

Master's Degree in Engineering in Computer Science
Sapienza University of Rome
Academic Year 2021-2022

Project Presentation for Data Management



SAPIENZA
UNIVERSITÀ DI ROMA

Course: Data Management
Professor: Prof. Lenzerini Maurizio

Candidate: Carmignani Federico

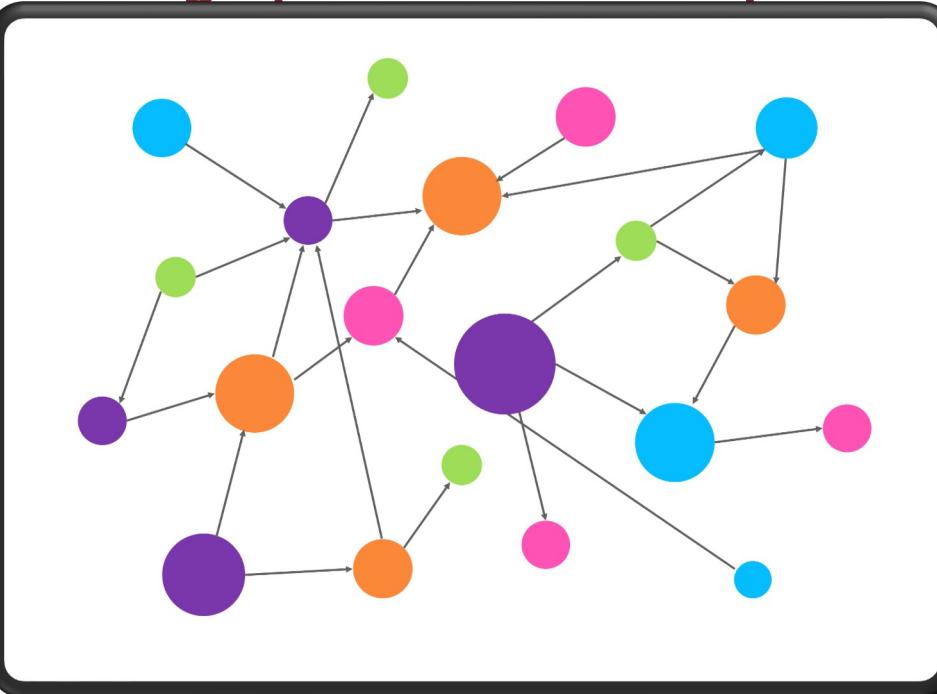
ID number: 1845479

E-mail: carmignani.1845479@studenti.uniroma1.it



GRAPH DATABASES in FOOTBALL ANALYSIS

“How graphs are used in Sports?”





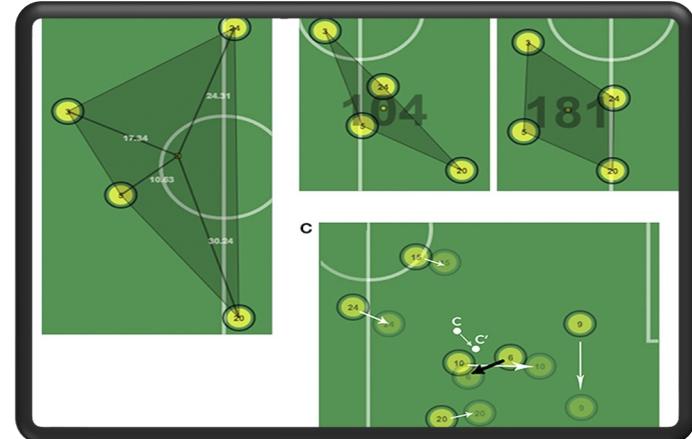
INDEX

Sports Professional Analysis:
Graph Databases

Dataset:
Composition and Content

Tool:
Neo4j and its usage

Experiments:
Applications to Football Analysis and a
quick Demo



OUR GOAL

Take a NoSQL database showing a relevant usage of them over the selected dataset.



GRAPH DATABASES IN SPORTS ANALYSIS

A FACTOR OF GROWTH

Graph databases will drive the next generation of:



- Sports Analytics
- Data Science



«Moneyball» by Michael Lewis as an inspiration story



Money

VS



Analysis

IMPROVEMENT

- Analytics are now important in **Sport Analysis**.
- **Private Departments** in Sports team: Football, Basketball and many others.
- The goal is to adopt **behaviours** or **actions** leading to higher probability of success or anticipating scenarios with the aim of **strategy**.



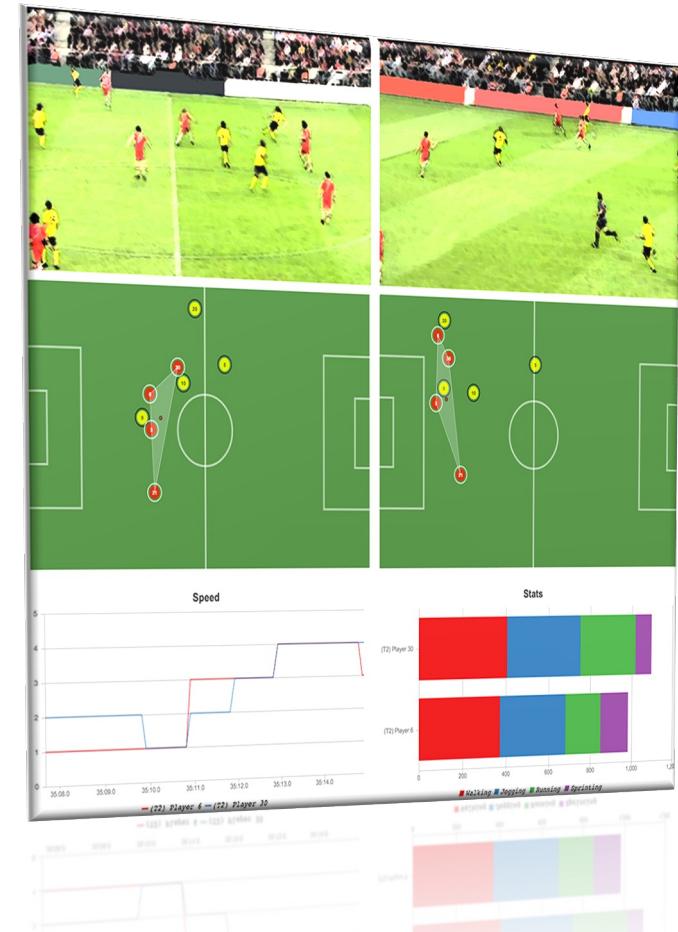
WHY GRAPH DATABASES IN SPORT ANALYTICS?

