PUBLIC SECTOR SUMMIT ONLINE

Building a protected data lake in AWS

Richard Tomkinson

Managing Director

Cloudten

aws PUBLIC SECTOR SUMMIT ONLINE

About Cloudten

Redefining Infrastructure. Releasing Potential. Empowering Innovation

Trusted Partner

Cloudten is a trusted partner to some of Australia's biggest names across enterprise and public sector. **We are a Prequalified Supplier to the federal government and a Preferred Supplier to all state governments in Australia.**

Proven Quality

We are an AWS Advanced Consulting Partner with a proven track record in delivering quality cloud solutions across a wide range of verticals, which can help realise your IT transformation programme.

Data Specialists

Cloudten has actively worked on some of the most cutting edge data platforms in the country. We run a team of experienced professionals with skillsets from data engineering through to data science and machine learning.

aws partner network

Advanced

Consulting Partner

- Security Competency
- DevOps Competency
- Government Competency
- Public Sector Partner

What does it mean to be PROTECTED?

CLOUDTEN

Protective marking	Business Impact Level	Compromise of information confidentiality would be expected to cause:
UNOFFICIAL	No business impact	No damage. This information does not form part of official duty.
OFFICIAL	No or insignificant damage. This is the majority of routine information.	No or insignificant damage. This is the majority of routine information.
OFFICIAL: Sensitive	2 Low to medium business impact	Limited damage to an individual, organisation, or government generally if compromised.
PROTECTED	3 High business impact	**Damage** to the national interest, organisations, or individuals.
SECRET	4 Extreme business impact	**Serious damage** to the national interest, organisations, or individuals.
TOP SECRET	5 Catastrophic business impact	**Exceptionally grave damage** to the national interest, organisations, or individuals.

What does it mean to be PROTECTED?

CLOUDTEN

The Australian government uses the Information Security Manual (ISM) as the standard security framework to protect its data and systems.

For cloud-based workloads the ISM lists over 800 controls relating to recommended security best practices across a range of areas.

Whilst not every control needs to be met for every workload or system, an Information Security Registered Assessors Program (IRAP) assessment needs to completed before the workload can be certified to store PROTECTED data.

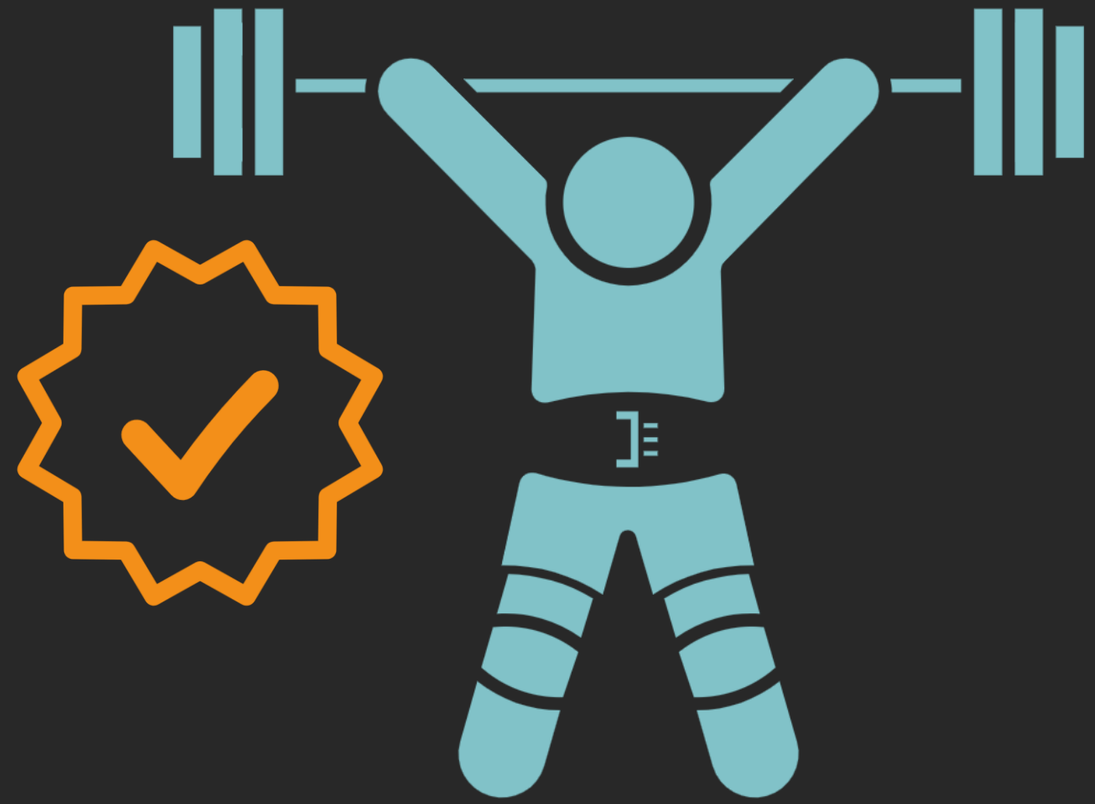
CLOUDTEN

AWS PROTECTED services

AWS has done quite a bit of the heavy lifting for us in advance.

