NAKIT – Google Cloud Platform integration & dashboards

Národní agentura pro komunikační a informační technologie,

s. p. Revolt BI s.r.o.

## Effective date 2024-07-15



**Table of Contents** Executive summary Requirements & solution

Google Cloud deployment location Activities & deliverables

Timeline

Roles and Responsibilities

Risks, Assumptions, Issues, and Dependencies Out of Scope

Success Criteria Pricing Signature Block

Appendix: Pricing Breakdown

# Executive summary

This Statement of Work (SoW) covers data integration from various data sources, developing a data warehouse on the Google Cloud Platform (GCP), and visualization of KPIs using Looker Studio for Národní agentura pro komunikační a informační technologie (NAKIT) (<https://nakit.cz/en/>).

Národní agentura pro komunikační a informační technologie (NAKIT) is helping the Czech Republic achieve a successful digital transformation of public administration. That makes NAKIT a strategic partner of the state providing communication and information services for public administration and eGovernment. NAKIT's mission is to provide useful and secure services.

Revolt BI is a data consulting and development agency leveraging data to achieve business results. We manage the data through its whole lifecycle – from data collection process design to data storage, reporting, and statistical modeling or AI.

By implementing data warehouse and visualization reports, we aim to enhance the efficiency and transparency of state digitalization efforts. The scope of the project includes the deployment of key performance indicator (KPI) dashboards across various government portals for the Czech Republic (CZ). To do so, we have to integrate various data sources.

The success criteria is, that all identified data sources from various state departments must be fully integrated into the Google Cloud Platform, ensuring seamless data flow and storage. Secondly, a data model must be developed and implemented to organize Key Performance Indicator (KPI) data effectively, facilitating straightforward access and in-depth analysis. Thirdly, we aim to develop interactive dashboards and reports in Looker Studio that not only accurately represent the KPI data but are also user-friendly and visually engaging. Finally, the accuracy and reliability of the KPI data stored in the data warehouse will be validated against historical data to ensure precision and correctness, cementing the project's integrity and success.

The work on this project is expected to begin on 2024-07-15.

# Requirements & solution

The business goal is to publicly publish reports/visualizations on the prepared website of NAKIT or DIA (Digitální a informační agentura) to present metrics related to the efficiency of state digitalization. We will achieve this by integrating various data sources to create a data warehouse in GCP, from which we will create visualizations of KPIs using Looker Studio reports. We also use around 10 different data source types from various departments/offices as a data source.

We identified two use cases: data integration & data warehouse, and data visualization in Looker Studio. The project's technical architecture is designed to support these business goals through a dual-use case approach. We will integrate data from approximately ten distinct types from various government departments and offices into Google Cloud Platform (GCP). This integration involves consolidating diverse datasets into a unified data warehouse, facilitating centralized data management and accessibility. Our approach includes setting up secure and efficient data pipelines that ensure timely and accurate data transfers.

The data warehouse will not only serve as the storage repository but also as the analytical base where data is cleaned, processed, and structured for reporting purposes. This foundational step is critical as it dictates the quality and reliability of the outputs in subsequent stages.

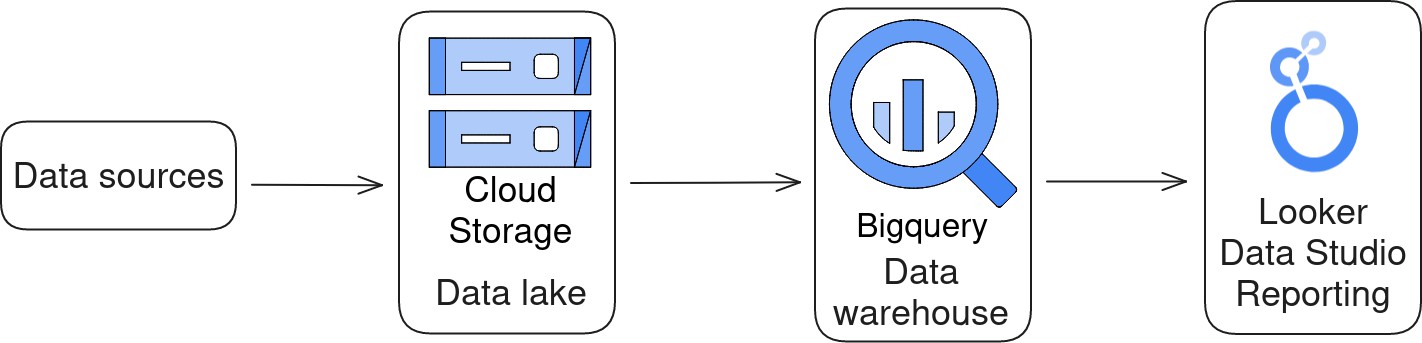
Upon establishing a data warehouse, we will utilize Looker Studio to develop interactive, intuitive reports and dashboards that visually represent the KPIs. These visualizations are designed to be user-friendly, providing clear insights into the state's digitalization efforts and their effectiveness.