1. Organization
2. Interaction
3. Extraction of Value



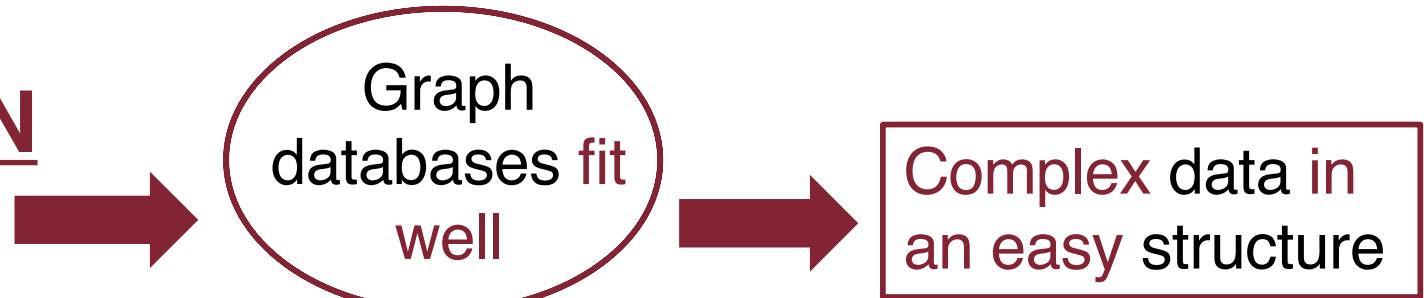
Improving in Sports Data

- Event-based
- Temporal relevance

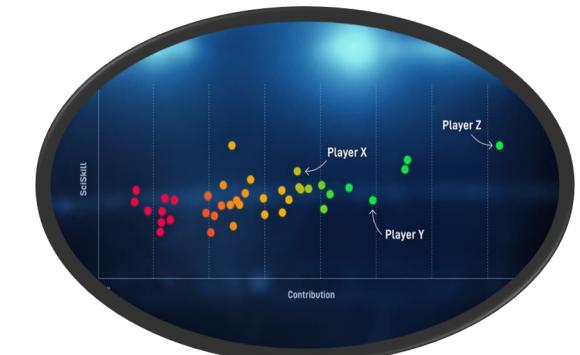




STATISTICS IN SPORTS



Used to maximize quality for analysis



Player can move from a league to another

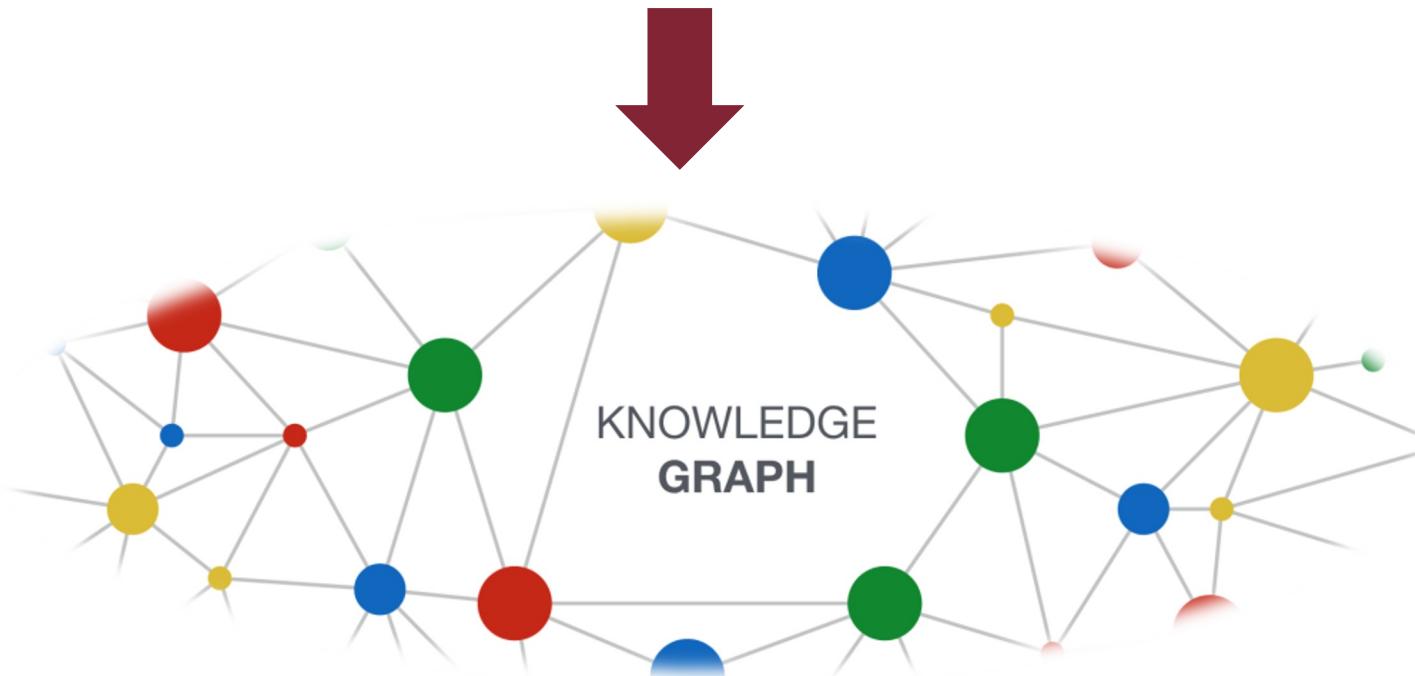
EVENT TYPES:

Players can get injured

Players can fill different positions in the squad



Hard to organize data in a scalable and transparent way



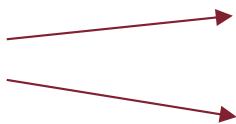
Graph-schema enables *queries* and **aggregations** not simple in a traditional data store (**RDBMS**).

Joins are costly whereas here it is possible to go from a relation to another in **Constant Time**.



ADVANTAGES:

- Complex queries



*Has the player been used in the
correct position during the match?*

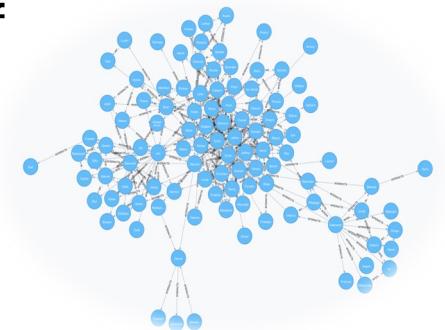
*Which players played together in
Serie A team AS Roma?*

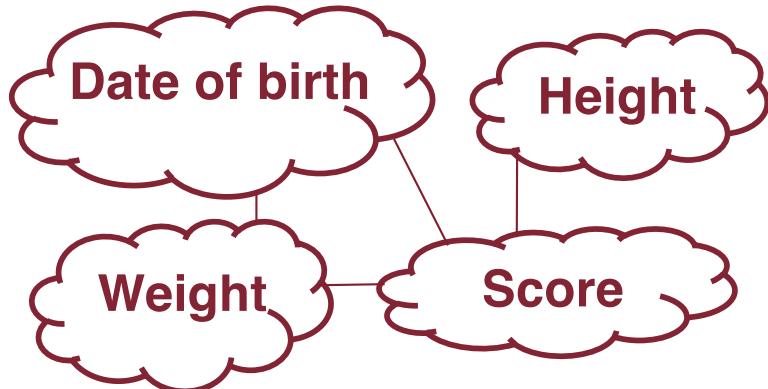


- Add properties to each node and edge (relationships)



- Pattern matching to find sequences (e.g. a sequence of actions leading to a goal)





It is possible to answer to:

«Players under 25 with the best winning rate in Serie A?»

