



UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania
Matematycznego i Komputerowego
icm.edu.pl

Trovares xGT

Akademia ICM – warsztaty

Jakub Jałowiec

Interdyscyplinarne Centrum Modelowania

ICM UW = superkomputery na Uniwersytecie Warszawskim:

- Okeanos
- Topola
- **Rysy**

Granty obliczeniowe: https://kdm.icm.edu.pl/Podrecznik_grantowy/zakladanie_konta/

Agenda

Cel warsztatów:

- pokazać działanie Trovares xGT i nauczyć podstaw wykorzystywania tego narzędzia
- udostępnić „wędkę do łowienia ryb”

Plan warsztatów:

- wprowadzenie – 1h
- demonstracja – 30min
- część praktyczna – 2h
- ewaluacja

Przerwy?

Pytania?

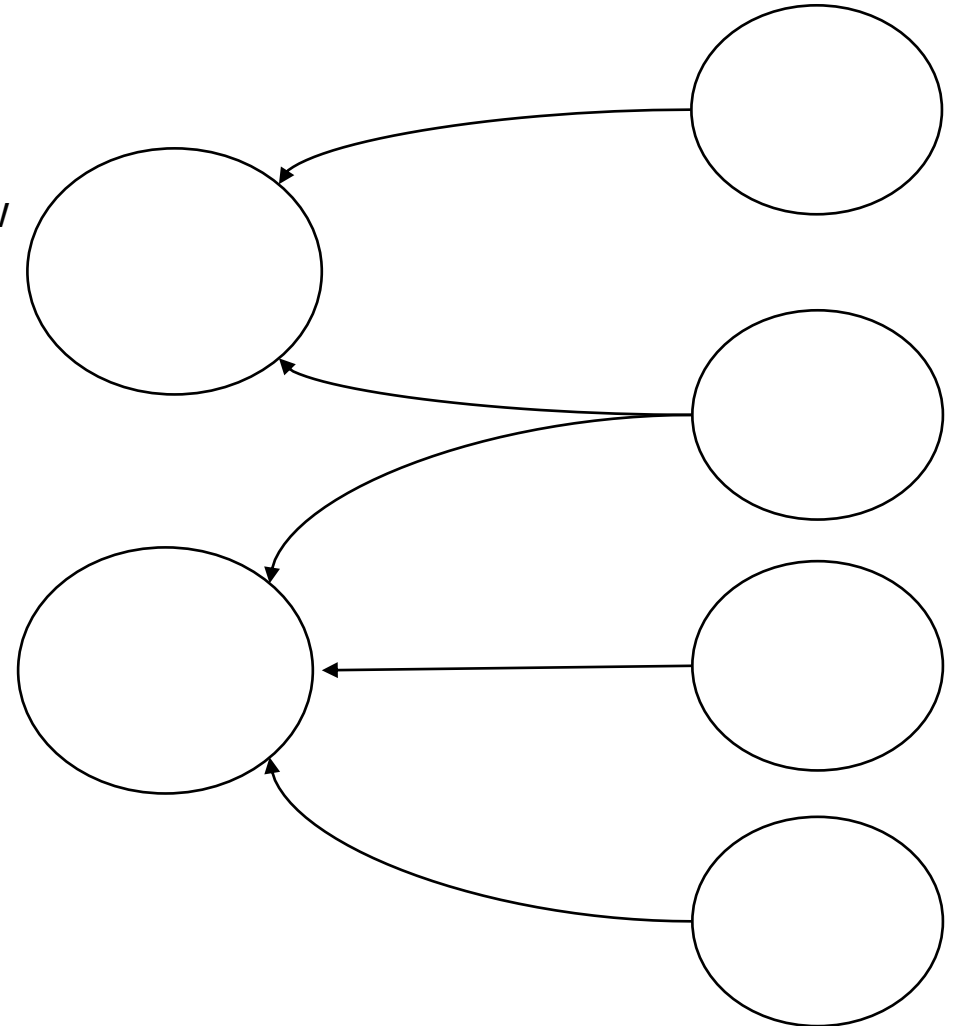
Czym są grafy?

