End-to-End Machine Learning with Apache AsterixDB

Datasources

High Speed and Large Volume of Data

Expensive Analytics

Distributed Environment

Analysts w/o computer expertise

Wail Alkowalieet (UC Irvine), Sattam Alsubaiee (Center for Complex Engineering Systems at KACST and MIT), Michael Carey, Chen Li (UC Irvine), Heri Ramampiaro (Norwegian University of Science and Technology), Phanwadee Sinthong, Xikui Wang (UC Irvine)
Data Analytics with AsterixDB

Storage

Raw Tweets

Training Data

Machine Learning Libraries

Models

Query Evaluation Pipeline

Query

Feed

Ingestion Pipeline

{..., text: “I’m upset.”, ...}

{..., text: “I’m upset.”, sentiment: “Negative”, ...}

Data Analysts

Apache Zeppelin

Jupyter

010101010
010101010
010101010