Data-Driven Architecture: A Comprehensive Overview

Luis Daniel Benavides Navarro

17-11-2023

Motivation

- In today's data-centric era, businesses and applications require:
  - Sophisticated and specialized storage solutions.
  - Rapid innovation
  - Value generation

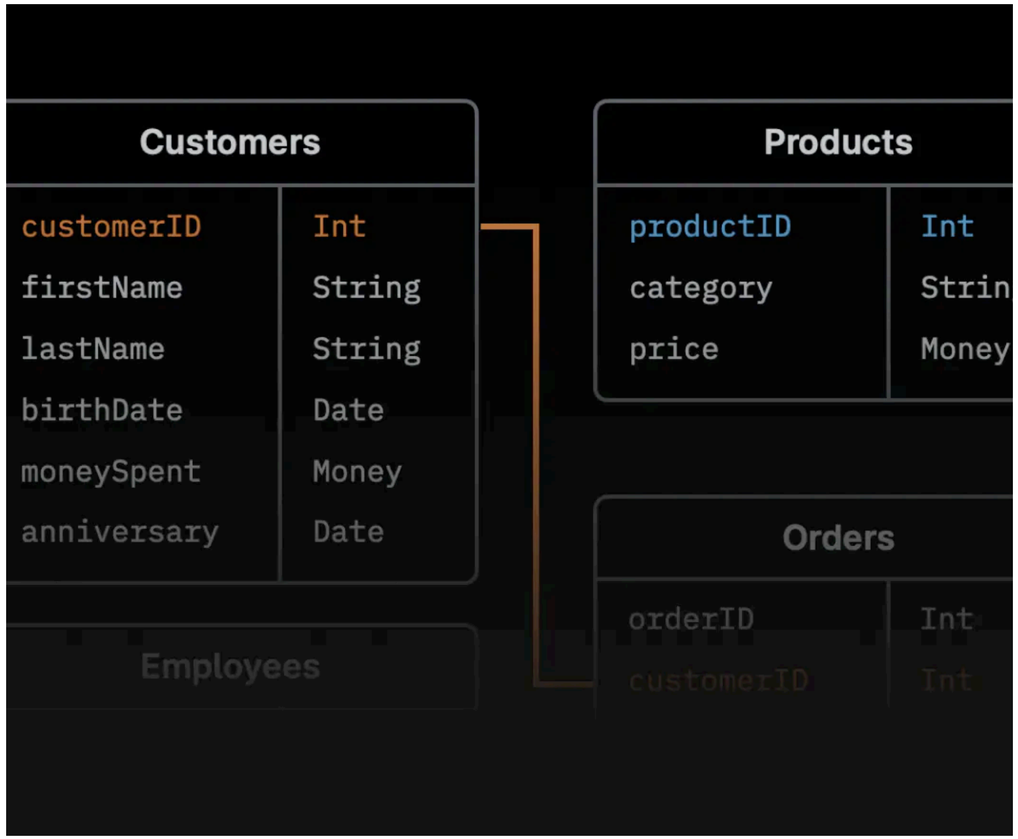
Data at the center

- Data is now the driven force behind the creation of new products and solutions

The old good SQL

Relational Databases

- Organize data into structured tables with predefined schemas.
- Structured Query lanaguages: SQL
- They excel in scenarios requiring strict data integrity and complex queries.
- Widely adopted for traditional business applications.
- Examples include MySQL, PostgreSQL, and Oracle.