64 AWS services across a range of categories have been IRAP certified by the Australian Cyber Security Centre (ACSC) up to PROTECTED level.

Whilst this does not automatically mean that any workload deployed using these services is inherently certified, using these services as building blocks it makes our job a lot easier.



How to leverage PROTECTED services

The self-assessment process

Access the IRAP PROTECTED Package on AWS Artifact

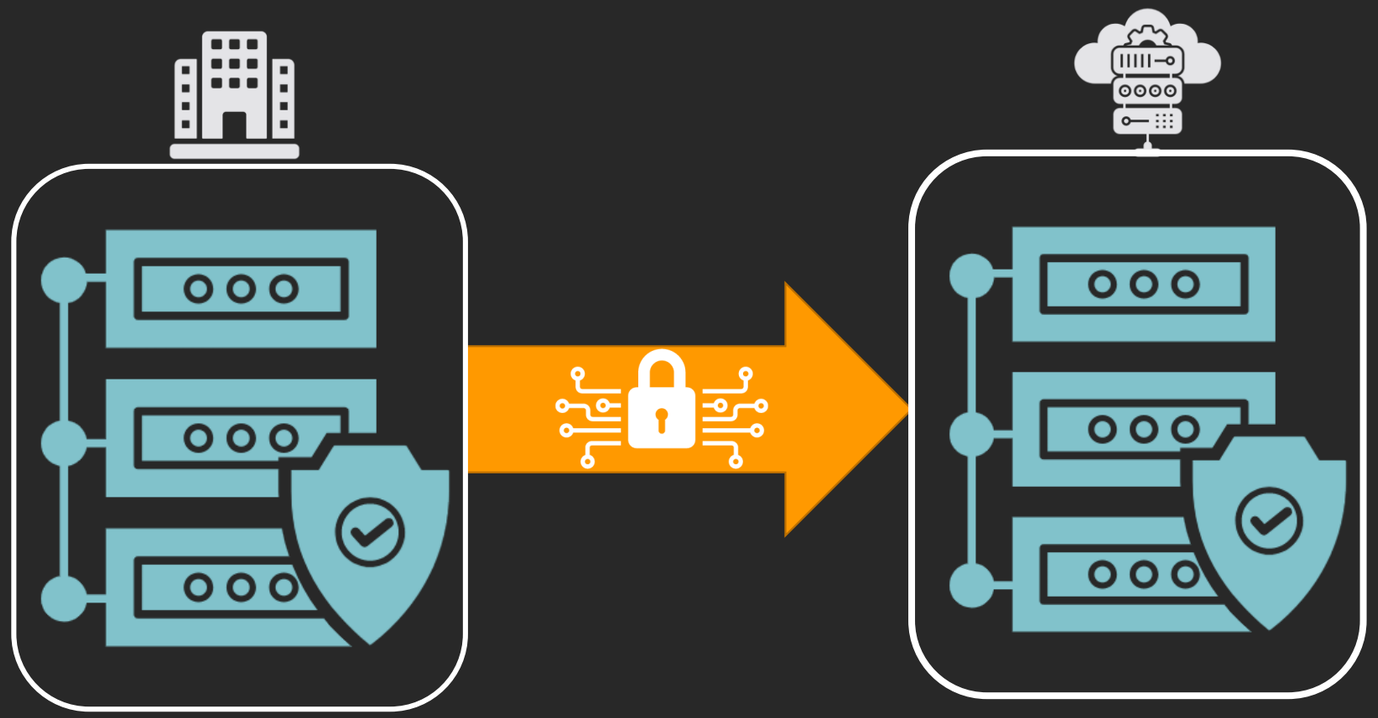
Review and assess system documentation against requirements

Certification Authority assesses system to process PROTECTED data

The business problem

…moving whilst maintaining compliance

- Significant time, effort, and money has been spent meeting compliance obligations in existing datacentres
- Existing staff may not be trained in cloud specific governance
- Business needs to keep running with minimal/no downtime
- Data security and compliance needs to be maintained at all stages



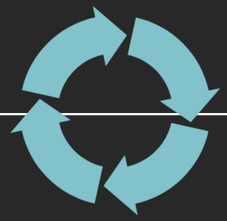
What are IRAP assessors looking for?

