INTERNSHIP PRESENTATION

MEDESSI Seithondji Dieudonné

Capgemini

Developer of new data technologies

March 4th – September 3rd 2024

Supervisor: **Mr Youness BOUAROUROU**

AGENDA

01 Know Capgemini

02 Experience overview

03 Google Cortex Framework

04 SAP Datasphere

05 Generative AI project

06 Conclusion

Know Capgemini

CAPGEMINI GROUP IN BRIEF

Founded in
1967

22.5 billion euros
in revenue in 2023

More than
340 000
employees

More than
50
country
With more than
120 nationalities

Capgemini's Brands

Capgemini
Major global company providing comprehensive digital transformation services

Capgemini invent
Focused on strategy and digital innovation

Capgemini engineering
Specializes in advanced engineering and R&D

sogeti
Part of Capgemini
Provides IT services with a personnalized approch, often catering to SMEs and local or regional projects

CAPGEMINI STRASBOURG

More than
150
employees

Entity
PBS Package-Based Solutions
Specialized in ERP deployment and integration

SAP BTP Team
10
employees

ERP

Enterprise
Ressource
Planning

ERP for instances

SAP S/4 HANA

Microsoft Dynamics

ORACLE NETSUITE

SAP BUSINESS TECHNOLOGY PLATFORM

Secure and compliant
Open and flexible
Unified and simple

Broad ecosystem
Enterprise grade and scalable
Business centric

App Dev	Automation	Integration	Data and Analytics	AI
→ Visual low-code/ no-code experience	→ Workflow management	→ Process integration	→ Analytics and planning	→ Pretrained AI models
→ Pro-code tooling	→ Robotic process automation	→ API-led integration	→ Data management	→ Generative AI and AI workload management
→ Digital experience	→ Process monitoring and analytics	→ Event-driven integration	→ Operational database	→ Business data and context
→ DevOps	→ Automated document processing	→ Hybrid integration	→ Data warehouse and data lake	→ Responsible AI
		→ B2B integration		
		→ Data integration		

Experience overview

EXPERIENCE OVERVIEW

Onboarding phase

Designed to help new arrivals to integrate

Learn about Capgemini in details

Construct a network

SAP Business Technology Platform

SAP BTP Discovery

SAP Datasphere exploration

Cloud technologies

Cloud certification

Study of Cortex Framework architecture with hands-on experience

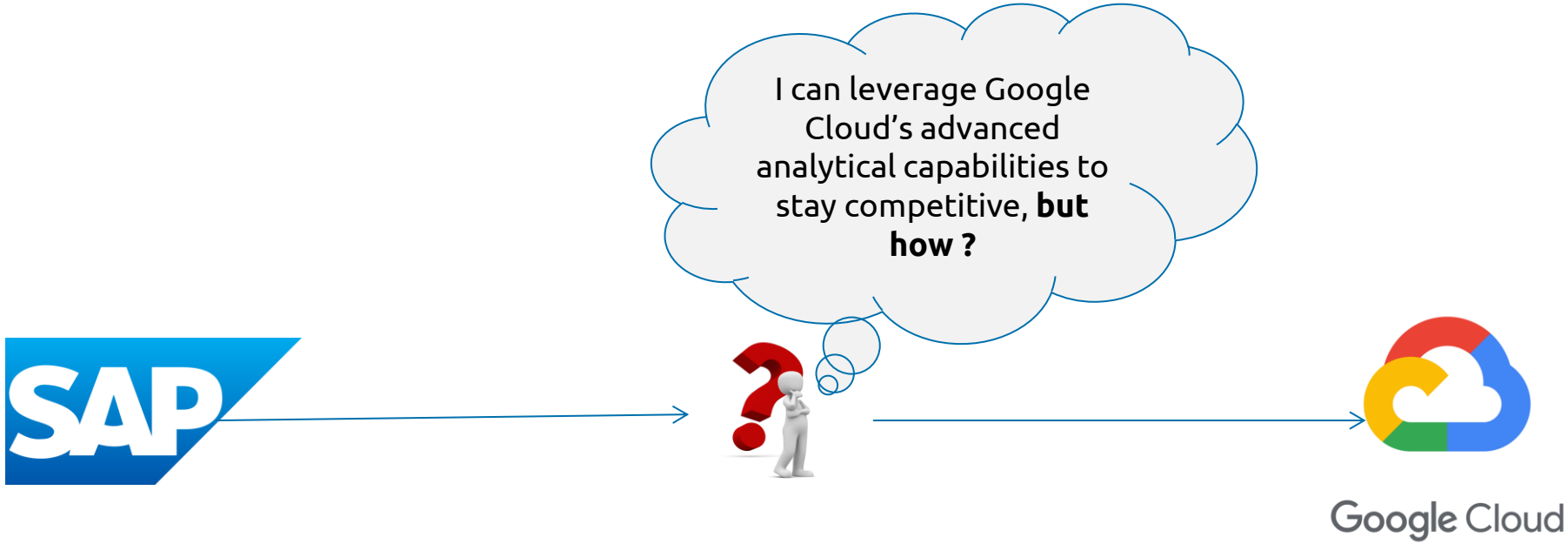
Generative AI project

Upskilling in GenAI technologies

Work on a project

Google Cortex Framework

WHAT IS GOOGLE CORTEX FRAMEWORK ?

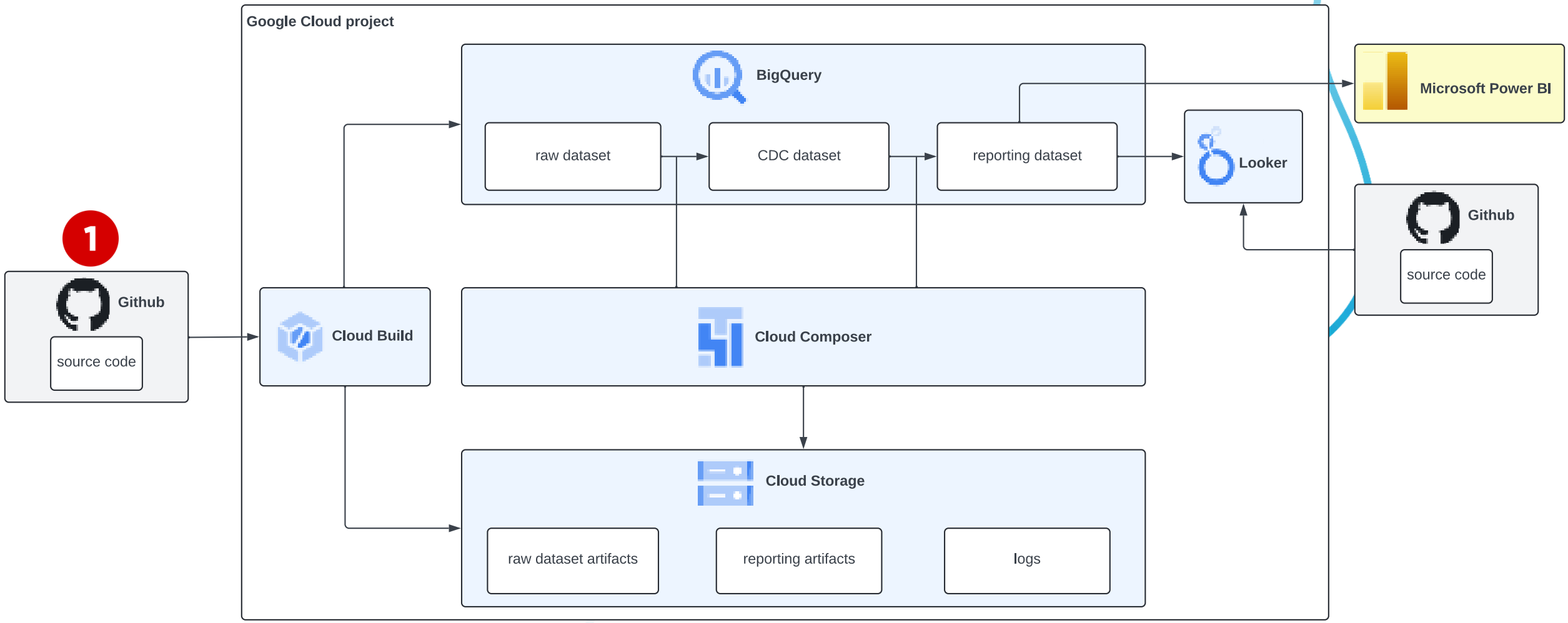


WHAT IS GOOGLE CORTEX FRAMEWORK ?

Google Cortex Framework

Quickly
Minimal Risk
Reduced Complexity
Lower Cost

SAP WORKLOAD DEPLOYMENT ARCHITECTURE



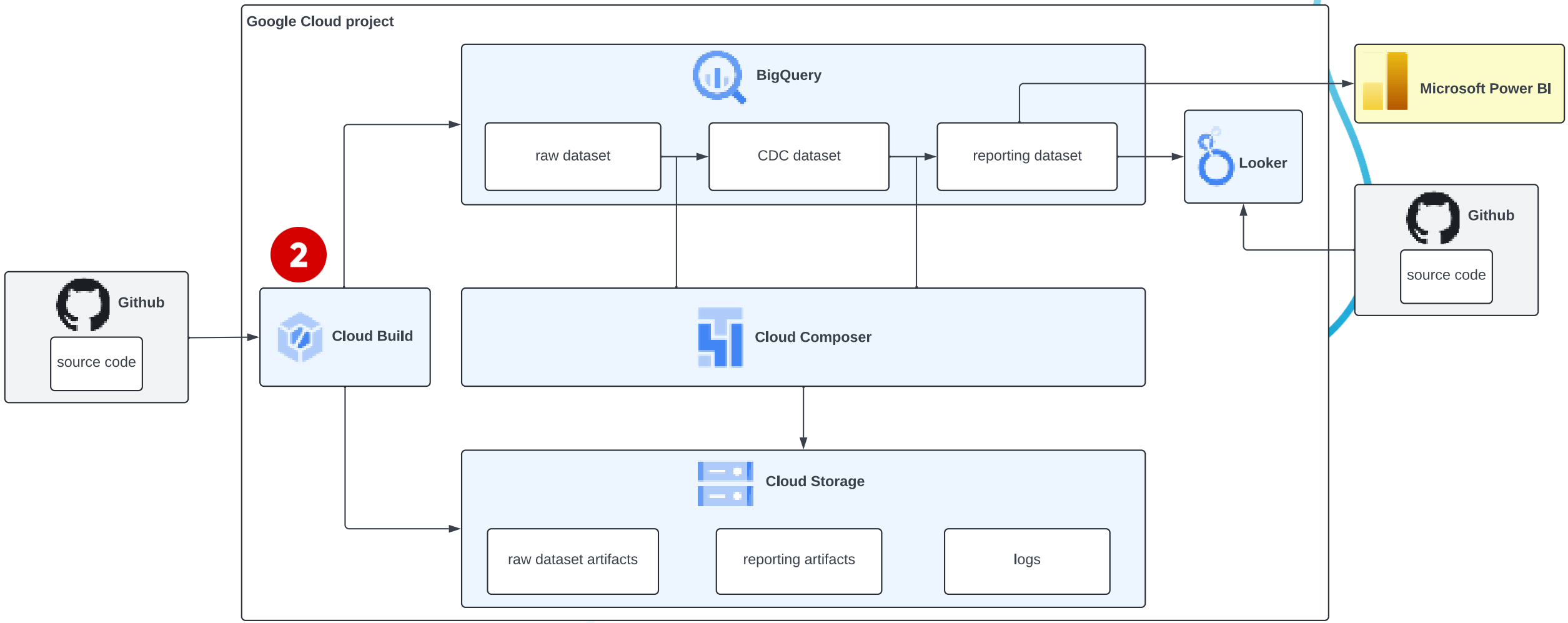
PRECONFIGURATION 1

```
{
    "testData": true,
    "deploySAP": true,
    "deploySFDC": false,
    "deployMarketing": false,
    "turboMode": true,
    "projectIdSource": "cortex-demo-423814",
    "projectIdTarget": "cortex-demo-423814",
    "targetBucket": "cortex-demo-dags-bucket",
    "location": "EU",
    "languages": [ "E" ],
    "currencies": [ "USD" ],
    "testDataProject": "kittycorn-public",
    "k9": {
        "datasets": {
            "processing": "K9_PROCESSING",
            "reporting": "K9_REPORTING"
        }
    },
    "SAP": {
        "deployCDC": true,
        "datasets": {
            "cdc": "SAP_CDC_PROCESSED",
            "raw": "SAP_RAW_LANDING",
            "ml": "SAP_ML_MODELS",
            "reporting": "SAP_REPORTING"
        },
        "SQLFlavor": "ecc",
        "mandt": "100"
    },
    "shareWithCredly": false,
    "userInfo": {
        "email": "",
        "firstName": "",
        "lastName": ""
    }
}
```

```
data_to_replicate:
{% if sql_flavour.upper() == 'ECC' %}
  - base_table: anla
    load_frequency: "@hourly"
  - base_table: ankt
    load_frequency: "@hourly"
  - base_table: faglflexa
    load_frequency: "@hourly"
  - base_table: konv
    load_frequency: "@daily"
    #
    # OTHER TABLES AND REFRESH FREQUENCIES
    #
{% endif %}
  - base_table: adr6
    load_frequency: "@daily"
  - base_table: adrc
    load_frequency: "@daily"
  - base_table: adrct
    load_frequency: "@daily"
  - base_table: afko
    load_frequency: "@hourly"
  - base_table: afpo
    load_frequency: "@hourly"
  - base_table: aufk
    load_frequency: "@hourly"
  - base_table: vbep
    load_frequency: "@hourly"
  - base_table: vbfa
    load_frequency: "@hourly"
  - base_table: vbpa
    load_frequency: "@hourly"
  - base_table: vbrk
    load_frequency: "@hourly"
  - base_table: vbrp
    load_frequency: "@hourly"
  - base_table: mch1
    load_frequency: "@daily"
  - base_table: mska
    load_frequency: "@weekly"
  - base_table: mslb
    load_frequency: "@weekly"
  - base_table: msku
    load_frequency: "@weekly"
  - base_table: mkol
    load_frequency: "@weekly"
    #
    # OTHER TABLES AND REFRESH FREQUENCIES
    #
```

```
bq_independent_objects:

  - sql_file: NetDueDateCalc.sql
    type: script

  - sql_file: CurrencyConversion.sql
    description: "Exchange Rate for Currency Conversion"
    type: view

  - sql_file: SalesOrders.sql
    description: "Sales Orders Header and Items"
    type: table
    table_setting:
      load_frequency: "*/5 * * * *"

  - sql_file: CustomersMD.sql
    description: "Customer Master Data"
    type: table
    table_setting:
      load_frequency: "@daily"

  - sql_file: SalesOrderStatus.sql
    description: "Sales Order Status"
    type: table
    table_setting:
      load_frequency: "*/10 * * * *"
  #
  # OTHER OBJECTS
  #

bq_dependent_objects:
  - sql_file: BusinessPartnersMD.sql
    description: "Business Partners and Addresses Master Data"
    type: table
    table_setting:
      load_frequency: "@daily"

  - sql_file: Deliveries.sql
    description: "Deliveries Master Data"
    type: table
    table_setting:
      load_frequency: "@daily"

  - sql_file: SalesOrders_V2.sql
    description: "Sales Orders Header and Items"
    type: table
    table_setting:
      load_frequency: "*/5 * * * *"

  - sql_file: Billing.sql
    description: "Billing Master Data"
    type: table
    table_setting:
      load_frequency: "@hourly"
  #
  # OTHER OBJECTS
  #
```

