A Precise Model for Google Cloud Platform

Stéphanie Challita | Faiez Zalila | Christophe Gourdin | Philippe Merle

Inria Lille – Nord Europe & University of Lille

6th IEEE International Conference on Cloud Engineering (IC2E 2018)

Google Cloud Platform
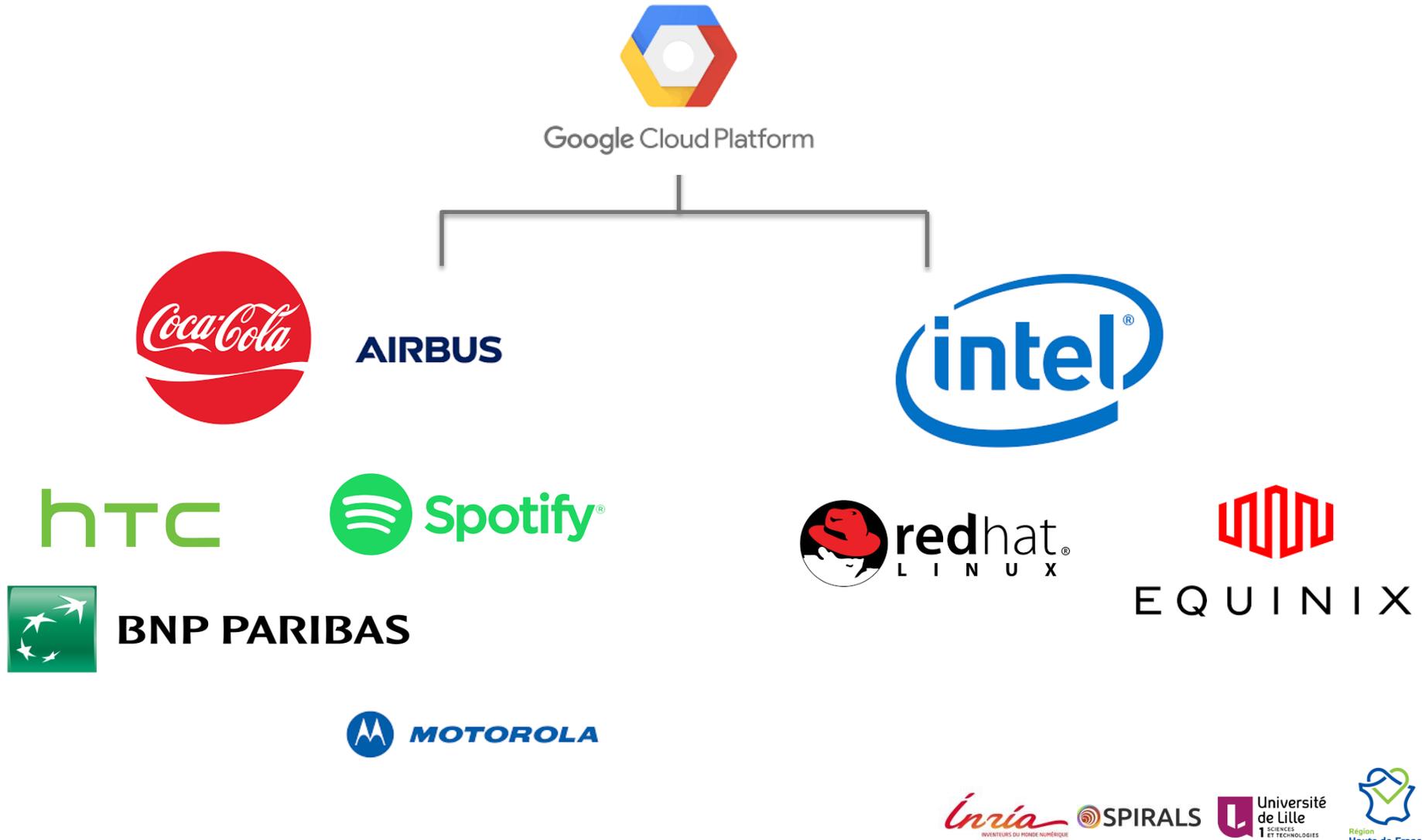
Google Cloud Platform

