Data Analytics & Business Intelligence

FINANZA.TECH

IL NUOVO MODO DI FARE FINANZA D'IMPRESA

HELLO

Davide D’Antonio

Senior IT Engineer & Tech Lead @ FINANZA.TECH

Quiz Time!

http://survey.finanza.tech

AGENDA

Introduzione al mondo dei dati

Ruoli e competenze di un professionista dei dati

Strumenti per la classificazione dei dati

Il “Triangolo degli Strumenti” dell’analista

files, formati e tipi

Pipelines: i dati dalla collezione all’analisi

PARTE I

Introduzione al mondo dei dati

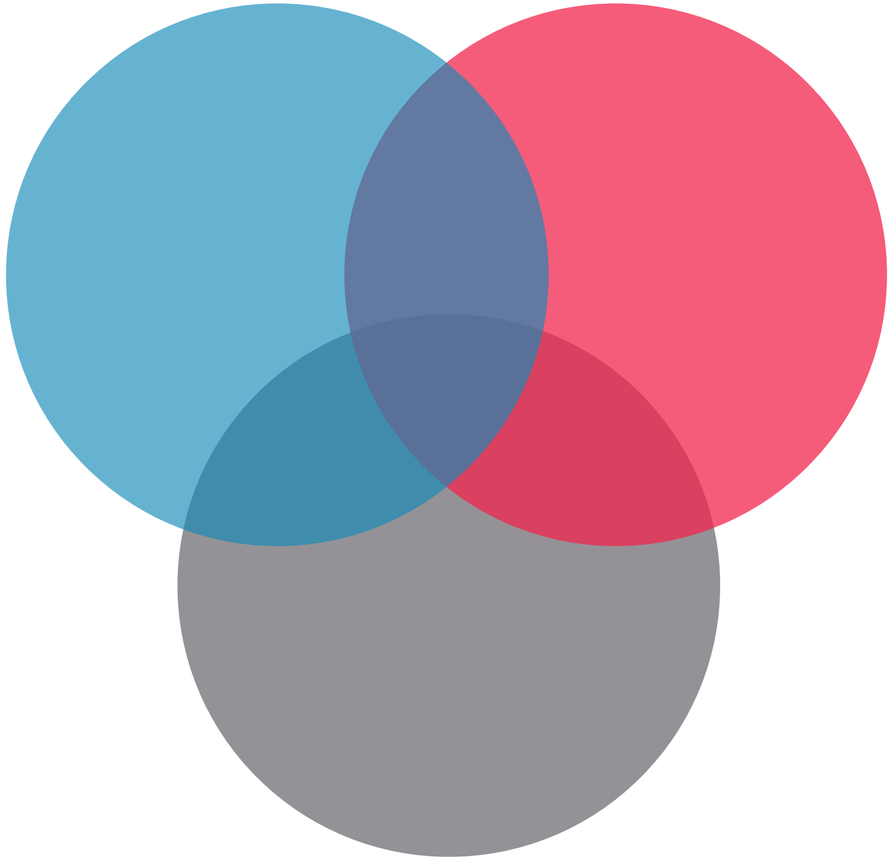
Cinque semplici domande

DOMANDA #1	Perché i dati, perché adesso?
DOMANDA #2	Cosa sono i dati?
DOMANDA #3	Da dove provengono i dati?
DOMANDA #4	Chi utilizza i dati?
DOMANDA #5	A cosa servono i dati?

DOMANDA #1

Perché i dati, perché adesso?

TRAIETTORIA
Volume / Varietà



BENEFICI
Società / Business

ABBASSAMENTO COSTI
Convenienza / Accessibilità

PERCHÉ I DATI, PERCHÉ ADESSO?

I dati contengono potenziale, e la dimensione di quel potenziale diventa sempre più grande e sempre più accessibile

DOMANDA #2

Cosa sono i dati?

Cosa sono i dati?

Filosoficamente

Praticamente

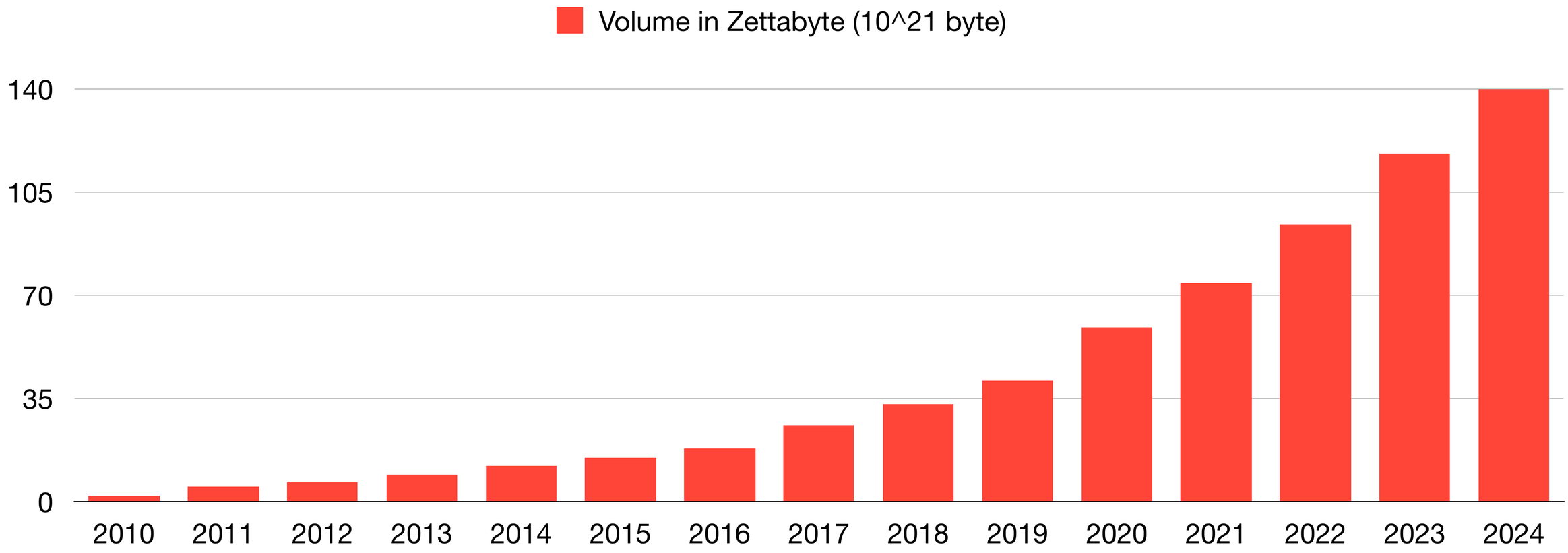
Filosoficamente, cosa sono i dati?

F
E
H
C

Fall Spice

Filosoficamente, i dati sono un MINDSET

Praticamente, cosa sono i dati?


Volume in Zettabyte (10^21 byte)
140
105
70
35
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

Praticamente, i dati sono tutte le cose che vengono catturate dalle tecnologie moderne

DOMANDA #3

Da dove provengono i dati?

Da dove provengono i dati?

TECNOLOGIA	ENTITÀ	DATO
Social Network		
Wikipedia		
Open Data		
Questionari	Google Forms	

DOMANDA #4

Chi utilizza i dati?

Chi utilizza i dati?

TECNOLOGIA	ENTITÀ	DATO	Settore
Social Network			Internet/Advertising
Wikipedia			Non-Profit
Open Data			Fintech
Questionari	Google Forms		Governo

DOMANDA #5

A cosa servono i dati?

Decisioni

Prodotti

Storie

DATI

I dati sono tutte le cose che vengono registrate che potrebbero essere utilizzate come materia prima per prendere le decisioni della vita, aiutare a costruire prodotti migliori e condividere storie straordinarie.

PARTE III

Ruoli e competenze di un professionista dei dati

DATA ANALYTICS

DATA SCIENCE

DATA ENGINEERING

DATA ANALYTICS

Decisioni con i dati

DATA ANALYTICS

Decisioni con i dati

1. Cosa accade

Come sono andate le vendite?

Quanti utenti si sono registrati?

Quante sottoscrizioni sono state annullate?

DATA ANALYTICS

Decisioni con i dati

2. Perché è accaduto

Quali auto sono state vendute di più il mese scorso?

Su quale piattaforma si registrano gli utenti?

Quali film hanno visto prima di dismettere?

