NORTHROP GRUMMAN
UNIVERSITY RESEARCH SYMPOSIUM

Research in Applications for Learning Machines (REALM) Consortium

Situational Knowledge On Demand (SKOD)

23rd October 2019

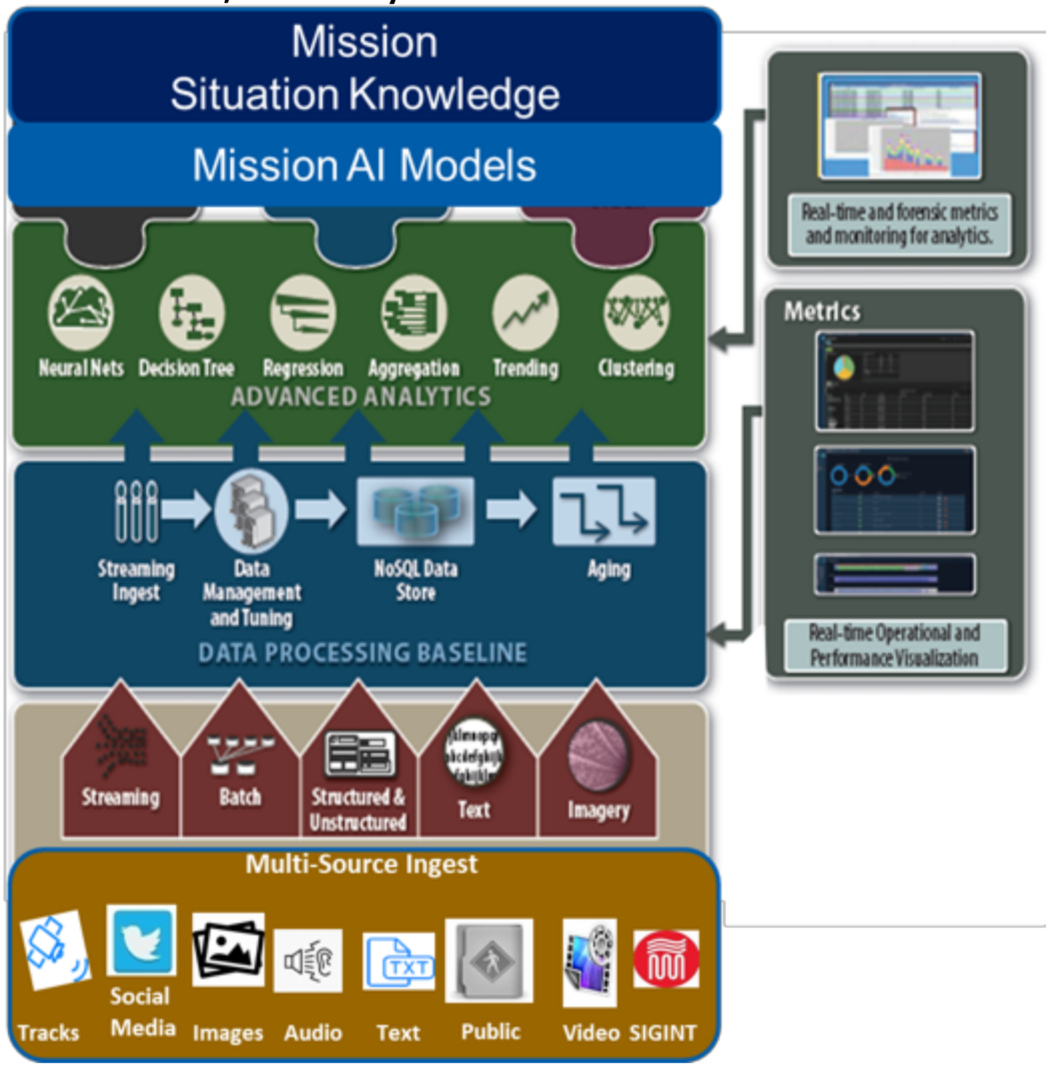
Bharat Bhargava
Purdue University

Technical Champion: Dr. James MacDonald

Collaborations

- **Primary Researchers**
  - Bharat Bhargava (Purdue)
  - Michael Stonebraker (MIT)
  - Michael Cafarella (UMich)
  - Aarti Singh (CMU)
  - Peter Bailis (Stanford)
- **Students**
  - KMA Solaiman
  - Servio Palacios
  - Alina Nesen
  - Pelin Angin
  - Zachary Collins (MIT)
  - Aaron Sipser (MIT)
  - Miguel Villarreal-Vasquez
  - Ganapathy Mani
  - Aala Oqab Alsalem
  - Tunazzina Islam
  - Denis Ulybyshev
  - Daniel Kang (Stanford)

The project is applicable across a variety of industries, military to commercial to academic.



Principal Investigators:

- *Bharat Bhargava, Purdue University Research*
  - *Extract and identify patterns related to significant mission needs*
  - *Develop algorithms to establish situational awareness*
  - *Connect disaggregate knowledge sources*
- *Michael Stonebraker, Massachusetts Institute of Technology Research*
  - *Information Value*
  - *Push relevant information efficiently to interested parties (e.g. analysts, experts, and decision makers)*
- *Aarti Singh, Carnegie Mellon University Research*
  - *Context Aware Machine Learning*
  - *Metadata Tagging*
- *Peter Bailis, Stanford University* Research
  - *Extract Knowledge Patterns from Streams*
  - *Real-time Content Reduction &*
  - *Object Association*

Integration with Paradigm

Multiple Data Sources
SKOD
Novel Sources

Integration with Paradigm

Multiple Data Sources
SKOD
Novel Sources

→

Ingestion & Preprocessing
SKOD
Data Processing Pipeline

Integration with Paradigm

Multiple Data Sources
SKOD
Novel Sources

Ingestion & Preprocessing
SKOD
Data Processing Pipeline

Analytic Post-Processing
SKOD
Relevant Tweet Extraction
Object Detection
Video Feature Extraction
Title & Entity Extraction
Subj, Verb, Obj Extraction
Knowledge Graph
Indexing

Integration with Paradigm

Multiple Data Sources
SKOD
Novel Sources

Ingestion & Preprocessing
SKOD
Data Processing Pipeline

Analytic Post-Processing
SKOD
Relevant Tweet Extraction
Object Detection
Video Feature Extraction
Title & Entity Extraction
Subj, Verb, Obj Extraction
Knowledge Graph
Indexing

Alerting
SKOD
User Modeling
Data Profiling

Integration with Paradigm

Multiple Data Sources
SKOD
Novel Sources

Ingestion & Preprocessing
SKOD
Data Processing Pipeline

Analytic Post-Processing
SKOD
Relevant Tweet Extraction
Object Detection
Video Feature Extraction
Title & Entity Extraction
Subj, Verb, Obj Extraction
Knowledge Graph
Indexing

Alerting
SKOD
User Modeling
Data Profiling

Alerts

Outline

- Possible Scenarios
- Objectives
- Problem Statement
- Datasets
- SKOD Architecture
- Summary
- Deliverables and Demo
- Future Plans

Outline

- Possible Scenarios
- Objectives
- Problem Statement
- Datasets
- **SKOD** Architecture
- Summary
- Deliverables and Demo
- Future Plans

Architecture Modules

- Data Streaming
- **Feature Extraction**
- **Knowledge Graph**
- **User Profiling**
- PostgreSQL Database
- Graph-based Indexing Layer
- Front End

Achievements

Relevant Publications:

1. S. Palacios and K. Solaiman, P. Angin, A. Nesen, B. Bhargava, Z. Collins, A. Sipser, M. Stonebraker, J. Macdonald. ***SKOD: A Framework for Situational Knowledge on Demand.*** In *Polystores and other Systems for Heterogeneous Data (**Poly 2019**),* at **VLDB 2019**, LA, California, August 30, 2019.

2. K. Solaiman, B. Bhargava, J. MacDonald. ***Multi-modal Information Retrieval via Joint Embedding.*** (To be submitted)

3. A. Nesen, B. Bhargava, J. MacDonald. ***Explainable Anomaly Detection in Surveillance Video With Deep Learning and Knowledge Graphs.*** (To be submitted)

4. M. Kabir and S. Madria. ***A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management.*** In *27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, Illinois, Nov 7, 2019.

5. D. Kang, P. Bailis, and M. Zaharia. ***Blazeit: Fast exploratory video queries using neural networks.*** (2018).

6. Peter Bailis, et al. ***Infrastructure for Usable Machine Learning: The Stanford DAWN Project.*** (2017).

Achievements

Third Party Funding:

- DARPA award on *Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON)* initiative of DoD
  - Generating Novelty in Open-world Multi-agent Environments (GNOME)
- Several white papers have been submitted for DoD

Possible Scenario: Child Left Alone in Car in heat or cold

- In 2019, 51 children died from heatstroke after being left in a hot vehicle, 2 in Indiana.*

* https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/

Possible Scenario: Child Left Alone in Car in heat or cold

- In 2019, 51 children died from heatstroke after being left in a hot vehicle, 2 in Indiana.*

Context & User	Mission	Contextual Info. Propagation
Normal Day & Regular Petrol	Finding an Unattended Child in Car	Send to Appropriate User
During an Earthquake & Rescue Personnel	Finding an Unattended Child in Car	Send to Appropriate User

* https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/

Possible Scenario: Child Left Alone in Car in heat or cold

- In 2019, 51 children died from heatstroke after being left in a hot vehicle, 2 in Indiana.*

Context & User	Mission	Contextual Info. Propagation
Normal Day & Regular Petrol	Finding an Unattended Child in Car	**Bad** Send to Appropriate User
During an Earthquake & Rescue Personnel	Finding an Unattended Child in Car	Send to Appropriate User

* https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/

Possible Scenario: Child Left Alone in Car in heat or cold

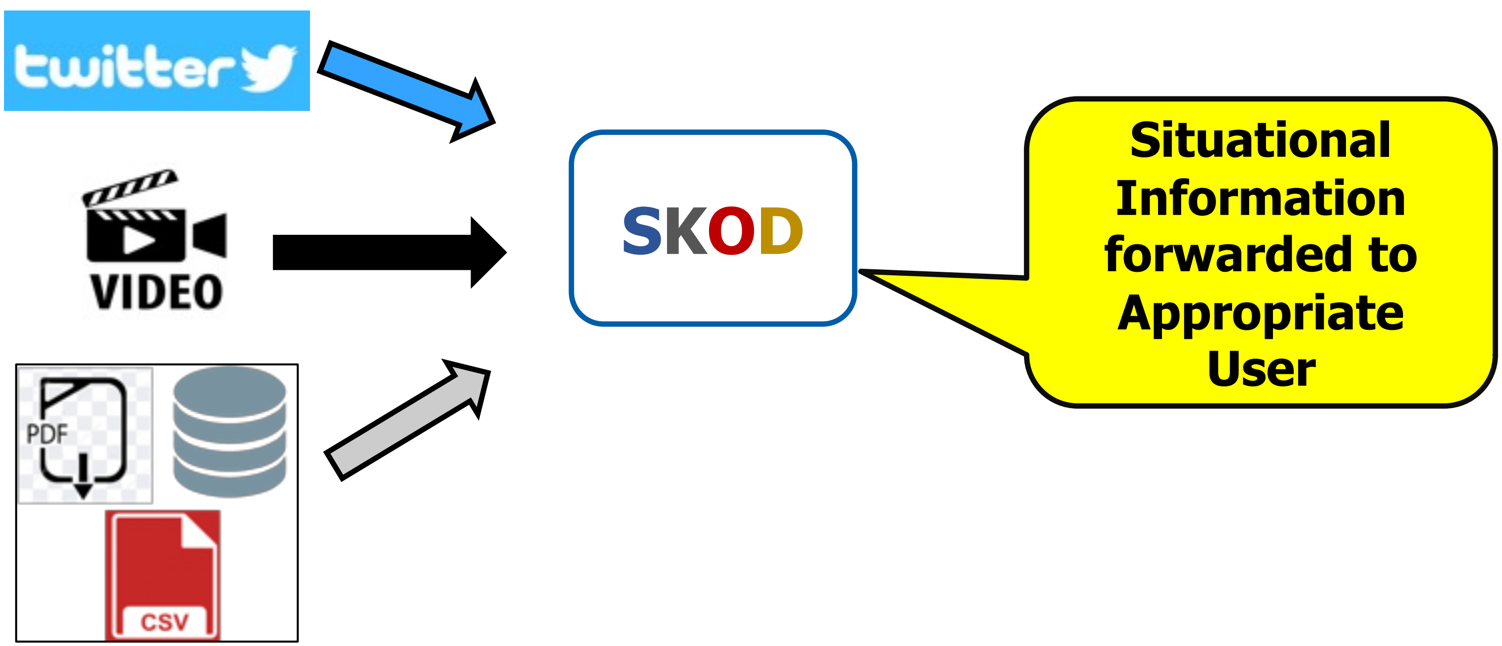
- In 2019, 51 children died from heatstroke after being left in a hot vehicle, 2 in Indiana.*

Context & User	Mission	Contextual Info. Propagation
Normal Day & Regular Petrol	Finding an Unattended Child in Car	**Bad** Send to Appropriate User
During an Earthquake & Rescue Personnel	Finding an Unattended Child in Car	**Good** Send to Appropriate User

* https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/

Possible Scenario: Child Left Alone in Car in heat or cold

- In 2019, 51 children died from heatstroke after being left in a hot vehicle, 2 in Indiana.*



City Data

* https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/hotcars/

Possible Scenario: Suspected Person for Violence

ATF Records
- Record of people buying guns and ammunitions in an area

BMV Records
- Record of DUI Convictions

crimemapping.com
- Is involved in Assault / Disturbing the peace / Homicide / Vandalism

Census Records
- No Family Connection to NYC or close by

GPS tracking
- Headed to NYC times square

Suspected Person

Possible Scenario: Suspected Person for Violence

ATF Records

- Record of people buying guns and ammunitions in an area

BMV Records

- Record of DUI Convictions

crimemapping.com

- Is involved in Assault / Disturbing the peace / Homicide / Vandalism

Census Records

- No Family Connection to NYC or close by

GPS tracking

- Headed to NYC times square

Suspected Person

Context:
New Years Evening

Possible Scenario: Suspected Person for Violence

ATF Records
- Record of people buying guns and ammunitions in an area

BMV Records
- Record of DUI Convictions

crimemapping.com
- Is involved in Assault / Disturbing the peace / Homicide / Vandalism

Census Records
- No Family Connection to NYC or close by

GPS tracking
- Headed to NYC times square

Suspected Person

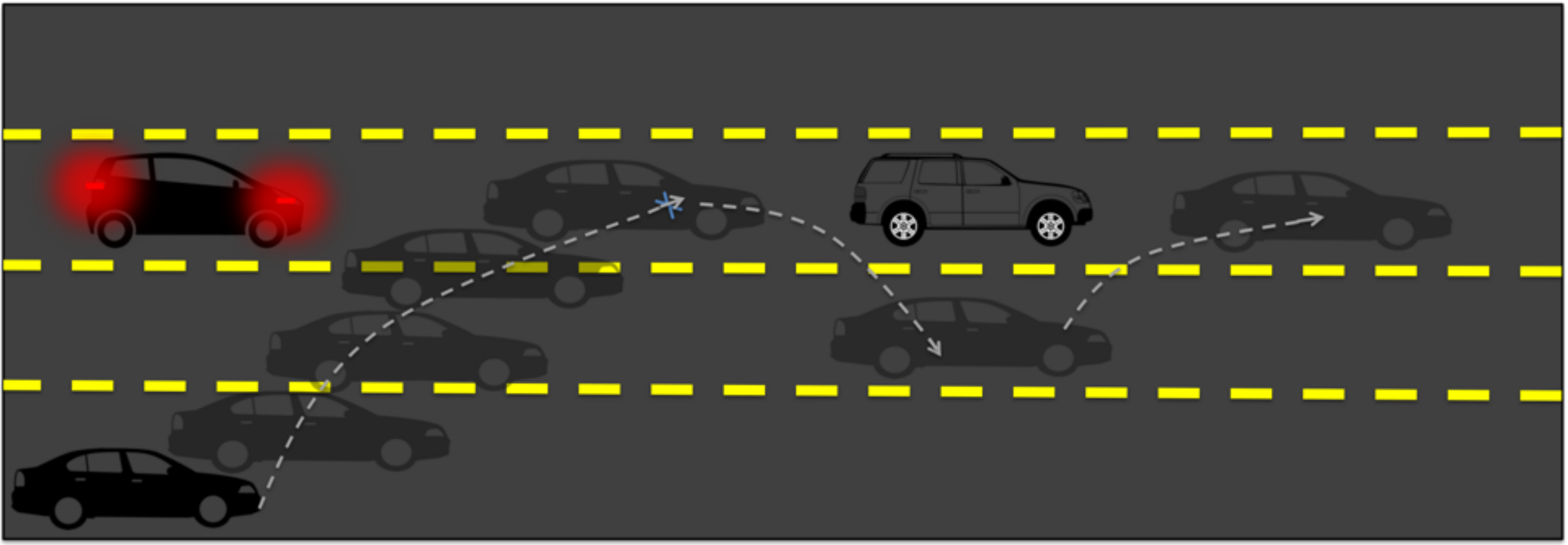
Context: *New Years Evening*

NY Police needs to Know

Possible Scenarios

Possible Scenarios

Identify Unsafe Lane Changes



Possible Scenarios

Identify Jaywalking

SKOD Framework : Agents

- Numerous agents with different missions in a city (i.e., Cambridge)
  - Cambridge police
  - University (Harvard, MIT) police
  - TRANSIT police
  - Cambridge public works
  - Citizens
  - FEMA (Emergency personnel)
  - Homeland Security

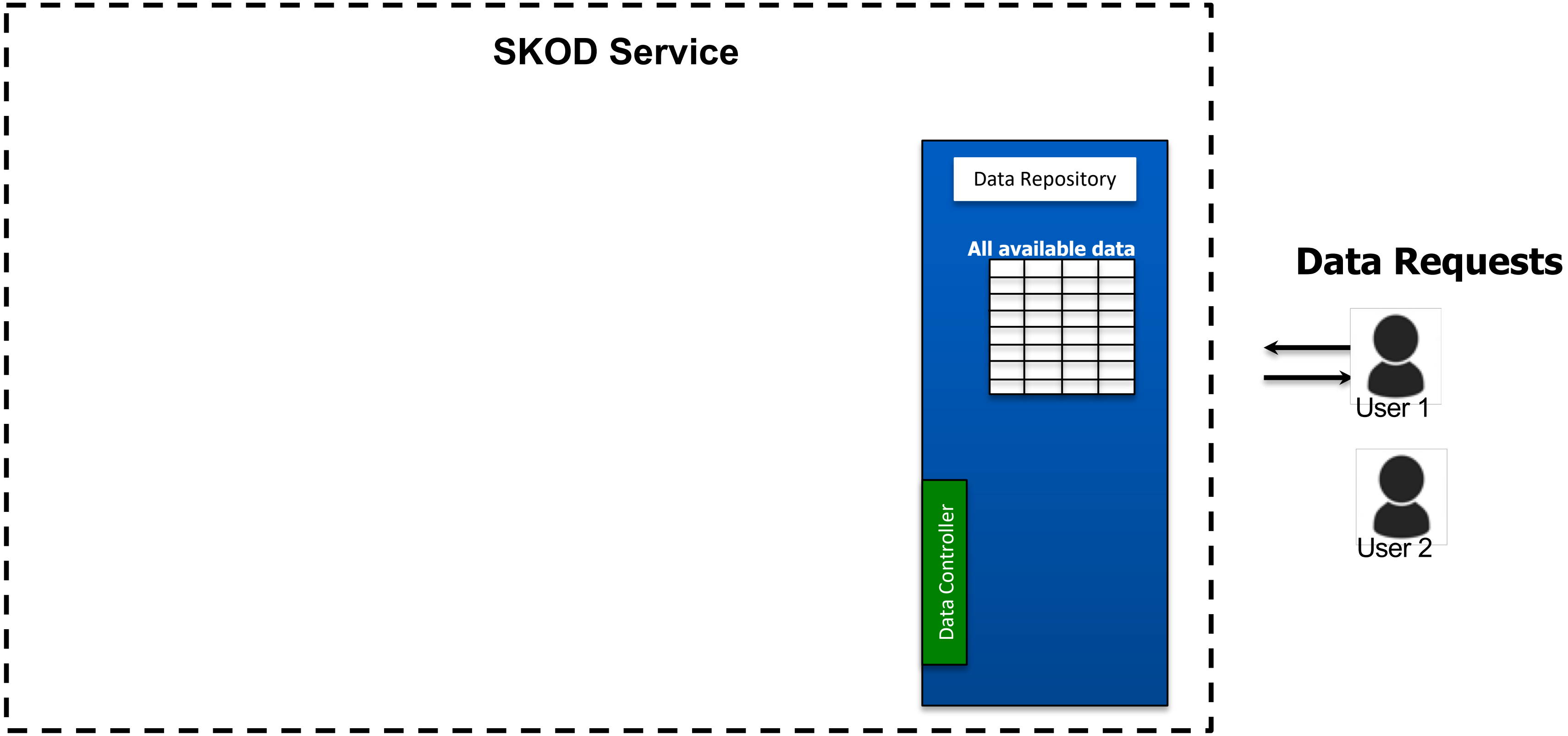
SKOD Framework : Missions

- Missions with various needs for information
  - MIT police (pedestrians in the middle of the road, unsafe lane changes, ”choke” points, Child left alone in parked car, purple Cadillac used by a bad guy identified …)
  - Cambridge public works (potholes, down or occluded street signs)
  - Citizens (crane or car illegally blocking the sidewalk in front of house)
- SKOD framework consists of
  - Multimodal data with Multiple Users with different needs
  - Streaming and Restful data