An agreement with the developer on exactly how the system will operate

GCP documentation

Read

Develop

conforms to

Cloud developer/designer   **Cloud provider**
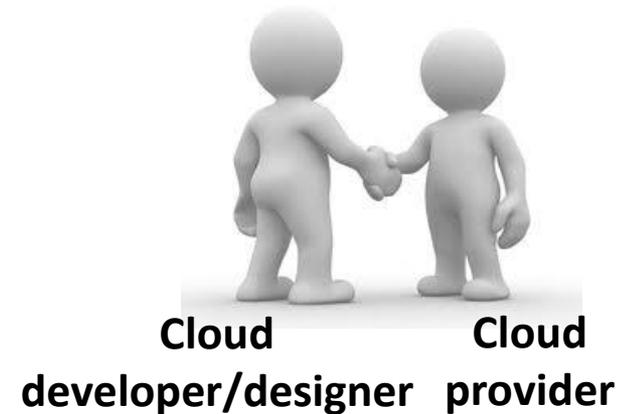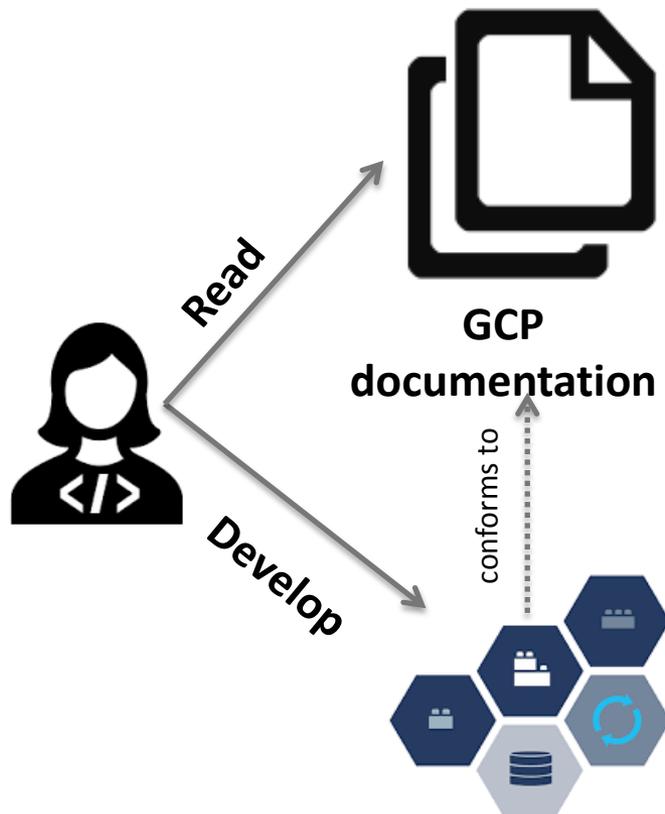
GCP documentation is written in natural language → human errors and/or semantic confusions

Agenda

- **Drawbacks & Motivations**

- Contribution

- Perspectives

Drawbacks & Motivations

List of Drawbacks

- Informal Heterogeneous Documentation

- Imprecise Types

- Implicit Attribute Metadata

- Hidden Links

- Redundancy

- Lack of Visual Support

Informal Heterogeneous Documentation

Property name	Value	Description	Notes
IPv4Range	string	The range of internal addresses that are legal on this network. This range is a CIDR specification, for example: 192.168.0.0/16. Provided by the client when the network is created.	

Available at
https://cloud.google.com/compute/docs/reference/latest/networks

Vs.

Fields	
projectId	string
	Required. The Google Cloud Platform project ID that the cluster belongs to.