CLOUDTEN

Policy

What do I need to meet on ISM?
- ✓ Encryption in transit
- ✓ Backup and restore
- ✓ etc…

AWS
- ✓ Region lockdown
- ✓ Data security
- ✓ IRAP services lockdown



Process

How will I meet the policy ?
- ✓ Cloud native services
- ✓ Commercial tools
- ✓ Custom configurations

Do I need to meet the policy?
- ✓ Is the control in scope?
- ✓ If not why not?

Evidence

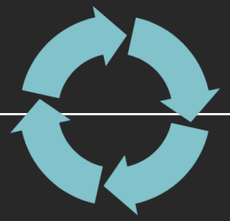
Proof of compliance
- ✓ Monitoring
- ✓ Alerting
- ✓ Reporting

Exceptions
- ✓ Risk mitigation
- ✓ Risk acceptance

Example:

Policy

Security Control: 0459; Revision: 3*; Updated: Sep-18; Applicability: O, P Encryption software used for data at rest implements full disk encryption, or partial encryption where access controls will only allow writing to the encrypted partition.*



Process

Use AWS KMS with AES-256 encryption to secure all block and object stores.

Enforce by AWS Organizations Security Control Policy (SCP).

Evidence / Compliance

Provide historical and real time reports from AWS Config Rules to prove compliance and alert on any non-conformity.

Use Palo Alto Prisma for alerting, monitoring, and reporting.

The PROTECTED landscape

CLOUDTEN

Secure Foundations

Secure Network Connectivity
- ✓ Private
- ✓ Redundant
- ✓ Encrypted

Establish Consistent Baseline
- ✓ Account Governance
- ✓ Centralised Auditing
- ✓ Shared Services
- ✓ Enforceable Guardrails

Templated Deployment

Fully Automated Delivery
- ✓ Repeatable
- ✓ Consistent
- ✓ Immutable

Auditable Configuration
- ✓ Automatic Registration
- ✓ Out-of-the-box Compliant
- ✓ Guardrail Reporting

Integrations & Extensions

Integrate to Existing Services
- ✓ IDAM
- ✓ SIEM
- ✓ ITSM

Extending the Gateway
- ✓ NextGen FW
- ✓ IPS
- ✓ WAF
- ✓ Web Proxy

Continuous Compliance

Posture Management
- ✓ Automated Inventory
- ✓ Compliance Monitoring
- ✓ Data Security

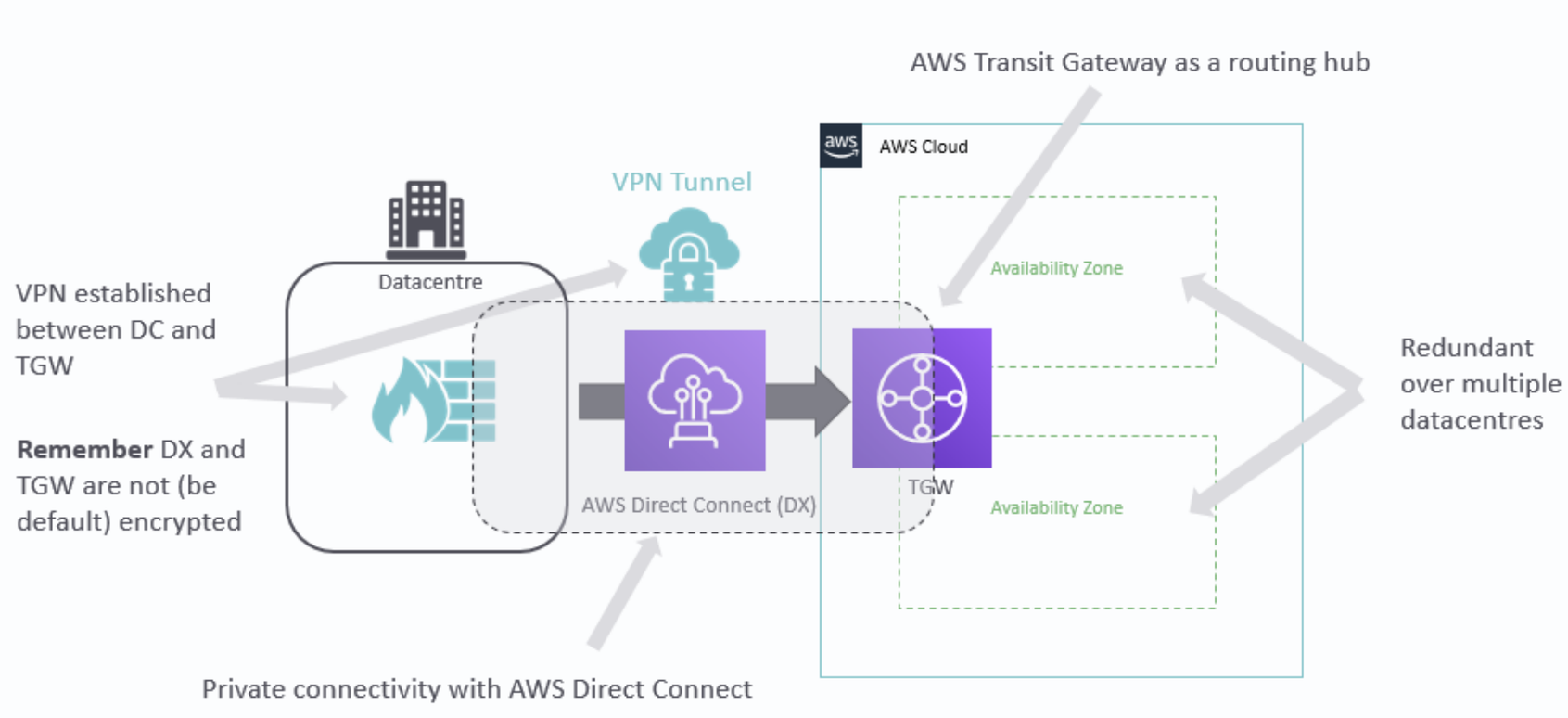
Workload Protection
- ✓ Vulnerability Management
- ✓ Workload Security
- ✓ Activity Monitoring
- ✓ IAM Governance

