Airflow – Cross Environment Triggers
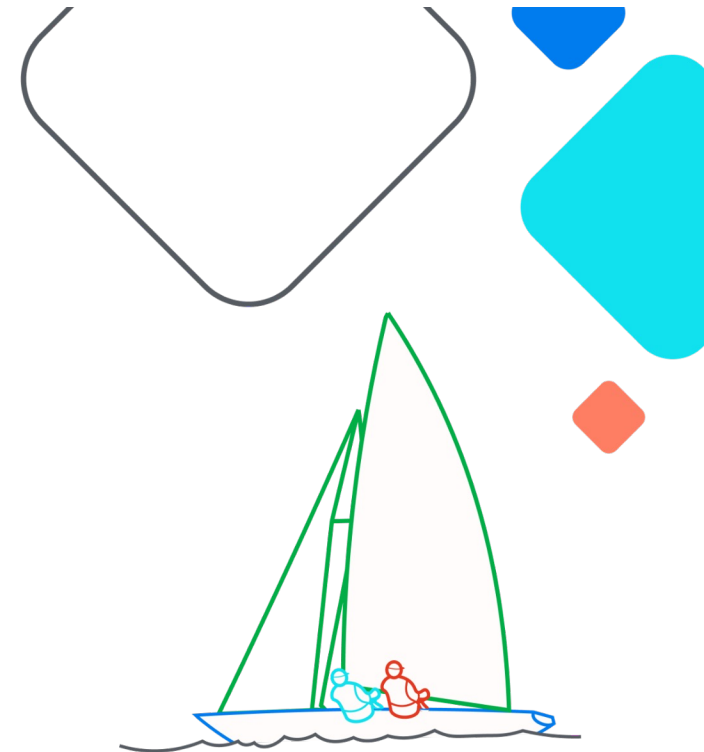
Kunal Jain, Staff Engineer
American Express

Technology

Airflow Summit
Let's flow together

September 19-21, 2023,
Toronto, Canada

About Me

I am a Software Architect working for American Express. My expertise has been on Data Engineering and Governance space

Recently I spent significant amount of my time understanding and solutioning Apache Airflow for our enterprise needs
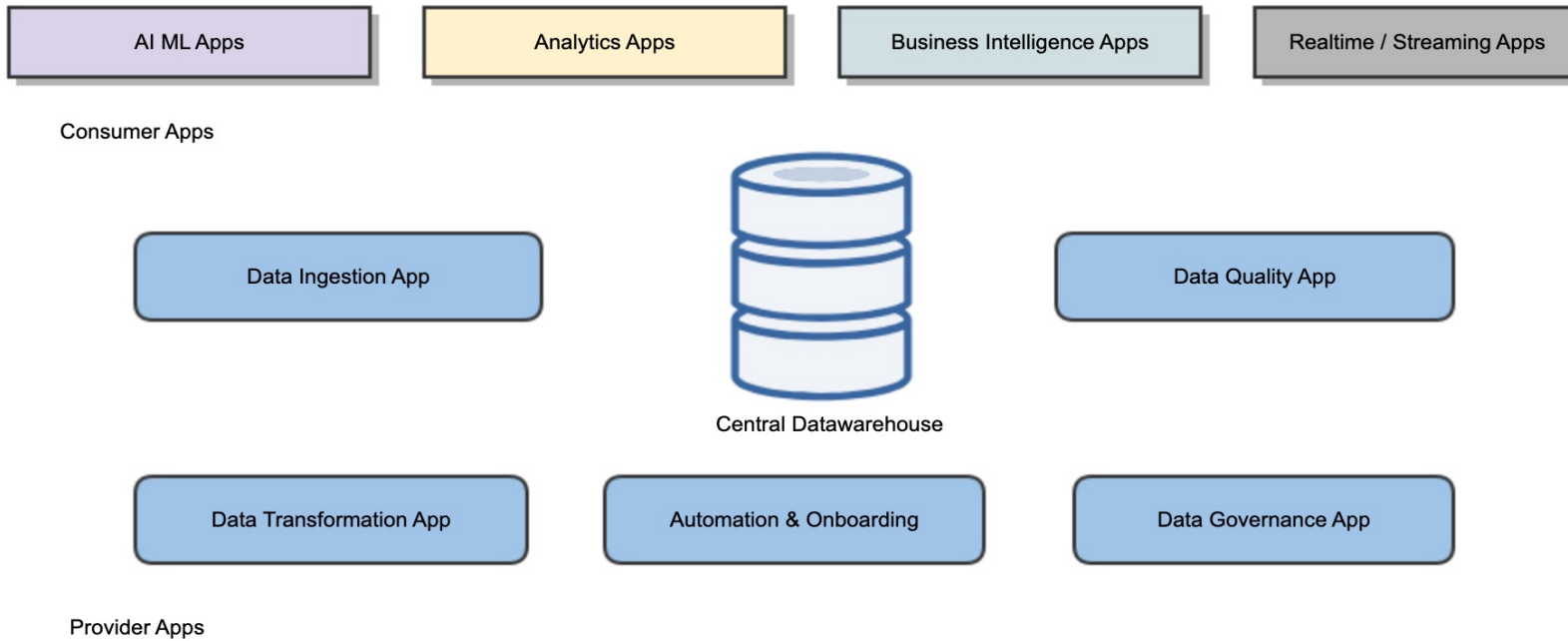
I live in Phoenix , Az with my family