Available at
https://cloud.google.com/dataproc/docs/reference/rest/v1/projects.regions.clusters

Imprecise Types

Lack of Visual Support

- Only **descriptive information** in GCP documentation
  - A huge time to be properly understood and analyzed

- Visual diagrams easily highlight in short but catchy view the concepts of the API
  - Help to avoid wastage of time
  - Logical sequence and comparative analysis can be undertaken

Contribution

Global Vision

- Precise model for GCP

- Work of reverse-engineering, HTML ➡ Model

- GCP model refinement

GCP Snapshot

GCP HTML pages

Snapshot
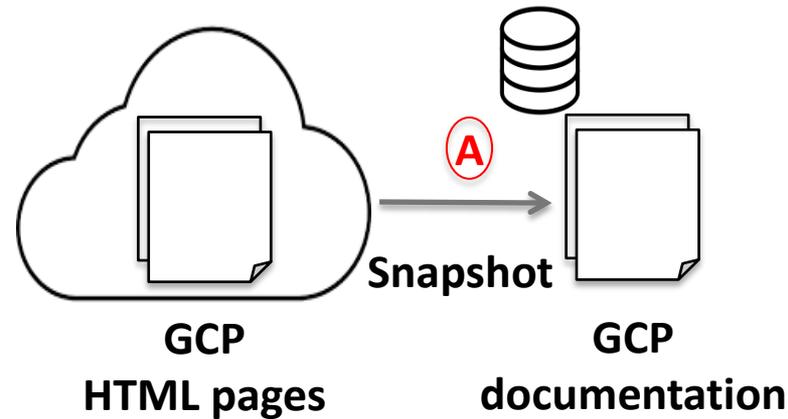
GCP documentation

- GCP engineers could update/correct GCP documentation

- Continuously following up with GCP documentation is costly

- Snapshot of GCP API

GCP Crawler & GCP Model
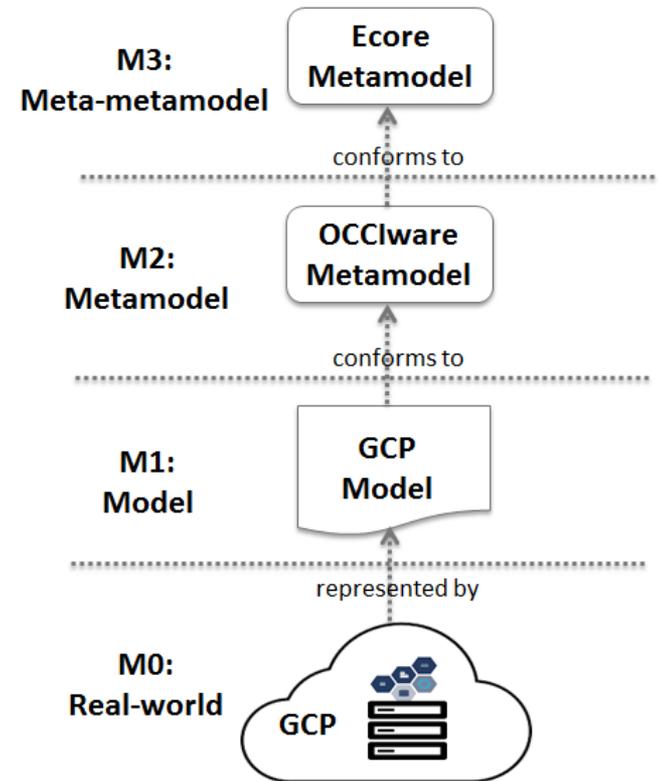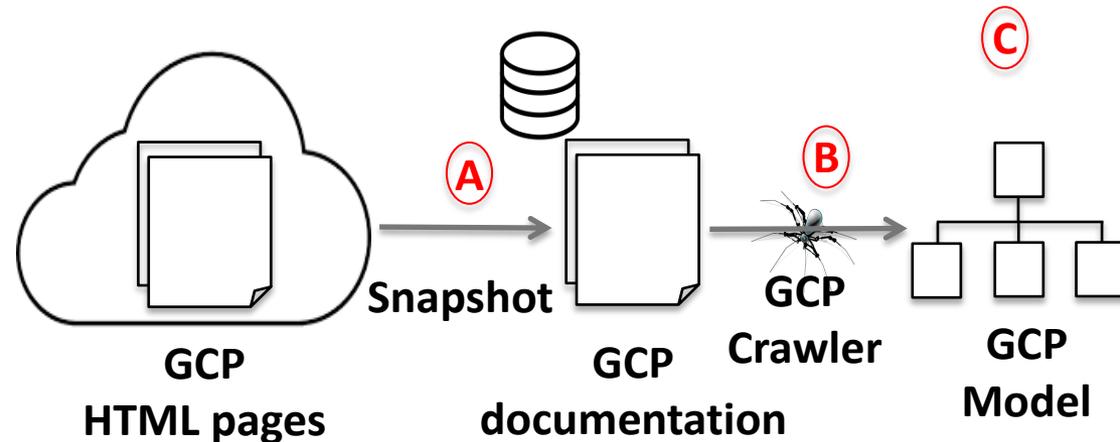
- **GCP Crawler** to extract all GCP resources, their attributes and actions