Data governance

- Set of practices to manage data (in the prevailing theory)
- Control data representation types and correctness rules
- Data ownership
- Data cleaning
- Control data access
- Control data sources
- Create a hierarchical structure of control

Business Intelligence

- Data sources -> Ingest (Centralize) -> Clean -> Analyze -> Create Value

The emergence of Big Data

Big Data

- Extremely large sets of highly diverse complex data that cannot be managed with the tools based on the prevailing theory.
- This data is growing very fast

Characteristics of Big data

- Volume (GB-TB-Petabytes-Exabytes)
- Velocity (Internet speed and proliferation of connected devices)
- Variety (unestructured)
    - Types
    - Origin (voice, songs, movies, transactions, emails)
    - Source (Phones, IoT, Browsers, Searches, other devices)
- Veracity (Messy data, risky data, fake news)

Use cases

- Monitoring and tracking
  - Sentiments tracking
  - Inventory Tracking
  - Prediction of Energy consumption
  - Track website performance
  - predict financial market behavior

Use cases II

- Analysis and insight
  - Predict election results
  - Prevent crime
  - to better diagnose diseases and make medicine prescriptions.

Use cases III

- Digital product development
  - Stock markets feeds
  - Google Search
  - Image identification
  - Imagination is the limit.

NoSQL Databases

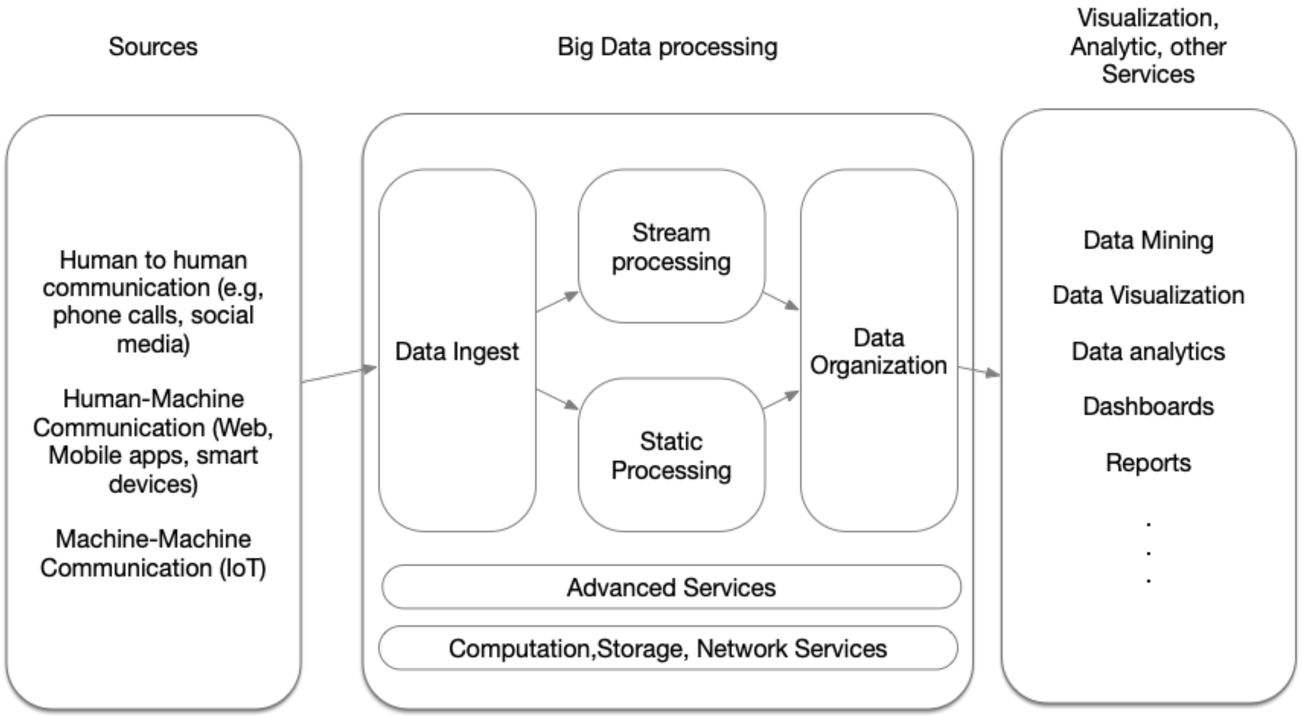
- The NoSQL movement emerged to address limitations in relational databases,
- Offering more flexibility and scalability.
- Types: document, key-value, and graph databases cater to different data structures.
- Examples: MongoDB, a document database, exemplifies the agility and scalability NoSQL databases provide.

