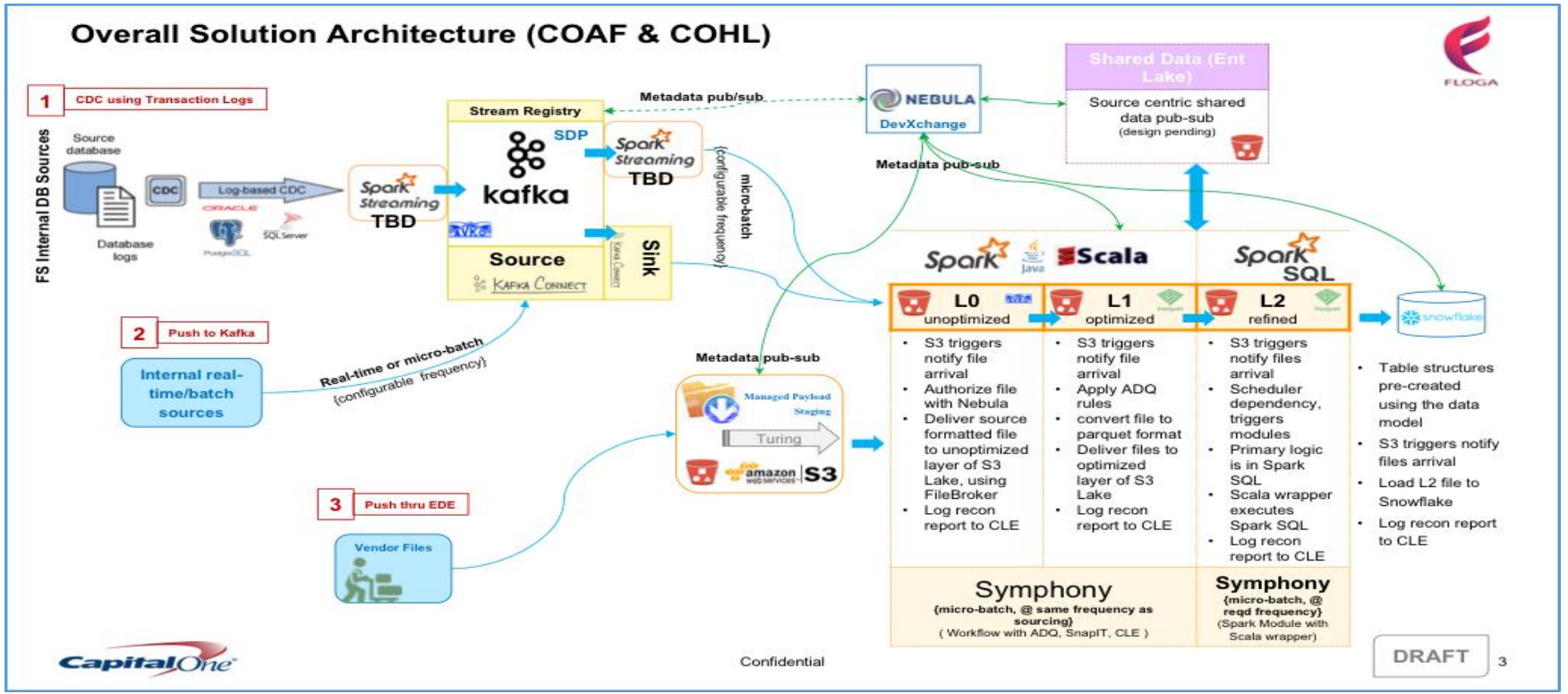
FLOGA

FLOGA is the enterprise-wide Data Transformation program tasked with creating a brand new data ecosystem that will simplify the producer and consumer experience. This program is responsible for creating this new data ecosystem for Financial Services by rebuilding data ecosystem from the ground up, in the cloud, using modern tools and technologies.

Benefits:

1. 100% Data in the Cloud
2. Enablement of Real-Time Data
3. Separation of Data and Computing
4. Self-Service Data Products
5. Accurate and Complete Metadata

FLOGA Architecture



As part of FLOGA, frameworks are developed to establish and automate data migration pipeline.

1. Streaming Data Ingestion
   - Platform to stream real time data from auto finance source systems into AWS S3
2. Code Generator
   - Micro-services to automate generation of SQL for loading data to Snowflake.
3. Scheduler Framework
   - Event based framework to execute SQLs generated by Code Generator.
4. Squadron
   - REST application to track progress of all the job executions.