- **GCP Model** for a better description of the GCP resources and for reasoning over them

No more Informal Heterogeneous Documentation 👍

Model Transformations

GCP HTML pages → **Snapshot** (A) → GCP documentation → **GCP Crawler** (B) → GCP Model (C)

Model Transformations:
- Type Refinement
- Implicit Attribute Metadata Detection
- Link Identification
- Redundancy Removal
- Model Visualization

Model Transformations
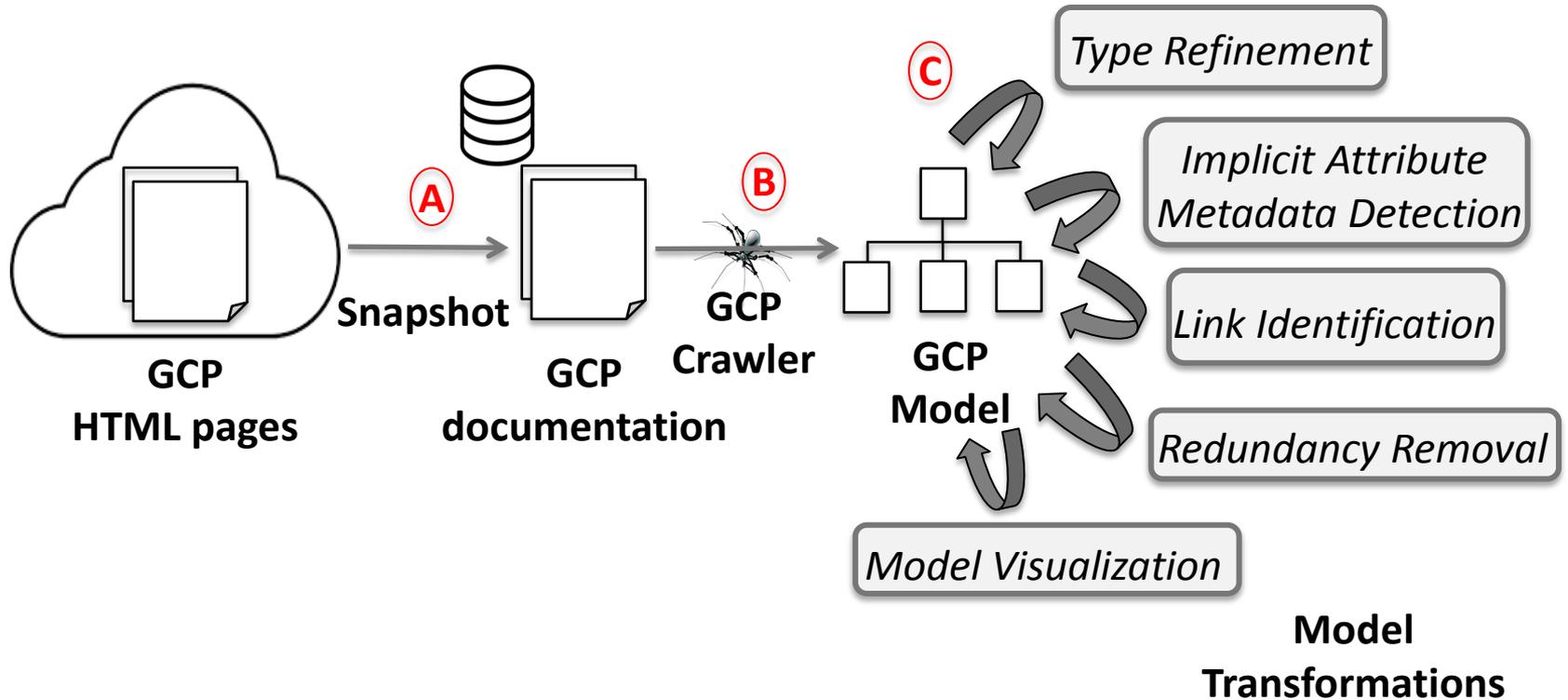
Type Refinement

- By adopting the data type system proposed by OCCIware metamodel
  - defining regular expressions,
