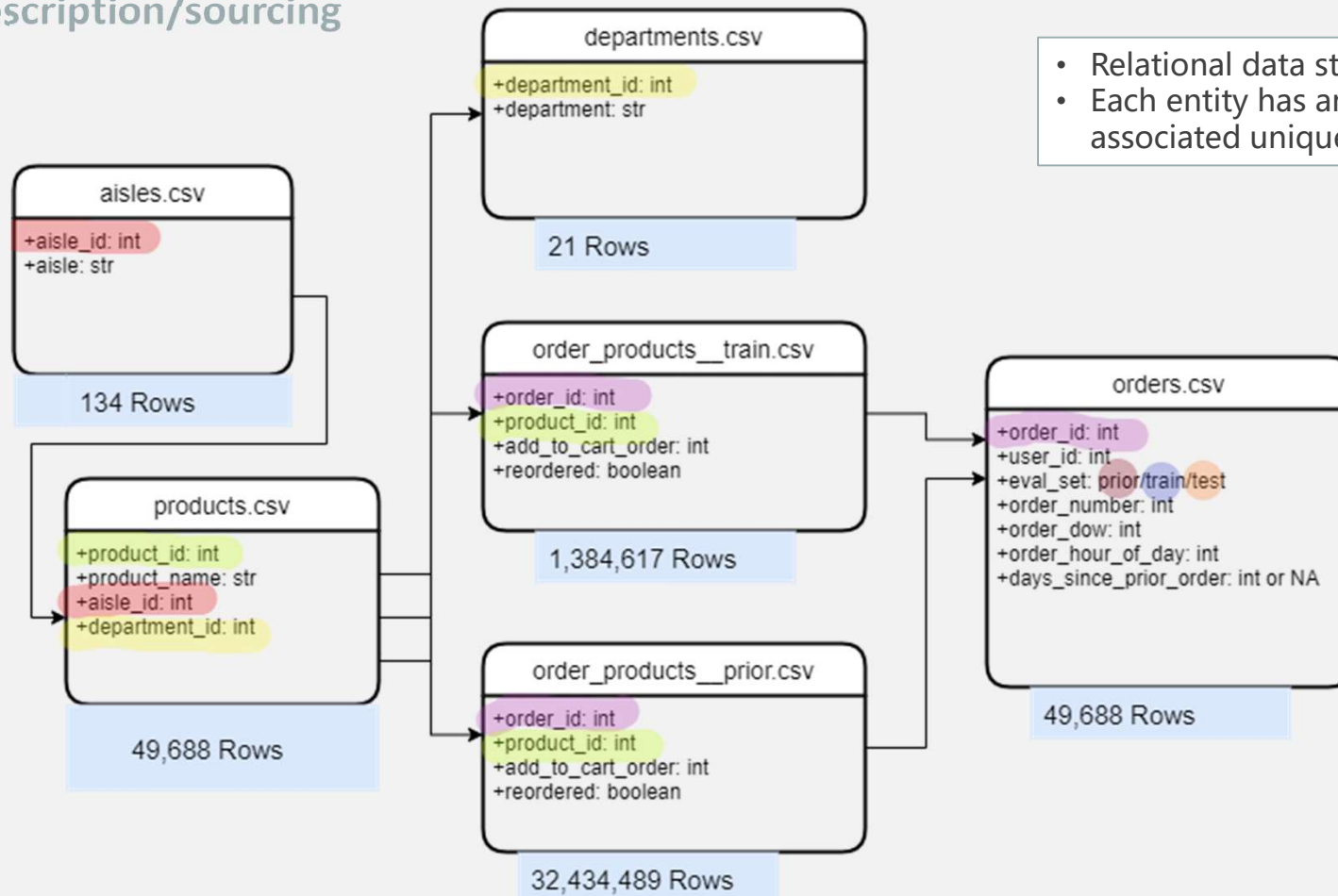