Big Data Management

- Privacy policies
- Safe defaults
- Confidentiality and security enforced by technology
- Protect against internal attacks, "Encryption on storage"
- Legal frameworks
- Compliance
- Deal with Realtime

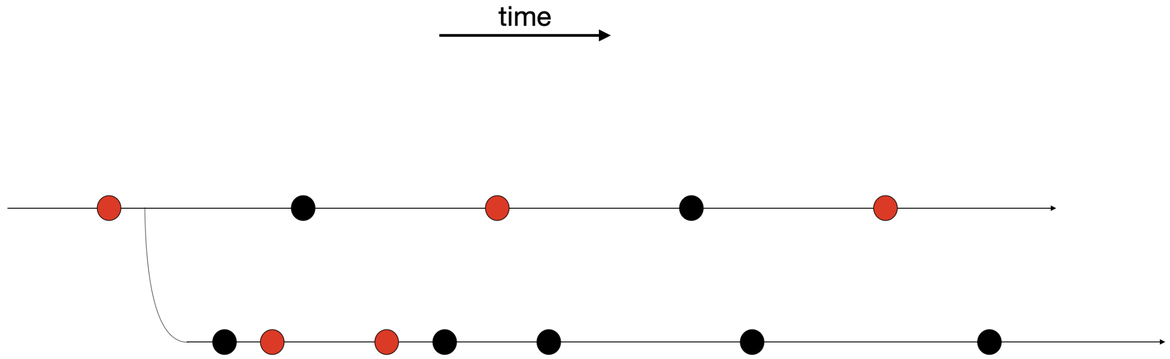
Analyzing BigData

- Big data in motion (Streams of data)
  - Clicks on Adds, most streamed series
- Big data at rest (Static big data)
  - FDC over Biomedical Data



Processing big data in realtime

- Events are the atomic
- Are more interesting in distributed applications
- Events describe actions over objects
- events happen at specific times
- Time is relative
- Time is clock-dependent as in relativity
- Causality and concurrency are difficult to detect
- There are many events



Multi threaded application in one computer

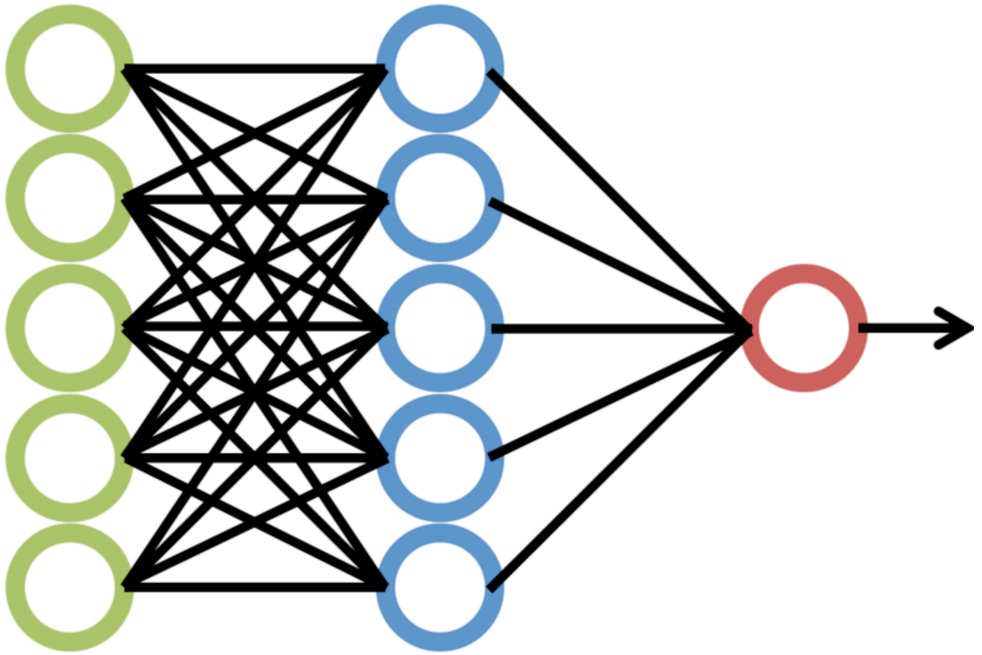
The rise of AI

AI

- Narrow
- AGI (Turing test?)

