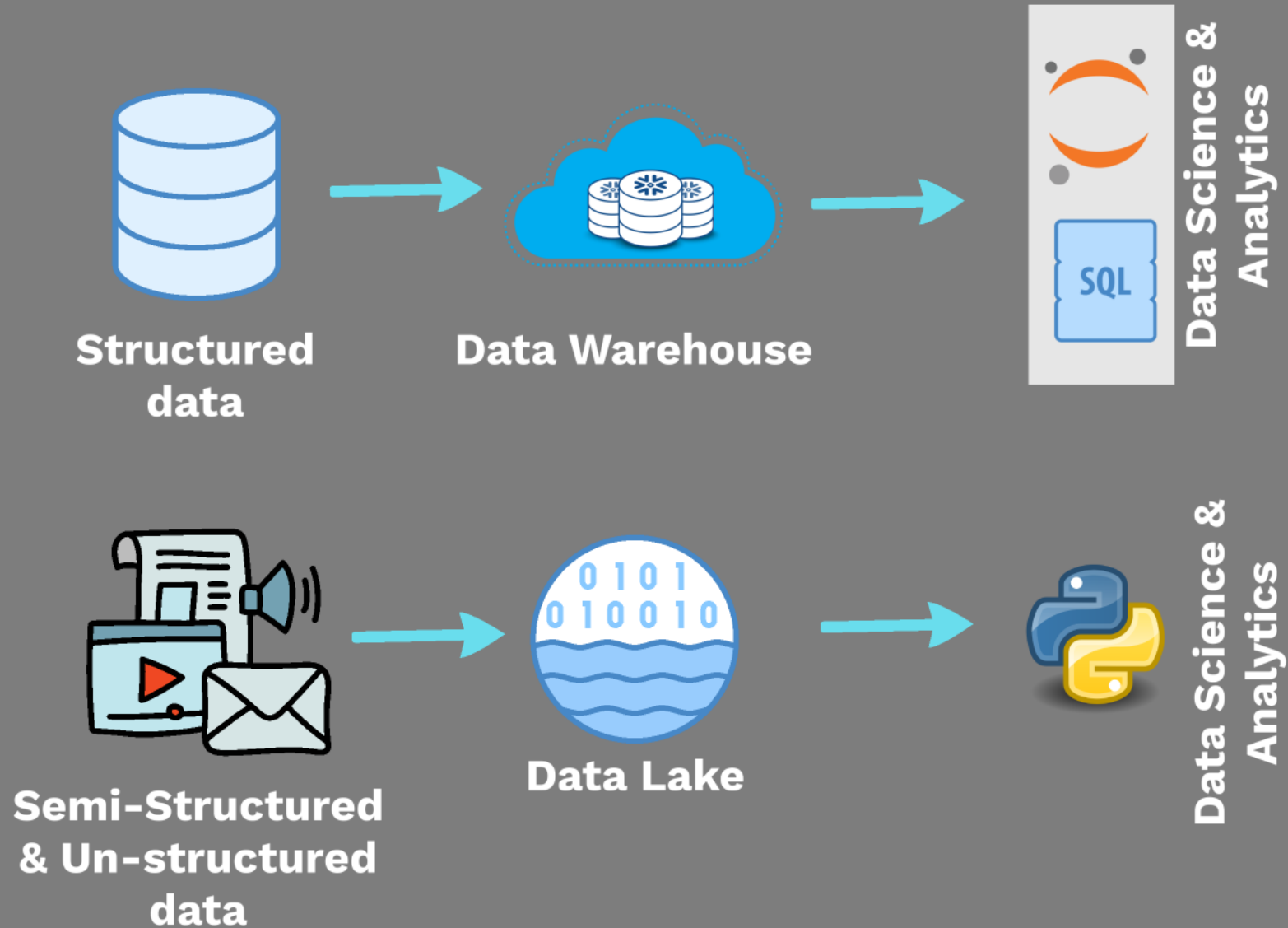
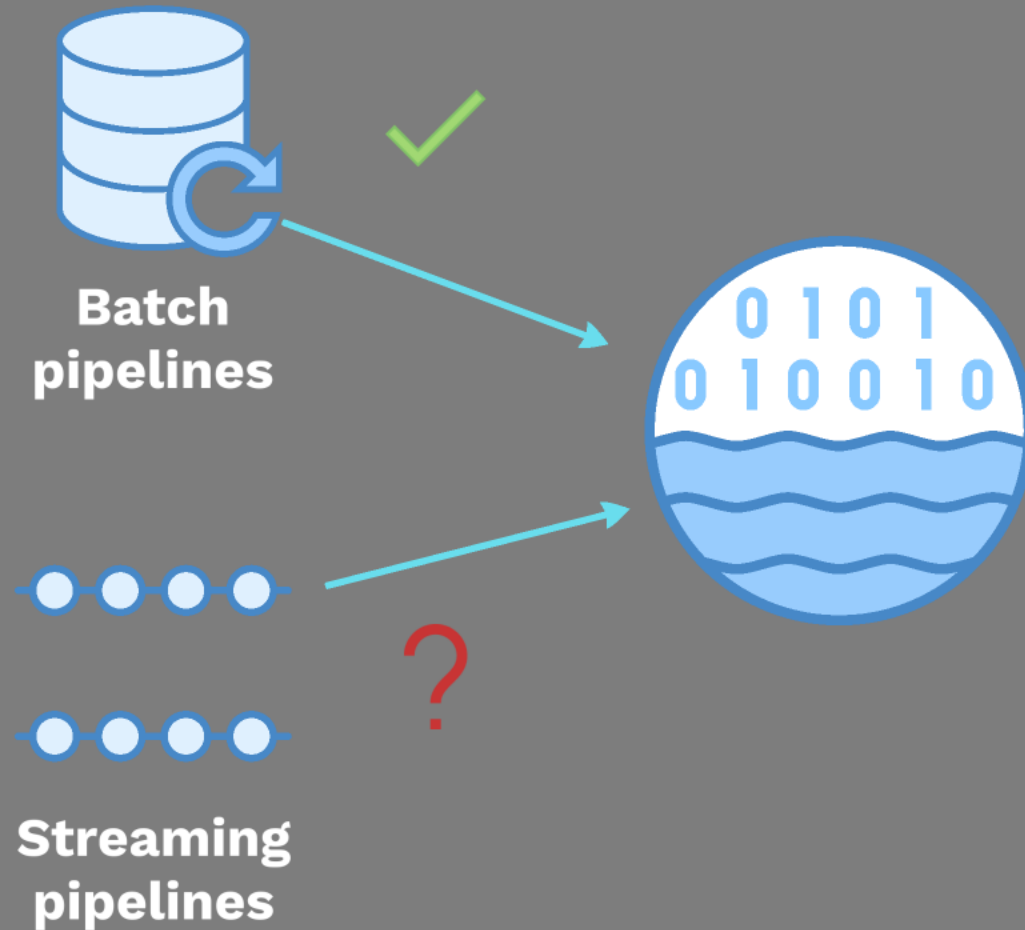