DATA ANALYTICS

Decisioni con i dati

3. Cosa accadrà

Quali saranno le vendite?

Quanti utenti avremo?

Diminuiranno le cancellazioni?

DATA ANALYTICS

Decisioni con i dati

4.Che azione intraprendere

Come aumentiamo le vendite?

Come raggiungiamo i 2 miliardi di iscritti?

Come riduciamo le cancellazioni del 10%?

DATA ANALYTICS

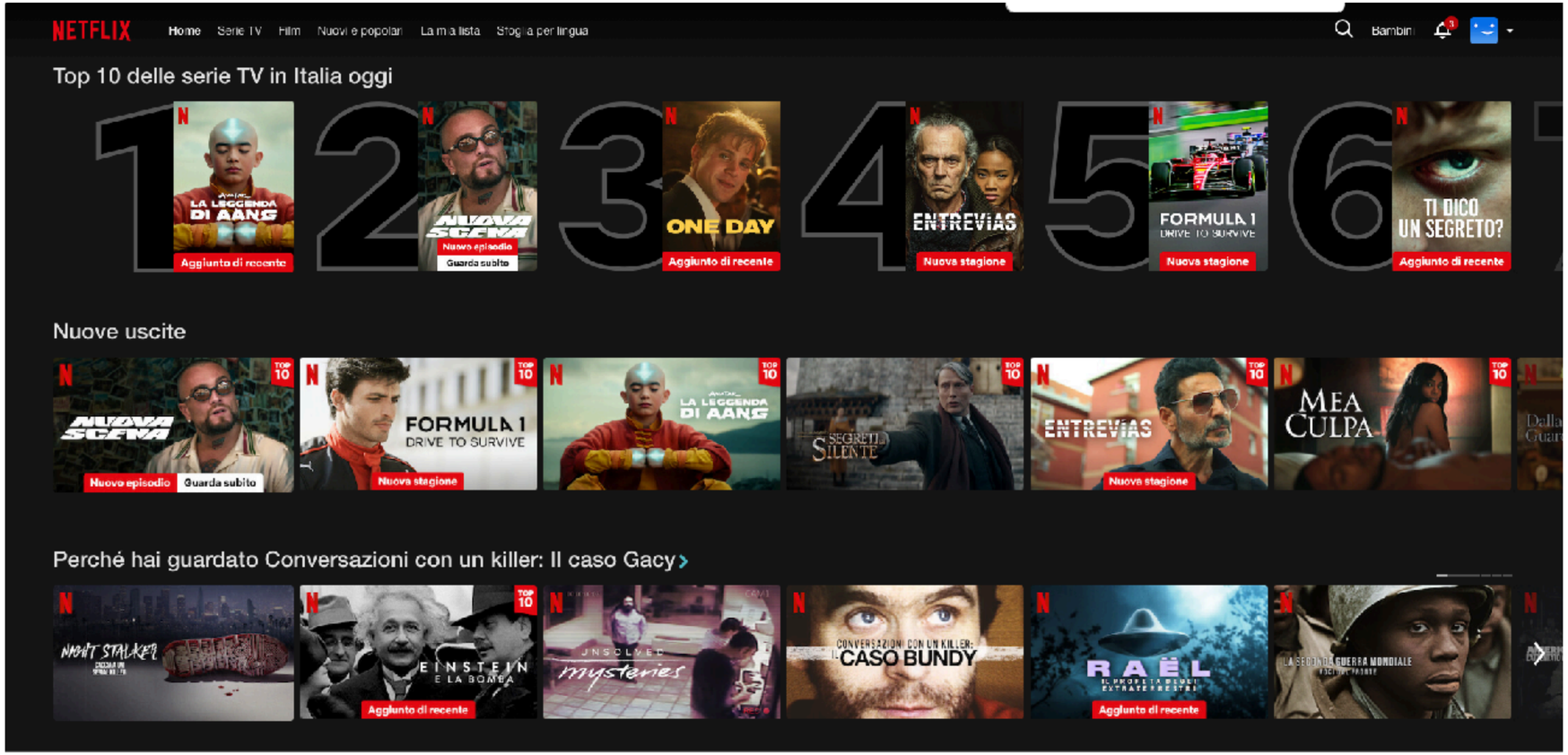
Decisioni con i dati

1. Cosa accade
2. Perché è accaduto
3. Cosa accadrà
4. Che azione intraprendere

DECISION-BASED
Modelli statistici avanzati

DATA SCIENCE

FEATURE-BASED
Costruire prodotti con i dati


NETFLIX
Home
Serie TV
Film
Nuovi e popolari
La mia lista
Sfoglia per lingua
Bambini
Top 10 delle serie TV in Italia oggi
LA LEGGENDA DI AANG
Aggiunto di recente
NUOVA SCENA
Nuovo episodio
Guarda subito
ONE DAY
Aggiunto di recente
ENTREVIAS
Nuova stagione
FORMULA 1
DRIVE TO SURVIVE
Nuova stagione
TI DICO UN SEGRETO?
Aggiunto di recente
Nuove uscite
NUOVA SCENA
Nuovo episodio
Guarda subito
FORMULA 1
DRIVE TO SURVIVE
Nuova stagione
LA LEGGENDA DI AANG
SEGRETO SILENTE
ENTREVIAS
Nuova stagione
MEA CULPA
Perché hai guardato Conversazioni con un killer: Il caso Gacy
NIGHT STALKER
EINSTEIN E LA BOMBA
Aggiunto di recente
UNSOLVED mysteries
CONVERSAZIONI CON UN KILLER: IL CASO BUNDY
RAËL
Aggiunto di recente
LA SECONDA GUERRA MONDIALE

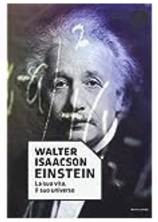
I clienti che hanno visto questo articolo hanno visto anche

Pagina 1 di 9

Elon Musk. (Italiano)
Walter Isaacson
4.4 251
Copertina rigida
25,65 €
prime Consegna GRATUITA
in un giorno Ricevilo
Domani, feb 27

L'arte della vittoria. Autobiografia del fondatore della Nike
Phil Knight
4.4 4.526
Copertina flessibile
12,82 €
prime Consegna GRATUITA
in un giorno Ricevilo
Domani, feb 27

Elon Musk. Tesla, SpaceX e la sfida per un futuro fantastico
Ashlee Vance
4.4 1.895
Copertina flessibile
14,15 €
prime Consegna GRATUITA
in un giorno Ricevilo
Domani, feb 27



Einstein. La sua vita, il suo universo
Walter Isaacson
4.4 380
Copertina flessibile
19,95 €
prime Consegna GRATUITA
in un giorno Ricevilo
Domani, feb 27

Sergio Marchionne. Nuova Edizione Aggiornata
Tommaso Ebhardt
4.4 2.790
Copertina flessibile
17,95 €
prime Consegna GRATUITA
in un giorno Ricevilo
Domani, feb 27

DATA ENGINEERING

Implementa le infrastrutture

(che consente l'analisi e il data science)

DATA ENGINEERING

Implementa le infrastrutture

(che consente l'analisi e il data science)

1. Raccolta

Qual è l'obiettivo della raccolta?

Quali tipi di dati sono necessari?

Quali fonti si possono utilizzare?

DATA ENGINEERING

Implementa le infrastrutture

(che consente l'analisi e il data science)

2. Conservazione

Quali tipi di dati devo conservare?

Esistono requisiti normativi?

Per quanto tempo devo conservarli?

Quali sono le migliori pratiche di conservazione?

DATA ENGINEERING

Implementa le infrastrutture

(che consente l'analisi e il data science)

3. Elaborazione

Quali sono i dati coinvolti?

Qual è l'obiettivo dell'elaborazione?

Che tipo di dato è coinvolto?

Come verranno interpretati i dati?

DATA ENGINEERING

Implementa le infrastrutture

(che consente l'analisi e il data science)

4. Elaborazione

Che strumenti vengono utilizzati?

Che dati devo trasferire?

Chi sono i destinatari dei dati?

Che misure di sicurezza vengono utilizzati?