Flexibility

Power

Complex questions involving relationships with every aspect of the data





Let's think on obtaining the players of a team

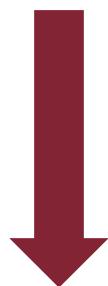
Constant Time in
a Graph



But the schema can be deep

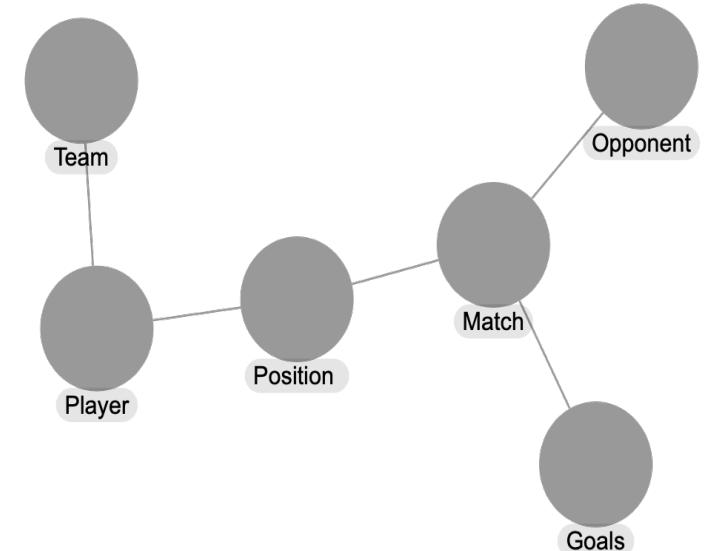


Player's
performance



Player's
opponents

Transformative and valuable
way for Sports Data:
numbers and relationships

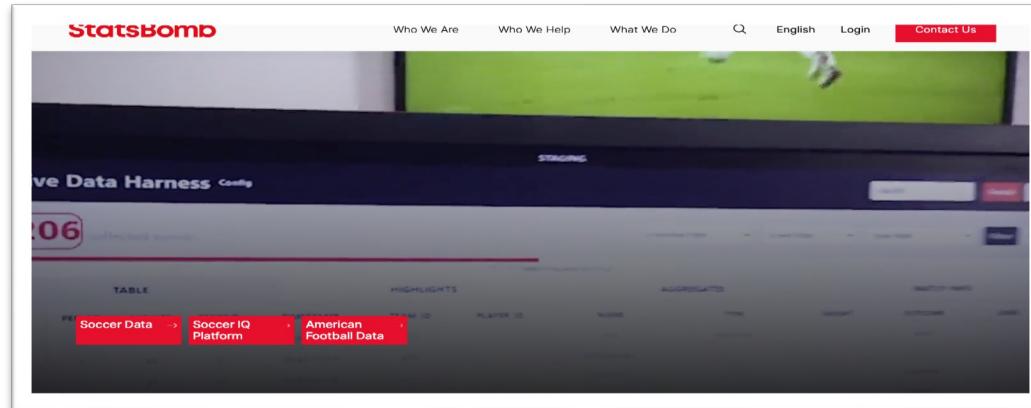




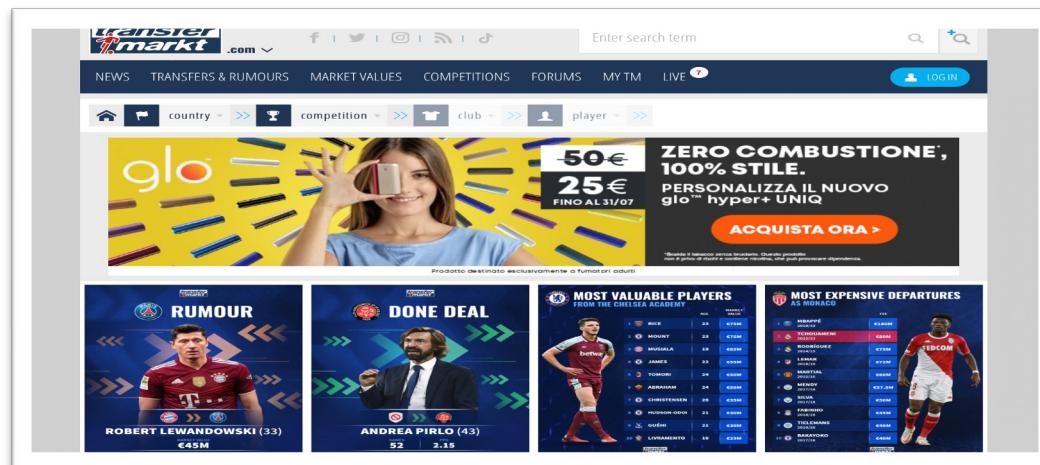
DATASET: COMPOSITION AND CONTENT

The dataset used in this project is composed of 2 main open-source (sub)datasets:

1. StatsBomb dataset



2. Transfermarkt dataset





- The goal is to import and analyse **football transfers**.
 - A bunch of transfers has been inserted into a **CSV file** to make practice with this dataset.

It is possible to retrieve it using the following URL:



https://s3-eu-west-1.amazonaws.com/football-transfers.neo4j.com/transfers-all.csv



Transfermarkt dataset structure as a CSV file:



100% ▾

Visualizza Ridimensiona Aggiungi categoria Inserisci Tabella Grafico Testo Forma Multimedia Commento Collabora Formattazione Organizza

Foglio 1

season	playerUri	playerName	playerPosition	playerAge	sellerClubUri	sellerClubName	sellerClubCountry	buyerClubUri
2009/2010	/douglas-costa/profil/spieler/75615	Douglas Costa	Left Wing	19	/gremio-foot-ball-porto-allegrense/startseite/verein/210	Grêmio	Brazil	/shakhtar-donetsk/startseite/verein/660
2009/2010	/florent-sinama-pongolle/profil/spieler/4216	Florent Sinama-Pongolle	Centre-Forward	25	/atletico-madrid/startseite/verein/13	Atlético Madrid	Spain	/sporting-lissabon/startseite/verein/336
2009/2010	/keisuke-honda/profil/spieler/66521	Keisuke Honda	Right Wing	23	/vvv-venlo/startseite/verein/1426	VVV-Venlo	Netherlands	/zaka-moskau/startseite/verein/2410
2009/2010	/alex-teixeira/profil/spieler/66515	Alex Teixeira	Attacking Midfield	19	/vasco-da-gama-rio-de-janeiro/startseite/verein/978	Vasco da Gama	Brazil	/shakhtar-donetsk/startseite/verein/660
2009/2010	/younes-kaboul/profil/spieler/27114	Younès Kaboul	Centre-Back	23	/fc-portsmouth/startseite/verein/1020	Portsmouth FC	England	/tottenham-hotspur/startseite/verein/14
2009/2010	/alessandro-budel/profil/spieler/25314	Alessandro Budel	Defensive Midfield	28	/fc-parma/startseite/verein/130	Parma	Italy	/brescia-calcio/startseite/verein/19
2009/2010	/daniele-vantaggiato/profil/spieler/22155	Daniele Vantaggiato	Centre-Forward	25	/fc-parma/startseite/verein/130	Parma	Italy	/ac-padova/startseite/verein/3037
2009/2010	/rever/profil/spieler/72489	Réver	Centre-Back	24	/gremio-foot-ball-porto-allegrense/startseite/verein/210	Grêmio	Brazil	/vfl-wolfsburg/startseite/verein/82
2009/2010	/aleksandr-samedov/profil/spieler/15076	Aleksandr Samedov	Right Midfield	25	/fk-moskau/startseite/verein/5268	FK Moskau	Russia	/dinamo-moskau/startseite/verein/121
2009/2010	/eder-luis/profil/spieler/53078	Éder Luís	Right Wing	24	/atletico-mineiro/startseite/verein/330	Atlético-MG	Brazil	/benfica-lissabon/startseite/verein/294
2009/2010	/ererton/profil/spieler/76120	Everton	Left Wing	21	/desportivo-brasil-rla-sp-/startseite/verein/17300	Desp. Brasil	Brazil	/tigres-uanl/startseite/verein/7055
2009/2010	/christian-gimenez/profil/spieler/28085	Christian Giménez	Attacking Midfield	28	/cf-pachucua/startseite/verein/4035	CF Pachuca	Mexico	/cd-cruz-azuil/startseite/verein/3711
2009/2010	/mario-boiatti/profil/spieler/52343	Mario Boiatti	Defensive Midfield	24	/fc-porto/startseite/verein/720	FC Porto	Portugal	/ac-florenz/startseite/verein/430
2009/2010	/jonathan-pereira/profil/spieler/53239	Jonathan Pereira	Centre-Forward	22	/fc-villarreal/startseite/verein/1050	Villarreal CF	Spain	/real-betis-sevilla/startseite/verein/150
2009/2010	/michel/profil/spieler/57818	Michel	Defensive Midfield	24	/sporting-gijón/startseite/verein/2448	Sporting Gijón	Spain	/birmingham-city/startseite/verein/337
2009/2010	/jao-pereira/profil/spieler/13217	João Pereira	Right-Back	25	/sc-braga/startseite/verein/1075	Braga	Portugal	/sporting-lissabon/startseite/verein/336
2009/2010	/felipe/profil/spieler/17127	Felipe	Centre-Back	25	/udinese-calcio/startseite/verein/410	Udinese Calcio	Italy	/ac-florenz/startseite/verein/430
2009/2010	/ruben-micael/profil/spieler/84491	Ruben Micael	Attacking Midfield	23	/cd-nacional/startseite/verein/982	Nacional	Portugal	/fc-porto/startseite/verein/720
2009/2010	/jackson-martinez/profil/spieler/74418	Jackson Martínez	Centre-Forward	23	/independiente-medellin/startseite/verein/10093	Independ. Medellín	Colombia	/chiapas-fc/startseite/verein/4774
2009/2010	/airton/profil/spieler/772861	Airton	Defensive Midfield	19	/flamengo-rio-de-janeiro/startseite/verein/614	Flamengo	Brazil	/benfica-lissabon/startseite/verein/294
2009/2010	/dario-dainelli/profil/spieler/5970	Dario Dainelli	Centre-Back	30	/ac-florenz/startseite/verein/430	Florentina	Italy	/fc-genoa-1893/startseite/verein/252
2009/2010	/alex-geijo/profil/spieler/15981	Álex Geijo	Centre-Forward	27	/racing-santander/startseite/verein/630	Racing	Spain	/udinese-calcio/startseite/verein/410
2009/2010	/alexandru-epureanu/profil/spieler/45340	Alexandru Epureanu	Centre-Back	23	/fk-moskau/startseite/verein/5268	FK Moskau	Russia	/dinamo-moskau/startseite/verein/121
2009/2010	/alan-kardec/profil/spieler/53922	Alan Kardé	Centre-Forward	20	/vasco-da-gama-rio-de-janeiro/startseite/verein/978	Vasco da Gama	Brazil	/benfica-lissabon/startseite/verein/294
2009/2010	/sung-yong-ki/profil/spieler/81796	Sung-Yong Ki	Defensive Midfield	20	/fc-seoul/startseite/verein/6500	FC Seoul	Korea, South	/celtic-glasgow/startseite/verein/371
2009/2010	/edin-cocalic/profil/spieler/46205	Edin Cocalic	Centre-Back	22	/fk-zeljeznica-sarajevo/startseite/verein/2573	Zeljeznicar	Bosnia-Herzegovina	/panionios-athen/startseite/verein/169
2009/2010	/jared-jeffrey/profil/spieler/69426	Jared Jeffrey	Central Midfield	19	/fc-brugge/startseite/verein/2282	Club Brugge	Belgium	/1-fsv-mainz-05/startseite/verein/39
2009/2010	/mattias-adelstam/profil/spieler/75447	Mattias Adelstam	Centre-Forward	27	/angelholms-ff/startseite/verein/3625	Angelholm	Sweden	/treleborgs-ff/startseite/verein/701
2009/2010	/david-silva/profil/spieler/86759	David Silva	Left Midfield	23	/zkska-sofia/startseite/verein/208	CSKA Sofia	Bulgaria	/cd-castellon/startseite/verein/2502
2009/2010	/rafael-caroca/profil/spieler/67917	Rafael Caroca	Defensive Midfield	20	/spartak-moskau/startseite/verein/232	Spartak Moscow	Russia	/vasco-da-gama-rio-de-janeiro/startseite/verein/978
2009/2010	/germano-valiati/profil/spieler/2010	Germano Valiati	Keeper	29	/fc-sion/startseite/verein/321	FC Sion	Switzerland	/fc-st-gallen/startseite/verein/257
2009/2010	/marko-popovic/profil/spieler/28634	Marko Popović	Centre-Back	27	/mk-maribor/startseite/verein/790	NK Maribor	Slovenia	/ms-ashdod/startseite/verein/6105
2009/2010	/andreas-samaris/profil/spieler/140032	Andreas Samaris	Defensive Midfield	20	/panachaiki-patra/startseite/verein/1837	Panachaiki	Greece	/panionios-athen/startseite/verein/169
2009/2010	/bartosz-iwan/profil/spieler/14897	Bartosz Iwan	Attacking Midfield	25	/gks-katowice/startseite/verein/6456	GKS Katowice	Poland	/piast-glivice/startseite/verein/6112
2009/2010	/sebastian-andersen/profil/spieler/62445	Sebastian Andersen	Right Midfield	21	/hb-koge/startseite/verein/23297	HB Køge	Denmark	/esbjerg-fb/startseite/verein/3426
2009/2010	/ivan-djokovic/profil/spieler/35984	Ivan Djoković	Centre-Back	27	/borac-cacak/startseite/verein/2209	Borac Čačak	Serbia	/mfk-kosice/startseite/verein/779
2009/2010	/miguel-garcia/profil/spieler/14514	Miguel García	Right-Back	26	/sporting-olhanense/startseite/verein/4750	Olhãoense	Portugal	/sc-braga/startseite/verein/1075
2009/2010	/radenko-kamberovic/profil/spieler/17198	Radenko Kamberović	Right-Back	26	/fk-sloboda-uzice/startseite/verein/11066	Sloboda Užice	Serbia	/fk-partizan-belgrad/startseite/verein/8