Data Silos



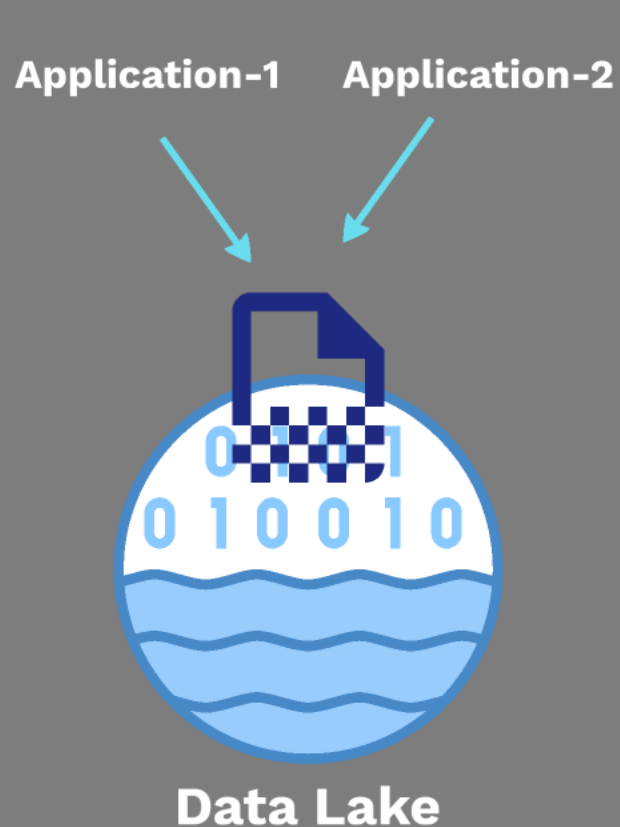
Batch & Streaming support



Files in data lake are **immutable**. Once written files cannot be edited. New data means new files or replacing files. Hence storing data from streaming data sources in a data lake poses serious challenges.

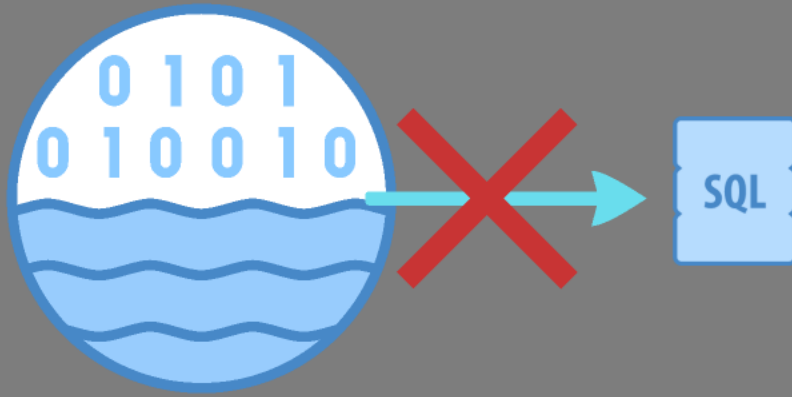
ACID guarantee on the data lake

- **Many applications expect ACID guarantees from data transactions (read or write) made on the data lake.**



