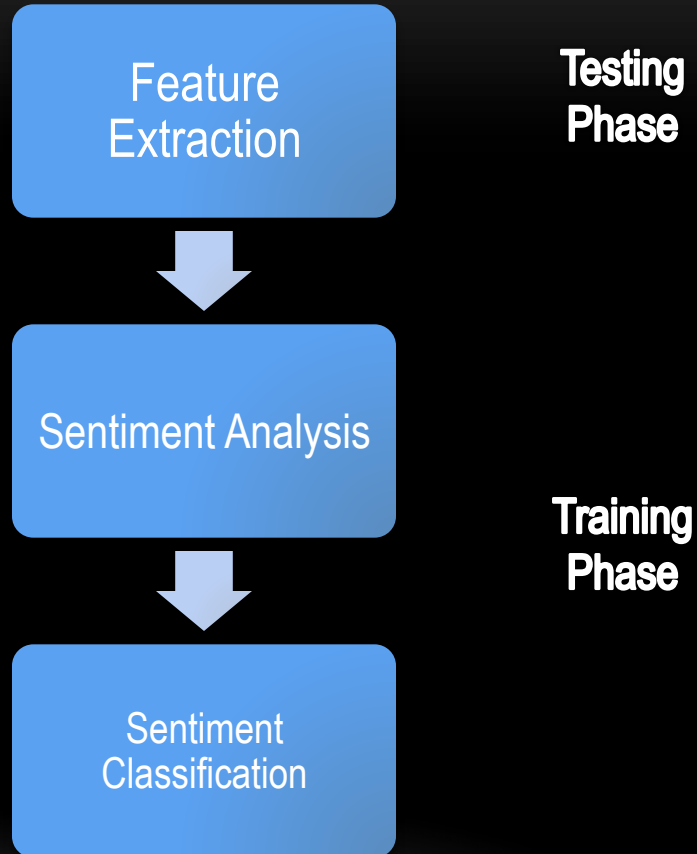# REVIEW SUMMARY SYSTEM

Team – 22

M Manoj Kumar – Srinath Ravichandran - Dharmesh kakadia – Sandhya S

(201107502)   -   *(201107625)*   -   *(201107616)*   -   *(201107617)*
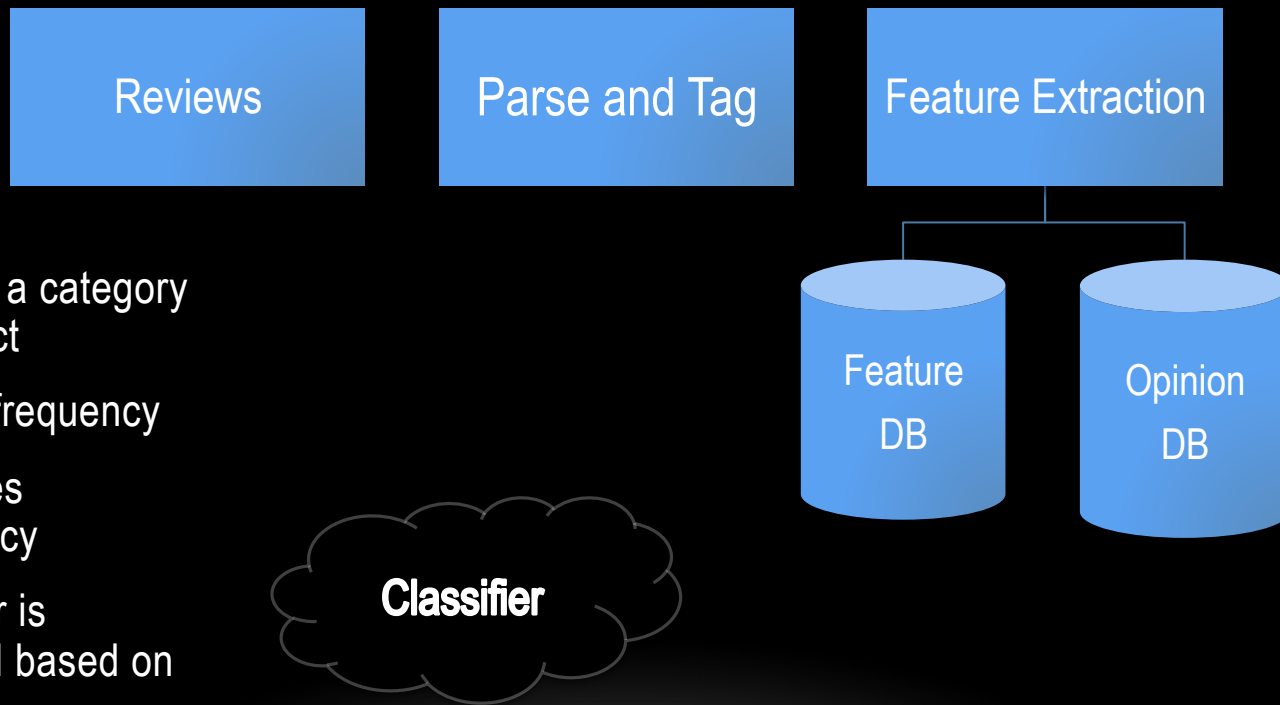
OVERVIEW