Secure network foundations

Secure Foundations

Secure Connectivity
- ✓ **Private**
- ✓ **Redundant**
- ✓ **Encrypted**

Establish Consistent Baseline
- ✓ Account Governance
- ✓ Centralised Auditing
- ✓ Shared Services
- ✓ Enforceable Guardrails



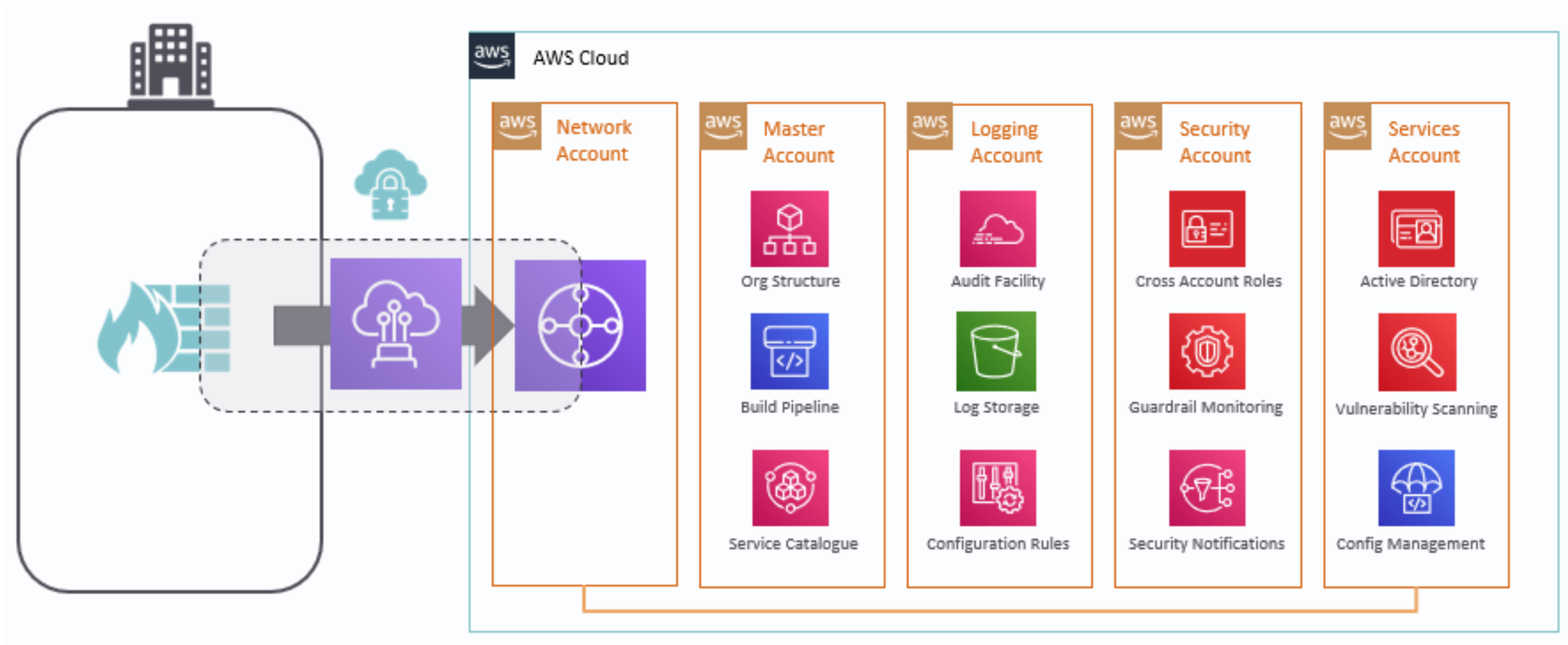
Establishing a baseline (landing zone)