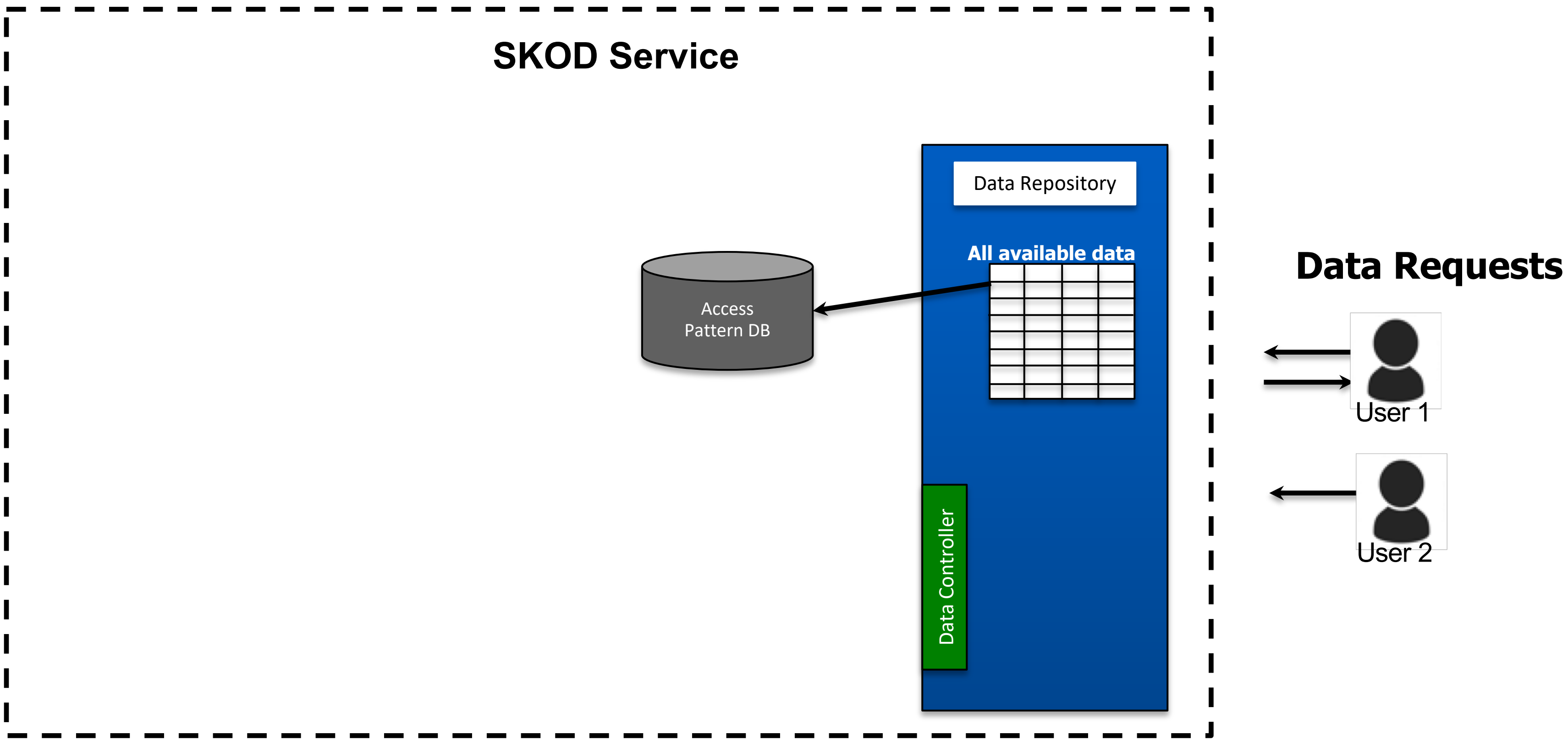
SKOD Objectives

- Retrieve knowledge needed by multiple users with *changing* needs based on Situational Awareness