- System to summarize reviews from various sources

- Users can view and compare products based on features

- Results exposed as RESTful web-service

- Ability to cater to different products

OVERALL WORK FLOW

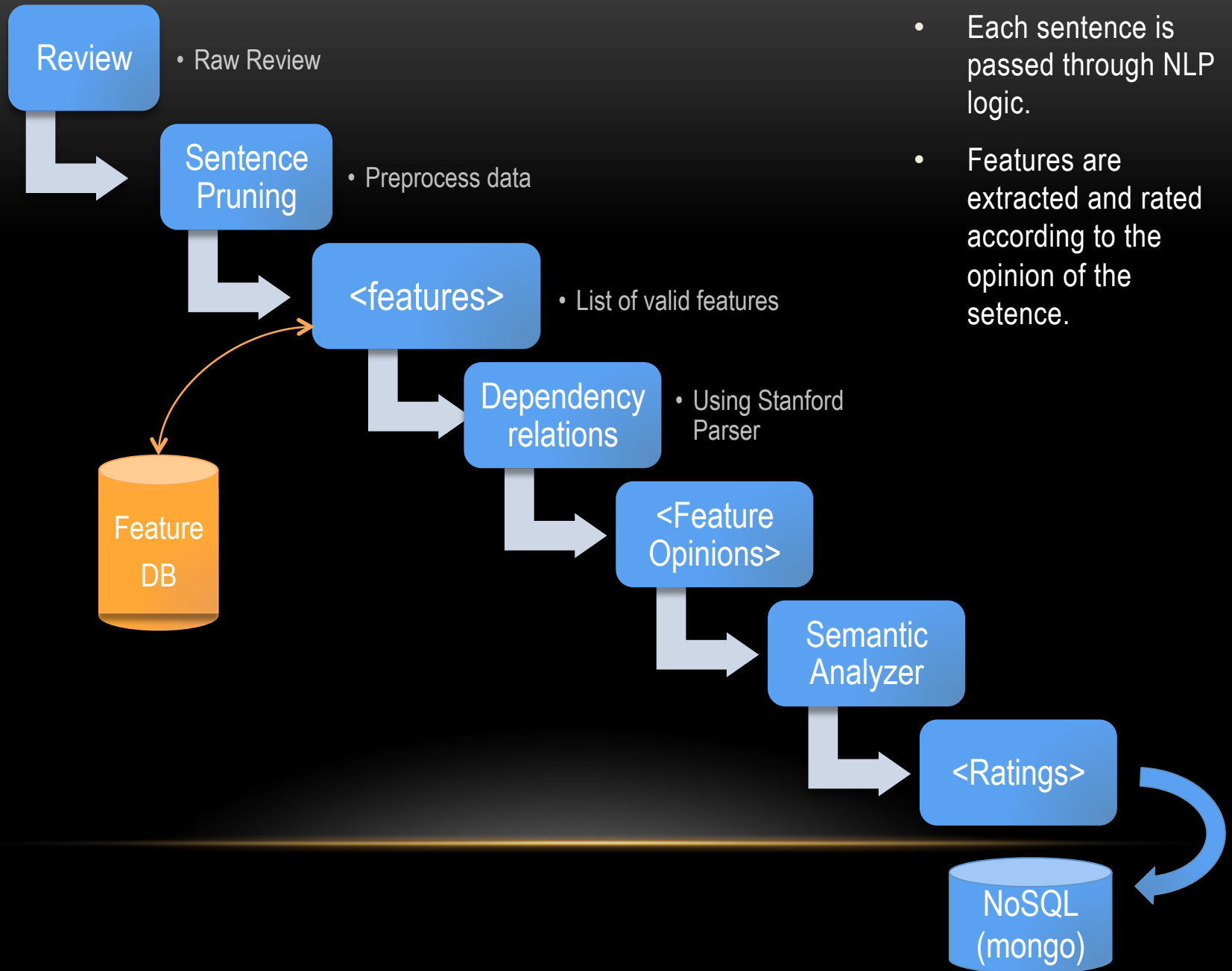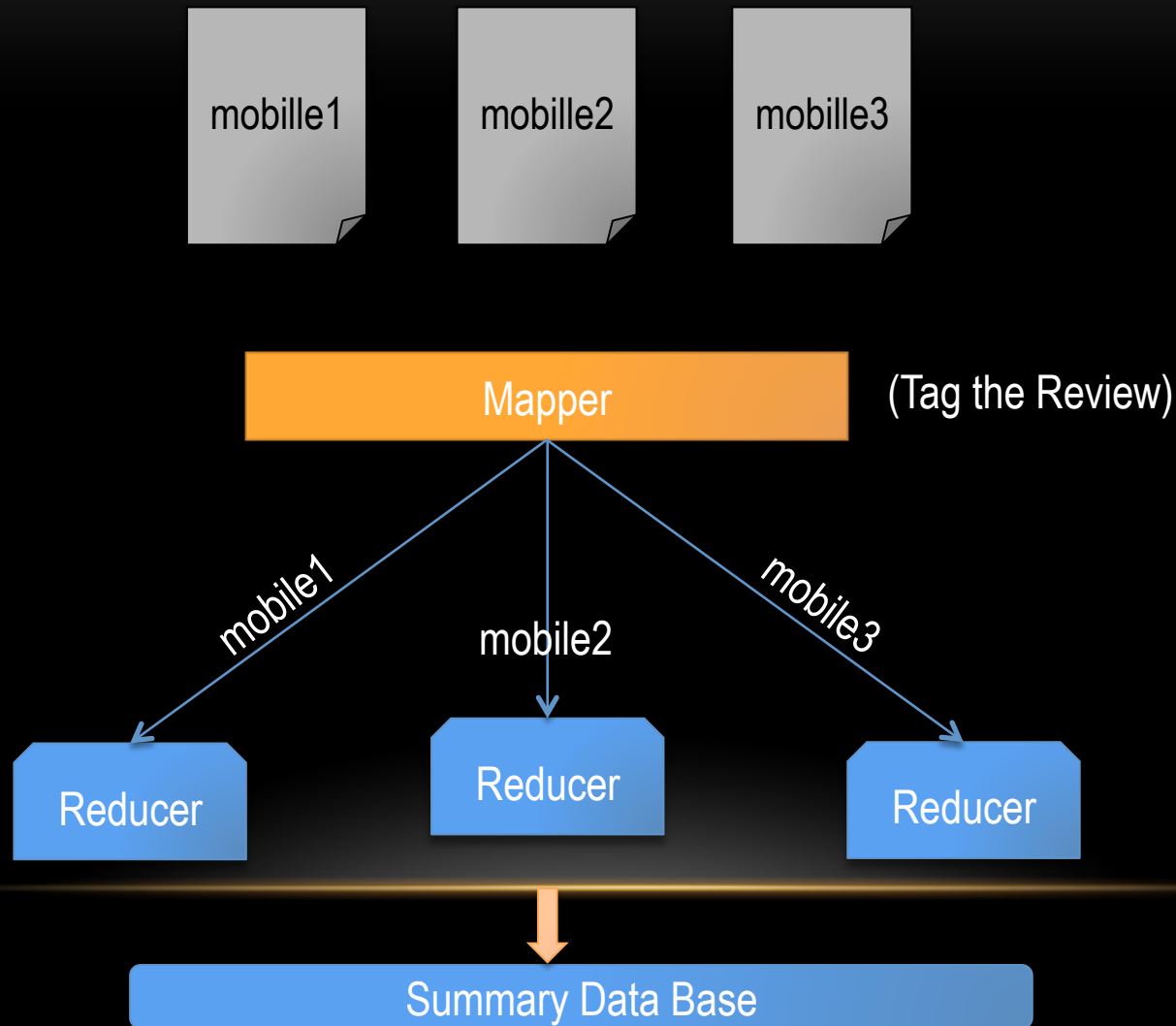# DETAILED FLOW CHART