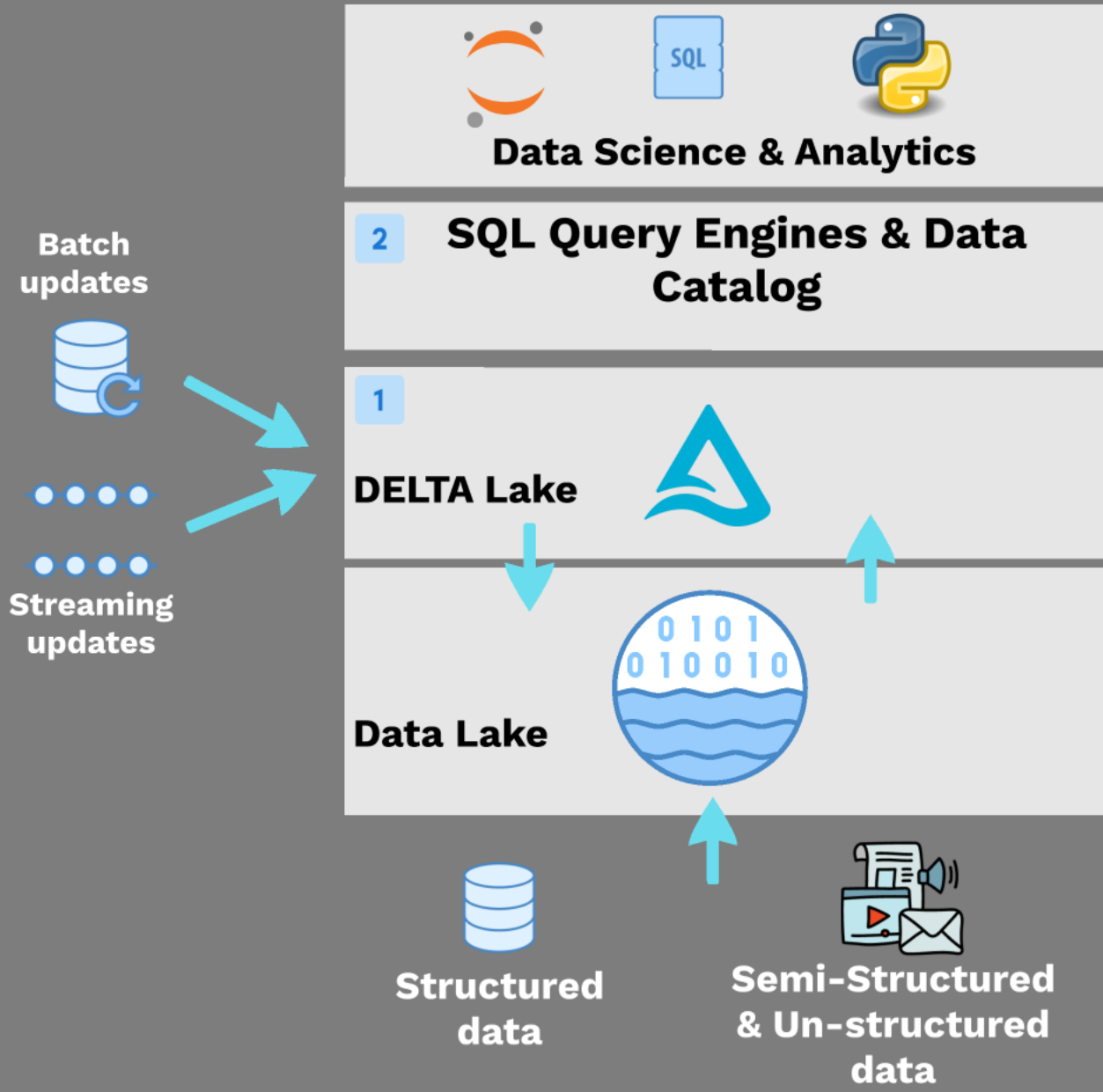
- **A** - *Atomicity* - transactions should succeed or fail completely. No partial transactions.
- **C** - *Consistency* - Even though multiple applications are updating a file, the state of the data should appear uniform .
- **I** - *Isolation* - Simultaneous operations should not conflict with each other.
- **D** - *Durability* - committed changes should be permanent.

No in built SQL engine



- Data analysts prefer SQL to query and process data.
- A data lake does NOT have an inbuilt SQL engine.

Data Lakehouse



2 SQL query engines provide analysts opportunity to query data with SQL queries. Data Catalog helps data scientists to locate useful datasets in the data lake.

1 Delta lake is an open source software that helps maintain a transaction log for writes being made on files stored in a data lake. Delta lake helps the data lake become

- ACID compliant
- Allows Time travel with versioning
- Schema enforcement like a data warehouse
- supports merge, update and delete operations on data stored in the data lake.