Graf to formalnie para (*zbiór wierzchołków, zbiór krawędzi*).

Każda krawędź musi zaczynać się w jednym z wierzchołków i kończyć się w drugim z wierzchołków.

Przykład po prawej:

- 6 wierzchołków
- 5 krawędzi

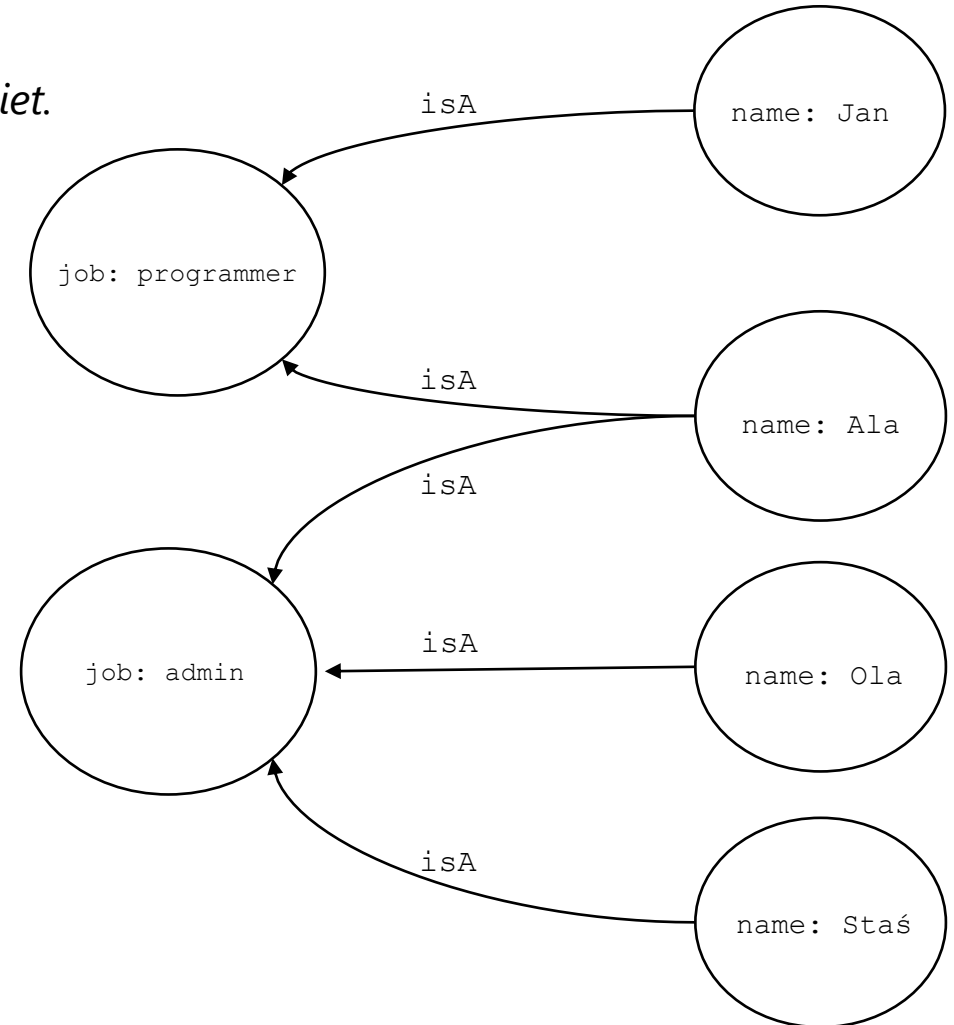


Etykiety i atrybuty w grafach

Prostą definicję grafu można wzbogacić o pojęcia *atrybutów* i *etykiet*.

Atrybut – para (*klucz, wartość*) przechowywana w wierzchołku lub krawędzi.

Etykieta – pojedyncza wartość przechowywana w wierzchołku lub krawędzi.



Czym jest Trovares xGT?

