

PGX Graph Size Calculator And Data Studio Load Tests

Presented By : Mohamed Boukhlif

School supervisors

- Mrs. Nezha Benmoussa
- Mr. Mohammed Qbadou

Company mentor

- Mrs. Rania Medraoui

2019/2020

Agenda

- 1 Introduction
- 2 PFE Host Company
- 3 Data Studio
- 4 Our Project
 - 4.1 Graph Size Calculator
 - 4.2 Data Studio Load Tests
- 5 Conclusion

1. Introduction

PFE internship

2. Host Company

Oracle, Oracle Labs, Oracle Labs Morocco

Host Company

Oracle



Founded in 1977, Oracle Corporation is an American multinational computer technology corporation headquartered in Redwood Shores, California. The company sells database software and technology, cloud engineered systems, and enterprise software products.

Host Company

Oracle Labs



Oracle Labs is a research and development branch of Oracle Corporation. The labs were created when Oracle acquired Sun Microsystems. Sun Labs was established in 1990 by Ivan Sutherland and Robert Sproull. The initial locations were in Menlo Park, California and Burlington, Massachusetts, United States.

Host Company

Oracle Labs Morocco

Oracle Labs Morocco represents a new branch of Oracle laboratories. Oracle Labs has offices in Redwood Shores, California, Cambridge in the United Kingdom, Brisbane in Australia, Vienna in Austria, Zurich in Switzerland.

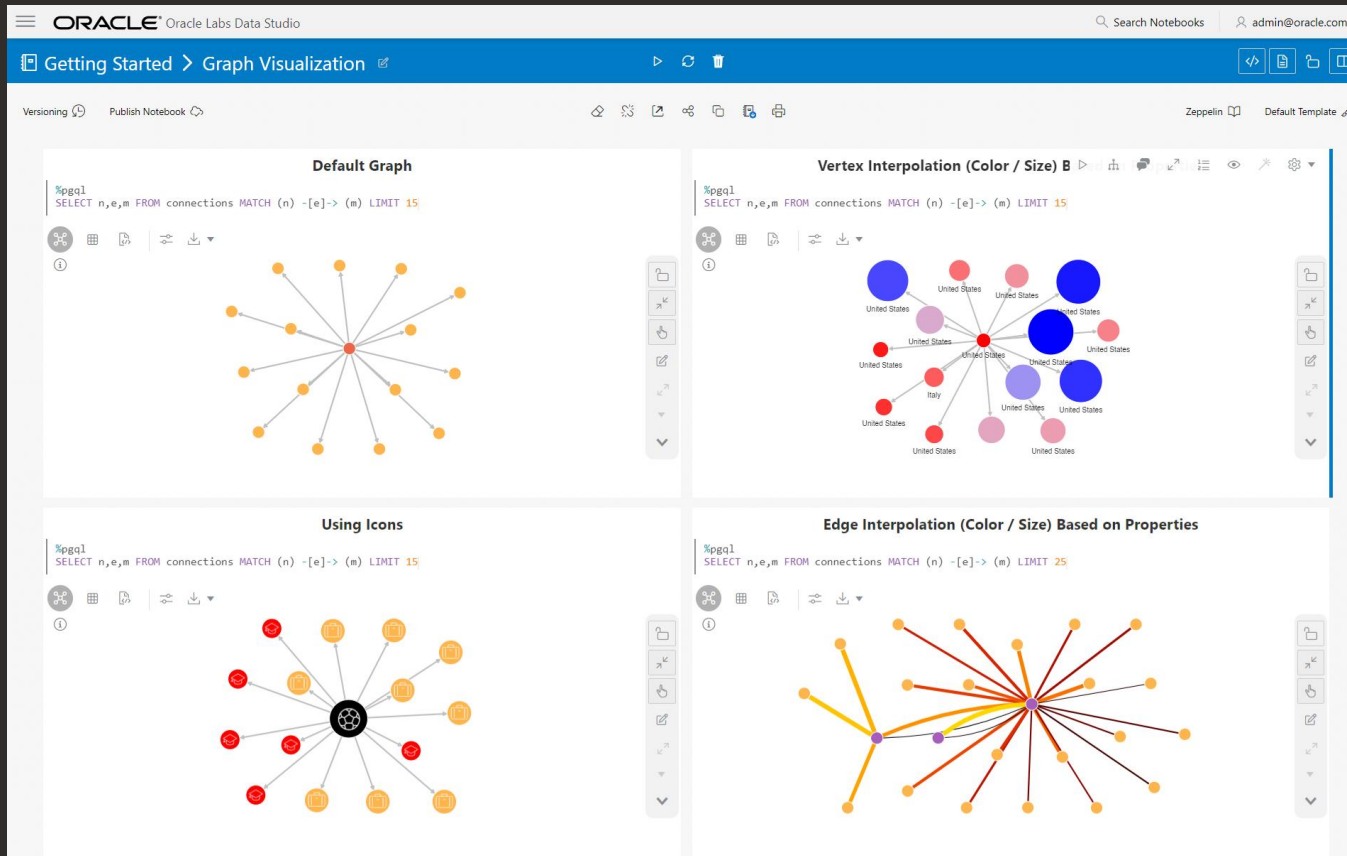
3. Data Studio

What is Data Studio ?