CLOUDTEN

Templated Deployment

Fully Automated Delivery
- ✓ **Repeatable**
- ✓ **Consistent**
- ✓ **Immutable**

Auditable Configuration
- ✓ Automatic Registration
- ✓ **Out-of-the-box Compliant**
- ✓ Guardrail Reporting



* New accounts are rolled out via an automated vending machine process

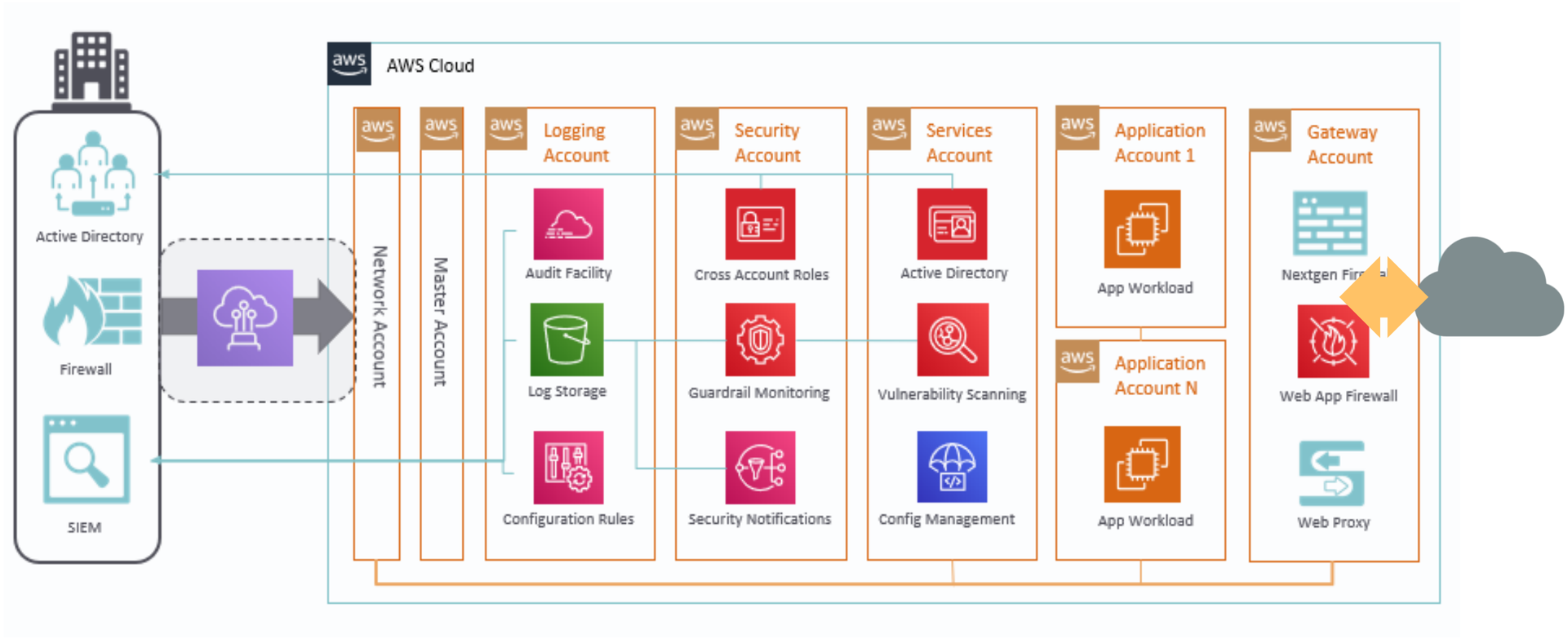
Integrations & extensions

CLOUDTEN

Integrations & Extensions

Integrate to Existing Services
- ✓ IDAM
- ✓ SIEM
- ✓ ITSM

Extending the Gateway
- ✓ NextGen FW
- ✓ IPS
- ✓ WAF
- ✓ Web Proxy



PROTECTED data lake in AWS

CLOUDTEN

Australian Government

aws

What is a data lake?