- grafowe narzędzie analityczne typu *in memory*
- języki wejściowe:
 - Python – administracja (np. wczytywanie/zapisywanie danych)
 - *Trovares Query Language* – analityka (ile jest wierzchołków...?)
- formaty wejściowe/wyjściowe danych: CSV

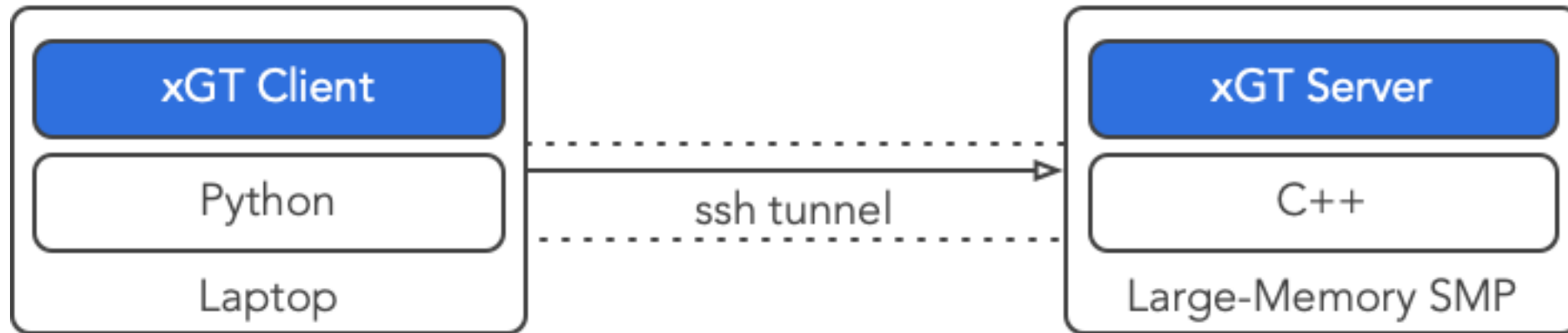


Do czego można zastosować Trovares xGT?

- wyszukiwanie wzorców w dużych grafach
- zastosowania:
 - *fraud detection*
 - nauki społeczne
 - ...

Co wchodzi w skład Trovares xGT?

- *oprogramowanie klienta* – publicznie dostępna biblioteka o nazwie *xgt* w języku Python
- *oprogramowanie serwera* – licencjonowany program napisany w C++ do wydajnego przetwarzania grafów
- *klient „zleca” wykonywanie prac serwerowi*

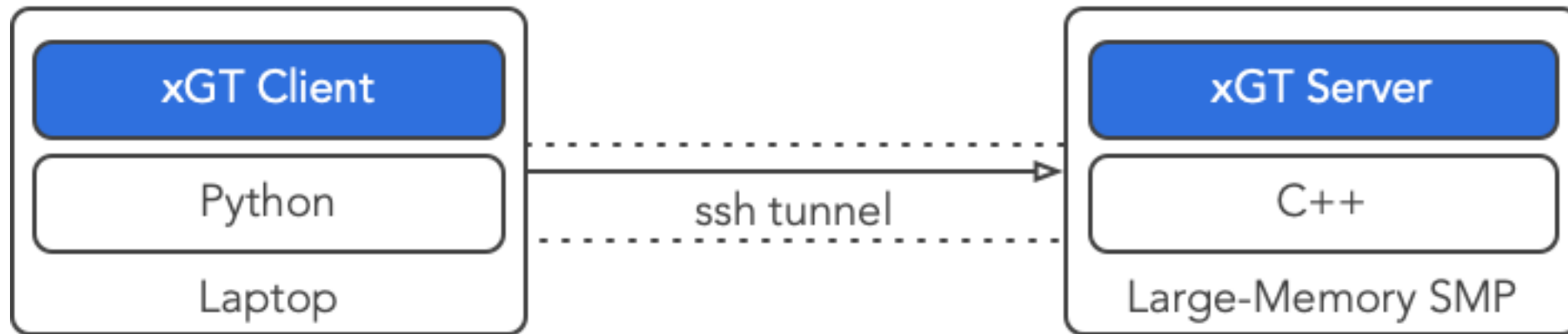


[źródło](#)

Serwer Trovares xGT

Serwer Trovares xGT – komputer o odpowiednio dużej pamięci RAM z zainstalowanym licencjonowanym oprogramowaniem Trovares xGT.

