Google Cloud: Automating Cloud Data Governance

A conversation with

Mose Tronci
Industry Solution Architect
Google Cloud

Mark Tomlinson
Principal Architect
Google Cloud
Poll #1

What type of organisation do you work for?

- Enterprise
- Public sector
- Technology provider / Hyperscaler
- Consulting
- Other
Cloud Data Management Capabilities (CDMC)

Cloud Challenges

• **Regulatory Mandate:** safeguard cloud computing services to protect customers’ sensitive information
• **Inefficiency:** data, technology, regulatory and planning challenges on nearly every cloud implementation
• **Complexity:** 89% of firms use 2 or more cloud providers*

CDMC Group Objectives

1. **Best Practices** for a hybrid-cloud world
2. **Cloud Data Controls** to meet regulatory obligations for protecting Sensitive Data
3. **Accelerate Trusted Cloud Adoption** modeled after the DCAM data management industry framework

* Source: Flexera, 2022 State of the Cloud
Cloud Data Management Capabilities (CDMC) Industry Engagement

100+ Organizations – 300+ SME participants – Cross Industry Leading Organizations

CDMC Working Group

- LSEG Morgan Stanley
- J.P. Morgan
- Standard Chartered
- HSBC
- Societe Generale
- Wells Fargo
- Deutsche Bank
- BARCLAYS
- TP ICAP
- Lloyds Bank
- DTCC
- Tradeweb
- PayPal
- TD Bank
- BNP Paribas
- Nasdaq
- Freddie Mac
- KPMG
- CAPCO
- EY

Cloud & Technology Providers

- Microsoft Azure
- IBM Cloud
- Google Cloud
- AWS
- colibra
- snowflake
- Informatica
- BigID
- data.world
- PRIVITAR
- Solidatus
- securiti

Regulatory Engagement

- US: Federal Reserve, SEC, CFTC, FDIC
- Canada: OSFI
- UK: BoE, FCA, ICO
- EU: ECB, ESMA (pending)
- Germany: BaFin
- Switzerland: FinMA
- Australia: APRA
- Singapore: MAS
- Israel: Bank of Israel
- India: RBI, SEBI (pending)
- Africa/Middle East: 20+ Regulators
- Others in process…

CDMC Adoption Support

- Training Courses
- Cloud Service Certification
- Open Source Tools
- CDMC Authorized Partner Program

CDMC Working Group Delivered release v1.1

28 September 2021

CDMC Framework Validated

Cross Industry

- Life Sciences
- Telecommunications
- Manufacturing
- Retail / Services / CPG
- Consumer Tech
- Government / Defense

© 2023 EDM Council Inc.
Google Cloud becomes the next CDMC Cloud Certified Solution

How to migrate sensitive data with confidence using Google Cloud’s CDMC-certified architecture

June 26, 2023

EDM Council Designates Google Cloud as a Certified Cloud Solution

Mon, Jun 26, 2023

Google Cloud clients migrating sensitive data to the cloud can do so with greater confidence now that Google Cloud has been certified against EDMC’s CDMC control framework.
Automating Cloud Data Governance

Mose Tronci, Industry Solution Architect
Mark Tomlinson, Principal Architect
Poll #2

What is your role?

- Data Management professional
- CDO
- Data Engineer
- Cloud Architect
- Project Manager
- Consultant
- Sales
- Other
CDMC Main Elements Recap

**CDMC 1.1.1 Framework**
Defines the 14 capabilities and 37 sub-capabilities across 6 domains.

**CDMC 1.1.1 Assessment Template**
Spreadsheet to facilitate sub-capability assessments using scoring guide of 1 to 6.

**CDMC 1.1.1 Key Controls**
Specifies 14 key controls across the 6 framework domains.

**CDMC Controls Procedures and Test Specifications**
Outlines tests and evidence required for third-party CDMC platform/solution certification.
Cloud Data Management Capabilities

14 Key Controls for Managing Data Risk

- Data Sovereignty and Cross-Border Movement of sensitive data
- Authoritative Data Sources and Provisioning Points
- Data Ownership
- Data Control Compliance
- Cost Metrics
- Data Lineage
- Data Quality Measurement
- Data Retention, Archiving and Purging
- Data Privacy Impact Assessments
- Data Lifecyle
- Protection & Privacy
- Accessibility & Usage
- Cataloguing & Classification
- Governance & Accountability
- Entitlements and Access for Sensitive Data
- Data Classification
- Data Cataloguing
- Data Consumption Purpose
- Security Controls
# CDMC Assessment types