CLOUDTEN

A centralized repository of business data (structured, unstructured, or semi-structured)

(In this case) Cloud-hosted (AWS) system of data, tools, and processes that act together to provide a powerful analytics platform for IT and business users

Data is collated from various sources (internal & external) and processed. Processed data products are then made available to the business for any data-led initiatives or decisions

Offers virtually unlimited data storage and compute power

Data ingestion and processing pipeline, governance, connectivity, security, analytics and visualisation features – all in one place

CLOUDTEN

Structured Data

- Data that is rigidly formatted with fixed length fields and types
- Can generally directly queried by SQL
- Examples include database systems (Oracle, SQL Server, MySQL etc)

Semi-Structured Data

- Doesn’t reside in a database but does have some formatting properties that make it easier to analyse
- Can’t usually be directly queried by SQL
- Examples include file types such as JSON, XML and CSV

Non-Structured Data

- Data that has no defined format and can not easily be analysed by a machine
- Needs to be converted before it can queried
- Examples include PDF documents, audio and video files

Data lake overview

CLOUDTEN

Advanced Analytics for Secure Data Insights

Multiple Data Formats

- ✓ **Structured (Oracle, SQL)**
- ✓ **Semi-Structured (XML,CSV)**
- ✓ **Non-Structured (PDF,MP3)**

Diverse Datasources

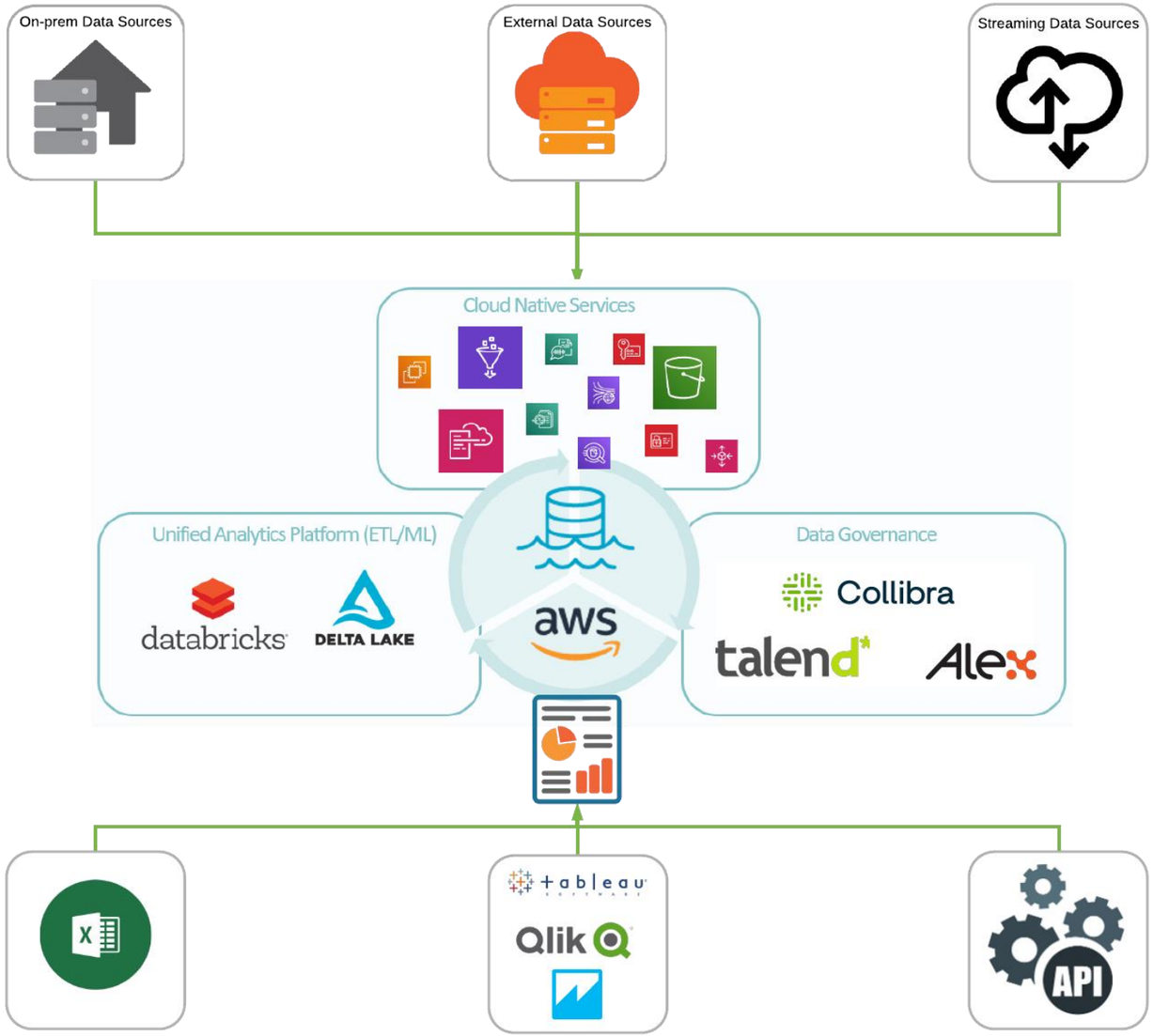
- ✓ **On-premises (DB, Sharepoint)**
- ✓ **External (SFTP, Subscriptions)**
- ✓ **Streaming Data (RSS, Twitter)**