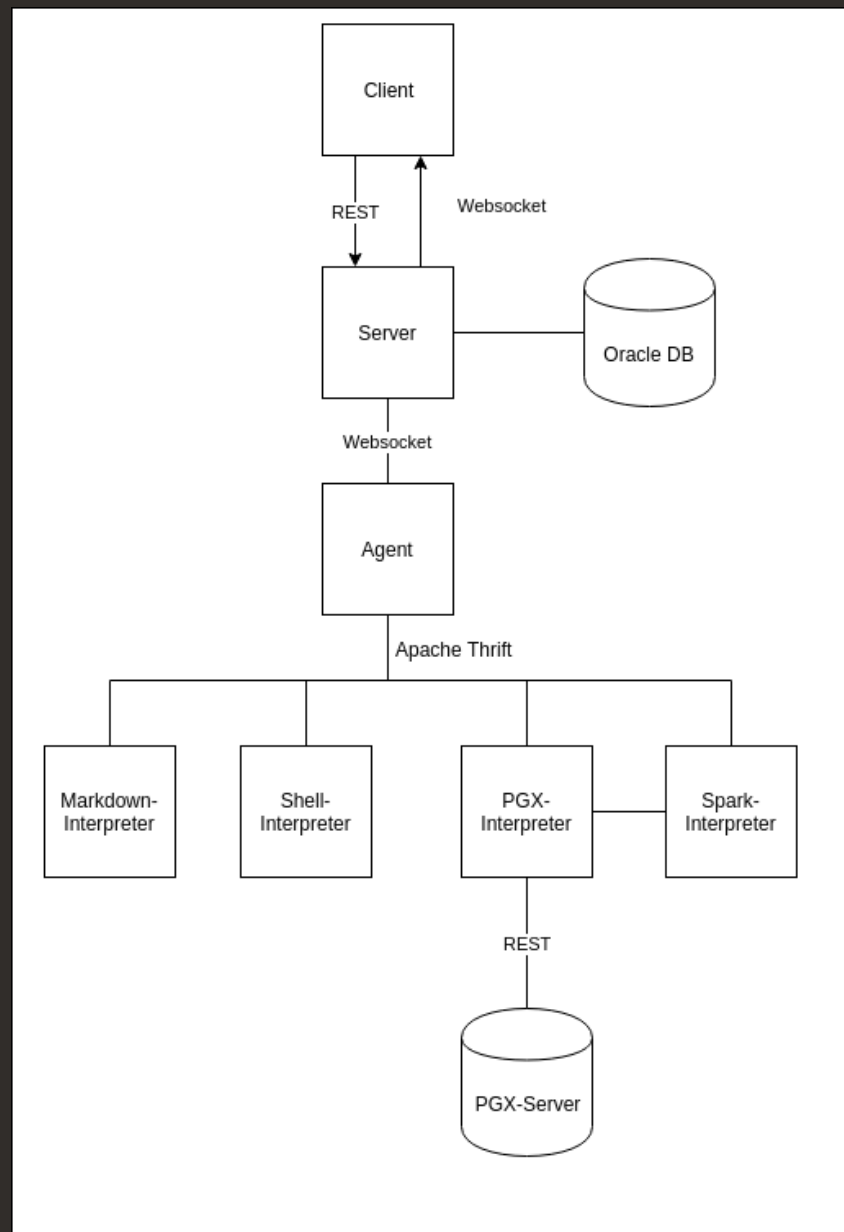
Data Studio

Data What?

Oracle Labs Data Studio is a web-based notebook platform for data scientists. By combining live code collaboration in multiple programming languages with rich, interactive graphical analysis and visualization, Data Studio accelerates the process of exploring and acquiring information from your data.



Data Studio Architecture



4. Our Project

- PGX Graph Size Calculator
- Data Studio Load Tests

Our project

PGX Graph Size Calculator

- Presentation
- Analysis and Design
- UI Mocks
- Development
- Tests
- Accuracy Tests

Data Studio Load Tests

- What is load tests?
- Goal
- Our test plans
- Results

4.1. Graph Size Calculator

- Presentation
- Analysis and Design
- UI Mocks
- Development
- Tests and Deployment
- Accuracy Tests

Graph Size Calculator

Presentation

The problem we're trying to solve :

To estimate the size of a graph, the PGX and Data Studio team were using an Excel file, where it is necessary for each graph to manually insert all the characteristics (partitioned / non-partitioned graph, number of vertices, number of edges, type of each property...)

Drawbacks :

- For each graph you need an Excel file
- Manual entry of the characteristics of the graph
- Difficult to scale the graph
- Difficult to maintain calculation formulas between files

Graph Size Calculator

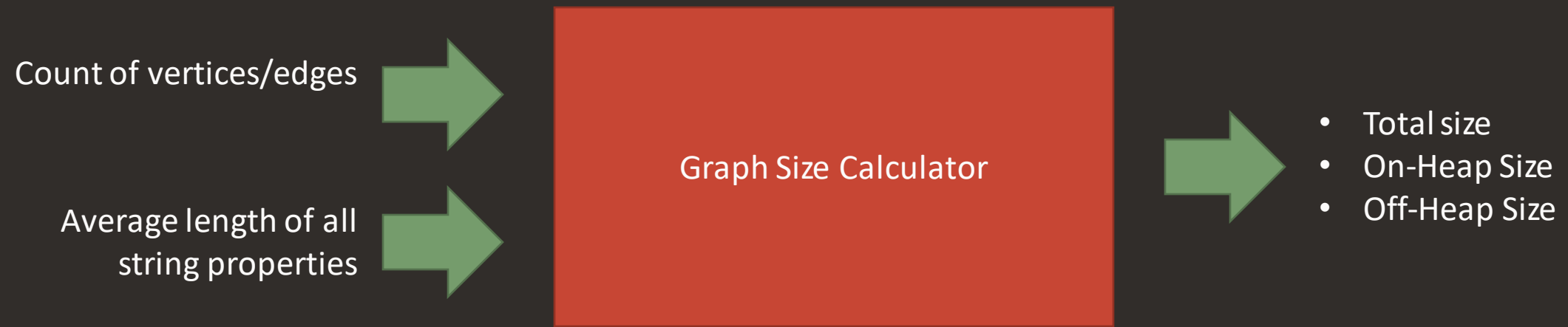
Presentation

Goal :

The goal is to create a tool capable of estimating the memory size of PGX graphs.

The tool will take as input a configuration of a graph, as well as the number of vertices and edges, and output the memory estimation for On-Heap, Off-Heap and the total memory.

