# Data Innovation Overview

Data Innovation Team
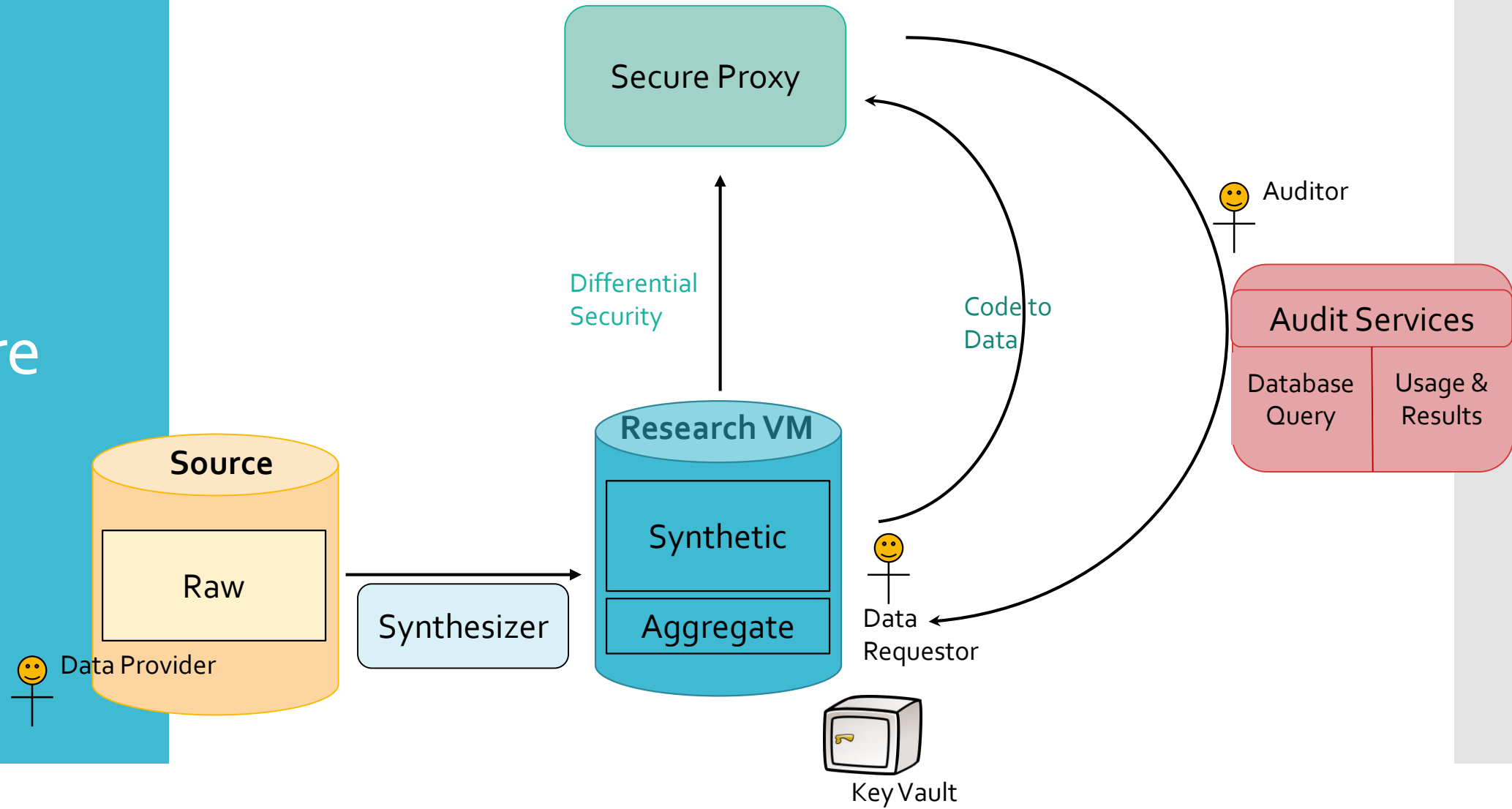
# Work Flow

## Ingestion Workflow

Raw Data Provided → Categorize Data → Run through Synthesizer Process → Create Synthesized Data