Predictive Analytics

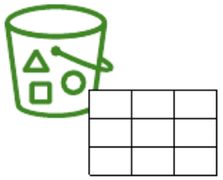
- ✓ **Advanced ETL**
- ✓ **Machine Learning**
- ✓ **Data Science Lab**

IRAP Certified

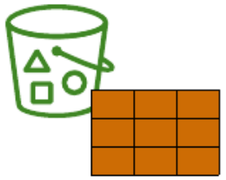
- ✓ **End-to-End Governance**
- ✓ **CCSA Encryption**
- ✓ **Independently Audited**



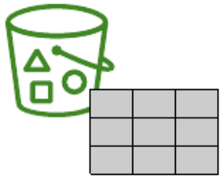
Data buckets



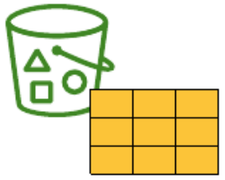
Raw is the Amazon S3 bucket where data is first stored after ingestion. In Databricks, this bucket is represented like a relational table, called the raw table.



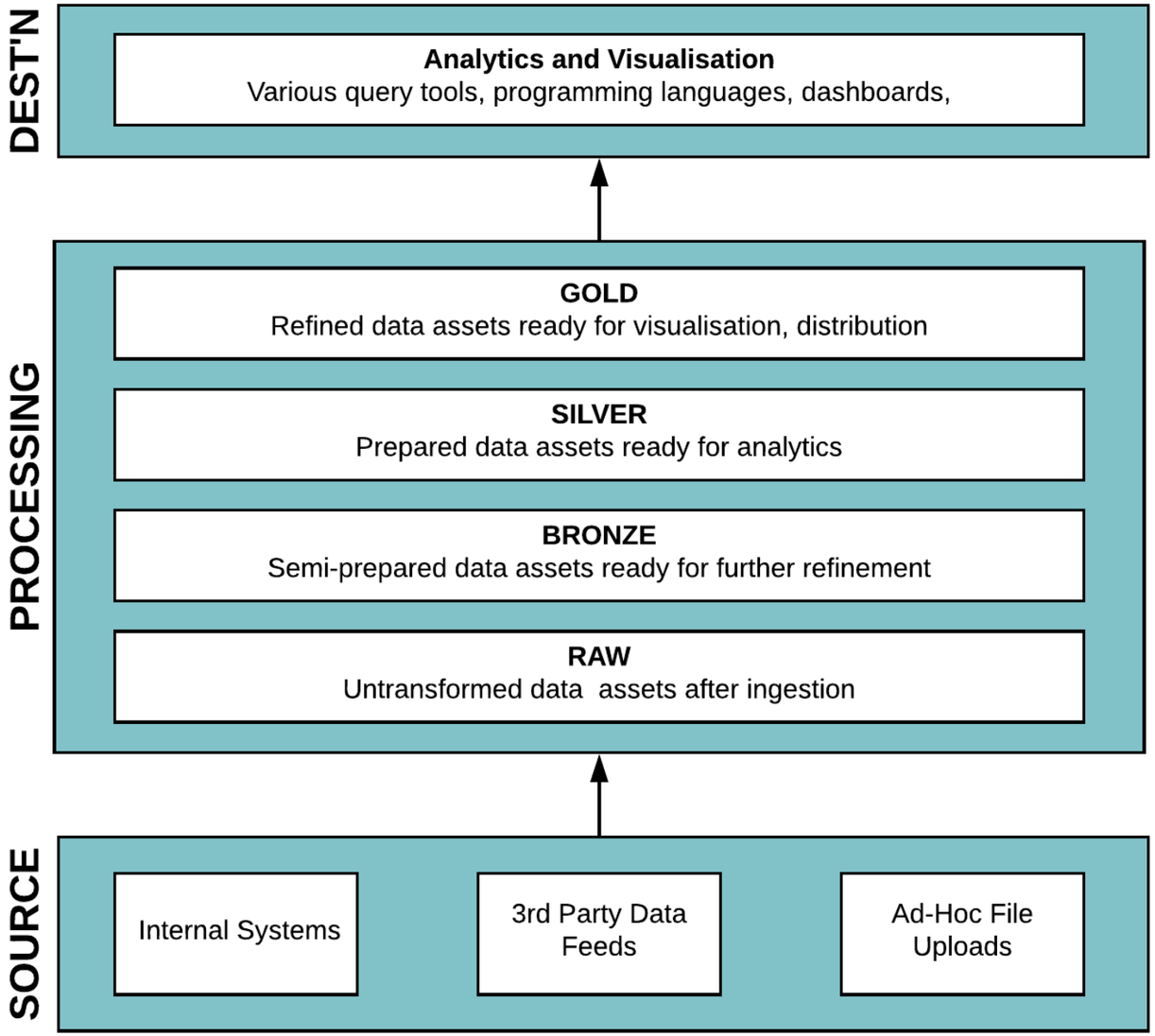
Bronze is the Amazon S3 bucket where the semi-processed raw data is stored. In Databricks, this bucket is represented like a relational table, called the bronze table.