  - and using the EMF validator to check the type constraints that are attached to the attributes

- If the type of an attribute in the documentation is string and the description explains that this is an email address, we apply the **email validation constraint** :
  - STRINGTYPE + this **regular expression**:

    ^[A-Z0-9._%+-]+@[A-Z0-9.-]+\\.[A-Z]{2,6}$

  - **No more Imprecise Types** 👍

Model Visualization

For an easier understanding and analysis of the API, because the insights become obvious
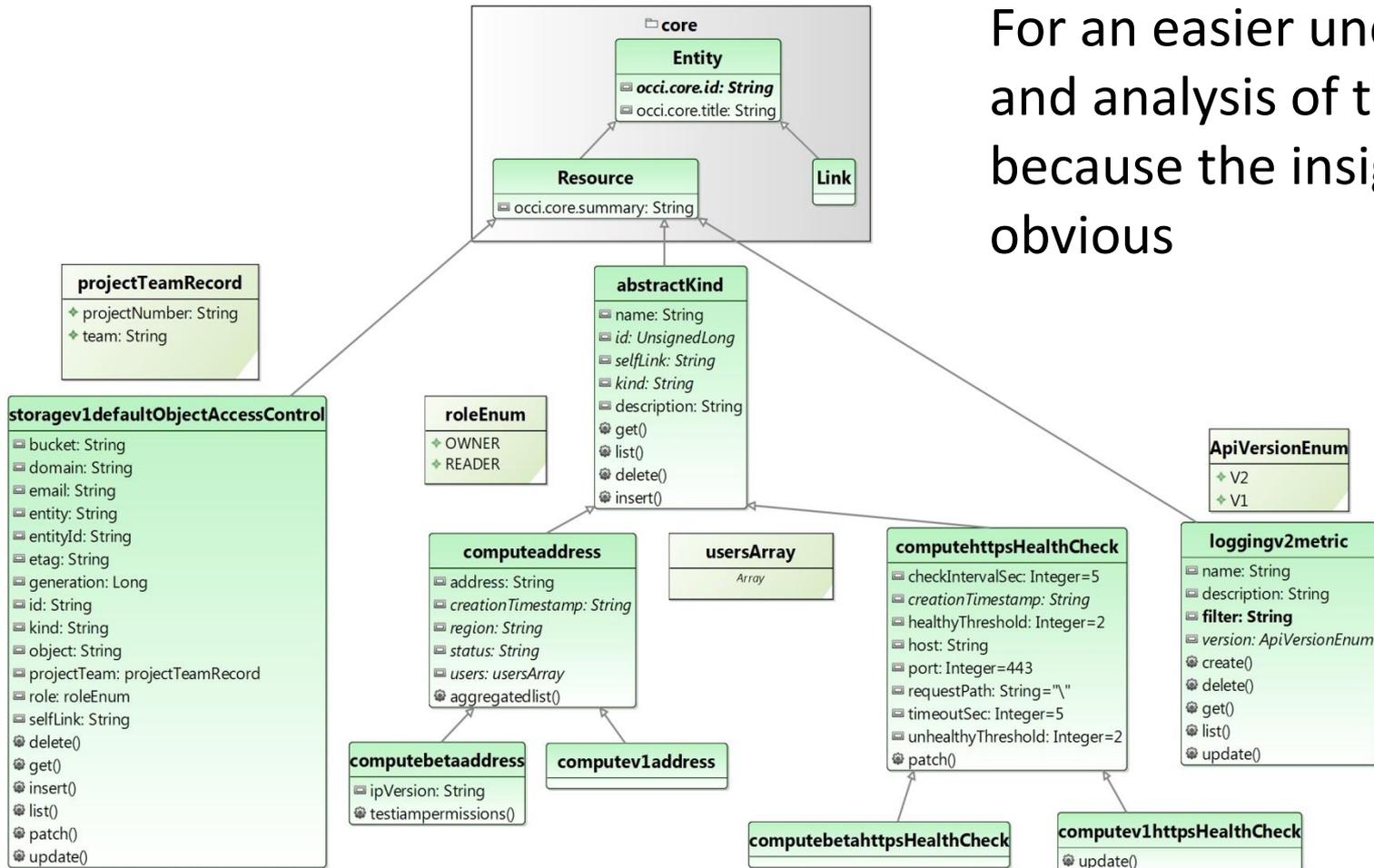
No more Lack of Visual Support 👍

