



## Il caso Ti-fone

# La compagni telefonica Ti-fone

---

- ▶ Due archivi:
  - ▶ **Clienti**, con dati relativi a
    - ▶ Numero di telefono e SIM
    - ▶ Anagrafica
    - ▶ Tipo contratto (abbonamento, ricaricabile, ecc.)
    - ▶ ...
  - ▶ **Telefonate**, con dati relativi a
    - ▶ Numero chiamato
    - ▶ Orario inizio e fine (durata)
    - ▶ Costo
- ▶ Archivi collegati in qualche modo:
  - ▶ Numero di telefono
  - ▶ Puntatori
  - ▶ Codice
  - ▶ ...

# Le richieste

---

- ▶ Al fine di migliorare i propri servizi, comprendere meglio le abitudini della propria clientela, ecc., il team di Ti-fone ha bisogno di poter soddisfare, tra le altre, le seguenti richieste:
  - 1) Trovare i dati anagrafici dato un numero di telefono (es. 348-1122334)
  - 2) Trovare i numeri di un cliente (es. Mario Rossi)
  - 3) Trovare i dettagli di tutte le chiamate di un giorno superiori a una certa durata (es. 20/09/2017, durata almeno 0:10.00 = 600 secondi)
  - 4) Trovare tutte le chiamate di un numero in un periodo (es. 348-1122334 a Settembre 2017)
  - 5) Trovare, per tutti i clienti residenti a Bologna, le chiamate eseguite un dato giorno di costo superiore a un dato importo (es. 20/09/2017, costo > 1.50€)
  - 6) Calcolare il costo totale in un dato giorno
  - 7) Trovare la coppia di clienti che si sono chiamati più spesso in un dato periodo
  - 8) Determinare quanti clienti chiamano in media in un mese più di 10 volte almeno 3 numeri diversi

# Algoritmi?

---

- ▶ Per ogni richiesta si può pensare di sviluppare uno specifico algoritmo, più o meno semplice (ma non è questo il punto!)
- ▶ Anche per le richieste più semplici la scelta tra diversi algoritmi non è scontata, e può dipendere dalle "condizioni al contorno":
  - 1.a) Scandire l'archivio dei clienti
  - 1.b) Se ordinato, eseguire una ricerca binaria
  - 1.c) Se disponibile un indice, usarlo per localizzare nell'archivio la posizione dei dati relativi al numero cercato
- ▶ ... oppure da vere e proprie modalità di esecuzione diverse
  - 5.a) Trovare tutti i clienti di Bologna;  
per ognuno trovare le chiamate del 20/09/2017 con costo > 1.50€
  - 5.b) Trovare le chiamate del 20/09/2017 con costo > 1.50€;  
per ognuna recuperare il cliente e mantenerla se di Bologna

# E quindi?

---

- ▶ Sperare di gestire la complessità dei dati odierni mediante lo sviluppo di algoritmi ad hoc è senza speranza
- ▶ Occorre quindi cambiare approccio, ovvero

rinunciare a trattare "direttamente" i dati  
e basarsi su un sistema sw (DBMS) pensato allo scopo,  
che oltre a occuparsi degli aspetti fisici (data storage)  
metta anche a disposizione degli strumenti "non algoritmici"  
per esprimere le varie richieste

▶ **COSA TI SERVE?** Elenco ordinato dei clienti di Bologna



Utente

▶ **COME LO OTTIENI?**

- ▶ Prima si eliminano i clienti non di Bologna, poi si ordina
- ▶ Prima si ordina, poi si eliminano i clienti non di Bologna
- ▶ Mentre si ordina si eliminano i clienti non di Bologna
- ▶ ...



DBMS

## Le basi...

- ▶ Contare, sia in senso progressivo che regressivo;
- ▶ confrontare raggruppamenti di oggetti rispetto alla loro quantità;
- ▶ leggere e scrivere i numeri naturali almeno entro il cento;
- ▶ eseguire con precisione e rapidità semplici calcoli mentali di addizioni e sottrazioni;
- ▶ raggruppare oggetti a due a due contando per due, e così via;
- ▶ calcolare il doppio/la metà, il triplo/il terzo, il quadruplo/il quarto, ecc.;
- ▶ eseguire, almeno entro il cento, addizioni e sottrazioni, moltiplicazioni e divisioni (con moltiplicatori e divisori di una cifra)

## ...e la continuazione

- ▶ Premesse:  $N, Z, Q, R$ , relazioni di ordine: minimo e massimo, estremi superiore e inferiore di un sottoinsieme di  $R$  [...]
- ▶ Numeri complessi
- ▶ Limiti
- ▶ Continuità
- ▶ Derivazione e applicazioni
- ▶ Integrazione e applicazioni
- ▶ Equazioni differenziali

# Le basi...

- ▶ Concetto di modello dei dati
- ▶ Il modello relazionale
- ▶ Le "4 operazioni" sui dati
- ▶ Non proceduralità delle richieste
- ▶ Come modellare i dati
- ▶ Vincoli sui dati e come farli rispettare

# ...e la continuazione

- ▶ Peculiarità e caratteristiche dei Big Data
- ▶ Caratteristiche delle infrastrutture IT per la raccolta e la gestione dei Big Data: Map Reduce, Pulizia e standardizzazione dei dati
- ▶ Introduzione alla piattaforma Hadoop
- ▶ Real time data analysis
- ▶ Modelli e tecniche statistiche applicate ai Big Data: regressione lineare e logistica, regressione ad albero, componenti principali, cluster, analisi di serie storiche, social networks analysis, machine learning.