I like to talk about anything data. In my free time, I love to catch up on movies, travelling to mountains, listening to music

AMEX Technology

Table of Contents

- Context
- Problem Statement
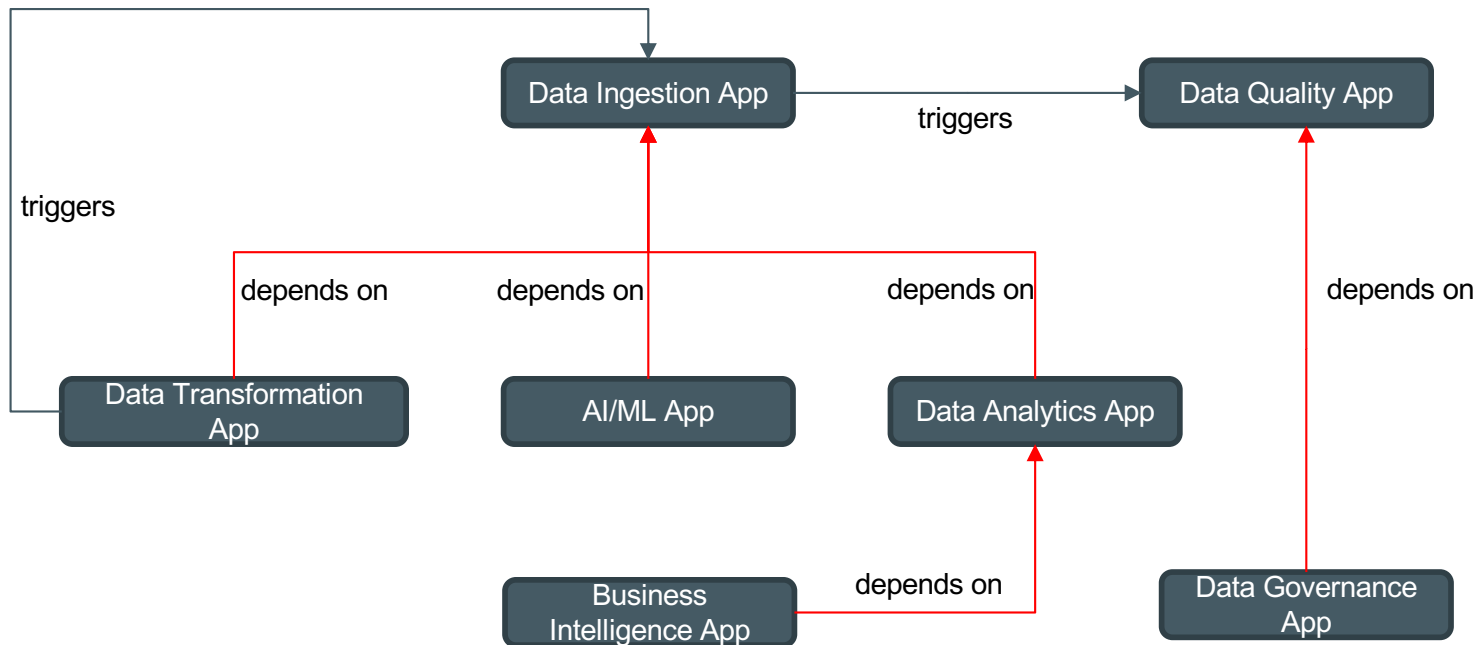- Solution I
- Solution II
- Solution III
- Key Takeaways
- Q & A

