

miuul

Veri Mühendisliği Workshop

Airflow ile ETL Süreci



Seda Kayademir | Data Engineer, DataLonga (VBO)

Ajanda

- 1** Veri Mühendisliği Temelleri
- 2** Docker ile Airflow'un kurulması
- 3** Airflow ile PostgreSQL'e veri transferi

1.1 Veri Mühendisliği Nedir?

Veri Mühendisliği verilerin;

- toplanması,
- işlenmesi,
- depolanması,
- ve analiz edilmesi

için gereken altyapının oluşturulduğu bir disiplindir.

Özetle: Verinin A noktasından B noktasına taşınması işlemidir.

Bu işlemleri yapan kişilere de Veri Mühendisi (Data Engineer) denir.

1.2 Veri Mühendisi ne iş yapar?

- Veri altyapısı oluşturma
- Veri entegrasyonu
- Veri temizleme ve düzenleme
- Veri depolama ve yönetimi
- Veri işleme ve analitiği için altyapı oluşturma
- Veri güvenliği ve gizliliği
- Veri raporlama ve görselleştirme
- Veri akışı ve otomasyon

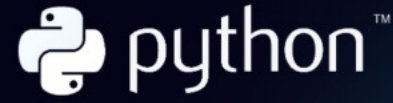
1.3 Veri Mühendisi hangi konularda bilgi sahibi olmalı?

- SQL ve ilişkisel veri tabanı bilgisi.
- OLTP, OLAP, DWH, ETL konularında bilgi sahibi olmak.
- Veri modelleme
- Linux, bash script
- Fonksiyonel Programlama yetenekleri. (Python, Scala)
- JSON ile parse işlemleri

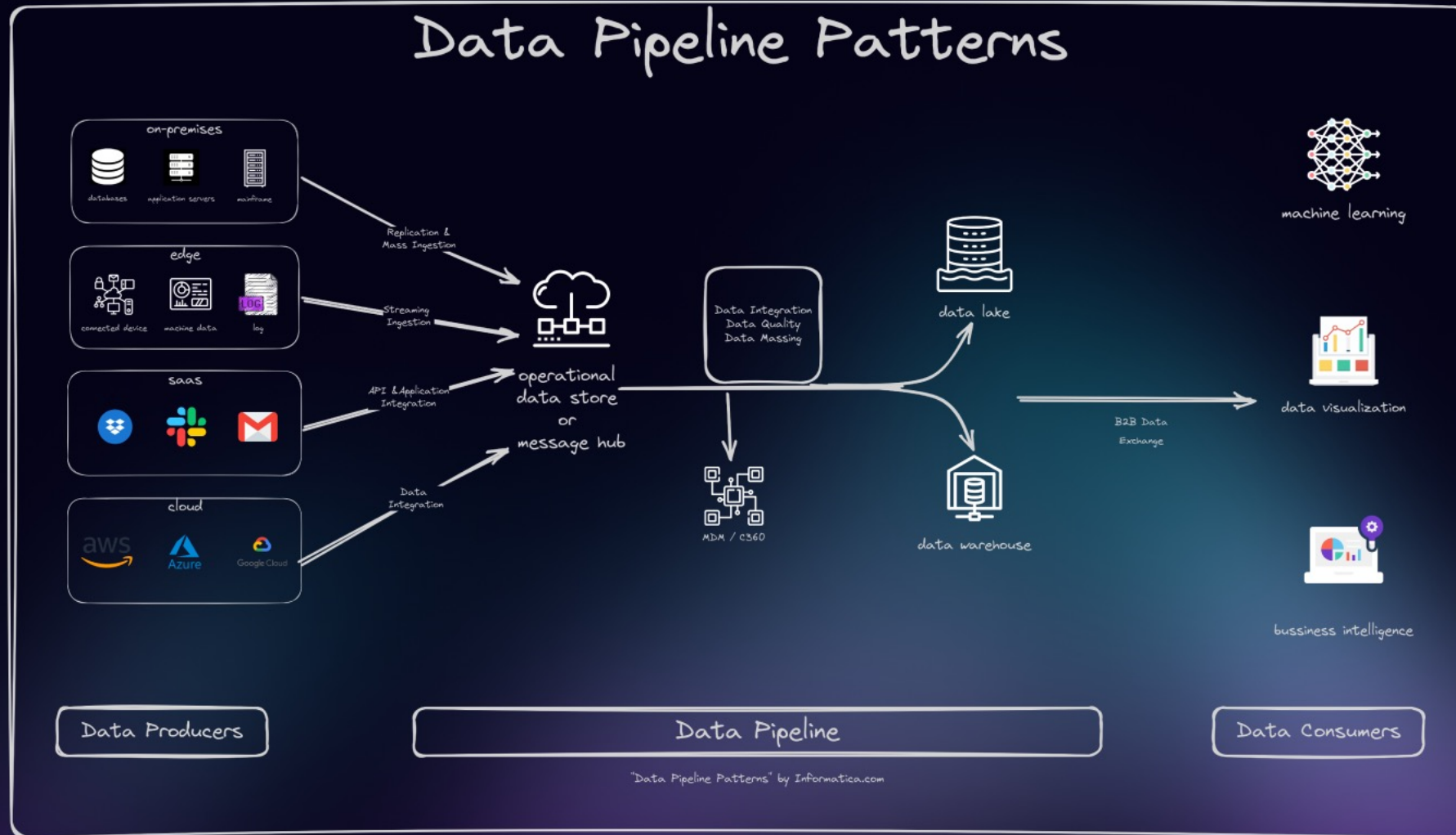
Biraz daha ileri seviyesi için;