To serwer xGT przetwarza dane a nie klient xGT!

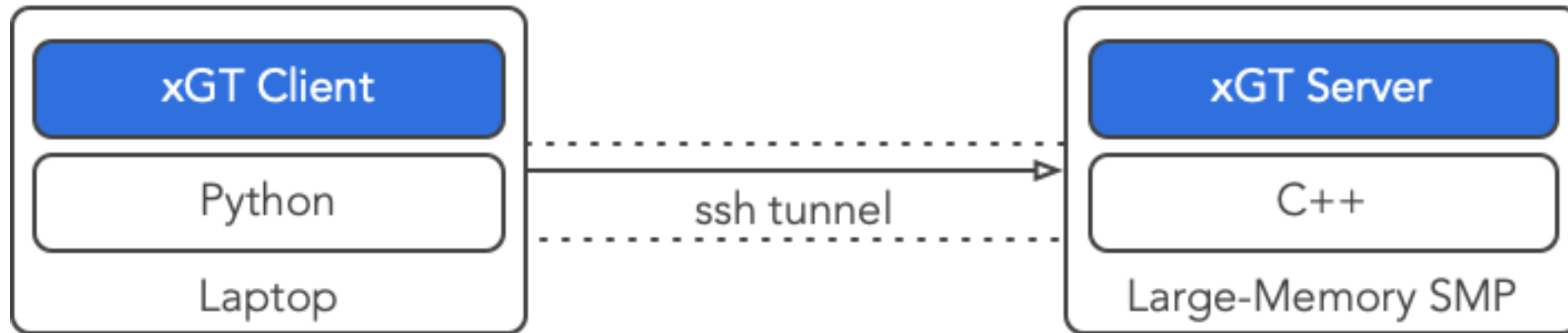


[źródło](#)

Klient Trovares xGT

Klient Trovares xGT – komputer z zainstalowanym środowiskiem Python 3.5.+, działającym połączeniem z internetem oraz zainstalowaną biblioteką *xgt* („*pip install xgt*”).

Biblioteka *xgt* w języku Python dostarcza zestaw klas do nawiązywania połączenia z serwerem xGT, wczytywania/zapisywania danych do i z serwera oraz zlecania mu „*prac*”.



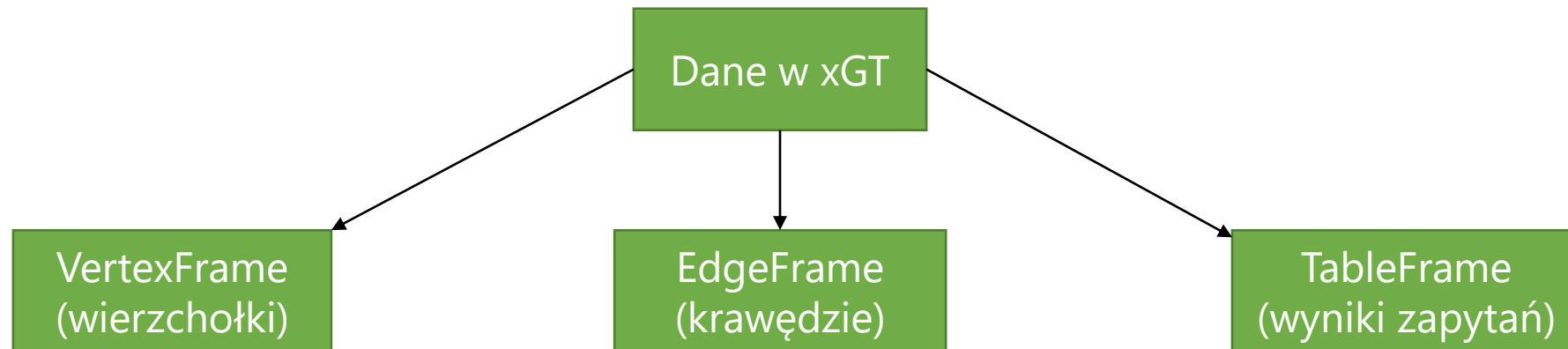
[źródło](#)

Jak reprezentowane są dane na serwerze Trovares xGT?

