LeMans - IT Solution Design - Release 3 - Conditional Approved

Owner	@Hari Lakshmip @shweta dixit (I
Reviewer	@Amrita Mukhe @Ubeyde Omer @Victoria Martir @Keith Sinclair
Status	CONDITIONAL APPRO
Creation Date	26 Jul 2024
Conditionally Approval Date	31 Jan 2025
Conditions to fully approved	SNC Completion f [EDMLLEM-6489] Conditions from S Conditions from D Conditions from R Conditions from G
JIRA Ticket	EDMLLEM-6 retrieve issue permission o authenticate EDMLLEM-6 retrieve issue permission o authenticate ITSD Sign Off

Version History

Version History

Author	Version	Date	Reason for Change
@Pinky Jain (Unlicensed) @Anil Gogia (Unlicensed)	0.1	26 Jul 2024	• Initial draft

Author	Version	Date	Reason for Change
@Pinky Jain (Unlicensed) @Anil Gogia (Unlicensed)	1.0	18 Sep 2024	• Conditionally Approved (See Comment)
@Kristine Tuyo (Deactivated)	1.1	20 Nov 2024	• Updated SNC references • Replaced SNC Reference SNC-2023-3901 to New SNC SNC-2024-6114
@shweta dixit (Deactivated) @Hari Lakshmipathi (Deactivated)	1.2	31 Jan 2025	• Updated ITSD in line with the feedback from Risk • Low level designs and implementation documents are linked • Updated the references to outdated and unused patterns • Added latest ARC submission of Filestore as storage solution • Added new entries in RAID, Assumptions and Decision log

Approvals & Distribution List

Approvals

Name	Role	Approved	Email Approval (Y/N)
Alastair Frame	LeMans Product Owner	Conditional Approved	See comment
Ubeyde Omer	DML Le Mans Engineering Lead	Conditional Approved	See comment
Amrita Mukherjee	Google Lead	Approved	See comment
Keith Sinclair	Security Consultant	Conditional Approved	See comment

Please read the comments on the conditional approval provided.

ITSD Summary

Click here to expand...

Requirement	IT Solution Design for deployment of LeMans SAS environment into Google Cloud Platform along with the feasible compute (SAS 9.4M8) and storage services(Filestore). It will also reflect on the migration of LeMans data, processes and users to GCP.
Scope	This document is intended to describe the deployment of the LeMans SAS environment into Google Cloud Platform. It will also highlight the associated activities required to satisfy the Banks' governance and approvals processes outlined by Cloud Services. The following solution areas will be covered in detail with design patterns. 1. Solution Design Target (GCP) Architecture 2. Infrastructure Architecture 3. Network Architecture 4. Environment Sizing 5. SAS Architecture 6. Data Architecture 7. Security Architecture

<table>
<tr><td>Requirement</td><td>IT Solution Design for deployment of LeMans SAS environment into Google Cloud Platform along with the feasible compute (SAS 9.4M8) and storage services(Filestore). It will also reflect on the migration of LeMans data, processes and users to GCP.</td></tr>
<tr><td></td><td>8. Resiliency
9. Logging, Monitoring and Alerting
10. FinOps

Each pattern will define the solution and the required components, along with the network and security configuration necessary to secure the application and content within.</td></tr>
<tr><td>Out of Scope</td><td>1. Data Migration for LeMans (This will be scope of migration team).
2. Transformation of existing processes.
3. Full adoption of SAS Viya.
4. Decommissioning of On-Premises LeMans solution. This will be covered in future releases.
5. Migration of Tape Backups from On-Premises to GCP
6. Test data creation and strategy.</td></tr>
<tr><td>Owner</td><td>@Pinky Jain (Unlicensed) @Anil Gogia (Unlicensed) @shweta dixit (Deactivated) @Hari Lakshmipathi (Deactivated)</td></tr>
<tr><td>Lifecycle status</td><td>Draft</td></tr>
<tr><td>Associated JIRA Ticket</td><td>[EDMLLEM-5850] ITSD Draft - Lloyds Banking Group Jira
[EDMLLEM-6552] ITSD Fully Approved - Lloyds Banking Group JIRA</td></tr>
<tr><td>Approved Decisions from DWG and ARC</td><td>Approvals from DWG on 29 June 2023 recorded in the Confluence page below for previous releases

DWG 2023 Inputs and Outcome - Data Migrations - Lloyds Banking Group Confluence

DD190 - LeMans Migration - R2 - Option consideration for Data Storage - Data Migrations - Lloyds Banking Group Confluence