AMEX Technology

Context – Multiple Airflow Environments

| AI ML Apps | Analytics Apps | Business Intelligence Apps | Realtime / Streaming Apps |

Consumer Apps

| Data Ingestion App | | Data Quality App |

Central Datawarehouse

| Data Transformation App | Automation & Onboarding | Data Governance App |

Provider Apps

Each app runs in its own project. Projects cannot share resources

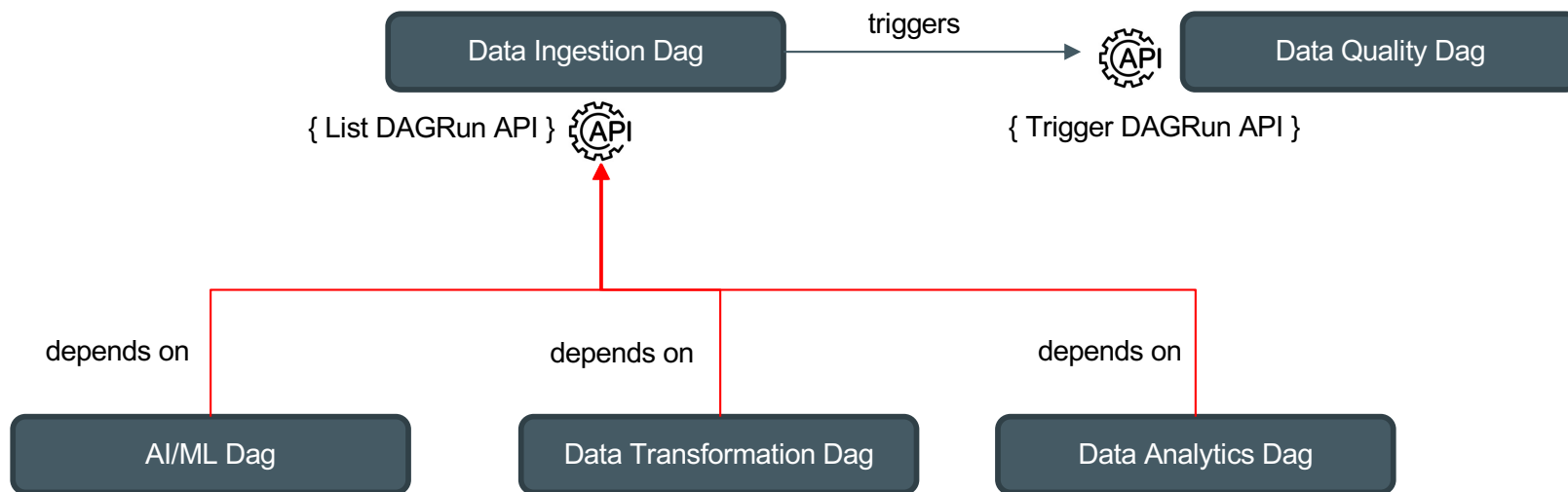Each project has its own Apache Airflow instance

Technology

4

Problem – Cross Environment Triggers & Dependencies

Problem1: A dag should be able to trigger a dag in another project

Problem 2: A dag should be able to build a dependency on the dag in another project

Solution I - Leverage Airflow API's

Data Ingestion Dag → triggers → (API) **Data Quality Dag**

{ List DAGRun API } (API)