DATA ANALYTICS
Decisioni con i dati

DECISION-BASED
Modelli statistici avanzati

DATA SCIENCE

FEATURE-BASED
Costruire prodotti con i dati

DATA ENGINEERING
Implementa le infrastrutture
(che consente l'analisi e il data science)

DATA ANALYTICS	DATA SCIENCE	DATA ENGINEERING
BACKGROUND *Mentalità aperta*	*BACKGROUND* *Informatica / Fisica / Matematica*	*BACKGROUND* *Informatica / Ingegnere del Software*
SKILLS *SQL / Comunicazione*	*SKILLS* *Programmazione / Statistica*	*SKILLS* *Programmazione / Sistemista Database*
RUOLI *Analista / Marketing*	*RUOLI* *Data Scientist / ML Engineer*	*RUOLI* *Data Engineer / Data Architect*

PARTE III

Strumenti per la classificazione dei dati

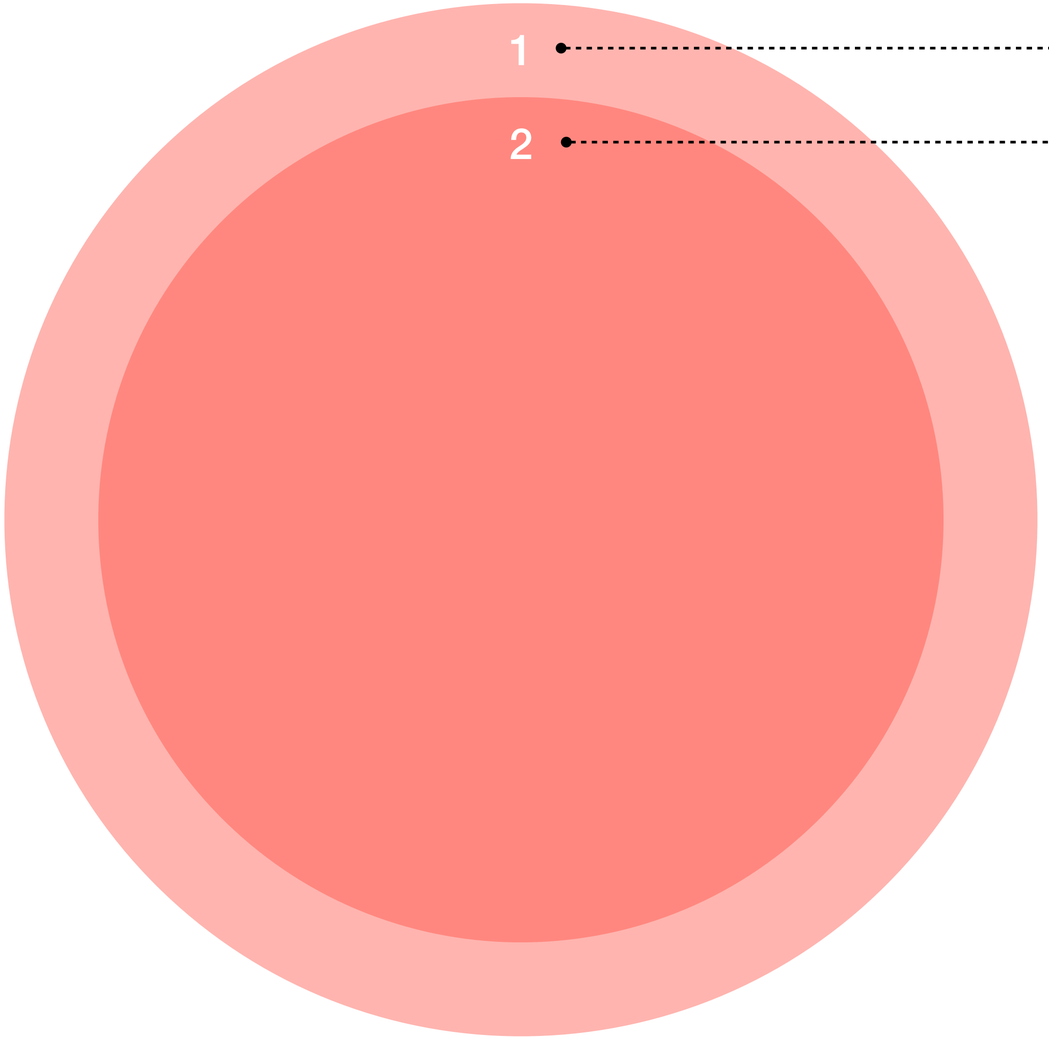
1

TECNOLOGIA

Hardware e Software

Hardware

Software



TECNOLOGIA
Hardware e Software

CAPACITÀ
Aspetti funzionali del software (features)

Abilità di immagazzinare dati

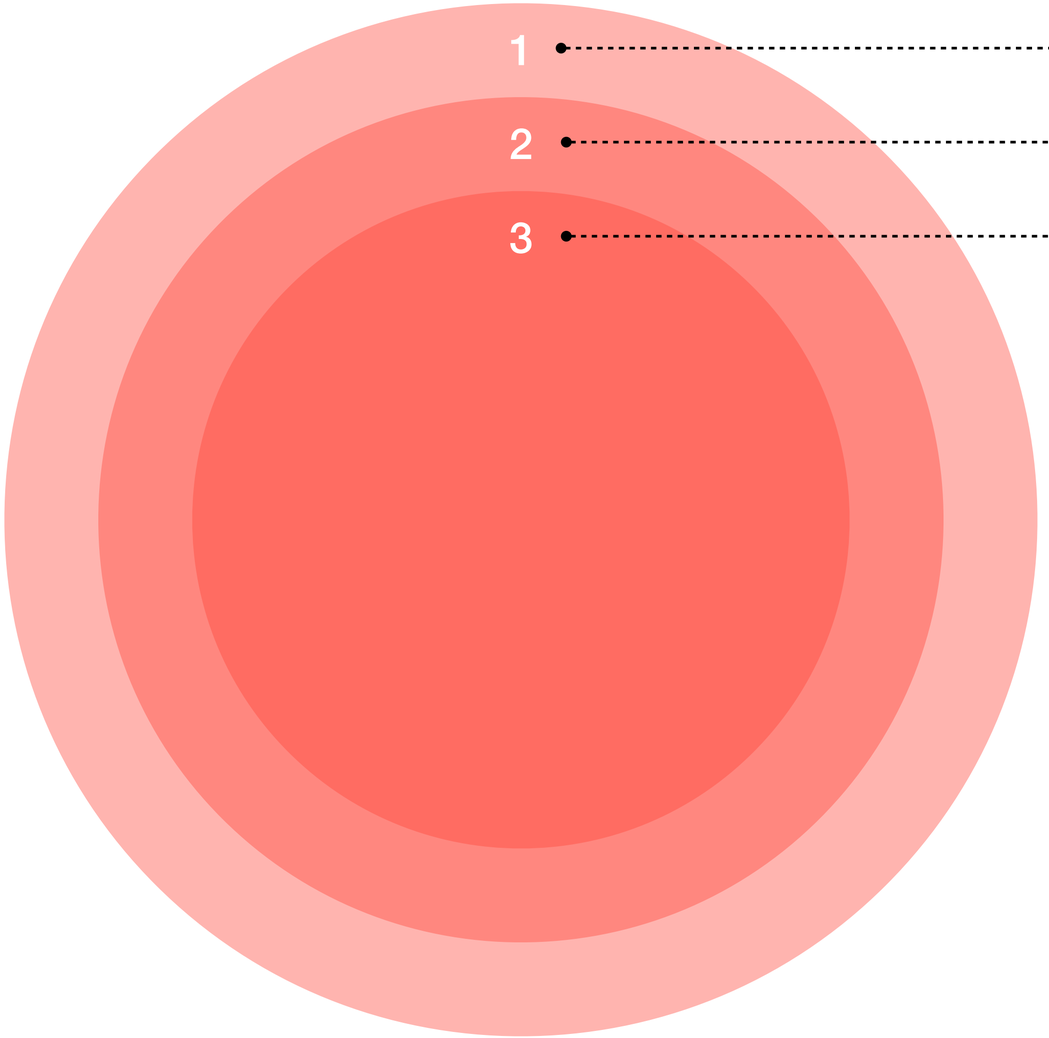
Eredita La struttura dei dati

Vista e/o estrazione dati

Calcoli, modifiche, ristrutturazione,…

Reports, visualizzazioni, annotazioni,…

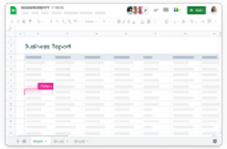
STORAGE **STRUTTURA** **ACCESSO** **MANIPOLAZIONE** **CONDIVISIONE**



TECNOLOGIA
Hardware e Software

CAPACITÀ
Aspetti funzionali del software (features)