ARC Submission - 16/10/2024</td></tr>
<tr><td>Approved Patterns</td><td>Approved Platform Patterns
<table>
<tr><th>Item</th><th>Pattern description</th><th>Pattern</th></tr>
<tr><td>1</td><td>ITSD for deployment of SAS 94 on GCP</td><td>LeMans - IT Solution Design - SAS 94</td></tr>
<tr><td>2</td><td>ITSD for deployment of Filestore</td><td>LeMans - IT Solution Design - Release 2</td></tr>
<tr><td>3</td><td>SAS Installation</td><td>LeMans SAS Installation</td></tr>
<tr><td>4</td><td>On-Premises Alignment</td><td>LeMans Custom Configuration Changes</td></tr>
<tr><td>5</td><td>LeMans Scaling</td><td>LeMans SAS Scaling</td></tr>
<tr><td>6</td><td>SAS 9.4 Environment Sizing</td><td>SAS 9.4 - Environment Sizing</td></tr>
<tr><td>7</td><td>LeMans SAS Storage</td><td>LeMans SAS Storage</td></tr>
<tr><td>9</td><td>LeMans Least Privileged Access</td><td>LeMans LPA Design</td></tr>
<tr><td>10</td><td>User Access and Authentication</td><td>LeMans User Access Management</td></tr>
<tr><td>11</td><td>LeMans Connect: Direct Pattern</td><td>LeMans Connect: Direct Pattern</td></tr>
<tr><td>12</td><td>LeMans Security Design</td><td>AA Endorsed: LeMans Security Design - Release 3</td></tr>
<tr><td>13</td><td>LeMans Network Design</td><td>LeMans - Network Design</td></tr>
<tr><td>14</td><td>LeMans Operating Model</td><td>LeMans Operating Model</td></tr>
<tr><td>15</td><td>LeMans On-prem to AP BigQuery Connectivity</td><td>POC - Le Mans On Prem to EDH GCP Connectivity</td></tr>
<tr><td>16</td><td>LeMans DCX to AP BigQuery Connectivity</td><td>SAS Design: SAS/Access to Google BigQuery</td></tr>
<tr><td>17</td><td>Filestore Backup and Recovery Options</td><td>LeMans Filestore Backup and Recovery - Options</td></tr>
</table></td></tr>
<tr><td>In Flight Patterns</td><td>LeMans Filestore Logging and Monitoring - LeMans Logging and Monitoring</td></tr>
<tr><td>RAID</td><td>Refer the RAID section for details on Risks, Assumptions, Issues and Dependencies. Project RAID log is available in JIRA.</td></tr>
<tr><td>Tech Debt Acknowledged/</td><td>Le Mans Tech Debt Registry - Data Migrations - Lloyds Banking Group Confluence</td></tr>
</table>

Requirement	IT Solution Design for deployment of LeMans SAS environment into Google Cloud Platform along with the feasible compute (SAS 9.4M8) and storage services(Filestore). It will also reflect on the migration of LeMans data, processes and users to GCP.
Recommendation	
FinOps	LeMans FinOps - Data Migrations - Lloyds Banking Group Confluence

1. Project Overview

The project objective is to migrate the current LeMans data, processes and users to Google Cloud Platform, hosted in the LBG tenant under Enterprise Data.

LeMans is an Advanced Analytics and Modelling SAS Platform initially stood up in 2009. Built around a SAS Grid Architecture, SAS Metadata Server, with Regulatory and Financial reporting capabilities for the Bank. LeMans has continued to grow since it was installed and is currently around:

- 24 Grid nodes
- 400+TB of data (SAS SPDS and SAS Datasets)
- 1000+ Users
- 100+ data sources

The platform supports both ad-hoc user and controlled batch processes, both managed by the Retail Risk AESM business support team. LeMans supports several regulatory and reporting processes critical to the Banks' industry compliance. Currently, the managed data for the platform is sourced through golden source systems and internal databases (EDH. GDW, etc.), but the business also has the capability to source ad-hoc data into the platform. This is with the understanding that this ad-hoc data is moved into a managed IT controlled process to support consistency, resilience, latency and reconciliation of data where it is used on a regulatory or permanent basis.

LeMans currently faces a series of risks which Lloyds want to address, key being the one related to the server infrastructure which is outdated and currently on a rolling 3-month IBM support extension from April 2021 until December 2023.Therefore, to mitigate the risk, the strategic choice and approved approach are to migrate this to Google Cloud and this has been aligned to other Verticals under this Enterprise Data programme.

