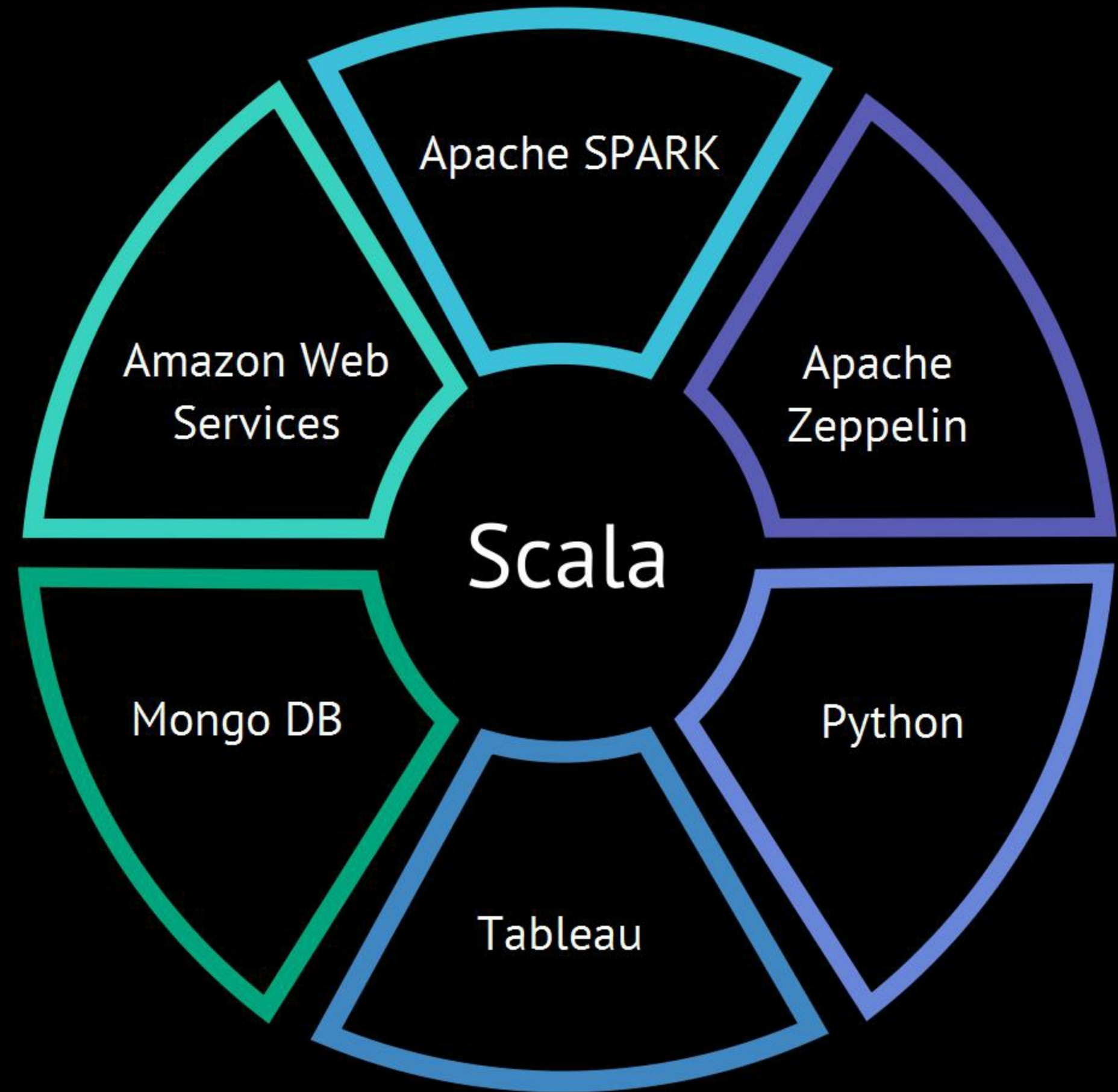


CRITIQUE MINING (E-COMMERCE REVIEWS)



Bala Gopalakrishnan

Malika Thapa

Rachan Hegde

AGENDA

01

Data Set

02

Goals

03

Scala Components

04

Big Data Analytics

05

Future Scope

DATA SET

Amazon Product Reviews

- Information about the products and reviews are in JSON format
- The volume of data set under consideration is ~10GB (> 10 Million Records)
- The details are grouped into the following data sets
 - Product Metadata (asin, product title, brand ...)
 - Product Reviews (asin, review, rating ...)