Machine learning

- Supervised
  - Linear regression, decision trees, random forest, neural networks
- unsupervised
  - K-means clustering, Hierarchical Clustering, etc.
- reinforcement learning.
  - Reinforcement learning (RL) algorithms are designed to enable agents to learn by interacting with an environment and receiving feedback in the form of rewards or penalties.



Deep learning

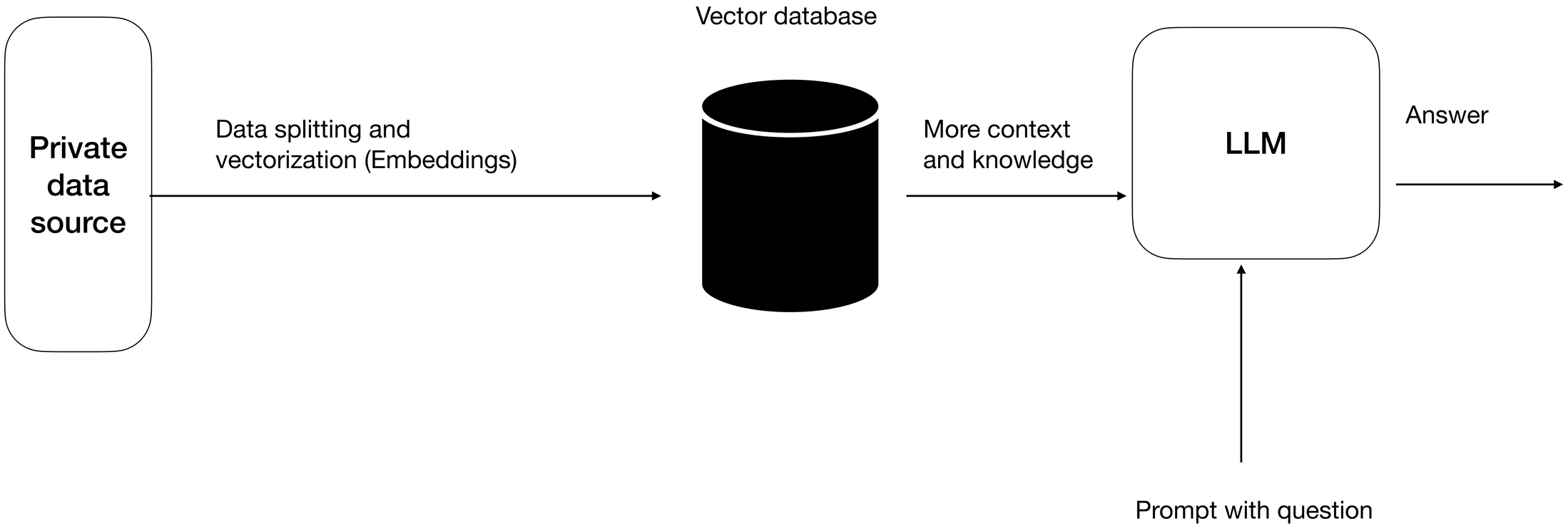
- Neural networks with many layers
- LLMs
- ChatGPT
- Text generation
- Image generation
- Audio generation

Technologies

- Large Language Models
- APIs
- Vector databases

A simple example of App architecture

Retrieval-augmented generation (RAG)



ML Management and Architecture

- Awareness
- Modernization
- Unification
- Innovation

Questions?