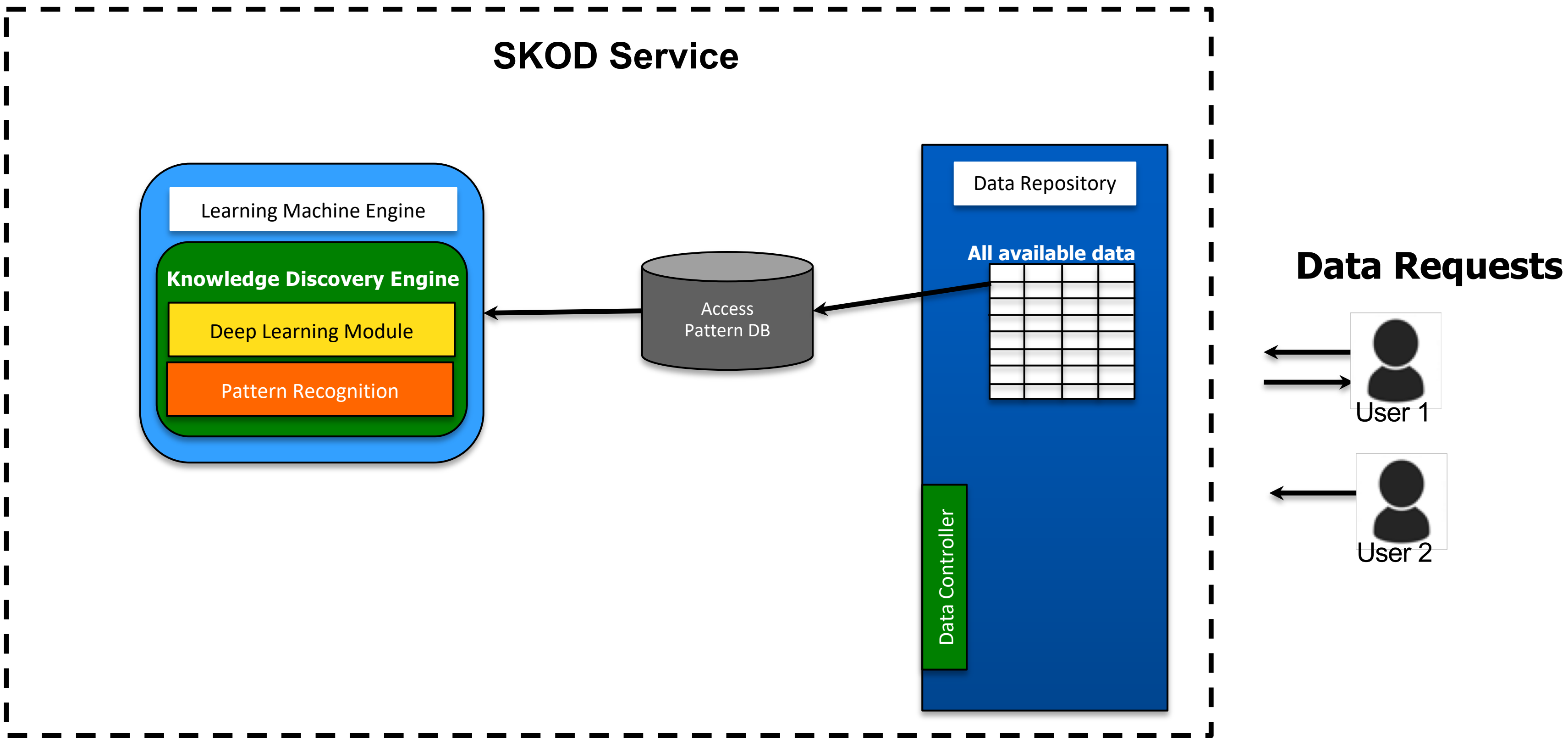
SKOD Objectives



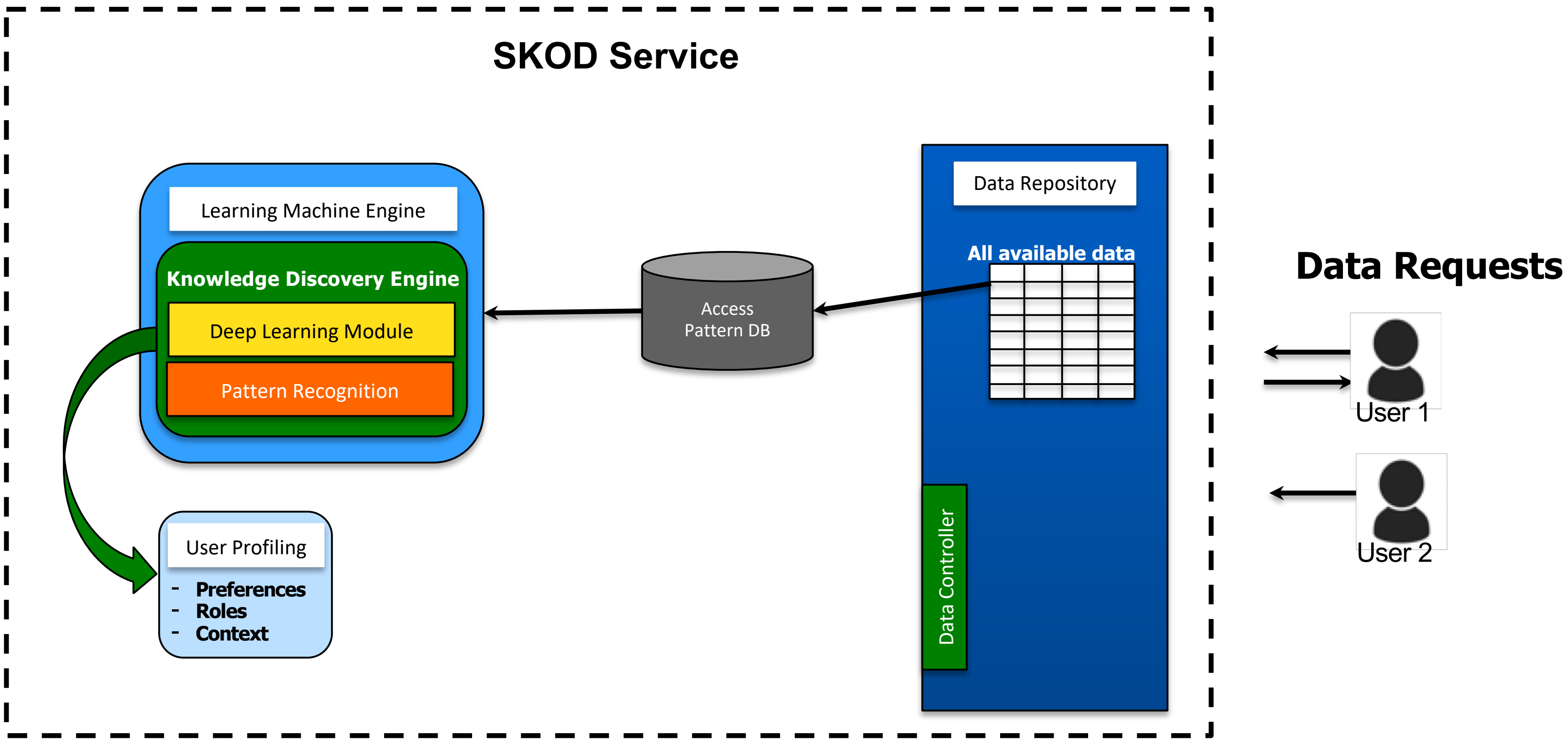
SKOD Objectives



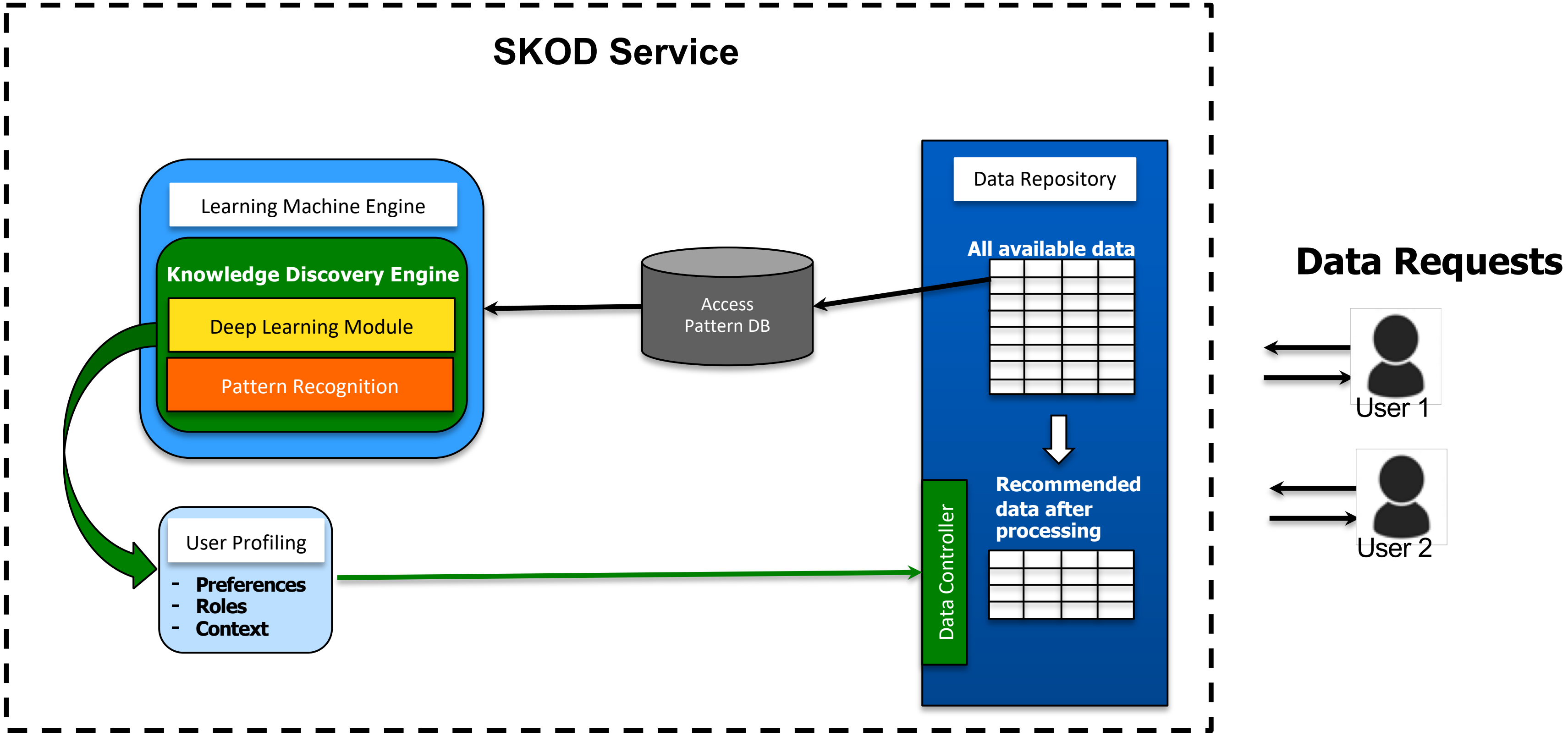
SKOD Objectives



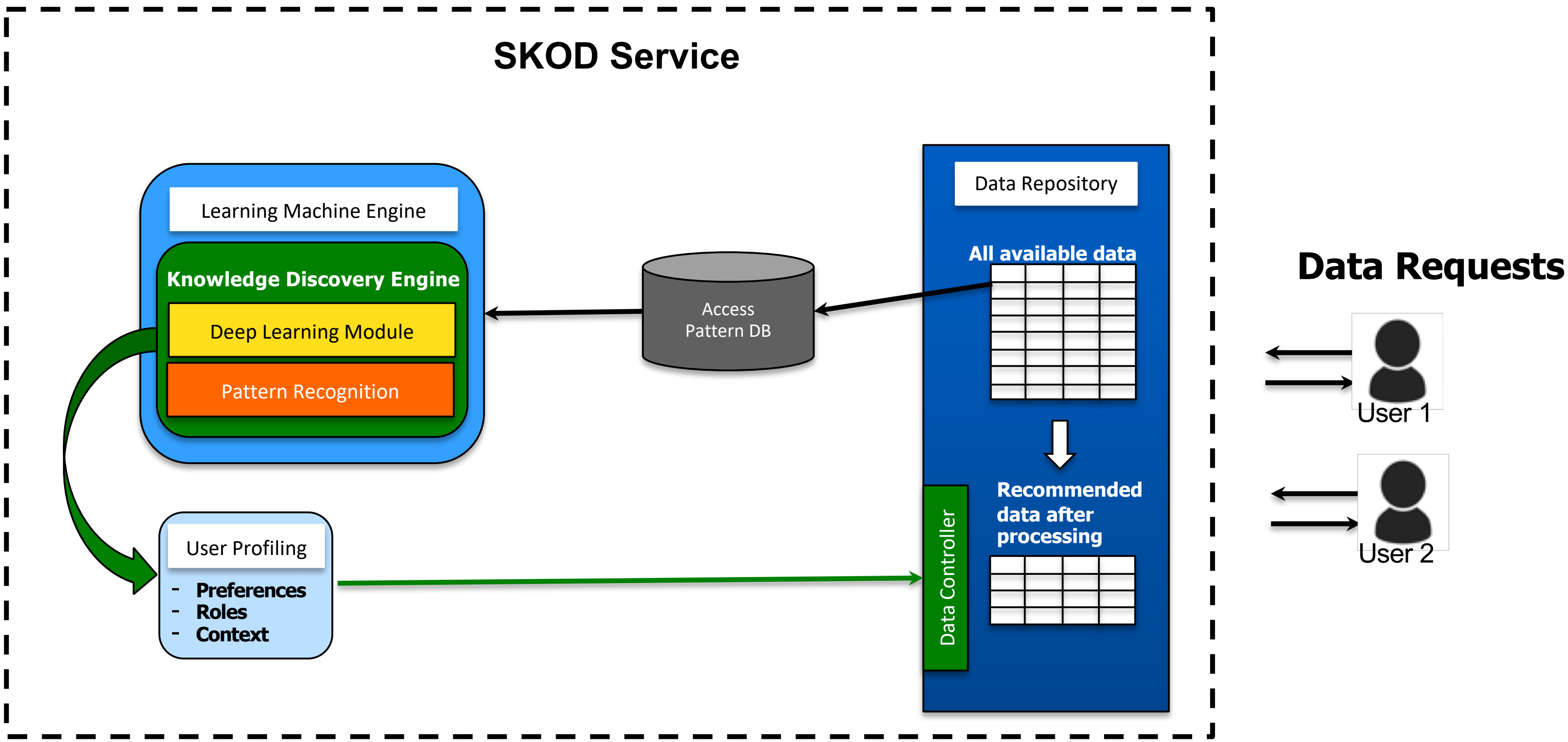
SKOD Objectives



SKOD Objectives



SKOD Objectives

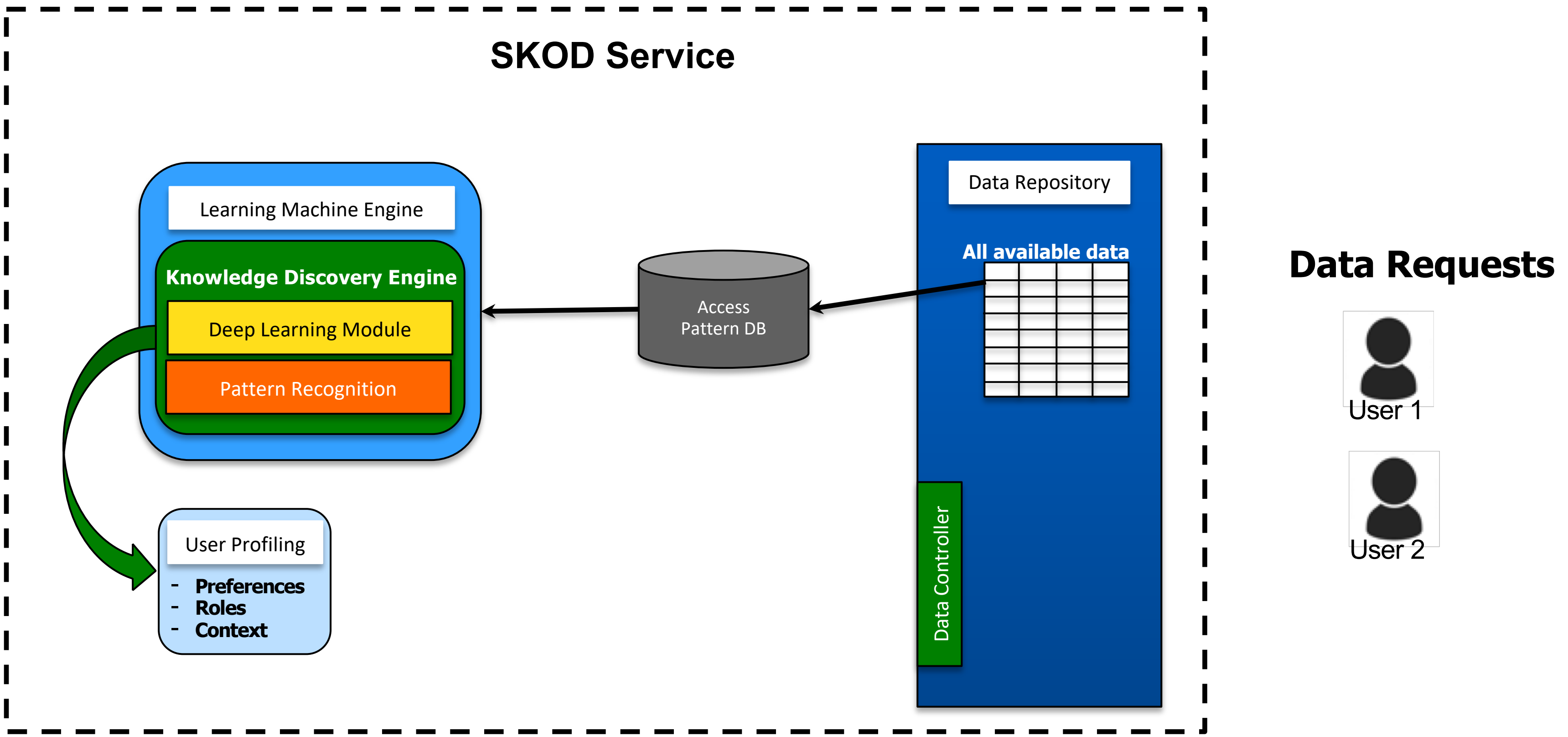


Objective 1: Relevant data is efficiently passed to users based on their requests

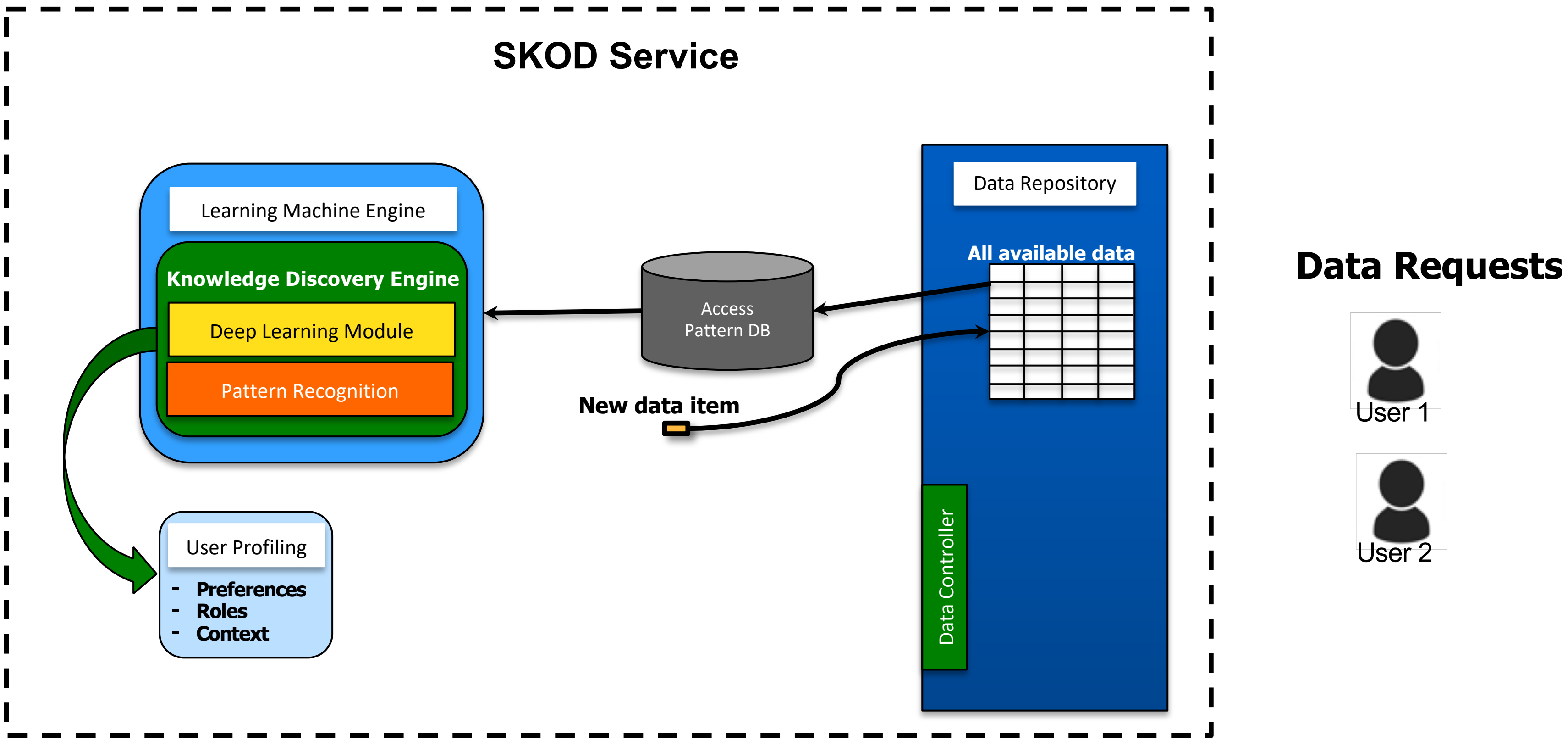
SKOD Objectives

- Retrieve knowledge needed by multiple users with *changing* needs based on Situational Awareness
- Relate multi-modal data and update the knowledge for users
- Integrate new *streaming data* with queries already used by mission
- Complete the unfulfilled data needs for missions based on the Situation and User Preference

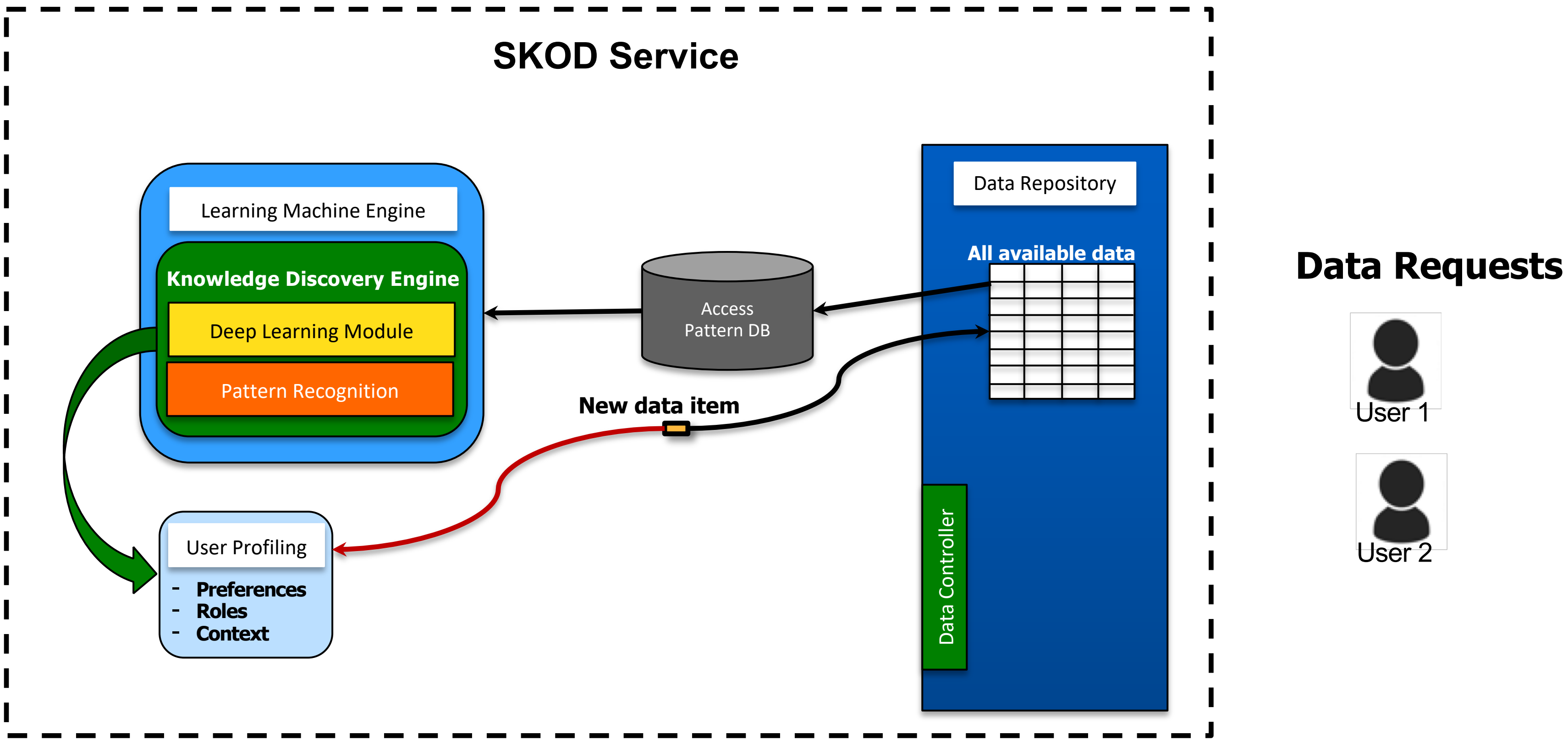
SKOD Objectives



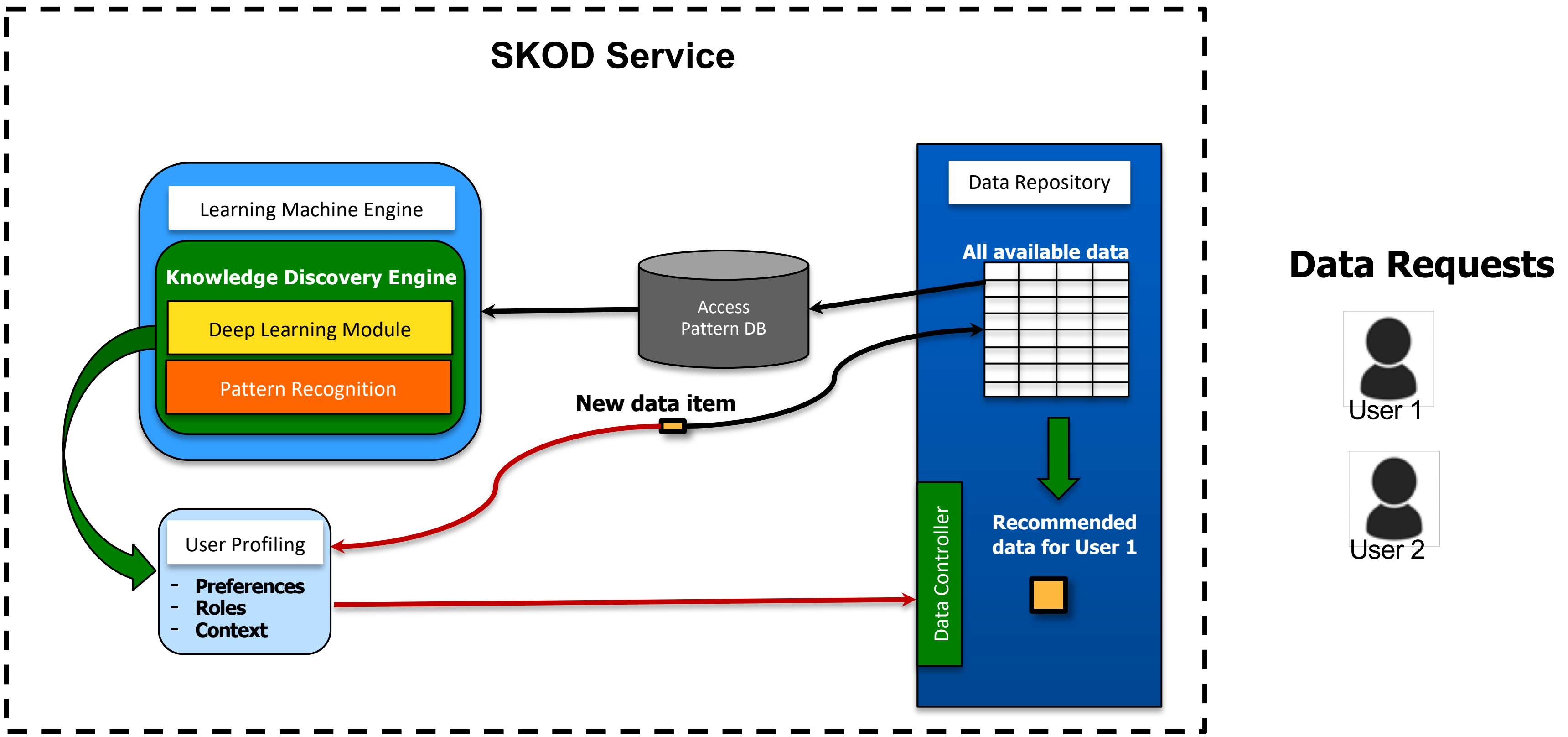
SKOD Objectives



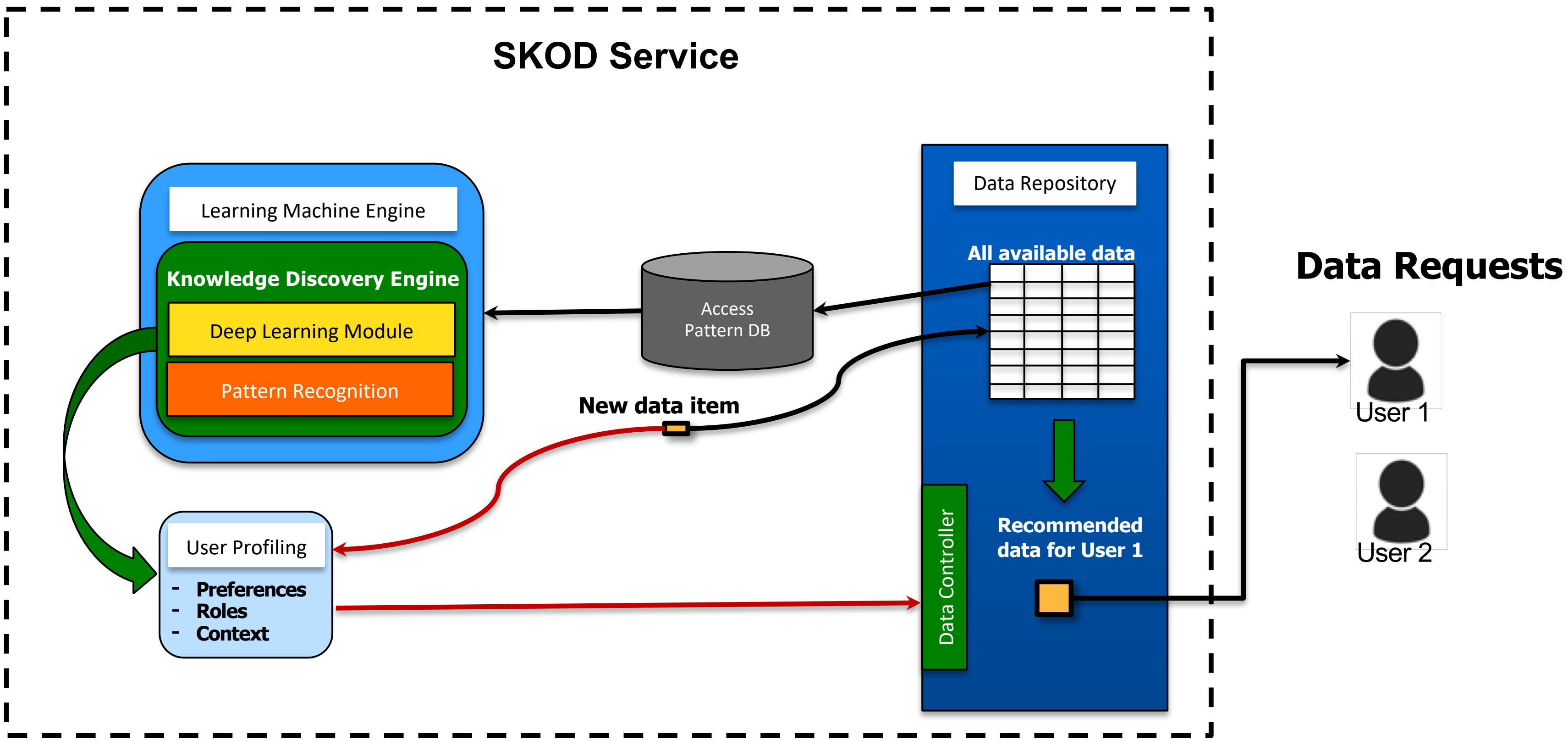
SKOD Objectives



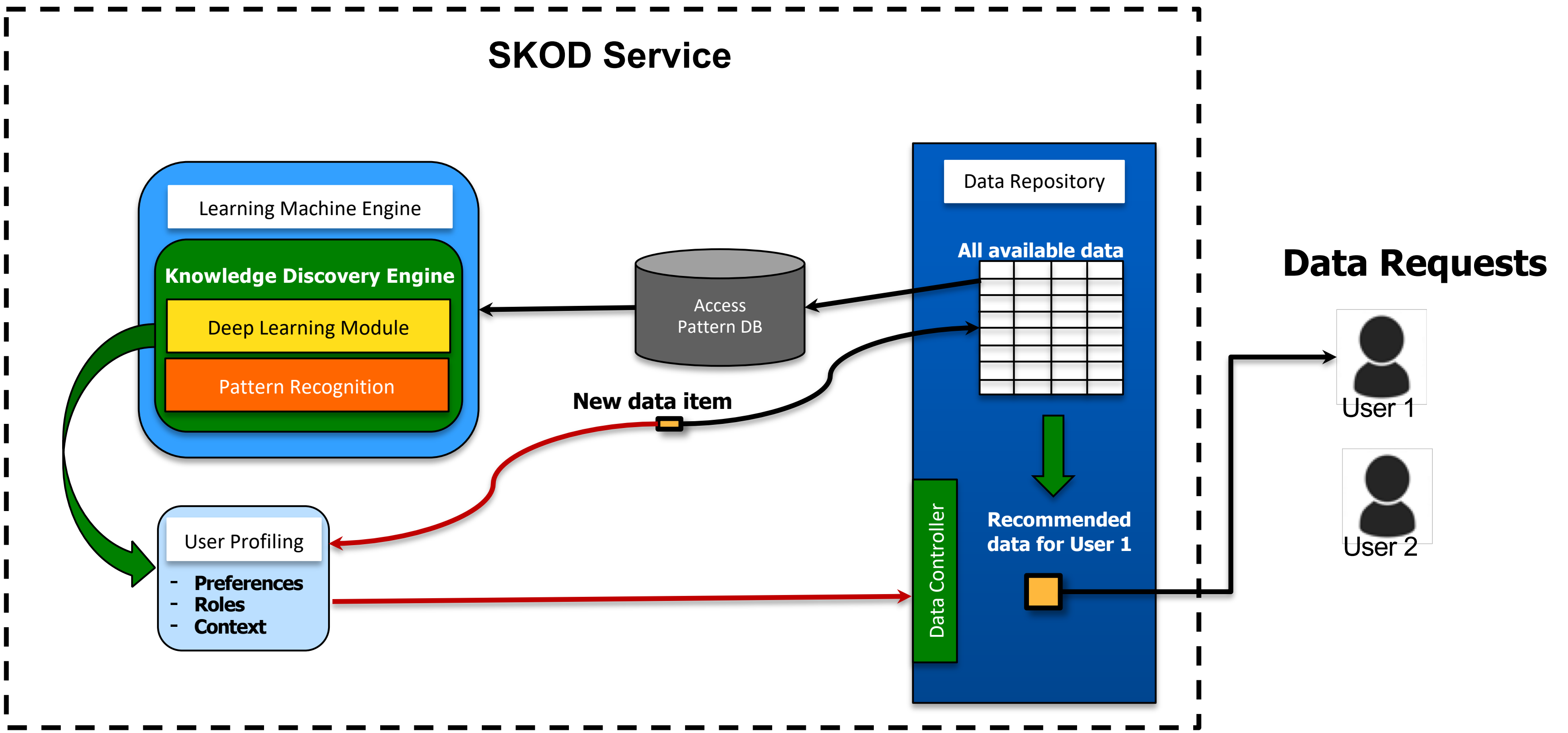
SKOD Objectives



SKOD Objectives



SKOD Objectives



Objective 2: New data items are directed to interested users based on User Profiling.