CATEGORIE
Categorie di strumenti utilizzati per l'analisi dei dati



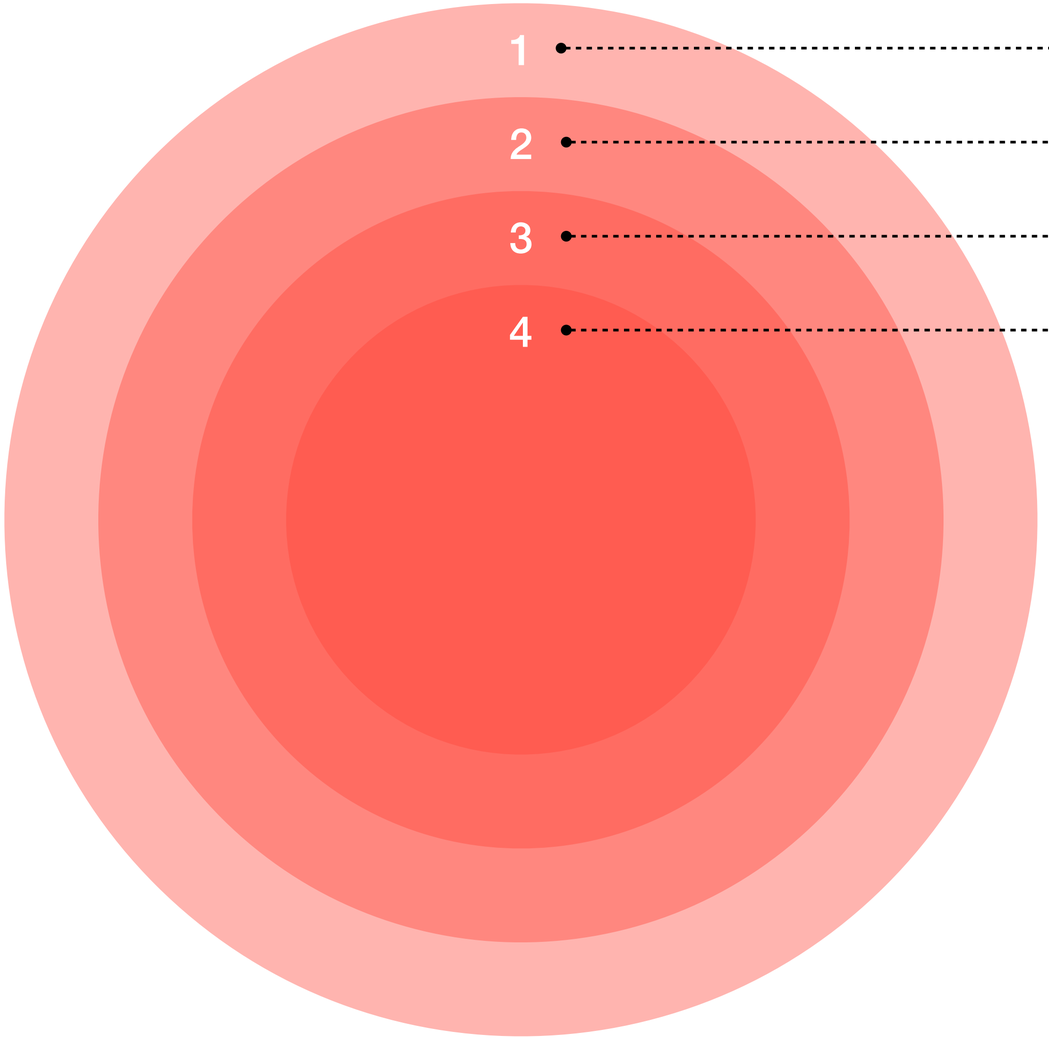
MySQL mongoDB

JS

SPREADSHEET | **DATABASE + QL** | **BI SOFTWARE** | **PROGRAMMAZIONE**

	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE	✓	✓			
QUERY LANGUAGE			✓	✓	
BI SOFTWARE			✓	✓	✓
PROGRAMMAZIONE		✓	✓	✓	✓

	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE + QL	✓	✓	✓	✓	
BI SOFTWARE			✓	✓	✓
PROGRAMMAZIONE		✓	✓	✓	✓



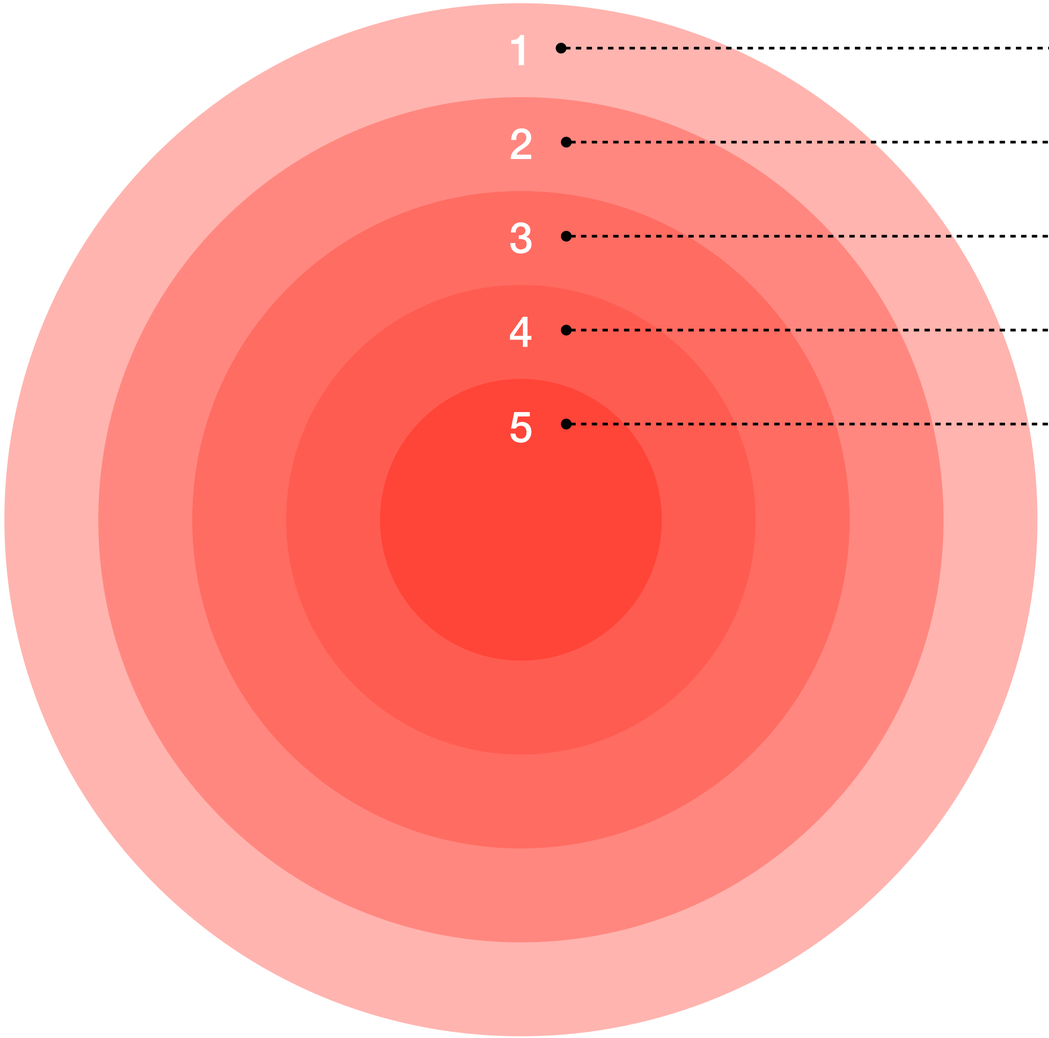
TECNOLOGIA
Hardware e Software

CAPACITÀ
Aspetti funzionali del software (features)

CATEGORIE
Categorie di strumenti utilizzati per l'analisi dei dati

CREATORI
Entità che creano uno strumento di analisi dei dati

	VENDOR	VENDOR	VENDOR
SPREADSHEET	Google		Microsoft
DATABASE	amazon	ORACLE	Microsoft
QUERY LANGUAGE	MySQL		Microsoft
BI SOFTWARE	Qlik	Grafana	Microsoft
PROGRAMMAZIONE	python™	R	Microsoft