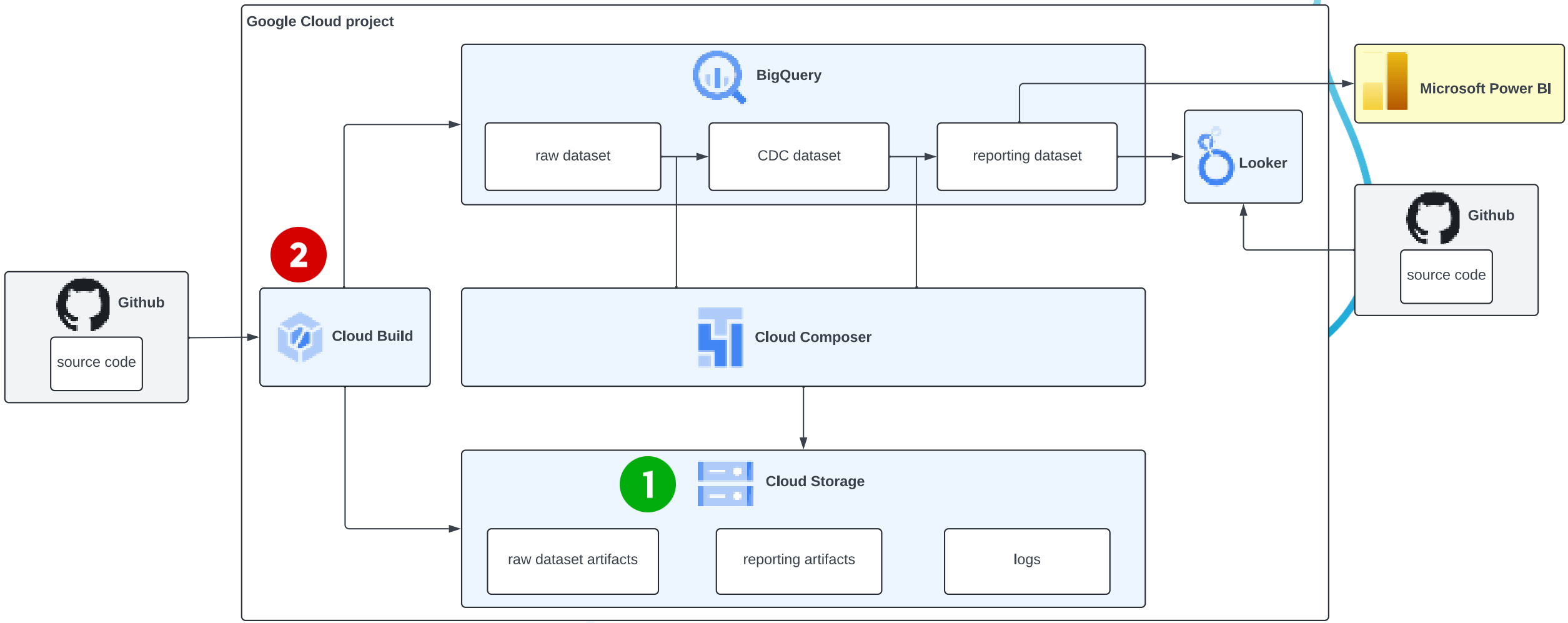
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



DEPLOYMENT ACCELERATORS 2

```yaml
steps:
  - name: gcr.io/cloud-builders/git
    entrypoint: "bash"
    id: 'init_git_env'
    waitFor: ['-']
    args:
      - "-c"
      - |-
        echo "Cleaning /workspace"
        rm -rf /workspace/*
        echo "Cloning the repository from Github"
        git clone --recurse-submodules "${_GIT_URL}" "${_GIT_PATH}"

  - name: gcr.io/kittycorn-public/deploy-kittycorn:v2.0
    entrypoint: "bash"
    id: 'init_deploy_config'
    waitFor: ['init_git_env']
    args:
      - "-c"
      - |-
        set -e
        echo "Initial configuration ${_CONFIG_FILE}:"
        cat ${_CONFIG_FILE}

        python3 src/common/init_deployment_config.py --config-file "${_CONFIG_FILE}" \
              --sub-validator "src/SAP/SAP_REPORTING" \
              --sub-validator "src/SFDC/src" \
              --sub-validator "src/marketing/src"

        echo "Processed configuration:"
        cat ${_CONFIG_FILE}
        echo -e "\n--------------------------------"
    dir: ${_GIT_PATH}

    #
    # OTHER STEPS HERE
    #

logsBucket: "gs://$_GCS_BUCKET"
timeout: 32400s
substitutions:
  _CONFIG_FILE: "config/config.json"
  _DEPLOY_SAP_ML_MODELS: "false"
  _NO_TEST_DATA: "false"
  _GIT_URL: "https://github.com/DonnescoPablo/cortex-data-foundation.git"
  _GIT_PATH: "cortex-data-foundation"
  _AIRFLOW_BUCKET: "southamerica-east1-cortex-d-3a0e66a5-bucket"
options:
  substitution_option: "ALLOW_LOOSE"
tags: ["cortex"]
```


SAP WORKLOAD DEPLOYMENT ARCHITECTURE

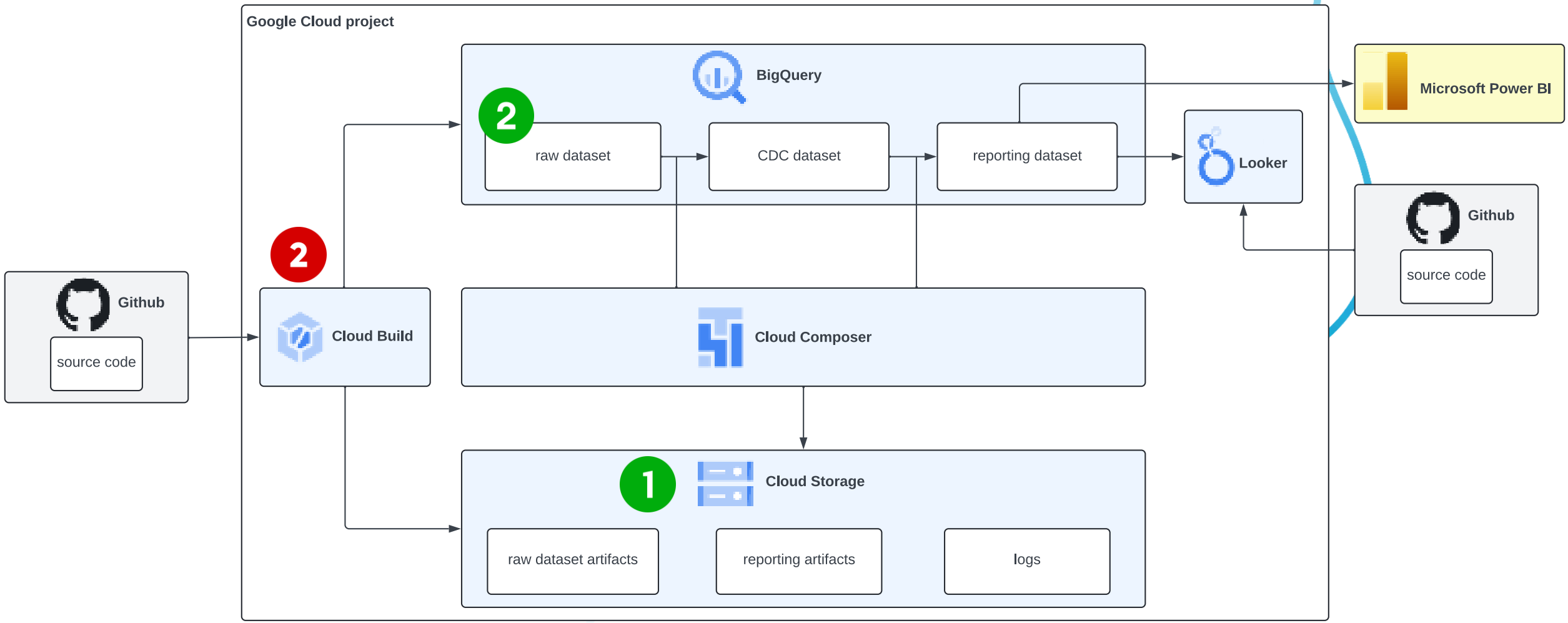


ARTEFACTS 1

```
MERGE `${target_table}` AS T
USING (
  WITH
    S0 AS (
      SELECT * FROM `${base_table}`
      WHERE recordstamp >= (
        SELECT IFNULL(MAX(recordstamp), TIMESTAMP('1940-12-25 05:30:00+00'))
        FROM `${target_table}`)
    ),
    -- To handle occasional dups from SLT connector
    S1 AS (
      SELECT * EXCEPT(row_num)
      FROM (
        SELECT *, ROW_NUMBER() OVER (PARTITION BY ${keys}, recordstamp ORDER BY recordstamp) AS row_num
        FROM S0
      )
      WHERE row_num = 1
    ),
    T1 AS (
      SELECT ${keys}, MAX(recordstamp) AS recordstamp
      FROM `${base_table}`
      WHERE recordstamp >= (
        SELECT IFNULL(MAX(recordstamp), TIMESTAMP('1940-12-25 05:30:00+00'))
        FROM `${target_table}`)
      GROUP BY ${keys}
    )
  SELECT S1.*
  FROM S1
  INNER JOIN T1
    ON ${p_key_sub_query}
      AND S1.recordstamp = T1.recordstamp
  ) AS S
ON ${p_key}
-- ## CORTEX-CUSTOMER You can use "`is_deleted` = true" condition along with "operation_flag = 'D'",
-- if that is applicable to your CDC set up.
WHEN NOT MATCHED AND IFNULL(S.operation_flag, 'I') != 'D' THEN
  INSERT (${fields})
  VALUES (${fields})
WHEN MATCHED AND S.operation_flag = 'D' THEN
  DELETE
WHEN MATCHED AND S.operation_flag IN ('I','U') THEN
  UPDATE SET ${update_fields};
```

```
from __future__ import print_function
from airflow.operators.dummy_operator import DummyOperator

from datetime import timedelta, datetime
import airflow
from airflow.contrib.operators.bigquery_operator import BigQueryOperator
from airflow.version import version as AIRFLOW_VERSION


default_dag_args = {
   'depends_on_past': False,
   'start_date': datetime(${year}, ${month}, ${day}),
   'catchup': False,
   'retries': 1,
   'retry_delay': timedelta(minutes=30),
}

with airflow.DAG("CDC_BigQuery_${base_table}",
                 template_searchpath=['/home/airflow/gcs/data/bq_data_replication/'],
                 default_args=default_dag_args,
                 catchup=False,
                 max_active_runs=1,
                 schedule_interval="${load_frequency}") as dag:
    start_task = DummyOperator(task_id="start")
    if AIRFLOW_VERSION.startswith("1."):
        copy_records = BigQueryOperator(
            task_id='merge_query_records',
            sql="${query_file}",
            create_disposition='CREATE_IF_NEEDED',
            bigquery_conn_id="sap_cdc_bq",
            use_legacy_sql=False)
    else:
        copy_records = BigQueryOperator(
            task_id='merge_query_records',
            sql="${query_file}",
            create_disposition='CREATE_IF_NEEDED',
            gcp_conn_id="sap_cdc_bq",
            use_legacy_sql=False)
    stop_task = DummyOperator(task_id="stop")
    start_task >> copy_records >> stop_task
```

