

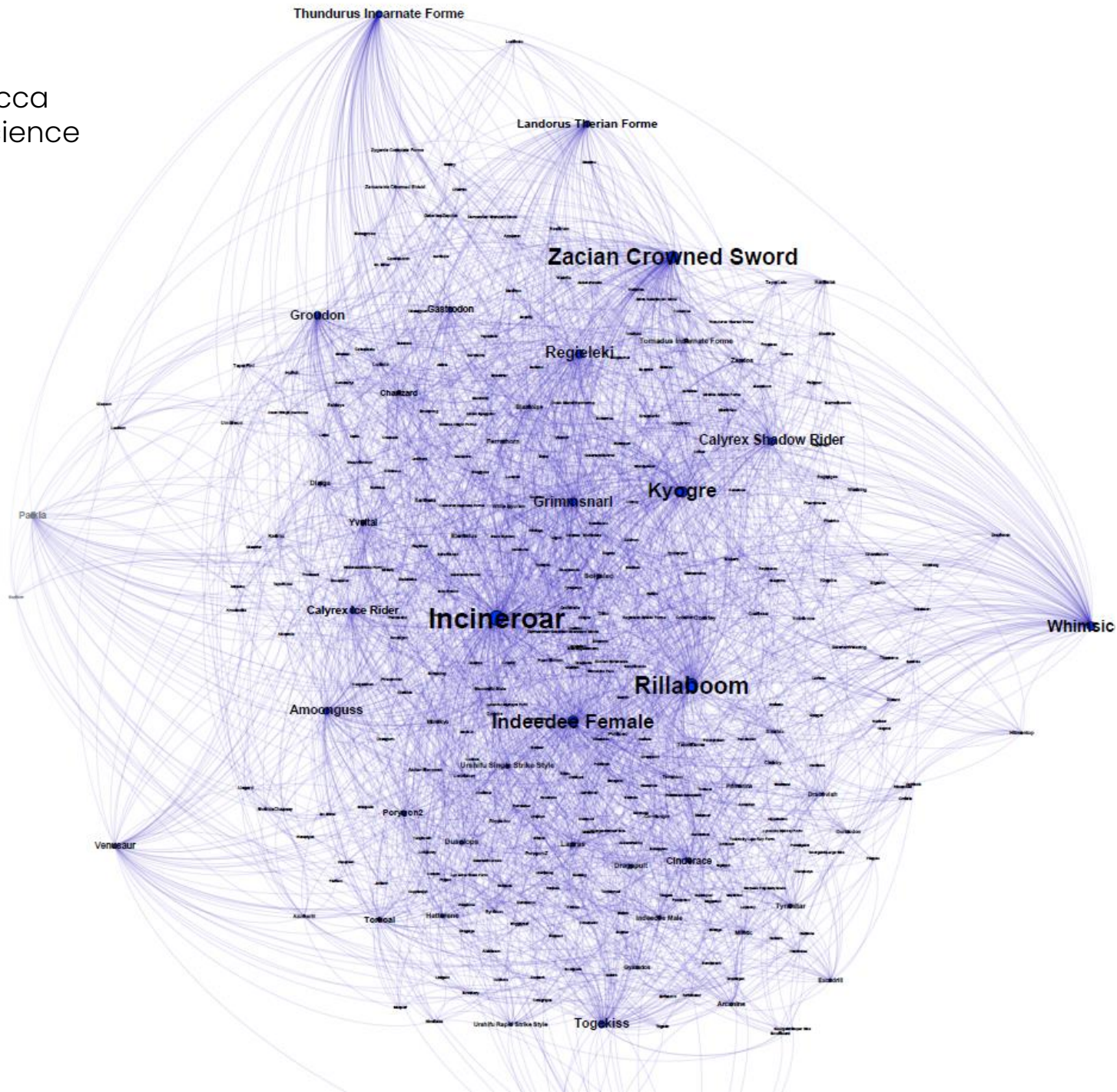
COMPETITIVE **POKÉMON** GRAPH DATABASE

Elaborato di:

Giorgio CARBONE matricola n. 811974

Gianluca CAVALLARO matricola n. 826049

Remo MARCONZINI matricola n. 883256



/ Idea e obiettivi

❑ Contesto:

- Componente competitiva dei videogiochi Pokémon
- Video Game Championship 2022 Series 12 → **Pokémon World Championships** di Londra 2022

❑ Un **database a grafo** per supportare i giocatori del competitivo Pokémon:

- Pokémon, mosse, strumenti, abilità e altri elementi e meccaniche di gioco
- Relazioni tra gli elementi di *gameplay* nei videogiochi
- Relazioni tra i Pokémon e i compagni di squadra, le mosse, gli strumenti, le abilità e le nature ad essi più frequentemente abbinati nei match competitivi online di Marzo 2022



DATA ACQUISITION: **API & WEB SCRAPING**

/ API: PokéAPI

❑ PokéAPI:

- Interfaccia RESTful API
- **Elementi e meccaniche di gioco** della serie di videogiochi Pokémon

❑ Processo implementato:

1. Individuazione API endpoints delle risorse relative all'entità di interesse
2. Creazione di una funzione che passi in modo automatizzato gli endpoints alla funzione `.get()` della libreria **Python Requests**, per inviare le GET request al server
3. La funzione ritorna una lista di dizionari, convertita in un unico **dizionario di dizionari**
4. Salvataggio dizionario di dizionari in formato `.json`



The RESTful Pokémon API

Serving over 250,000,000 API calls each month!

All the Pokémon data you'll ever need in one place,
easily accessible through a modern RESTful API.

/ API: dataset ottenuti

Species API

898 righe × 9 colonne

	ID	Name	Varieties	Generation	Evolves_from	Has_gender_diff	Is_baby	Is_legendary	Is_mythical
0	1	Bulbasaur	{'is_default': True, 'pokemon': {'name': 'bul...	{'name': 'generation-i', 'url': 'https://pokea...	None	False	False	False	False
1	2	Ivysaur	{'is_default': True, 'pokemon': {'name': 'ivy...	{'name': 'generation-i', 'url': 'https://pokea...	{'name': 'bulbasaur', 'url': 'https://pokeapi....	False	False	False	False

Items API

1606 righe × 4 colonne

	Name	Attributes	Category	Effect
0	Master-ball	{'name': 'countable', 'url': 'https://pokeapi...	{'name': 'standard-balls', 'url': 'https://pok...	{'effect': 'Used in battle : Catches a wild...
1	Ultra-ball	{'name': 'countable', 'url': 'https://pokeapi...	{'name': 'standard-balls', 'url': 'https://pok...	{'effect': 'Used in battle : Attempts to ca...

Abilities API

327 righe × 5 colonne

	Name	Generation	Is_Main_Series	Effect_entries	Pokemon
0	Stench	3	True	This Pokémon's damaging moves have a 10% chanc...	{'is_hidden': True, 'pokemon': {'name': 'gloo...
1	Drizzle	3	True	The weather changes to rain when this Pokémon ...	{'is_hidden': True, 'pokemon': {'name': 'poli...

