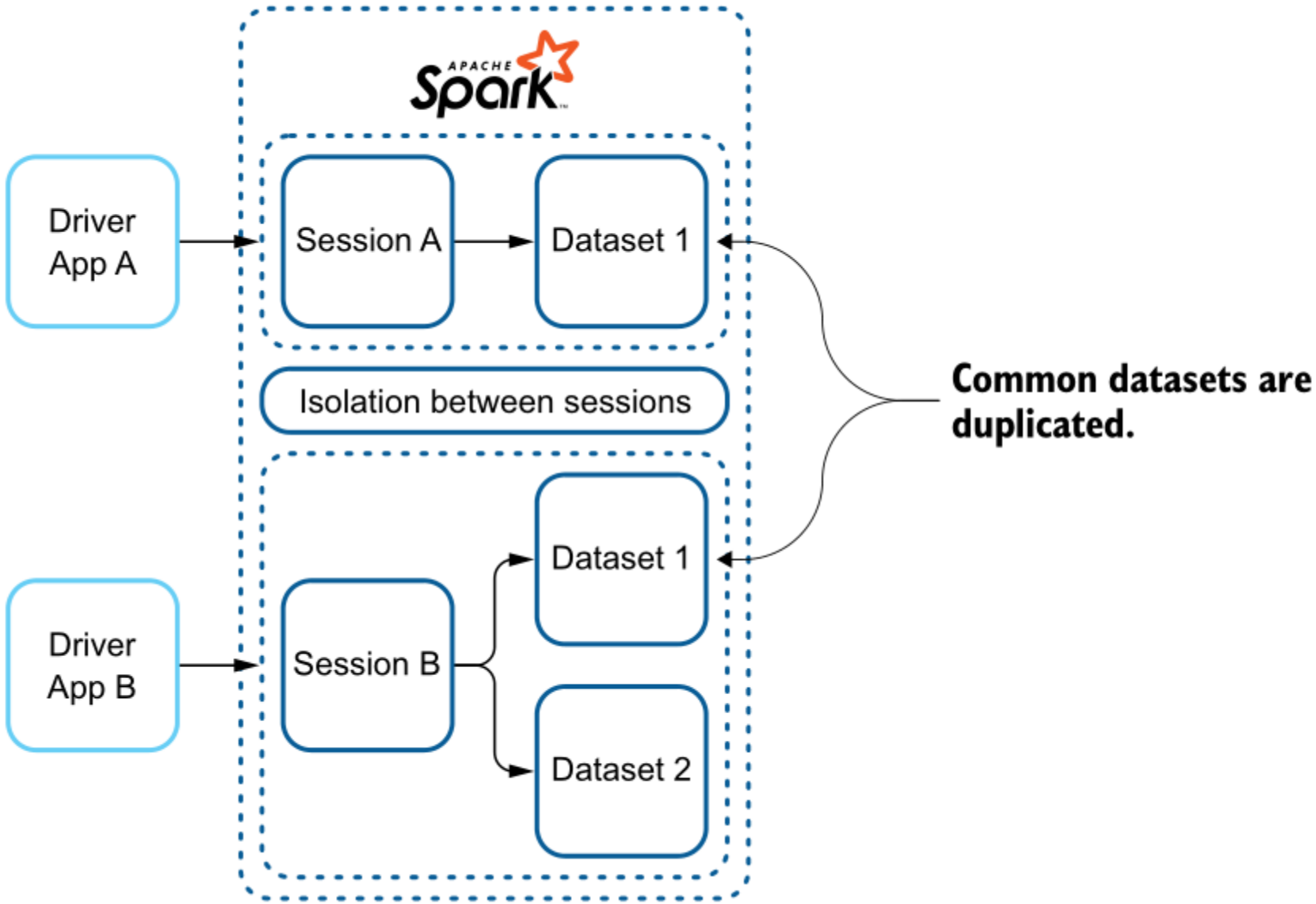
