**Delta & Databricks for the DBA**

Once upon a time we had the Data Warehouse, life was good but it had its limitations, particularly around loading/storing complex data types. As data grew larger and more varied, the warehouse became too rigid and opinionated.  
  
So we started using Data Lakes to store data, and tools such as Databricks to do our compute. Things were good, but we missed some of the good times that the Data Warehouse had given us. The lake had become too flexible, we needed stability in our life. In particular, we needed A.C.I.D (Atomicity, Consistency, Isolation, and Durability) Transactions.  
  
Delta Lake, hosted by the Linux Foundation, is an open-source file layout protocol for giving us back those good times, whilst retaining all of the flexibility of the lake. Delta has gone from strength to strength, and in 2022 Databricks finally open-sourced the entire code-base, including lots of advanced features that were previously Databricks-only.

Databricks was developed in 2013: A Man named Matei Zaharia, along with colleagues in UC Berkley College, invented a distributed data compute tool called Spark in 2010, and after donating this to the Apache Foundation he created Databricks, a more commercial (paid for) offering of Spark. Both Spark and Databricks have proved invaluable for data processing, and ultimately incredibly popular.

This is all great, but how do those who have been using traditional Warehousing tools, in particular SQL Server, make the leap to Delta and Databricks? In this session we will explore this question. We will do direct comparisons between features/functionality, illustrating how these different tools are ultimately the same and very different at the same time.