TECNOLOGIA
Hardware e Software

CAPACITÀ
Aspetti funzionali del software (features)

CATEGORIE
Categorie di strumenti utilizzati per l'analisi dei dati

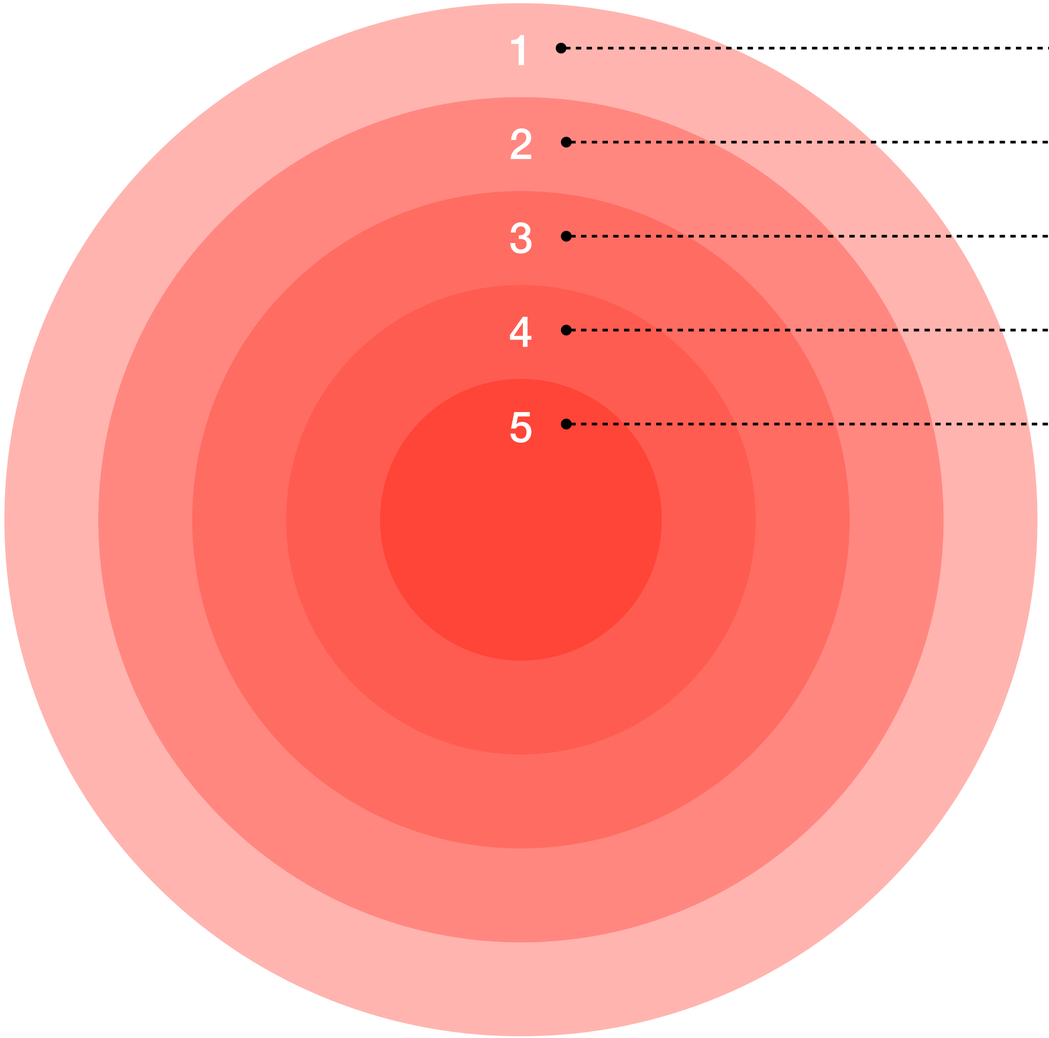
CREATORI
Entità che creano uno strumento di analisi dei dati

TOOL
Il prodotto che infine viene utilizzato e definito strumento di analisi dati

	VENDOR	TOOL
SPREADSHEET	Microsoft	
DATABASE + QL	amazon	amazon RDS | MySQL
BI SOFTWARE	Grafana Labs	Grafana
PROGRAMMAZIONE	python SOFTWARE FOUNDATION	python™

