



# TIPOLOGIE

- Progetto relativo alle tecnologie di data warehouse
  - Progetto relativo alle tecnologie di big data processing
  - Progetto relativo alle tecnologie di OLAP serveless
  - Progetto relativo alle tecnologie di LLM applicate ai dati
- 
- In tutti i casi
  - Si possono formare gruppi da uno a tre studenti
  - PRIMA di procedere allo svolgimento del progetto il gruppo deve comunicare al docente il dominio e le tipologie di analisi da effettuare
  - Di seguito si riportano alcune proposte, sono anche accettate altre proposte relative agli argomenti del corso



TRACCIA DW

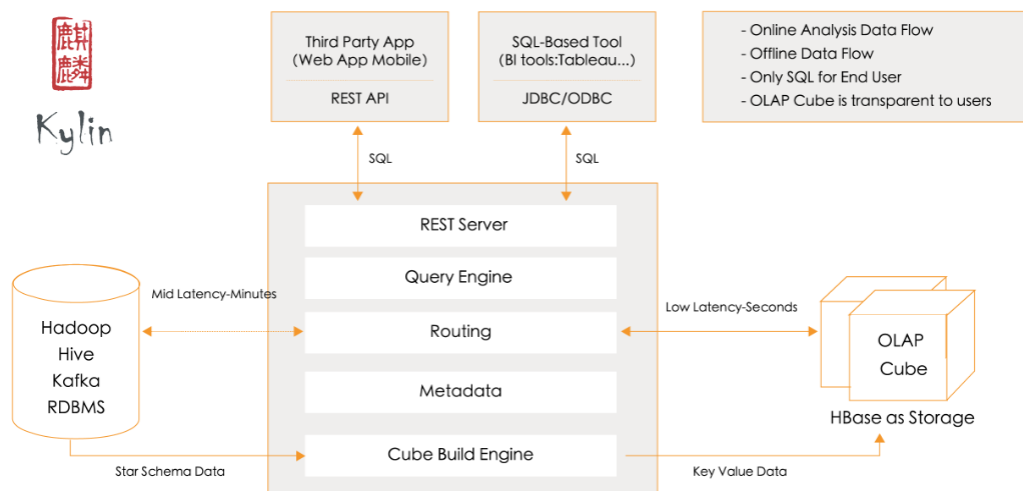


# TRACCIA

- Sviluppare un datawarehouse
  - con tecnologia SAP HANA
- Sviluppare un DW su sistemi noSQL
- Altri progetti
- HANA È un datawarehouse della SAP di tipo colonnare in memory
  - Molto usato nelle grandi aziende che usano SAP

# DATA WAREHOUSE SU NOSQL/OBJECT STORE

- Studiare strumenti esistenti o adattabili per costruire un data warehouse memorizzando i dati in formato nosql (documentale)



- One Big Table Technology

# ALTRI PROGETTI

- Costruire un dw con tecnologie tradizionali
- Una volta scelta la traccia è necessario
  - Identificare sorgenti dati disponibili. Per ciascuna sorgente
  - Valutazione della qualità
  - Tecniche da adottare per ETL
- Identificare (o dedurre) schemi locali delle sorgenti e mostrare il processo che porta alla realizzazione dello schema riconciliato.
- Proporre un glossario dei requisiti ed una stima del carico di lavoro preliminare.
- Definire lo schema di fatto relativo ad almeno un fatto riportato nel glossario dei requisiti. Utilizzare, se opportuno, la procedura traduci e tecniche di editing sul risultante albero degli attributi.
- Definire e realizzare lo schema a stella (o lo schema snowflakes) relativo allo schema di fatto creato.
  - Delineare una strategia per la gestione delle dimensioni dinamiche
- Effettuare l'ETL (con strumenti di mercato o con soluzioni custom).
- Creare interactive reports o dashboards, utilizzando uno strumento esistente incentrati sulle interrogazioni incluse nella stima del carico di lavoro preliminare.

# CONSEGNA E DISCUSSIONE TRACCIA DW

- Consegna di una tesina che documenti l'attività di analisi e progettazione svolta. In particolare:
  - Scenario ed analisi delle sorgenti
  - Procedura utilizzata per la definizione dello schema di fatto
  - Realizzazione della base di dati
  - Procedura utilizzata per ETL ed eventuali wrapper sviluppati
  - Realizzazione di un report/dashboard
- All'orale discussione della tesina e demo live della creazione di un report/dashboard.



# TRACCIA BIG DATA PROCESSING





# TRACCE PROPOSTE

- Gestione dati in grafi di grandi dimensioni
  - Datawave, janusgraph...
- Raccolta e analisi in tempo reale
- Analisi dei formati tabulari per data lakehouse

# PROPOSTE

- Analizzare piattaforma per la gestione di KG di grandi dimensioni
  - Linguaggio di modellazione/interrogazione
  - Gestione della scalabilità
  - Gestione della sicurezza
- Costruzione di un KG a partire da fonti relazionali
- Costruzione di un KG a partire da fonti documentati
- Descrivere uno scenario nel quale emerga la necessità di attività di big data analysis
  - Calcolo distribuito usando SPARK
  - Costruzione cruscotti KIBANA
  - Definire il processo di ingestion dei dati
  - Definire il processo di integrazione/arricchimento/controllo qualità dei dati
- Valutare le performance dei sistemi tabulari
  - Iceberg, hudi, delta lake

# CONSEGNA E DISCUSSIONE TRACCIA BDP

- Consegna di una tesina che documenti l'attività di analisi e progettazione svolta. In particolare:
  - Scenario ed analisi delle sorgenti
  - Procedura utilizzata l'ingestion
  - Procedura per la manipolazione e memorizzazione dei dati
  - Realizzazione di un analisi/report/dashboard
- All'orale discussione della tesina e demo live della creazione di un analisi/report/dashboard.



# TRACCIA LLM & DATI



# TRACCIA

- Architettura RAG e tecniche di prompting per documentazione non finanziaria e tassonomia ESG
- Modelli fine tuning mondo finance
- Costruzione di benchmark per valutazione performance

# TRACCIA LLM

- Architettura RAG
  - tools
    - Scegliere il LLM da usare
    - Scegliere il vector DB
    - Scegliere il framework di integrazione
    - Usare anche un KG applicato ai KG
  - Prompting
    - Provare varie tecniche di prompting
  - Costruzione di modelli fine tuned
  - Costruire il dataset
    - Anche con arricchimento
  - Scegliere tecniche di fine tuning (Lora, Qlora)
  - Prompting
  - Valutazione rispetto ad una base line costruita dal modello non trainato
- 
- In entrambi i casi il dominio di riferimento è quello dell'analisi della documentazione non finanziaria con particolare attenzione alla tassonomia ESG

# TRACCIA LLM

- Costruzione di un benchmark per la valutazione delle diverse scelte architetturali di un LLM applicato in ambito finanziario

# CONSEGNA E DISCUSSIONE TRACCIA LLM

- Consegna di una tesina che documenti l'attività di analisi e progettazione svolta. In particolare:
  - Descrizione dell'architettura
  - Descrizione del dataset utilizzato
  - Descrizione del LLM usato
  - Tecniche di prompting e test
- All'orale discussione della tesina e demo live





E INFINE





SE AVETE ALTRE IDEE



Ogni proposta può  
essere estesa a Tesi