Moves API

327 righe × 8 colonne

	Name	Type	Power	Accuracy	PP	Damage_class	Introduced_in	Learned_by
0	pound	{'name': 'normal', 'url': 'https://pokeapi.co/...	40.0	100.0	35	{'name': 'physical', 'url': 'https://pokeapi.c...	{'name': 'generation-i', 'url': 'https://pokea...	{'name': 'clefairy', 'url': 'https://pokeapi....
1	karate-chop	{'name': 'fighting', 'url': 'https://pokeapi.c...	50.0	100.0	25	{'name': 'physical', 'url': 'https://pokeapi.c...	{'name': 'generation-i', 'url': 'https://pokea...	{'name': 'mankey', 'url': 'https://pokeapi.co...

Types API

20 righe × 3 colonne

	Name	Introduced_in	Damage_relations
15	Dragon	{'name': 'generation-ii', 'url': 'https://pokea...	{'double_damage_from': [{'name': 'ice', 'url': '...
16	Dark	{'name': 'generation-ii', 'url': 'https://poke...	{'double_damage_from': [{'name': 'fighting', '...

/ Web Scraping

❑ Siti *target*

- **Pokémon Database**
informazioni aggiornate, in formato tabellare, sui videogiochi della serie principale
- **Pikalytics**
analisi e statistiche relative ai **match competitivi**
- **Serebii.net**
notizie e informazioni aggiornate relative al mondo Pokémon
- **VictoryRoad**
gestito da organizzatori e giudici del Pokémon VGC. Informazioni, **regolamenti** e *know-how* relativi al competitivo
- **Bulbapedia:**
enciclopedia *community-driven* a tema Pokèmon

❑ **Processo** implementato e **librerie** utilizzate in **Python**

- Esplorazione pagina *target* e ispezione del documento HTML
- Richieste HTTP GET alla pagina target mediante il metodo `.get()` di **Python Requests**
- Parsing, navigazione e estrazione dei dati dai documenti HTML mediante **Beautiful Soup**
- Organizzazione dati in DataFrame **Pandas** e salvataggio in .csv e .JSON
- Se dati strutturati in *HTML Tables* → funzione `pandas.read_html()`
- **Selenium**, con *Microsoft Edge Webdriver*, per simulare l'azione di scrolling su Pikalytics

/ Scraping su Pokémon Database

- ❑ Aspetti tecnici e statistiche relativi a **elementi** e **meccaniche di gioco**
- ❑ **Obiettivo** → integrazione e miglioramento della qualità dei dati ottenuti mediante API
- ❑ Ottenuti **5 dataset** e **una tabella ponte**

Type Chart

Tabella ponte a doppia entrata

18 righe × 19 colonne

	Atk_Type/Pokemon_Type	Normal	Fire	Water	Electric	Grass	Ice	Fighting	Poison	Ground	Flying	Psychic	Bug	Rock	Ghost	Dragon	Dark	Steel
0	Normal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	½	0	NaN	NaN	½
1	Fire	NaN	½	½	NaN	2	2	NaN	NaN	NaN	NaN	NaN	2	½	NaN	½	NaN	2
2	Water	NaN	2	½	NaN	½	NaN	NaN	NaN	2	NaN	NaN	NaN	2	NaN	½	NaN	NaN
3	Electric	NaN	NaN	2	½	½	NaN	NaN	NaN	0	2	NaN	NaN	NaN	NaN	½	NaN	NaN
4	Grass	NaN	½	2	NaN	½	NaN	NaN	½	2	½	NaN	½	2	NaN	½	NaN	½

/ Scraping su Pokémon Database

Complete Pokémon Pokédex

1075 righe × 12 colonne

	#	Name	Type1	Type2	Variant	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	sprite_url
0	1	Bulbasaur	Grass	Poison	None	318	45	49	49	65	65	45	https://img.pokemondb.net/sprites/sword-shield...
1	2	Ivysaur	Grass	Poison	None	405	60	62	63	80	80	60	https://img.pokemondb.net/sprites/sword-shield...

Items

960 righe × 4 colonne

	Name	Category	Effect	sprite_url
0	Ability Capsule	Hold items	A capsule that allows a Pokémon with two Abili...	https://img.pokemondb.net/sprites/items/abilit...
1	Ability Patch	General items	Changes a Pokémon's ability to its Hidden Abil...	https://img.pokemondb.net/s.png

Abilities

266 righe × 4 colonne

	Name	Pokémon	Description	Gen.
0	Adaptability	16	Powers up moves of the same type.	4
1	Aerilate	2	Turns Normal-type moves into Flying-type moves.	6

Moves

865 righe × 9 colonne

	Name	Type	Cat.	Power	Acc.	PP	TM	Effect	Prob. (%)
0	10,000,000 Volt Thunderbolt	Electric	NaN	195	—	1	NaN	Pikachu-exclusive Z-Move. High critical hit ra...	—
1	Absorb	Grass	NaN	20	100	25	NaN	User recovers half the HP inflicted on opponent.	—

Natures

23 righe × 5 colonne)

	Nature	Increases	Decreases	Likes_berry	Dislikes_berry
0	Adamant	Attack	Sp. Atk	Spicy	Dry
1	Bashful	Sp. Atk	Sp. Atk	Dry	Dry

/ Scraping su Pikalytics

- ❑ Acquisizione dei dati relativi ai **match competitivi online** svolti a **Marzo 2022** e disputati sulla piattaforma **Pokémon Showdown!**
- ❑ **Obiettivo** → individuare **rank** di utilizzo di ogni pokémon e le **relazioni** che li legano ai compagni di squadra, gli oggetti, le mosse, le abilità e le nature più frequentemente abbinate
- ❑ Ottenuti **5 tabelle ponte** e un **dataset**

Pokémon Usage

Dataset

342 righe × 4 colonne

	Pokemon	Monthly Usage (k)	Usage Percent (%)	Monthly Rank
0	Zacian-Crowned	1088	65	1
1	Incineroar	1124	59	2
2	Kyogre	744	35	3
3	Grimmsnarl	577	30	4
4	Regieleki	603	29	5

/ Scraping su Pikalytics: Tabelle Ponte

	Pokemon	Teammate	Use_Percentage (%)
0	Zacian-Crowned	Incineroar	54.189%
1	Zacian-Crowned	Kyogre	31.467%

Teammates Usage

Tabella ponte Pokémon/Pokémon (3422 righe × 3 colonne)

Moves Usage

Tabella ponte Pokémon/Moves (3618 righe x 4 colonne)

	Pokemon	Move	Type	Use_Percentage (%)
0	Zacian-Crowned	Behemoth Blade	steel	99.996%
1	Zacian-Crowned	Protect	normal	98.672%

	Pokemon	Item	Use_Percentage (%)
0	Zacian-Crowned	Rusted Sword	100.000%
1	Incineroar	Shuca Berry	36.082%

Items Usage

Tabella Ponte Pokémon/Items (1892 righe × 3 colonne)

Abilities Usage

Tabella Ponte Pokémon/Abilities (773 righe × 3 colonne)

	Pokemon	Ability	Use_Percentage (%)
0	Zacian-Crowned	Intrepid Sword	100.000%
1	Incineroar	Intimidate	99.927%

	Pokemon	Nature	HP/Atk/Def/SpA/SpD/SpE	Use_Percentage (%)
0	Zacian-Crowned	Jolly	0/252/0/0/4/252	10.388%
1	Zacian-Crowned	Adamant	252/28/4/0/12/212	4.504%