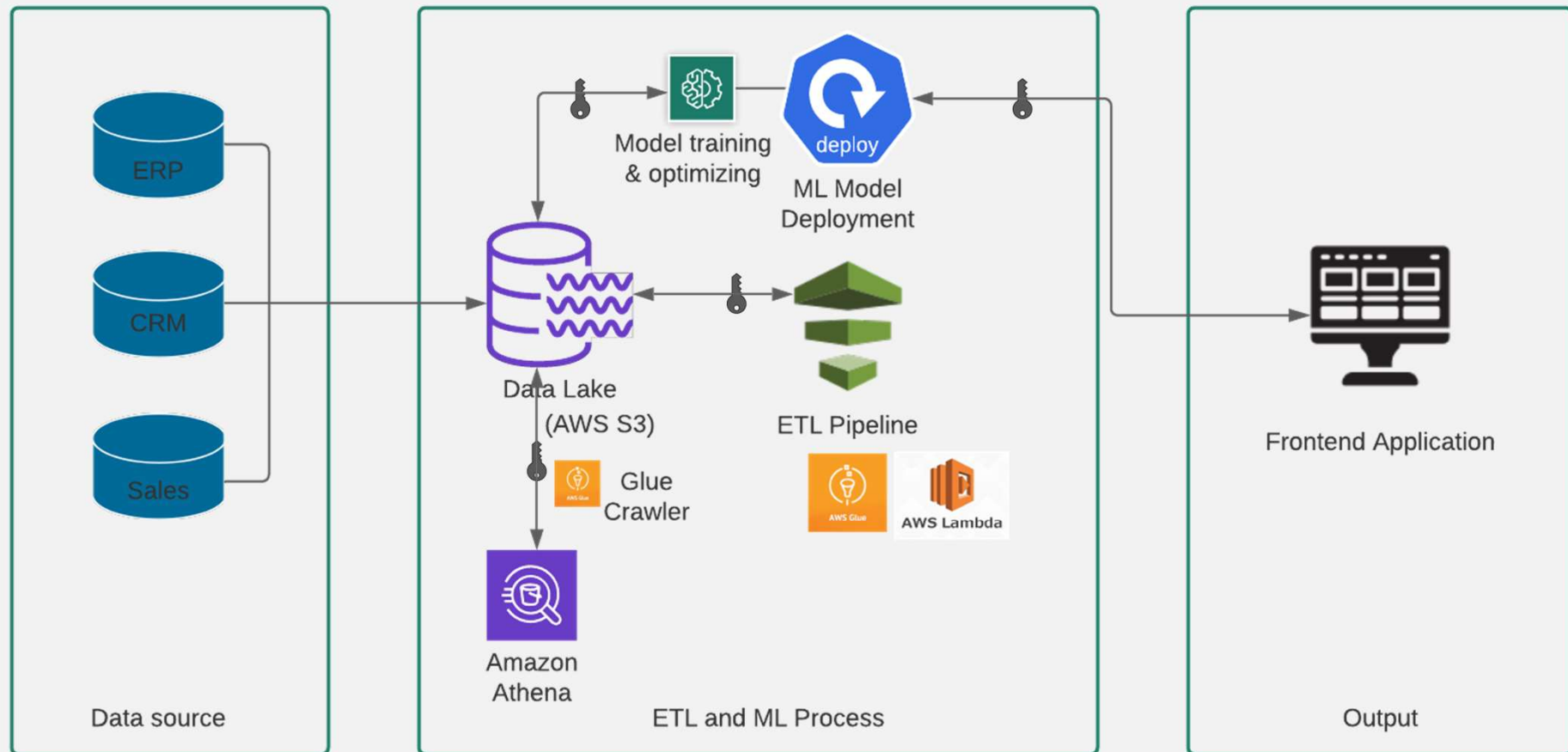