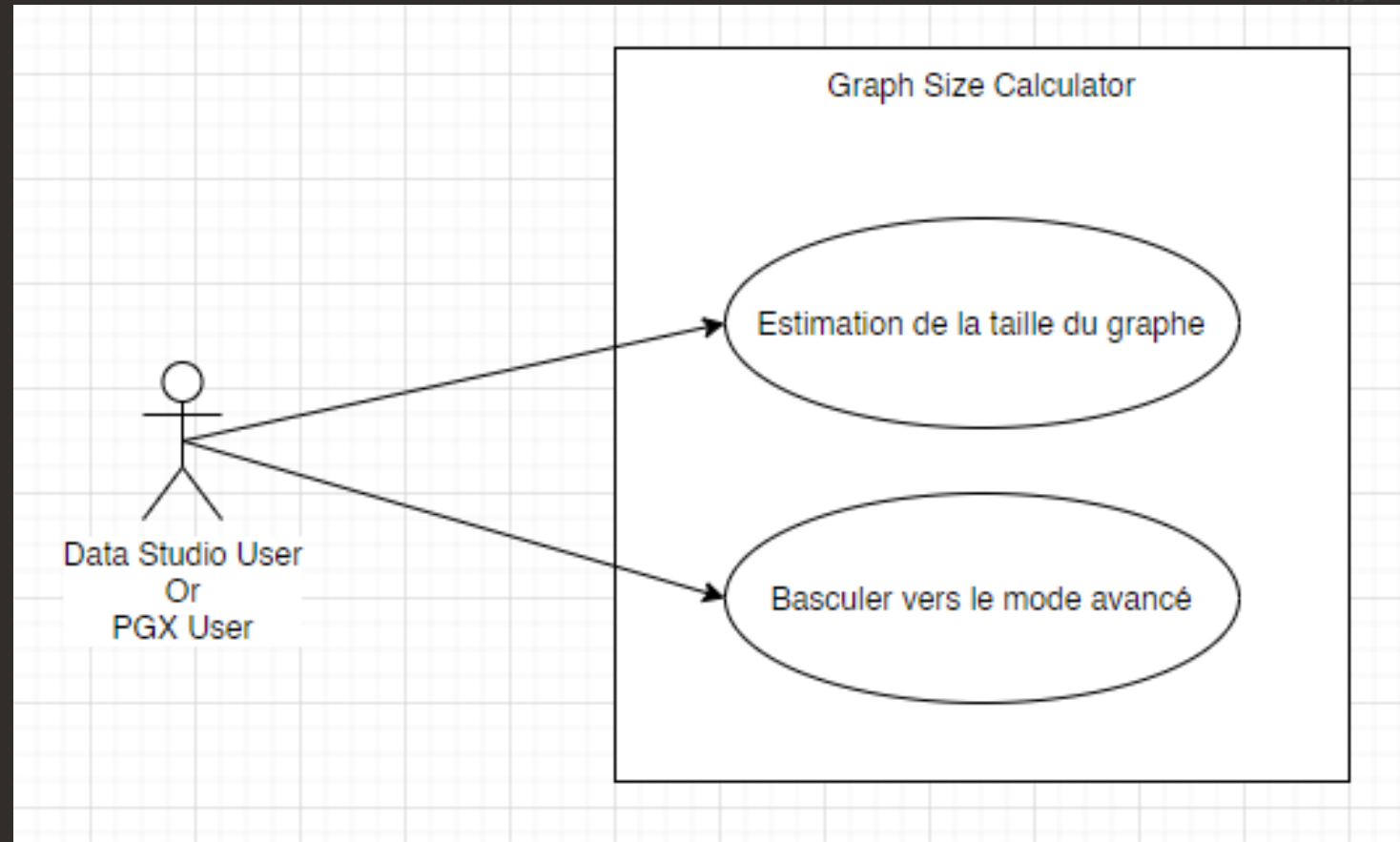
The tool will be an Oracle JET web component.



Graph Size Calculator

Analysis & Design

Use case diagram

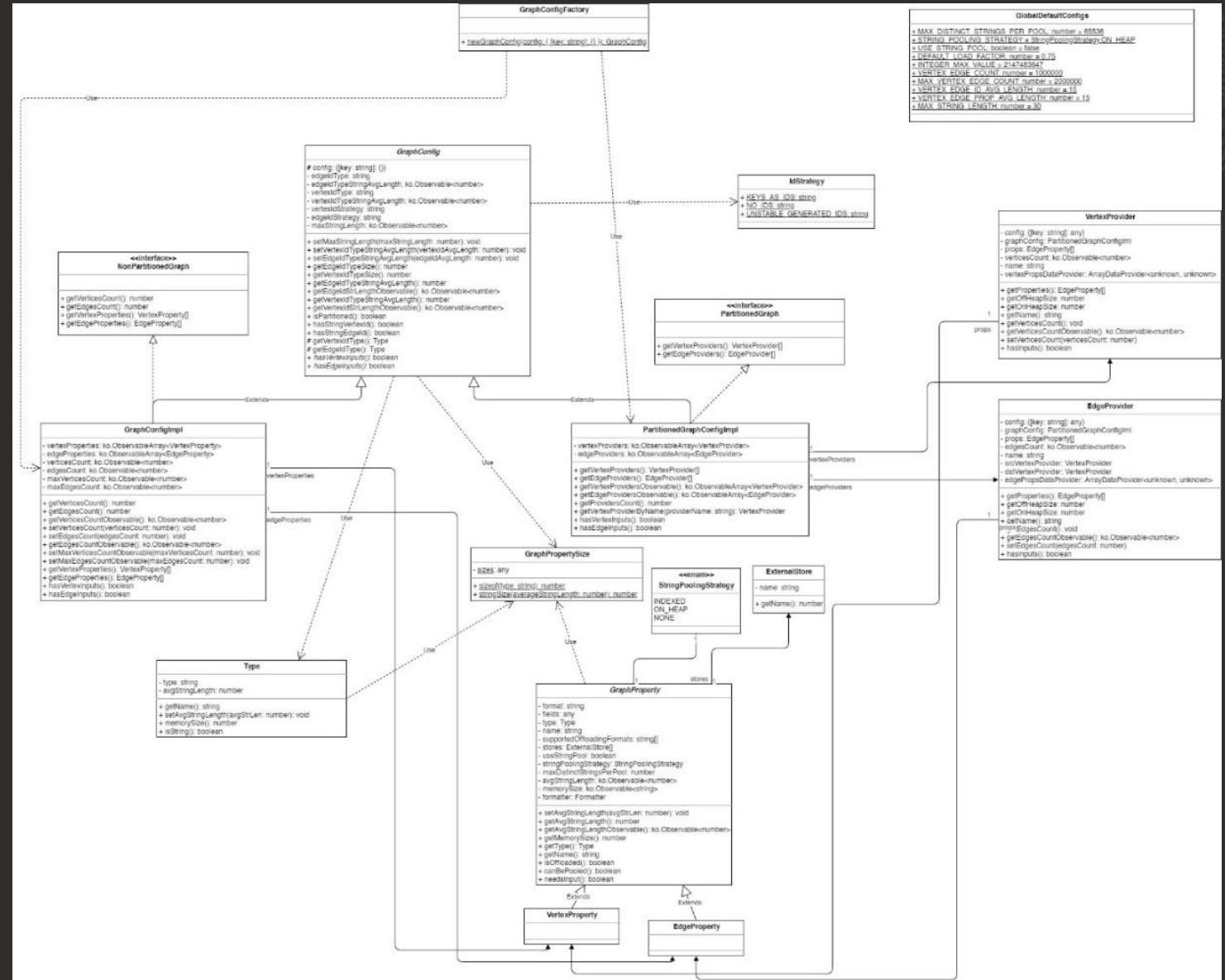


Graph Size Calculator

Analysis & Design

Class diagrams :