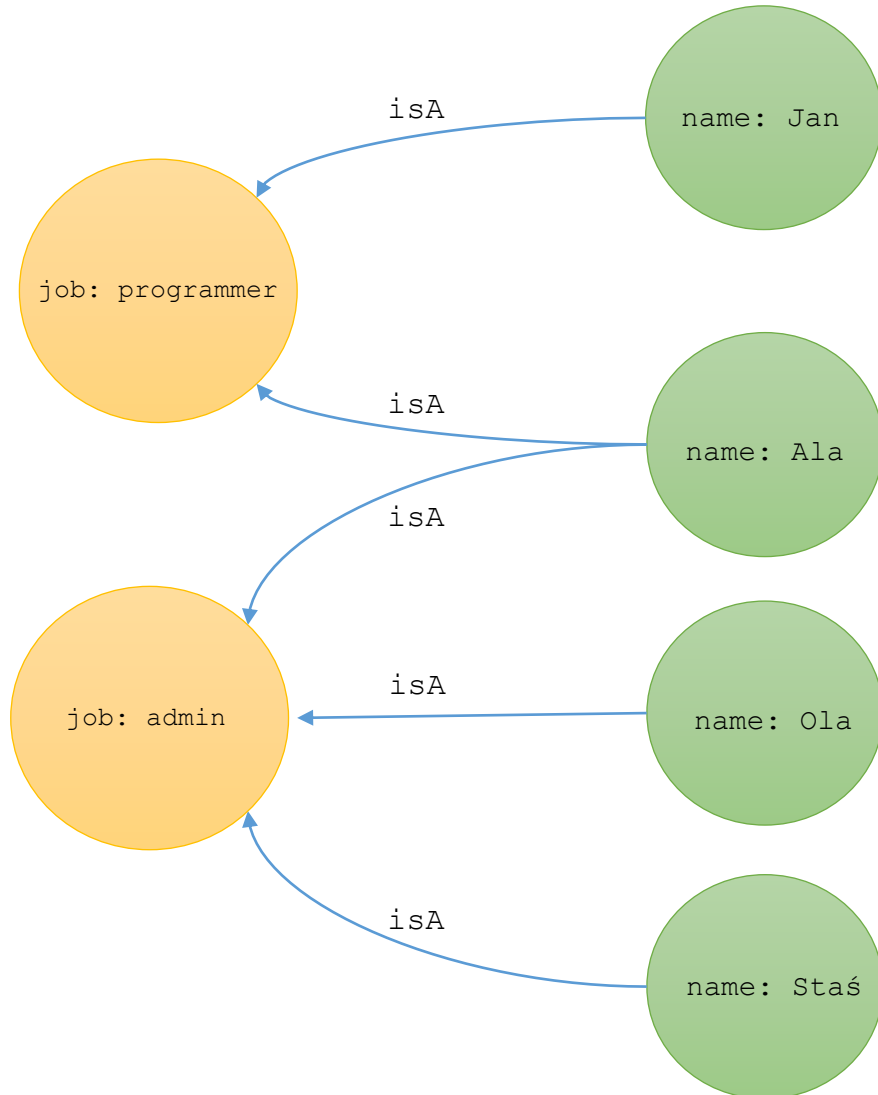
Sposób reprezentacji danych w xGT

Dane przechowywane są na serwerze Trovares xGT w tzw. ramkach (*frames*), których są 3 typy:

- ramki wierzchołków (*VertexFrame*),
- ramki krawędzi (*EdgeFrame*),
- wyniki zapytań (*TableFrame*).

