Technical Integration Document

April 30th 2018

INTRODUCTION

This document outlines the different scenario's in which the KeenCorp software can integrate with a client infrastructure.

KeenCorp has two commercial offerings, which are often combined in a Proof of Concept (PoC):

- Retrospective Analysis
- Live installation

A typical PoC process consists of an analysis on a selected part of the organization. The KeenCorp software will be used to analyze a minimum of 2 years of historic data (Retrospective Analysis) after which the software will run in real-time for a 12-month period (Live Installation).

During the PoC, the KeenCorp signal will measure employee engagement levels as expressed in the KeenCorp Index of a minimum of around 500 employees. Digital communication will be measured from emails (or optionally company chat, if available).

To perform a retrospective analysis a client provides a data set of digital communication, for example e-mail (KeenCorp can assist in this process). This will render initial insight into the client's organizational, engagement and culture dynamics and provide a baseline for ongoing measurement. This will take place after the Retrospective analysis and will be performed on a continuous basis.

The Retrospective analysis is conducted on the whole data set and on defined 'clusters' of employees, according to client specifications. KeenCorp will support a client in defining the relevant clusters based on client priorities (to be defined), PoC objectives and best practices. Each cluster will contain a minimum of 10 employees (privacy by design) and employees can be part of multiple clusters (gender, age, educational level, office, seniority, project team, etc.)

We distinguish two sub-processes to perform the KeenCorp analysis:
- pre-processing, where digital communication is grouped and anonymized, and
- KeenCorp Scoring that deals with the analysis.

Pre-processing (client side)
The first pre-processing step happens before any data is sent. This first step concerns detection and anonymization of the sender and receiver of the message. Every individual message will solely be linked to a predefined cluster (or clusters) so that an individual message can never be retraced to an individual sender. The portion of the KeenCorp software concerning the handling of the anonymized digital communication data is available as source-code, in order to provide transparent insight and verification of the process followed regarding handling of anonymized digital communication.

KeenCorp Scoring (optionally client side – see slides 5-9)
In order to perform analysis on the pre-processed digital communication, the KeenCorp Server will automatically perform the following actions:
1. Detection of the language of the message;
2. Analysis by a NER (Named Entity Recognition): the names, places and company names within the message will be anonymized (an example of a processed message can be found in Addendum A);
3. Processing of the anonymized message by the KeenCorp algorithm. The message will be processed and the result reduced to a numerical value which in no way can be related back to the original message.

After all data has been processed the resulting scores will be transferred to KeenCorp in order for our professional experts to interpret the results and report back to the client. Reporting will be delivered in a digital or physical format.

The exact method used in performing some of the aforementioned steps is proprietary to KeenCorp and consequently this is not disclosed to third-parties. It is possible to provide the intermediate results of the aforementioned steps to provide more insight in the process and make transparent which data eventually will be stored on the KeenCorp servers. These intermediate results can be provided in the form of the specifically created trace output of the processing method. In addition, we can provide (parts of) the source code concerning the non-confidential part of the software.

The original digital communications in their original form are never in any way stored by KeenCorp; they only exist in the server's volatile memory (RAM) for the duration of a message's processing (typically milliseconds). After the processing of a message is finished, the message is irretrievably erased from the KeenCorp server's volatile memory.

Information to be stored consist of a single numerical value denoting the message score, the "KeenCorp Index", which is combined with a moving average index value per day; the updated average index value is the only data which is stored into the database. This removes even the theoretical possibility of linking any individual measurement to an individual message or employee, as the constructed average cannot be mathematically factorized back into its constituent values.

This process is elaborated on a video clip: https://vimeo.com/298983960

THREE TECHNICAL SCENARIOS FOR THE KEENCORP SERVER PROCESS

KeenCorp offers three technical integration scenarios for the integration of KeenCorp software with the clients infrastructure. Solutions can be tailored to suit the requirements of a client specific infrastructure and security protocols.

