsoft Fabric have build-in Catalogs and SQL Endpoints – yeah ☺

SQL KONFERENZ 2024

Transactions



ACID transactions

- Were unavailable in the early days of Data Lakes.
- Delta Table to the rescue!*

*One caveat: Transactions are always table scoped!



Identity



Losing my identity?

- Spark and also Fabric Data Warehouse have no possibility to create „true identities“
- Workarounds exist for the lack of identity columns:
  - Hashes
  - logic to caluculate the last value and increase it

Example Fabric Warehouse:

```
--Assign the last identifier value to a variable
--If the table doesn't contain any rows, assign zero to the variable
DECLARE @MaxID AS BIGINT;

IF EXISTS(SELECT * FROM [dbo].[Orders_with_Identifier])
    SET @MaxID = (SELECT MAX([Row_ID]) FROM [dbo].[Orders_with_Identifier]);
ELSE
    SET @MaxID = 0;
```



And now? Getting Started



You can do it!



- **SQL Knowledge** is a good starting point!
- Learn **Python** for Data Engineering – you can focus on basic python and **Spark** to start
- **Graphical components** for pipelines and orchestration still exist!

SQL KONFERENZ 2024


SQL
KONFERENZ
2024

Thank you

SQL KONFERENZ 2024



Claudia Huber

Teamlead Data & Analytics

adesso Austria GmbH