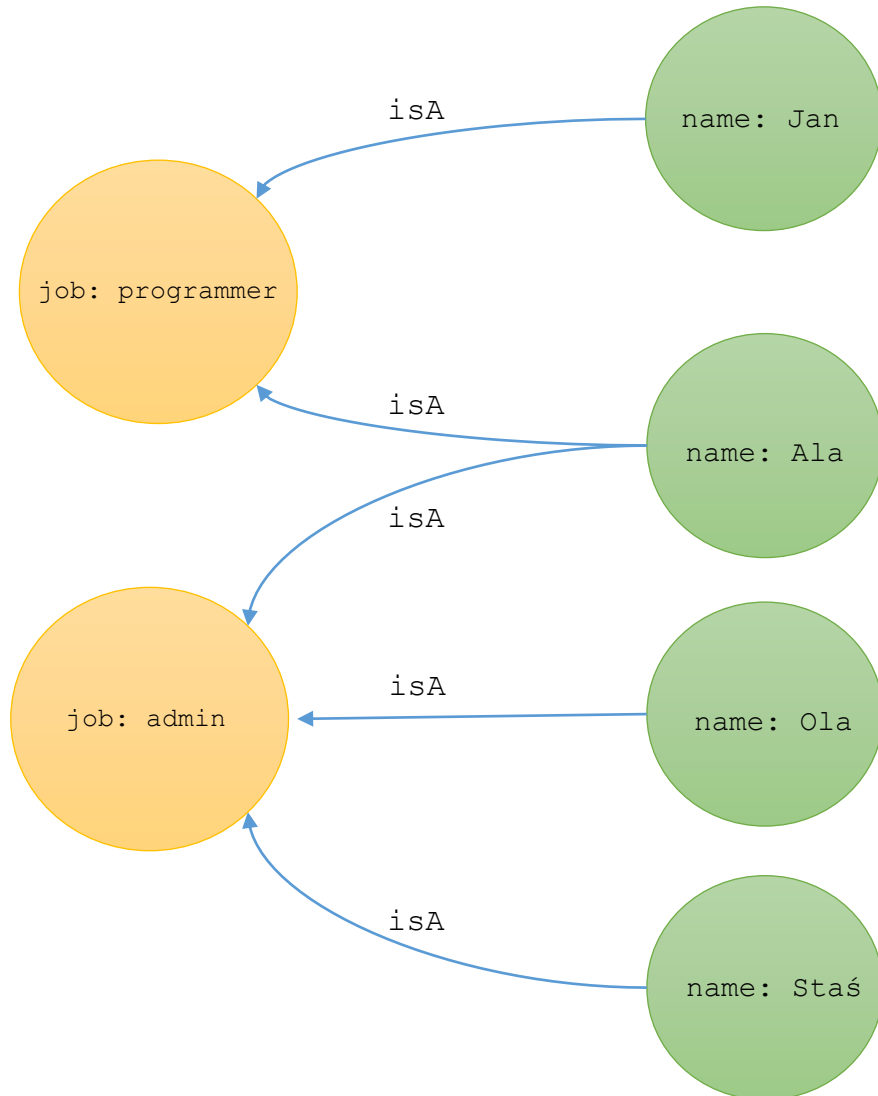


Mapowanie grafów na ramki



2 typy wierzchołków?
1 typ krawędzi?

Mapowanie grafów na ramki



EdgeFrame: isA

source	target
1	1001
2	1001
2	1002
3	1002
4	1002

VertexFrame: Persons

id	name
1	Jan
2	Ala
3	Ola
4	Staś

VertexFrame: Jobs

id	job
1001	programmer
1002	admin

Dostępne typy danych

- `xgt.BOOLEAN`
- `xgt.INTEGER`
- `xgt.FLOAT`
- `xgt.DATE`
- `xgt.TIME`
- `xgt.TEXT`
- `xgt.DATETIME`
- `xgt.IPADDRESS`

Schemat ramki

Schemat ramki to zbiór rodzajów pól, jakie przechowuje dany typ wierzchołka lub krawędzi.