The three scenario's are developed to facilitate the legal obligations some of our clients have in the handling of digital communication:
1. Direct secure integration with the KeenCorp private cloud solution;
2. Secure integration with a dedicated hosting server ("private cloud");
3. On-premise integration of a KeenCorp server

Each scenario is Privacy-by-design and GDPR compliant, designed to make sure every step of the process is secure and guaranteeing the highest degree of anonymity.

The scenario's follow the same steps and procedures as described in the previous pages and laid out on the next slide.
The difference between the options is based on the physical location at which different steps of processing take place; this can be either a client's location or a KeenCorp (hosted) location. The choice of scenario influences the hardware requirements, staff capacity of local systems administrators and KeenCorp staff, and maintenance procedures and costs.

The scenarios are described in slides 7, 8 and 9 respectively. Slide 9 gives an overview of the differences in a table to support decision-making.

KEENCORP

KeenCorp can extract data from multiple digital communication platforms: (e.g.)

Office 365 · Exchange · mimecast · G

1 Using extraction scripts ,data is initially downloaded locally (client side). Only sender, receiver and email body content is downloaded (attachments are ignored). Only internal email is processed.

2 In the first step of anonymization the sender and receiver pair are anonymized to (a) specific cluster(s). The result of this is a pseudonymized file.

3 4 5 The pseudonymized file is transferred to KeenCorp. Depending on the scenario this is done
(a) directly (data transfer protocol),
(b) indirectly (to a dedicated hosting environment) or
(c) to a on-premise KeenCorp server (on-site integration).

On the KeenCorp infrastructure the pseudonymized file is processed according to the following steps:

1 Extraction script

2 First step of anonymization

3 Pseudonymized file

4 Data transfer

5 Pseudonymized file
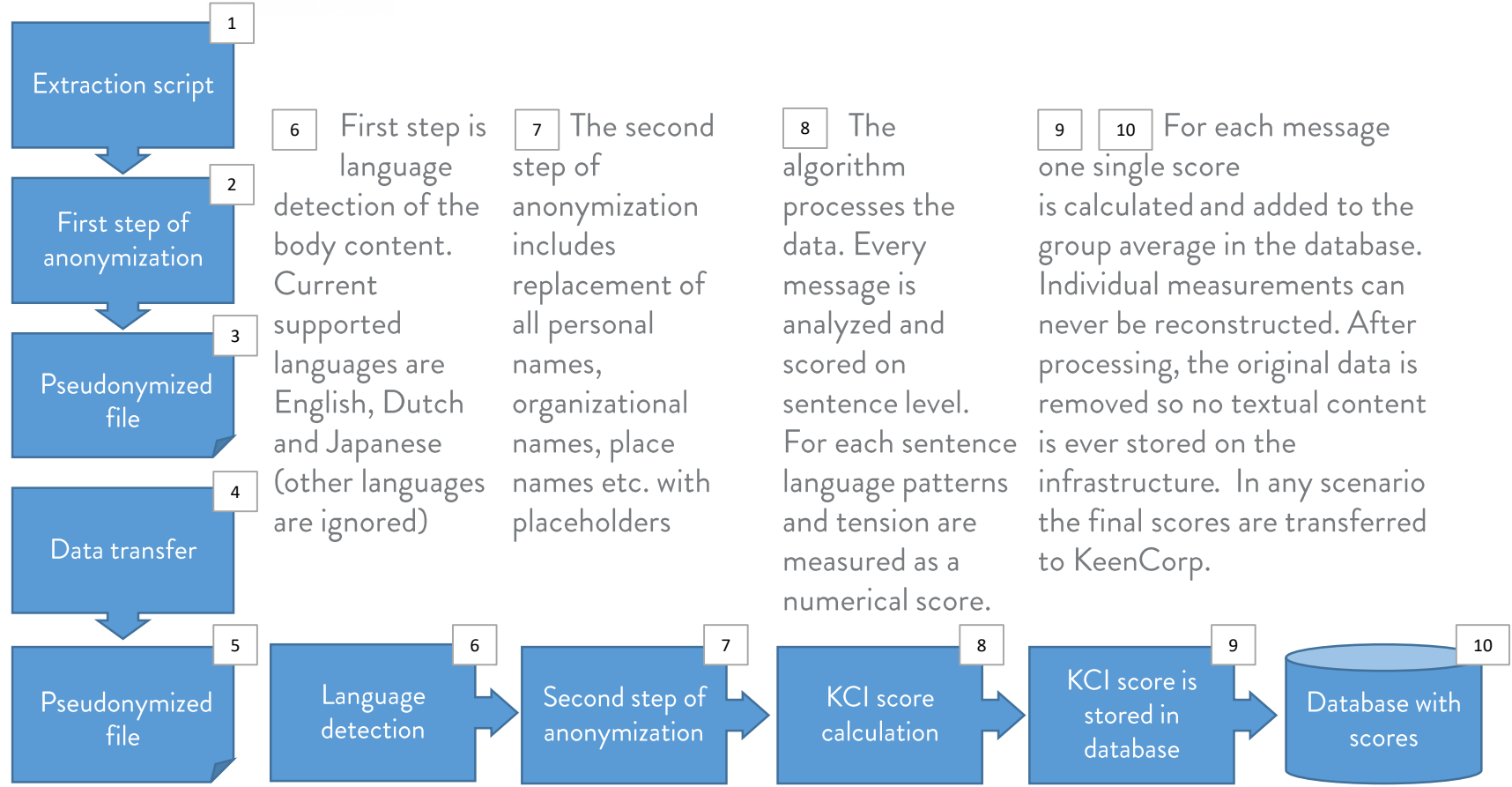
6 First step is language detection of the body content. Current supported languages are English, Dutch and Japanese (other languages are ignored)