- Docker
- Apache Airflow
- Apache Spark
- Apache Kafka
- Bulut Sistemleri (Amazon Web Services, Google Cloud, Microsoft Azure)

ML Engineer hangi toolları kullanabilmeli?



1.5 Örnek bir Veri Akışı

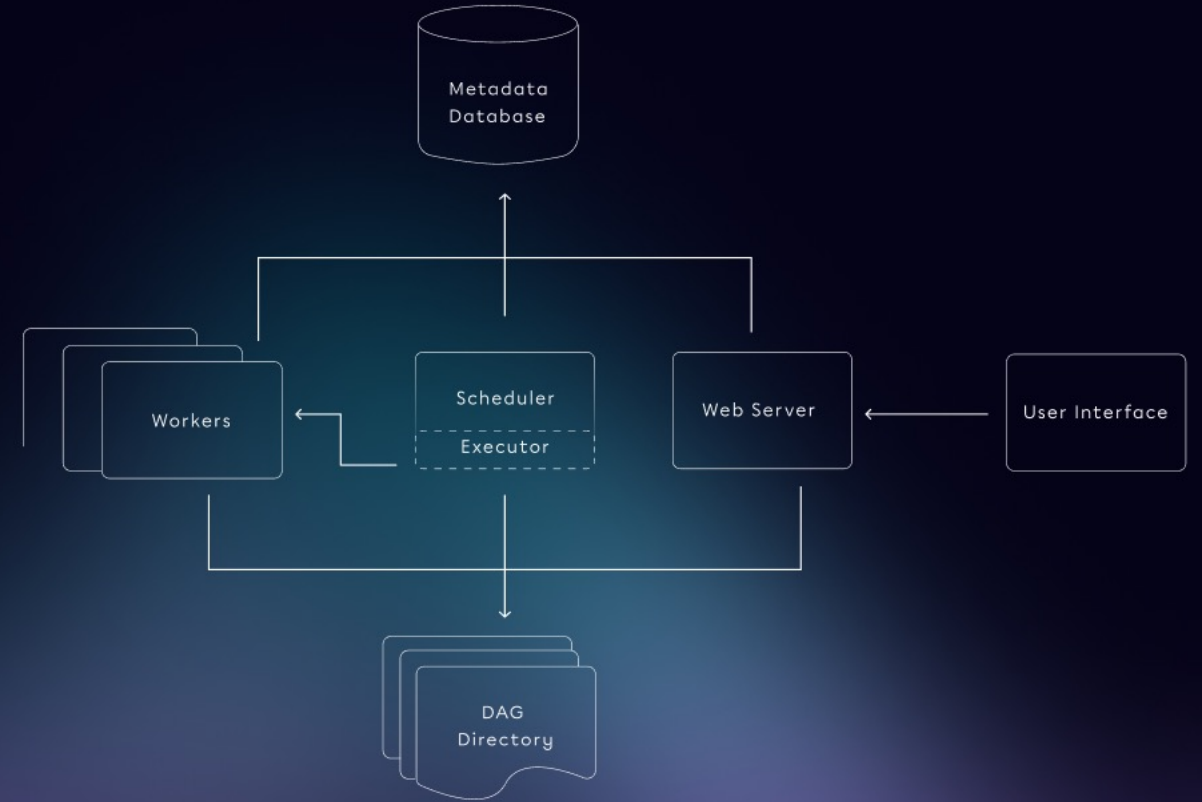


Demo | Airflow ile Veri Transferi

- Airflow - Genel Bakış
- Airflow - Temel Kavramlar
- Demo - Mimari

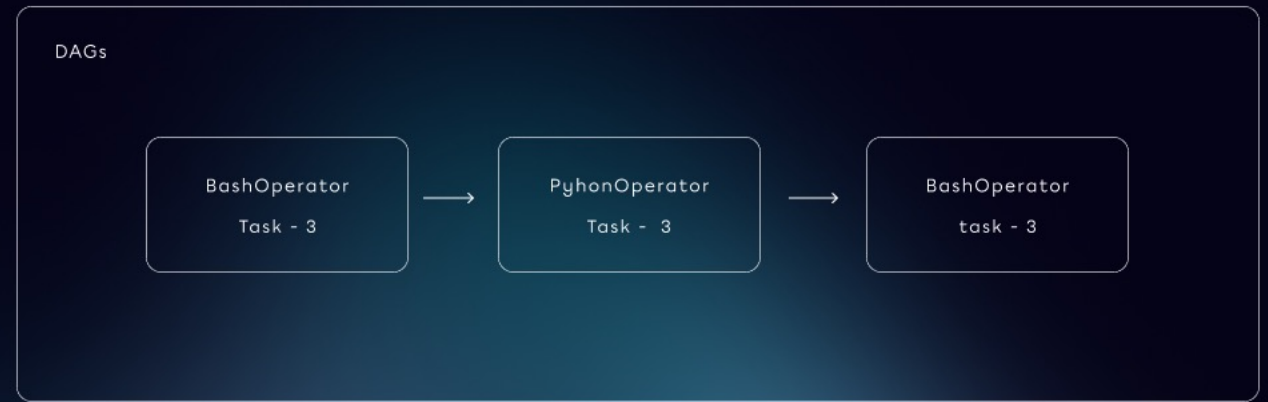
Airflow - Genel Bakış

- **Schedule:** Zamanlama işlemlerini yönetir.
- **Executor:** Görevlerin nasıl çalışacağına karar verir.
- **Metadata Database:** Airflow'un tüm verilerini depolar ve yönetir.
- **Message Broker:** Görevler arasında mesajlar gönderir ve alır.
- **Web UI:** Görevlerin izlenebildiği ve yönetilebildiği bir arayüz sağlar.
- **Log Storage:** Airflow'un tüm log kayıtlarını depolar.
- **Workers:** Görevleri işlemek için kullanılır.
- **Cluster Management:** Airflow'un birden fazla sunucuda dağıtık bir şekilde yönetilmesini sağlar.



Airflow - Temel Kavramlar

- **DAGs:** Görevlerin sırasını ve bağımlılıklarını gösteren yönlendirilmiş döngüsüz bir grafik yapısı.
- **Task:** İş akışında tamamlanması gereken belirli bir görev veya işlem birimi.
- **Operator:** Görevlerin nasıl çalışacağını belirleyen ve gerçekleştiren bileşen.
- **Sensor:** Bir durumu veya koşulu kontrol eden ve görevlerin çalışmasını tetikleyen bileşen.



Demo - Mimari



miuul

MIUUL™ Copyright© Miuul, Inc. All Rights Reserved

www.miuul.com