{ Trigger DAGRun API }

depends on — **AI/ML Dag**

depends on — **Data Transformation Dag**

depends on — **Data Analytics Dag**

Pros
Easy to implement with minimal code
Good option for 1-2 instances

Cons
Tightly coupled approach
Thousands of sensors hitting Airflow API's

AMEX Technology

6

Solution II – Centralized Event based triggers
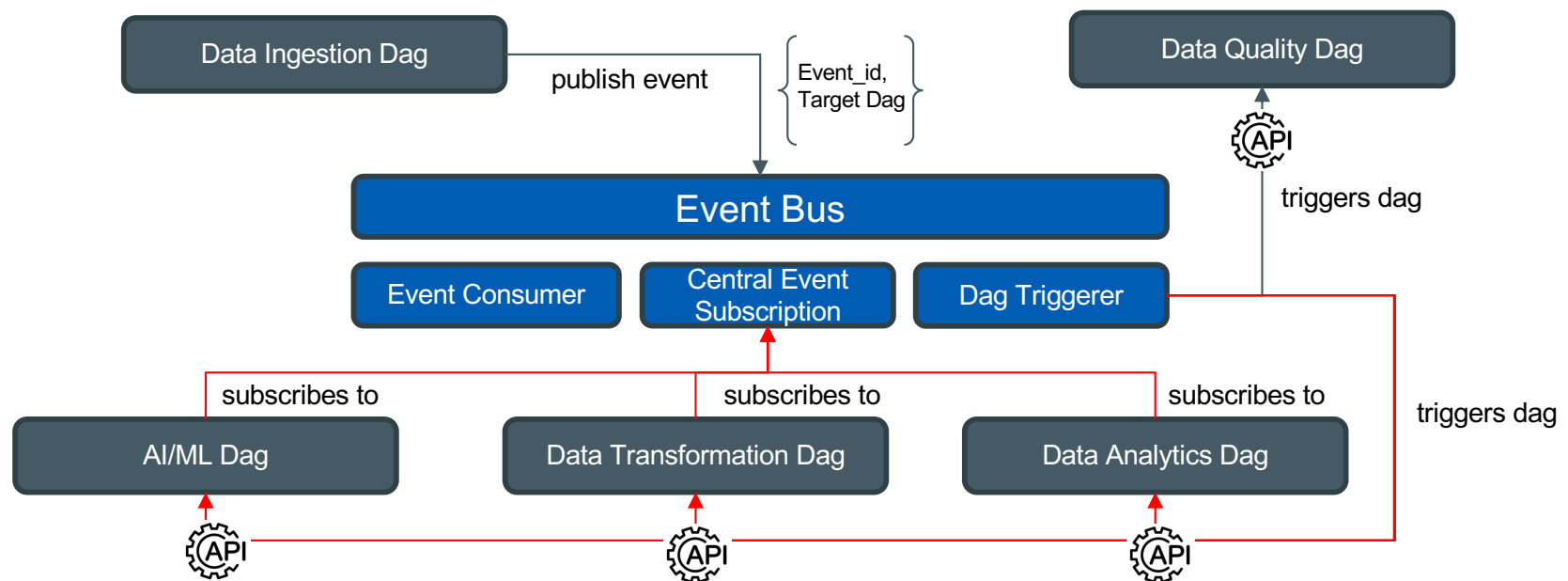
Pros	Cons
Decoupled Approach, each environment can evolve independently	Need central application and additional team to manage the events
Good option for more than 2-3 airflow instances	Higher development effort