1. Configuration module

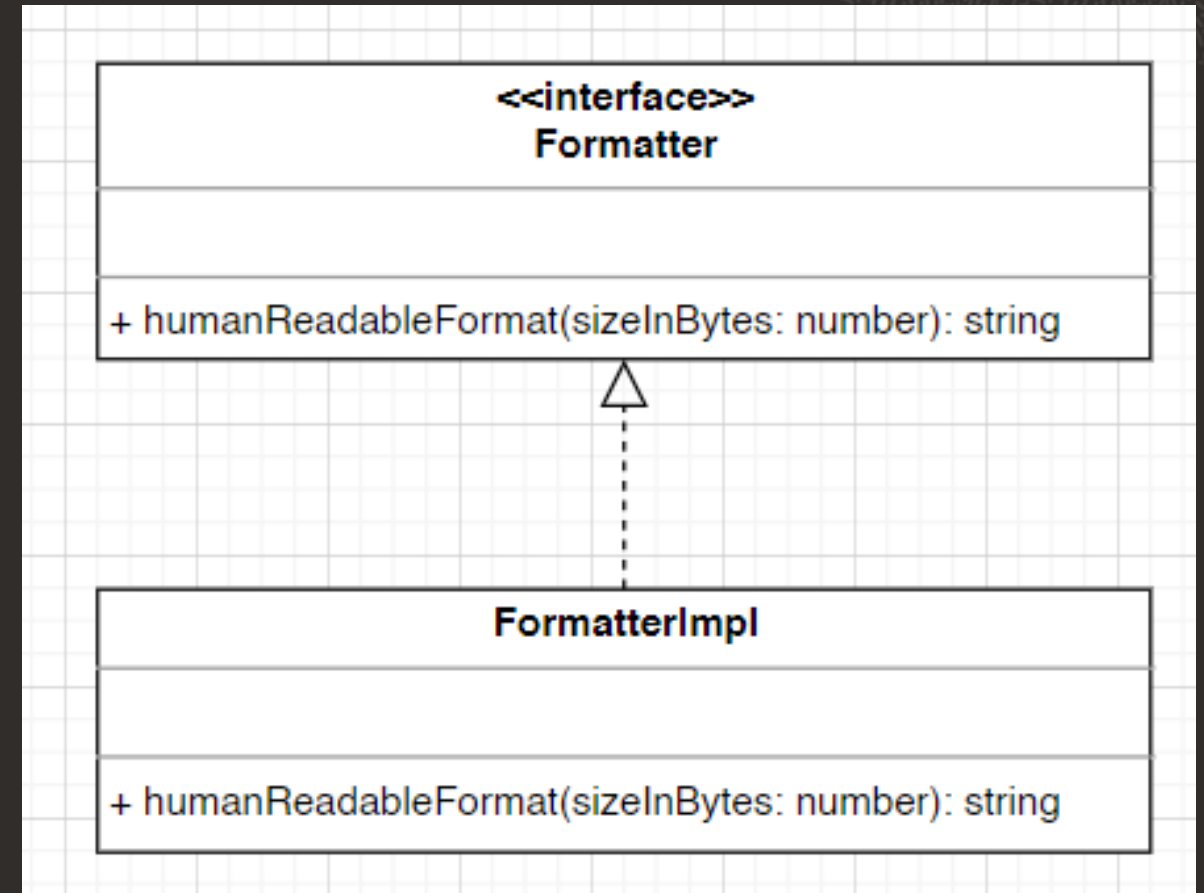


Graph Size Calculator

Analysis & Design

Class diagrams :

