Data Engineering Project

Module 7
Data Visualization and Business Intelligence Pipeline

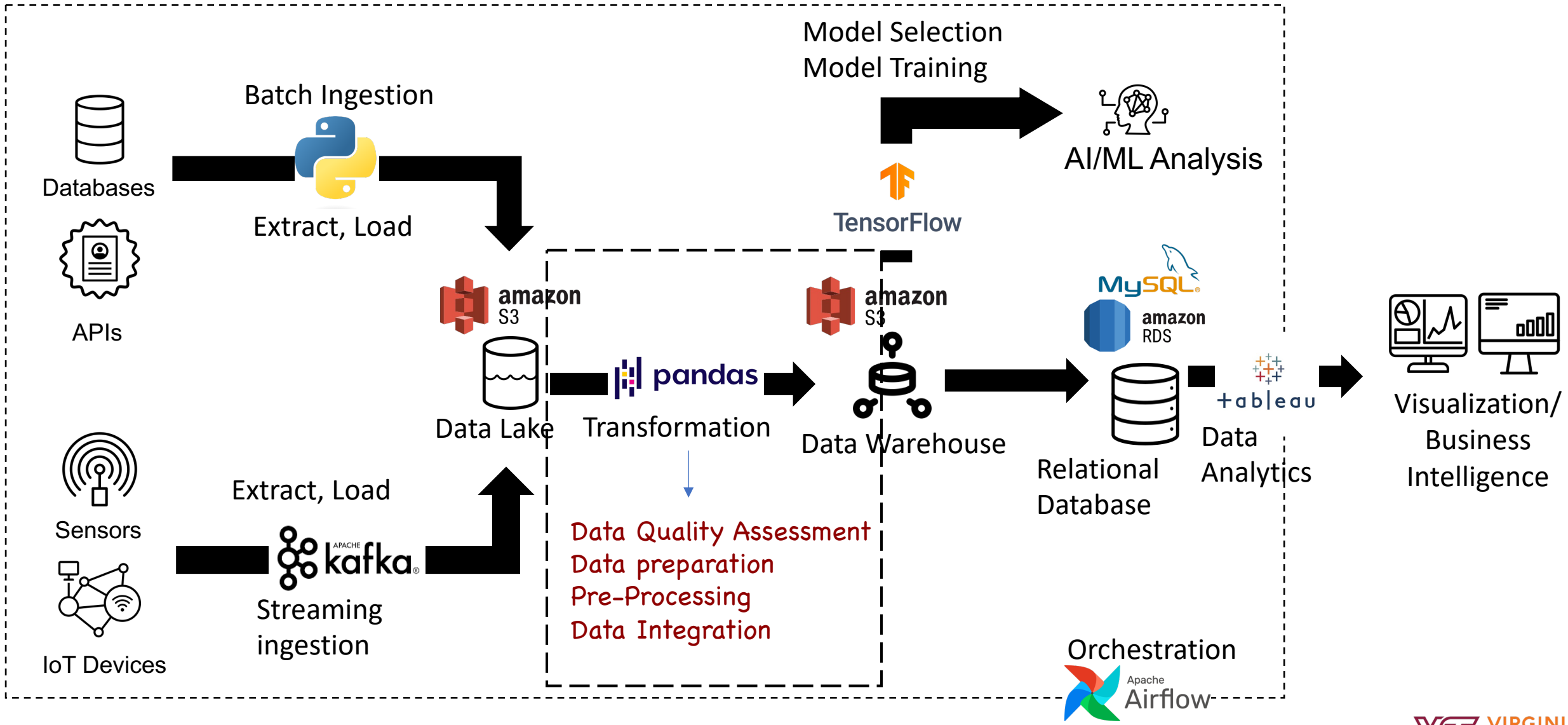
Nektaria Tryfona, PhD

Electrical and Computer Engineering
Virginia Tech

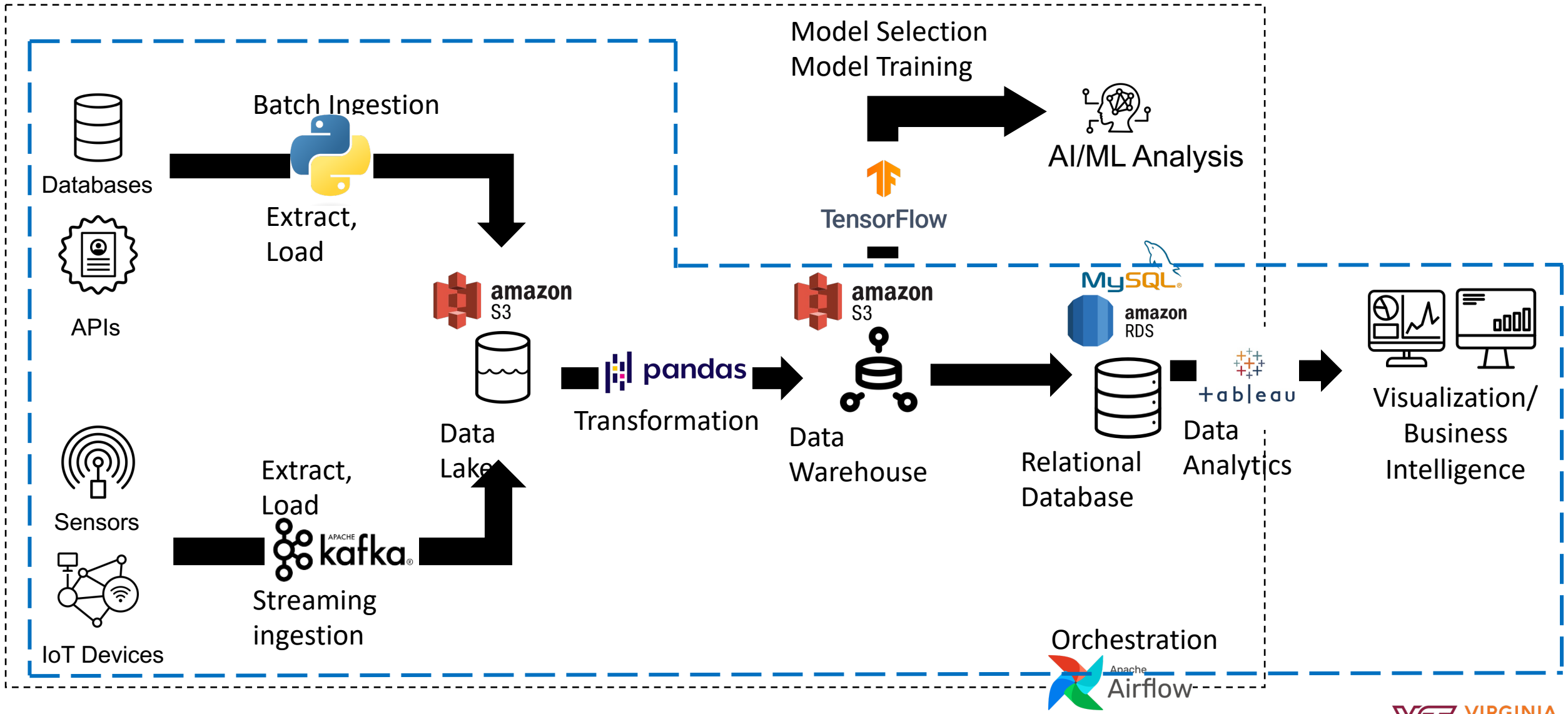
Objectives

- Data Visualization and the Business Intelligence Pipeline
- Cloud Database as part of the Pipeline

Custom Data Engineering Pipeline



Custom Data Engineering Pipeline



Data Visualization/BI Pipeline overview

Step 1
Acquire the needed dataset and perform EDA/data cleaning on it

Step 2
Perform transformations based on the insights we want to showcase and transform the model based on a database schema

Step 3
Push the transformed data from Data Warehouse to the Database

Step 4
Connect the data visualization tool to the database

Step 5
Make data visualizations based on certain features/columns of the data in the database and make certain business intelligence dashboards

Cloud Database as part of the Pipeline

- Cloud database: runs in a public or hybrid cloud environment
  - to help organize, store, and manage data within an organization

- AWS RDS comes with a DataBase (DB) instance
  - instance is an isolated database environment in the AWS Cloud

- Amazon RDS currently supports the following engines:
  - Microsoft SQL Server
  - **MySQL**
  - Oracle
  - PostgreSQL

Note: A DB engine is the specific relational database software that runs on your DB instance

SQL: Structured Query Language

Allows to handle the information within databases using tables

A language to query these tables and other objects related (views, functions, procedures, etc.), e.g.,