SKOD Framework : Research Directions

- CNN based Neural Networks and Transfer Learning for objects from Video.
- Generative and Deep Learning (encapsulating Word2Vec) models for topics, ontologies and triplets (KG) from Text.
- DL model combining attention based Bi-LSTM and CNN [4] to classify tweets for Disaster Resource Management and similar scenarios.
- BlazeIt [5] for complex queries over video related to objects of interest.
- Research DAWN’s End-to-End ML Systems [6] for Recommendation.
- Research reinforcement learning and active learning for User Profiling.
- Apply models to other NG large databases (sensors, signals, text, phone calls, videos, images, voice)

Problem Statement

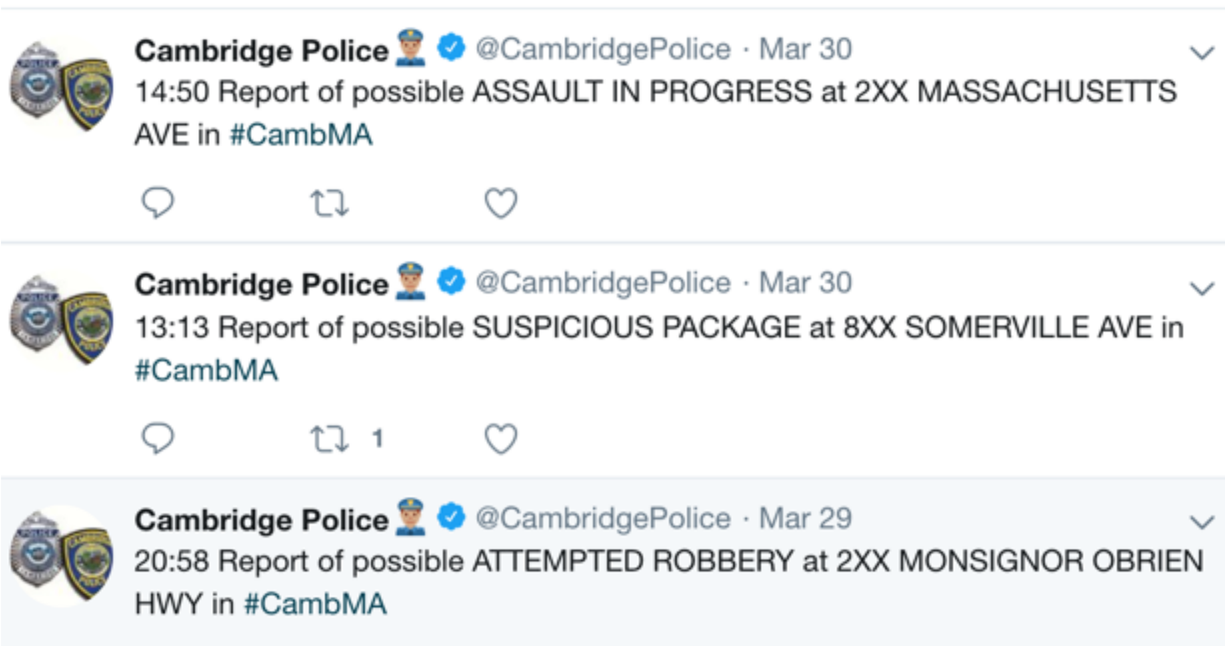
Determine relevant information from heterogeneous data at rest and data streams, and deliver it to the right user based on situational awareness. Build context-aware knowledge on top of relational database utilizing user queries and deliver missing information to fulfill mission requirements.

Datasets

- **Video**
  - 100+ hours of dashcam video collected at MIT
  - Raw video can be retrieved from MIT database at Cambridge
    - Split into chunks of 30 seconds
    - Metadata collected: geolocation and timestamp for each 30 seconds
- **Unstructured Text** (Twitter data)
  - Collected ~200K tweets (Target ~ 1 million)
  - Automatic tweet parsing and recording system into Postgres in place
- **Structured data**
  - Cambridge public datasets
  - Automatic weekly updates into Postgres in place
- **Data from drones and dashcams**

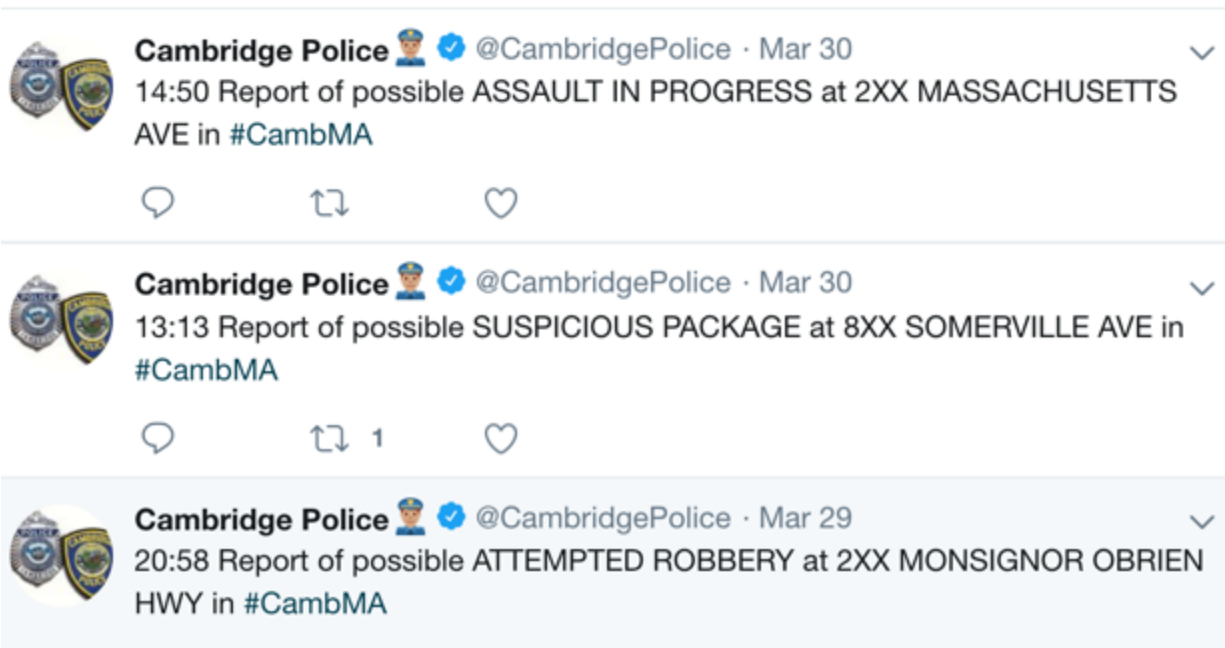
Datasets Example

- Tweets from Cambridge Police
- A video that has a bicyclist without helmet on it 00:01 to 00:27



Datasets Example

- Tweets from Cambridge Police
- A video that has a bicyclist without helmet on it 00:01 to 00:27

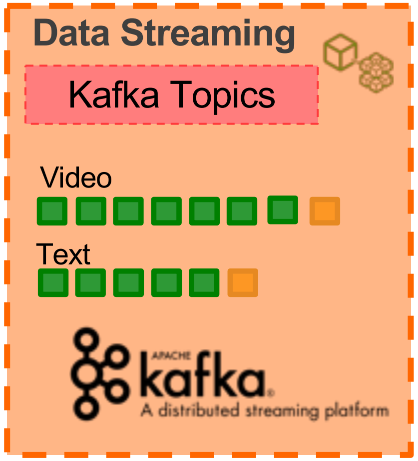


Future Datasets

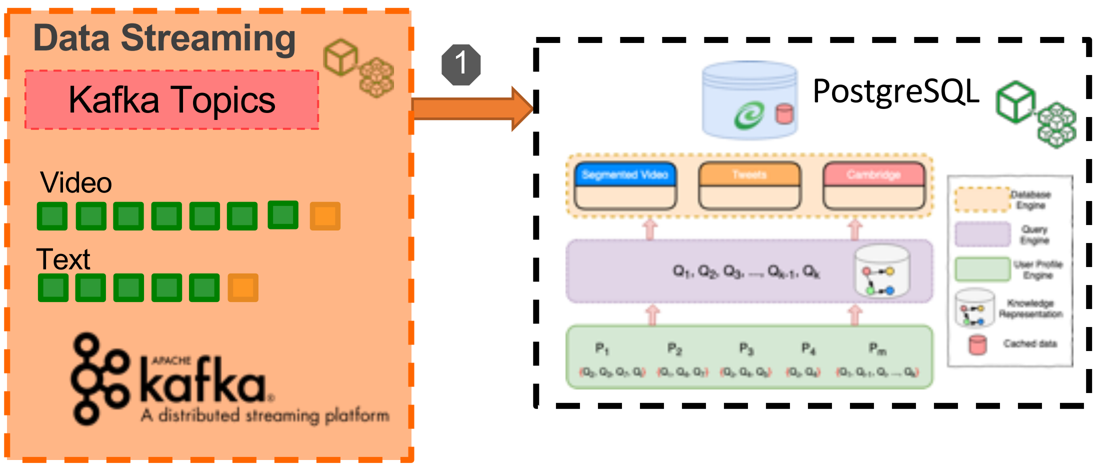
- Waymo Open Dataset
  - Sensor data
    - Synchronized lidar and camera data from 1,000 segments (20s each)
  - Labeled data
    - Labels for 4 object classes - Vehicles, Pedestrians, Cyclists, Signs
- Yelp Dataset
  - Reviews
  - Businesses
  - Pictures
  - Metropolitan Areas
- News Articles
  - https://www.cambridgema.gov/news?page=2&ResultsPerPage=10
  - Google News

https://waymo.com/open/; https://www.yelp.com/dataset

Architecture



Architecture



Microservice

Users' queries

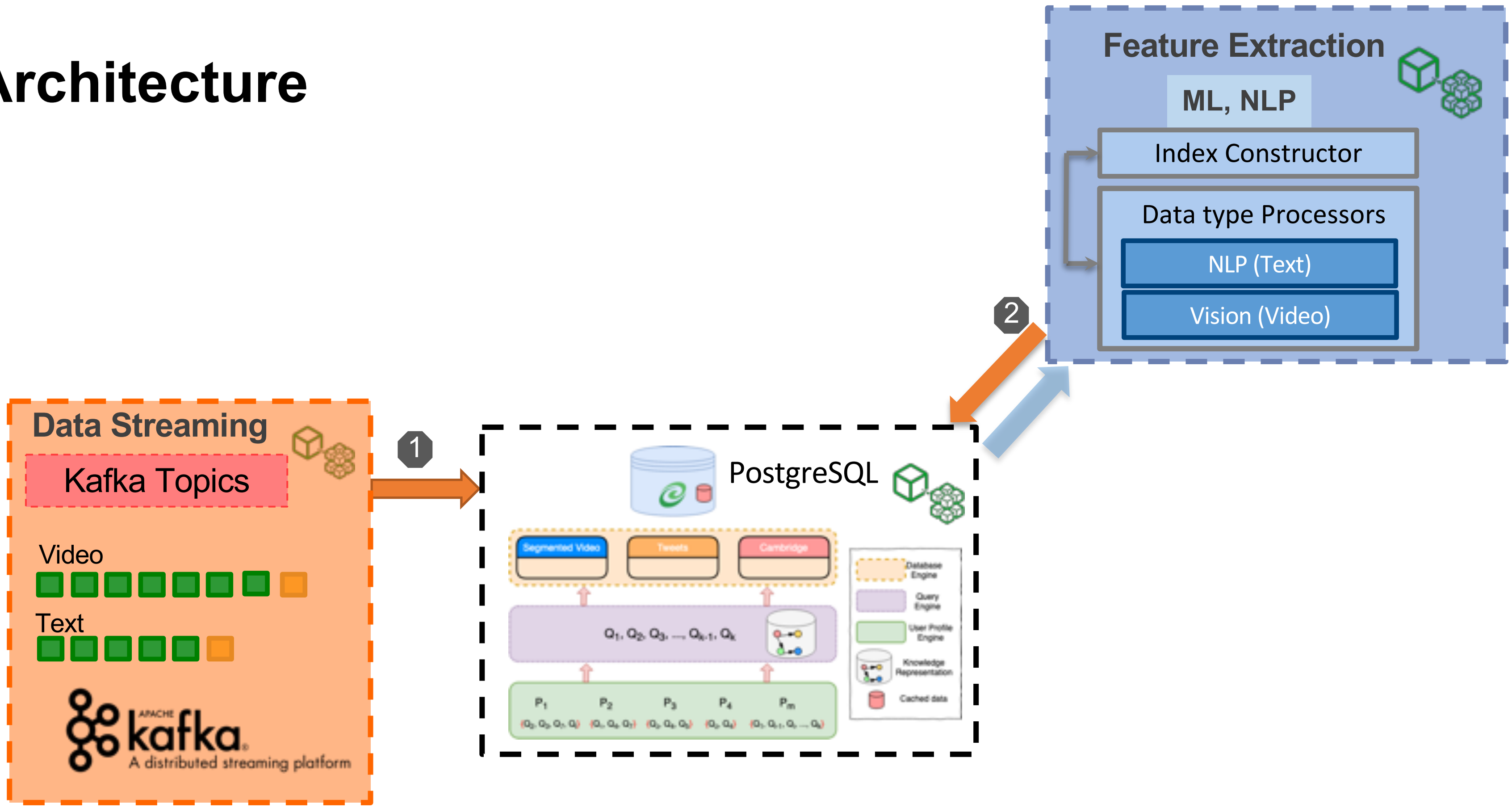
Heterogeneous Data Streams

Knowledge derived from queries

Situational Aware Indexed Data

Relevant patterns of data

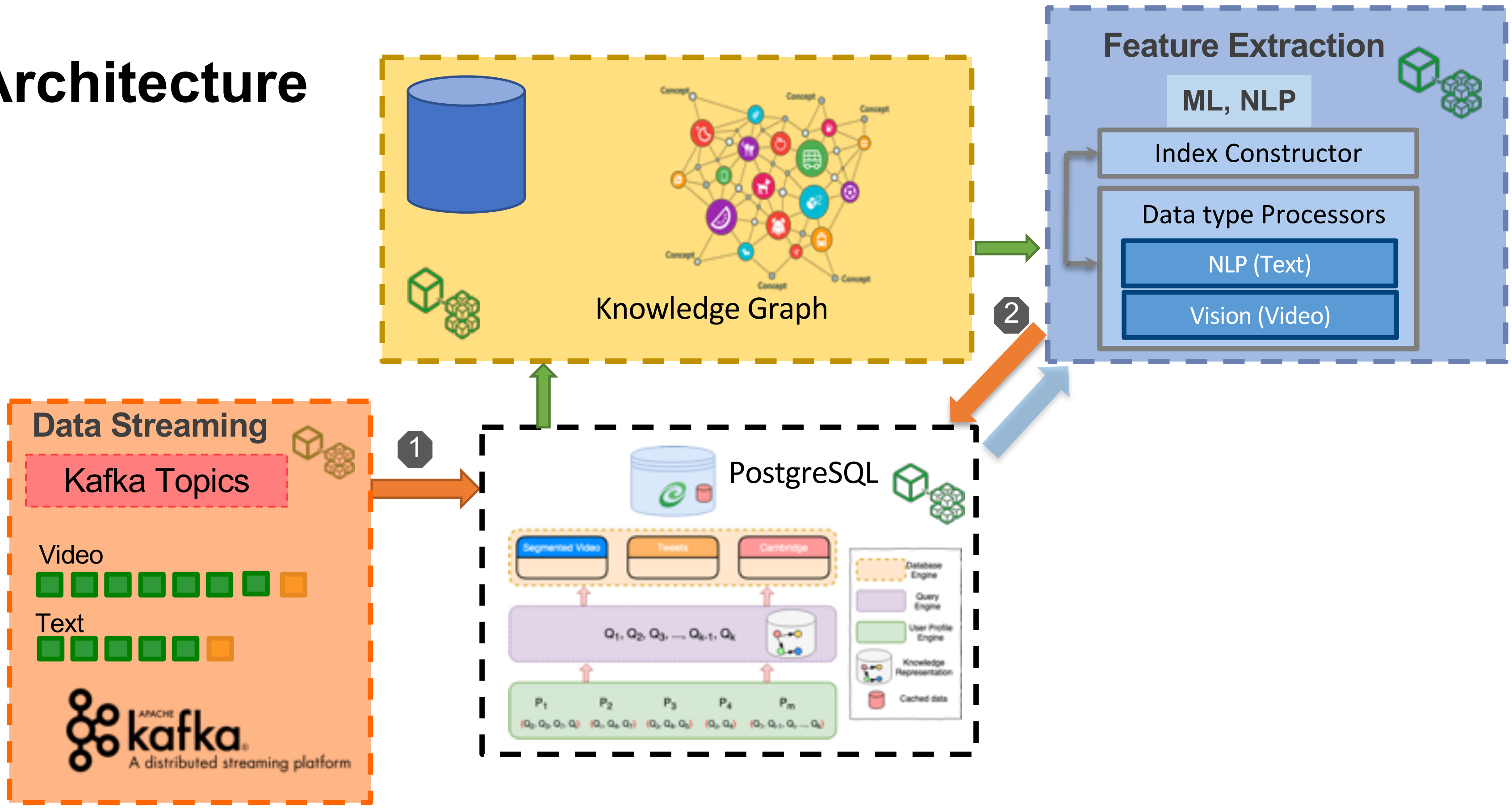
Architecture



Microservice

- Users' queries
- Heterogeneous Data Streams
- Knowledge derived from queries
- Situational Aware Indexed Data
- Relevant patterns of data