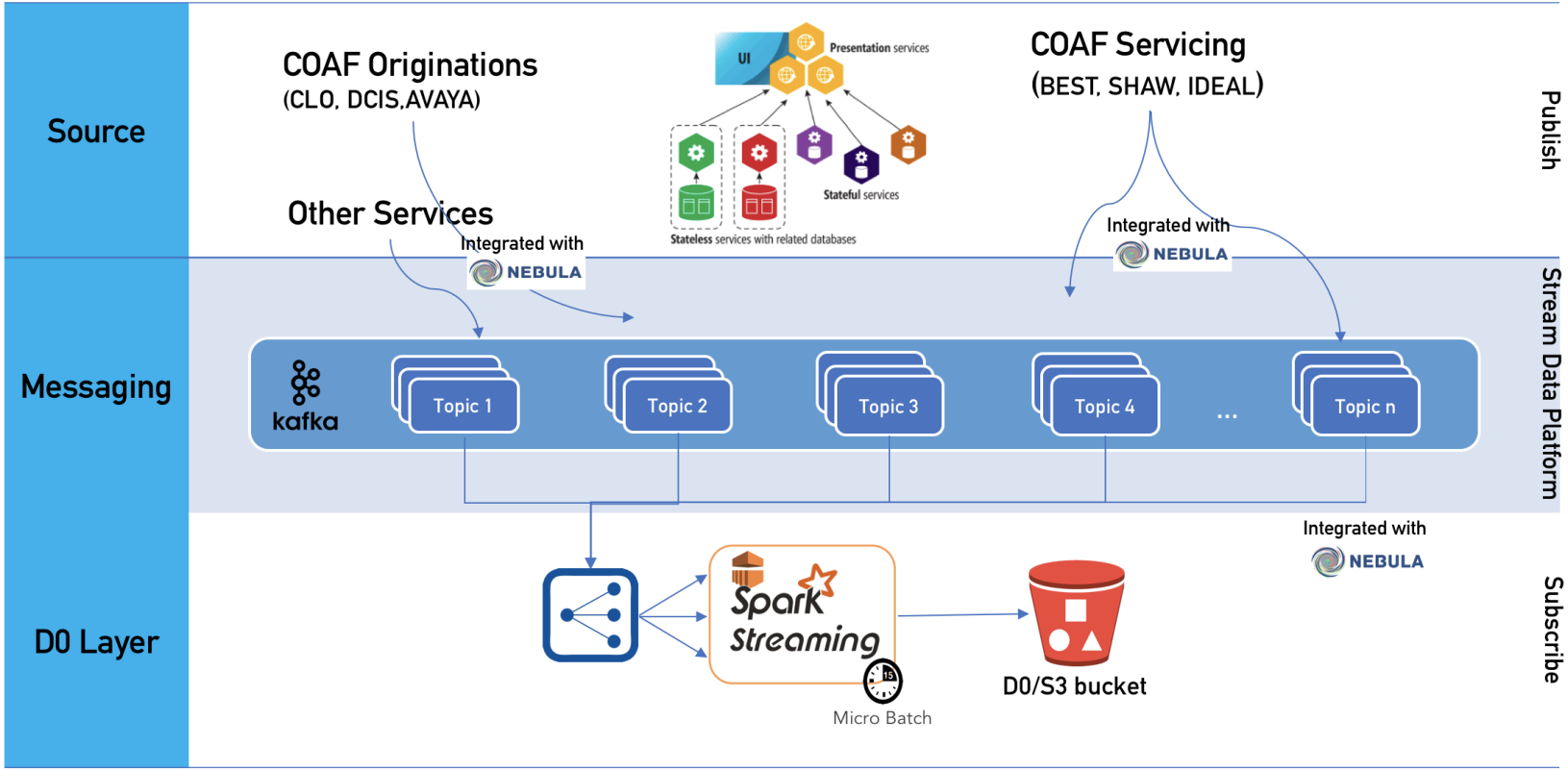
Streaming Data Ingestion

Build an experience intended to consume data from Capital One’s Auto Loan origination and servicing systems(CLO, BEST, SHAW, IDEAL), process and deliver data for Operational and Analytical consumption of Data using Amazon’s AWS cloud.

Technologies used:

Apache Spark Streaming, Kafka, AWS S3, Scala

Streaming Data Ingestion



Code Generator

Code Generator is microservice architecture implementation primarily to automate sql code generation.