2. Formatting module

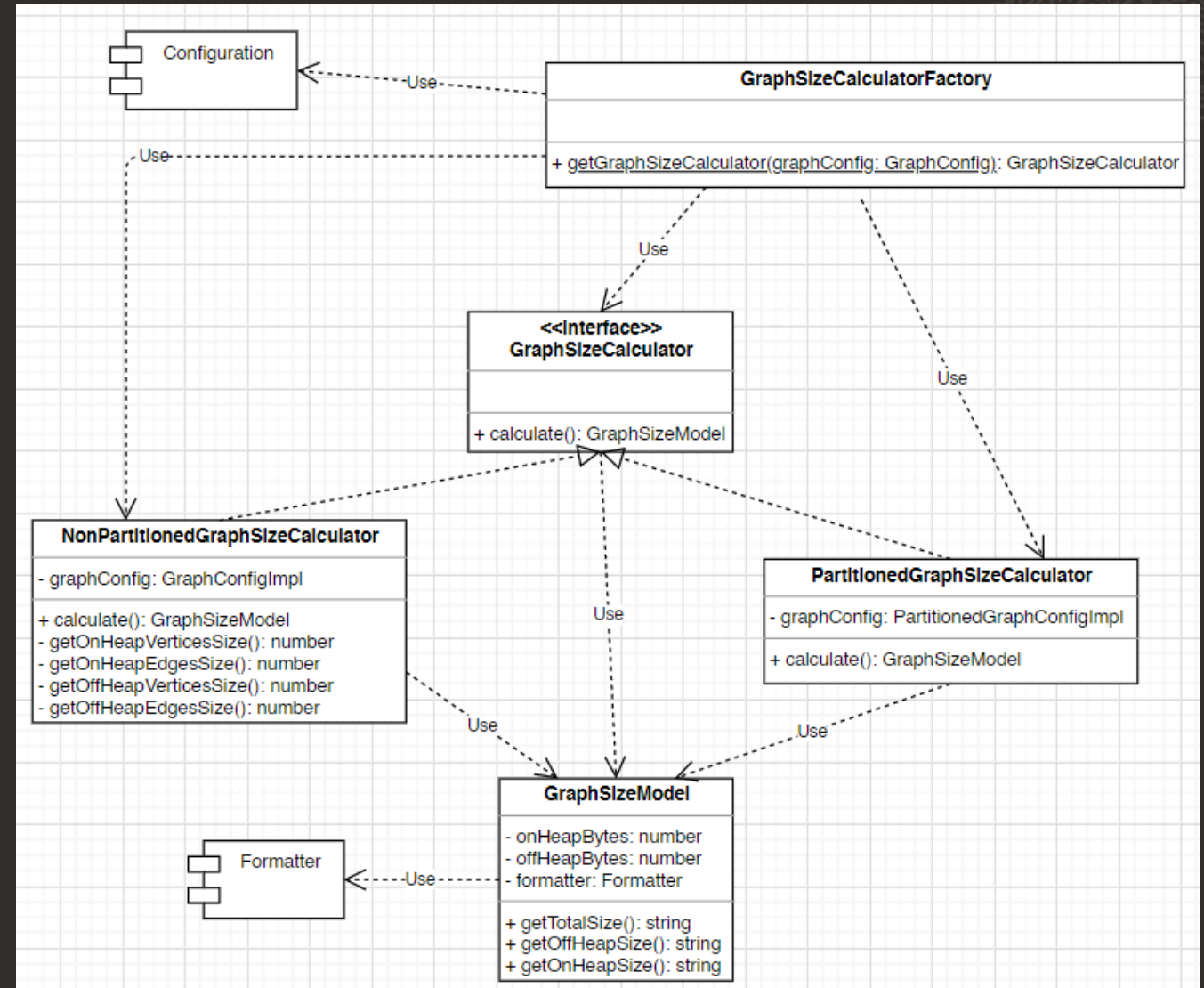


Graph Size Calculator

Analysis & Design

Class diagrams :

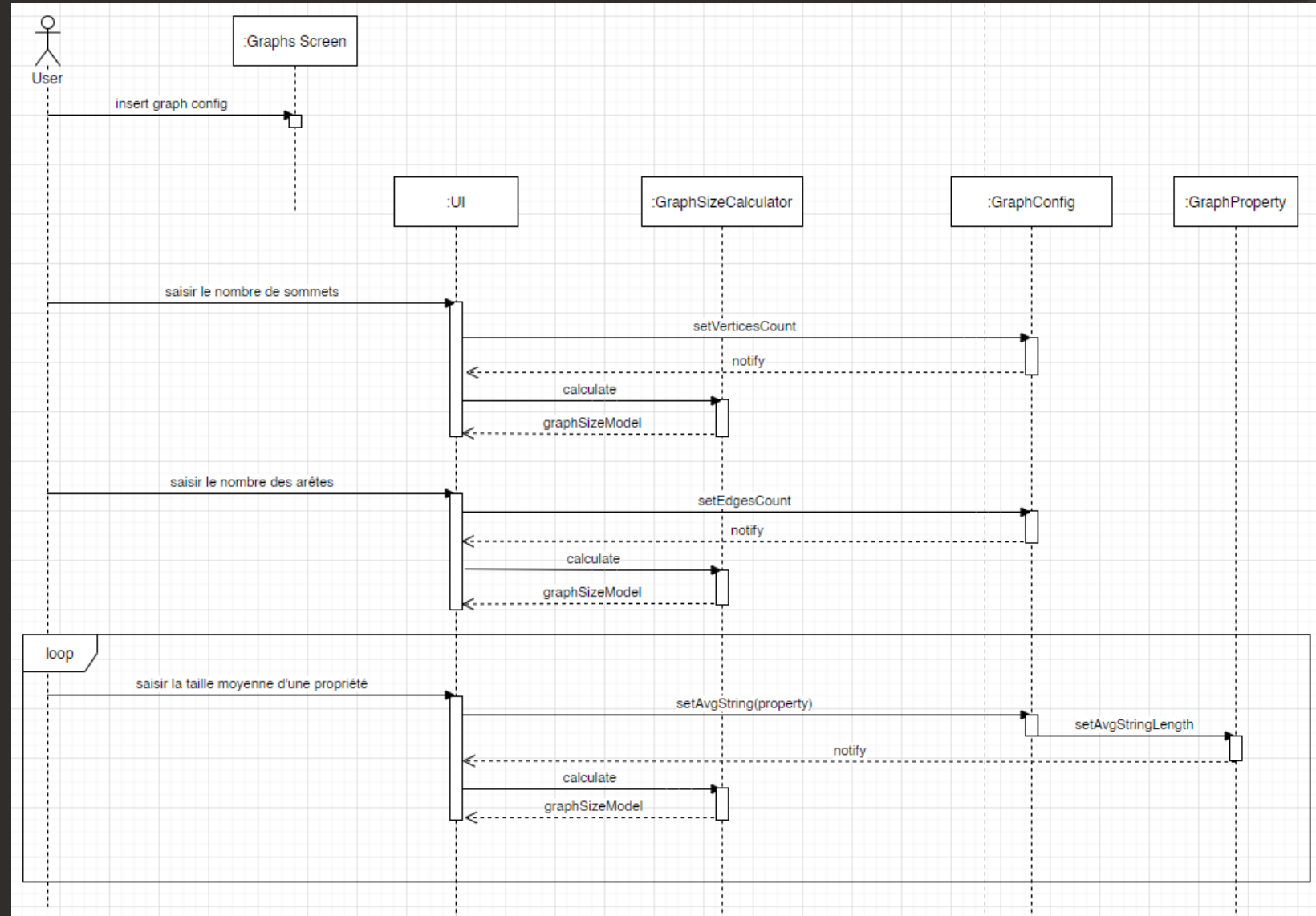
3. Main Module (entry point)