AM EX Technology

Solution III – Decentralized Event based triggers
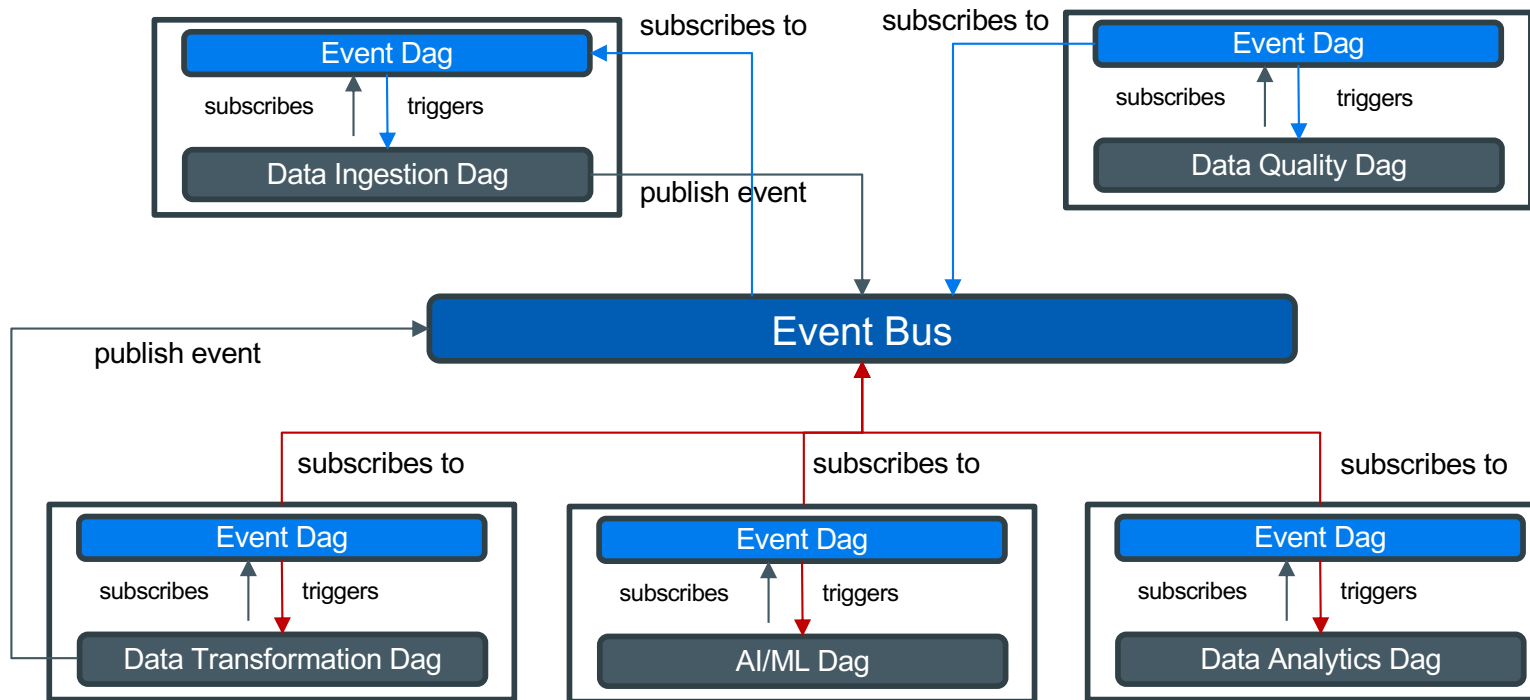
Pros	Cons
Decoupled Approach and decentralized approach	Need good inner-sourcing practice
Good option for more than 2-3 airflow instances	Initial development effort

AMEX Technology

8

Key Takeaways

Yes, we can build event-based triggers and dependencies with airflow

Features	Leverage Airflow API's	Centralized Event based Triggers	Decentralized Event based Triggers
Number of Airflow Environments	1-3	> 1	> 1
Decoupled Approach	No	Yes	Yes
Development Effort	Low	High	Medium
Ongoing Maintenance	Low	High	Low
Cost	Low	High	Medium
Scalability	No	Yes	Yes
Extendable beyond Airflow	No	Yes	Yes
Inner sourcing	Low	Low	High

AM EX Technology

Questions?

We're Hiring!

americanexpress.com/techcareers

Kunal Jain

LET'S CONNECT

kunal.jain@aexp.com

kunal-jain-1ba84313

#TeamAmex   #AmexTech   #ViewsMyOwn

AMEX Technology