-- Training Phase --

| Reviews | Parse and Tag | Feature Extraction |

- Once for a category of product
- Nouns #frequency
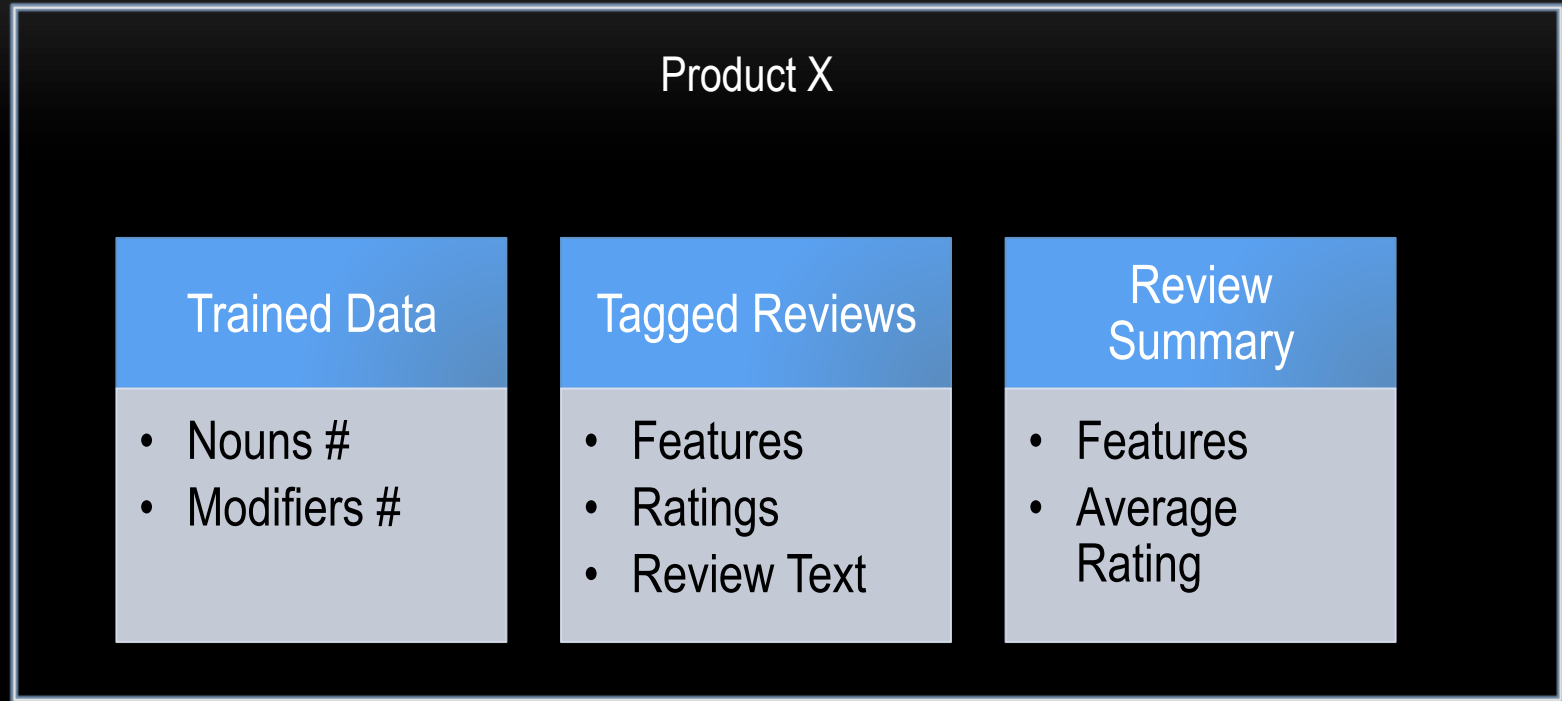- Adjectives #frequency
- Classifier is designed based on this data.