CLASSIFICAZIONE SPREADSHEET

MICROSOFT EXCEL



TECNOLOGIA
Software

CAPACITÀ
Storage / Struttura / Accesso / Manipolazione / Condivisione

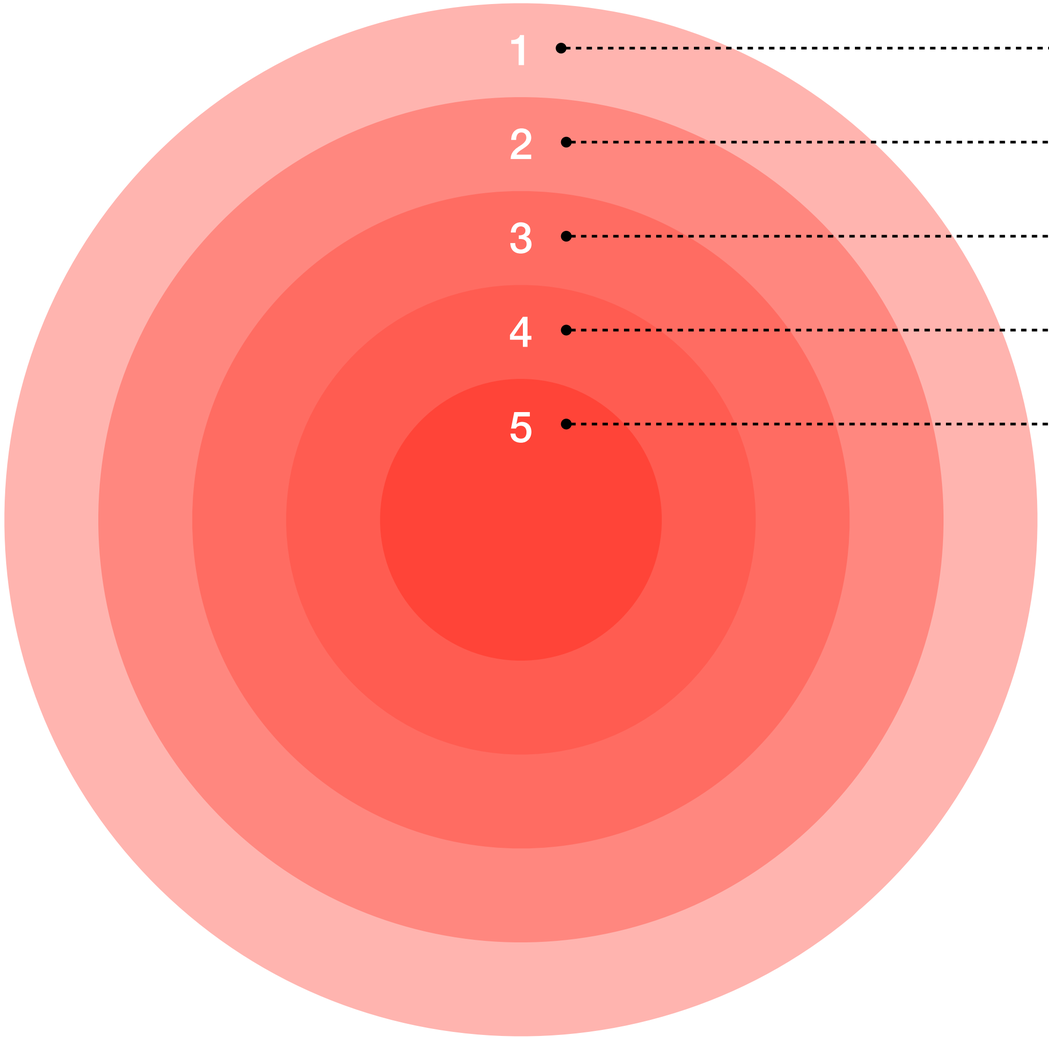
CATEGORIE
Spreadsheet

CREATORI
Microsoft

TOOL
Excel per Mac

CLASSIFICAZIONE DATABASE + QL

AMAZON RDS, MySQL



TECNOLOGIA
Software

CAPACITÀ
Storage / Struttura / Accesso / Manipolazione

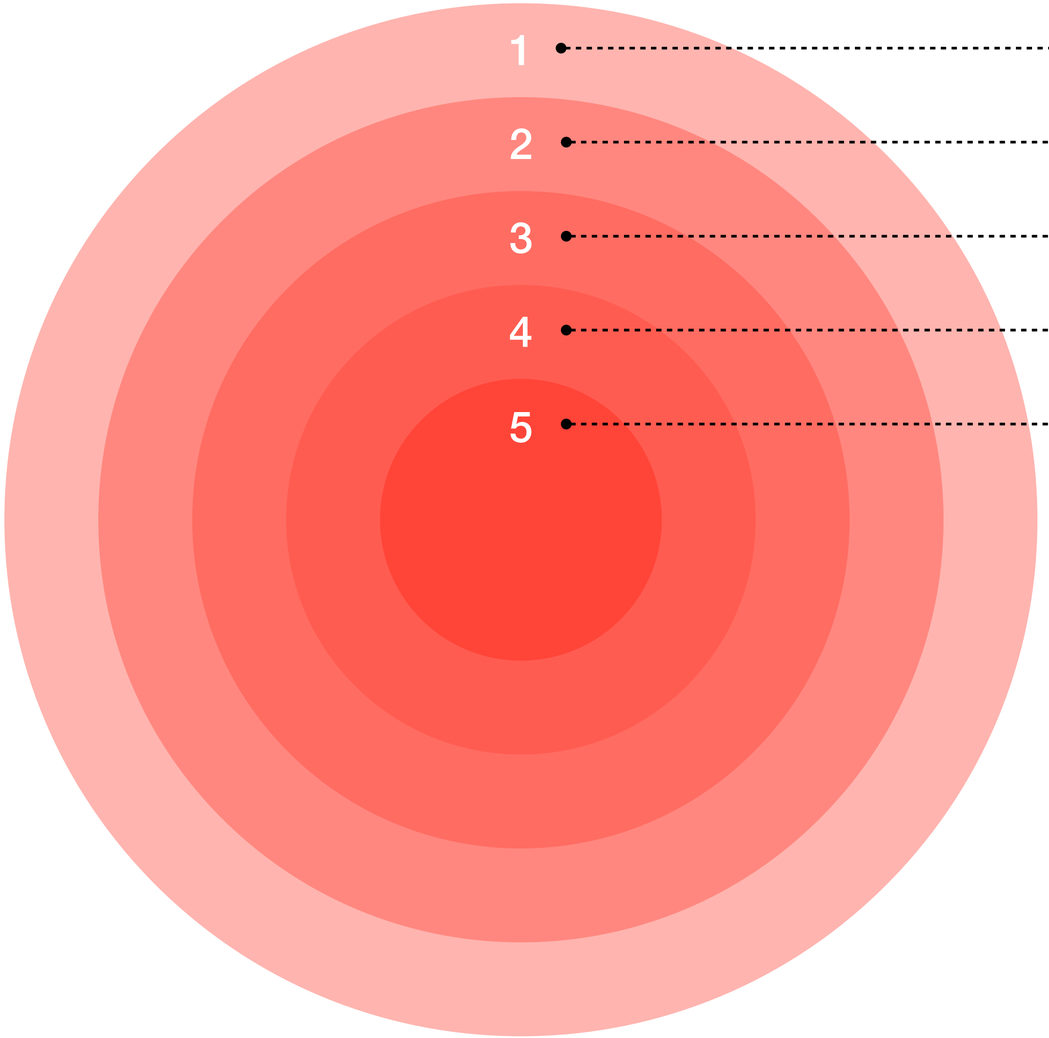
CATEGORIE
Database + QL

CREATORI
Amazon + MySQL

TOOL
Amazon RDS, MySQL + SQL

CLASSIFICAZIONE BI SOFTWARE

GRAFANA



TECNOLOGIA
Software

CAPACITÀ
Accesso / Manipolazione / Condivisione

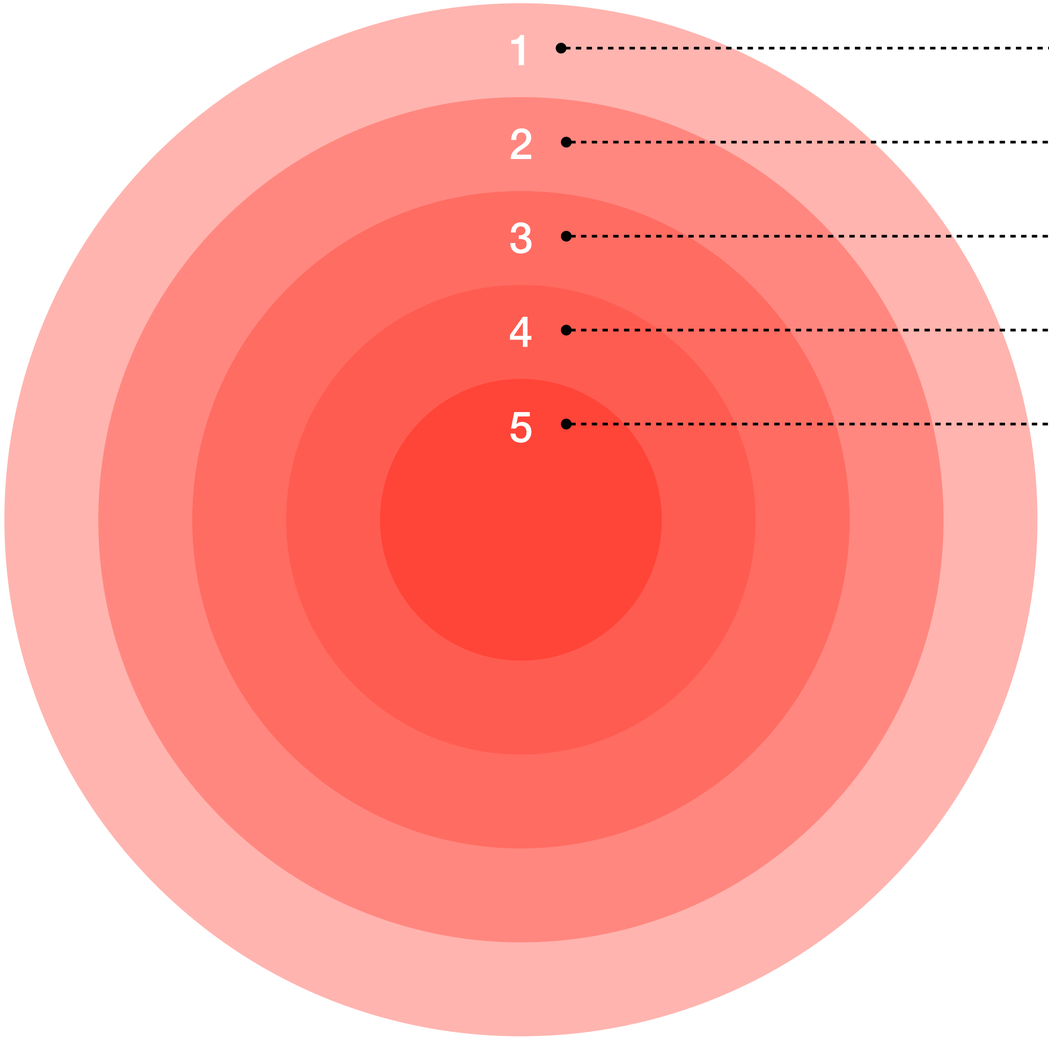
CATEGORIE
Software di Business Intelligence