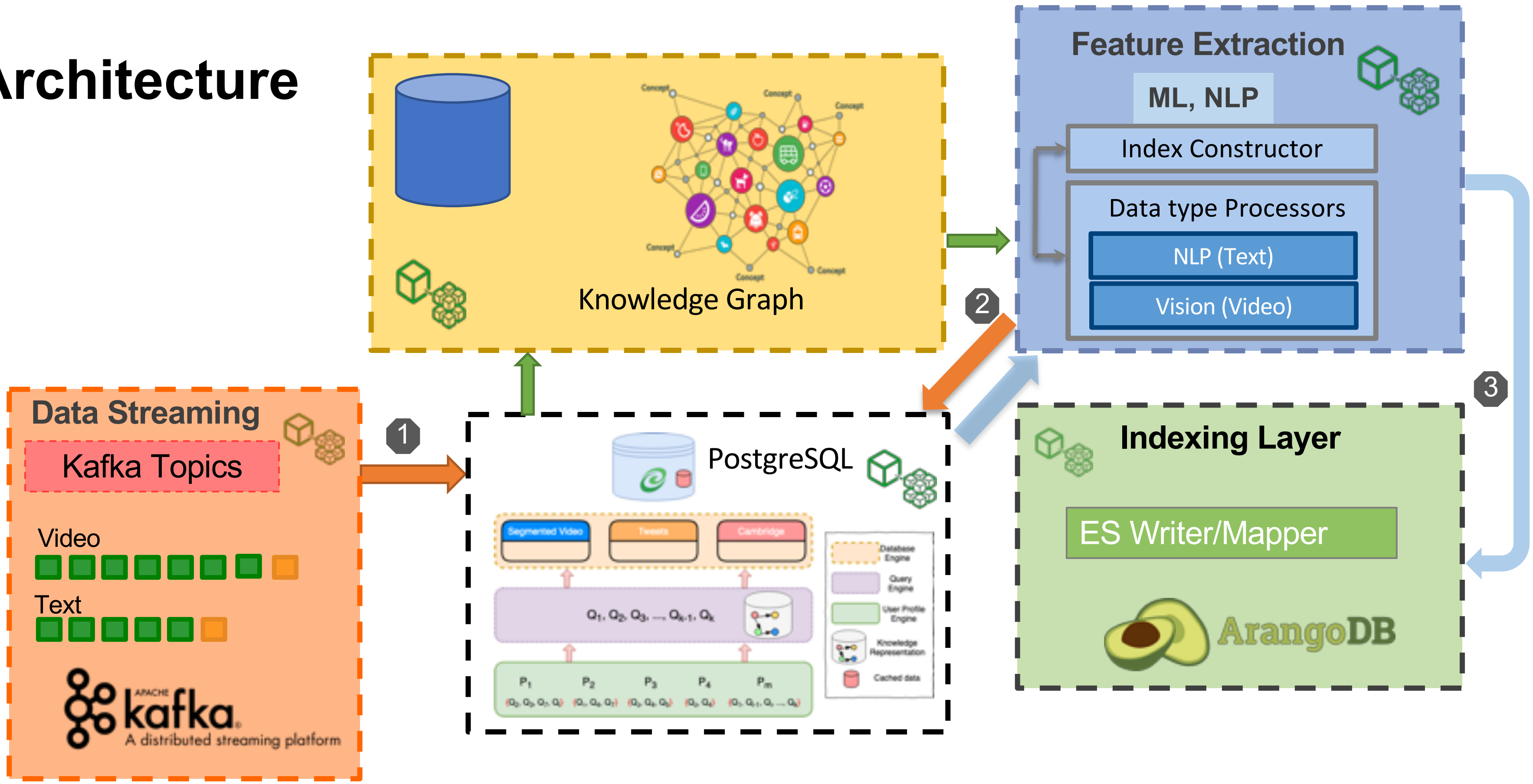
Architecture



Microservice

- Users' queries
- Heterogeneous Data Streams
- Knowledge derived from queries
- Situational Aware Indexed Data
- Relevant patterns of data

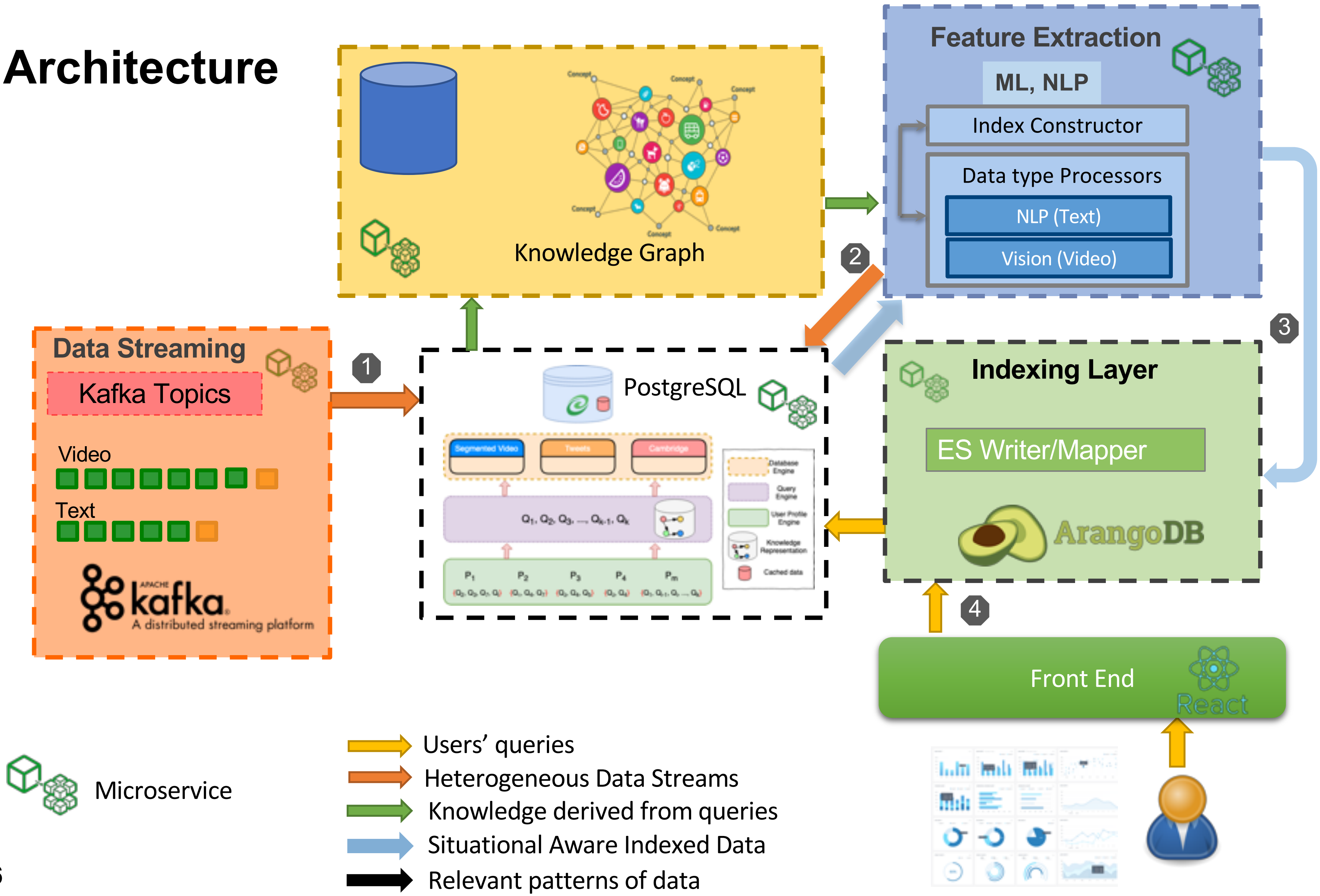
Architecture



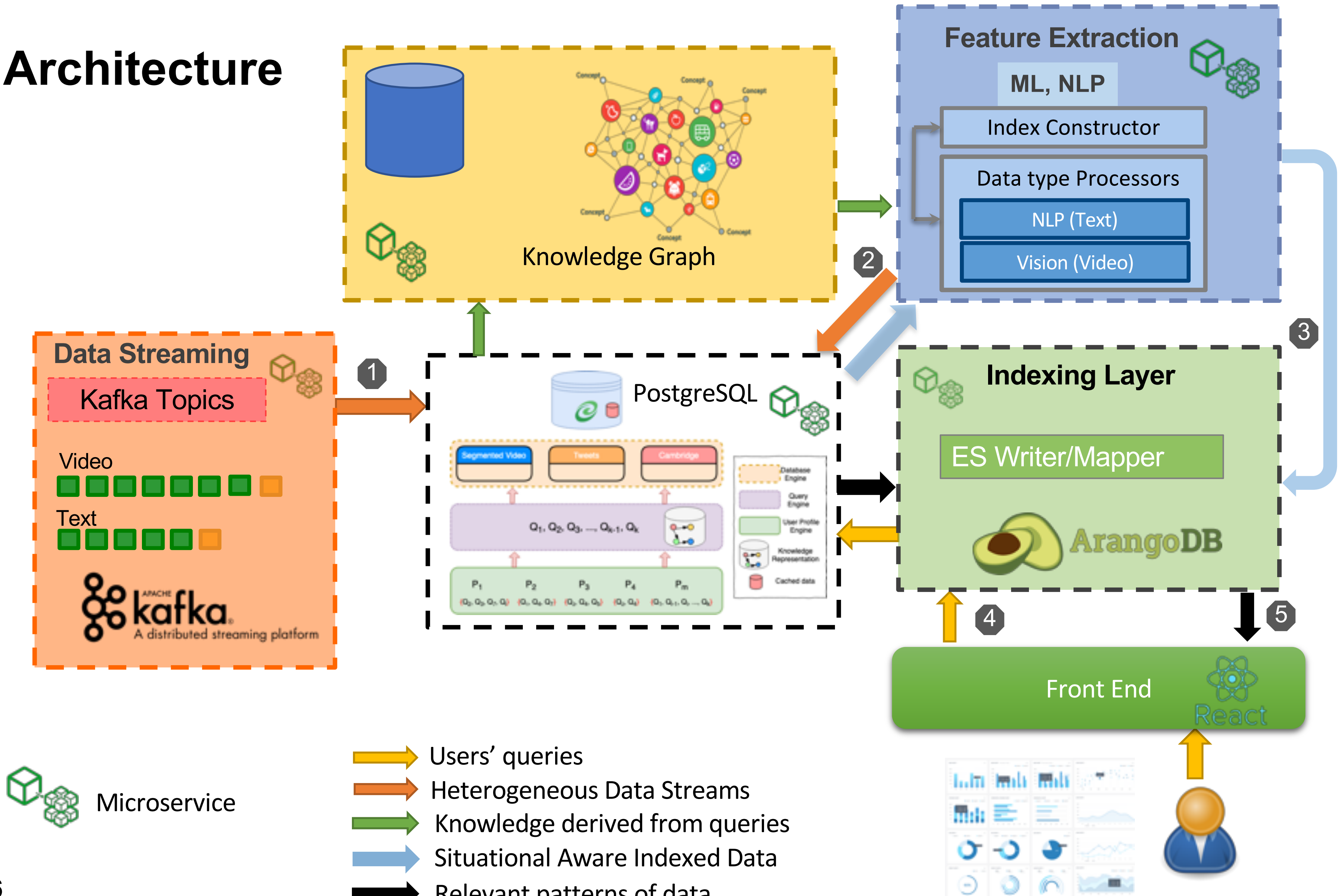
Microservice

- Users' queries
- Heterogeneous Data Streams
- Knowledge derived from queries
- Situational Aware Indexed Data
- Relevant patterns of data

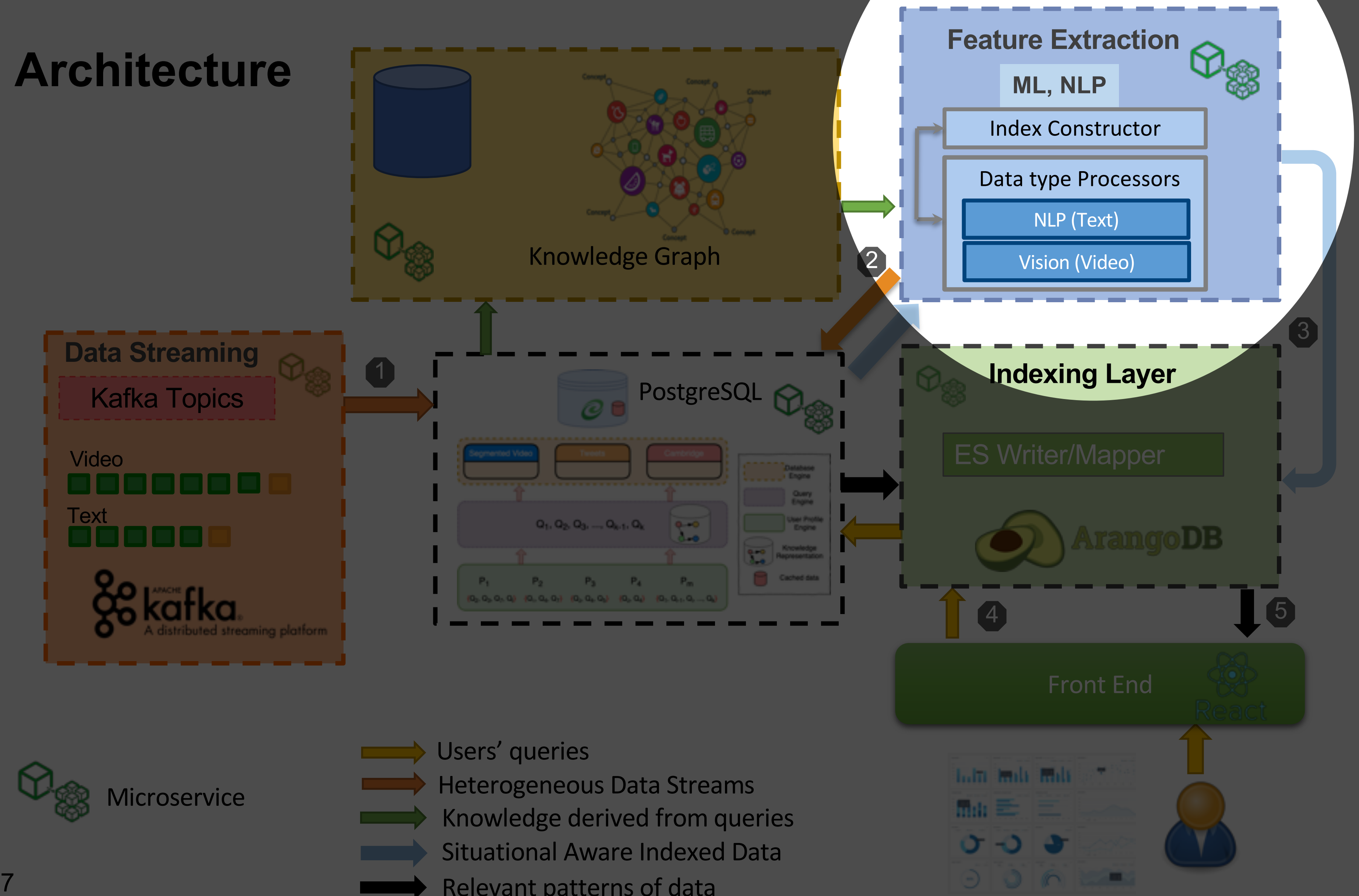
Architecture



Architecture



Architecture



Feature Extraction Module

- Example Query

```
Select * from tweets, videos where
tweets.objects_discussed == "car" and
tweets.objects_discussed == "child" and
videos.objects_detected == "car" and
videos.objects.detected == "child"
```

- Answer queries such as above
- Find interesting features from incoming data and data at rest
- Relate data from different modalities

Extracting Features from Video with Deep Learning

- Object detection and classification: best result achieved with deep learning architectures:
  - Faster RCNN
  - YOLO
  - SSD
- Manual annotation and labeling
  - Time-consuming and expensive for large datasets
  - Outsourced human labor can be employed (MTurk)
- We use *pre-trained* YOLO neural network to extract knowledge, detect and label objects in video
- Retrain YOLO with Transfer Learning for detecting classes outside of pretrained ones

Neural Network For Object Detection and Classification

- YOLO detects 100+ classes
- Our raw video dataset contains about 15 of the objects from these classes
- YOLOv3 object detection algorithm
  1. Regions of interests (ROI) proposals are generated
  2. For each region, features are extracted and classified with Convolutional Neural Network
  3. Apply non-maximum suppression: all candidate regions where probability of certain object detection is not max are dismissed

YOLO (You Only Look Once) v3 Architecture

1. The image is split into an *SxS* grid of cells.

2. Each grid predicts *B* bounding boxes with *C* class probabilities

- *SxSxBx5* outputs in total

3. Conditional class probabilities are predicted *Pr(Class(i)/Object):*

- *SxSxC* class probabilities
- *SxSx(B*5+C)* output tensor
- *S=7, B=2, C=20 => (7,7,30)*
- Train a CNN to predict (7,7,30) tensor

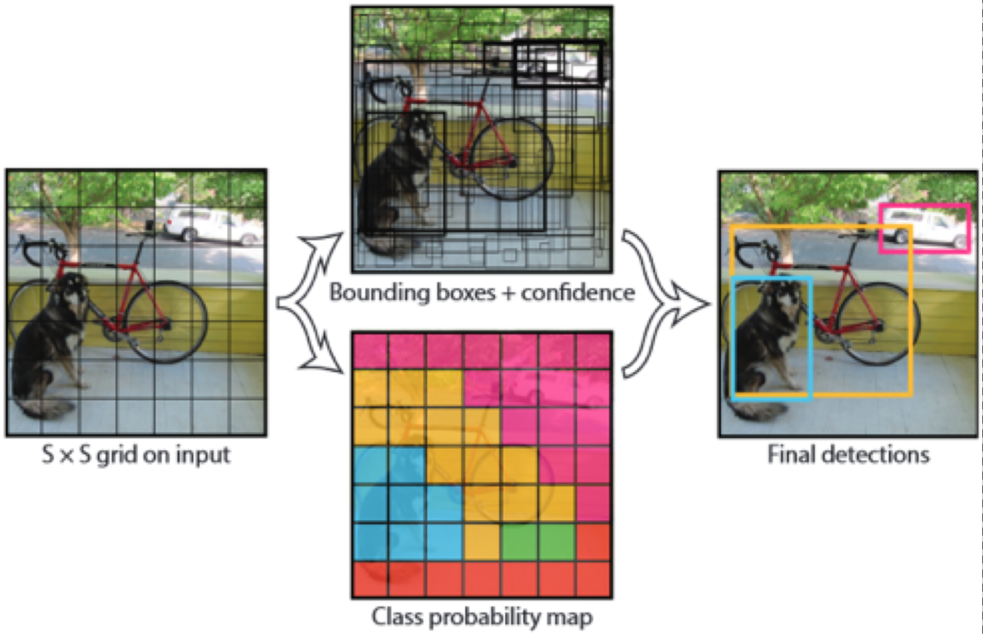


Image source: **You Only Look Once: Unified, Real-Time Object Detection**
Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi
https://arxiv.org/abs/1506.02640

Detected Classes In the MIT Video Dataset

CAR

TRUCK

PERSON

BICYCLE

TRAFFIC LIGHT

STOP SIGN

FIRE HYDRANT

PARKING METER

... AND MORE!