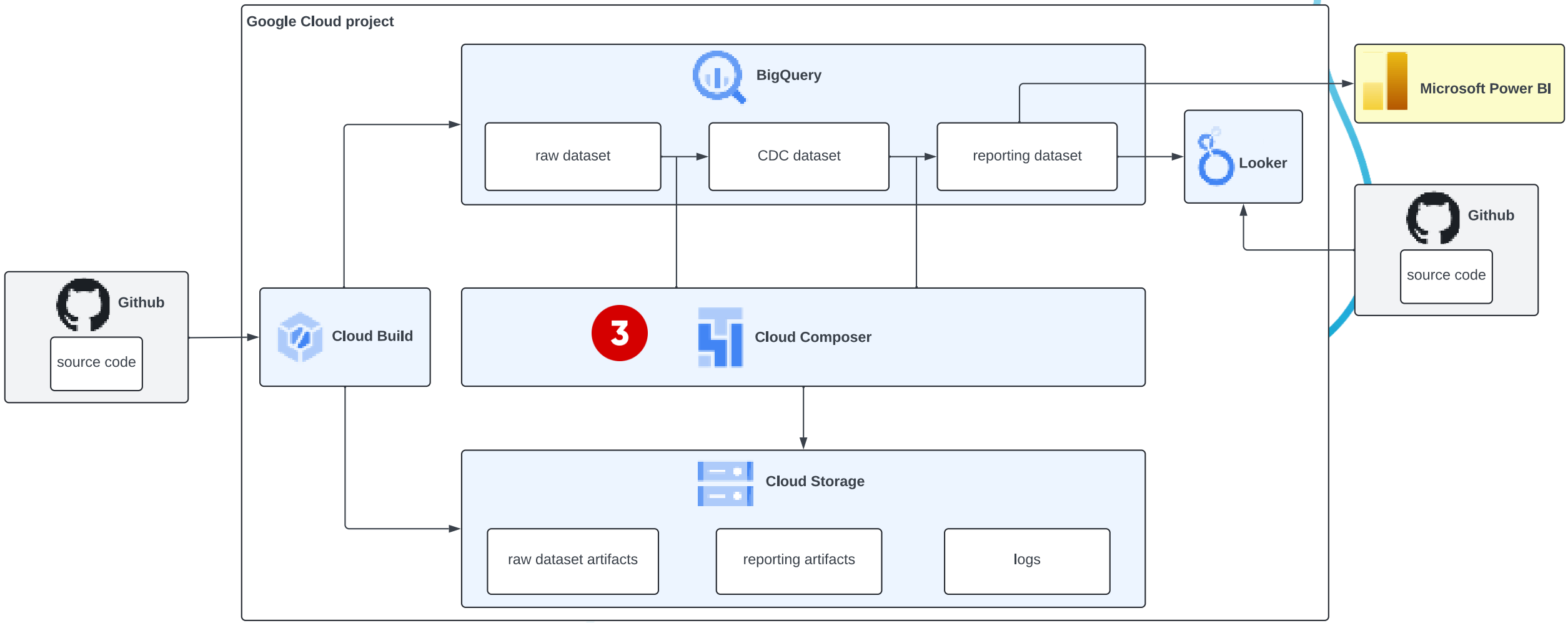
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



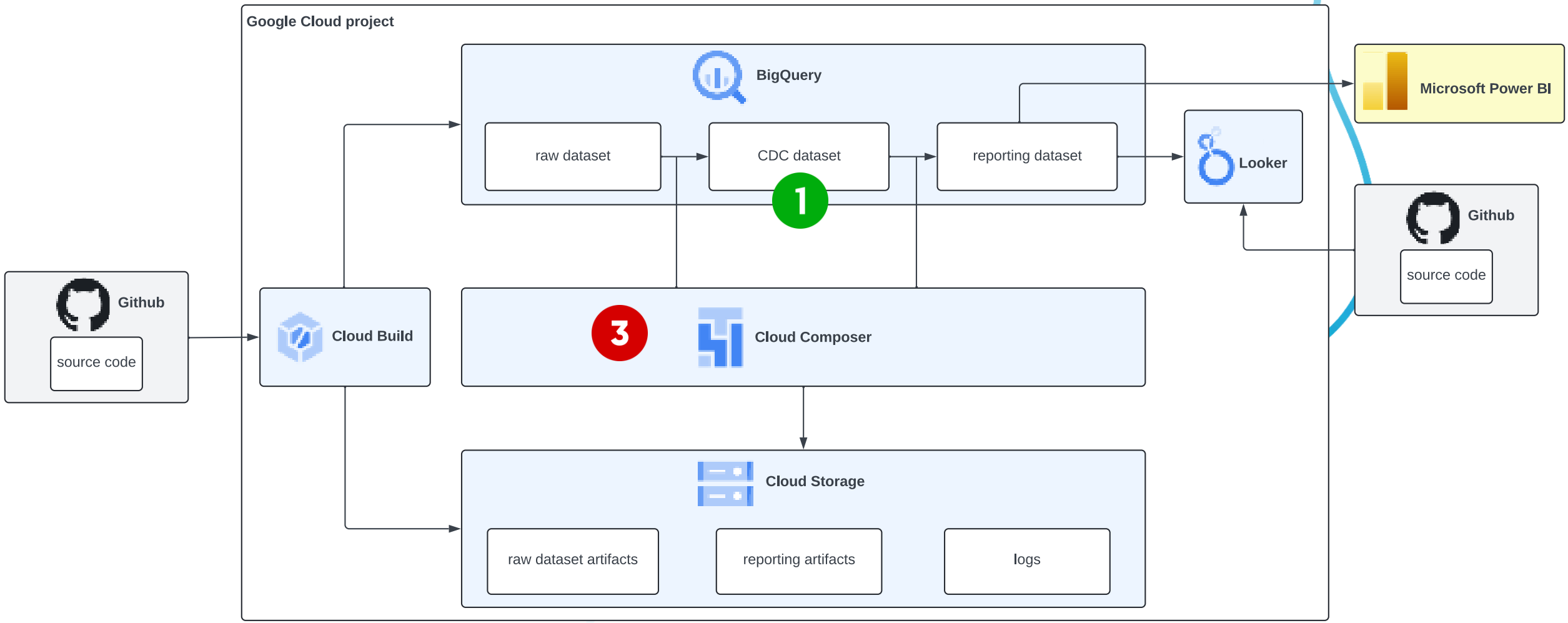
OVERVIEW OF A TABLE IN RAW DATASET 2

Row	partner	type	bu_sort1	title	chusr	chdat	chtim	operation_flag	recordstamp
1	LUCIA	1	UPDATE1	0001	UNICORN	2021-07-01	23:36:38	U	2021-07-01 21:36:39.618124 UTC
2	LUCIA	1	UPDATE2	0001	UNICORN	2021-07-01	23:39:57	U	2021-07-01 21:39:59.356904 UTC
3	LUCIA	1	LATEST	0001	UNICORN	2021-07-01	23:40:21	U	2021-07-01 21:40:22.381178 UTC
4	LUCIA	1	NEW RECORD	*null*	*null*	*null*	*null*	I	2021-06-23 21:26:44.511611 UTC
5	RIZ	1	NEW RECORD	*null*	*null*	*null*	*null*	I	2021-07-01 21:51:26.950546 UTC
6	RIZ	1	UPDATE 1	*null*	UNICORN	2021-07-01	23:51:39	U	2021-07-01 21:51:40.132487 UTC
7	RIZ	1	UPDATE 2	*null*	UNICORN	2021-07-01	23:51:48	U	2021-07-01 21:51:49.835927 UTC
8	RIZ	1	LAST UPDATE	*null*	UNICORN	2021-07-01	23:52:08	U	2021-07-01 21:52:08.929096 UTC

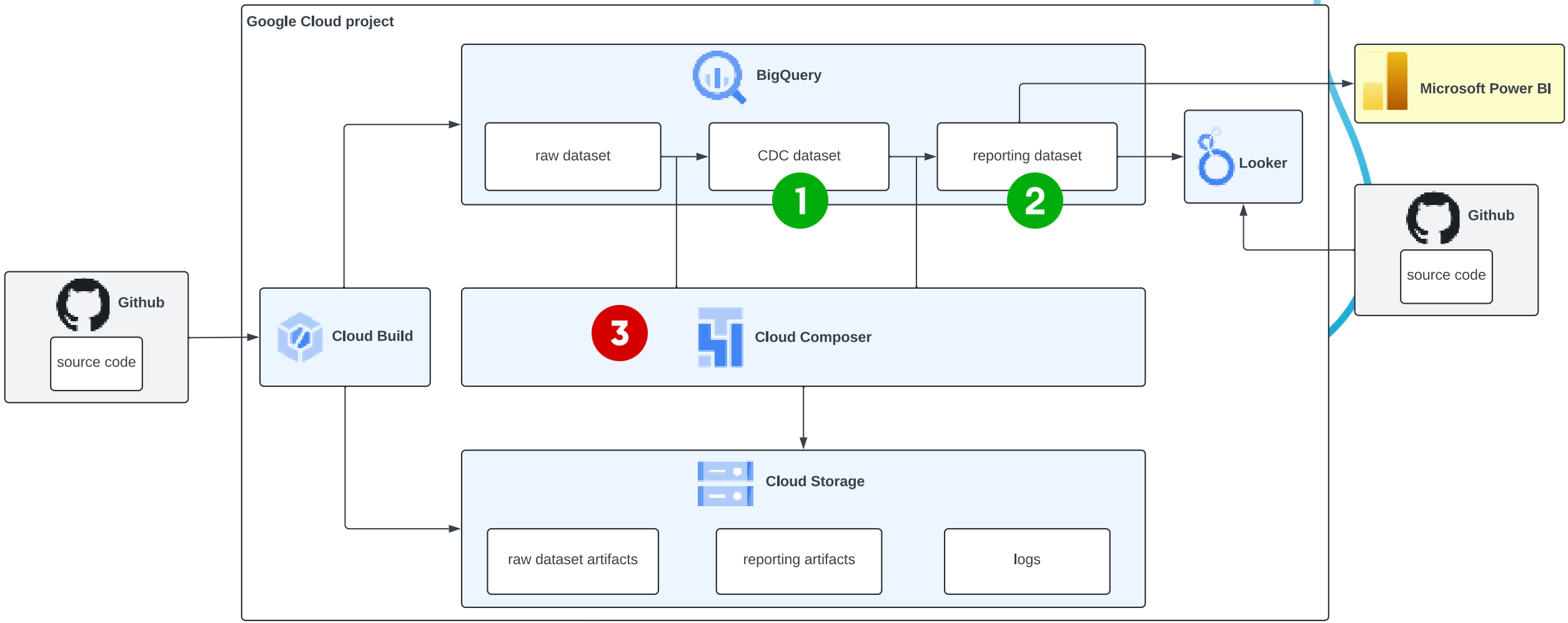
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