Graph Size Calculator

Analysis & Design

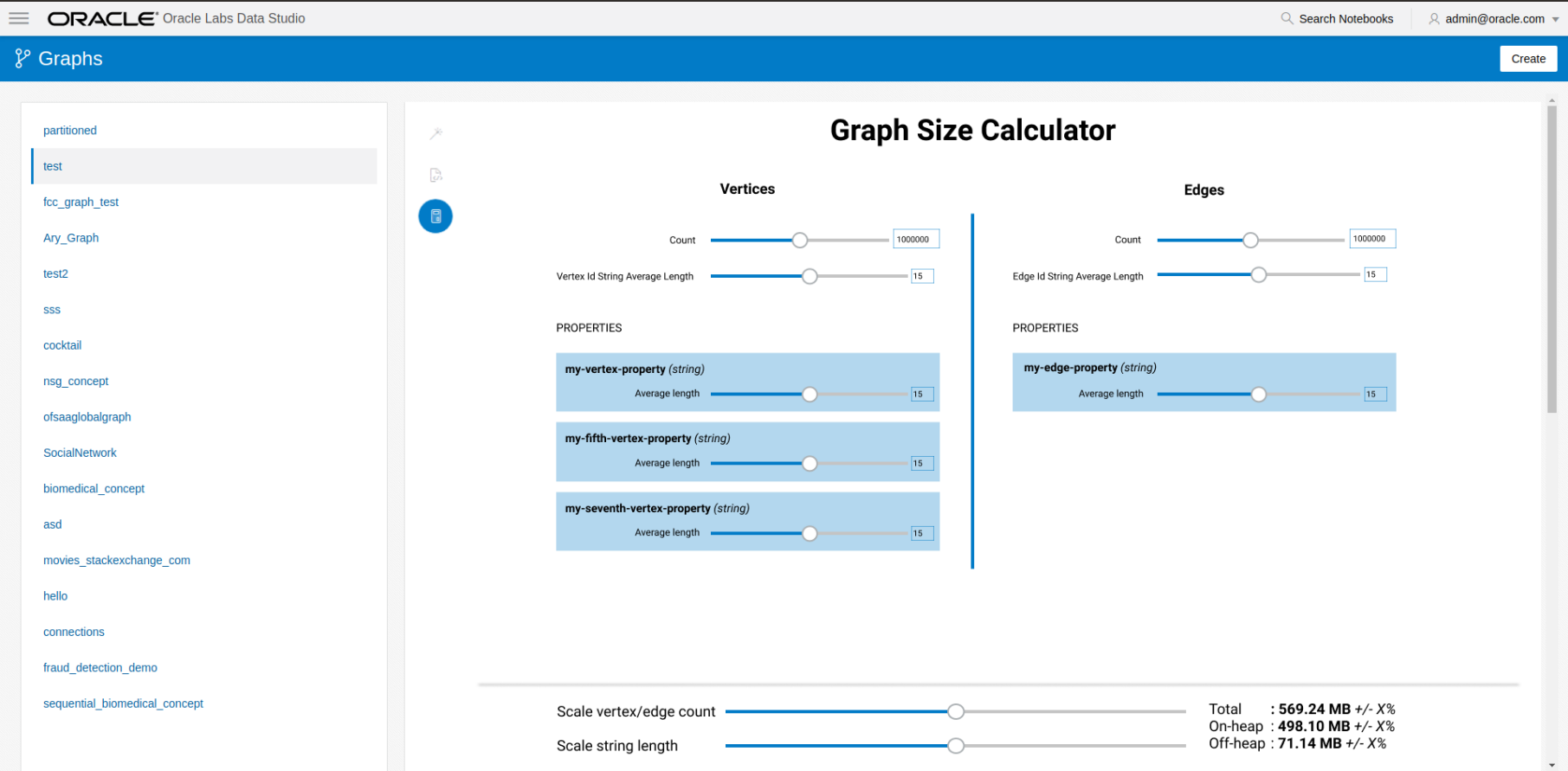
Sequence Diagram



Graph Size Calculator

UI Mocks

Example of non partitioned graph (homogeneous graph)



Graph Size Calculator

UI Mocks

Example of partitioned graph (heterogeneous graph)

ORACLE Oracle Labs Data Studio

Search Notebooks admin@oracle.com

Graphs>Create

partitioned

test

fcc_graph_test

Ary_Graph

test2

sss

cocktail

nsq_concept

ofsaaglobalgraph

SocialNetwork

biomedical_concept

asd

movies_stackexchange_com

hello

connections

fraud_detection_demo

sequential_biomedical_concept

Graph Size Calculator

Vertices

Vertex Id String Average Length15

Account provider

Count1000000

PROPERTIES

Name (string)

Average length15

Country (string)

Average length15

Customer provider

Count1000000

PROPERTIES

Given Name (string)

Average length15

Email (string)

Average length15

Edges

Edge Id String Average Length15

HasAccount provider

Count1000000

PROPERTIES

Label (string)

Average length15

Currency (string)

Average length15

Another property (string)

Average length15

Scale vertex/edge count

Scale string length

Total : 569.24 MB +/- X%

On-heap : 498.10 MB +/- X%

Off-heap : 71.14 MB +/- X%

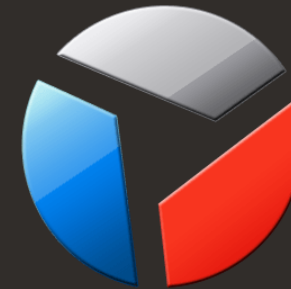


Graph Size Calculator

Development



TypeScript is an open-source programming language developed and maintained by Microsoft. It is a strict syntactical superset of JavaScript and adds optional static typing to the language. TypeScript is designed for development of large applications and transcompiles to JavaScript.



Oracle JavaScript Extension Toolkit (Oracle JET) is a modular JavaScript development toolkit helping developers build engaging user interfaces. Oracle JET is based on KnockoutJs which is a ViewModel-ModelView library

Graph Size Calculator

Tests



Jest is a JavaScript testing library maintained by Facebook. with a focus on simplicity. It works with projects using: Babel, TypeScript, Node.js, React, Angular and Vue.js. It aims to work out of the box and config free.