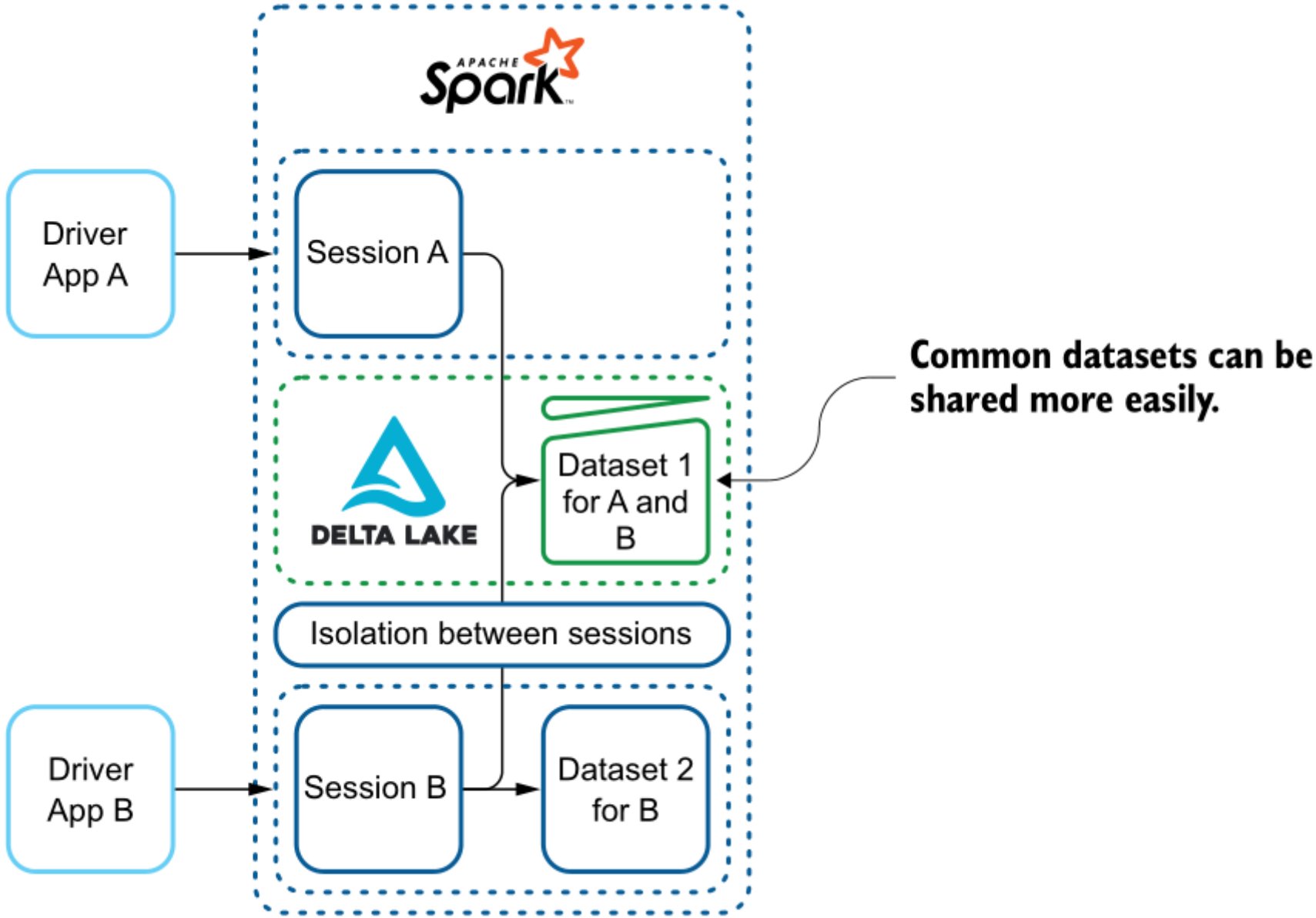
Build Lakehouses with Delta Lake