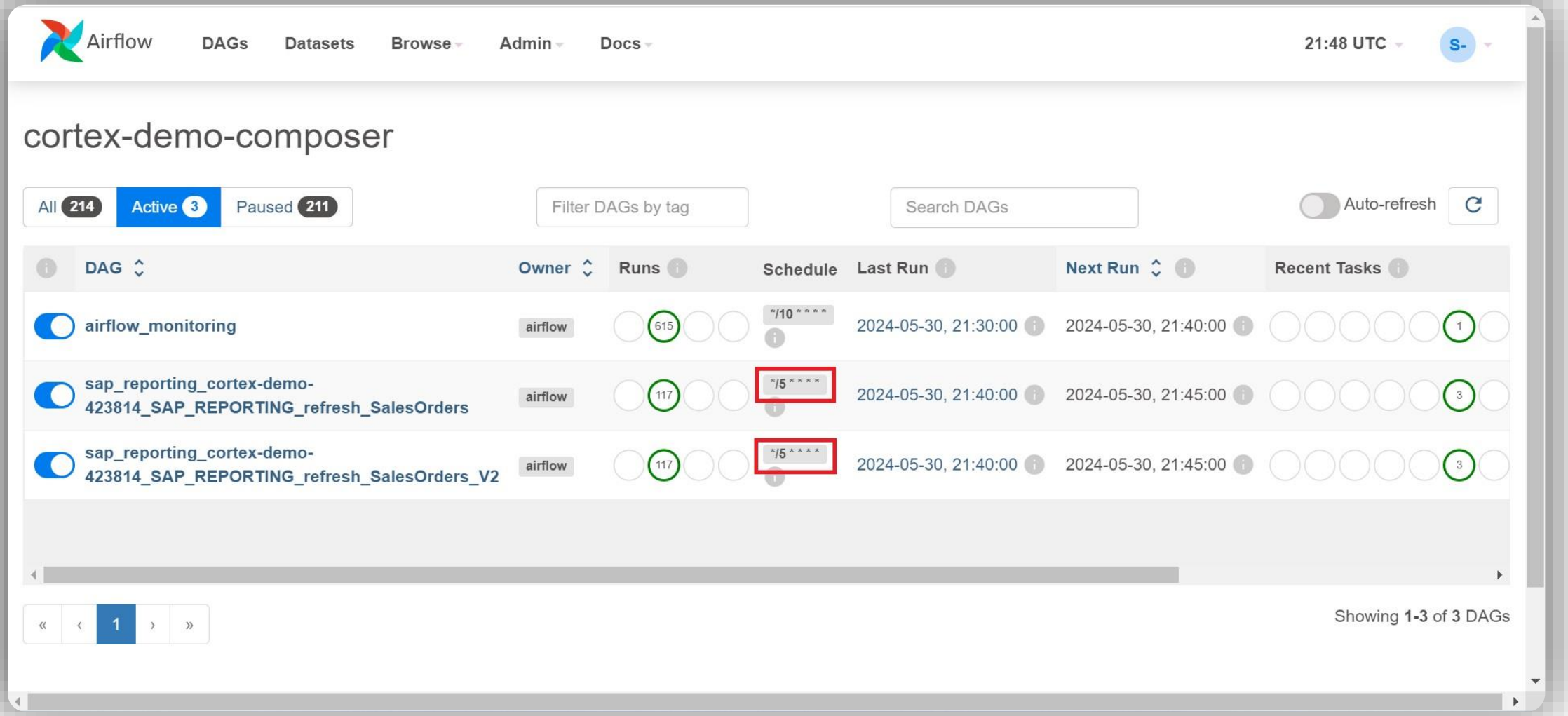
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



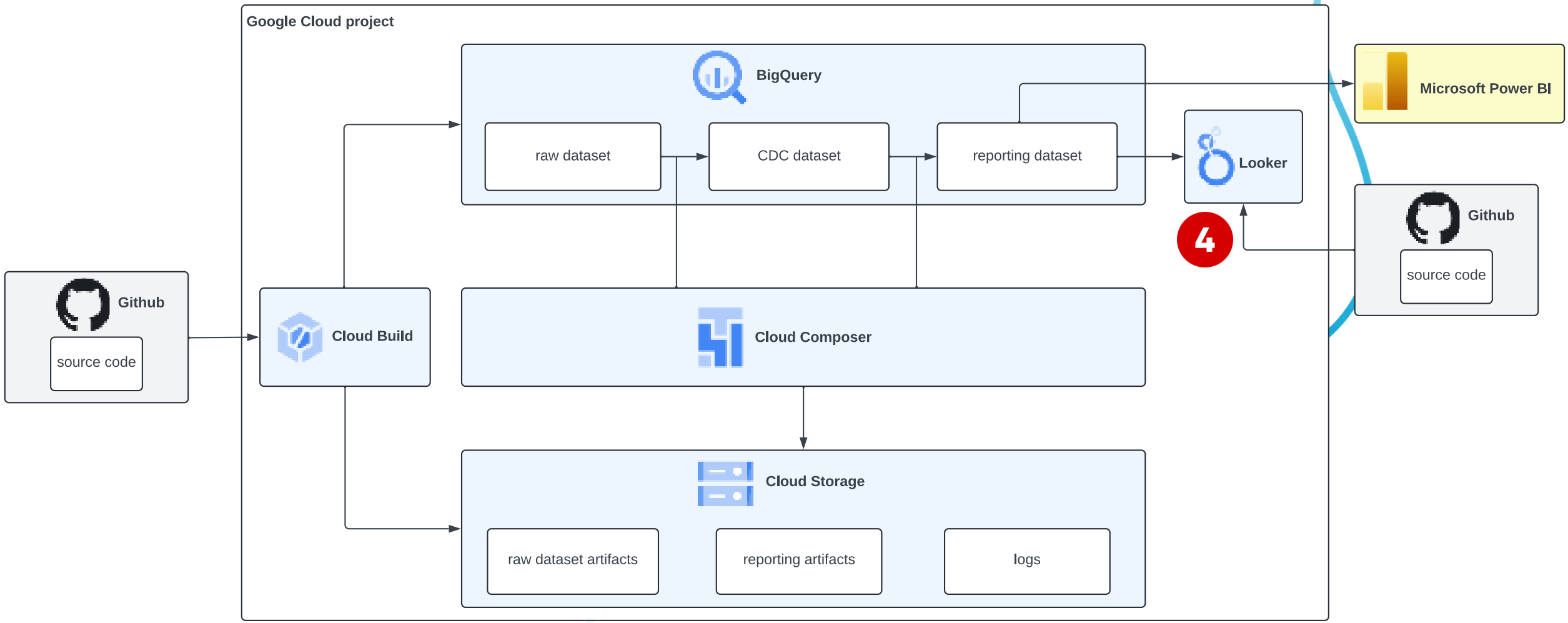
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



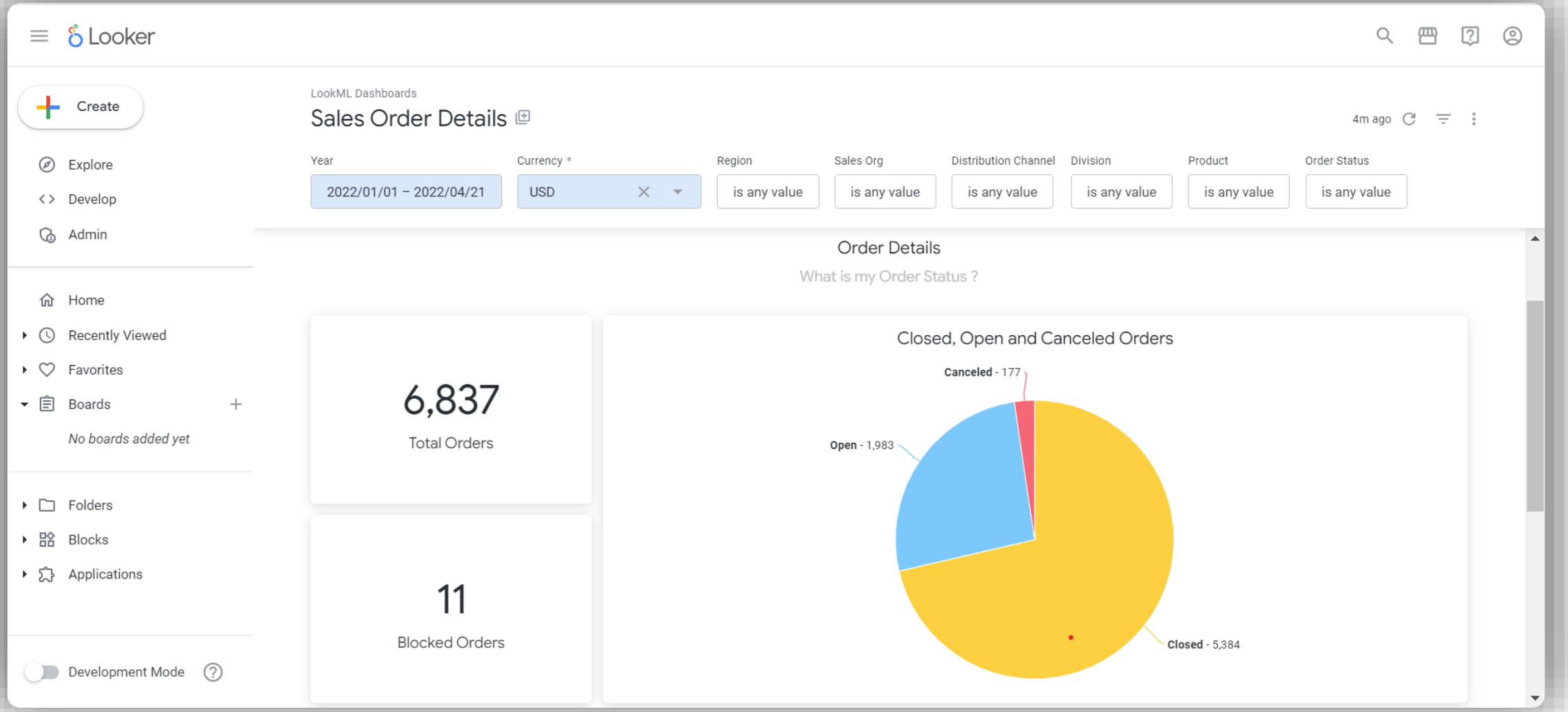
CLOUD COMPOSER MONITORING 3



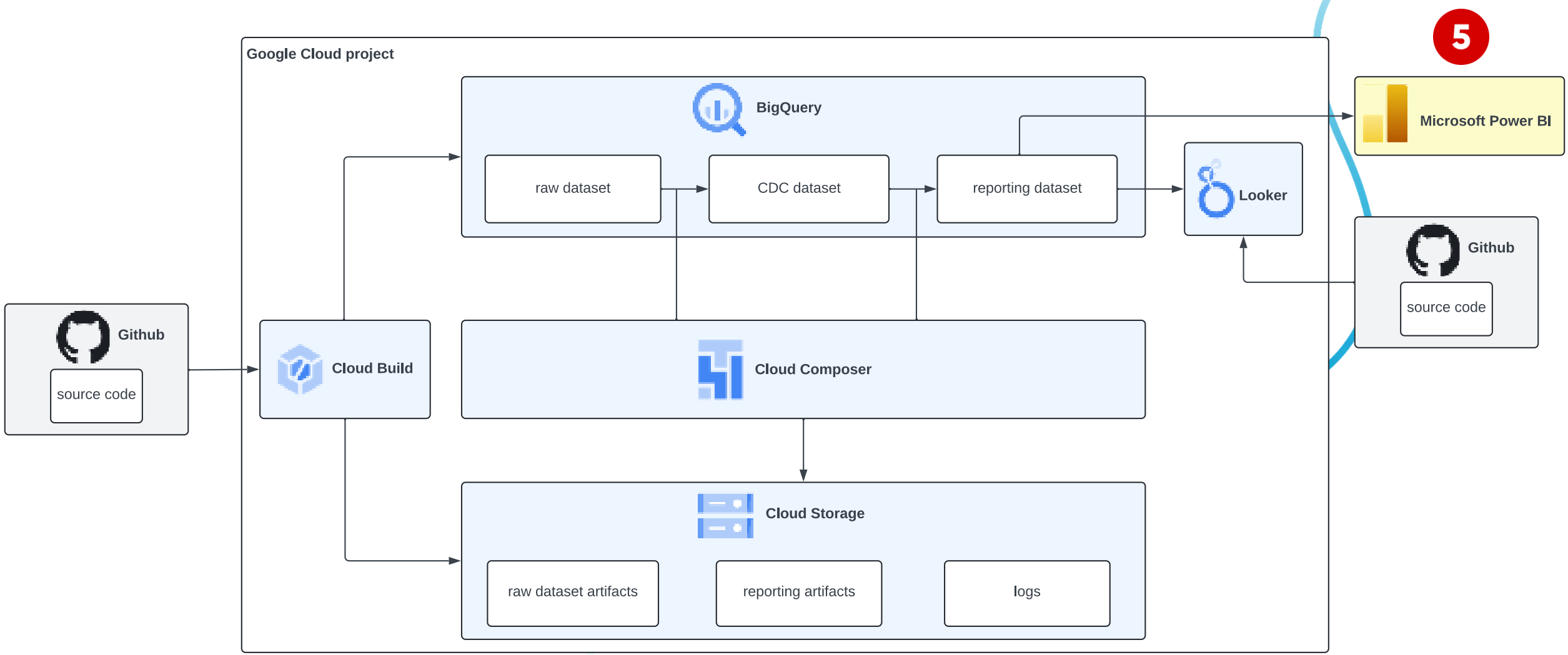
SAP WORKLOAD DEPLOYMENT ARCHITECTURE



LOOKER DASHBOARD 4

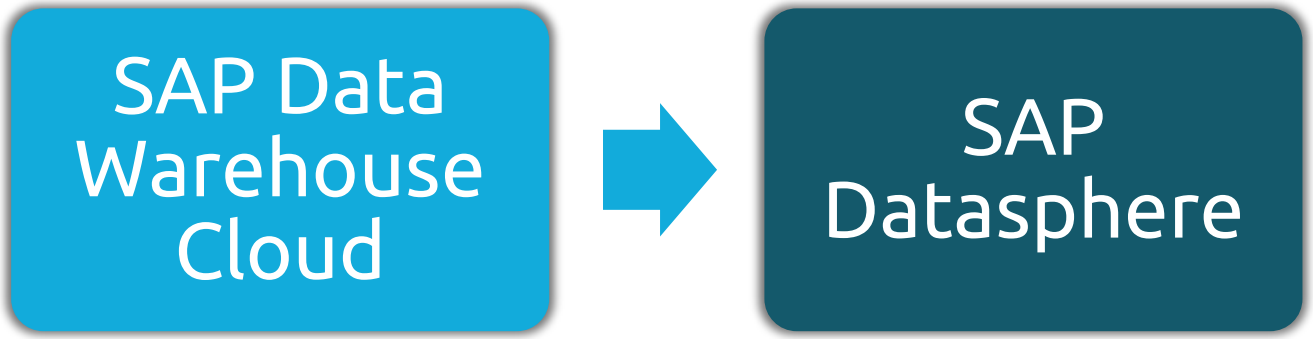


SAP WORKLOAD DEPLOYMENT ARCHITECTURE



SAP Datasphere

SAP DATASPHERE



Data Warehousing Strategy

Data Fabric Strategy

Data Fabric

A data architecture (a plan to implement) that:

- simplifies storage
- real-time access
- data sharing

SAP DATASPHERE

Data consumers — Planning and analytics | Intelligent data apps | Data science

SAP Datasphere running SAP BTP

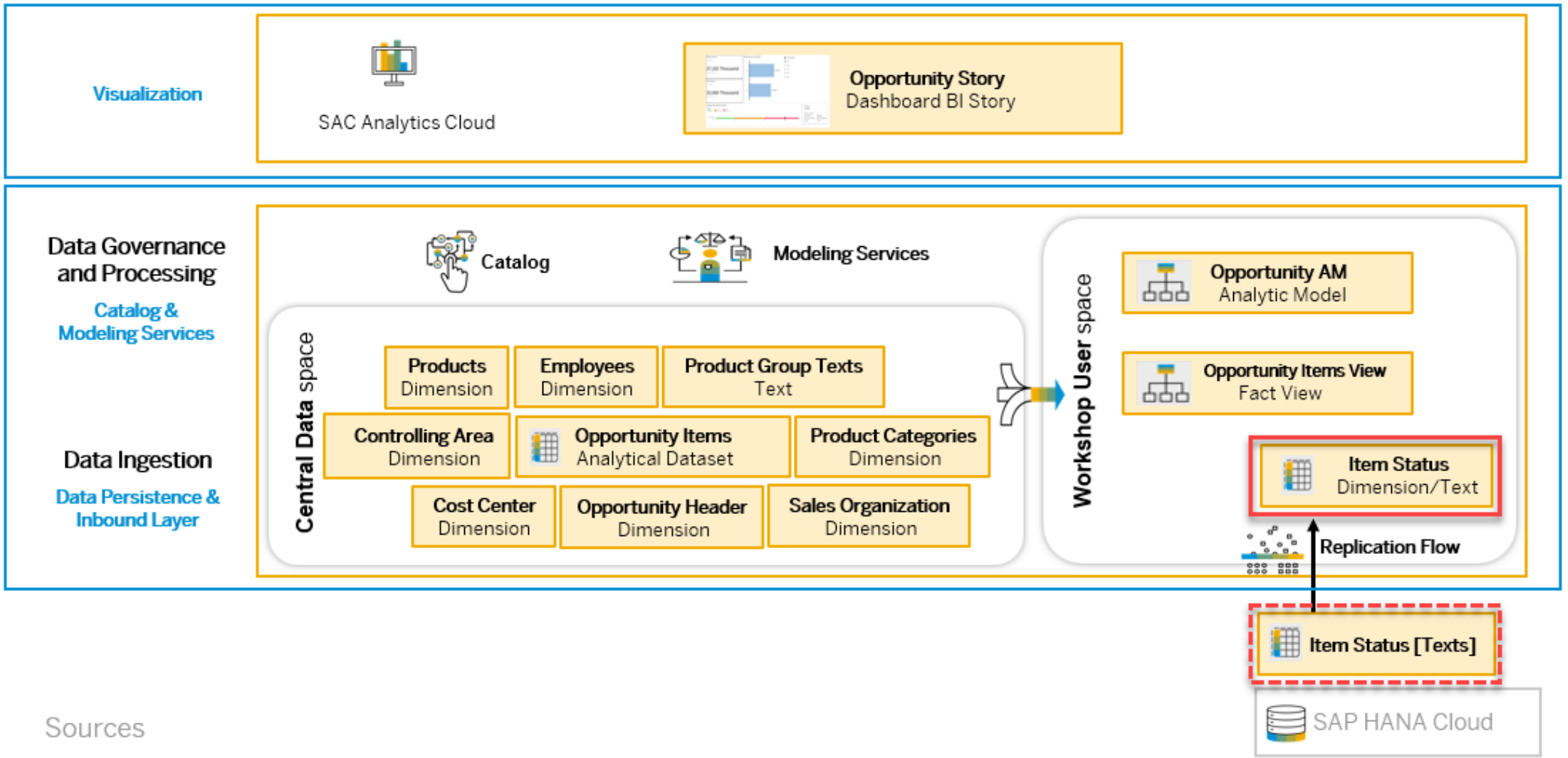
- **Discovery, consumption, sharing & self-service**
  space, data product, marketplace
- **Governance**
  cataloging, metadata management, lineage, privacy, quality
- **Business semantics**
  semantic onboarding, entity import, business content
- **Modelling**
  relational & rich analytic models, lifecycle management, authorizations
- **Data integration & preparation**
  real-time replication & federation, transformation, scheduling, monitoring

SAP and non-SAP data — Apps | On-premises Data Warehouses | Cloud Data Warehouses and Lakehouses | Relational Databases | Unstructured/ Semi-structured Data