Graph Size Calculator

Accuracy tests

Config	Vertices	Edges	Actual String length	Experimental size	String length used	Estimated size
Social cloud	2880870	12658205	11.088	<ul style="list-style-type: none">Total : 926 MBOn-Heap : 210 MBOff-Heap : 716 MB	11	<ul style="list-style-type: none">Total : 916.21 MBOn-Heap : 175.83 MBOff-Heap : 740.38 MB
Customers Transactions	56856811	94453354	14.872	<ul style="list-style-type: none">Total : 11.14 GBOn-Heap : 3.93 GBOff-Heap : 7.21 GB	15	<ul style="list-style-type: none">Total : 11.17 GBOn-Heap : 4.21 GBOff-Heap : 6.96 GB
Page links	8637721	165049964	10.862	<ul style="list-style-type: none">Total : 5.93 GBOn-Heap : 2.81 GBOff-Heap : 3.13 GB	11	<ul style="list-style-type: none">Total : 6.26 GBOn-Heap : 2.20 GBOff-Heap : 4.06 GB
Frappe RDBMS	5018583	63819065	15.418	<ul style="list-style-type: none">Total : 4.57 GBOn-Heap : 989.91 MBOff-Heap : 3.60 GB	15	<ul style="list-style-type: none">Total : 5.28 GBOn-Heap : 1.23 GBOff-Heap : 4.05 GB

Graph Size Calculator

Demo

The screenshot displays the Oracle Labs Data Studio interface for managing graphs. The browser address bar shows the URL `100.107.236.236:30648/?root=graphs&graph=ds8aG53K`. The page header includes the Oracle Labs Data Studio logo and a search bar. The main content area is titled "Graphs" and features a sidebar on the left with a list of graphs: `biomedical_concept`, `cocktail` (selected), `connections`, `fraud_detection_demo`, `movies_stackexchange_com`, and `sequential_biomedical_concept`. The right panel shows the configuration for the selected `cocktail` graph. It includes fields for the graph's name, format, and URI, along with sections for defining vertex and edge properties. The `cocktail` graph is configured with the format `Edge list`, vertex property `name (String)`, and edge properties `quantity (Double)` and `unit (String)`. The URI is set to `https://artifacthub.oraclecorp.com/olds`. The `Advanced Mode` toggle is currently disabled. At the bottom of the configuration panel are buttons for `Update`, `Share`, and `Delete`. The Windows taskbar at the bottom shows the time as 7:25 PM on 06/07/2020.

Pull Request #2929: [WIP] Featur... x Graphs | Oracle Labs Data Studio x +

Not secure | 100.107.236.236:30648/?root=graphs&graph=ds8aG53K

ORACLE® Oracle Labs Data Studio

Search Notebooks admin@oracle.com

Graphs Create

biomedical_concept

cocktail

connections

fraud_detection_demo

movies_stackexchange_com

sequential_biomedical_concept

* Name cocktail

Format Edge list

Vertex Properties name (String) Add Vertex Property +

Edge Properties quantity (Double) unit (String) Add Edge Property +

* URI https://artifacthub.oraclecorp.com/olds

Attributes Add Attributes +

Advanced Mode

Update Share Delete

4.2 Data Studio Load Tests

-
- What is load tests?
 - Goal
 - Our test plans
 - Results

Data Studio Load Tests

What is load tests

Load testing is the process of putting demand on a system and measuring its response.

Load testing generally refers to the practice of modelling the expected usage of a software program by simulating multiple users accessing the program concurrently.

Data Studio Load Tests

Goal

The goal is to test the performance of Data Studio when users are dealing with PGX graphs.

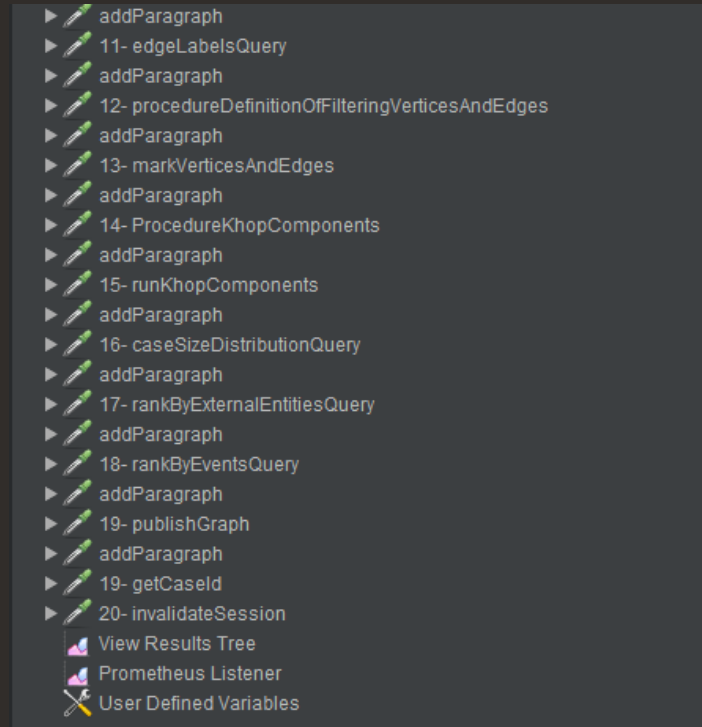
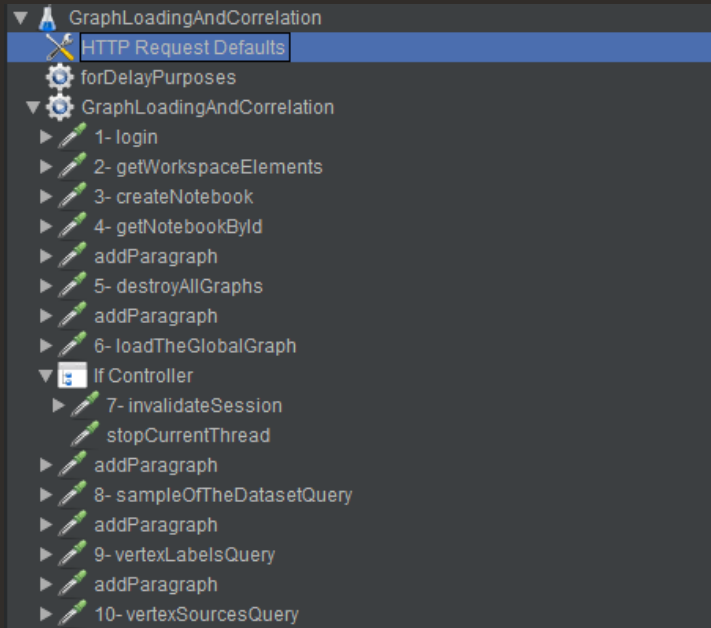
How?

Create a benchmark using 2 notebooks, the first will generate a graph which represents the transactions of a given bank, it also generates cases that needs investigation.

The second one will investigate those cases in order to detect frauds.