Each dashboard will be tailored to showcase specific metrics that reflect the efficiencies gained through digitalization initiatives, supporting decision-makers and public stakeholders in understanding the impact of these measures.

## High-level steps to go live:

* Data source ingestion
* Metrics (KPIs) definition and documentation
* Data cleaning and preparation

## End-state architecture:



[End-state architecture diagram (.png file)](https://drive.google.com/file/d/1s34xy7EgyyVLxKjJAhudSE2igSPKHOIo/view?usp=sharing) [End-state architecture diagram (.svg file)](https://drive.google.com/file/d/1jop_PMLKjlZzMGLbF-Rhqtv3eTf1pb6x/view?usp=sharing)

[End-state architecture diagram (.excalidraw file)](https://drive.google.com/file/d/1Jjnh6Ldr2VtPlIcpNqszlMcbssMhUOP8/view?usp=sharing)

**Functional requirement**

1 Data increment frequency: Ideally, data aggregation is expected to occur daily with minimal latency.

Data volume: The estimated data volume from metrics is up to 500 GB

2 per month, while the pilot dashboard will require up to 3 TB of data per month.

3 The data warehouse consolidates KPI data from multiple state departments, facilitating analysis and reporting.

4 Looker Studio manages dynamic reports visualizing KPI data for enhanced decision-making.

**Google Cloud's components**

|  |  |  |
| --- | --- | --- |
|  | | **Used components** |
| 1 | Cloud Storage |  |
| 2 | BigQuery |  |
| 3 | App Engine |  |
| 4 | Cloud Logging |  |

# Google Cloud deployment location

The solution infrastructure will be deployed in the customer’s Google Cloud Platform tenant assigned to Revolt BI's billing account and the customer will be billed monthly for the consumption. NAKIT receives

$4,500 worth of credits from Google which will be utilized for the GCP project usage related to this implementation and running the project for the specified time range.

To optimize costs, we assume the utilization of on-demand service rather than the Standard edition in BigQuery. Also, we assume to use location type: region and location: Frankfurt in most of the services.

# Activities & deliverables

The deliverable for each milestone is a documented and deployed component accepted by the customer on the status meeting or via email.

At the client's (NAKIT) request, it was agreed to work in a consume-as-you-go mode, which means that there are no fixed deliverables for this project (for example, the number of reports produced). Work on the project will be subject to available budget. The deliveries below are therefore indicative only.

There is no timeline specified, Revolt BI will provide requested assistance (support) until the full agreed number of MDs defined in this SoW (23 MDs) is consumed.

The workload running on Google Cloud is mostly Cloud Storage and BigQuery.

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone** | **Description** | **Deliverable** | **Effort**  **(man-days)** |
| Connection of the data sources and GCP set-up | Integrate state department KPI data sources with Google Cloud Platform. | Integrated data environment on GCP. | 10 |
| Data model set-up &  implementation | Develop and implement a data model to organize and manage the KPI data effectively in the data warehouse. | An operational data model in the data warehouse. | 4 |
| Data visualization | Create dynamic reports and dashboards in Looker Studio to visualize the KPI data. | Interactive dashboards and reports in Looker Studio. | 3 |
| Data validation | Ensure the accuracy and integrity of the KPI data through validation processes. | Validation report confirming data accuracy. | 2 |
| Project management, Documentation, and Consulting | Manage project activities, document processes, and configurations, and provide expert consulting. | Comprehensive project documentation. | 2 |
| Handover | Final review and handover of the complete system and documentation to the client. | Final project deliverables, including access credentials. The recommendations are in line with the expected outcomes. | 2 |
| **Total** |  |  | **23** |

**Total estimated effort:** 23 man-days (184 man-hours)

# Roles and Responsibilities

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Role** | **Role description** | **Organization** |
| xxx | Project Manager | Manages the project to meet the budget and timeline, ensuring all project phases are coordinated. | Revolt BI |
| xxx | Solution Architect | Outlines the solution to fit technical requirements and implements the proposed solution. | Revolt BI |
| xxx | Associate Partner | Enables customer success and ensures the project aligns with the success criteria. | Revolt BI |
| xxx | Data Engineer | Responsible for the data integration, architecture/design, construction, and maintenance of GCP and its components. | Revolt BI |
| xxx | Data Engineer | Responsible for the data integration, architecture/design, construction, and maintenance of GCP and its components. | Revolt BI |
| xxx | Data Analyst | Responsible for developing the data model and creating data visualizations in Looker Studio. | Revolt BI |
| xxx | Project Owner | Oversees the project objectives, ensuring alignment with agency goals and stakeholder expectations. | NAKIT |
| xxx | Technical Consultant | Provides expert guidance on the technical aspects of the project, ensuring the solutions are effectively implemented. | NAKIT |
| xxx | Tech Exec | Oversees the technological strategy and execution, ensuring the technical deliverables meet the organization's standards and strategic objectives.  Account to be opened under [xxx](mailto:sysadmins@nakit.cz) | NAKIT |

The Revolt BI services delivery team is in Prague, Czechia. The delivery of this service to NAKIT will be through remote delivery.