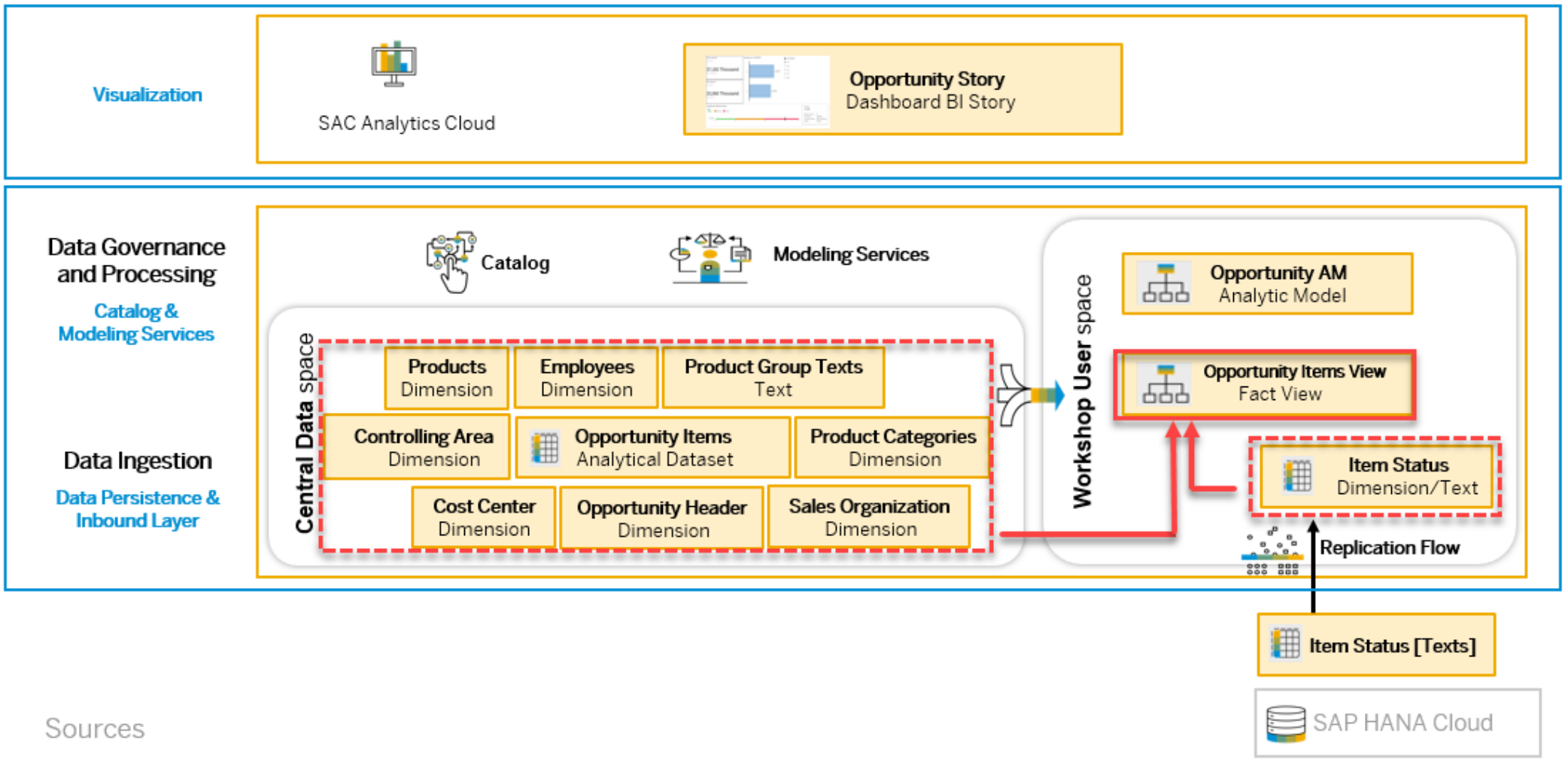
Open data ecosystem partners

- Collibra
- CONFLUENT
- DataRobot
- databricks
- Google Cloud

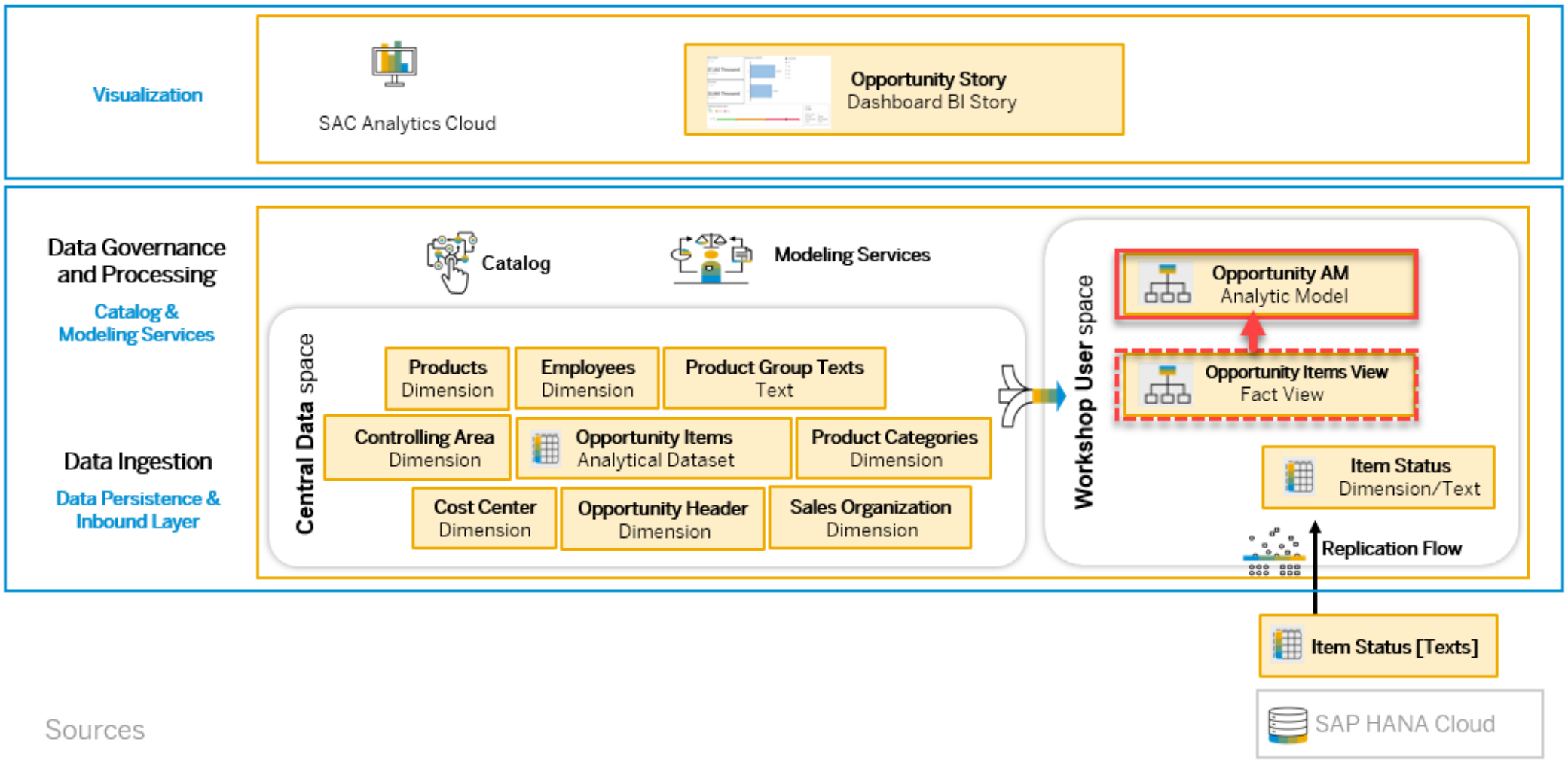
WORKSHOP – Hands-on with SAP Datasphere



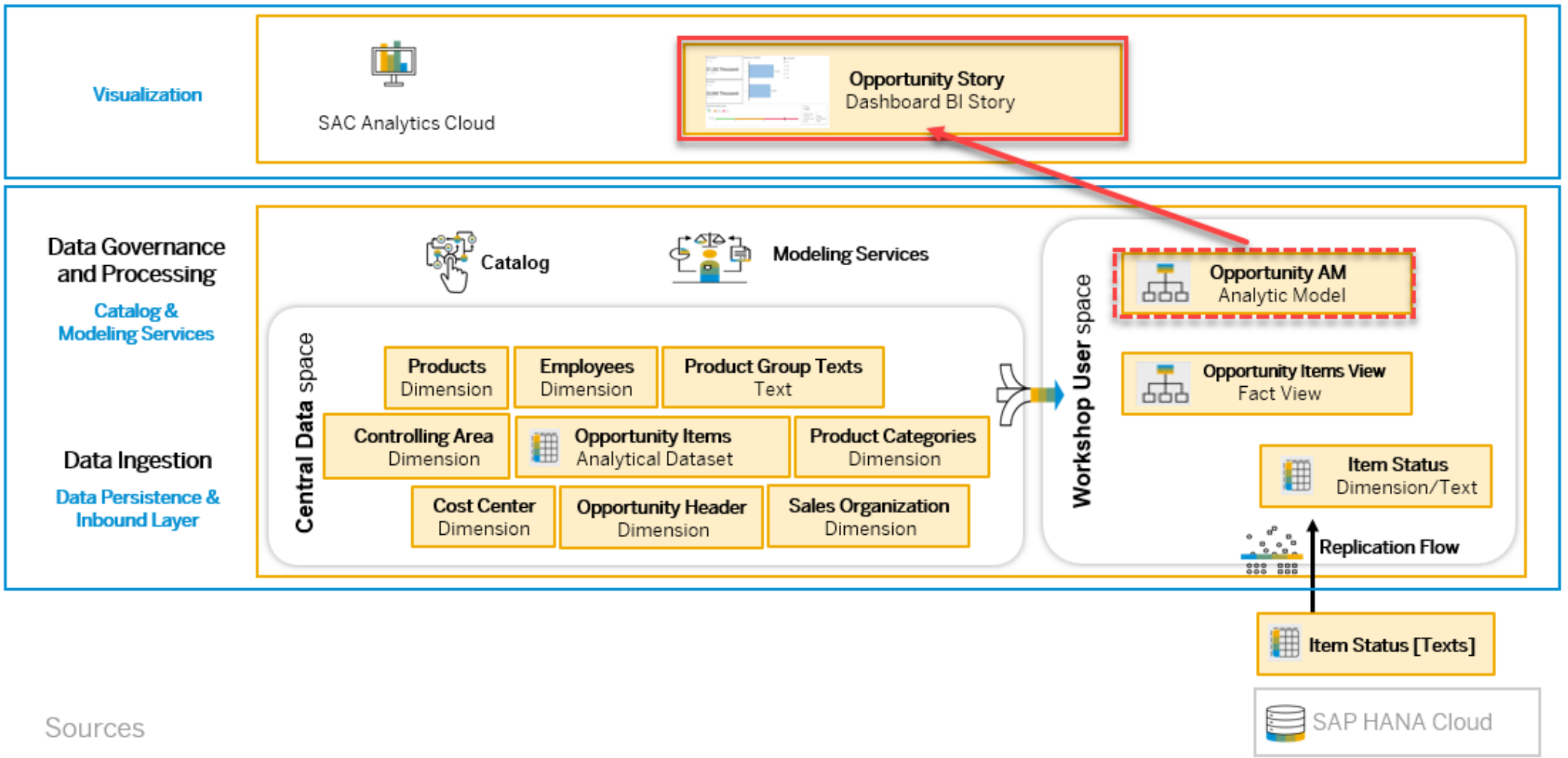
WORKSHOP – Hands-on with SAP Datasphere



WORKSHOP – Hands-on with SAP Datasphere

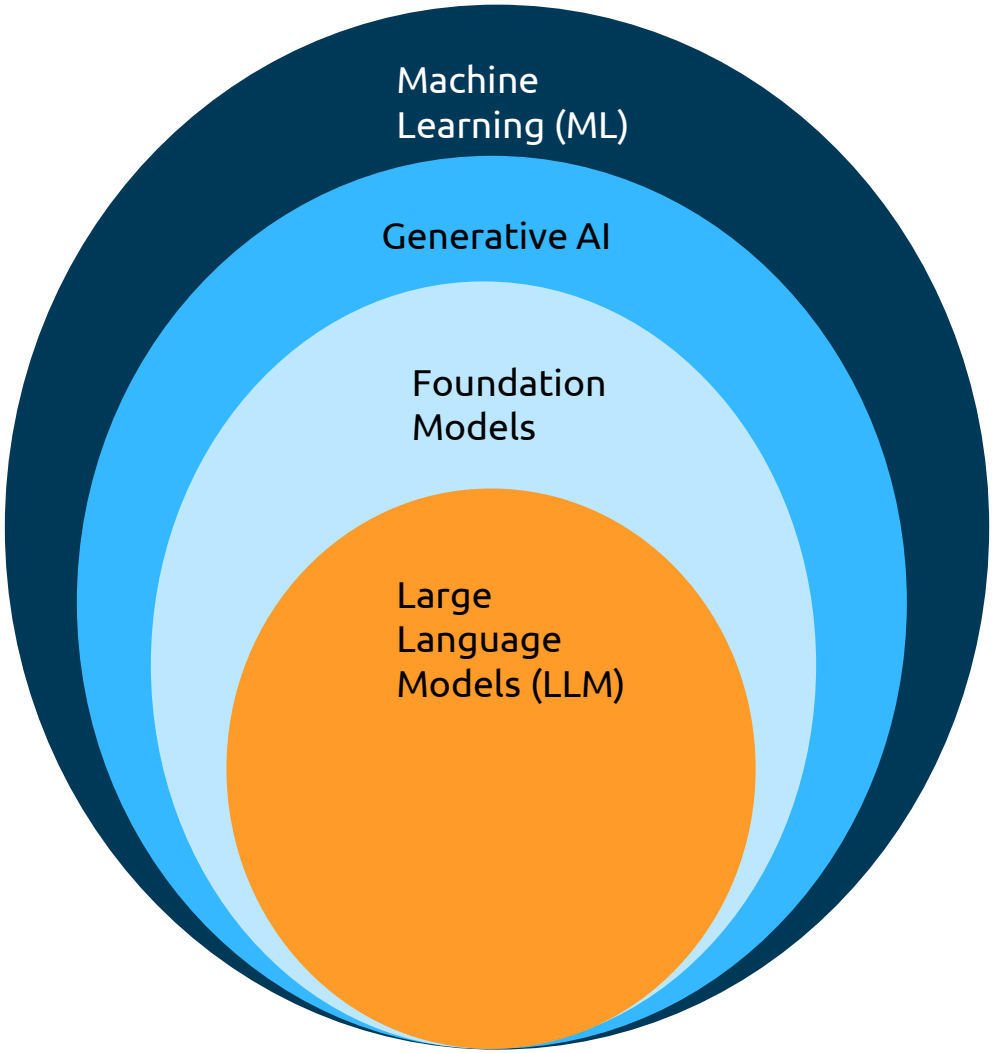