GOALS

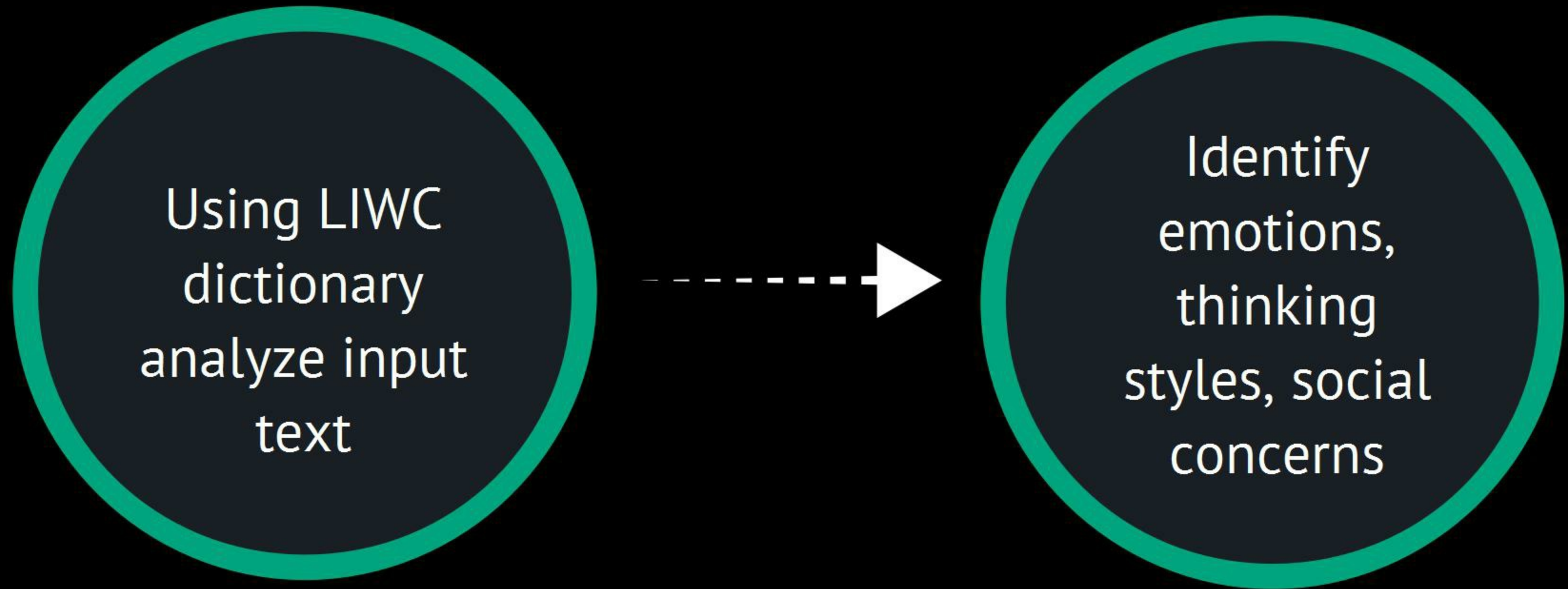
IT

- Use NoSQL to store product reviews (JSON)
- Create a LIWC dictionary
- Compare reviews with LIWC dictionary to generate features set for ML
- Predict the actual rating using Spark ML
- Parallelize using SPARK

Business

- Find products with better customer satisfaction
- Neglect Fake Reviews
- Dashboards for informed decisions

LIWC (LINGUISTIC INQUIRY AND WORD COUNT)



SCALA COMPONENTS

01

NoSQL Data Loader

Configurable data loader (Import, Update, Upsert) for MongoDB

02

Search Index Generator

Converts LIWC text dictionary set into TRIE to improve search performance

03

LIWC Feature Generator

Enterprise will continue to exist in a foreseeable future.

04

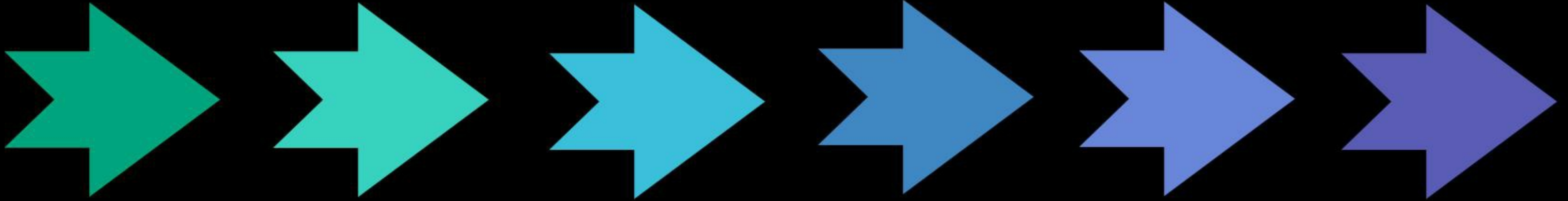
ML Pipelines

Expenditure which brings into existence asset or benefit of a long term nature.

BIG DATA ANALYTICS



Big Data Analytics



Load Data into Mongo DB	Analyze data for inconsistadncie s	Create LIWC Search Index	Using LIWC search index, create feature set for each review	Apply spark ML using feature set generated	Update Mongo DB with features and predicted rating
Scala		Scala		Scala	
Mongo DB	Mongo Shell	JSON Util		SPARK ML	
Casbah	Mongo Compass		Scala	Regression	Scala
EmbedMongo	Tableau		JSON Util	PCA Pipeline	Casbah
				Cross Validator	EmbedMongo
				Metrics	
				Zeppelin	



Future Scope

**Aggregated Product
Sentiments**

Across
Dimensions

**Product
Recommender**

Associative
Rule Mining

**Seamless Integration
between Modules**

Integration is
currently manual

Thank you