7 The second step of anonymization includes replacement of all personal names, organizational names, place names etc. with placeholders

8 The algorithm processes the data. Every message is analyzed and scored on sentence level. For each sentence language patterns and tension are measured as a numerical score.

9 **10** For each message one single score is calculated and added to the group average in the database. Individual measurements can never be reconstructed. After processing, the original data is removed so no textual content is ever stored on the infrastructure. In any scenario the final scores are transferred to KeenCorp.

6 Language detection → **7** Second step of anonymization → **8** KCI score calculation → **9** KCI score is stored in database → **10** Database with scores

SCENARIO 1: DIRECT SECURED INTEGRATION WITH THE KEENCORP PRIVATE CLOUD SOLUTION
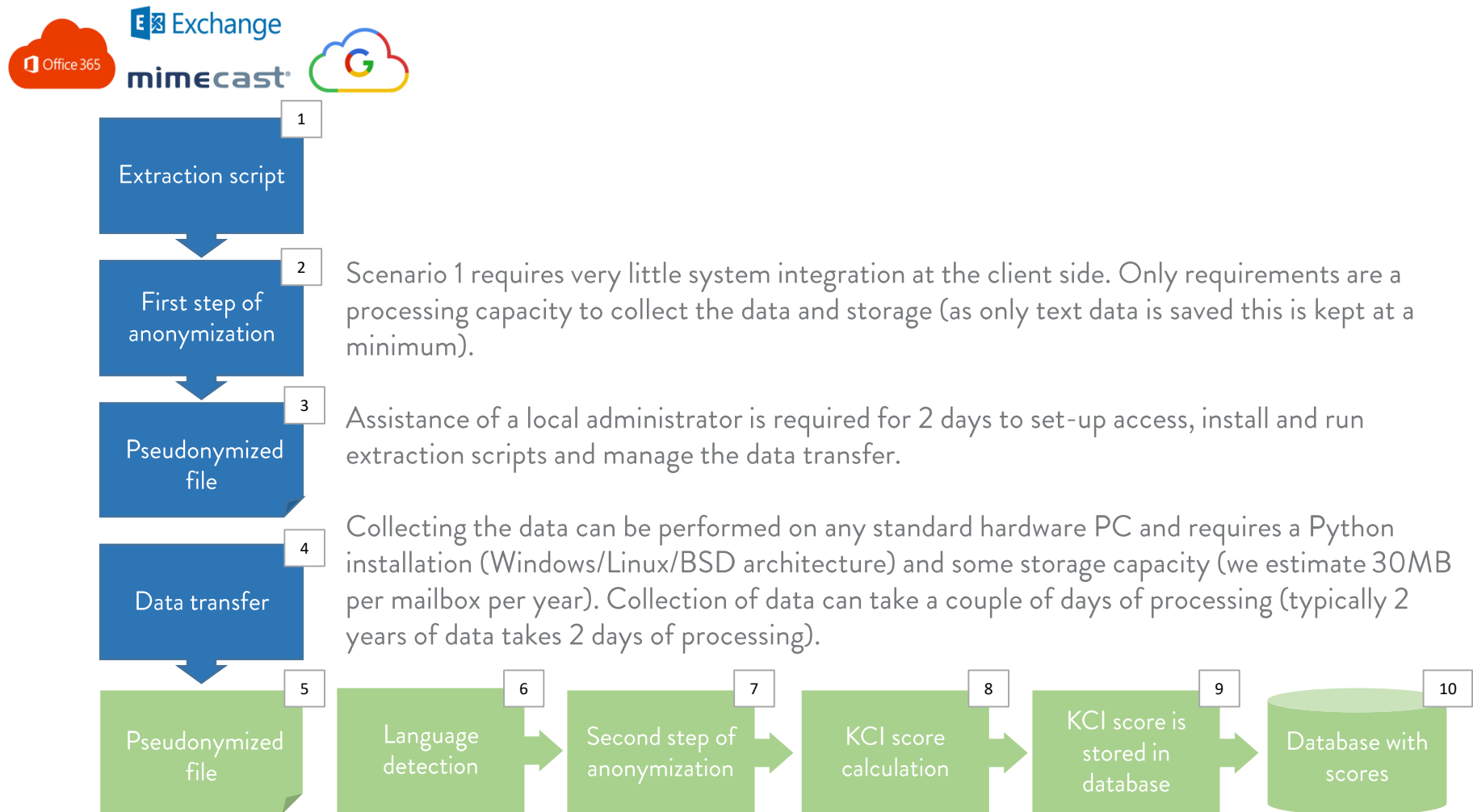
In the first scenario a direct integration with KeenCorp's private cloud is set-up in a secure way.

At client's side the data is collected and the first step of anonymization is performed (replacement of sender-receiver pair with a cluster).

After the data has been collected the data is transferred to KeenCorp and processed on KeenCorp servers.

Data transfer can be performed physically (encrypted USB), digitally (Secure file transfer), or in a fully automated fashion (secured connection).

1 Extraction script

2 First step of anonymization

Scenario 1 requires very little system integration at the client side. Only requirements are a processing capacity to collect the data and storage (as only text data is saved this is kept at a minimum).

3 Pseudonymized file

Assistance of a local administrator is required for 2 days to set-up access, install and run extraction scripts and manage the data transfer.

4 Data transfer

Collecting the data can be performed on any standard hardware PC and requires a Python installation (Windows/Linux/BSD architecture) and some storage capacity (we estimate 30MB per mailbox per year). Collection of data can take a couple of days of processing (typically 2 years of data takes 2 days of processing).

