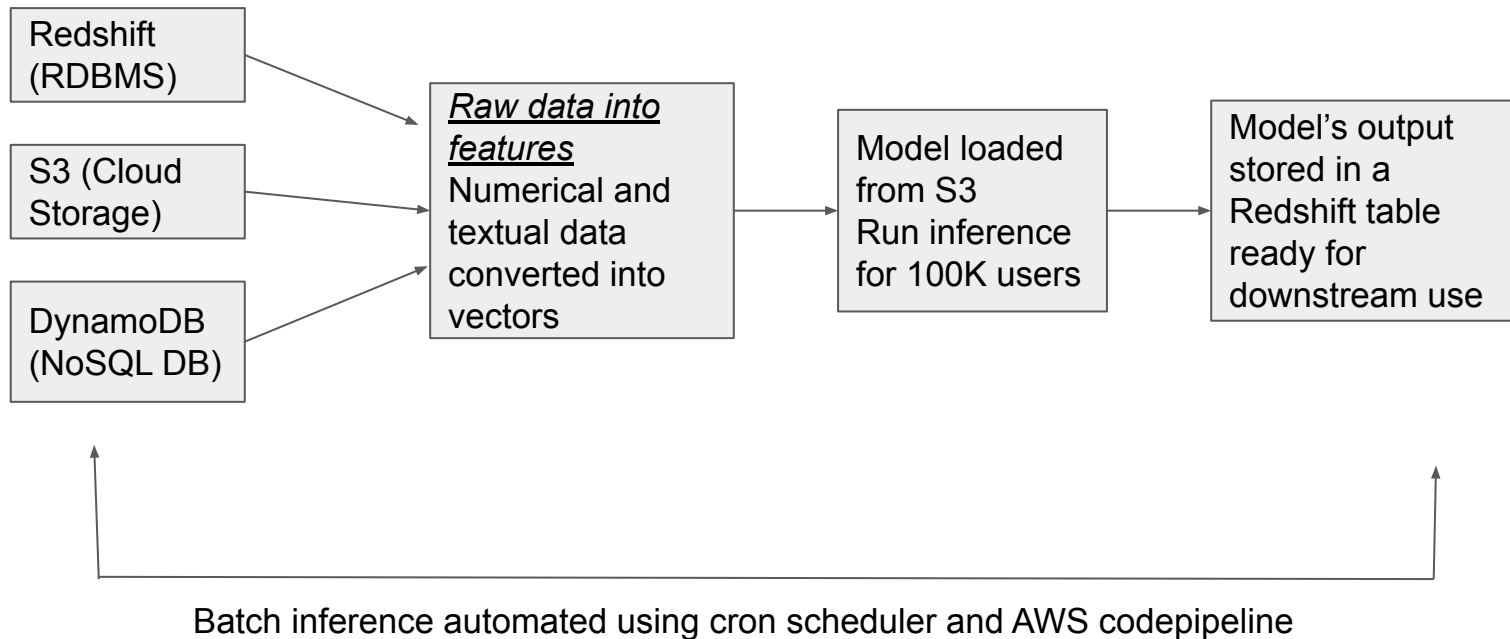# Hi! I'm Megha Agarwal

- Senior Data Scientist at CRED (a fast-growing fintech startup in India)
  - Worked for 2 years; using data science to solve problems in the world of finance
  - Generating actionable insights from data, building predictive models that drive company metrics upwards were key aspects of my role
- Software Engineer in Machine Learning at Samsung R&D
  - Worked for another 2 years; developed deep neural networks for computer vision problems
  - Object Detection, Instance Segmentation, Semantic Segmentation, etc
- Graduated from IIT Guwahati with a major in Electronics & Electrical Engineering and minor in Computer Science & Engineering

Credit Risk Model Inferencing

- A decision-tree based model that predicts the probability of a user defaulting on a loan (not paying back on time)
- Input features come from various sources of data
- Credit risk scores are temporal in nature; a user's credit riskiness can change with time
- Task: 100,000 users to be scored daily; Automate batch inference of ML model

Redshift
(RDBMS)

S3 (Cloud
Storage)

DynamoDB
(NoSQL DB)

_Raw data into
features_
Numerical and
textual data
converted into
vectors

Model loaded
from S3
Run inference
for 100K users

Model's output
stored in a
Redshift table
ready for
downstream use

Batch inference automated using cron scheduler and AWS codepipeline

Multi class text classification

Dataset
Unstructured text
Each line with 5-15 words

Preprocessing
Removal of stop words
Normalization techniques (trees -> tree, 567 -> 000 -> 0)

Vector representation of text
Bag of words
Word2Vec
TF-IDF NMF
fastText embeddings

Model Training and Evaluation
Logistic Regression, SVM etc
Accuracy metrics

Model Deployment
Real -time inference
RESTful API using Flask
Low latency and high throughput constraints
20+ RPS with response time under 100 ms