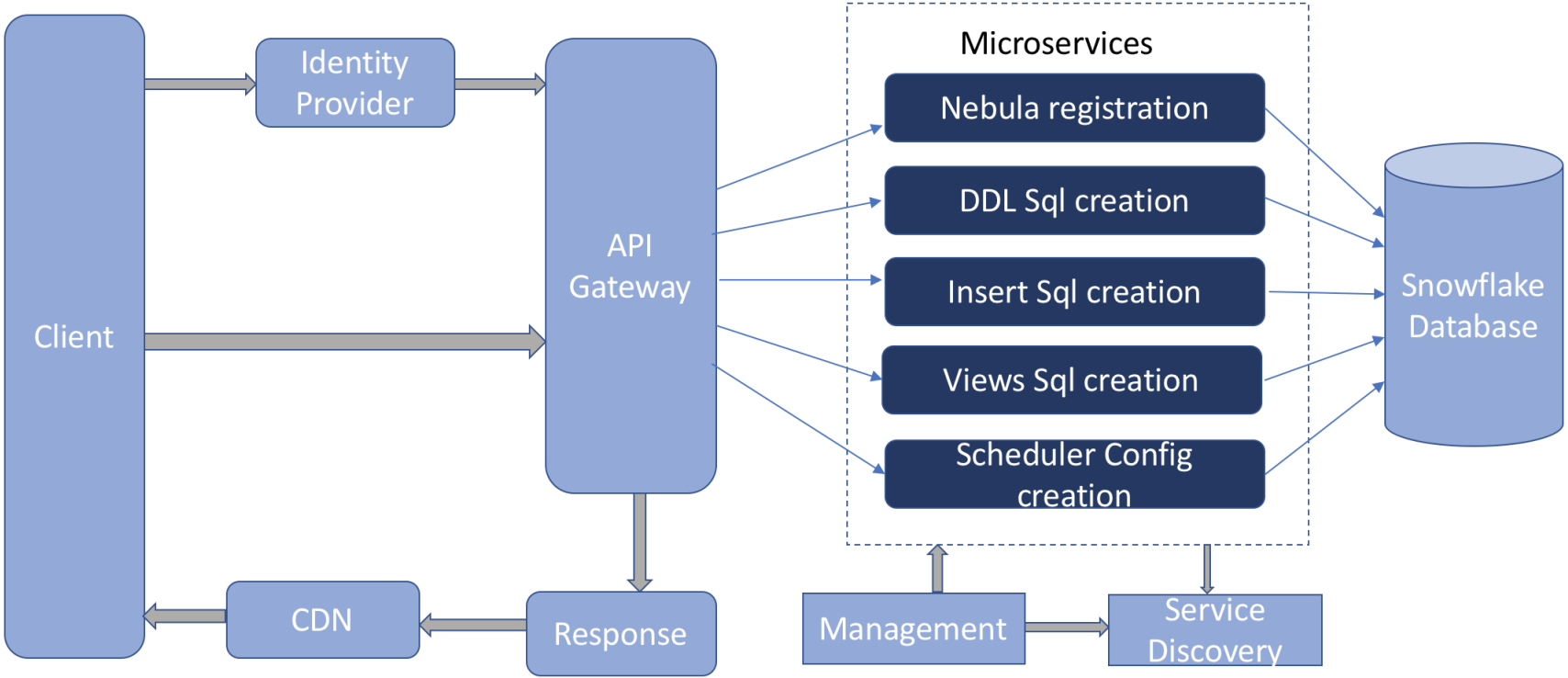
Services provided in Code Generator:

1. Mapping document generator from ILDM document.
2. Metadata registration in Nebula.
3. DDL Sql creation and check in to GitHub
4. Insert Sql Creation and check in to GitHub
5. Views creation based on data classification
6. Config file creation for Scheduler framework

Technologies used:

Java, REST, Spring boot, micro-services, AWS EC2, Github

Code Generator Microservices Architecture



Event based load- Scheduler

“Event based triggers for continuous load” provides a service to consume D1 table load and publish D2A tables accordingly. The experience will build new state of design to consume data sources at all frequencies ranging from Intraday to daily batch.