Nome foglio Foglio 1

Sfondo

Duplica foglio

Elimina foglio



Transfermarkt dataset structure as a CSV file:



100% Visualizza Ridimensiona Aggiungi categoria

Inserisci Tabella Grafico Testo Forma Multimedia Commento Collabora Formattazione Organizza

Foglio 1

buyerClubName	buyerClubCountry	transferUri	transferFee	playerImage	playerNationality	timestamp
Shakhtar D.	Ukraine	/jumplist/transfers/spieler/75615/transfer_id/376095	£7.20m	https://tmssl.akamaized.net/images/portrait/medium/75615-1471439432.JPG?lm=1471439625	Brazil	1262304000
Sporting CP	Portugal	/jumplist/transfers/spieler/4216/transfer_id/372427	£5.85m	https://tmssl.akamaized.net/images/portrait/medium/4216-1448635837.jpg?lm=1448635855	France	1262304000
CSKA Moscow	Russia	/jumplist/transfers/spieler/66521/transfer_id/373312	£5.40m	https://tmssl.akamaized.net/images/portrait/medium/66521-1455538386.jpg?lm=1455538398	Japan	1262304000
Shakhtar D.	Ukraine	/jumplist/transfers/spieler/66515/transfer_id/368828	£5.40m	https://tmssl.akamaized.net/images/portrait/medium/66515-1454413290.jpg?lm=1454413353	Brazil	1262304000
Spurs	England	/jumplist/transfers/spieler/27114/transfer_id/385673	£5.31m	https://tmssl.akamaized.net/bilder/spielerotos/s_27114_148_2012_1.jpg?lm=0	France	1262304000
Brescia	Italy	/jumplist/transfers/spieler/25314/transfer_id/380060	£4.50m	https://tmssl.akamaized.net/images/portrait/medium/25314-1470211564.jpg?lm=1470211585	Italy	1262304000
Padova	Italy	/jumplist/transfers/spieler/22155/transfer_id/380226	£4.50m	https://tmssl.akamaized.net/images/portrait/medium/22155-1488543409.jpg?lm=1488543418	Italy	1262304000
VfL Wolfsburg	Germany	/jumplist/transfers/spieler/72489/transfer_id/384674	£4.50m	https://tmssl.akamaized.net/images/portrait/medium/72489-1447749786.jpg?lm=1447749805	Brazil	1262304000
Dinamo Moscow	Russia	/jumplist/transfers/spieler/15076/transfer_id/366510	£4.05m	https://tmssl.akamaized.net/images/portrait/medium/15076-1464945350.jpg?lm=1464945365	Russia	1262304000
Bentita	Portugal	/jumplist/transfers/spieler/53078/transfer_id/372793	£3.60m	https://tmssl.akamaized.net/bilder/spielerotos/s_53078_978_2012_2.jpg?lm=0	Brazil	1262304000
Tigres	Mexico	/jumplist/transfers/spieler/76120/transfer_id/376525	£3.60m	https://tmssl.akamaized.net/bilder/spielerotos/s_76120_679_2012_1.jpg?lm=0	Brazil	1262304000
CD Cruz Azul	Mexico	/jumplist/transfers/spieler/28085/transfer_id/368270	£3.15m	https://tmssl.akamaized.net/images/portrait/medium/28085-1488301496.jpg?lm=1488301510	Mexico	1262304000
Florentina	Italy	/jumplist/transfers/spieler/52343/transfer_id/380643	£3.15m	https://tmssl.akamaized.net/bilder/spielerotos/s_52343_6600_2010_3.jpg?lm=0	Argentina	1262304000
Real Betis	Spain	/jumplist/transfers/spieler/53239/transfer_id/380020	£3.06m	https://tmssl.akamaized.net/images/portrait/medium/53239-1447254811.jpg?lm=1447254887	Spain	1262304000
Birmingham	England	/jumplist/transfers/spieler/57818/transfer_id/377746	£2.97m	https://tmssl.akamaized.net/bilder/spielerotos/s_57818_3709_2013_10_17_1.jpg?lm=0	Spain	1262304000
Sporting CP	Portugal	/jumplist/transfers/spieler/13217/transfer_id/368424	£2.70m	https://tmssl.akamaized.net/images/portrait/medium/13217-1502397946.png?lm=1502398056	Portugal	1262304000
Florentina	Italy	/jumplist/transfers/spieler/17127/transfer_id/372702	Loan fee: £2.70m	https://tmssl.akamaized.net/images/portrait/medium/17127-1449131926.jpg?lm=1449131969	Brazil	1262304000
FC Porto	Portugal	/jumplist/transfers/spieler/84491/transfer_id/379939	£2.70m	https://tmssl.akamaized.net/images/portrait/medium/84491-1412952690.jpg?lm=1433145011	Portugal	1262304000
Chiapas FC	Mexico	/jumplist/transfers/spieler/77441/transfer_id/379478	£2.52m	https://tmssl.akamaized.net/images/portrait/medium/74418-1447152748.jpg?lm=1447152788	Colombia	1262304000
Benfica	Portugal	/jumplist/transfers/spieler/72861/transfer_id/371494	£2.34m	https://tmssl.akamaized.net/bilder/spielerotos/s_72861_6600_2012_2.jpg?lm=0	Brazil	1262304000
Genoa	Italy	/jumplist/transfers/spieler/59707/transfer_id/377974	£2.25m	https://tmssl.akamaized.net/images/portrait/medium/59707-1455535333.jpg?lm=1455535351	Italy	1262304000
Udinese Calcio	Italy	/jumplist/transfers/spieler/15981/transfer_id/386706	£2.25m	https://tmssl.akamaized.net/images/portrait/medium/15981-1474109931.png?lm=1474109943	Spain	1262304000
Dinamo Moscow	Russia	/jumplist/transfers/spieler/45340/transfer_id/366505	£2.25m	https://tmssl.akamaized.net/images/portrait/medium/45340-1474384609.jpg?lm=1474384622	Moldova	1262304000
Benfica	Portugal	/jumplist/transfers/spieler/53922/transfer_id/369461	£2.25m	https://tmssl.akamaized.net/bilder/spielerotos/s_53922_1023_2012_1.jpg?lm=0	Brazil	1262304000
Celtic	Scotland	/jumplist/transfers/spieler/81796/transfer_id/368815	£2.16m	https://tmssl.akamaized.net/images/portrait/medium/81796-1469701725.jpg?lm=1469701745	Korea, South	1262304000
Panionios	Greece	/jumplist/transfers/spieler/46205/transfer_id/381107	£0.98k	https://tmssl.akamaized.net/images/portrait/medium/46205-1484750736.jpg?lm=1484750755	Bosnia-Herzegovina	1262304000
1.FSV Mainz 05	Germany	/jumplist/transfers/spieler/69426/transfer_id/373065	£0.98k	https://tmssl.akamaized.net/images/portrait/medium/69426-1489536808.jpg?lm=1489536931	United States	1262304000
Trelleborg	Sweden	/jumplist/transfers/spieler/75447/transfer_id/380208	£0.93k	https://tmssl.akamaized.net/bilder/spielerotos/s_75447_701_2010_1.jpg?lm=0	Sweden	1262304000
CD Castellón	Spain	/jumplist/transfers/spieler/86759/transfer_id/379816	Loan fee: £63k	https://tmssl.akamaized.net/bilder/spielerotos/s_86759_515_2014_03_24_1.jpg?lm=0	Cape Verde	1262304000
Vasco da Gama	Brazil	/jumplist/transfers/spieler/67917/transfer_id/371104	Loan fee: £63k	https://tmssl.akamaized.net/images/portrait/medium/67917-1454587236.jpg?lm=1454587195	Brazil	1262304000
FC St. Gallen	Switzerland	/jumplist/transfers/spieler/20104/transfer_id/375202	£59k	https://tmssl.akamaized.net/images/portrait/medium/20104-1473155068.jpg?lm=1473155086	Switzerland	1262304000
FC Ashdod	Israel	/jumplist/transfers/spieler/28634/transfer_id/385484	£59k	https://tmssl.akamaized.net/bilder/spielerotos/s_28634_790_2009_1.jpg?lm=0	Serbia	1262304000
Panionios	Greece	/jumplist/transfers/spieler/140032/transfer_id/478846	£54k	https://tmssl.akamaized.net/images/portrait/medium/140032-1474546790.jpg?lm=1474546798	Greece	1262304000
Piast Gliwice	Poland	/jumplist/transfers/spieler/14897/transfer_id/377673	£50k	https://tmssl.akamaized.net/bilder/spielerotos/s_14897_428_2013_09_06_1.jpg?lm=0	Poland	1262304000
Esbjerg fB	Denmark	/jumplist/transfers/spieler/62445/transfer_id/378400	£50k	https://tmssl.akamaized.net/bilder/spielerotos/s_62445_3426_2011_1.jpg?lm=0	Denmark	1262304000
Kosice	Slovakia	/jumplist/transfers/spieler/35984/transfer_id/388245	£45k	https://tmssl.akamaized.net/bilder/spielerotos/s_35984_779_2011_1.jpg?lm=0	Serbia	1262304000
Braga	Portugal	/jumplist/transfers/spieler/14514/transfer_id/373009	£45k	https://tmssl.akamaized.net/bilder/spielerotos/s_14514_1075_2010_2_02_1.jpg?lm=0	Portugal	1262304000
Partizan	Serbia	/jumplist/transfers/spieler/17198/transfer_id/1521784	£45k	https://tmssl.akamaized.net/bilder/spielerotos/s_17198_669_2009_1.jpg?lm=0	Serbia	1262304000
Anag. Karditsas	Greece	/jumplist/transfers/spieler/25835/transfer_id/381418	£45k	https://tmssl.akamaized.net/bilder/spielerotos/s_25835_5572_2012_1.jpg?lm=0	Brazil	1262304000