5 Pseudonymized file

6 Language detection

7 Second step of anonymization

8 KCI score calculation

9 KCI score is stored in database

10 Database with scores
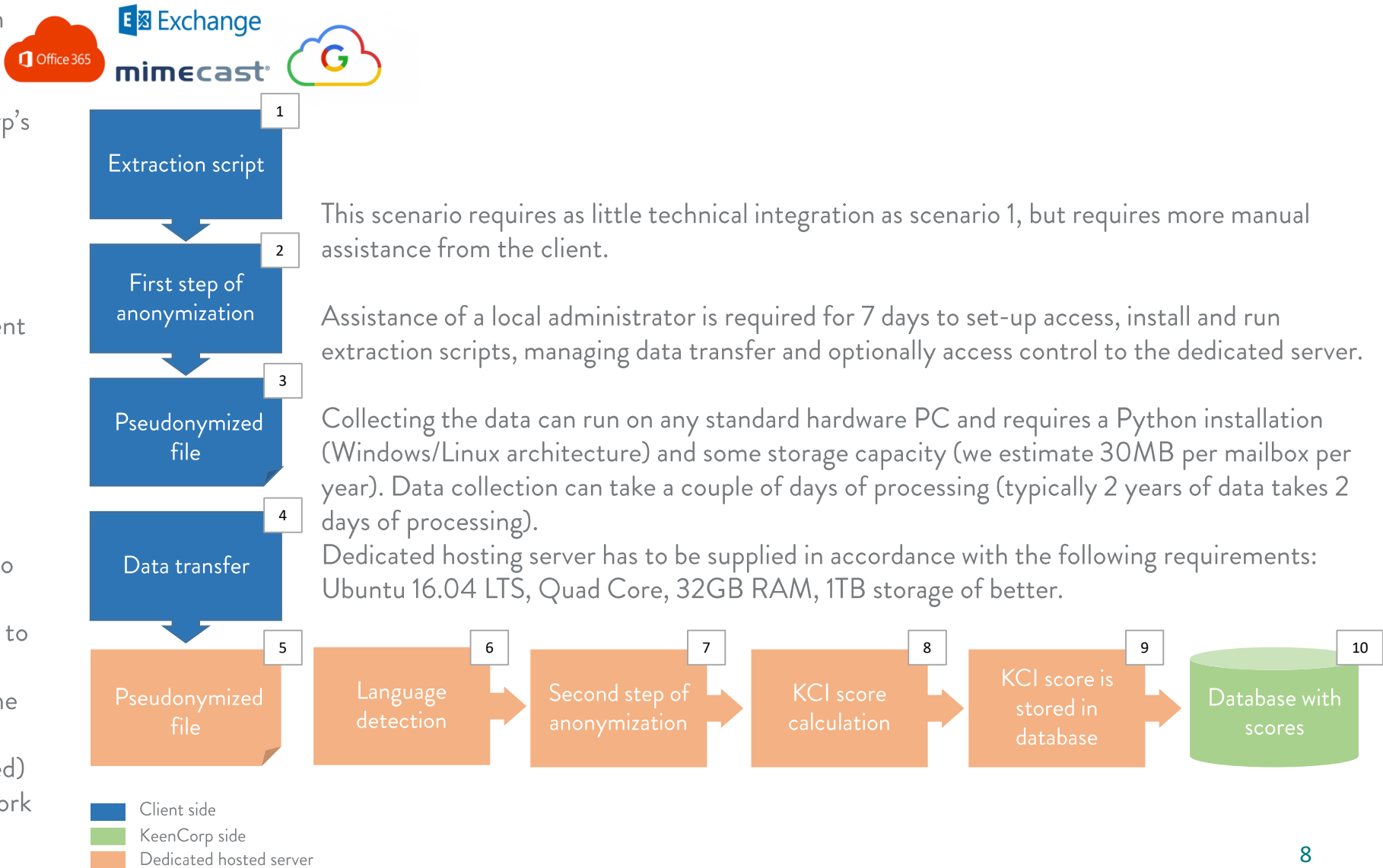
- Client side
- KeenCorp side

SCENARIO 2: SECURED INTEGRATION WITH A DEDICATED HOSTED SERVER

In the second scenario a hosted solution is provisioned at a third-party hosting provider (in consultation with the client). A dedicated version of KeenCorp's platform is installed at the hosted infrastructure.

As in scenario 1 data is collected at the client location and the first step of anonymization is performed (replacement of sender-receiver pair with a cluster).

After the data is collected this is transferred to the dedicated hosted solution.