Natures Usage

Tabella Ponte Pokémon/Natures (6819 righe × 4 colonne)

/ Scraping su VictoryRoad

- ❑ Acquisizione informazioni dal **regolamento ufficiale** attuale del gioco competitivo (**Series 12 VGC 2022**)
- ❑ Ottenute **7 liste**

Banned Pokémon

Pokémon non ammessi

['Mew', 'Celebi', 'Jirachi', 'Victini', 'Keldeo (both forms)', 'Genesect (all forms)', 'Diancie', <...>

Restricted Pokemon

Ammesso uno per team

[150, 249, 250, 382, 383, 384, 483, 484, 487, 643, 644, 646, 716, 717, 718, 789, 790, 791, 792,

Permitted Pokémon

4 liste. Pokémon ammessi

crowntundra_permitted	209	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
galar_permitted	397	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
isleofarmor_permitted	210	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
national_permitted	48	[243, 244, 245, 252, 253, 254, 255, 256, 257, 258, 259, 260, 380, 381, 480, 4

Allowed Gigantamax

Pokémon di cui è ammessa la trasformazione Gigantamax

['Venusaur', 'Charizard', 'Blastoise', 'Butterfree', 'Pikachu', 'Meowth', 'Machop', 'Gengar',

/ Scraping su serebii.net

- ❑ Parte delle liste ottenute da VictoryRoad fanno riferimento ai Pokémon usando l'ID nei Pokédex Regionali
- ❑ Acquisizione dei tre **Pokédex regionali** di Galar, Isle of Armor e The Crown Tundra

No.	Name	Abilities	HP	Att	Def	S.Att	S.Def	Spd	sprite_url
0 #001	Grookeyサルノリ	Overgrow Grassy Surge	50	65	50	40	40	65	https://www.serebii.net/swordshield/pokemon/s...
1 #002	Thwackeyパチンキー	Overgrow Grassy Surge	70	85	70	55	60	80	https://www.serebii.net/swordshield/pokemon/s...

Galar Pokédex

400 righe × 10 colonne

Isle of Armor Pokédex

211 righe × 10 colonne

No.	Name	Abilities	HP	Att	Def	S.Att	S.Def	Spd	sprite_url
0 #001	Slowpokeヤドン	Gluttony Own Tempo Regenerator	90	65	65	40	40	15	https://www.serebii.net/swordshield/pokemon/s...
1 #002	Slowbroヤドラン	Quick Draw Own Tempo Regenerator	95	100	95	100	70	30	https://www.serebii.net/swordshield/pokemon/s...

No.	Name	Abilities	HP	Att	Def	S.Att	S.Def	Spd	sprite_url
0 #001	Snomユキハミ	Shield Dust Ice Scales	30	25	35	45	30	20	https://www.serebii.net/swordshield/pokemon/s...
1 #002	Frosmothモスノウ	Shield Dust Ice Scales	70	65	60	125	90	65	https://www.serebii.net/swordshield/pokemon/s...

The Crown Tundra Pokédex

210 righe × 10 colonne

/ Scraping su Bulbapedia

- ❑ Alcuni Pokémon possono **trasformarsi temporaneamente** in battaglia in una variante detta **Gigantamax**
- ❑ Ogni Pokémon Gigantamax possiede una **mossa esclusiva** che prende il nome di **G-Max Move**
- ❑ Obiettivo → ottenere i due dataset relativi ai Pokémon Gigantamax e alle G-Max Move

	Max Move	Gigantamax Pokémon	Type	Additional effect	image_url
0	G-Max Vine Lash	Venusaur	Grass	Inflicts damage for four turns on non-Grass-ty...	https://archives.bulbagarden.net/media/upload/...
1	G-Max Wildfire	Charizard	Fire	Inflicts damage for four turns on non-Fire-typ...	https://archives.bulbagarden.net/media/upload/...

G-Max Moves

32 righe × 5 colonne

Gigantamax Pokémon

32 righe × 5 colonne

	Pokémon	Type	Height	G-Max Move	G-Max Move Type
0	Charizard	Fire Flying	91'10" + (28.0+ m)	G-Max Wildfire	Fire
1	Butterfree	Bug Flying	55'09" + (17.0+ m)	G-Max Befuddle	Bug

DATA CLEANSING, **DATA INTEGRATION &** **DATA ENRICHMENT**

/ Obiettivi

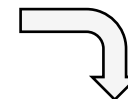
- ❑ Cleansing dei dati
- ❑ Integrare i 23 dataset acquisiti, creando i 6 dataset e le 13 tabelle ponte finali
- ❑ Garantire **consistenza** fra le diverse rappresentazioni delle stesse entità

/ Dataset API e scraping di Pokémon Database

❑ Cleansing dataset API:

- Normalizzazione attributi di tipo **lista** o **dizionario**
- Correzione di alcuni attributi

	ID	Name	Varieties	Generation	Evolves_from	Has_gender_diff	Is_baby	Is_legendary	Is_mythical
0	1	Bulbasaur	[{'is_default': True, 'pokemon': {'name': 'bul...	{'name': 'generation-i', 'url': 'https://pokea...	None	False	False	False	False
1	2	Ivysaur	[{'is_default': True, 'pokemon': {'name': 'ivy...	{'name': 'generation-i', 'url': 'https://pokea...	{'name': 'bulbasaur', 'url': 'https://pokeapi....	False	False	False	False



	ID	Name	Varieties	Generation	Evolves_from	Has_gender_diff	Rarity
0	1	Bulbasaur	[{'is_default': True, 'pokemon': {'name': 'bul...	1	NaN	False	Common
1	2	Ivysaur	[{'is_default': True, 'pokemon': {'name': 'ivy...	1	Bulbasaur	False	Common

❑ Creazione delle tabelle ponte **Abilities-Pokémon** e **Moves-Pokémon**

Abilities-Pokémon

2533 righe × 3 colonne

	Ability	Pokemon	Hidden
0	Adaptability	eevee	False
1	Adaptability	corphish	True

Moves-Pokémon

78182 righe × 2 colonne

	Move	Pokemon
0	Absorb	zubat
1	Absorb	golbat

/ Dataset API e scraping di Pokémon Database

❑ Integrazione fra API e Pokémon Database:

- Utilizzo dell'algoritmo **Levenshtein distance** per il record linkage
- Controllo manuale dei record scartati

Pokémon integrato

1066 righe × 18 colonne

#	Name	Generation	Rarity	Evolves_from	Has_gender_diff	Type1	Type2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Variant	
0	1	Bulbasaur	1	Common	None	False	Grass	Poison	318	45	49	49	65	65	45	NaN
1	2	Ivysaur	1	Common	Bulbasaur	False	Grass	Poison	405	60	62	63	80	80	60	NaN

❑ Estrazione del dataset **G-Max Moves** da **Moves** di Pokémon Database

G-Max Moves

33 righe × 8 colonne

	Name	Type	Power	Acc.	PP	Effect	Prob. (%)	Cat.
275	G Max Befuddle	Bug	—	∞	5	Butterfree-exclusive G-Max Move. Poisons, para...	100	—
276	G Max Cannonade	Water	—	∞	10	Blastoise-exclusive G-Max Move. Damages non-Wa...	—	—

