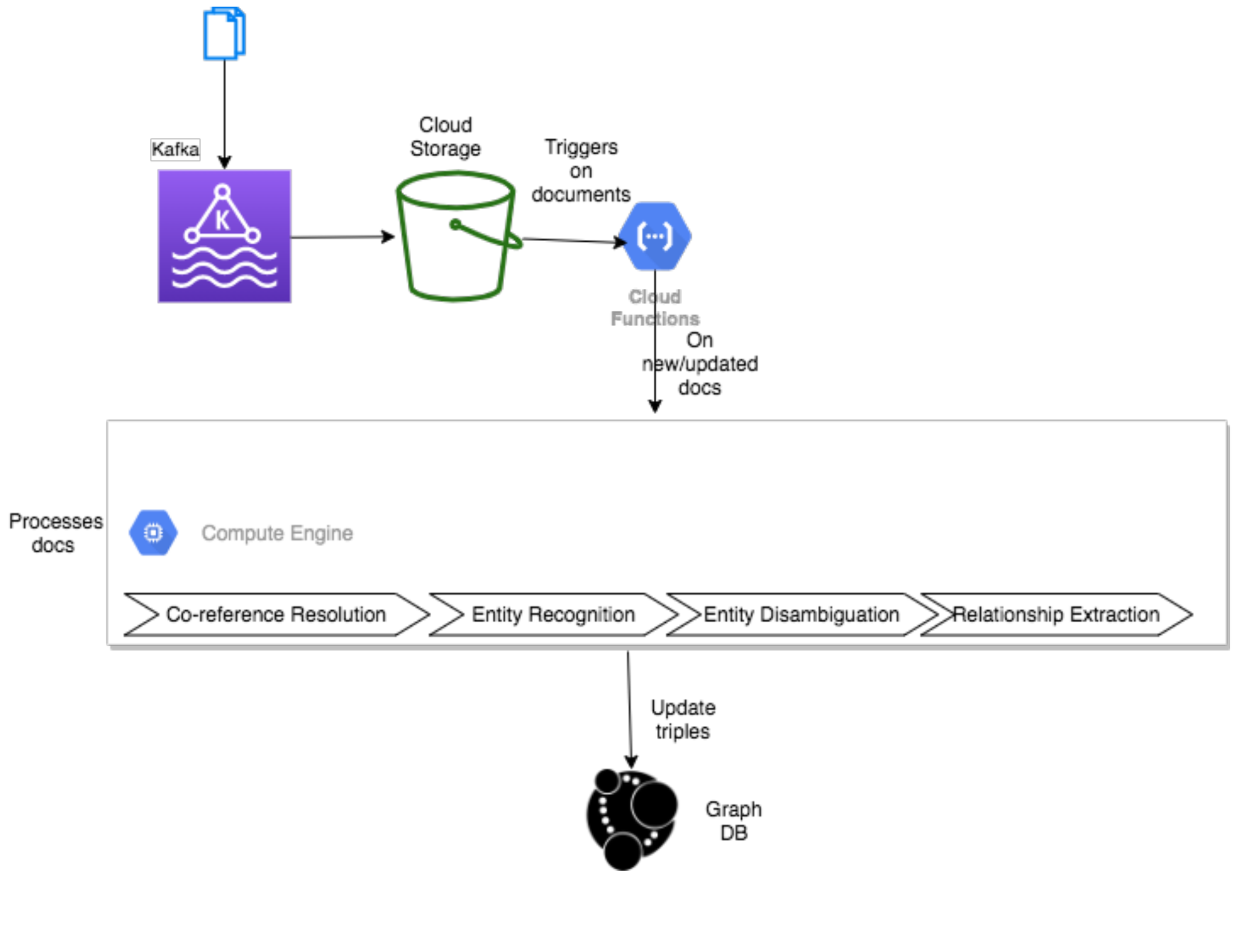
Kafka
Cloud
Storage
Triggers
on
documents
Cloud
Functions
On
new/updated
docs
Processes
docs
Compute Engine
Co-reference Resolution
Entity Recognition
Entity Disambiguation
Relationship Extraction
Update
triples
Graph
DB

Knowledge Graphs

- Knowledge graphs are used to search, store and present fact-based data and are also used to power search engines, recommendations and chatbots.
- Knowledge graphs contain triples of 2 entities and a relationship between them: subject, relation and object.
- Process for generating a knowledge graph -

1. Coreference Resolution - Task of finding all expressions that refer to the same entity in a text, like personal pronouns, possessive pronouns etc.
2. Extract Entity - Extract entities like Person, Organisation etc using systems like Spacy NER, Stanford NER or through Pretrained models like BERT. Entity recognition has its own limitations like different forms of the same name like Mahatma Gandhi, Gandhi ji.
3. Entity Disambiguation - Mapping the entity forms to a unique entity present in a target knowledge base.
4. Relationship extraction - Extract relations between entities, to finally form triples. This can be done using rule based extractors - Spacy or NLP based approaches

Triples are stored in a graph database like Neo4j, ArangoDB, FlockDB etc. Querying the Graph DB, gets the relationships of different levels, like son, grandfather, CEO etc.