- StatsBomb are committed to **sharing new data and research publicly** to enhance understanding of the game of Football. They want to actively encourage new research and analysis at all levels, making certain leagues of StatsBomb Data freely available for **public use for research projects** and genuine interest in football analytics.
- The **data** is provided as **JSON files** exported from the **StatsBomb Data API**, in the following structure:
 - **Competition and seasons** stored in competitions.json.
 - **Matches for each competition and season**, stored in matches.
 - Each **folder** within is named for a competition ID, each **file** is named for a season ID within that competition.
 - **Events and lineups for each match**, stored in events and lineups respectively.
 - Each **file** is named for a match ID.
 - **StatsBomb 360 data for selected matches**, stored in three-sixty.
 - Each **file** is named for a match ID.
- Some **documentation** about the meaning of different events and the **format** of the **JSON** can be found in the doc directory.



StatsBomb open-source dataset at:



<https://github.com/statsbomb/open-data>

Pagina 1 di 45

STATSBOMB

StatsBomb Data Specification v1.1

[StatsBomb Data Specification - last updated 13 May 2019]

Change Log - Updates for Data Version 1.1

Competition Data Updates

- "Gender" added for all competitions.

Match Data Updates

- Team section now includes variables for: "Gender", "Group", "Country" and "Managers"
- "Match Week" added as an integer denoting the match week in the competition.
- "Competition Stage" added to reference the phase of the competition this particular match is in.
- "Stadium" updated from a name to an object with a name and corresponding country and id.
- "Referee" updated from a name to an object with a name and corresponding country and id.
- The data-version object is no longer displayed as a property of the match.
- Each match has a metadata object containing information about the structure/versions of various data attributes in the match. This includes the data version. Note: the metadata tag can be empty for any match not in an "available" status. For matches that are available the metadata will contain a data version and may contain other attributes.

Lineups Data Updates

- "Nickname" added for all players in the lineup.

Page 1 of 45

outcome of the event type. The remainder of this document describes an exhaustive list of variables, descriptions, and potential values.