/ Dataset Gigantamax Pokémon e G-Max Moves

❑ Gigantamax Pokémon di Bulbapedia:

- Schema matching rispetto al dataset **Pokémon**
- Concatenazione a **Pokémon**
- Creazione del nuovo attributo **Name (Nome specie + Nome variante)**
- Correzione di alcuni valori

❑ G-Max Moves di Bulpapedia:

- Integrazione con **G-Max Moves di Pokémon Database**
- Concatenazione con **Moves**
- Creazione della tabella ponte **Pokémon-G-Max Moves**

Pokémon-G-Max Moves

33 righe × 2 colonne

	Max Move	Gigantamax Pokémon
0	G-Max Vine Lash	Venusaur Gigantamax
1	G-Max Wildfire	Charizard Gigantamax
2	G-Max Cannonade	Blastoise Gigantamax
3	G-Max Befuddle	Butterfree Gigantamax
4	G-Max Volt Crash	Pikachu Gigantamax

/ Dataset Type e relative tabelle ponte

❑ Cleansing di **Types** API

❑ Conversione di Type Chart (18x18) nella tabella ponte **Type-Type_MOVE_EFFECTIVENESS_ON_POKEMON**

**Type-
Type_MOVE_EFFECTIVENESS_ON_POKEMON**

324 righe × 4 colonne

	Atk. Move Type	Def. Pokemon Type	Damage Multiplier	Effectiveness
0	Normal	Normal	<NA>	Normal (100%)
1	Fire	Normal	<NA>	Normal (100%)
2	Water	Normal	<NA>	Normal (100%)

❑ Creazione delle tabelle ponte **Pokémon-Type_IS_OF_TYPE** e **Moves-Type_MOVE_IS_TYPE**

**Pokémon-
Type_IS_OF_TYPE**

1686 righe × 2 colonne

	Name	Type
0	Bulbasaur	Grass
0	Bulbasaur	Poison
1	Ivysaur	Grass

**Move-
Type_MOVE_IS_TYPE**

821 righe × 2 colonne

	Name	Type
0	Absorb	Grass
1	Accelerock	Rock
2	Acid	Poison

/ Tabelle ponte relative a Pokémon

- ❑ Creazione delle tabelle ponte **Pokémon-Pokémon_EVOLVES_FROM** e **Pokémon-Pokémon_HAS_VARIANT**

**Pokémon-
Pokémon_EVOLVES_
FROM**

555 righe × 2 colonne

	Name	Evolves_from
0	Ivysaur	Bulbasaur
1	Venusaur	Ivysaur
2	Mega Venusaur	Venusaur

**Pokémon-
Pokémon_HAS_VARIANT**

153 righe × 2 colonne

	Non-Variant Pkm Name	Variant Pkm Name
0	Venusaur	Mega Venusaur
1	Venusaur	Venusaur Gigantamax
2	Charizard	Mega Charizard Y

- ❑ Creazione delle tabelle ponte **Pokémon-Abilities_MAY_HAS** e **Pokémon-Moves_MAY_LEARN** (a partire da **Abilities-Pokémon** e **Moves-Pokémon**)

- Correzione dei nomi dei **Pokémon**, eseguita tramite record linkage con algoritmo **q-gram**

**Pokémon-
Abilities_MAY_HAS**

2533 righe × 2 colonne

	Ability	Hidden	Pokemon
0	Adaptability	False	Eevee
1	Anticipation	True	Eevee
2	Run Away	False	Eevee

**Pokémon-
Moves_MAY_LEARN**

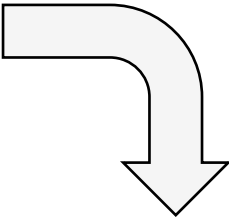
78182 righe × 2 colonne

	Move	Pokemon
0	Absorb	Zubat
1	Absorb	Golbat
2	Absorb	Oddish

/ Integrazione dati del competitivo

- ❑ Manipolazione delle liste acquisite da **VictoryRoad** con ottenimento delle liste **permitted_VGC2022**, **restricted_VGC2022**, **banned_VGC2022** e **gigantamax_allowed_VGC2022**

Name	Size	Value
gigantamax_allowed_list	32	['Venusaur', 'Charizard', 'Blastoise', 'Butterfree', 'Pikachu', 'Meowth', 'Machop', 'Gengar', 'K
banned	14	['Mew', 'Celebi', 'Jirachi', 'Victini', 'Keldeo (both forms)', 'Genesect (all forms)', 'Diancie', <...>
crowntundra_permitted	209	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 2 <...> 0,
galar_permitted	397	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 2 <...> 8,
isleofarmor_permitted	210	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 2 <...> 1,
national_permitted	48	[243, 244, 245, 252, 253, 254, 255, 256, 257, 258, 259, 260, 380, 381, 480, 481, 482, 485, 488,
national_restricted	23	[150, 249, 250, 382, 383, 384, 483, 484, 487, 643, 644, 646, 716, 717, 718, 789, 790, 791, 792,



Name	Value
banned_VGC2022	['Mega Venusaur', 'Mega Charizard Y', 'Mega Chari
gigantamax_allowed_VGC2022	['Venusaur Gigantamax', 'Charizard Gigantamax', 'E
permitted_VGC2022	['Raikou', 'Entei', 'Suicune', 'Treecko', 'Grovyle', 'Sce
restricted_VGC2022	['Mewtwo', 'Lugia', 'Ho-oh', 'Kyogre', 'Groudon', 'Ra

/ Integrazione dati del competitivo

- Integrazione delle informazioni delle 4 liste precedenti e dei dataset di **Pikalytics** ai dataset principali delle rispettive entità
 - Correzione dei nomi di **Pokémon** e **Moves** nei dataset di **Pikalytics**, eseguita tramite record linkage con algoritmo **q-gram**

Pokémon completo

1098 righe × 22 colonne

#	Name	Species	Variant	Generatio	Rarity	Evolves_from	Has_gender_diff	Type1	Type2	Total
1	Bulbasaur	Bulbasaur		1	Common		False	Grass	Poison	318
2	Ivysaur	Ivysaur		1	Common	Bulbasaur	False	Grass	Poison	405
3	Venusaur	Venusaur		1	Common	Ivysaur	True	Grass	Poison	525
3	Mega Venusaur	Venusaur	Mega	6	Common	Venusaur	True	Grass	Poison	625
3	Venusaur Gigantamax	Venusaur	Gigantamax	8		Venusaur		Grass	Poison	
4	Charmander	Charmander		1	Common		False	Fire		309
5	Charmeleon	Charmeleon		1	Common	Charmander	False	Fire		405
6	Mega Charizard Y	Charizard	Mega Y	6	Common	Charizard	False	Fire	Flying	634
6	Charizard Gigantamax	Charizard	Gigantamax	8		Charizard		Fire	Flying	
6	Mega Charizard X	Charizard	Mega X	6	Common	Charizard	False	Fire	Dragon	634
6	Charizard	Charizard		1	Common	Charmeleon	False	Fire	Flying	534

HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	image_url	VGC2022 rules	Monthly Usage (k)	Usage Percent (%)	Monthly Rank
45	49	49	65	65	45	https://img.poker	Permitted			
60	62	63	80	80	60	https://img.poker	Permitted			
80	82	83	100	100	80	https://img.poker	Permitted	204	7	24
80	100	123	122	120	80	https://img.poker	Banned			
						https://archives.b	Gigantamax Allowed			
39	52	43	60	50	65	https://img.poker	Permitted			
58	64	58	80	65	80	https://img.poker	Permitted			
78	104	78	159	115	100	https://img.poker	Banned			
						https://archives.b	Gigantamax Allowed			
78	130	111	130	85	100	https://img.poker	Banned			
78	84	78	109	85	100	https://img.poker	Permitted	344	19	12

DATA QUALITY

/ Currency

- ❑ Dati dello scraping più aggiornati ed affidabili rispetto all'API
- ❑ La **currency** diminuisce mensilmente
 - Aggiornamento mensile dei dati del competitivo
 - La currency diminuisce anche con l'uscita di nuovi giochi della serie

/ Completezza

❑ Presenza di **missing value** ampiamente giustificabile

- Eccezione per l'attributo **Usage_Attributes** di **Items**, particolarmente sparso sin dall'inizio

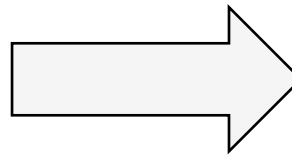
Completezza dei 6 dataset finali

Dataset	% di NaN	Totale di NaN	Totale valori
Pokémon	20,08	4851	24156
Items	14,27	839	5880
Moves	16,74	1237	7389
Abilities	0,12	1	801
Natures	0	0	0
Types	0	0	0

/ Consistenza

- ❑ Incoerenza fra le diverse rappresentazioni delle stesse entità nei 23 dataset di partenza
 - Esecuzione di numerose operazioni per ricondurre tutte le rappresentazioni allo stesso formato
 - Particolare attenzione rivolta al nome dei **Pokémon**, per cui è stato utilizzato il formato **Nome specie + Nome variante**

#	Name	Variant
0	1 Bulbasaur	<NA>
1	2 Ivysaur	<NA>
2	3 Venusaur	<NA>
3	3 Venusaur	Mega Venusaur
4	3 Venusaur	Gigantamax

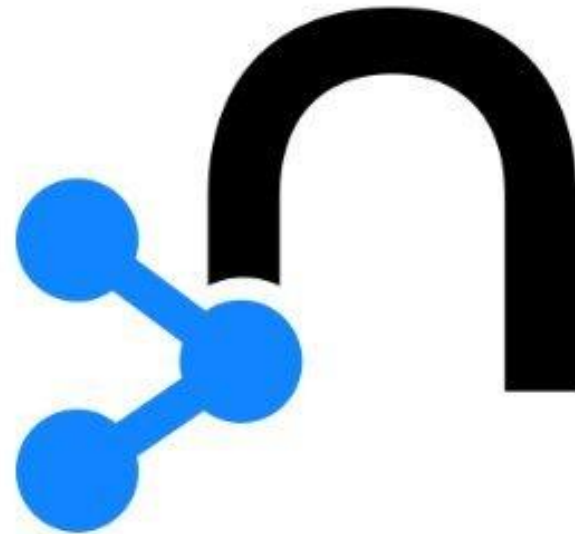


#	Name	Species	Variant
0	1 Bulbasaur	Bulbasaur	<NA>
1	2 Ivysaur	Ivysaur	<NA>
2	3 Venusaur	Venusaur	<NA>
3	3 Mega Venusaur	Venusaur	Mega
4	3 Venusaur Gigantamax	Venusaur	Gigantamax

DATA MODEL

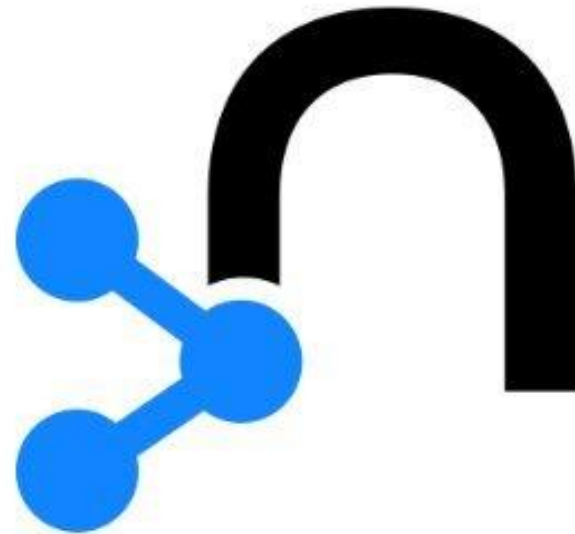
/ Data Model

- ❑ Perché un database a **grafo** ?
 - ❑ Per la natura relazionale dei dati
 - ❑ Per il rapporto tra operazioni di **scrittura** e **lettura**
 - ❑ Costo computazionale delle operazioni di scrittura maggiore di quelle di lettura



/ Neo4J

- ❑ Perché **Neo4J** ?
 - ❑ Piattaforma Open Source
 - ❑ Database a grafo nativo
 - ❑ Database totalmente transazionale

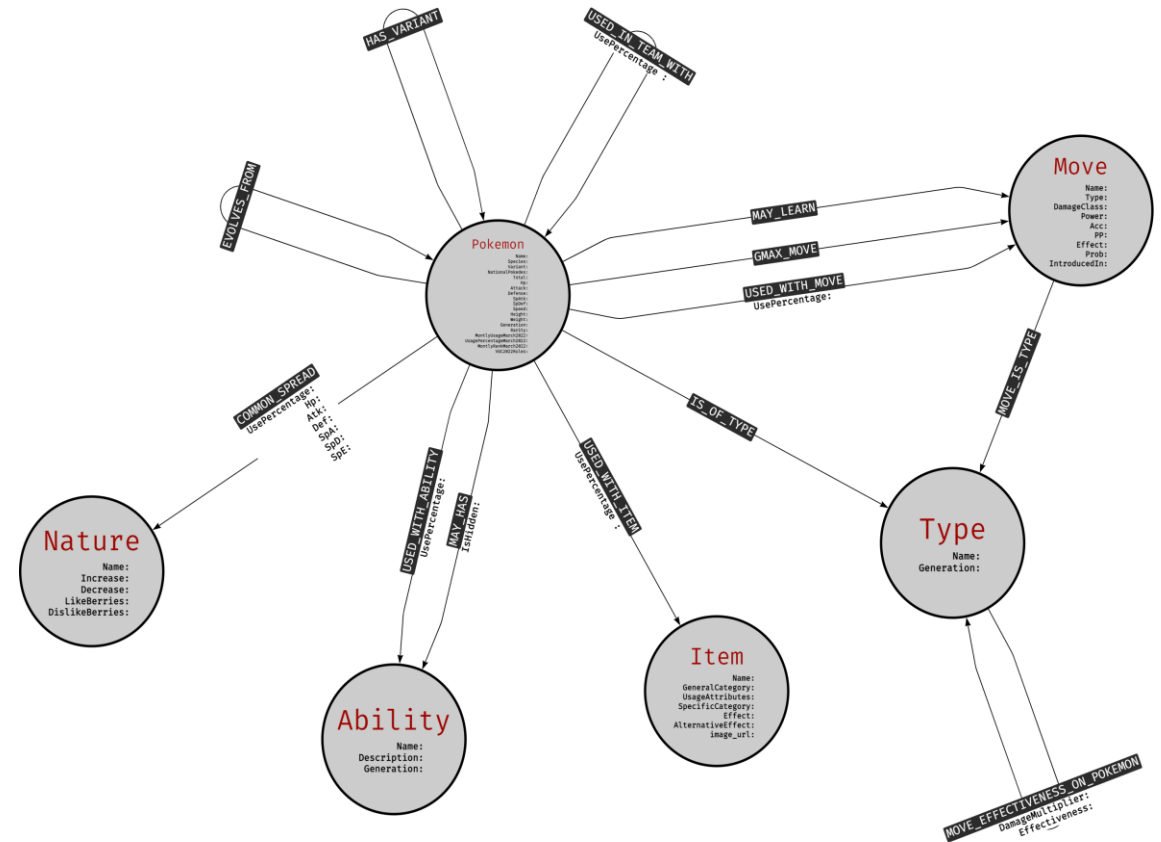


/ Schema 1

❑ La progettazione del DBMS ha visto la definizione di:

❑ 6 tipi di **nodi**: Pokémon, Moves, Items, Natures, Type, Abilities

❑ 13 tipi di **relazioni**: HAS_VARIANT, EVOLVES_FROM, IS_OF_TYPE, MAY_LEARN, G-MAX_MOVE, MAY_HAS, MOVES_IS_TYPE, MOVE_EFFECTIVENESS_ON_POKEMON, USED_IN_TEAM_WITH, USED_WITH_MOVE, USED_WITH_ABILITY, COMMON_SPREAD



/ Schema 2

- ❑ **Vincoli:** vincolo di **unicità** per la proprietà Name.
 - ❑ Consistenza e qualità dei dati
- ❑ **Indici:** indice per la proprietà Name.
 - ❑ Performance delle query

Constraints	
ON (abilities:Abilities) ASSERT (abilities.Name) IS UNIQUE	
ON (item:Item) ASSERT (item.Name) IS UNIQUE	
ON (moves:Moves) ASSERT (moves.Name) IS UNIQUE	
ON (nature:Nature) ASSERT (nature.Name) IS UNIQUE	
ON (type:Type) ASSERT (type.Name) IS UNIQUE	
ON (pokemon:Pokemon) ASSERT (pokemon.Name) IS UNIQUE	

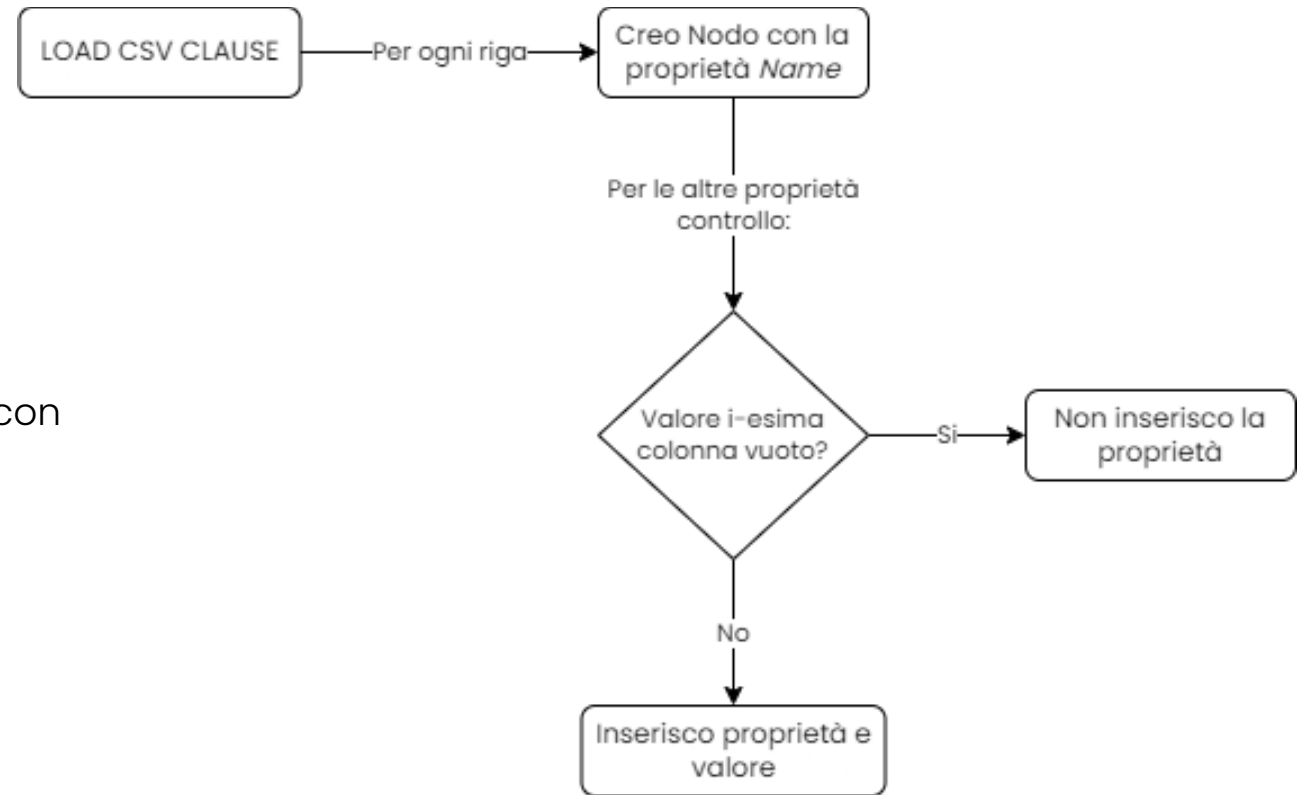
Index Name	Type	Uniqueness	EntityType	LabelsOrTypes	Properties	State
AbilitiesNameKey	BTREE	UNIQUE	NODE	["Abilities"]	["Name"]	ONLINE
ItemNameKey	BTREE	UNIQUE	NODE	["Item"]	["Name"]	ONLINE
MovesNameKey	BTREE	UNIQUE	NODE	["Moves"]	["Name"]	ONLINE
NatureNameKey	BTREE	UNIQUE	NODE	["Nature"]	["Name"]	ONLINE
TypesNameKey	BTREE	UNIQUE	NODE	["Type"]	["Name"]	ONLINE
index_343aff4e	LOOKUP	NONUNIQUE	NODE	[]	[]	ONLINE
index_f770d477	LOOKUP	NONUNIQUE	RELATIONSHIP	[]	[]	ONLINE
pokemonNameKey	BTREE	UNIQUE	NODE	["Pokemon"]	["Name"]	ONLINE