Silver is the Amazon S3 bucket where the processed data is stored, ready for consumption by other systems.

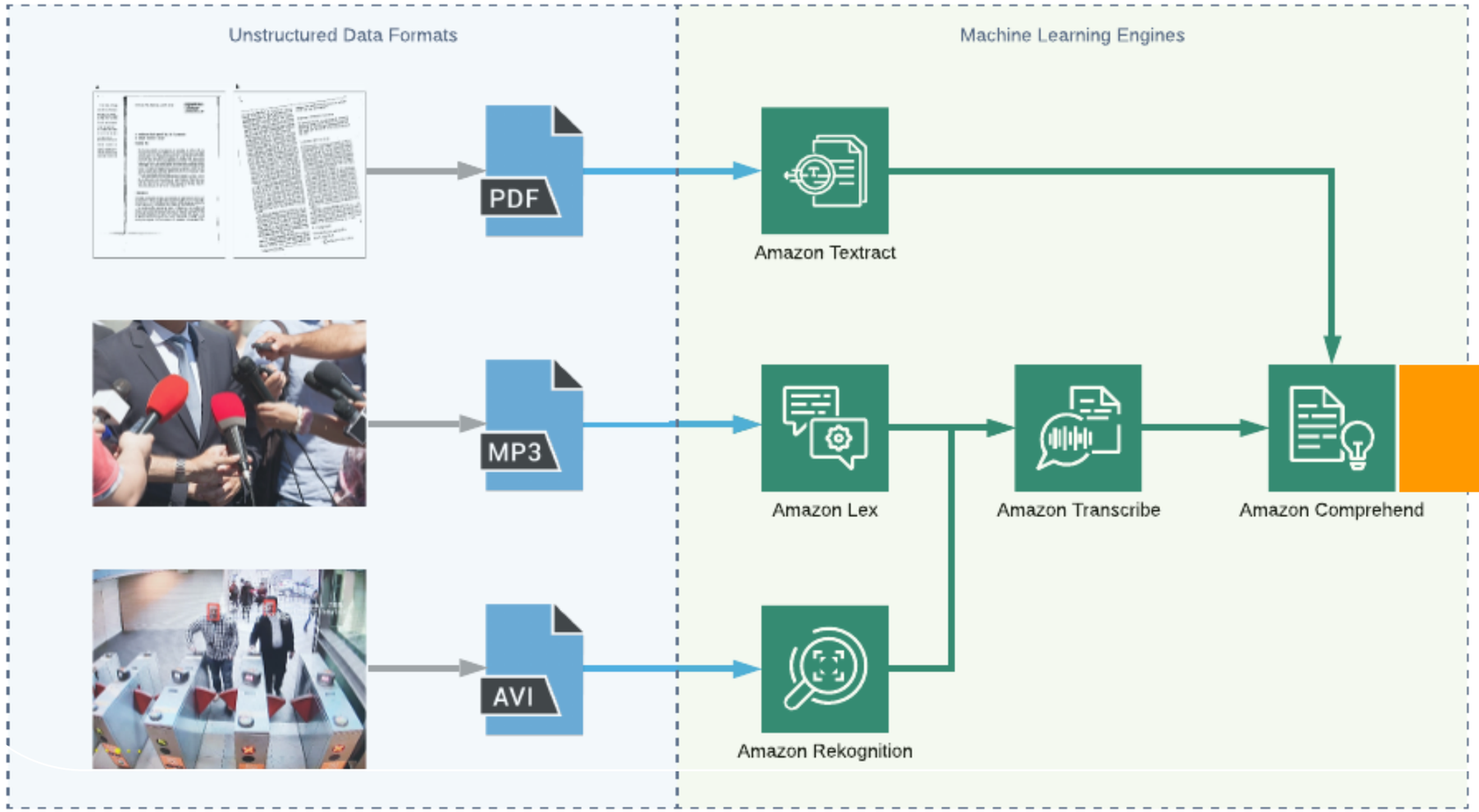


Gold is the Amazon S3 bucket where the ultimate refined data is stored, ready for consumption by other systems.



Roadmap

CLOUDTEN



databricks

aws

Thank you!

Richard Tomkinson

aws PUBLIC SECTOR SUMMIT ONLINE