- insert, delete, and update data
- create, delete, or alter database objects

Tableau Desktop

- **Tableau Desktop**: used to connect to data, explore data, do analytics, and create reports, dashboards and storyboards
- It is a powerful and fast growing data visualization tool used in the Business Intelligence Industry
- It helps create the data that can be understood by professionals at any level in an organization
- It also allows non-technical users to create customized dashboards

Summary

- Data Visualization and the Business Intelligence Pipeline
- Cloud Database as part of the Pipeline

Data Engineering Project

Module 7
Data Visualization and Business Intelligence Pipeline

Nektaria Tryfona, PhD

Electrical and Computer Engineering
Virginia Tech

Basic mySQL commands

The CREATE DATABASE statement is used to create a new SQL database

Syntax:

CREATE DATABASE *databasename*;

Creating a database does not select it for use; you must do that explicitly

Syntax:

USE *databasename;*

The DROP DATABASE statement is used to drop an existing SQL database

Syntax:

DROP DATABASE *databasename*;

Basic mySQL commands

The CREATE TABLE statement is used to create a new table in a database

Syntax:

```
CREATE TABLE table_name (
    column1 datatype,
    column2 datatype,
    column3 datatype,
   ....
);
```

The column parameters specify the names of the columns of the table.
The datatype parameter specifies the type of data the column can hold (e.g. varchar, integer, date, etc.)

The DROP TABLE statement is used to drop an existing table in a database.

Syntax:

```
DROP TABLE table_name;
```