StatsBomb Event Data General Attributes

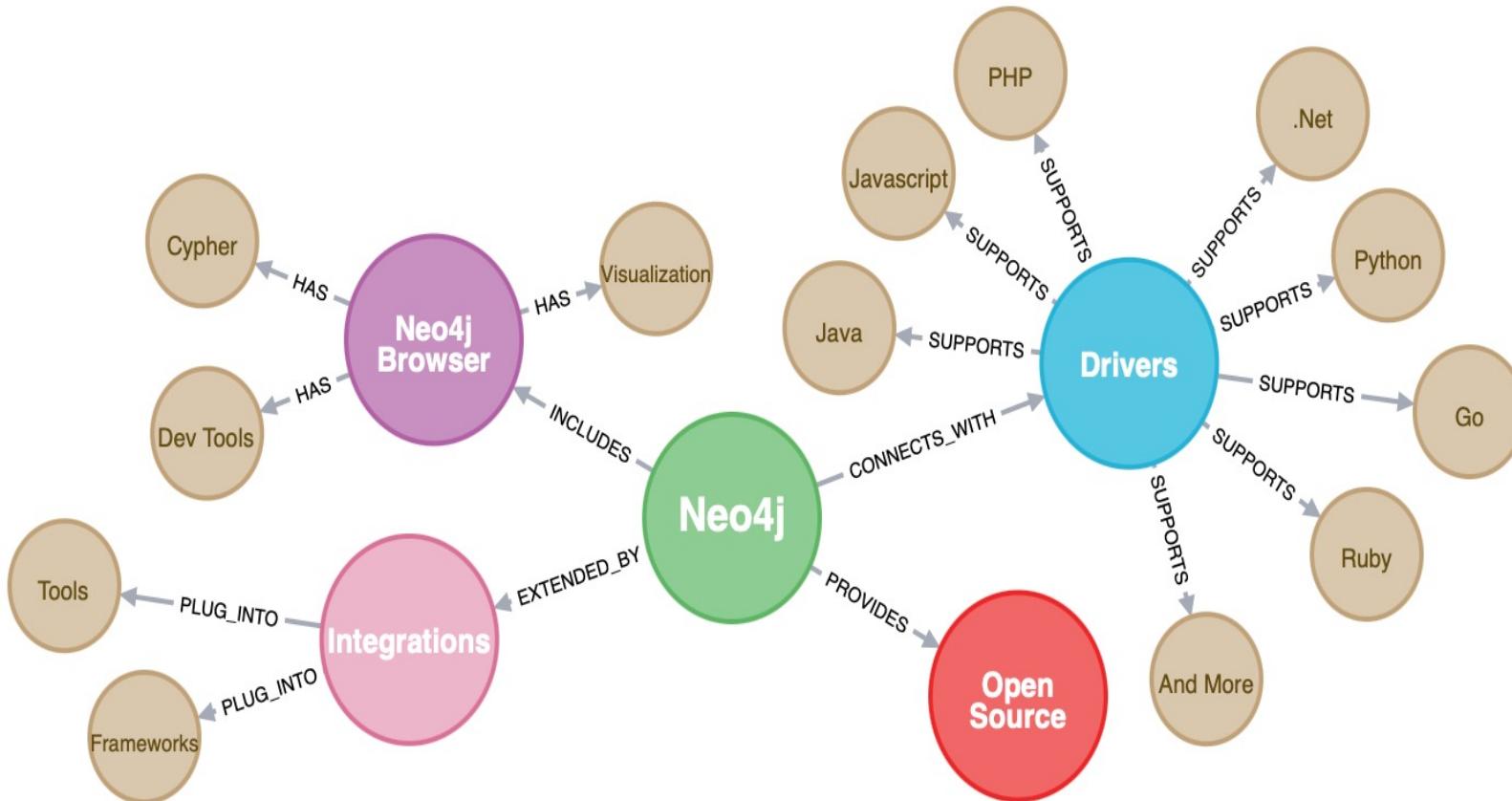
The general attributes are the most common variables for each observation in StatsBomb event data. These attributes are recorded depending on their applicability to the event type described in the next section.

Variable	Variable Type	Variable Description	Values	Value Description
id	uuid	The unique identifier for each event	e.g. 7052d1b5-e2b0-4629-bbea-c18c98aa103	
Index	integer	Sequence notation for the ordering of events within each match.	e.g. 1-4 of events	
period	integer	The part of the match the timestamp relates to.	1 2 3 4 5	1st Half 2nd Half 3rd Period 4th Period Penalty Shootout
timestamp	timestamp	Time in the match the event takes place, recorded to the millisecond.	e.g., 00:00:06.293	
minute	integer	The minutes on the clock at the time of this event. Resets to 45 at half-time, 0 at the start of extra time etc.	e.g., 40	
second	integer	The second part of the timestamp	e.g., 15	
type	object(id(integer), name(text))	The type of event the corresponds to.	See the second on Event attributes for a thorough list.	
possession	integer	Indicates the current unique possession in the game. A single possession denotes a period of play in which the ball is in play and a single team is in control of the ball.	e.g., 1 - # of unique possessions	New possession are triggered after a team demonstrate they've established control of the ball. A new possession can begin even if the same team has possession of the ball for example, a blocked pass goes out for a throw in for the same team, this would be a new possession for the same attacking team.

Page 9 of 45



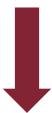
TOOL: NEO4J and ITS USAGE



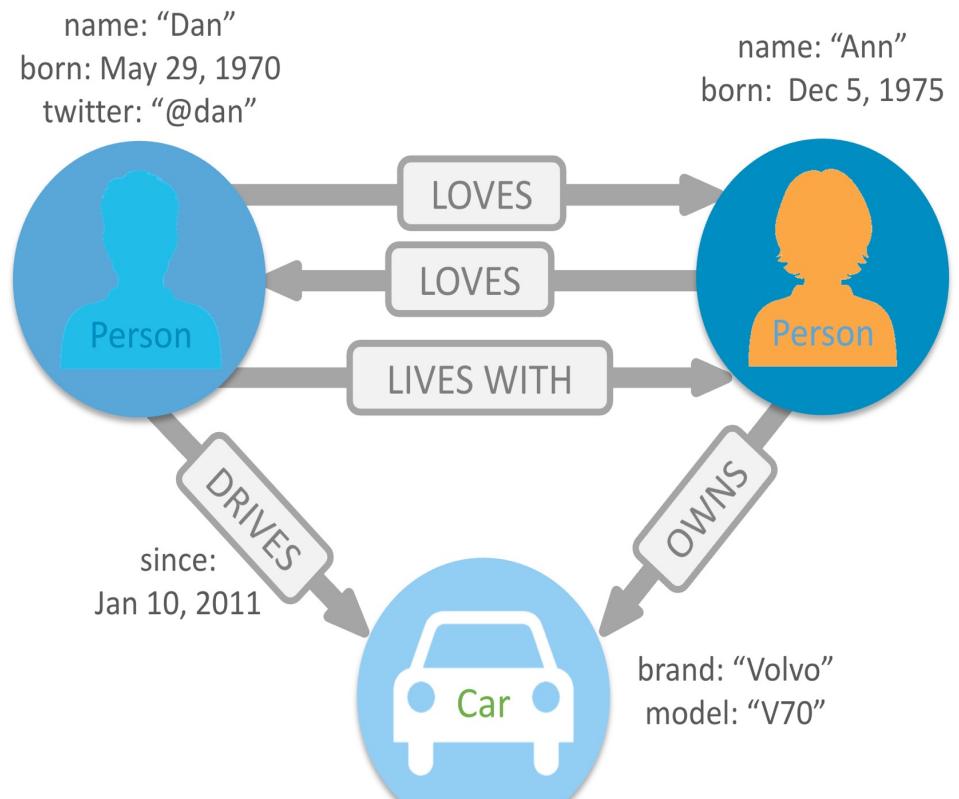


➤ **Neo4j**: an open-source **software for NoSQL databases** developed in Java.

- A new **Data model**: Graph-based model.
- The reason to use it is to **store, handle and query highly-connected data** in a data model based on **Graphs**.
- **Native graph database processing and storage**.
- **Transactional** based on folders to memorize data.
- **Stand-alone** and **integrated** in applications.



- Nodes: entities
- Edges: relationships
- Properties: [attribute, value]



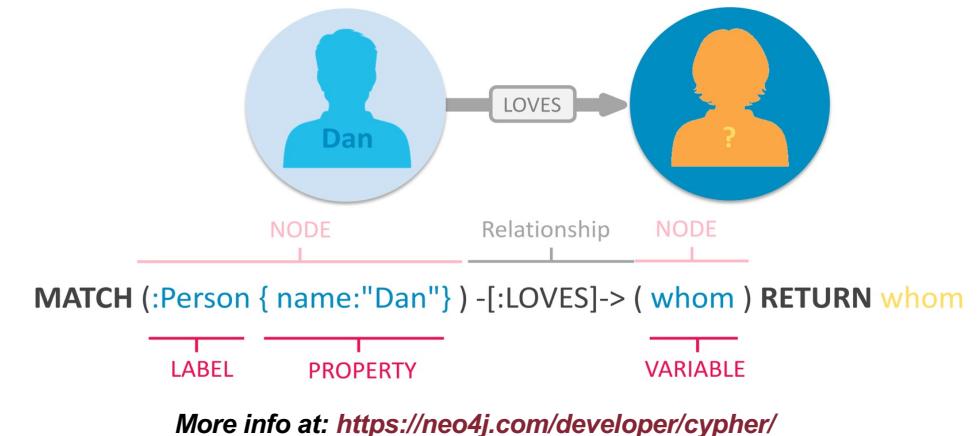


- Create **Graphs** (also importing Data from external sources such as CSV or RDB).
- **Query** Graphs with **Cipher**.
- **Data Science-driven analysis** on Neo4j-Graphs.
- Administration of Graphs in **Production and Deployment** (e.g. Docker or in the cloud).

CIPHER:

A next generation query language: human-friendly declarative language for graph patterns based on the power of SQL but optimized for Graphs following CRUD operations.

- ✓ Navigate deep hierarchies,
- ✓ Find hidden connections between distant items, and
- ✓ Discover inter-relationships between items.