Important consideration for this scenario is that, after installation at the hosted solution is complete, KeenCorp's access to the hosting solution can be limited until the data is transferred (by client) and the analysis is started (by client). After analysis is complete (and data is removed) KeenCorp's access can be restored to work with the remaining scores only.

This scenario requires as little technical integration as scenario 1, but requires more manual assistance from the client.

Assistance of a local administrator is required for 7 days to set-up access, install and run extraction scripts, managing data transfer and optionally access control to the dedicated server.

Collecting the data can run on any standard hardware PC and requires a Python installation (Windows/Linux architecture) and some storage capacity (we estimate 30MB per mailbox per year). Data collection can take a couple of days of processing (typically 2 years of data takes 2 days of processing).

Dedicated hosting server has to be supplied in accordance with the following requirements: Ubuntu 16.04 LTS, Quad Core, 32GB RAM, 1TB storage of better.

Flowchart:

1. Extraction script
2. First step of anonymization
3. Pseudonymized file
4. Data transfer
5. Pseudonymized file
6. Language detection
7. Second step of anonymization
8. KCI score calculation
9. KCI score is stored in database
10. Database with scores

Legend:
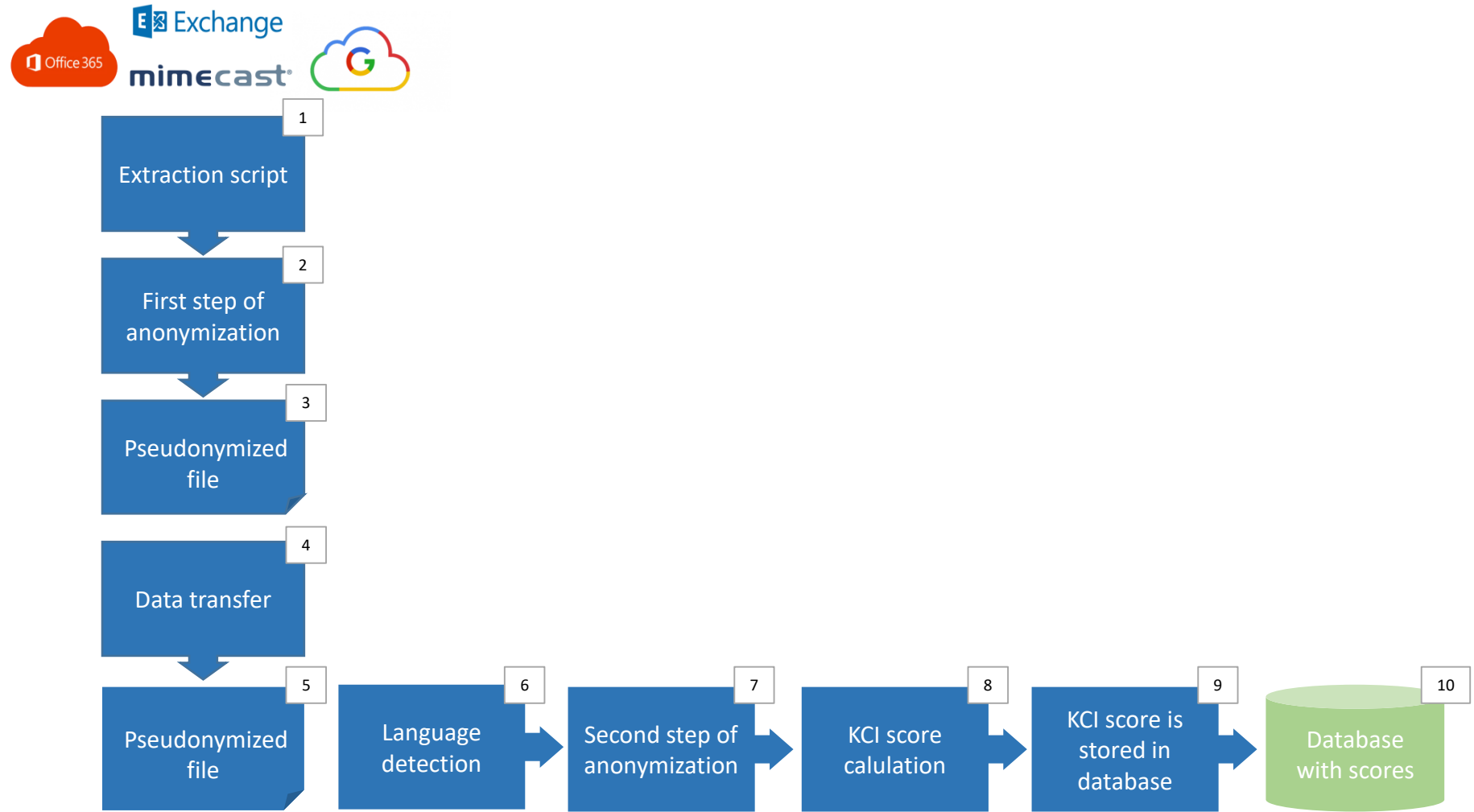- Client side
- KeenCorp side
- Dedicated hosted server