DATA STORAGE

/ Storage 1

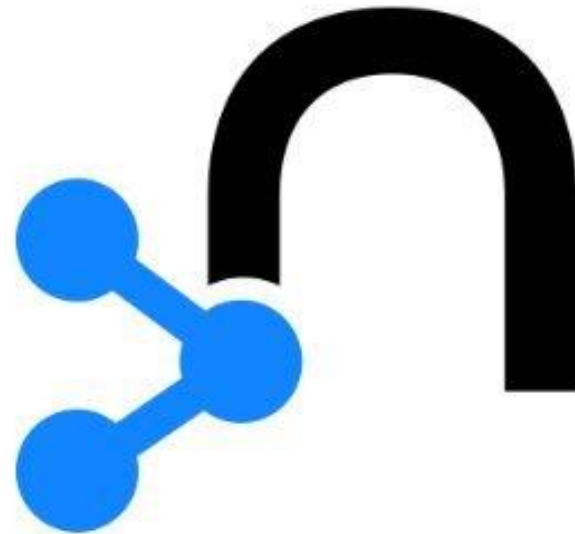
❑ Creazioni **nodi**:

- ❑ Driver Python ufficiale Neo4J
- ❑ Data pre-processing
- ❑ Utilizzo della funzione LOAD CSV integrata con una funzione di controllo



/ Storage 2

- ❑ Creazione **relazioni**:
 - ❑ Driver Python ufficiale Neo4J
 - ❑ Data pre-processing
 - ❑ Utilizzo delle tabelle ponte
 - ❑ MATCH preciso tra nodo di partenza e di arrivo



/ Summary

Node Type	Total	%
Pokemon	1098	35,75%
Moves	821	26,73%
Type	20	0,65%
Item	840	27,35%
Abilities	267	8,69%
Nature	25	0,81%

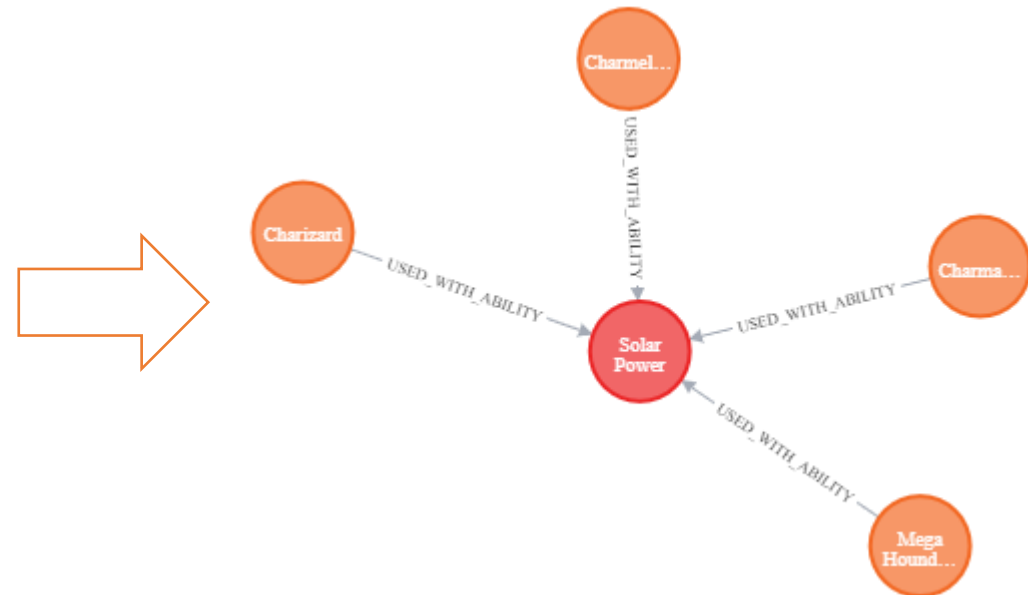
Relationship Type	Total	%
IS_OF_TYPE	3372	1,90%
EVOLVES_INTO	1074	0,60%
MAY_LEARN	136210	76,64%
USED_WITH_ABILITY	4350	2,45%
HAS_VARIANT	306	0,17%
USED_IN_TEAM_WITH	6822	3,84%
USED_WITH_ITEM	3080	1,73%
COMMON_SPREAD	13638	7,67%
USED_WITH_MOVE	6532	3,68%
MAY_LEARN_GMAX	64	0,04%
MOVES_IS_TYPE	1642	0,92%
MOVE_EFFECTIVENESS_ON_POKEMON	630	0,35%

GRAPH EXPLORATION

/ Query 1

- ❑ Quali sono i Pokémon di tipo **Fire** che vengono utilizzati con l'abilità **Solar Power** ?

```
query1 = ''
MATCH (p:Pokemon)-[r:IS_OF_TYPE]-> (t:Type {Name:'fire'})
CALL {
  WITH p
  MATCH (p)-[u:USED_WITH_ABILITY]->(a:Abilities {Name:'Solar Power'})
  RETURN u,a
}
RETURN p.Name AS Pokemon,a.Name AS Abilità, t.Name AS Tipo
''
```

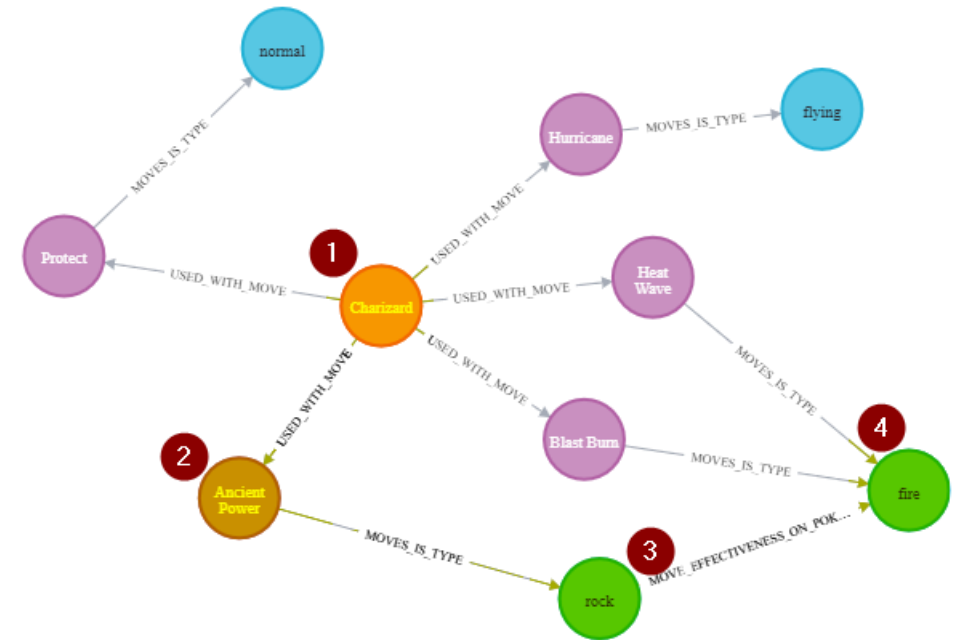
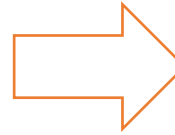


/ Query 2

❑ Quali, tra le mosse più utilizzate dal Pokémon **Charizard**, sono super-efficaci sul Pokémon **Arcanine**?