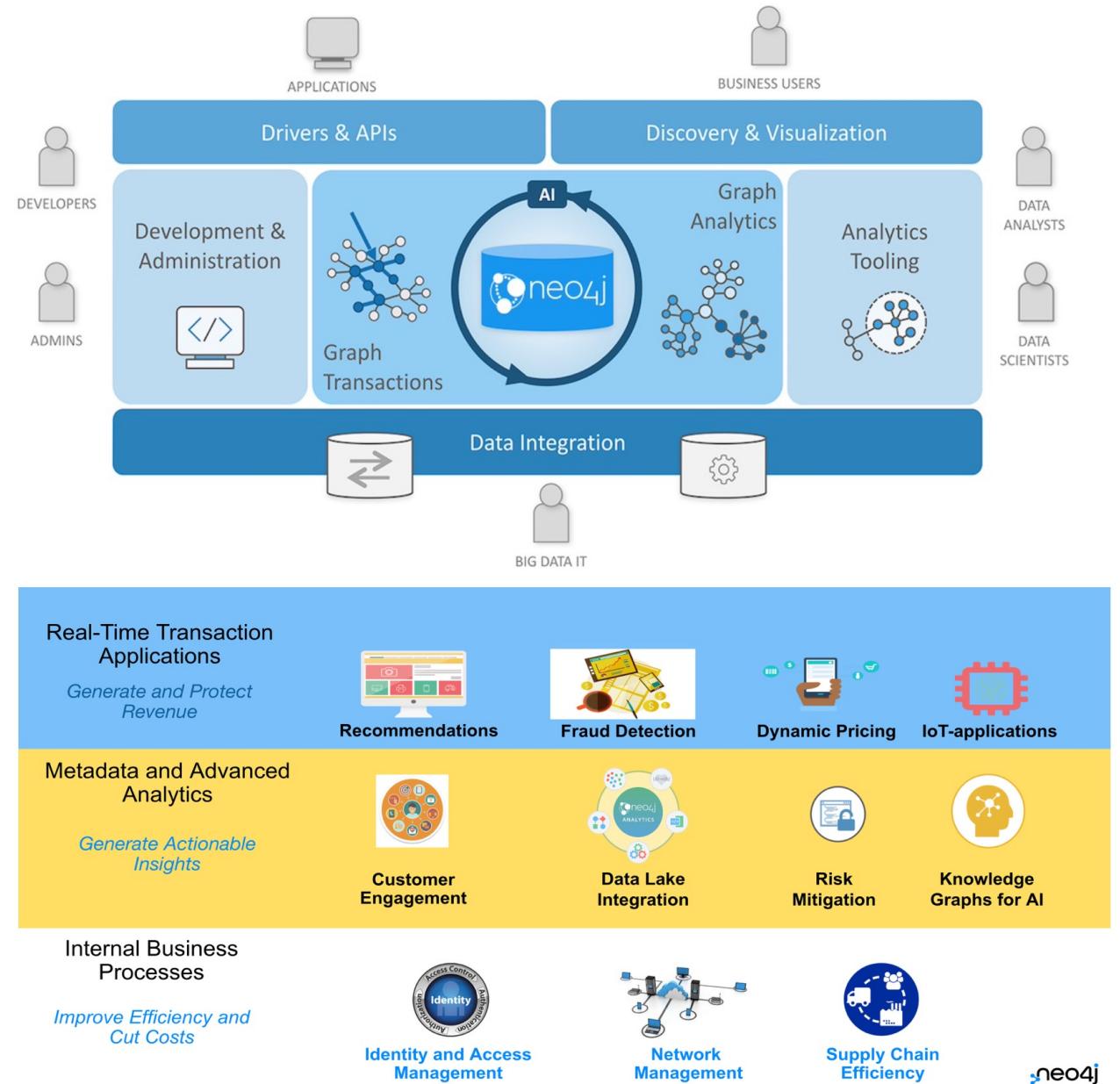


Neo4j Graph Database - our core graph database that is built to store and retrieve connected data. There are two editions - a Community Edition and an Enterprise Edition. Everything in the platform interacts with data stored in the database.

Neo4j Desktop - application to manage local instances of Neo4j.

Neo4j Browser - online browser interface to query and view the data in the database. Basic visualization capabilities using Cypher query language.

Neo4j Bloom - visualization tool for business users that does not require any code or programming skills to view and analyze data.





Neo4j Desktop:

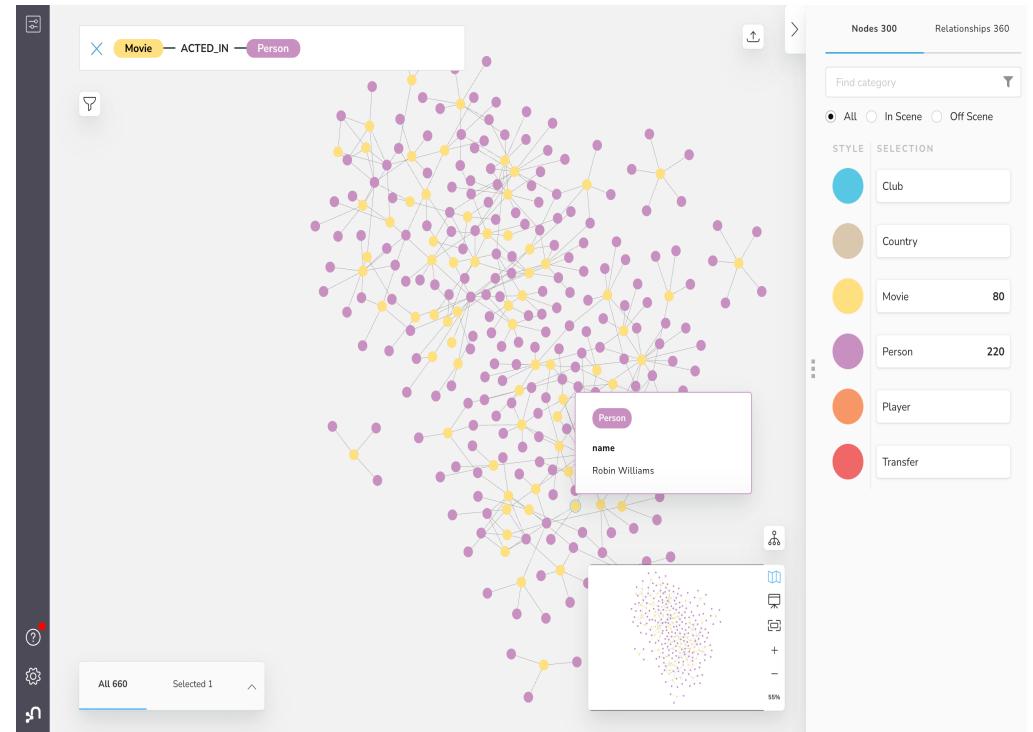
The image displays three side-by-side screenshots of the Neo4j Desktop application interface.

- Left Screenshot:** Shows the main project management screen. It features a sidebar with icons for Projects, New, Help, and Notifications. The main area is titled "Project" and shows a single entry: "Graph DBMS 4.4.5". Below it, there's a "File" section with a placeholder "Add project files to get started." and a "Reveal files in Finder" button.
- Middle Screenshot:** Shows a detailed view of a project. The top bar includes "Projects", "New", and a search icon. The main area is titled "Project" and shows the same "Graph DBMS 4.4.5" entry. Below it, a message says "This list of databases is cached, start the DBMS to refresh the list." followed by a list of databases: "system" and "neo4j (default)".
- Right Screenshot:** Shows the "Details" tab of a project configuration screen. The top bar includes "Projects", "New", and a search icon. The main area is titled "Project" and shows the "Graph DBMS 4.4.5" entry with a "start" button. To the right, there are tabs for "Details", "Plugins", and "Upgrade". The "Plugins" tab is selected, displaying a list of available plugins: "APOC", "Graph Data Science Library", "Neo4j Streams", and "Neosemantics (n10s)".



Neo4j Browser and Neo4j Bloom:

The screenshot shows the Neo4j Browser interface. On the left, there's a sidebar with "Database Information" containing sections for "Use database" (set to "neo4j"), "Node Labels" (Club, Country, Movie, Person, Player, Transfer), "Relationship Types" (ACTED_IN, FOLLOWED, PRODUCED, REVIEWED, WROTE), "Property Keys" (born, id, name, rating, released, roles, summary, tagline, title), "Connected as" (Username: neo4j, Roles: admin, PUBLIC, Admin: server user list, Disconnect: server disconnect), and "DBMS" (Version: 4.4.5, Edition: Enterprise, Name: neo4j, Databases: rdbms). The main area shows a terminal window with the command "\$:play intro" and a text block about the Neo4j Browser User Interface.