# Risks, Assumptions, Issues, and Dependencies

## Known issues:

* Dependence on many data providers (departments/institutions) and sources, which will be completely handled on the NAKIT side (see Assumptions), however, can easily lead to project delays.
* Unknown data source types (worst case: email attachment).
* Necessary adaptation to the data owners (both push and pull transmission methods will be used).

## Assumptions:

* Data availability: Structured and clean data with clear documentation will be available from various departments and institutions, only in electronic format and in a ready state to upload into a database.
* Internal team: NAKIT has its own in-house team that will be able to handle part of the project itself. Our team will be used mainly for solving more complex tasks in which the in-house team does not have the necessary expertise.
* Data sensitivity: Sensitive personal data will not be processed, minimizing requirements for personal data protection.
* Types of data sources: Data will originate from different systems and databases. Data will be delivered with various methods of delivery. Data will not be drawn directly from the primary databases. The data will be exported or sent to one of the repositories (e.g. blob storage).
* Data availability: Structured and clean data will be available from various departments and institutions, only in electronic format and in a ready state to upload into a database.
* Data owner's cooperation: NAKIT will ensure that inter-departmental/inter-institutional cooperation is ensured.

## Dependencies:

* Data public presentation: Data must be publicly presented on the prepared NAKIT or DIA website. However, it will not be up to Revolt BI to manage these pages or implement the visualizations in the web interface. Revolt BI will only prepare reports in Looker Studio.
* Project specifications and details: The project on the NAKIT's side is managed in an agile manner. Some specifications should be clarified soon (estimated 14 days) and should be communicated to us promptly.
* Project success criteria: Project success will depend on collaboration and timely data delivery from relevant institutions and departments, which will be coordinated by the representative of NAKIT.
* Data ETL process: For fast data retrieval, NAKIT wants to adapt to the data owners (both transmission methods will be used – push and pull). The involvement of Keboola is assumed by the request of NAKIT.

# Out of Scope

* Implementations on the side of the website (see Dependencies).
* Models reset in case the business use case (requirements) changes.
* Integration of other data sources.
* Solution support after delivery (is subject to potential follow-up phase).
* Advanced analytics like planning or any data science cases.
* Spend alerts are under the client’s competency.
* GDPR and other regulatory requirements compliance for user data collection.

# Success Criteria

* Data integration and data warehouse setup completion: All identified data sources from various state departments are successfully integrated into GCP.
* Data model implementation: A data model is developed and implemented that effectively organizes KPI data for easy access and analysis.
* Visualization and reporting: Interactive dashboards and reports are developed in Looker Studio that accurately represent the KPI data.
* Data accuracy and validation: The KPI data stored in the data warehouse is accurate and validated against historical data.

# Pricing

|  |  |
| --- | --- |
| **Services** | **Cost** |
| Source data connection | $6,500 |
| Data preparation & data model | $2,600 |
| Visualization | $2,000 |
| Data validation | $1,300 |
| Handover | $1,300 |
| Project Management + Documentation + Consulting | $1,300 |
| **Total cost including VAT** | $15,000\* |

\* Revolt BI reserves right to a publicly publishable reference from the client (NAKIT) in the agreed format within 30 days of project completion. This reference will include relevant information about the project and its outcomes, which can be shared with the public to promote our services.

|  |  |
| --- | --- |
| **Services** | **Cost** |
| Google PSF Investment | $15,000 |
| Revolt BI Investment | $0 |
| The remaining balance belonging to NAKIT | $0 |
| **Total cost** | $15,000 |

# Signature Block

xxx

Národní agentura pro komunikační a informační technologie, s. p. (NAKIT)

Xxx

Revolt BI s.r.o.

Signature Signature

xxx

Národní agentura pro komunikační a informační technologie, s. p. (NAKIT)

Signature

# Appendix: Pricing Breakdown

Revolt BI will charge you a fixed blended rate for all technical and project roles. All prices in this document are including VAT.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Role** | **Number of Resources** | **MDs** | **Rate** | **Cost** |
| Project Manager | 1 | 2,5 | 650/MD | $1,625 |
| Solution Architect | 1 | 2 | 650/MD | $1,300 |
| Associate Partner | 1 | 0 | 650/MD | $375 |
| Data Engineer | 2 | 12 | 650/MD | $7,800 |
| Data Analyst | 1 | 6 | 650/MD | $3,900 |
| **Total** | **6** | **23** | **650/MD** | **$15,000** |