EdgeFrame: isA

source	target
1	1001
2	1001
2	1002
3	1002
4	1002

VertexFrame: Persons

id	name
1	Jan
2	Ala
3	Ola
4	Staś

VertexFrame: Jobs

id	job
1001	programmer
1002	admin



Jakie typy nadać atrybutom?

isA.source: ???
isA.target: ???
Persons.id: ???
Persons.name: ???
Jobs.id: ???
Jobs.job: ???

Schemat ramki

Schemat ramki to zbiór rodzajów pól, jakie przechowuje dany typ wierzchołka lub krawędzi.

EdgeFrame: isA

source	target
1	1001
2	1001
2	1002
3	1002
4	1002

VertexFrame: Persons

id	name
1	Jan
2	Ala
3	Ola
4	Staś

VertexFrame: Jobs

id	job
1001	programmer
1002	admin



```
isA.source:    xgt.INT
isA.target:    xgt.INT
Persons.id:    xgt.INT
Persons.name:  xgt.TEXT
Jobs.id:       xgt.INT
Jobs.job:      xgt.TEXT
```

Reprezentacja grafu – kompletny obraz

- nazwa ramki – uwaga na prefiks!
- schemat
- wyróżnienie kluczy

EdgeFrame: workshop__isA

source: xgt.INT	target: xgt.INT
1	1001
2	1001
2	1002
3	1002
4	1002

VertexFrame: workshop__Persons

id: xgt.INT	name: xgt.TEXT
1	Jan
2	Ala
3	Ola
4	Staś

VertexFrame: workshop__Jobs

id: xgt.INT	job: xgt.TEXT
1001	programmer
1002	admin

Jak pracować z Trovares xGT?

Trovares Query Language (TQL)

Przykład zapytania w TQL „wybierz wszystkie wierzchołki *a* połączone krawędzią *e* z wierzchołkiem *b*, gdzie wierzchołek *b* ma atrybut *job* równy "admin". Poniżej widoczna jest ramka `TableFrame`, będąca rezultatem zapytania:

```
MATCH
(a:workshop__Persons)-[e: workshop__isA]->(b: workshop__Jobs)
WHERE b.job = "admin"
RETURN a.name,
       b.job
INTO results__Person_Job
```

TableFrame: workshop__Person_Job

name: xgt.TEXT	job: xgt.TEXT
Ala	admin
Ola	admin
Staś	admin

Trovares Query Language (TQL)

TQL to język zapytań do danych przechowywanych na serwerze Trovares xGT.

Ogólna charakterystyka TQL:

- wybieranie danych: `MATCH`
- filtrowanie danych: `WHERE`
- modyfikacja danych: `SET`
- zapisywanie danych w nowej ramce: `RETURN ... INTO`

Jak wywołać zapytanie na serwerze?

Biblioteka *xgt* w języku Python dostarcza klasę `xgt.Connection`, służącą do wykonywania wszelkich działań na serwerze.

Klasa `xgt.Connection` posiada metody takie jak: `run_job(...)`, `create_vertex_frame(...)`, `create_edge_frame(...)`, `drop_frame(...)`, itd.

Uwaga – metody wywoływane na obiekcie klasy `xgt.Connection` mają przede wszystkim działanie po stronie serwera!



Jak wywołać zapytanie na serwerze?

kod Python

```
server = xgt.Connection(...parametry połączenia...)
```

...wczytanie danych do Persons, isA oraz Jobs...

```
query = """  
MATCH (a:workshop__Persons)-[e: workshop__isA]->(b: workshop__Jobs)  
WHERE b.job = "admin" RETURN a.name, b.job INTO results__Person_Job  
"""
```

```
server.run_job(query)
```

...sprawdzenie wyników...



UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania
Matematycznego i Komputerowego
icm.edu.pl

Dziękuję za uwagę

Zapraszam na część praktyczną