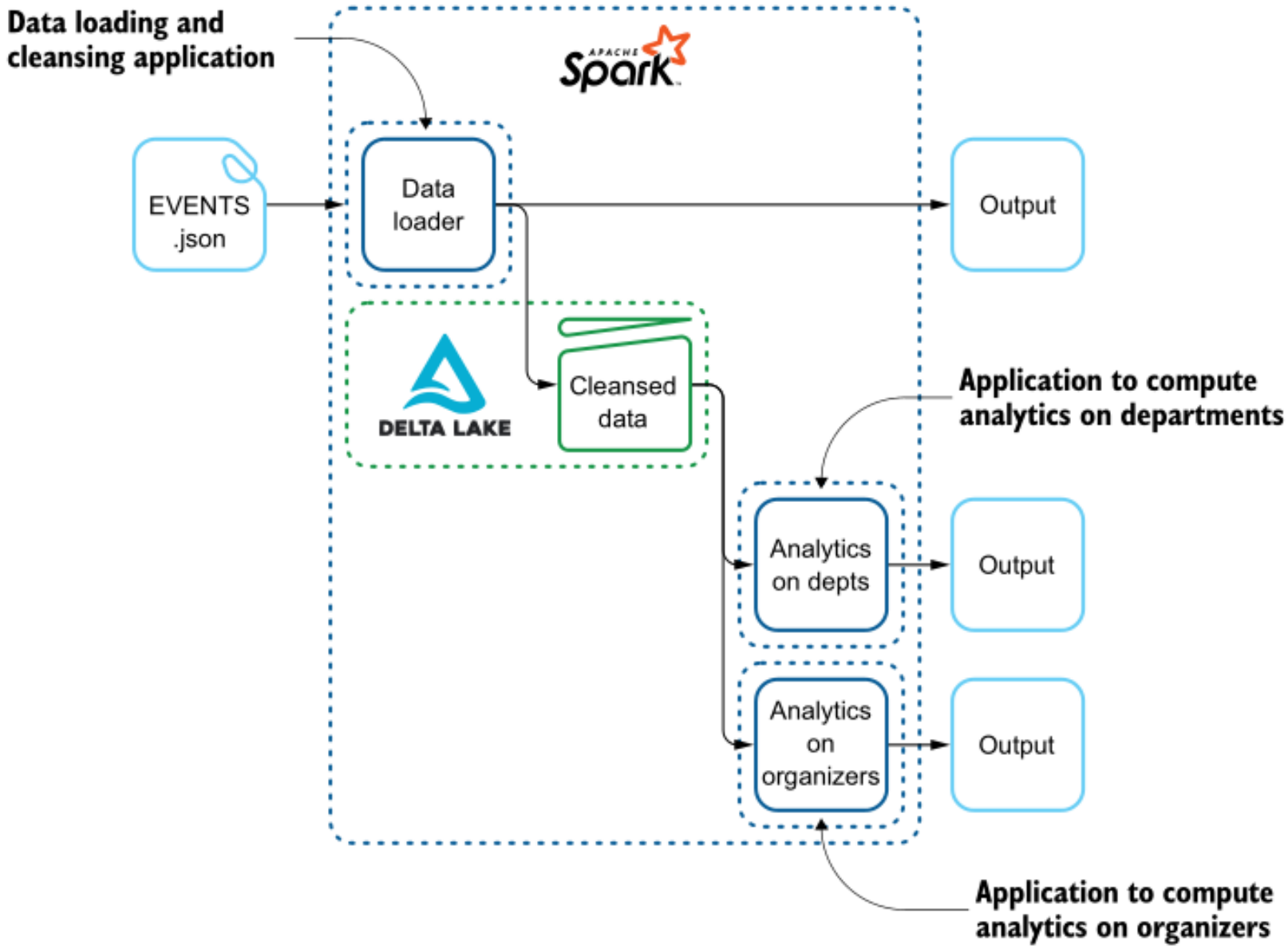
Delta Lake is an open-source storage framework that enables building a format agnostic Lakehouse architecture with compute engines including Spark, PrestoDB, Flink, Trino, Hive, Snowflake, Google BigQuery, Athena, Redshift, Databricks, Azure Fabric and APIs for Scala, Java, Rust, and Python. With Delta Universal Format aka UniForm, you can read now Delta tables with Iceberg and Hudi clients.