CREATORI
Grafana Labs

TOOL
Grafana

CLASSIFICAZIONE LINGUAGGI DI PROGRAMMAZIONE

PYTHON



TECNOLOGIA
Software

CAPACITÀ
Accesso / Manipolazione

CATEGORIE
Linguaggi di Programmazione

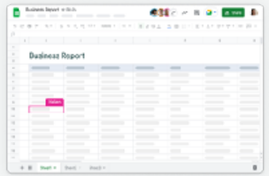
CREATORI
Python Software Foundation

TOOL
Python

PARTE IV

Il “Triangolo degli Strumenti” dell’analista

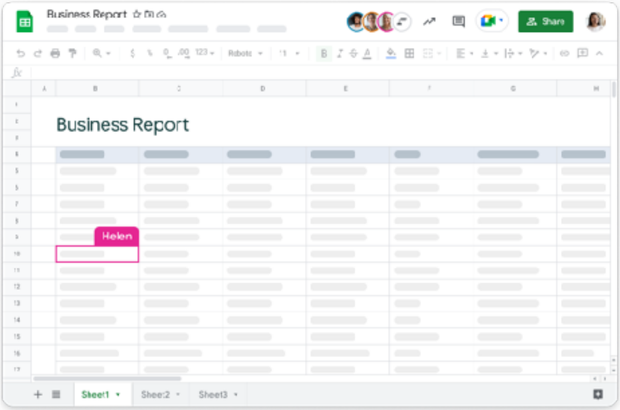
DATABASE & SQL



SPREADSHEET

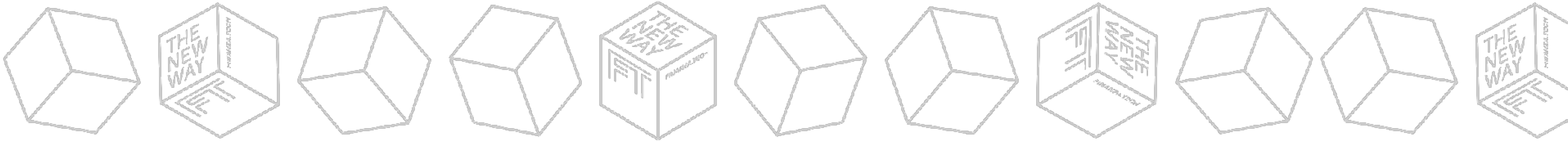
BI SOFTWARE

SPREADSHEET



	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE					
QUERY LANGUAGE					
BI SOFTWARE					

DATABASE & QL



	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE	✓	✓			
QUERY LANGUAGE			✓	✓	
BI SOFTWARE					

Salvataggio dei dati in uno
STORAGE

SQL

Accesso ai dati del nostro
DATABASE scrivendo **SQL**

Scriveremo il nostro **SQL** in
un **CLIENT SQL**

	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE + QL	✓	✓	✓	✓	
BI SOFTWARE					

BI SOFTWARE

BI SOFTWARE

SQL

DATABASE

	STORAGE	STRUTTURA	ACCESSO	MANIPOLAZIONE	CONDIVISIONE
SPREADSHEET	✓	✓	✓	✓	✓
DATABASE + QL	✓	✓	✓	✓	
BI SOFTWARE			✓	✓	✓

PARTE IV

files, formati e tipi

FILES DI DATI

il tipo di file in cui sono memorizzati i dati - solitamente riconosciuti dall’estensione del file

nome_file.xlsx

Estensione di un file Excel

Cerca
CALENDARIO DOCENZE CORSO IFTS BUSINESS DATA ANALYST .xlsx
Users > davide > Downloads > CALENDARIO DOCENZE CORSO IFTS BUSINESS DATA ANALYST .xlsx
Live Share
Riga 1, colonna 1
Spazi: 2
UTF-8
LF
Binary
Colorize: 0 variables
Colorize

nome_file.csv

Estensione di un file csv (comma separated values)

Nome,Cognome,età
Davide,D’Antonio,40
Mario,Rossi,35
John,Doe,43

nome_file.tsv

Estensione di un file tsv (tab separated values)

Nome	Cognome	età
Davide	D’Antonio	40
Mario	Rossi	35
John	Doe	43

CSV vs TSV

FORMATO DEI DATI

La struttura dei dati all’interno del file

CSV

Il file è formattato utilizzando il delimitatore virgola

TSV

Il file è formattato utilizzando il delimitatore tab

EXCEL

Il file è formattato utilizzando righe e colonne

FLAT

EXCEL

STRUTTURATO

Nome	Cognome	Età	Veicolo
Davide	D’Antonio	40	Hyundai Kona
Mario	Rossi	35	Jeep Renegade
John	Doe	43	Peugeot 208

Nome	Cognome	Età	Veicolo	Veicolo 2
Davide	D’Antonio	40	Hyundai Kona	Fiat Panda
Mario	Rossi	35	Jeep Renegade	
John	Doe	43	Peugeot 208	

CSV

Il file è formattato utilizzando il delimitatore virgola

TSV

Il file è formattato utilizzando il delimitatore tab

EXCEL

Il file è formattato utilizzando righe e colonne

JSON

Il file è formattato utilizzando coppie chiave valore

FLAT

EXCEL

JSON

STRUTTURATO

SEMI-

CHIAVE : VALORE

nome_file.json

Estensione di un file json (JavaScript Object Notation)

```
[
  {
    "nome": "Davide",
    "cognome": "D'Antonio",
    "età": 40,
    "veicolo": ["Hyundai Kona", "Fiat Panda"]
  },
  {
    "nome": "Mario",
    "cognome": "Rossi",
    "età": 35,
   "veicolo":["Jeep Renegade"]
  },
  {
    "nome": "John",
    "cognome": "Doe",
    "età": 43,
    "veicolo": ["Peugeot 208"]
  }
]
```

```
[
 {
  "nome": "Davide",
  "cognome": "D'Antonio",
  "età": 40,
  "veicolo": [
   { "marca": "Hyndai", "modello": "Kona" },
   { "marca": "Fiat", "modello": "Panda"}
  ]
 },
 {
  "nome": "Mario",
  "cognome": "Rossi",
  "età": 35,
  "veicolo":[
   { "marca": "Jeep", "modello": "Renegade"}
  ]
 },
 {
  "nome": "John",
  "cognome": "Doe",
  "età": 43,
  "veicolo": [
   { "marca": "Peugeot", "modello": "208"}
  ]
 }
]
```