Preprocessing Tweets

- Social media text has jargon, misspellings, special slangs, emojis

15:45 I luv my <3 iphone & you’re awsm apple, love you 3XXX. DisplayIsAwesome, sooo happppppy 🙂 🙏 http://www.apple.com #apple @sjobs

- Cleaning process –
  - HTML decoding
  - Expanding Contractions
  - Removing URL, Emoji, Reserved words, Smiley, User-mentions (or replace), hashtags
- Preprocessing before tokenization
  - Remove punctuation, space, stop word

Future Tasks: Preprocessing Tweets

- Normalization of Noisy Text
- Awsm ~ awesome, luv ~ love
- Methodologies
    1. Lexical normalization
    2. Normalization with edit scripts and recurrent neural embeddings
    3. Find balance between precision and recall

Topic Modeling with Tweets

- Latent Semantic Analysis, or LSA
  - Find document-term matrix with tf-idf
  - Topics are latent
  - Dimensionality reduction with SVD, gives our term-topic matrix
- Apply cosine similarity to evaluate:
  - the similarity of terms (or "queries") and documents (we want to retrieve passages most relevant to our search query).

occurrences of term in document

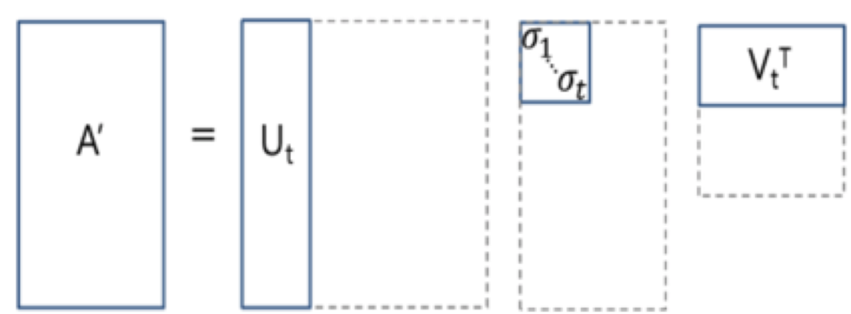
total documents

$$w_{i,j} = tf_{i,j} \times \log \frac{N}{df_j}$$

tf-idf score

documents containing word

$$A \approx U_t S_t V_t^T$$



Same User with Different Levels of Interest

U_1 TREE DOWN

U_2 PERSON with GUN

Data at Rest

D_0	D_1	D_2	D_3
D_4	D_5	D_6	D_7
D_8	D_9	D_{10}	D_{11}
D_{12}	D_{13}	D_{14}	D_{15}

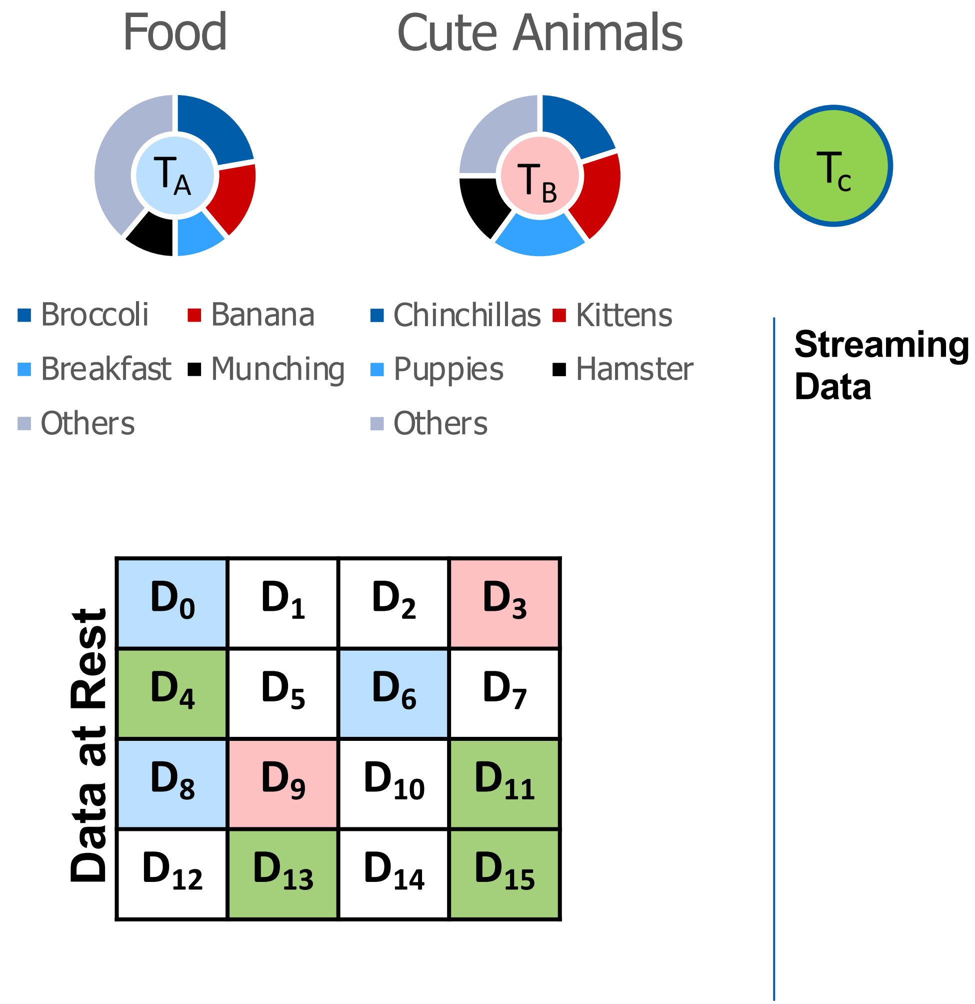
Same User with Different Levels of Interest

Topic Modeling for Ontologies (Generative Models)

- Even though LSA *finds* similar documents to user query, it has *less efficient* representation for topics.
- Topics are necessary for ontologies while building our knowledge graph
- LDA (Latent Dirichlet Allocation)
  - Generative Model
  - Uses Dirichlet priors for the document-topic and word-topic distributions
  - Results in better generalization for new documents
  - Allows online learning

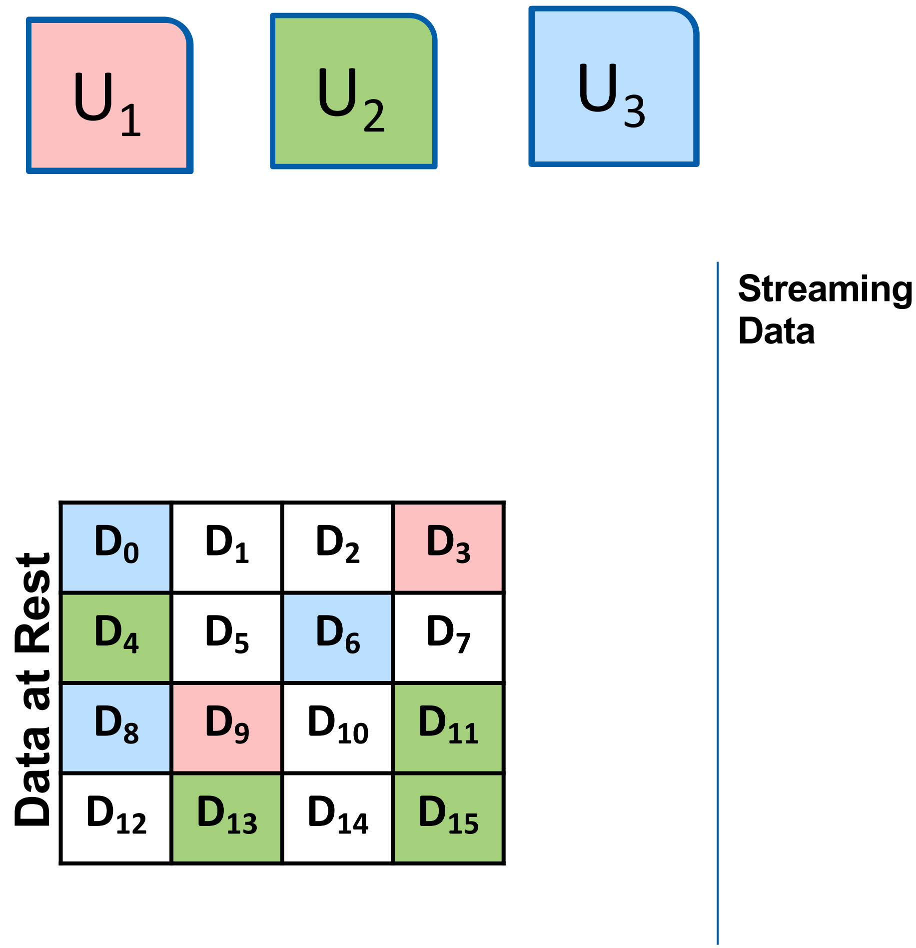
Multiple Data of Interest to Different Users

- Extract human-interpretable topics from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- Uses variational Bayes for inference, no need to re-train



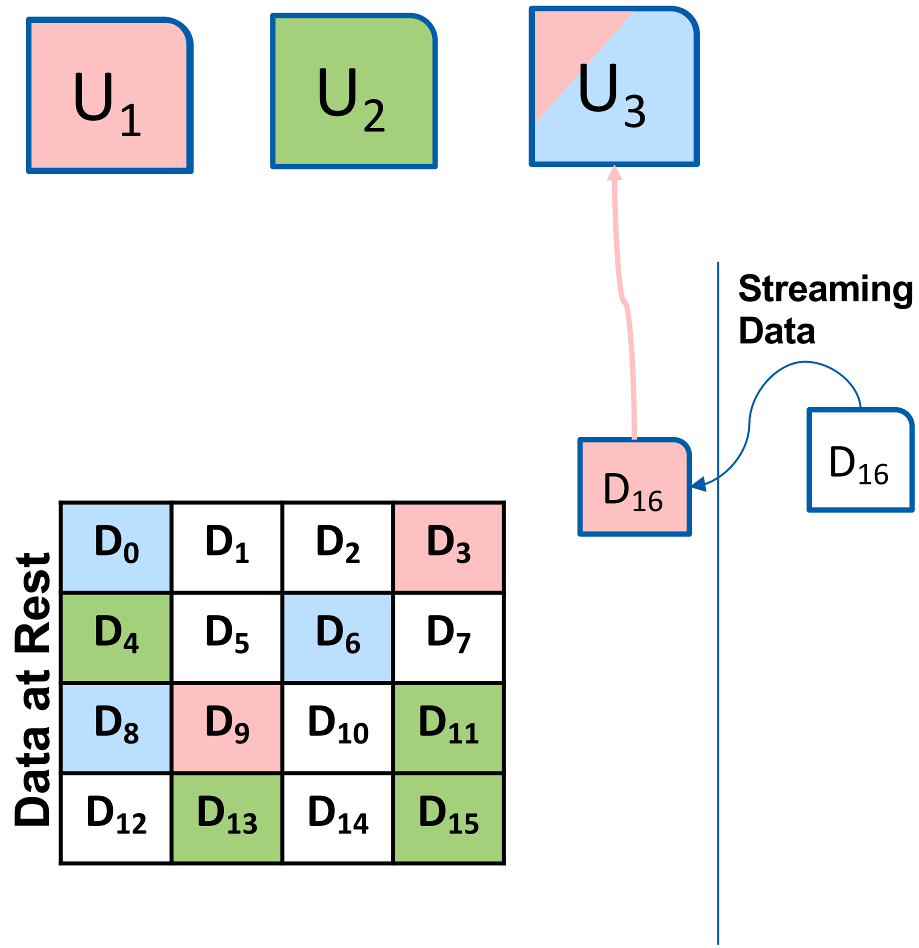
Multiple Data of Interest to Different Users

- Extract human-interpretable topics from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- Uses variational Bayes for inference, no need to re-train



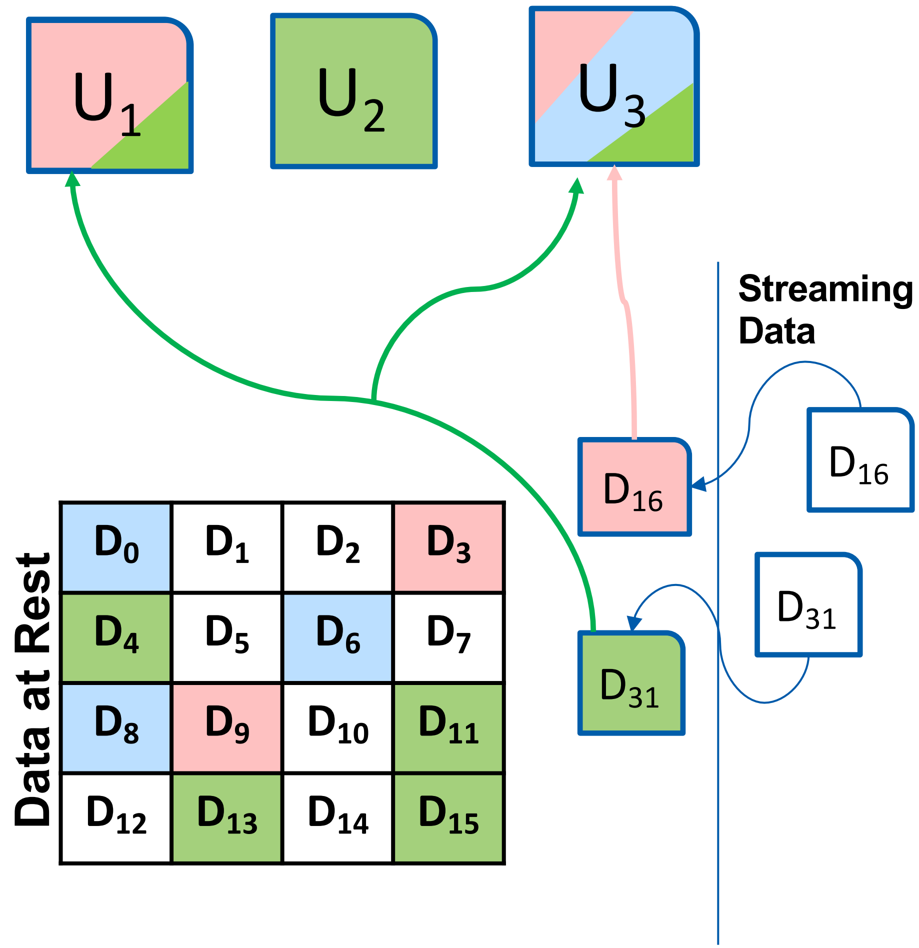
Multiple Data of Interest to Different Users

- Extract human-interpretable topics from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- Uses variational Bayes for inference, no need to re-train



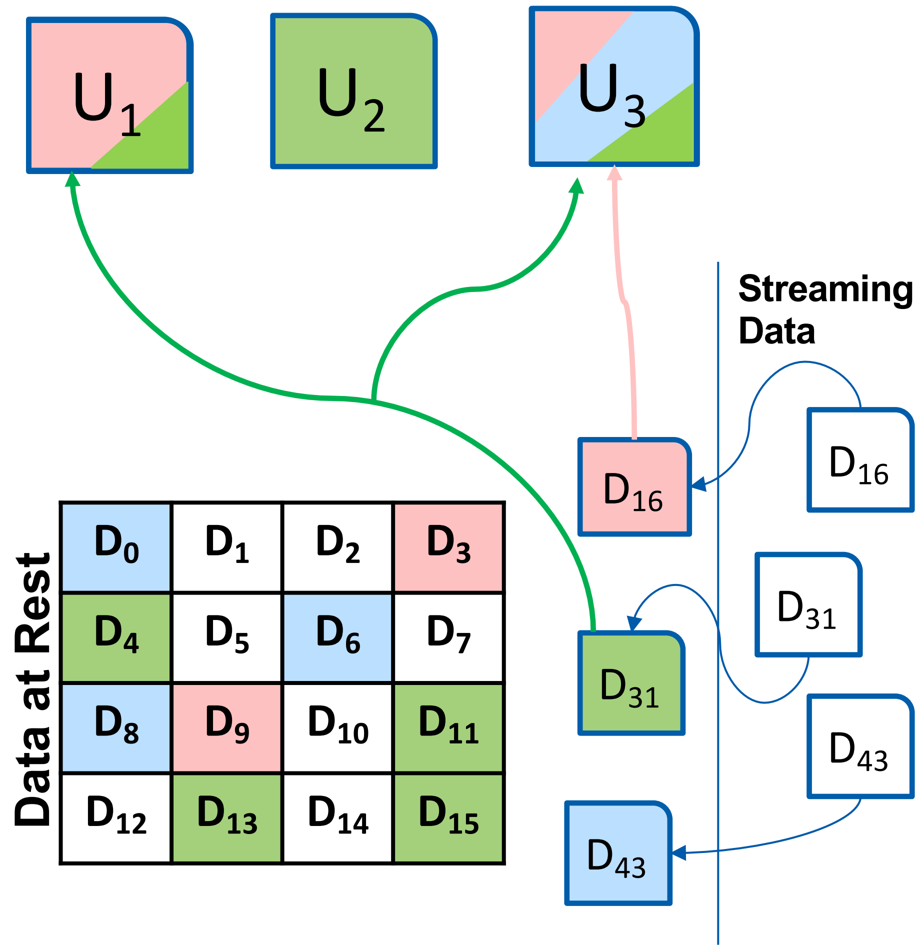
Multiple Data of Interest to Different Users

- Extract human-interpretable topics from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- Uses variational Bayes for inference, no need to re-train



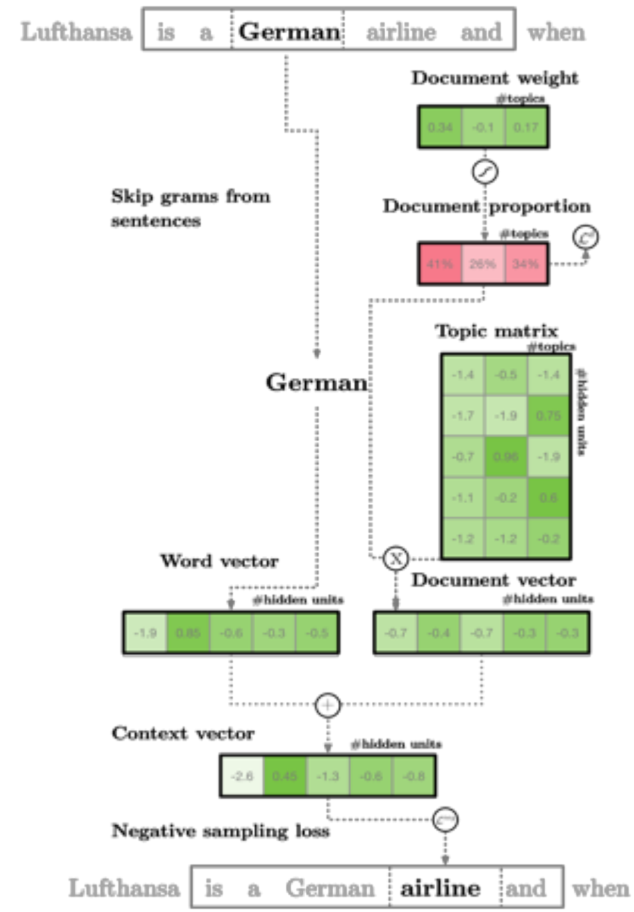
Multiple Data of Interest to Different Users

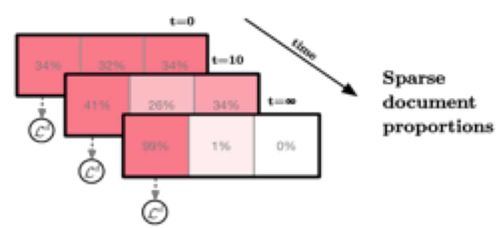
- Extract human-interpretable topics from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- Uses variational Bayes for inference, no need to re-train