Get Started

GitHub Releases Roadmap


APACHE Spark
Driver App A
Session A
Dataset 1
Isolation between sessions
Driver App B
Session B
Dataset 1
Dataset 2
Common datasets are duplicated.


APACHE Spark
Driver App A
Session A
DELTA LAKE
Dataset 1 for A and B
Common datasets can be shared more easily.
Isolation between sessions
Driver App B
Session B
Dataset 2 for B

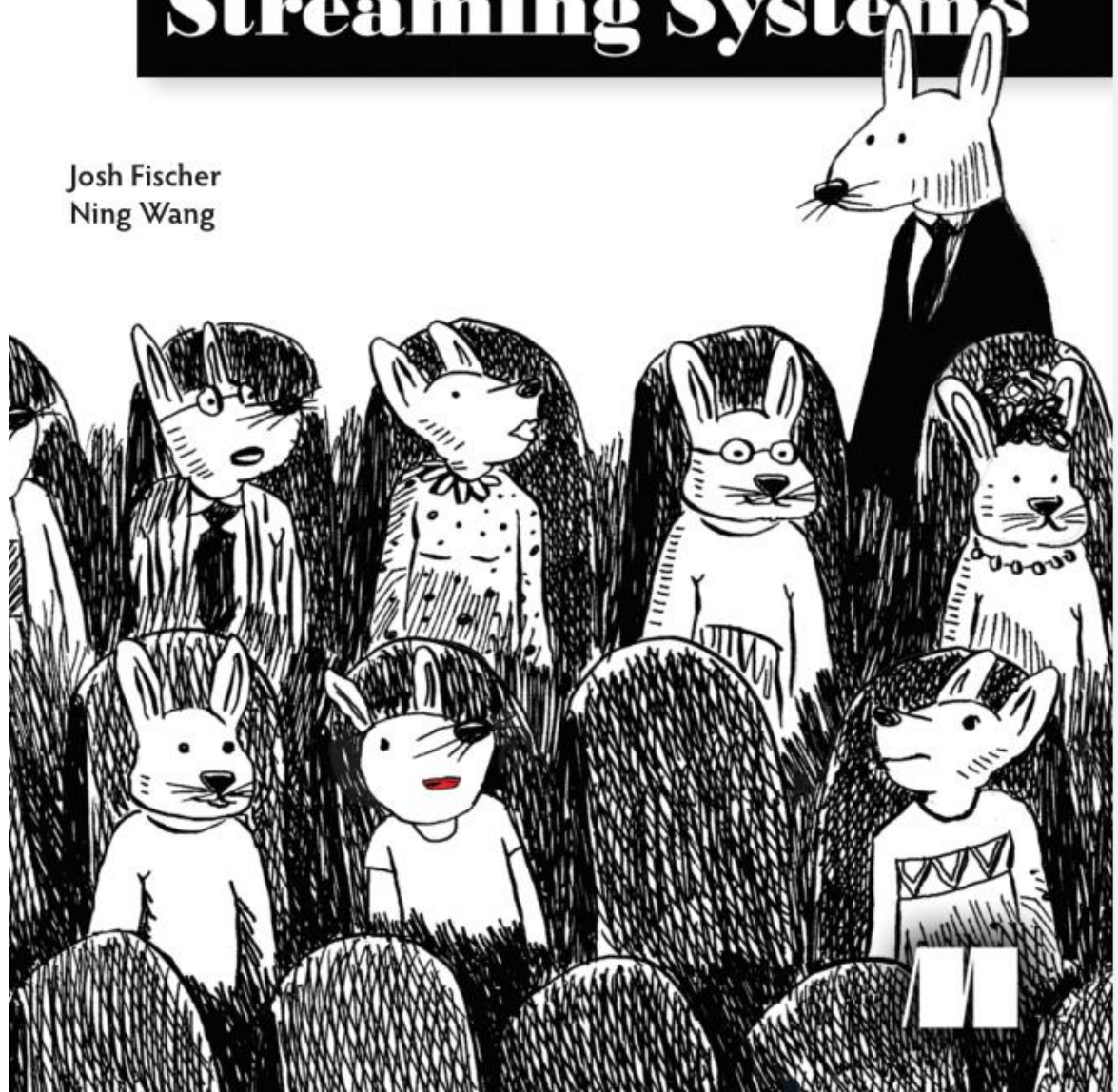

Data loading and cleansing application
APACHE Spark
EVENTS .json
Data loader
Output
DELTA LAKE
Cleansed data
Application to compute analytics on departments
Analytics on depts
Output
Analytics on organizers
Output
Application to compute analytics on organizers