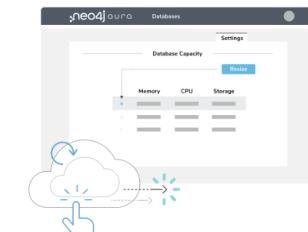


I will use a Python Notebook through Neo4j specific drivers to connect with it.

Neo4j Sandbox:

The screenshot shows the Neo4j Sandbox web interface. At the top, there's a navigation bar with 'neo4j' logo, 'New Project' button, and user 'Hi Federico'. Below the navigation is a table with one row: 'Name' (Blank Sandbox) and 'Status' (Running, Expires in about 3 days). There are 'Actions', 'Connection details', 'Connect via drivers', and 'Backups' buttons. Under 'Connection details', the following information is listed:
Username: neo4j IP Address: 44.201.167.70
Password: brain-console-forty HTTP Port: 7474
Bolt Port: 7687
Bolt URL: bolt://44.201.167.70:7687
Websocket Bolt URL: bolt+s://f63b851d2c498f0cb1d0e3768af11ffc.neo4jsandbox.com:7687

Then it can be added to the Neo4j Desktop as a remote database using the connection details but I will use also a local database through the Neo4j Browser.





EXPERIMENTS

2 TASKS:

The idea is to use **Neo4j** in two tasks:

1. The **first task** related to a dataset of [transfermarkt.com](#) and, using the graph as a data model, to represent the **transfers** that took place in a certain year or with particular properties; then it is about to show **queries** that can be advantageous in a representation of this type.
2. The **second task** is instead to show how a graph can be exploited in **match analysis**: this consists in taking a dataset from [statsbomb.com](#) relating to the events of a single match and showing how the graph can be structured and then some useful **queries**.



In both tasks the nature of data are based on **temporal sequences of events**.



Graph databases (Neo4j) are suitable!

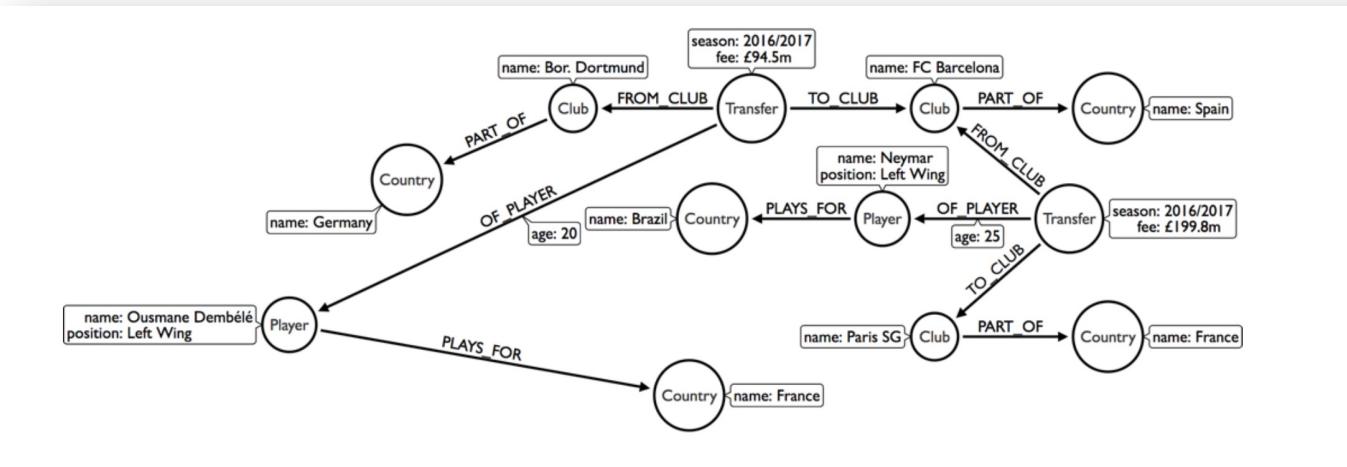


TASK 1: TRANSFERMARKT

1. Installation in Python (Pip3).
2. Driver for Neo4j: creation and query session.
3. Constraints setting and data loading.
4. Use of the usual **pattern for query**.
5. **APOC functions** for data cleaning and operations with strings.
6. **Aggregation queries and path-based queries**.
7. Other **keywords** used are: not exists, exists, starts with, ends with, case, when, end.

Some keywords used:

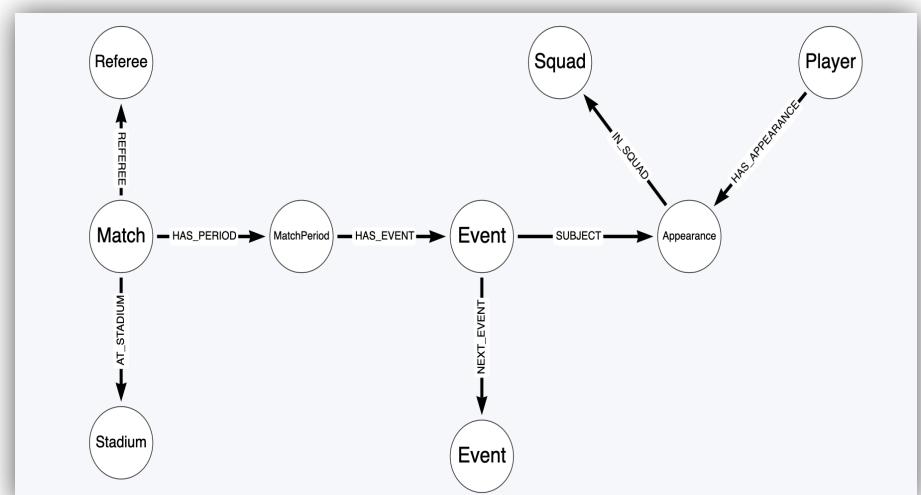
WITH
LOAD
ON CREATE SET
MATCH
WHERE
RETURN / MERGE
LIMIT
ORDER BY





TASK 2: STATSBOMB

1. Creation of local database and installation of APOC library.
2. Browser initialization for query session.
3. Constraints setting and data loading: all JSON files are uploaded using APOC functions.
4. Competitions are created, each one with different Seasons.
5. Matches are created with informations like Referee and Stadium.
6. For a particular Match the Players and Squads are inserted.
7. For a particular Match Events are created and linked in a linked list in chronological order.
8. Many Event types are defined such as Pass, Goal, Shot etc.
9. Use of the usual pattern for query and path-based queries: football analysis application.





AND NOW A QUICK DEMO..

TASK 1:

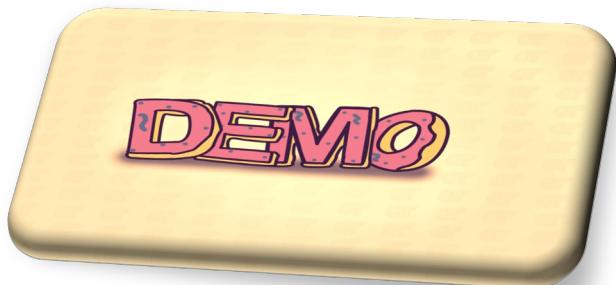
1. Sandbox Graph on this dataset.
2. Simple and more complex queries on it with Python.

} Neo4j Sandbox Environment and Python Notebook **(IN CLOUD)**

TASK 2:

1. How to create the local Graph on this dataset.
2. Simple and more complex queries on it.

} Neo4j Desktop and Browser **(ON PREMISE)**





CONCLUSIONS

POSSIBLE EXTENSIONS:

The idea to use Neo4j can be extended in different ways:

1. Integration in a **real application**.
2. Exploitation of the power of **Cipher** as Query Language and other **external libraries**.
3. Use of JavaScript frameworks (**D3.js**) for Data Analysis to **visualize** dynamically the **Graphs**.
4. Use of a **greater size** for the Datasets, also adding **ETL** operations.



Master's Degree in Engineering in Computer Science
Sapienza University of Rome
Academic Year 2021-2022

Thank you for your attention!



SAPIENZA
UNIVERSITÀ DI ROMA

Course: Data Management
Professor: Prof. Lenzerini Maurizio

Candidate: Carmignani Federico

ID number: 1845479

E-mail: carmignani.1845479@studenti.uniroma1.it