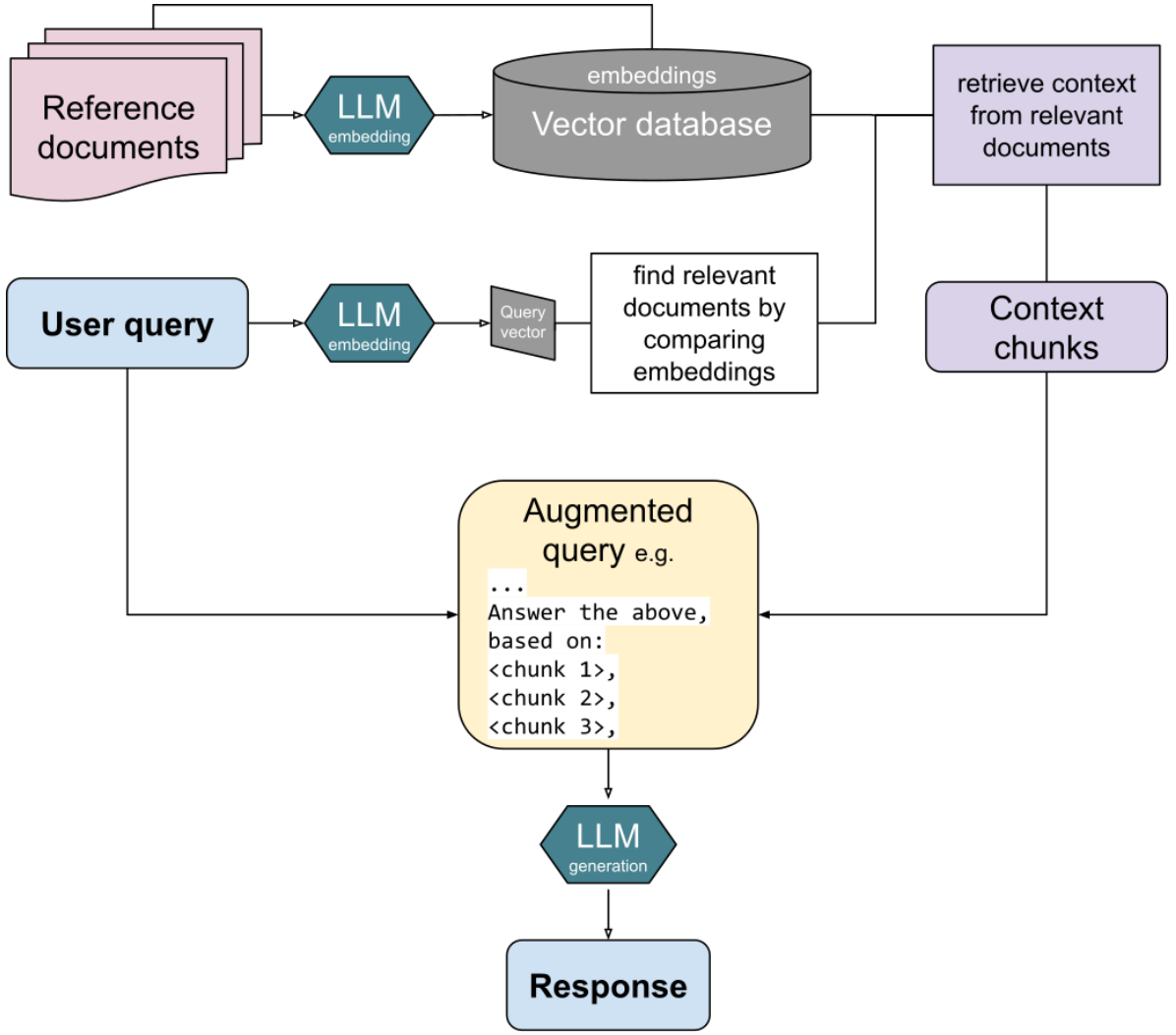
WORKSHOP – Hands-on with SAP Datasphere



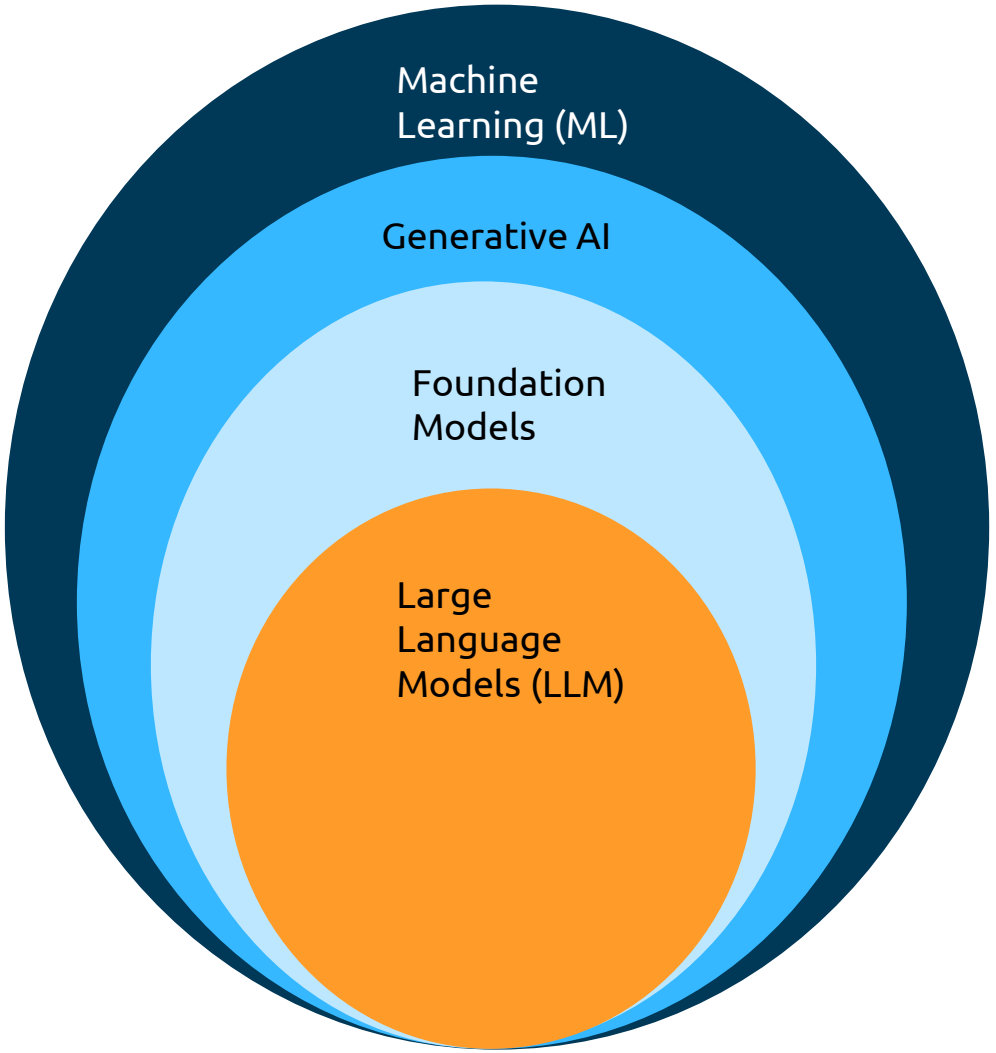
Generative AI project

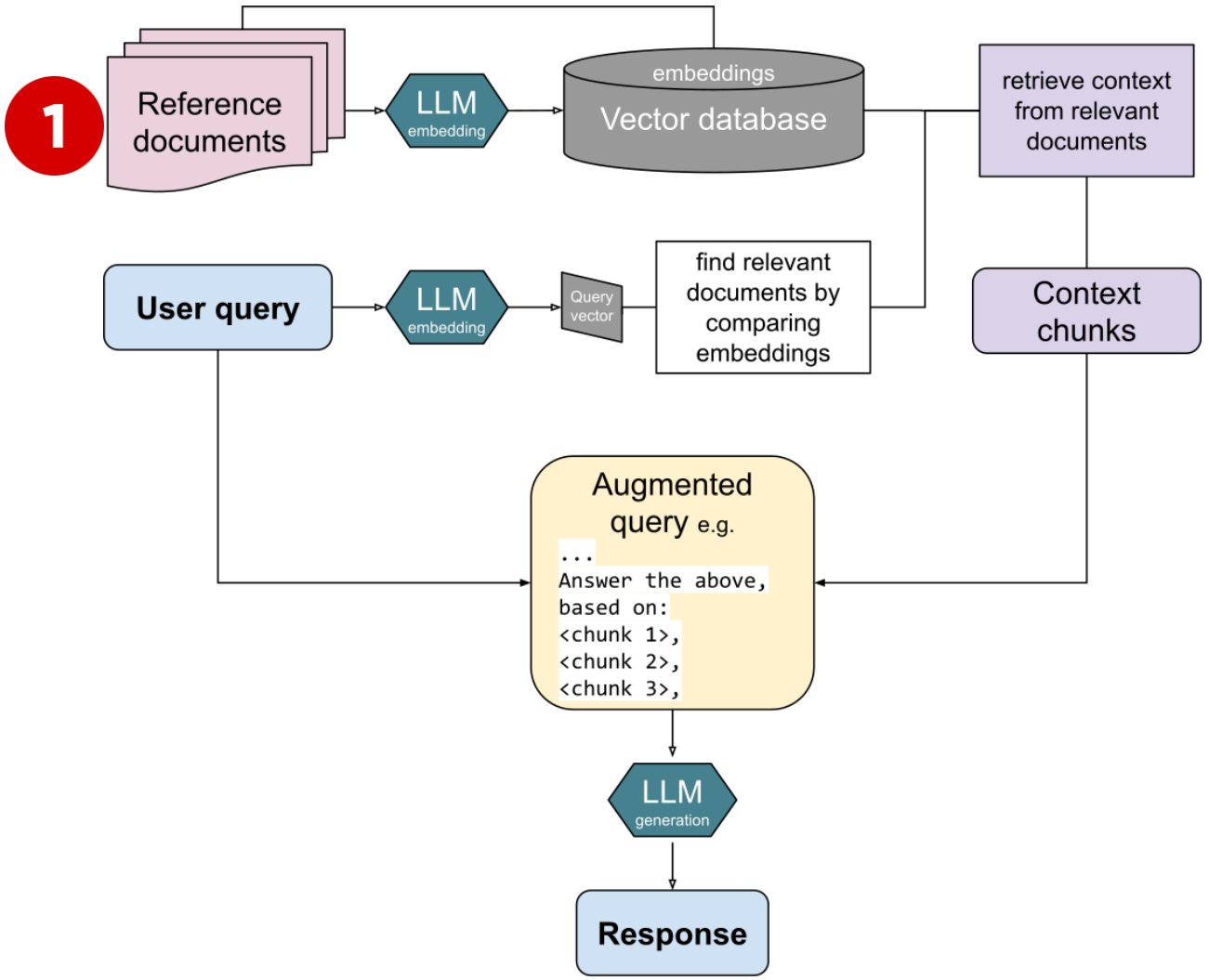
GENERATIVE AI & RETRIEVAL AUGMENTED GENERATION (RAG)



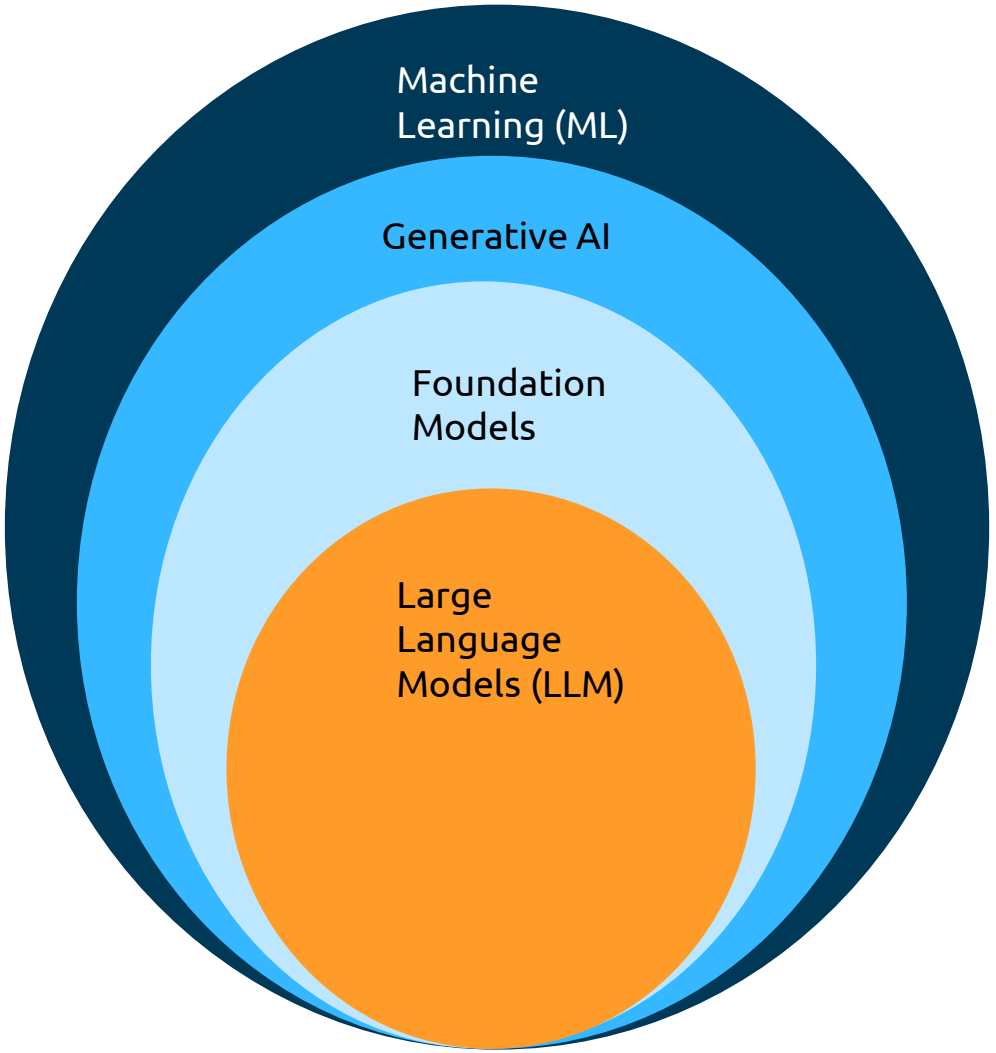


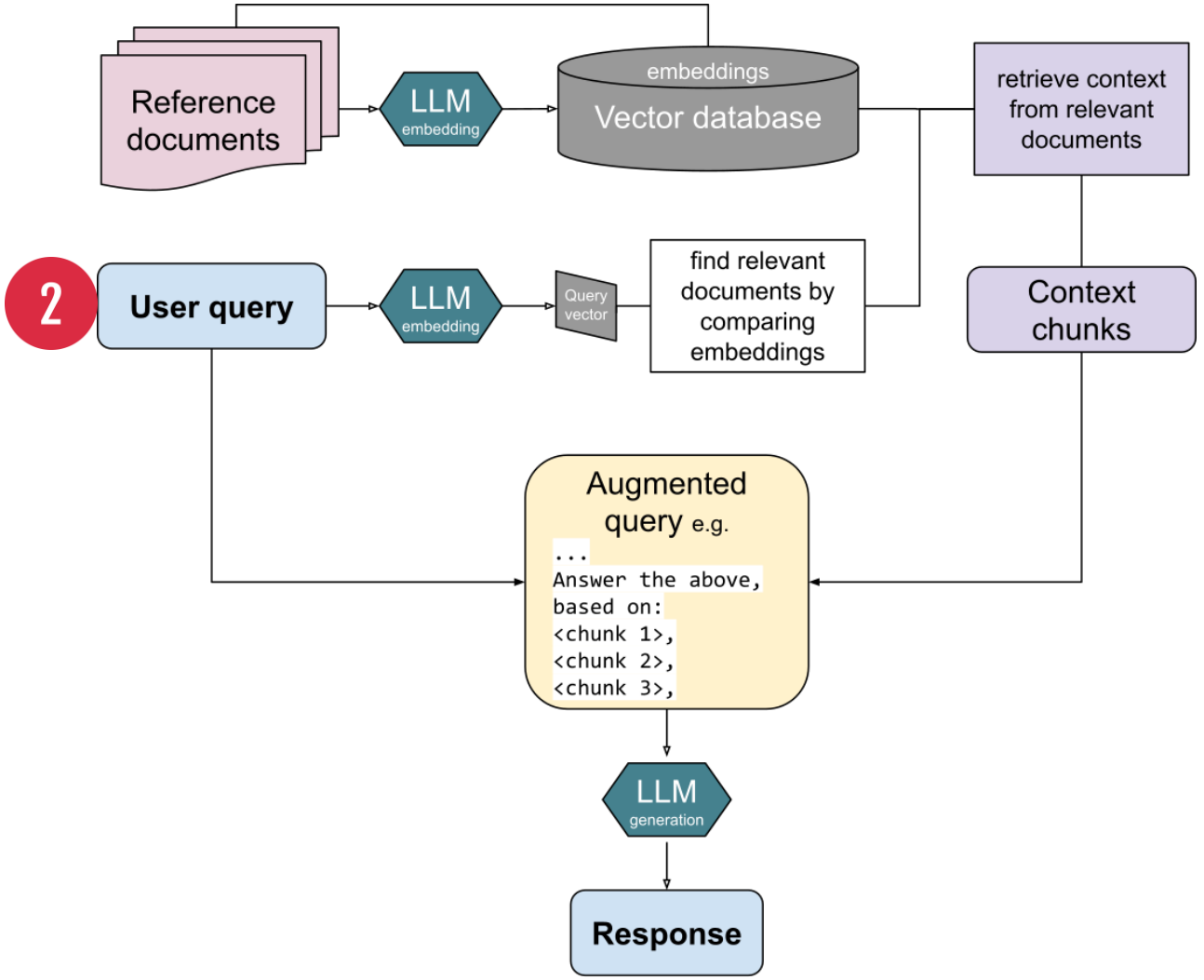
GENERATIVE AI & RETRIEVAL AUGMENTED GENERATION (RAG)