## Data description/sourcing

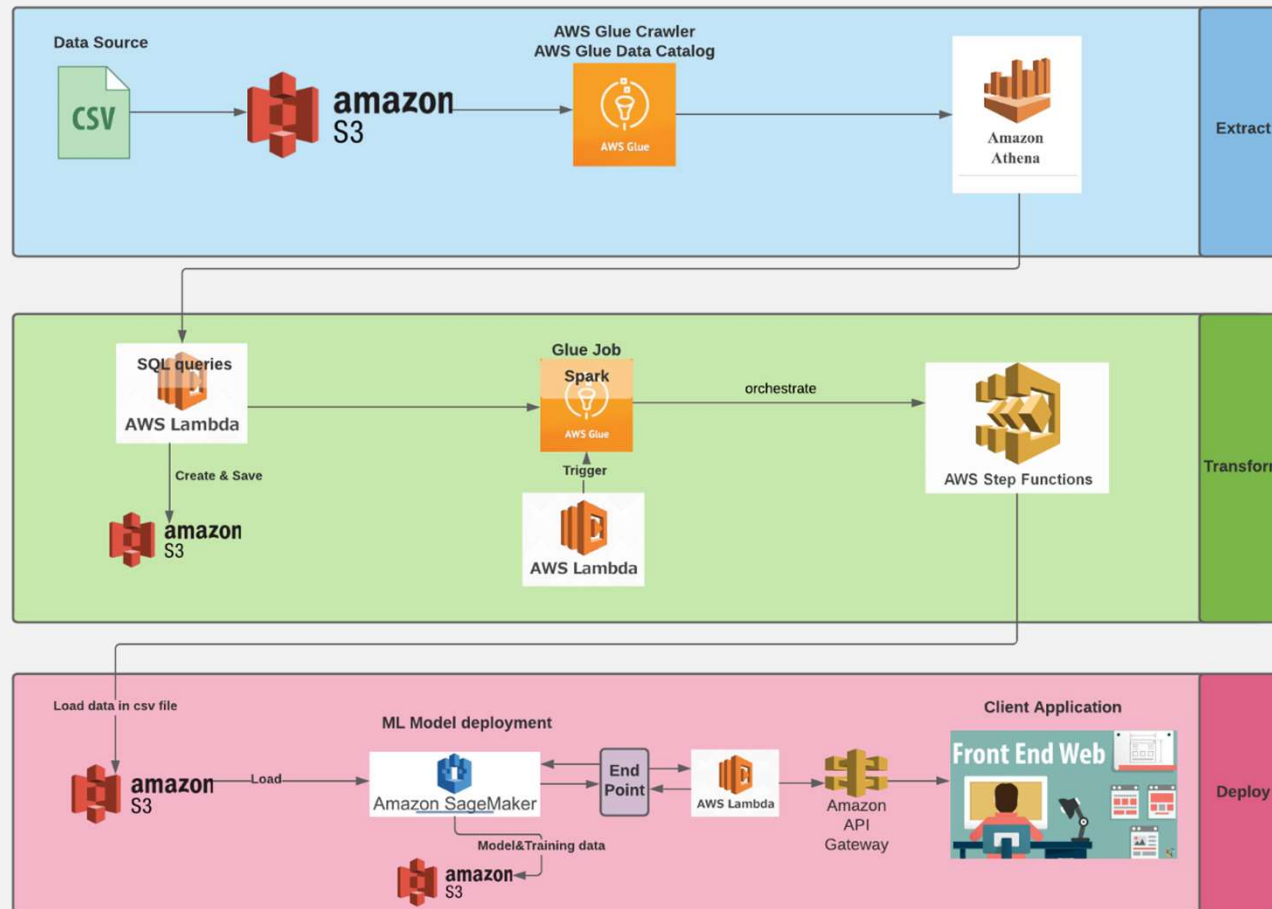


- Relational data structure
- Each entity has an associated unique id

## Data infrastructure



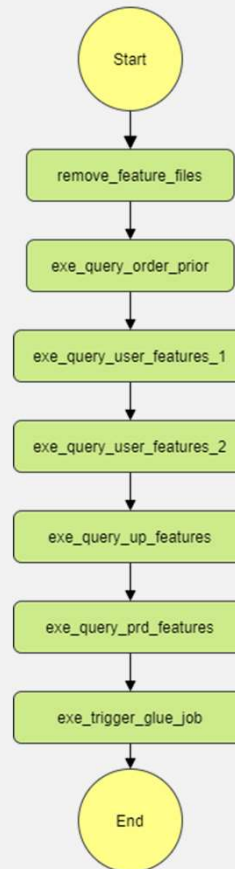
## Data Engineering



## ETL Pipeline



### AWS Step Function



Orchestrate multiple Lambda functions to an ETL pipeline



data.csv

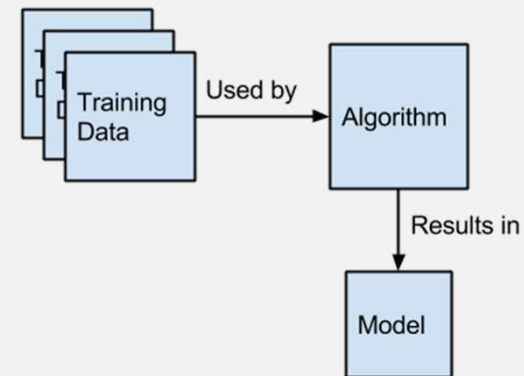
## Data Modelling

- **SageMaker**  
First fully integrated development environment for ML
- **XGBoost Classifier**
  - Parallel processing
  - Handling missing value
- **Feature Engineering**
  - Product reorder probability
  - Product reorder time
  - User average basket

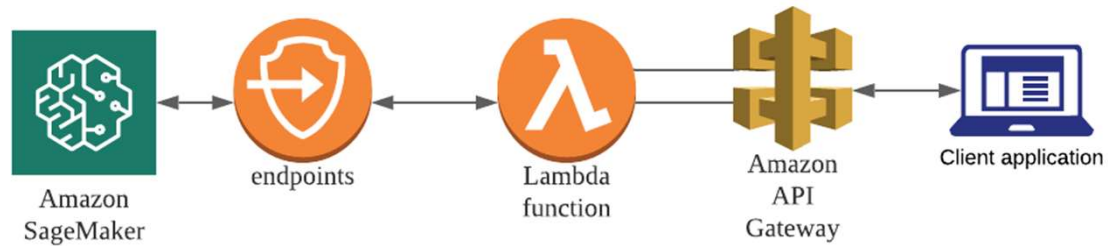


## Data Modelling

- **Build Model**
  - Create tuned XGBoost Model
  - Create hyperparameter tuner
- **Train Model**
  - Fit data into hyperparameter tuner
- **Test Model**
  - Create a transformer to do the inference
  - Output will be probability



## Model Deployment



Will this customer buy the product?

Enter the customer feature to find out...

**Customer feature:**

1,21.818181818181802,240,13,0.45454545454545503,18,5.0,1,3,1,54409,0.4872174312394754,2.7763943460733786,0.6398206179124777,6.0,0.3333333333333333,2,0.3333333333333333

Submit

Customer is likely to buy this product!