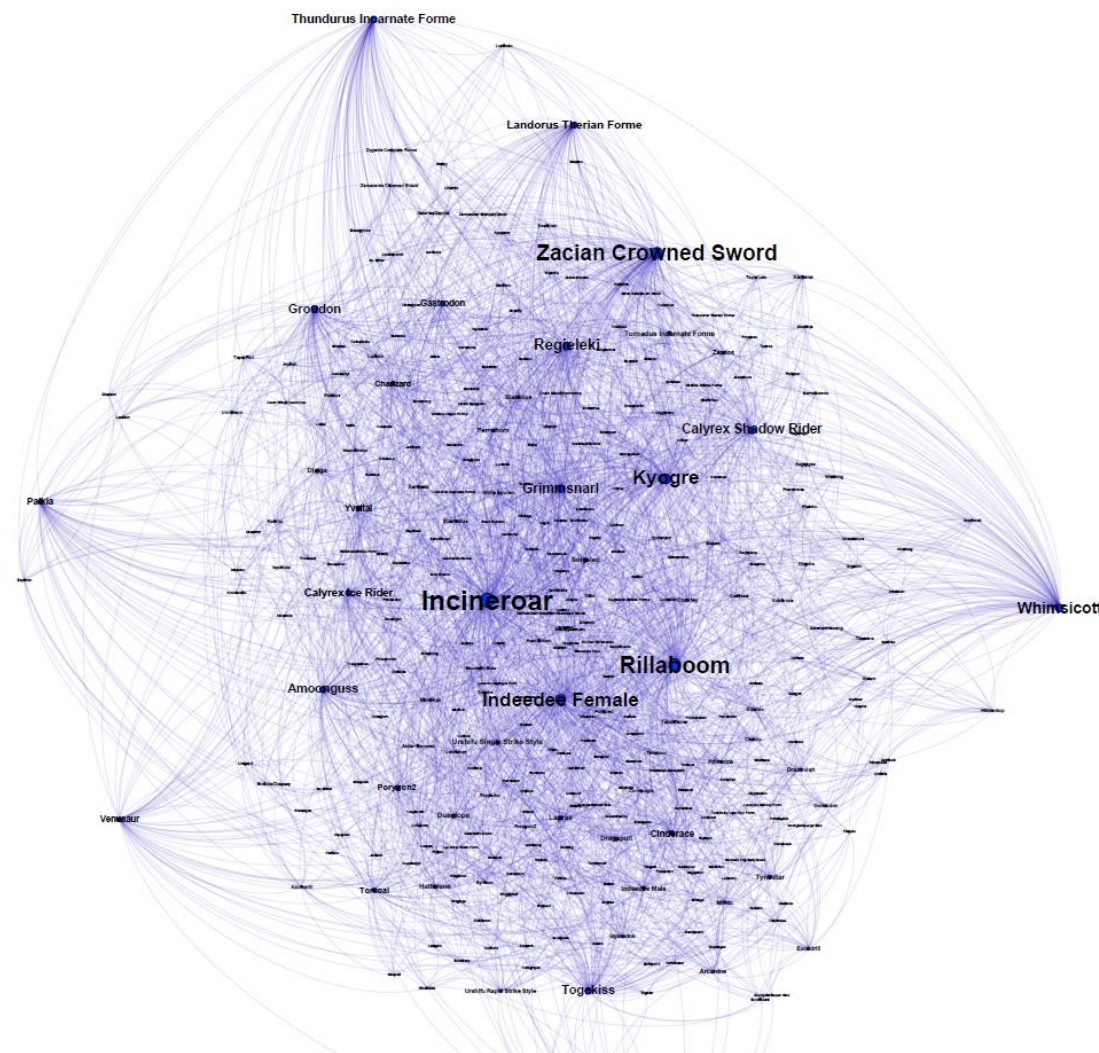
```
# Cerco il tipo del Pokémon Arcaine
query2=''
MATCH (p:Pokemon{Name:'Arcanine'})-[rt:IS_OF_TYPE]-> (t:Type)
RETURN p,rt,t
'''

# Cerco, fra le mosse più utilizzate da Charizard, quali sono quelle
# superefficaci su Pokémon di tipo Fire
query3=''
MATCH (p:Pokemon {Name:'Charizard'})-[r:USED_WITH_MOVE]->(m:Moves)
WHERE r.UsePercentage >30
MATCH (p)-[rt:IS_OF_TYPE]-> (tp:Type)
MATCH (m)-[rm:MOVES_IS_TYPE]->(tm)
CALL {
  WITH tm
  MATCH (t)-[eff:MOVE_EFFECTIVENESS_ON_POKEMON]->(teff:Type)
  WHERE teff.Name='fire' AND eff.Effectiveness='Super-effective (200%)'
  RETURN eff, teff
}
RETURN *
'''
```



/ Competitive Data Graph Visualization

- ❑ Tool **Gephi**
- ❑ Analisi del comportamento della **community**
- ❑ Compagni di squadra più frequenti
- ❑ Label proporzionali al **grado entrante**



Sitografia

- ❑ *PokeAPI: The RESTful Pokémon API*. URL: <https://pokeapi.co/>.
- ❑ *Requests: HTTP for Humans*. URL: <https://requests.readthedocs.io/en/master/>.
- ❑ *Pokémon Database: the fastest way to get your Pokémon information*. URL: <https://pokemondb.net/>.
- ❑ *Pikalytics: VGC 2022 Series 12 Stats Pokédex*. URL: <https://www.pikalytics.com/>.
- ❑ *Serebii.net – Where Legends Come To Life*. URL: <https://www.serebii.net/>.
- ❑ *Victory Road: 2022 Play! Pokémon Season Structure*. URL: <https://victoryroadvgc.com/2020-season/>.
- ❑ *Bulbapedia: the community-driven Pokemon encyclopedia*. URL: https://bulbapedia.bulbagarden.net/wiki/Main_Page.
- ❑ *Beautiful Soup. Libreria per web scraping*. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
- ❑ *html.parser – Simple HTML and XHTML parser*. URL: <https://docs.python.org/3/library/html.parser.html>.
- ❑ *Pandas: flexible and easy to use open source data analysis and manipulation tool*. URL: <https://requests.readthedocs.io/en/master/>.

Sitografia

- ❑ *Selenium with Python*. URL: <https://selenium-python.readthedocs.io/>.
- ❑ *Microsoft Edge WebDriver*. URL: <https://developer.microsoft.com/en-us/microsoft-edge/tools/webdriver/>.
- ❑ *Pokemon Showdown!. A web-based open-source Pokémon battle simulator*. URL: <https://play.pokemonshowdown.com/>.
- ❑ *Python Record Linkage Toolkit: a library to link records in or between data sources*. URL: <https://recordlinkage.readthedocs.io/en/latest/about.html>.
- ❑ *Levenshtein distance*. URL: https://en.wikipedia.org/wiki/Levenshtein_distance.
- ❑ *N-gram similarity and distance*. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.9369&rep=rep1&type=pdf>.
- ❑ *Neo4j Community Server*. URL: <https://neo4j.com/download-center/#community>.
- ❑ *Neo4j Driver*. URL: <https://neo4j.com/docs/api/python-driver/current/>.
- ❑ *LOAD CSV Procedure*. URL: <https://neo4j.com/docs/cypher-manual/current/clauses/loadcsv/>.
- ❑ *The Neo4j Cypher Manual*. URL: <https://neo4j.com/docs/cypher-manual/current/>.
- ❑ *The Py2neo Handbook*. URL: <https://py2neo.org/2021.1/>.
- ❑ *Data pretty printer*. URL: <https://docs.python.org/3/library/pprint.html>.
- ❑ *The Open Graph Viz Platform*. URL: <https://gephi.org/>.