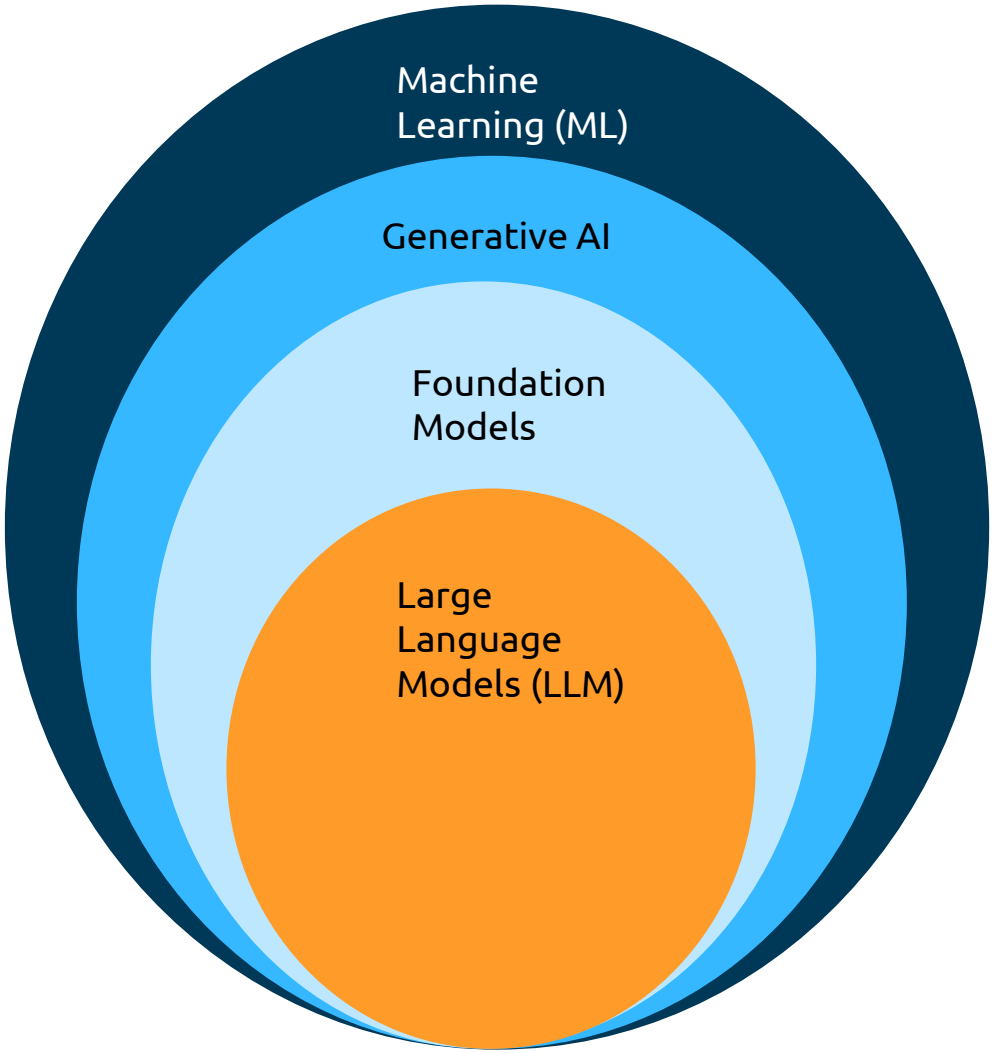


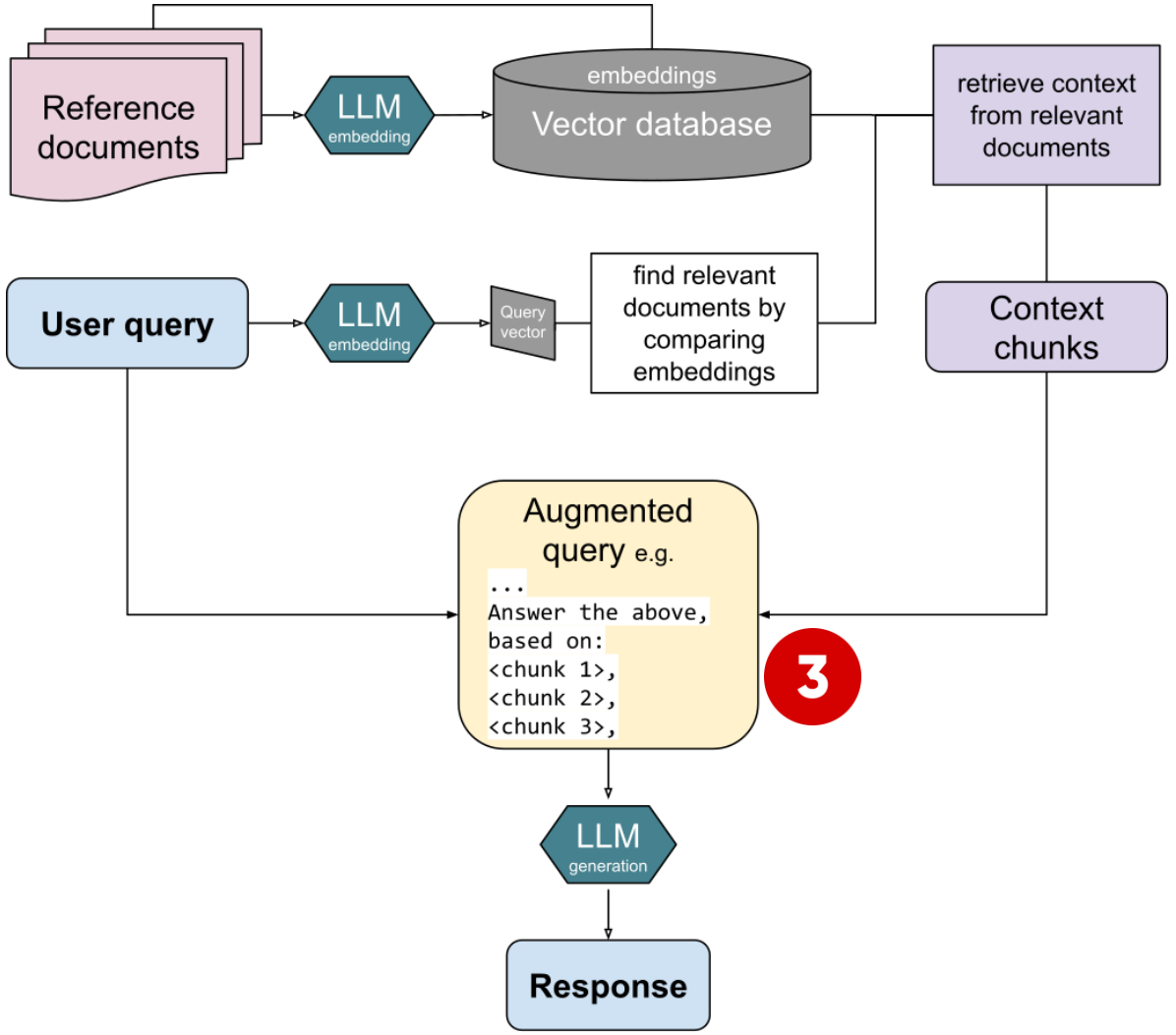
GENERATIVE AI & RETRIEVAL AUGMENTED GENERATION (RAG)





GENERATIVE AI & RETRIEVAL AUGMENTED GENERATION (RAG)





THE CLIENT

Meets the needs of approximately

725 000

customers
around the world

76 billion dollars
in revenue in 2023

More than

72 000

employees

Operates in

334

distribution centers
around the world

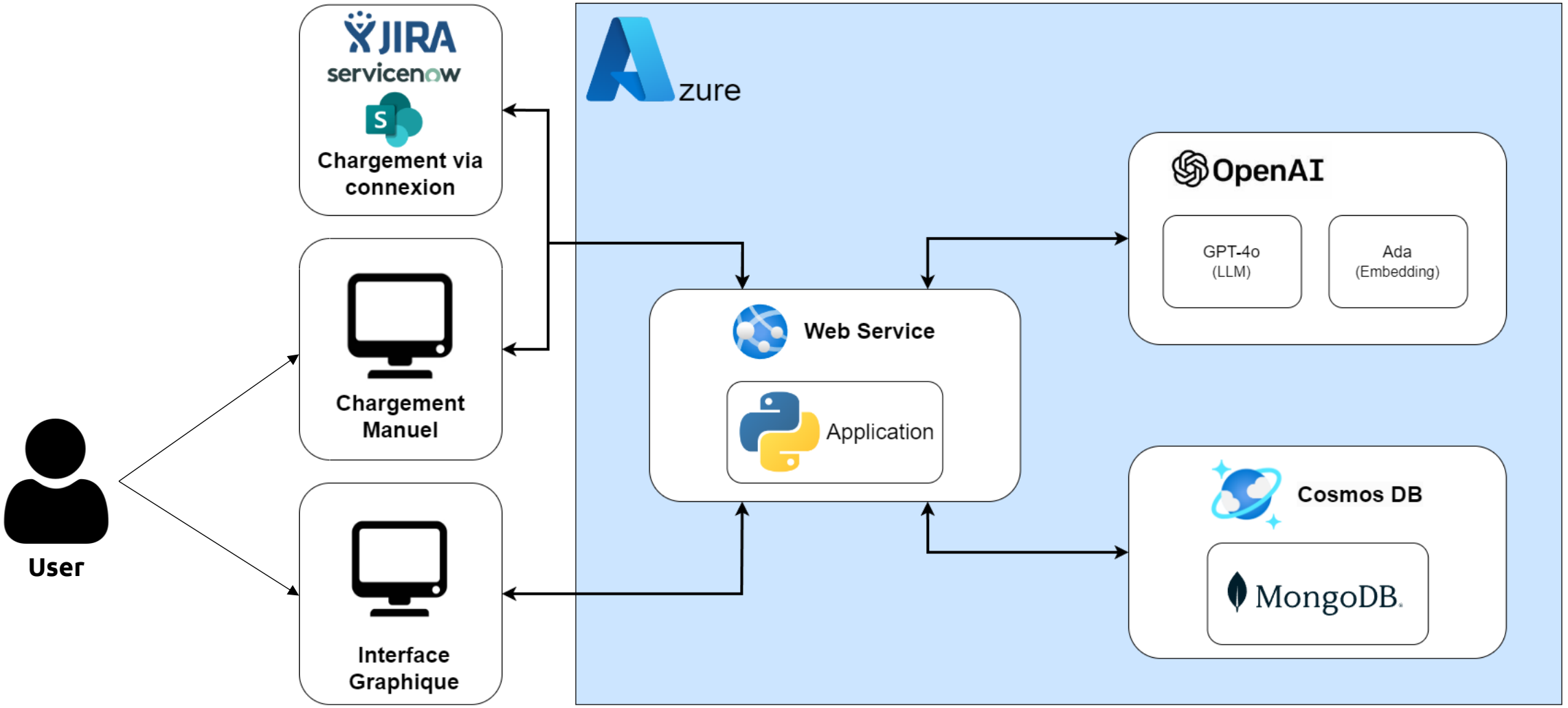
Sector

Worldwide Food processing.
Essential equipment and supplies for restaurant and hotel industries.

Need

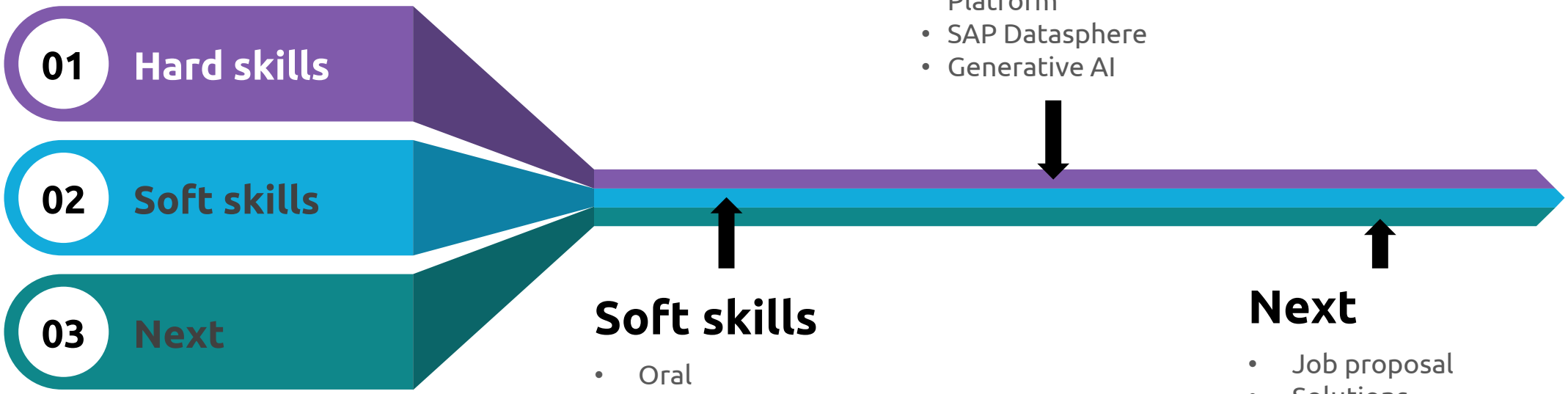
Provide ticket resolution assistance to SAP support technicians using a chatbot.

KM AGENT ARCHITECTURE



CONCLUSION

Conclusion



Hard skills

- Cloud computing
- SAP Business Technology Platform
- SAP Datasphere
- Generative AI

Soft skills

- Oral communication
- Powerpoint presentations
- English
- Teamwork

Next

- Job proposal
- Solutions Architect career

THANK YOU FOR PAYING ATTENTION !

Capgemini

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2023 global revenues of €22.5 billion.

Get the future you want | www.capgemini.com

Capgemini