Analysis

- "GCP documentation is developed by two separate clusters of development teams"
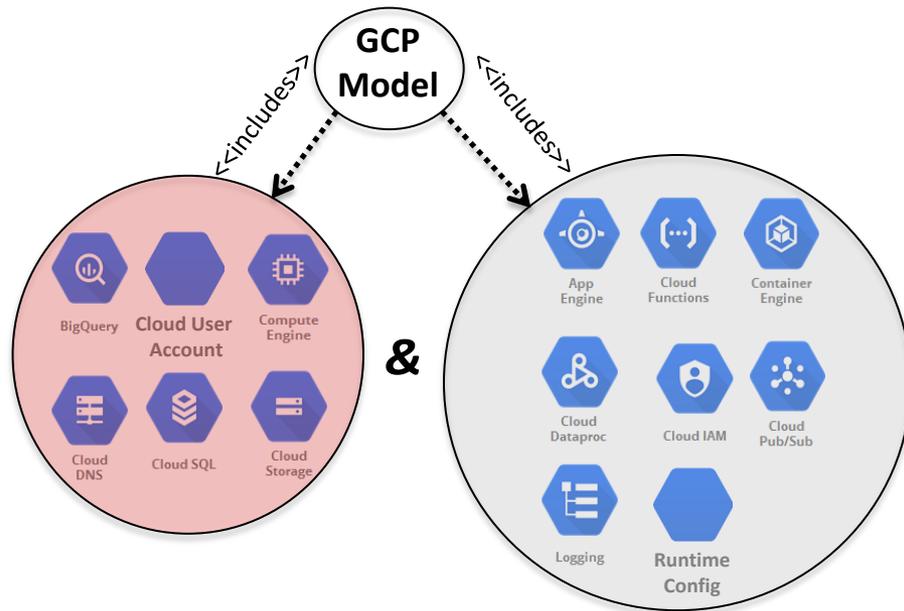
- "GCP documentation is redundant and not comprehensive"

Redundant Attributes	Before Abstraction		After Abstraction	
	# of occurrences	% of redundancy	# of occurrences	% of redundancy
name	92	64,79%	26	18,31%
id	80	56,34%	14	9,86%
selfLink	79	55,63%	13	9,15%
kind	79	55,63%	13	9,15%
description	75	52,82%	9	6,34%
Average		57,04%		10,56%
Redundant Actions	# of occurrences	% of redundancy	# of occurrences	% of redundancy
get	142	100,00%	76	53,52%
list	142	100,00%	76	53,52%
delete	140	98,59%	74	52,11%
insert	76	53,52%	10	7,04%
Average		88,02%		41,54%

Perspectives

In progress…

- Validation with Google

- **GCP Studio**, a dedicated model-driven environment for GCP

- Generation of **GCP artifacts** such as JSON files and CURL scripts

- Model-driven management of GCP systems

Thank you!

stephanie.challita@inria.fr
www.occiware.org
https://github.com/occiware/GCP-Model