<table>
<thead>
<tr>
<th>Assessment Type</th>
<th>Target</th>
<th>Use For</th>
<th>Basis</th>
<th>Uses CDMC Capability Framework?</th>
<th>Uses CDMC Key Controls Framework?</th>
<th>Uses CDMC Key Controls Test Procedures?</th>
<th>Assessed By</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDMC Certification Assessment</td>
<td>Individual</td>
<td>Personal skills, Planning</td>
<td>Training Attendance</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Online Exam</td>
</tr>
<tr>
<td>CDMC Readiness Assessment</td>
<td>Consumer</td>
<td>Planning and Progress Tracking</td>
<td>Sentiment or Evidence-based</td>
<td>Yes</td>
<td>Optional</td>
<td>Optional</td>
<td>Self- or CDMC Authorized partner</td>
</tr>
<tr>
<td>CDMC Certification Assessment</td>
<td>Consumer</td>
<td>Formal Audit, Regulatory Preparation</td>
<td>Evidence-based</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>CDMC Authorized partner</td>
</tr>
<tr>
<td>CDMC Certification Assessment</td>
<td>Cloud Providers</td>
<td>Planning and Independent Evidence</td>
<td>Evidence-based</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Independent CDMC Authorized partner</td>
</tr>
</tbody>
</table>

**Key Differences**

- **Focus today**
- **Relevant Lessons**

---

Focus today

Relevant Lessons
Our journey to certification

LESSON: Ensure sufficient time and investment is available
Introducing the Google CDMC Team

Executive Sponsors
- Prajakta Damle
  Director, Prod Mgmt (Data Gov)
  EDMC Board
- Andy Chang
  Group Prod Mgr (Security)
  Blueprint Sponsor

CDMC GCP Reference Architecture Team
- Mose Tronci
  Industry Solns Architect
  EDMC Lead
- Mark Tomlinson
  Principal Architect
  Project Lead
- Eduardo Marreto
  Cloud Data Engineer
- Erlander Lo
  Senior Product Manager, Security
  Blueprint Lead
- John Swain
  Customer Engineer, Smart Analytics
- Shirley Cohen
  Solutions Architect
  Tag Engine Lead
- Svitlana Gavrylova
  Data Management Practice Lead UK

CDMC 1.1 Specification Working Group Contributors
- Mose Tronci
  Industry Solns Architect
  EDMC Lead
- Andrew Ikonnikov
  Senior Product Manager, Data Governance
- Kishore Gopalan
  Principal Architect
- Prajakta Pandharkar
  Customer Engineer, ML
- Rochak Lamba
  Innovation Lead, Financial Services, PSO
- Svitlana Gavrylova
  Data Management Practice Lead UK
- Thinh Ha
  Cloud Data Engineer, PSO

LESSON:
Check your team has the right composition of skills
Google Cloud CDMC Reference Architecture

core principles

<table>
<thead>
<tr>
<th>#</th>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Build on prior work and maximize reuse</td>
</tr>
<tr>
<td>2</td>
<td>Implement all 14 CDMC controls</td>
</tr>
<tr>
<td>3</td>
<td>Align to product roadmap, but minimize dependencies on future items</td>
</tr>
<tr>
<td>4</td>
<td>Prefer native Google Cloud features over third party services</td>
</tr>
<tr>
<td>5</td>
<td>Strong preference for full automation, but pragmatic, iterative approach</td>
</tr>
<tr>
<td>6</td>
<td>Author with intent to publish externally for customers and partners</td>
</tr>
<tr>
<td>7</td>
<td>Provide value-add for both new and established Google Cloud clients</td>
</tr>
</tbody>
</table>

**LESSON:**
Consider bespoke vs. reusable implementation

**LESSON:**
Consider starting with limited automation and iterate
Delivery process overview

**Analysis & Design**
- Draft
- Review
- Rework

**Implementation**
- Control detector
- Tag Engine

**Evidence & Assessment**
- Evidence doc
- Review
- Rework

**Documentation & Publication**
- Github docs
- Architecture whitepaper

**Feedback**
- Capture feedback from users
- Iterate

**Lesson:** Identify sponsoring clients early and engage throughout

- **Control detector**
- **Tag Engine**

- **Evidence doc**
- **Review**
- **Rework**

---

**Control detector**

**Tag Engine**

---

**Evidence doc**

**Review**

**Rework**

---

**Github docs**

**Architecture whitepaper**
Poll #3

Where are you on your CDMC journey?

• Not started
• Currently investigating
• Project in progress
• Already assessed
• Already certified
Analysis & Design - key stats

The internal design for the reference architecture had over 100 pages with 60 figures, and 136 design decisions.