TIPI DI DATI

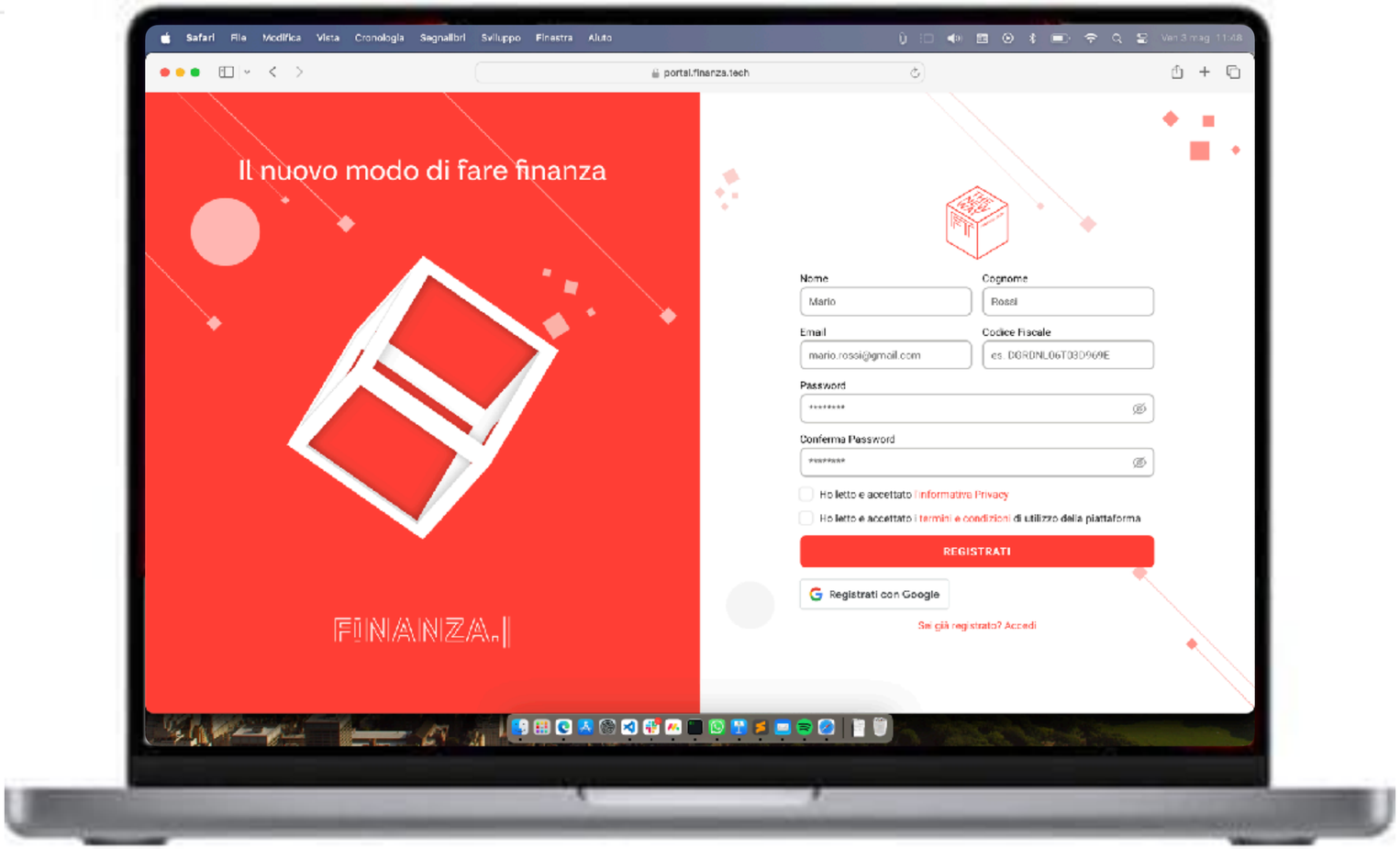
TESTO

NUMERI

DATE

TESTO

Spesso ci riferiamo ad esso come Stringa


Safari File Modifica Vista Cronologia Segnalibri Sviluppo Finestra Aiuto
portal.finanza.tech
Il nuovo modo di fare finanza
FINANZA.|
Nome
Mario
Cognome
Rossi
Email
mario.rossi@gmail.com
Codice Fiscale
es. DGRDNL06T03D969E
Password
Conferma Password
Ho letto e accettato l'informativa Privacy
Ho letto e accettato i termini e condizioni di utilizzo della piattaforma
REGISTRATI
Registrati con Google
Sei già registrato? Accedi

Numeri

Possono essere di diversi formati

Numeri

INTEGER

11

NUMERIC

234.22

BOOLEAN

1 | 0

DATE

Solitamente salvate in un database con il formato YYYY-MM-DD. In alcuni casi anche in formato intero (UNIX Time Stamp)

Che impatto hanno i tipi di dato?

$10.43

Numeric

$10

Integer

2024-04-01 23:34:33

DATE (TIMESTAMP)

UTC
PSD
EST

TIMEZONE

PARTE V

Pipelines: i dati dalla collezione all’analisi

Le data pipelines si riferiscono al processo di trasformazione, movimento e gestione dei dati da una fonte a una destinazione, seguendo una serie di passaggi predefiniti e spesso automatizzati.

Pipelines: da cosa sono composte

Dati

Analisi

Condividi

Esempio: transazioni e-commerce

Esempio: transazioni e-commerce

Collezione dati

Analisi + Condivisione

Difficoltà delle pipeline

BASE

INTERMEDIA

COMPLESSA

BASE

A → B

SORGENTE

e-commerce

RAW DATA

Interazioni utente

DATABASE

I dati vengono strutturati in un database

ANALISI + CONDIVISIONE

INTERMEDIO

A → B

SORGENTE

e-commerce

RAW DATA

Interazioni utente

DATABASE

I dati vengono strutturati in un database

ANALISI + CONDIVISIONE

EXTRACT

Estrarre i dati grezzi dal database

TRANSFORM

Fare “qualcosa” con i dati estratti (e.g. conversioni, filtri etc.)

LOAD

Salva i dati trasformati in un database separato

INTERMEDIO

A → B

SORGENTE

e-commerce

RAW DATA

Interazioni utente

DATABASE

I dati vengono strutturati in un database

ETL

I dati vengono puliti e trasformati

DATABASE ANALYTICS

I dati vengono salvati in un database separato

ANALISI + CONDIVISIONE

COMPLESSO

A → B

SORGENTE

e-commerce

RAW DATA

Interazioni utente

DATABASE

I dati vengono strutturati in un database

ETL

I dati vengono puliti e trasformati

DATABASE ANALYTICS

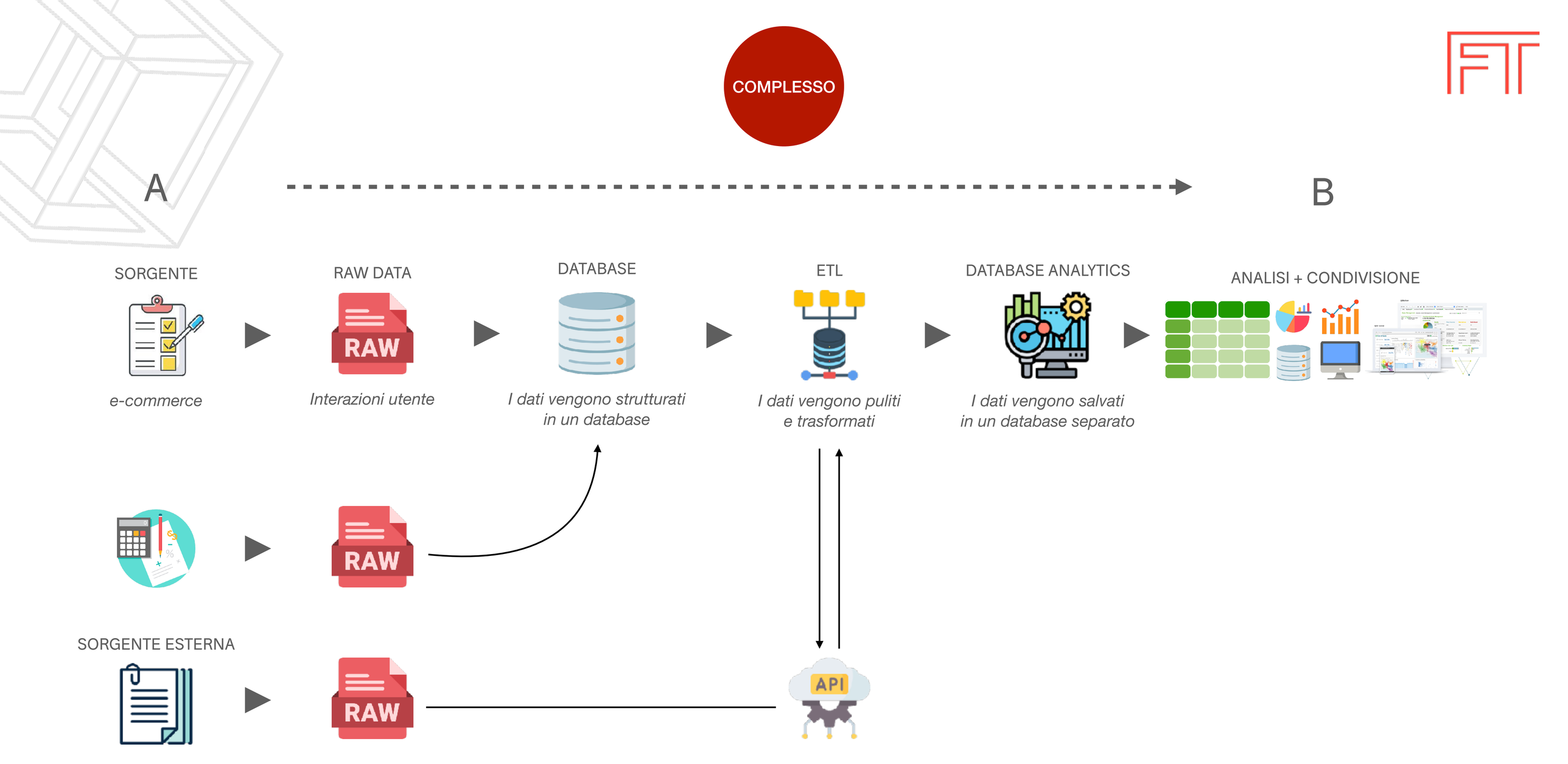
I dati vengono salvati in un database separato

ANALISI + CONDIVISIONE

SORGENTE ESTERNA

APPLICATION PROGRAMMING INTERFACE

Tecnologia che fa da collante tra diversi sistemi e consente di condividere dati



SOFTWARE ENGINEERING	DATA ENGINEERING	DATA ANALIST
Costruiscono cose che generano dati	*Muovono, manipolano e trasformano dati*	*Analizzano i dati e li condividono*

Senza i dati

la tua è solo un’opinione

MATERIALE DEL CORSO

http://data-analytics-bi.finanza.tech

https://github.com/davidedantonio/data-analytics

GOOD BYE

Davide D’Antonio

Senior IT Engineer & Tech Lead @ FINANZA.TECH