Classifier

Feature DB

Opinion DB

– Tagging and Summarizing –

```
Review                  • Raw Review

        Sentence
        Pruning         • Preprocess data

                <features>      • List of valid features

        Feature             Dependency
        DB                  relations       • Using Stanford
                                              Parser

                                    <Feature
                                    Opinions>

                                        Semantic
                                        Analyzer

                                            <Ratings>

                                    NoSQL
                                    (mongo)
```

- Each sentence is passed through NLP logic.

- Features are extracted and rated according to the opinion of the setence.

PARALLELIZING WITH HADOOP

DATABASE SCHEMA

Product X

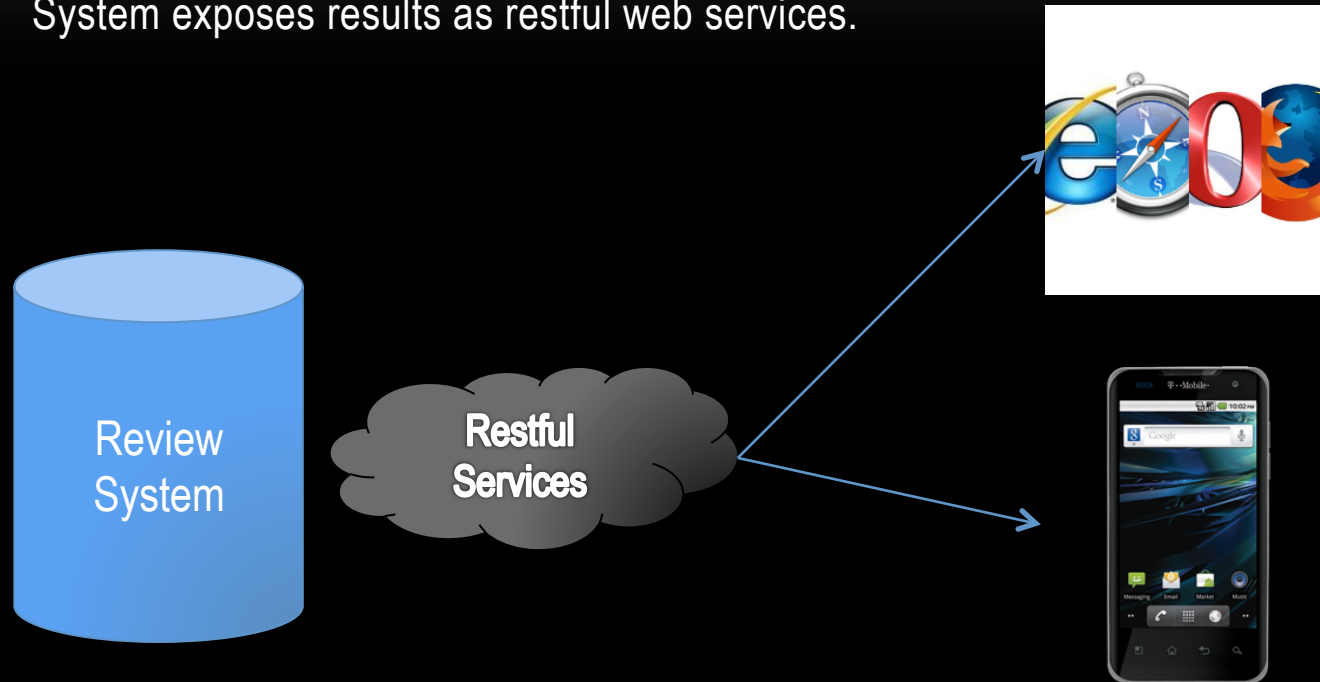| Trained Data	Tagged Reviews	Review Summary
• Nouns # • Modifiers #	• Features • Ratings • Review Text	• Features • Average Rating

RESTFUL WEB SERVICES

- System exposes results as restful web services.

EXPERIENCES & LEARNING

- NLP Dependency Relationships!!!

- REST is BEST

- SCHEMA defines EVERYTHING!!

FUTURE WORK

- Better feature Extraction.

- Synonym match can be extended with Wordnet::Similarity.

- Can be further optimized for blazing performance.

- Preprocess user query.

TOOLS USED

- NLP
  - Stanford Parser
  - Wordnet (Synonyms)
  - Sentiwordnet
- Hadoop 20.2
- Mongo DB