Key risks being addressed by migrations:

SNOW References	Description	Mitigation	Due Date
RK0026440 - Retired and combined to RK0042056	**Le Mans sits on IBM Power 6 machines which cannot be upgraded and IBM recommend moving to Power 10s, extended support confirmed to 31/12/23. SAS is on AIX 7.1 which cannot be used on Power 10s. Also uses IBM Spectrum Scale 4.2.3** There is a risk that infrastructure, applications/software, and associated assets are not supported or adequately maintained to guard against IT service or security issues in CIO Enterprise Risk due to the existing legacy IT estate and applications / software which is no longer within IT Support.	There is an extension expected for support of AIX, the Platform are analysing the processes which may depend on backed up data that would require the AIX machines to exist in order to land the backed-up data and onward migration into GCP. On this basis we request acceptance to 31-03-25 whilst this analysis can complete, and a plan put in place. Approval for acceptance and extension of end date recieved from Joe Soule through offline risk pack.	Acceptance End Date - 31-03-2025
RK0039102	**Le Mans Backup is reaching capacity, the current solution is out of support and can not be extended leading to a risk that without action being taken the application may not be able to complete its backups.** There is a risk that IT services and applications are impacted as a result of insufficient disaster recovery and backup & restore capability, especially relating to application and configuration data for key applications in the production environment. This could lead to extended service impact and/or loss of data which may result in reduced customer confidence, adverse media coverage and regulatory interest and fines.	Extension of TRD was approved at May 24 CIO Risk Committee with the given rationale: The LeMans back up has capacity to October 2024. The delay in the cutover of users to the 'Lift and Shift' of LeMans is placing this capacity at increased risk. The Risk Foundations Data Lab have transferred data to the DCX platform which would enable the purging of data on the backup to increase this capacity, however an incident related to purging prevents this from being possible at present. The tapes, drives and retrieval technology will be required until the data files have been transferred to DCX. The duration of this will be known once users have transferred to the DCX hosted LeMans application. The end date of the risk will be reviewed at the end of Q1 2025 where transfer rates will be proven, enabling an accurate view of closure dates.	30-09-2025 14:23:08

RK0047995	**CURRENCY : Red Hat Linux 6.9 is used by LeMans. The OS became obsolete in November '20, the bank has secured extended support until June 24 with no patches or support available after that time.** There is a risk that infrastructure, applications/software, and associated assets are not supported or adequately maintained to guard against IT service or security issues in CIO Enterprise Risk due to the existing legacy IT estate and applications / software which is no longer within IT Support.	Mitigate Risk of CURRENCY : Red Hat Linux 6.9 is used by LeMans. The OS became obsolete in November '20, the bank has secured extended support until June 24 with no patches or support available after that time.	25-11-2025 12:41:31

The Bank has committed to migrating the analytics platforms to Google Cloud Platform, this began this in 2021 with EDH and GDW. In 2022, this scope has increased to include LeMans as part of the wider programme, benefiting from the foundations laid by the proving completed and the horizontal teams to provide platform, business engagement and risk & regulatory support.

2. Requirements, Architecture Decisions, Assumptions, Dependencies, Future Roadmap, and Risks

2.1 Requirements

Below is list of high level requirements from requirement traceability matrix:

Click here to expand...

Requirement type	Reference	High level requirement	High level requirement description	Priority (MSCW)	Objective ID	Objective description	CSF ID	CSF description	Patterns linked	Scc this doc
Data	HLR.DR.001	Data Movement	The solution must migrate all current LeMans data via approved transfer routes to GCP.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.	LeMans Migration Approach - Data Migrations - Lloyds Banking Group Confluence	No- will of s for solu and of mig pro for
Data	HLR.DR.002	Data Quality & Reconciliation	The solution must replicate the current Data Quality and reconciliate processes running on LeMans to prove functionality.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.	LeMans Flow Analysis and Testing Process	No- will of s for solu and of mig pro for LeM
Data	HLR.DR.003	Use Case Movement	The solution must migrate use cases (current data and historical data) from LeMans on-premises to GCP.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.	LeMans Use Cases Mapping - Data Migrations - Lloyds Banking Group Confluence	No- will of s for solu and of mig pro for LeM
Data	HLR.DR.004	Data Accuracy	The solution must ensure that migrated data is accurately represented in the target system.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.	LeMans Migration Approach - Data Migrations - Lloyds Banking Group Confluence	No- will of s for solu and of mig pro for LeM
Functional	HLR.FR.001	User Access	The project team must retain the existing functionality of providing specific user access to all	M	OBJ.02	Strategic alignment	CSF.06	Consumers migrated to GCP.	LeMans User Access Management - Data Migrations - Lloyds Banking Group Confluence	Yes