Further Extension

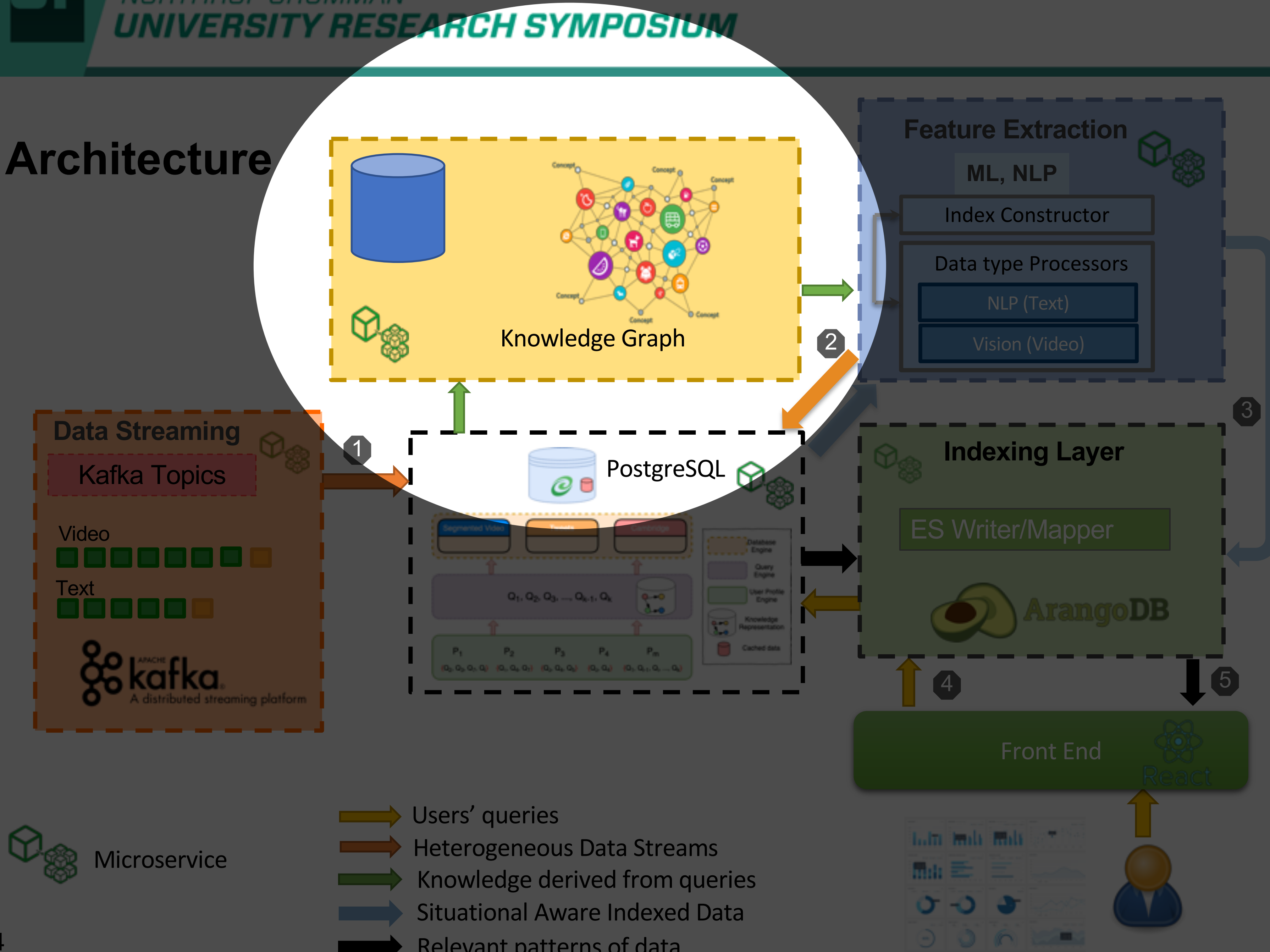
- Deep Learning model: Lda2Vec
- With lda2vec, leverages a context vector to make the predictions.
- Context : sum of the word vector and the document vector
- Context can be metadata in case of Twitter Data





https://multithreaded.stitchfix.com/blog/2016/05/27/lda2vec/

Architecture

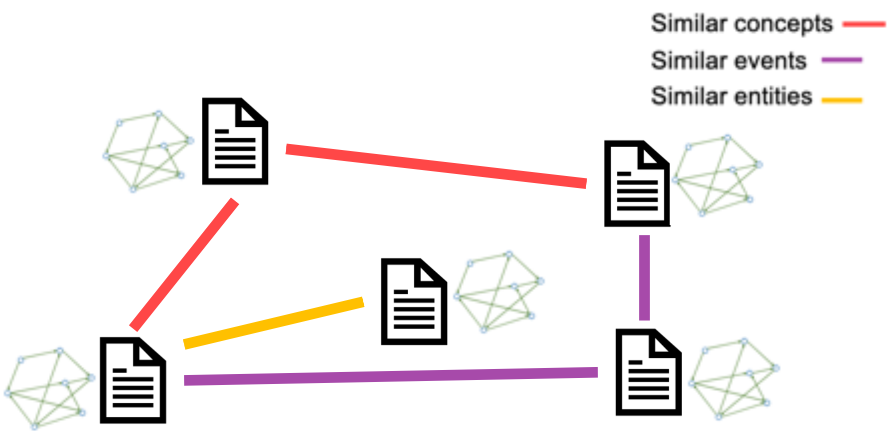


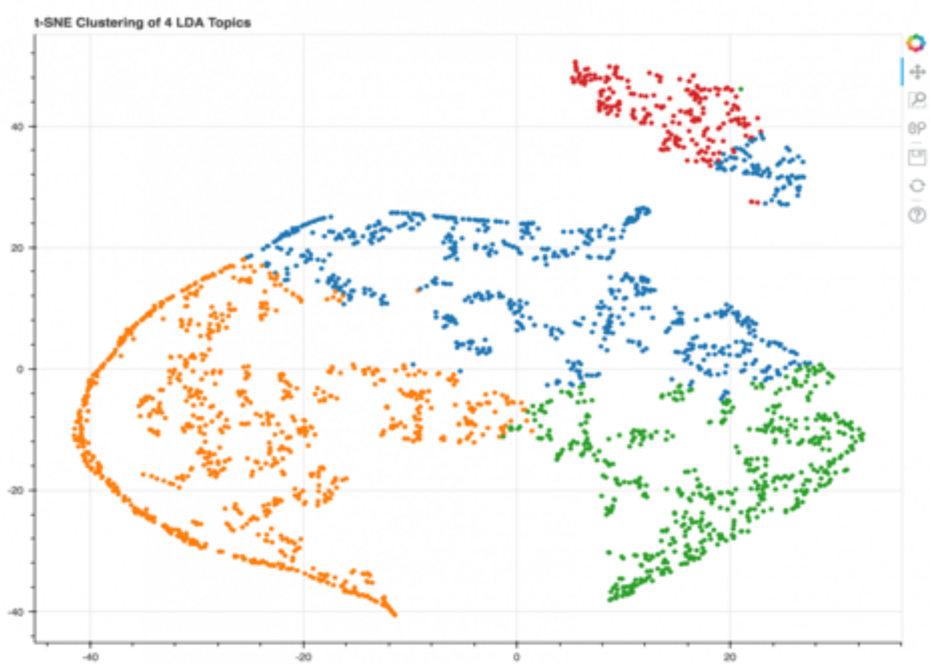
Knowledge Graph

- Ontologies / Concepts are extracted from LDA
- Extract Triplets <Subject, Relation, Object> to represent Events
- Entities are represented by Nodes
- Entities have Attributes (Labels)
- Entities are connected by Relations (Edges)

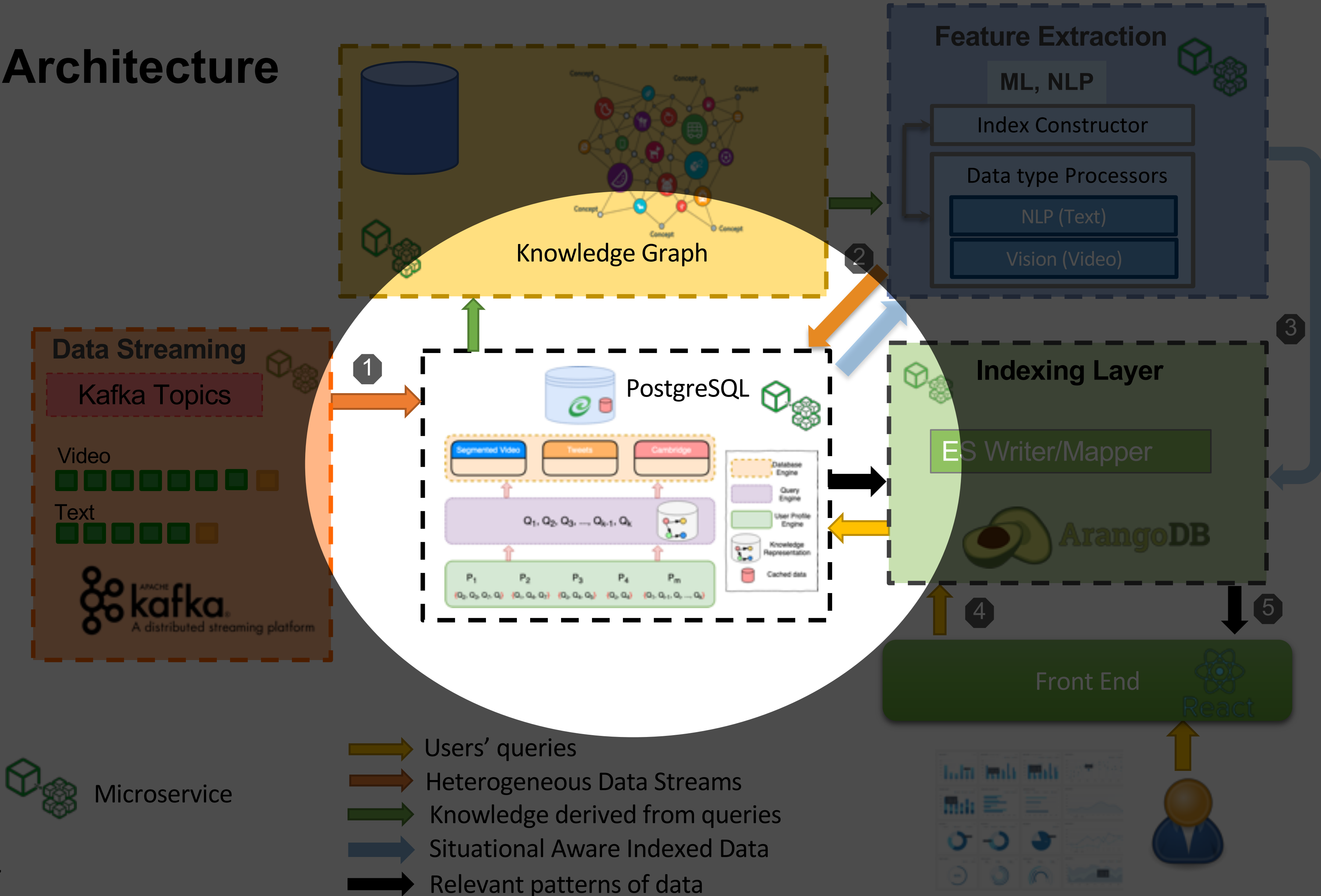
WIP with KG: Multi-modality

- Multi-modal Information Retrieval
- Poster represented In Northrop Grumman University Research Student Poster Competition





Architecture



User Modeling: Intention-aware Recommendation Engine

- Sends users streaming data that corresponds to their interests
- Builds User Profiles using the history of user queries
- Active Learning to narrow/expand intention model with more interaction
- Expands user queries with word embedding models to fetch relevant data from the database

User1

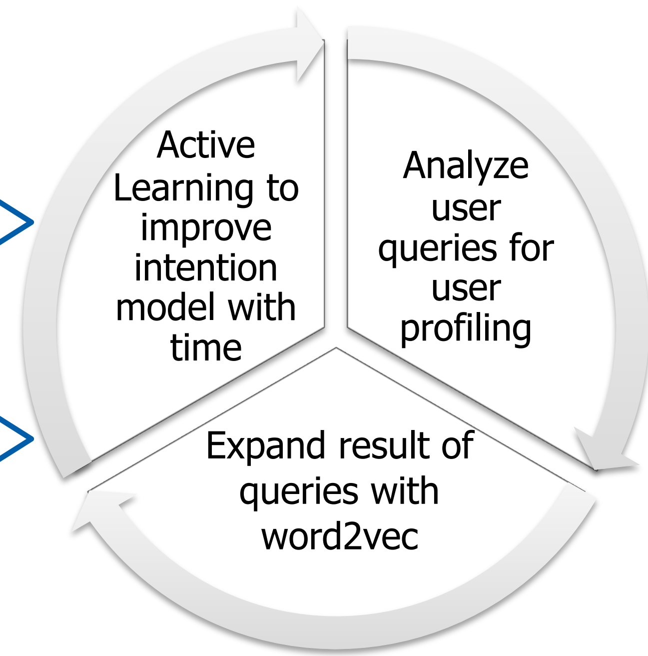
- Cars of specific make & model (purple Cadillac)
- Interested in info. about crimes in a specific district

SELECT * FROM video_data WHERE object = 'car' and attribute='purple'

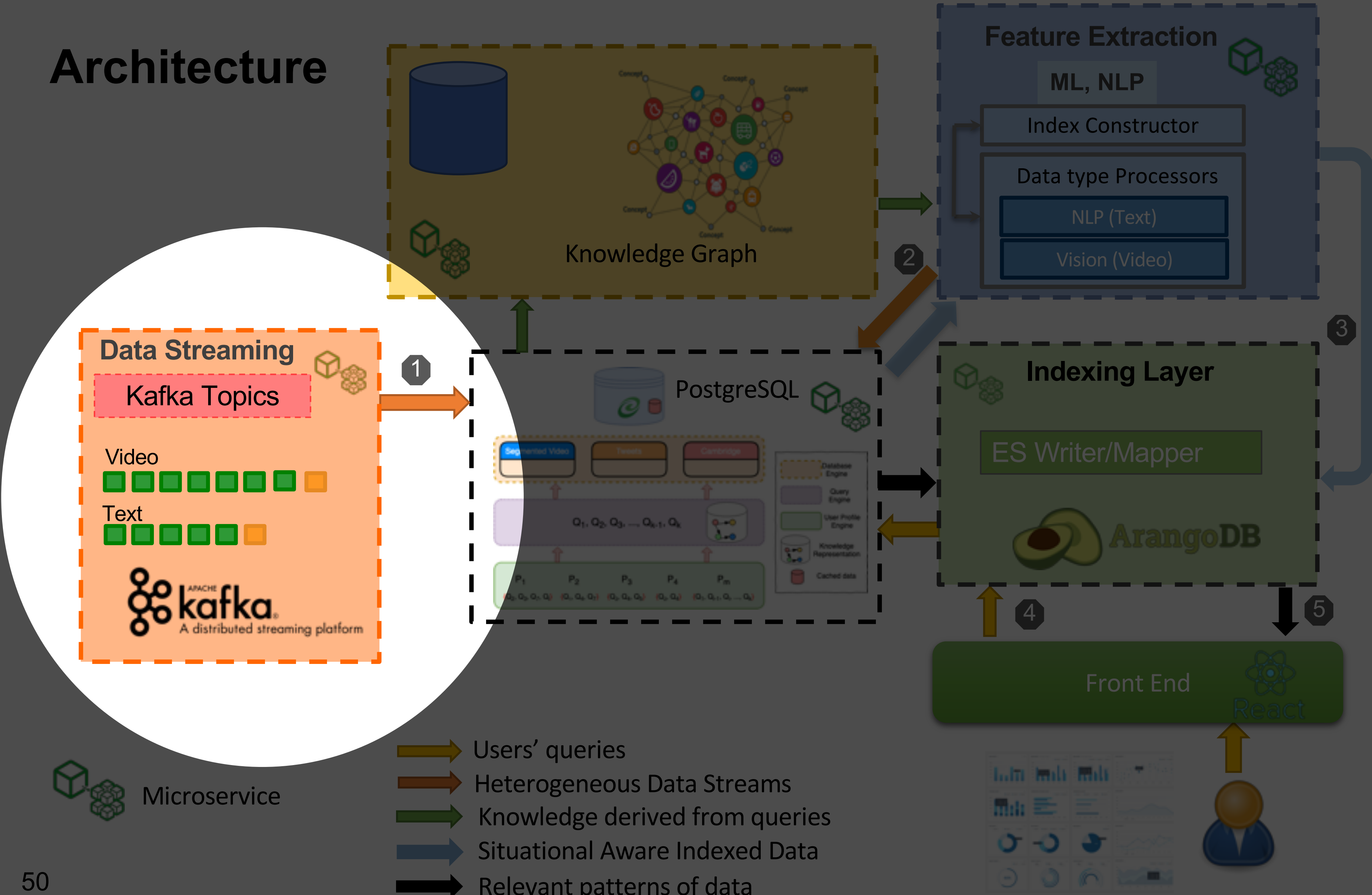
User2

SELECT * FROM crash_data WHERE date_hit = TODAY

- Looks for pedestrians in the video data
- Interested in traffic, accidents, violations



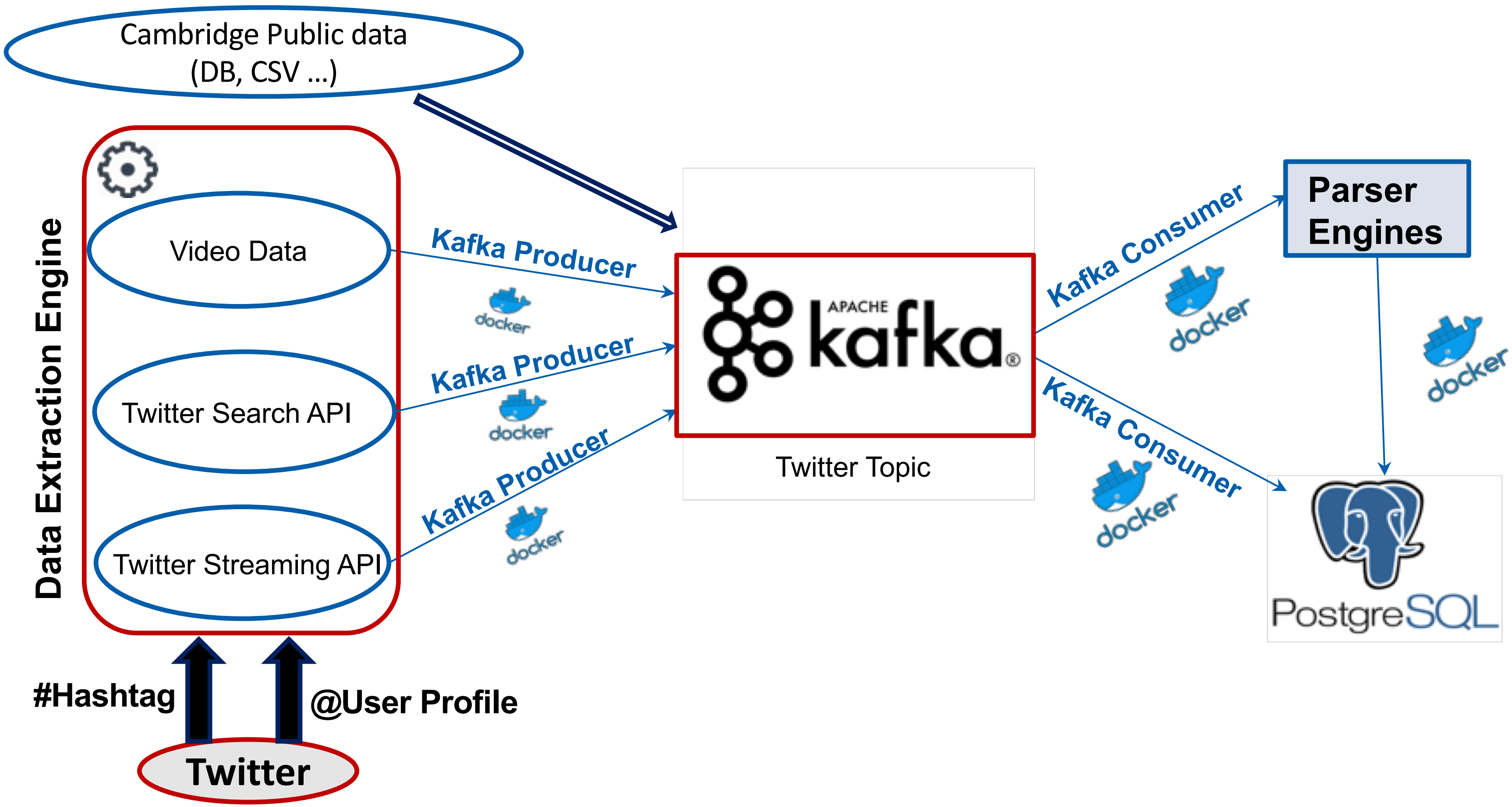
Architecture



Data Streaming Module

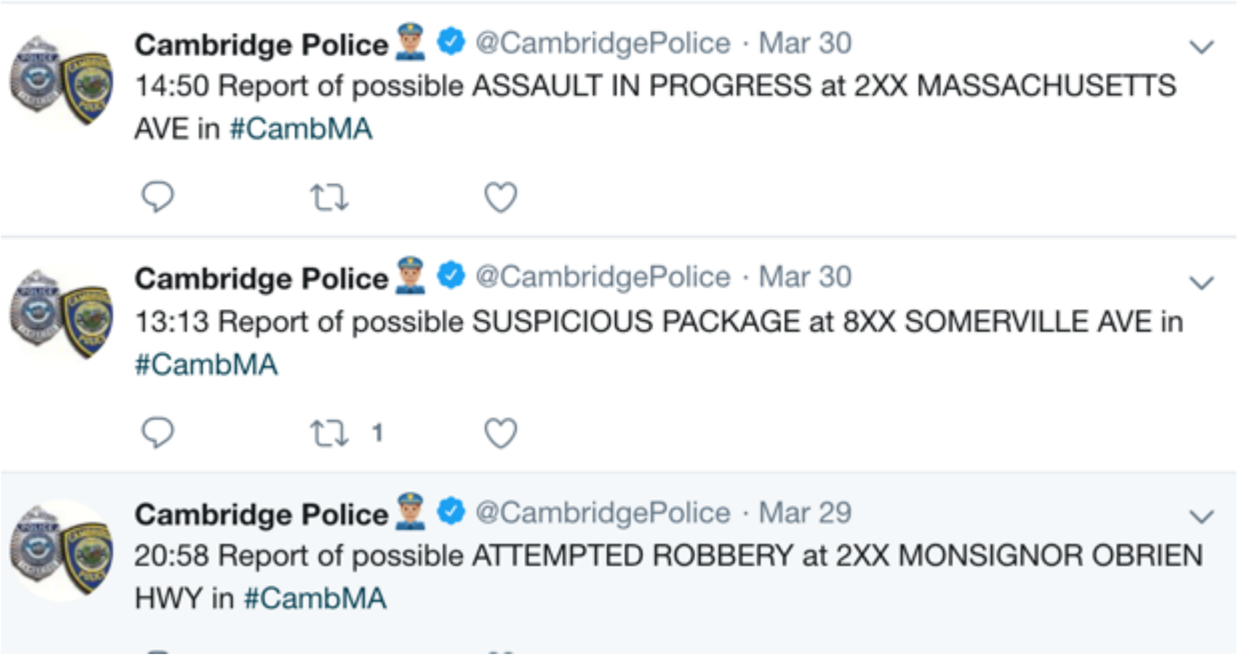
- Retrieve RESTFUL and Streaming Tweets
- Populate Postgres with all data
- *Parse collected metadata* to extract targeted information and *store in Postgres*
- Replicable, fault tolerant, scalable and continuous
- Build a Data Processing Pipeline with all features