Real-time event processing

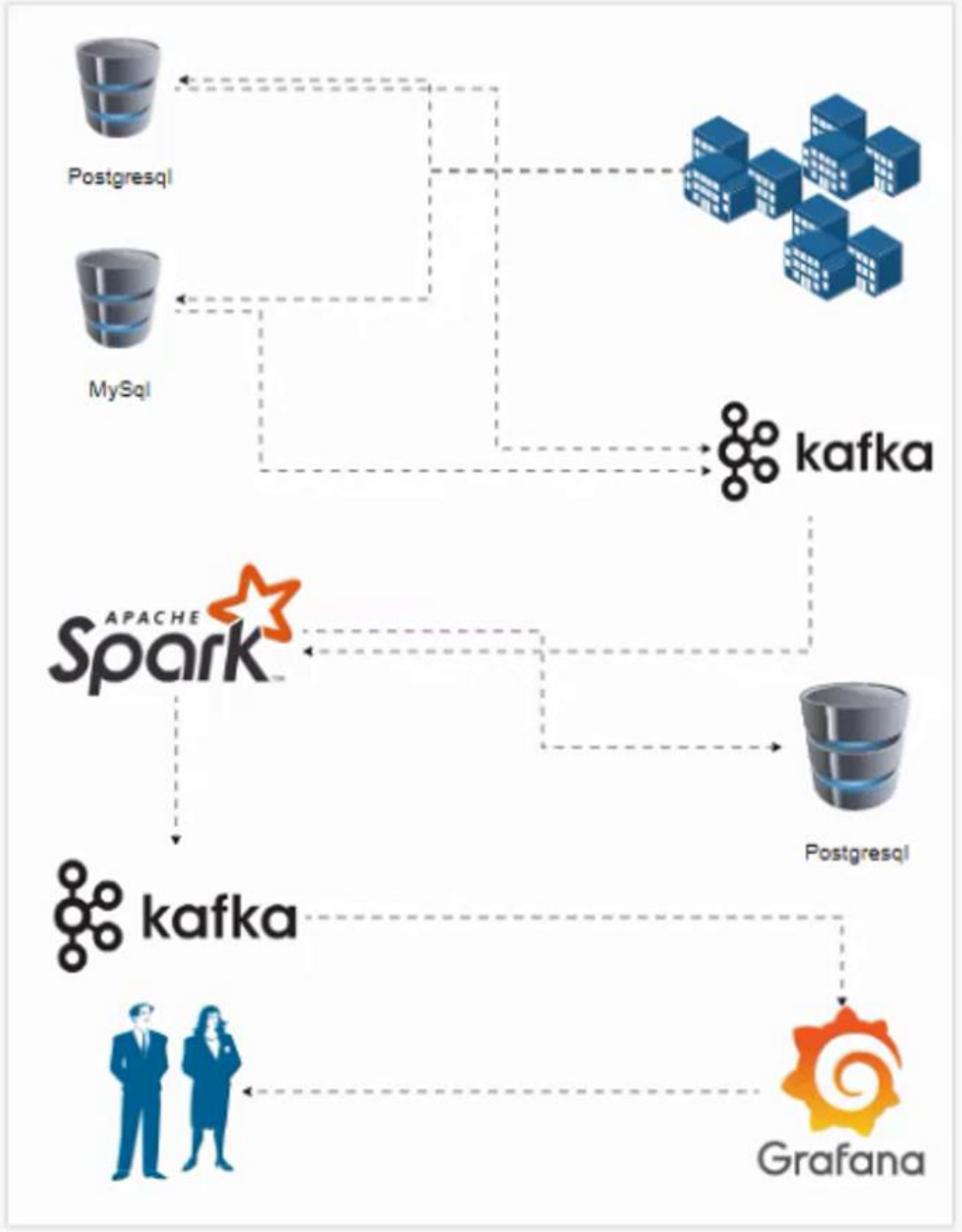
grokking

Streaming Systems

Josh Fischer
Ning Wang



One Interesting Project


Postgresql
MySql
kafka
APACHE
Spark
Postgresql
kafka
Grafana