SCENARIO 3: ON-PREMISE INTEGRATION OF A KEENCORP SERVER

In the third scenario an on-premise installation of the complete KeenCorp pipeline is performed at client side.

Every step of the processing is performed at the client location. There are different options to set this up.

From on-site assistance of KeenCorp staff to facilitate the complete process to KeenCorp providing a completely set-up hardware configuration.

This option is usually tailored to specific client needs and requirements.

1 Extraction script

2 First step of anonymization

3 Pseudonymized file

4 Data transfer

5 Pseudonymized file

6 Language detection

7 Second step of anonymization

8 KCI score calulation

9 KCI score is stored in database

10 Database with scores

■ Client side
■ KeenCorp side

SUMMARY OF 3 SCENARIOS

	Scenario 1	Scenario 2	Scenario 3
Hardware requirements	• Standard processing capacity (PC) running a Python script (Windows/Linux architecture), 3-5 GB storage	• Standard processing capacity (PC) running a Python script (Windows/Linux architecture), 3-5 GB storage • Third party hosting solution: Ubuntu 16.04 LTS, Quad Core, 32GB RAM, 1TB storage or better	• Standard processing capacity (PC) running a Python script (Windows/Linux architecture), 3-5 GB storage • Locally hosted solution: Ubuntu 16.04 LTS, Quad Core, 32GB RAM, 1TB storage (managed by client or provided by KeenCorp)
Degree of system integration required	Little	Little to Medium	High
Required client system administrator support	2 days	5-7 days	14 – 20 days
Required KeenCorp workload	2 days	7 days	10 days
Average project lead time	14 days	1 month	2 months
Estimated set-up one-off costs/scenario	nil	€ 5.000 - € 10.000	€ 10.000 - € 50.000
Est. annual maintenance costs year 2 onwards	nil	€ 3.000	€ 5.000 - € 25.000

Message Date: 06-25-2017
Message Sender: Sr. Management, Male

/PERSON, I am a cofounder of /COMPANY I live in /PLACE, actually /PERSON is my next door neighbour.

Our new Idea Manager solution was recently selected by /COMPANY to identify, develop and select new ventures. This was put on hold following the recent management change. Oddly enough, the Knowledge Management group, then housed in /PLACE, also thought we were the "best thing since sliced bread" and wanted to bring us in as an /COMPANY corporate offering. As you know, that group was disbanded.

I was interested in your participation in the /COMPANY. We are in discussions with /PERSON company, /COMPANY. It turned out that we have very similar clients and our software solution can be a natural enabler for /COMPANY approach to innovation (/COMPANY) is our latest corporate innovation client). Also, we can target our solution at venture management (/COMPANY), /COMPANY (/COMPANY) and new service development (/COMPANY).

I believe that we are an ideal solution for /COMPANY to identify and develop new business ventures, to enhance current products and services, and to share ideas and learnings across the company. I would like to meet with you to discuss our solution

I look forward to hearing from you.

Regards,

/PERSON