Data Processing Pipeline



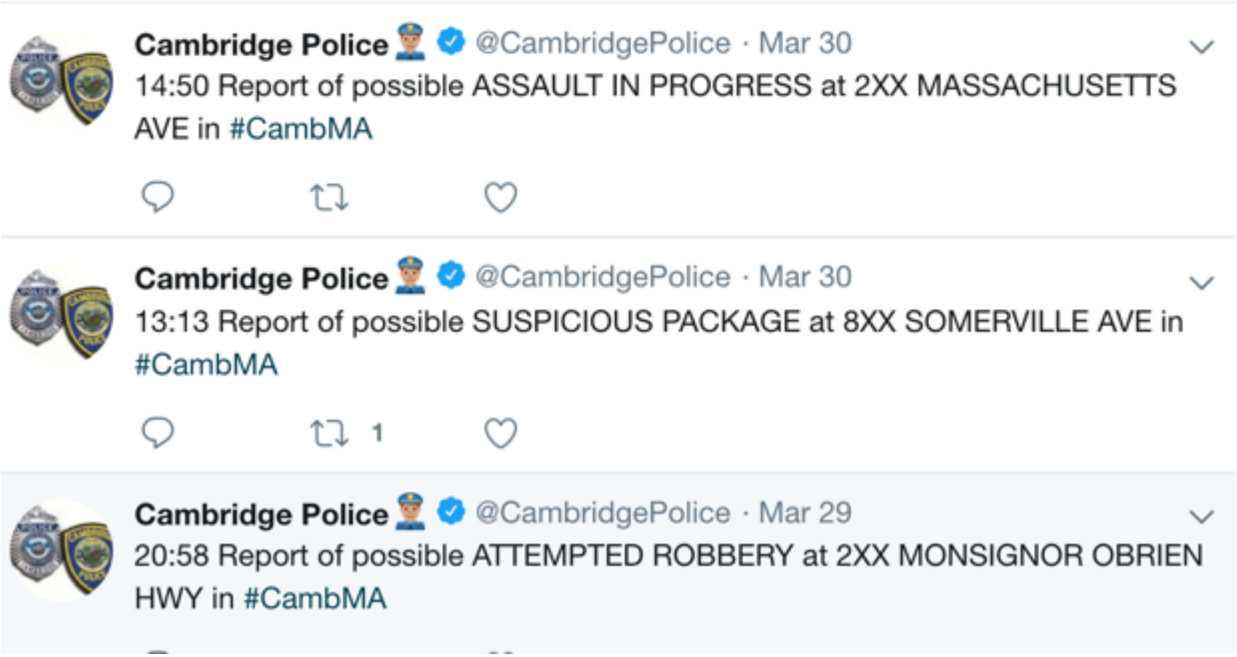
Retrieve Tweets : Implementation Choices

- Search tweets by
  - **Keyword / Hashtag** (i.e, CambMA)
  - **User Timeline** (i.e, CambridgePolice)



Retrieve Tweets : Implementation Choices

- Search tweets by
  - **Keyword / Hashtag** (i.e, CambMA)
  - **User Timeline** (i.e, CambridgePolice)



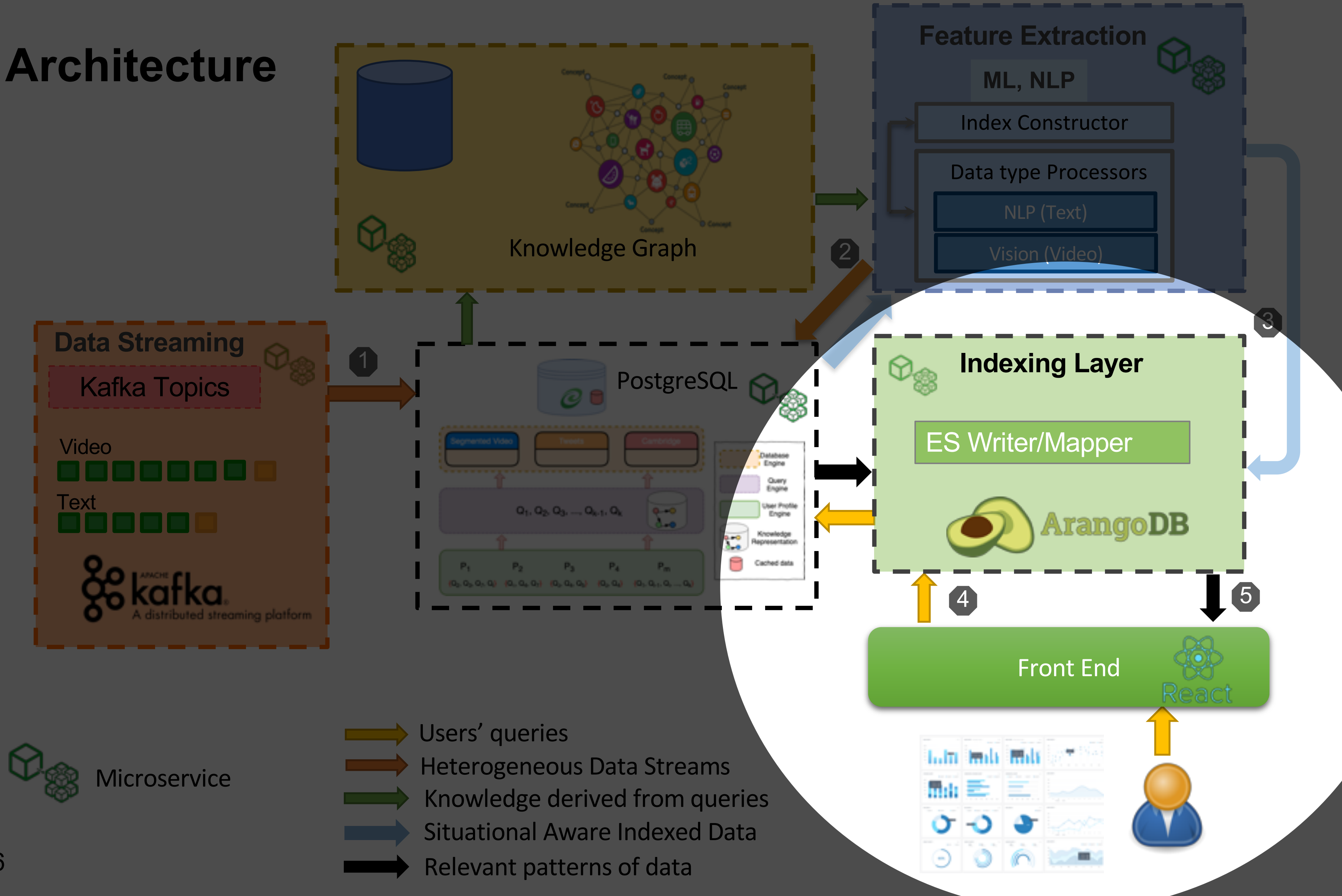
Compatibility with other sources of data

- Add new sources
  - JDBC
  - From file
  - Audio

- Kafka Connect provides a framework (extra layer between source and Kafka) to develop connectors importing data from various sources and exporting it to multiple targets

- Kafka Clients allow us to pass and retrieve messages directly to and from Kafka

Architecture

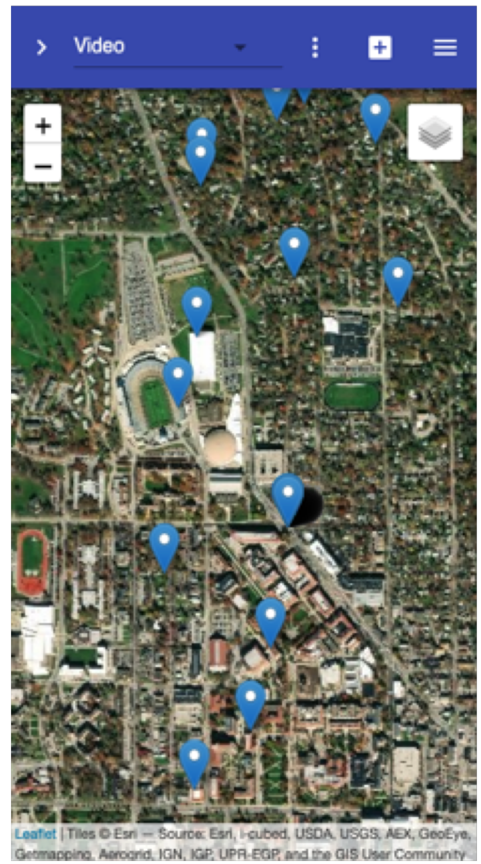


Representing Knowledge

- Build a tree for each index which point to the corresponding frames in Videos
  - Car, Person, Bicycle, Traffic light
- Build a tree for each index which point to the corresponding mentions in Tweets
  - Car, Person, Bicycle, Traffic light
- User Profiling: Built based on similar types of information
- Build triggers in Postgres
  - Data comes in with similar index
  - Deliver to User
- Model all our indices in GraphDB (ArangoDB)

SKOD Web Framework

- Extract data from Heterogeneous Sources and expose data via Apache Kafka Topics
- Consume data from Kafka Microservice and populate the RDBMS and the Index Layer (Elasticsearch and *Graph Database)*
- Utilizing geolocation to visualize real-time streams on Leaflet map
- Analyze data relationships through graph analytics (clustering)



kafka

Trellis

React

JS

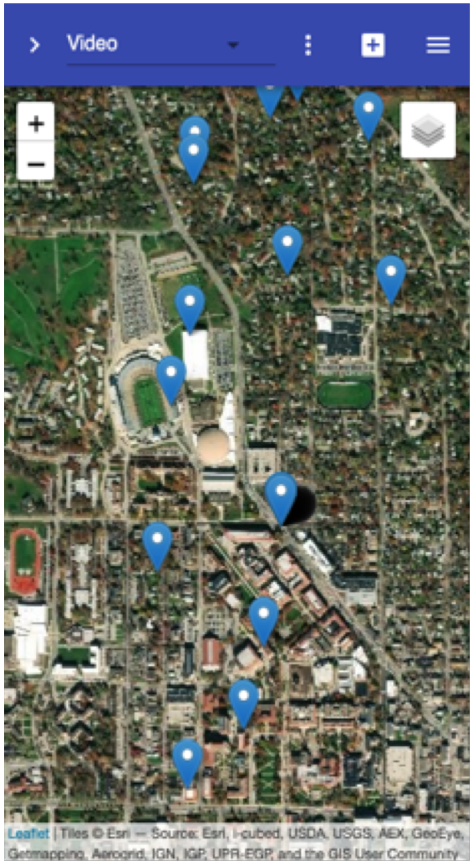
docker

JWT

We utilize the OADA/Trellis framework to build the PoC of the Web App.

SKOD Framework Features

- Open source @
- Distributed Compute Engine (Apache Spark GraphX) and Motif analysis
- ArangoDB Graph Database
- Multiple layers of Cache (PouchDB)@
- Eventual Consistent
- Easy to setup (using Docker containers)
- React based Analytics Web-UI



@ https://github.com/purdue-gask/skod/

@ https://github.com/OADA/oada-cache

Summary

- SKOD aims at delivering right information to the right user at the right time based on situational awareness
- There are numerous users with different missions
- Missions with various needs for information
- SKOD is an end-to-end system to empower such users with relevant knowledge from *streaming* or *stored* data
- SKOD is general purpose and can be specialized to NG use cases

https://www.cs.purdue.edu/news/articles/2019/bhargava-realm-ng.html

Deliverables

- Microservices for all modules
- Source Codes

https://github.com/purdue-gask
https://github.com/OADA

Demo Video

- Sequentially shows
  - How twitter data is consumed and processed via Data Streaming Module
  - Extracting objects from Videos
  - Extracts the tweets that discusses about *Object in Question*
  - Tie features from different modality using the Indexing Layer
    - Build Index on the objects from videos and tweets
  - Functionality of the Front End with Graph Analytics
  - User Profiling extracts other objects that can be of users' interest
  - Allows user to see the those objects from all modalities

Demo Video

- Simplified Query

```
Select * from tweets, videos where
tweets.objects_discussed == "car"
videos.objects_detected == "car"
```

NORTHROP GRUMMAN
UNIVERSITY RESEARCH SYMPOSIUM

Research in Applications for Learning Machines (REALM) Consortium

Situational Knowledge On Demand

SKOD

Future Plans for SKOD : Feature Identification

❖ Feature Identification from Video

- Pedestrians, Occluded traffic signs, Crane blocking a sidewalk, Child left in unattended car outside school
- Offline model construction (based on video and open street map)
- On-line execution

❖ Feature Identification from Text

- Interesting subset identification based on keywords
- Parse to an entity-attribute model of interesting info

More SKOD Benefit Scenarios

- Inform Drivers about
  - relevant obstacles and hazards: road closures, potholes, fallen trees and tree branches, ice, dumpster violations, downed road signs, not working traffic lights;
  - routes to avoid obstacles and hazards;
  - relevant POIs;
  - collision probability for a given date, time, weather conditions; recommend the speed.
- Inform blind / differently abled people via a mobile app about:
  - relevant obstacles and hazards;
  - routes to avoid obstacles and hazards;
  - relevant POIs.

More SKOD Benefit Scenarios

- Inform Law Enforcement about
  - suspicious activity (especially in crime-prevalent areas), illegal road constructions, downed road signs, blocked sidewalks, graffiti;
  - relevant obstacles and hazards;
  - routes to avoid obstacles and hazards;
  - collision probability for a given date, time, weather conditions; recommend the speed;
  - detected human faces in crime incidents and car accidents;
  - homeless people detected in certain areas.

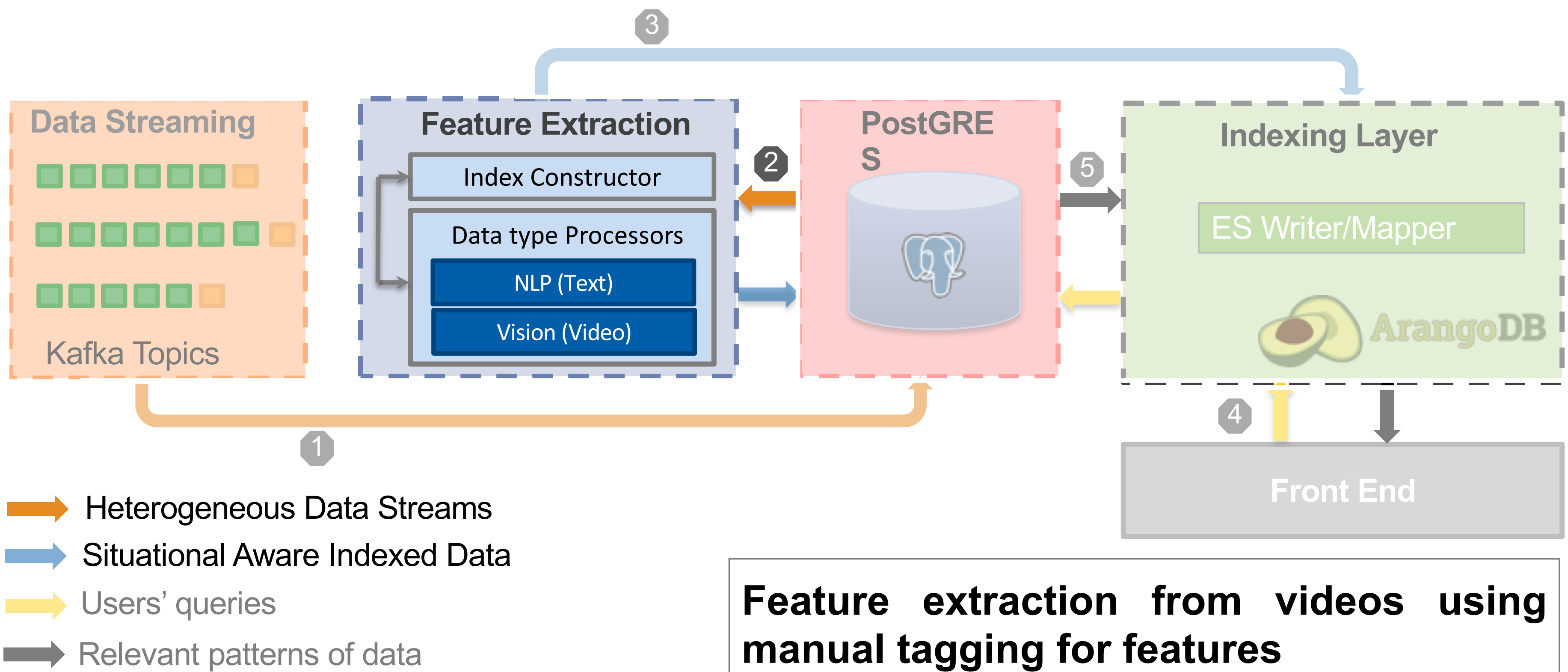

NORTHROP GRUMMAN
REALM
Research in Applications for Learning Machines

Backup Slides

Tweets-Parser-Engine

- Parses metadata to extract
  - Full tweet text
  - User Information
  - Hashtags, URLs, User mentions
  - Geolocation (latitude, longitude)
- Separates and processes
  - Original tweets
  - Retweets
  - Quoted tweets

Feature Extraction Module



Feature extraction from videos using manual tagging for features

Manual Feature Extraction from Videos

- Features targeted
  - Objects in Video
  - Attributes of the objects
- Amazon Mechanical Turk (Mturk)
  - For task design
  - For annotation collection
  - For task distribution
- Steps
  - Run Object detection algorithms
  - Segment video into frames
  - Modify the existing annotations

ksolaima@purdue.edu

Task Design Sample: Instance Segmentation

Instructions

View full instructions

View tool guide

Use the tools to label each instance of the requested items in the image

Color in each instance of the requested items in the image

Labels

Choose a class below to add its instance(s).

- Car
- Fire Hydrant
  - Fire Hydrant #1 1
  - Add instance
- Turn signals

Polygon | Brush | Eraser | Dimmer | Undo | Redo | Zoom in | Zoom out | Move | Fit image

Nothing to label

Submit

ksolaima@purdue.edu

Task Design Sample: Attribute Tagging

Instructions: Given a frame, describe the attributes of the marked object in the bounding box.

Attributes can include number plate, color of car, street name that can be used to describe the object.

Word/phrase 1
Number plate/SWW-14W

Word/phrase 2