→ Create Aggregate Data

## Consumption Workflow

Write/Test code against Synthetic DB → Auditor Services Review Code → Code Executed Against Raw Data
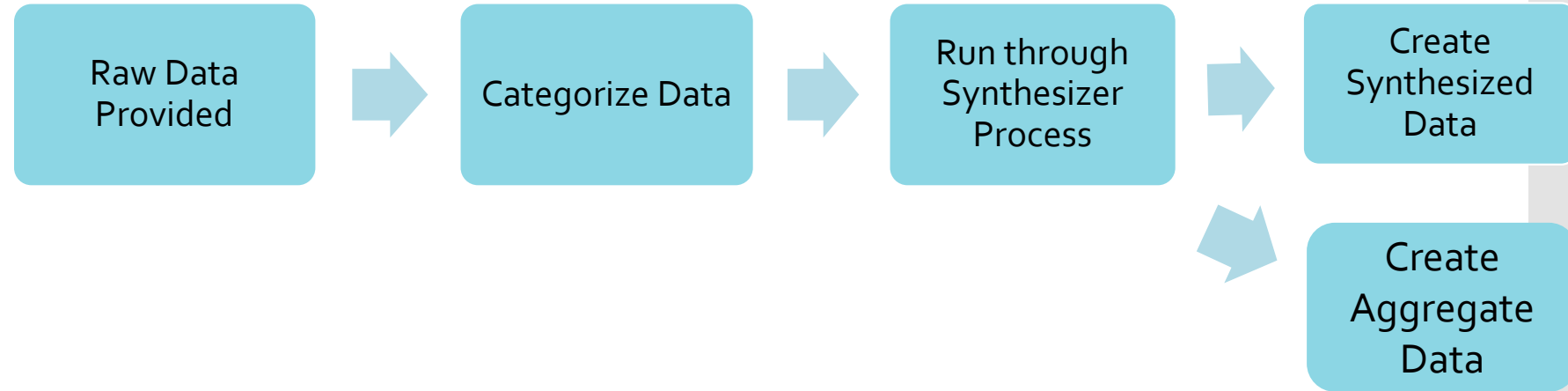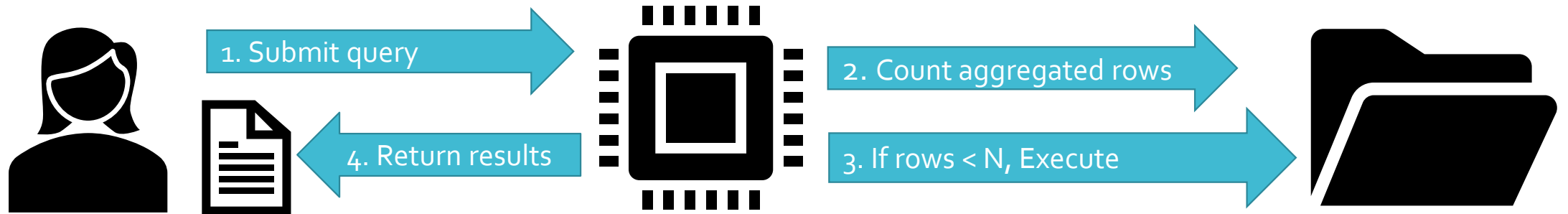
# Service Descriptions

- Synthesizer
  - a custom package built with the ability to accept the categories and also to run some commonly available static data anonymization, such as the Dynatrace Anonymization API.

- Secure Proxy for Differential & Privacy and Code to Data Security
  - designed to provide user access to a raw (unmasked, but anonymized) dataset in a controlled manner, by providing two mechanisms for querying the dataset which will restrict egress of PII data, described below

- Auditor
  - Database Query Auditor
    - Query monitor: Logs queries to the real database
    - Call stack authenticator: Reviews queries to sensitive information
  - Usage and Results Auditor
    - Automated tools: Detect schemes to leak sensitive information in results
    - Human auditor: Final review, paying extra attention to flagged items

# Differential Privacy

**Researcher** | **Secure Proxy** | **Raw Data (Likely in an RDBMS)**

1. Submit query

2. Count aggregated rows

3. If rows < N, Execute

4. Return results

# Code to Data Security

**Researcher** | **Secure Proxy** | **Raw Data (Likely in an RDBMS)**

1. Submit Code

2. Execute Code

3. Return results w/ Metadata

4. Return results

**Audit System**

4. Audits results