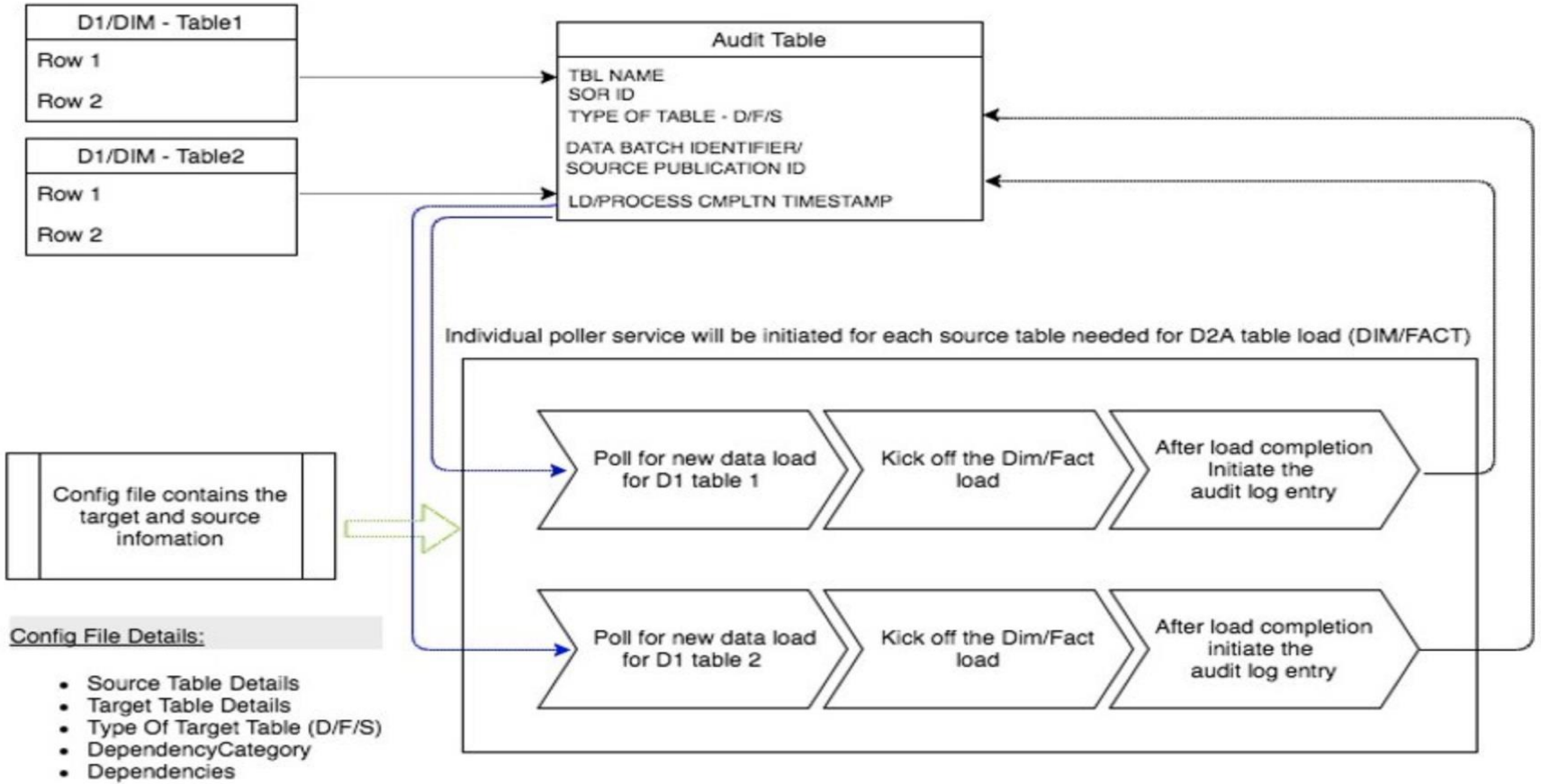
Framework offers full data compliance including –

- Lineage of Data flow, Post consumption using Polling service
- Data ingestion into D2A Snowflake
- Consuming D2A load SQL
- Failure Notification Process
- Restart able logic for failed Polling services
- Logging Snowflake connectivity issue

Technologies used:

Apache Spark, Scala, AWS EMR, AWS S3

Asynchronous/Event Based D2 A - DIM/FACT table load flow



Config File Details:

- Source Table Details
- Target Table Details
- Type Of Target Table (D/F/S)
- DependencyCategory
- Dependencies