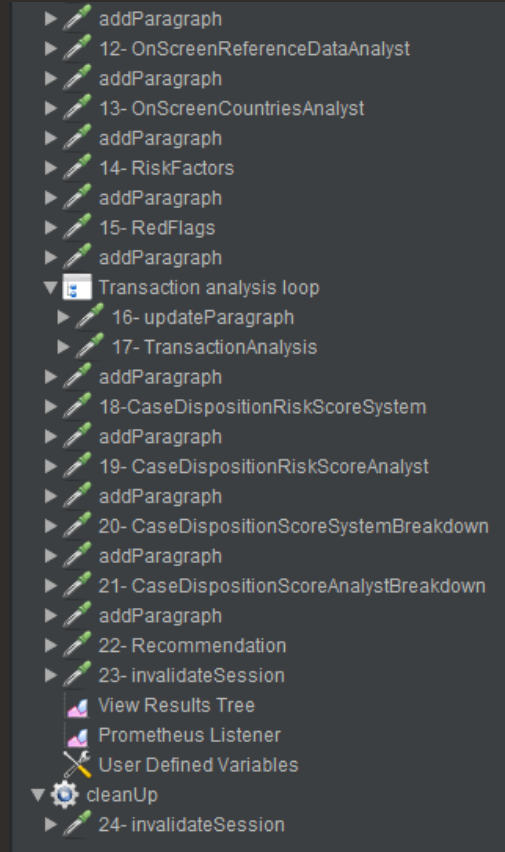
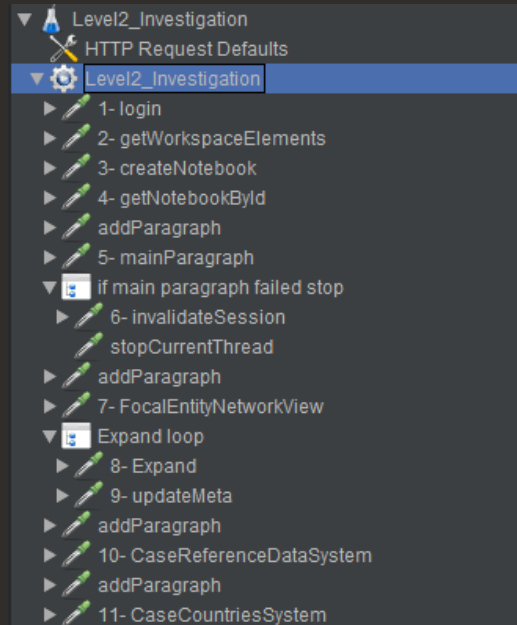
Data Studio Load Tests

1st Test Plan

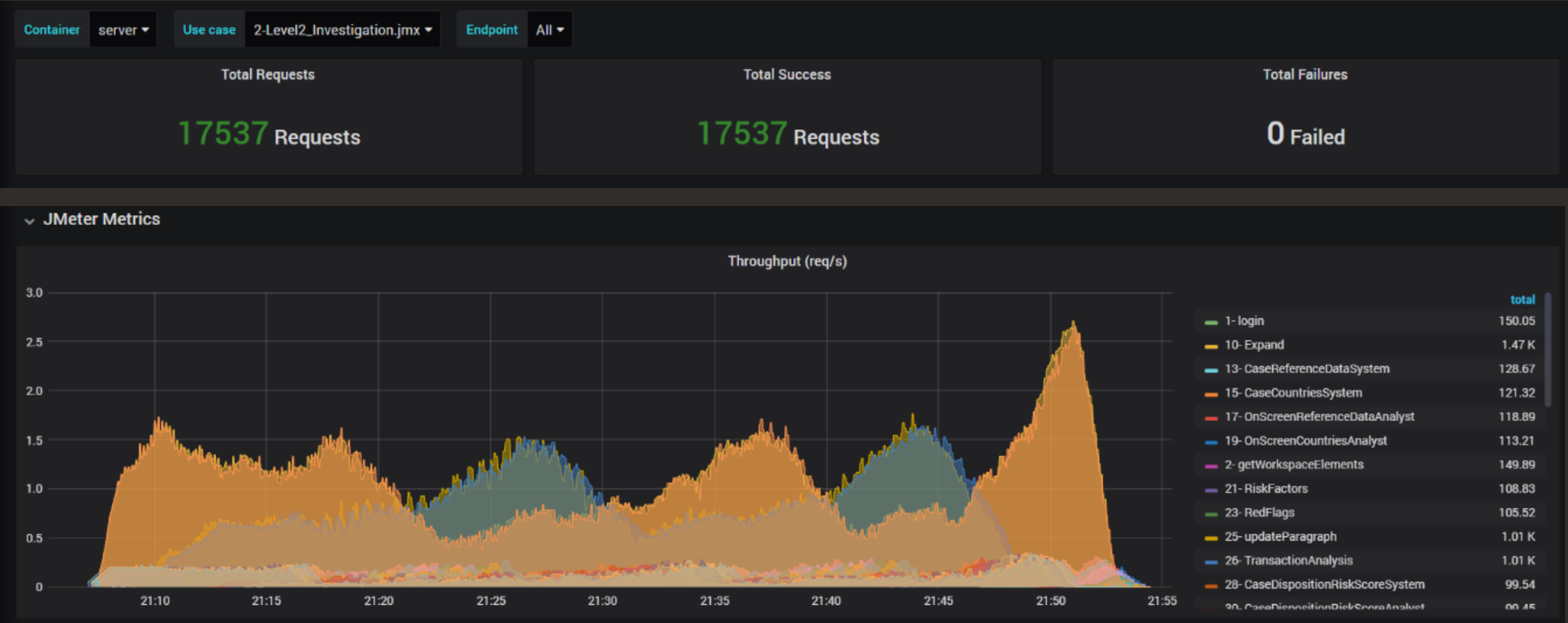


Data Studio Load Tests

2nd Test Plan



Data Studio Load Tests Results



5. Conclusion

- PFE in numbers
- Contribution of the internship

Conclusion

PFE in numbers



Meetings

33



Tickets

12



Working Days

108

(22 weeks)



Lines of code

15,874



Unit/integration Tests written

> 50



Reviewed PRs

1

Conclusion

Contribution of the internship



Thank you!



Q/A



Oracle PGX Graph Size Calculator And Data Studio Load Tests

By : Mohamed Boukhlif

School mentors

- Mrs. Nezha Benmoussa
- Mr. Mohammed Qbadou

Company mentor

- Mrs. Rania Medraoui