

Lakehouse for Data Science @ DoorDash

Hien Luu

Head of ML Platform, DoorDash

Agenda

- Use Cases – Analytics & Data Science
- Why Databricks?
- How is it working?
- Summary

About Me

Sr. Engineering Manager - Machine Learning Platform



Hien Luu

About Me

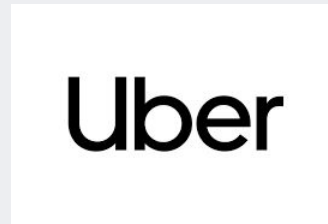
Career Journey



Cloud Infrastructure



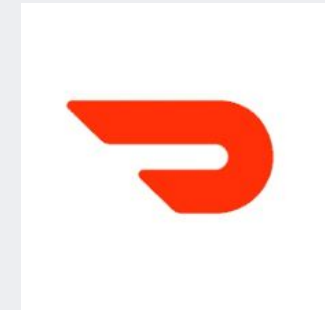
Big Data



Big Data



ML Infra

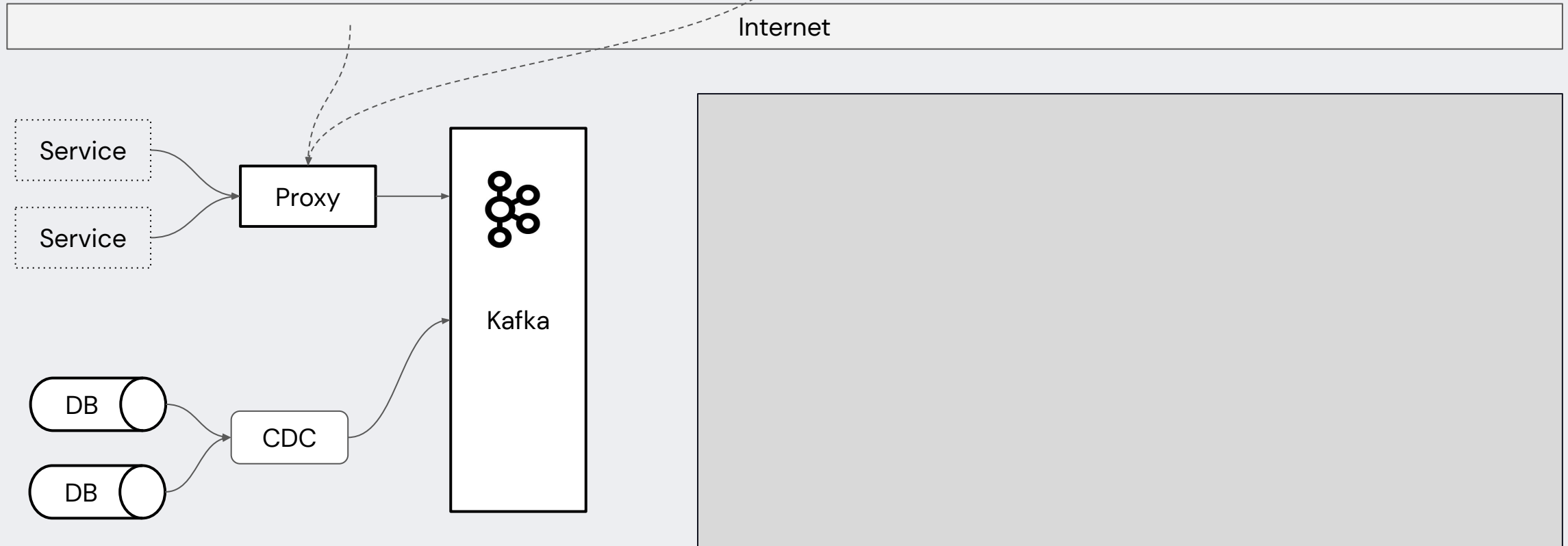
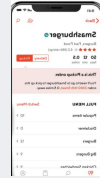