Squadron

Squadron is a process controller and monitoring application where each process like Data Quality check, Data Tokenization, L2 load in S3, MPP loads will be tracked at instance identifier level

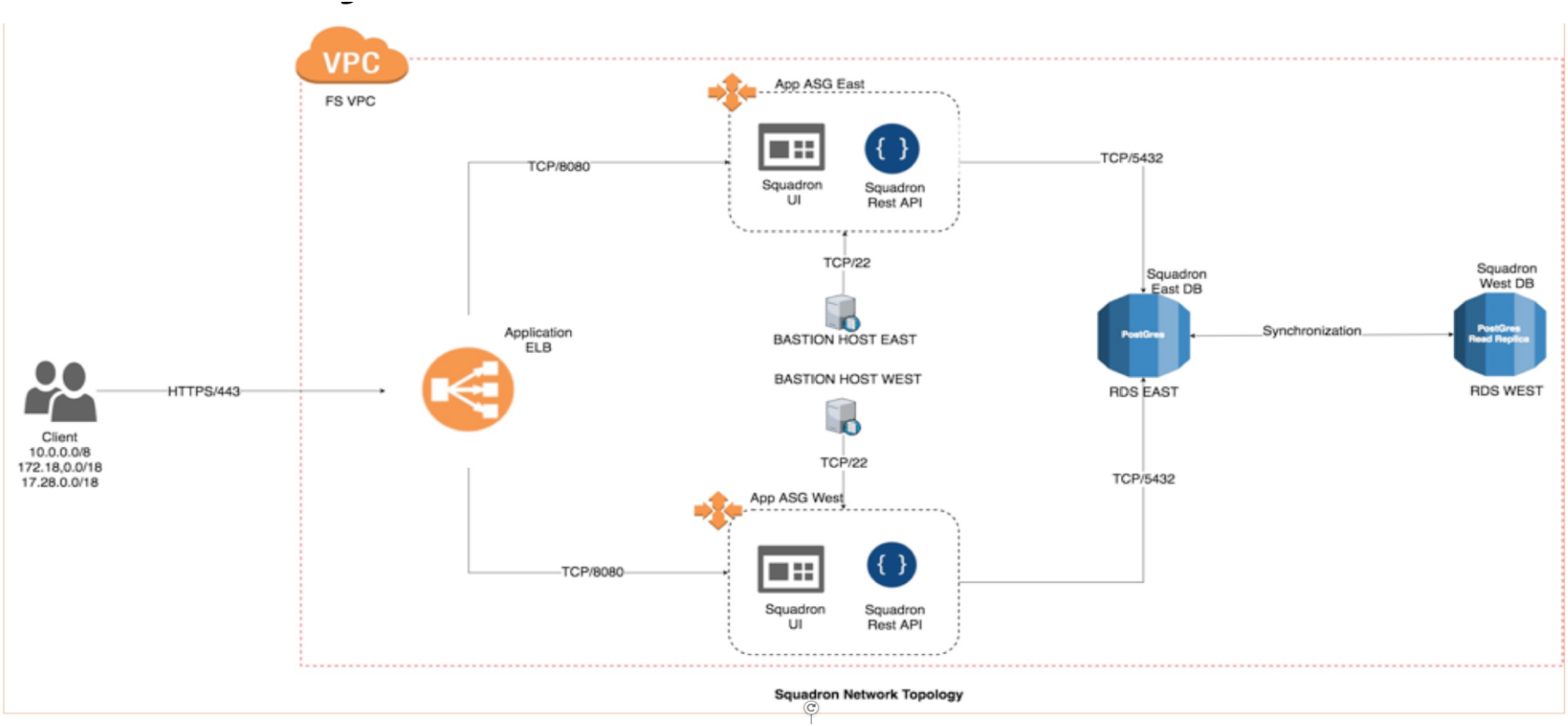
Components of Squadron

- Amazon RDS Postgres – Where entire Execution stats will be saved and used to control L2 executions.
- REST Services – Services to interact with Postgres tables.
- On-Boarding UI/Monitoring Dashboard – UI to onboard new tables, L2, Subject Areas & monitor the applications.
- Archival Job – Archival scheduled daily to remove old data from Postgres tables.
- Mail Service & Notification Job

Technologies used:

Java, Scala, REST APIs, Spring Boot, Amazon Web Services(EMR, S3, SNS, Lambda)

Squadron Architecture



Thank you