The CDMC framework specification extends to 165 pages and details 37 sub-capabilities of data management and governance.

LESSON:
Seek clarification in areas of ambiguity

No perfect answer - need to put a stake in the ground
Implementation: Overview and Control Map

LESSON: Common capability groupings regardless of implementation

LESSON: Anticipate and budget for the evolution of the platform
All findings in one place
Implementation: Test Datasets, subset of TPC-DI

Our TPC-DI-derived test data provides a large selection of data classifications and meaningful volumes to demonstrate our capabilities.
KPMG Assessment process

GCP CDMC RA Evidence Docs
Evidence documents for each Control validating alignment to criteria in test specification (300+ pages)

LESSON:
Validate the minimum evidence required to satisfy the assessor
Documentation & Publication: Our primary technical deliverables

Implement the CDMC key controls framework in a BigQuery data warehouse

---

GCP CDMC Whitepaper (40+ pgs)
Overview of the reference architecture and controls

GitHub Repo
Publication of example assets (scripts, detectors, setup guide) into new standalone repo.
CDMC Information Model
Controls Tests Mappings

Version 1.1
October 2022

edmcouncil.org/page/CDMC

Copyright © 2022 EDM Council Inc. All rights reserved.
Possession and application subject to CDMC™ Terms of Use.
Data Governance Project

Field-level CDMC Tag Templates

Prepopulated Technical Metadata
Type

Data Sensitivity Template
Sensitive Field
Sensitive Type

Security Policy Template
Platform De-Identification Method
Application De-Identification Method

Acceptable Threshold
Meets Threshold
Most Recent Run

Cost Metrics Template
Total Query Bytes Billed
Total Storage Bytes Billed
Estimated Query Cost
Estimated Storage Cost

CDMC Sensitive Data Taxonomy
Personal Identifiable Information
Personal Information
Sensitive Personal Identifiable Info
Sensitive Personal Information

Impact Assessment Template
Subject Locations
Is DPIA
Is PIA
Impact Assessment Reports
Most Recent Assessment
Oldest Assessment

LESSON:
Consider tactical metamodel vs. CDMC information model
# How to use our assets!

<table>
<thead>
<tr>
<th>Established Google Cloud Clients</th>
<th>New to Google Cloud</th>
<th>Users of Other Clouds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Read CDMC capability and key controls frameworks</td>
<td>1. Read CDMC capability and key controls frameworks</td>
<td>1. Read CDMC capability and key controls frameworks</td>
</tr>
<tr>
<td>2. Review our whitepaper</td>
<td>2. Review our whitepaper</td>
<td>2. Review our whitepaper to identify any architectural patterns that you could reuse in your own control design</td>
</tr>
<tr>
<td>3. Perform CDMC capability assessment</td>
<td>3. Evaluate and implement Secure Data Warehouse Blueprint</td>
<td>3. Reach out to your own provider and understand about how you can reuse their certified architecture</td>
</tr>
<tr>
<td>4. Perform CDMC controls gap analysis</td>
<td>4. Download and implement CDMC GitHub repo asset in its entirety</td>
<td></td>
</tr>
<tr>
<td>5. Download and selectively adapt CDMC GitHub repo assets to address control gaps in your environment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questions?

Google Cloud

EDM Webinar
Google’s Data and AI Cloud
Simplifying data and AI experiences for all users with Gen AI

Govern Data and AI with Dataplex

<table>
<thead>
<tr>
<th>Databases</th>
<th>Data Analytics</th>
<th>AI &amp; ML</th>
<th>Business Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlloyDB</td>
<td>BigQuery</td>
<td>Vertex AI</td>
<td>Looker</td>
</tr>
<tr>
<td>Spanner</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Google, Open Source, and 3rd party foundation models (text, chat, code, video, image, audio)

Google Cloud GPUs/TPUs
Run across clouds / on prem
Most open ecosystem
Recap of our lessons

1. Ensure sufficient time and investment is available
2. Check your team has the right composition of skills
3. Seek clarification in areas of ambiguity
4. Anticipate and budget for the evolution of the platform
5. Consider bespoke vs. reusable implementation
6. Consider starting with limited automation and iterate
7. Common capability groupings regardless of implementation
8. Test data selection critical
9. Identify sponsoring clients early and engage throughout
10. Validate the minimum evidence required to satisfy the assessor
11. Consider tactical metamodel vs. CDMC information model
Thank you!

FOR MORE INFORMATION:

Mose Tronci
Industry Solution Architect
Google Cloud
mtronci@google.com