ML Infra

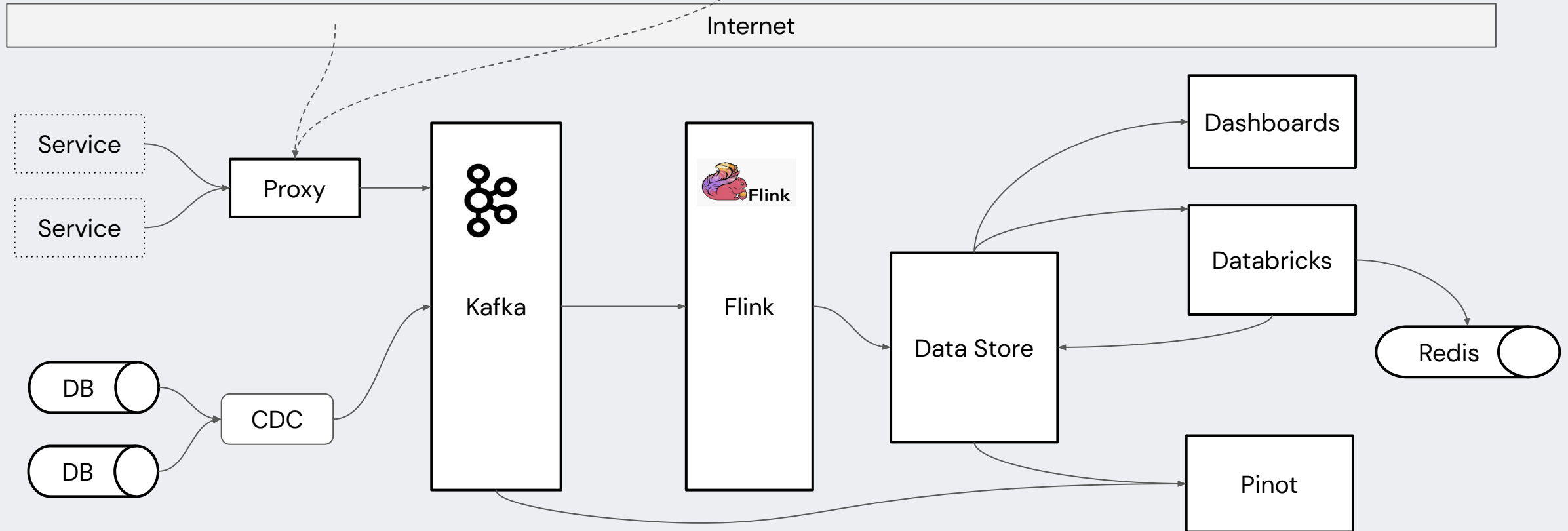
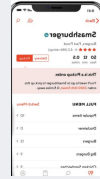
Use Cases – Analytics & Data Science

- Business analytics (metrics, reporting, dashboard)
- Data movement & curation:
 - Ingestion, ETL, data engineering
- Operational metrics
 - System monitoring & alerts
- Data Science & Machine Learning
 - Feature computation & generation
 - Model exploration and development

Use Cases - Analytics & Data Science



Use Cases - Analytics & Data Science



Why Databricks?

- Snowflake Data warehouse
- Evaluation factors
 - Cost, usability, timing, risks, scalability
- Quickly up and running
- Meeting business growth

Why Databricks?

Lakehouse at DoorDash

- Improved efficiency and reduced costs
- Data transparency in S3
- Data discovery & lineage
- Streaming and batch data ingestion

How is it working?

- More data & more users
- Allow us to focus on meeting the data related needs
 - Diverse set of internal teams
- Scaling computation & management

How is it working?

Databricks Impact

- Productivity
 - Data engineers, business analysts, data scientists
- Scaling
 - Feature computation and ML model development
 - Job scheduling & programmatic cluster provisioning
- Focus on core competencies and less on operation

Next Step & Future Plan

- Moving to latest version to reduce operational management
- Continuous performance improvement
- Cost attribution
- MLflow for ML development lifecycle
 - NLP & computer vision use cases

DATA+AI
SUMMIT 2022

Thank you