			LeMans users and service accounts both at platform level as well as folder (Teams and MDL) level.							
Functional	HLR.FR.002	Batch data scheduling and ingestion	The solution must allow running and monitoring of jobs via LSF (scheduling tool) and must retain the ability for users to schedule jobs using in-house scheduling tools run their jobs via teams scheduler and night scheduler.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.	LeMans Server Connections - Batch Server - Data Migrations - Lloyds Banking Group Confluence	No- will of s for solu and of mig pro for LeM
Functional	HLR.FR.003	Housekeeping	The solution must migrate the processes for archival of unused data and also the functionality to restore data from backups..	M	OBJ.02	Strategic alignment	CSF.01	GCP operational model agreed and in place.	LeMans Filestore LeMans Backup, Archival & Recovery - Options Paper Filestore Backup, Snapshots and Archival Design - Data Migrations - Lloyds Banking Group Confluence	Yes
Functional	HLR.FR.004	Storage	The solution must allow for store existing data and also be able to expand the storage to meet the business needs in future.	M	OBJ.02	Strategic alignment	CSF.01	GCP operational model agreed and in place.	LeMans SAS Storage - Data Migrations - Lloyds Banking Group Confluence	Yes
Functional	HLR.FR.005	User Roles	The solution must implement effective password management policy for all user service accounts and SSO for all user accounts, and service accounts and retain existing roles, ACL and policies.	M	OBJ.02	Strategic alignment	CSF.06	Consumers migrated to GCP.	LeMans User Access Management - Data Migrations - Lloyds Banking Group Confluence	Yes
Functional	HLR.FR.006	Performance Monitor & Alerts	The solution must retain capabilities to monitor, control, and investigate the jobs.	M	OBJ.02	Strategic alignment	CSF.02	Performance should be same or better compared to on-premises LeMans solution.	Lemans Monitoring Web Portal - Data Migrations - Lloyds Banking Group Confluence Lemans-Dude Web Portal - Data Migrations - Lloyds	No- will of s for solu and of mig pro for LeM

									Banking Group Confluence	
Functional	HLR.FR.007	Scheduling and orchestration of data pipelines	The system must enable scheduling and orchestration of data ingestion processes and extraction.	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.		No- will of s for solu and of mig pro for LeM
Functional	HLR.FR.008	Analytical / Reporting toolset connectivity	The solution must allow reporting tool - Model Maker(python tool) to connect directly to LeMans	M	OBJ.02	Strategic alignment	CSF.05	LeMans data (data sources including historic data) and use-cases migrated to GCP.		No- will of s for solu and of mig pro for LeM
Non-Functional	HLR.NFR.001	Reliability	The solution must implement AESM best practices the existing backup policies and should be able to retrieve data from backups in case of any disaster/failure. It must be able to spin up SAS compute nodes to maintain workloads.	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	LeMans Filestore LeMans Backup, Archival & Recovery Design Filestore Backup, Snapshots and Archival Design - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.002	Availability	The project team must ensure operational processes and procedures are in place to meet existing levels of service availability and in case of DR event. The solution also must ensure disaster recovery meets BIA standards.	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	LeMans Filestore LeMans Backup, Archival & Recovery Design Filestore Backup, Snapshots and Archival Design - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.003	Extensibility	The solution must retain the ability to connect to third party databases and external systems to transfer data between them.	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	CCOE-GCPAD-008 - LeMans Data Transfer Options - Cloud Platform Architecture Governance - Lloyds Banking Group Confluence LeMans - External Connections - Data Migrations - Lloyds Banking Group Confluence SAS Design:	Yes

									SAS/Access to Google BigQuery - Data Migrations - Lloyds Banking Group Confluence DTA - Data Transfer from On-Prem - Data Migrations - Lloyds Banking Group Confluence	
Non-Functional	HLR.NFR.004	Performance	The solution must provide performance equal to or greater than the current system capabilities	M	OBJ.03	Achieve scalability	CSF.02	Performance should be same or better compared to on-premises LeMans solution.	Performance Test Plan - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.005	Security	The system must ensure prevention of data change, destruction, or loss in an unauthorised or accidental manner to comply with regulatory obligations	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	AA Review: LeMans Security Design - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.006	Security	The solution must provide secure transfer of information between different systems and ensure relevant encryption for sensitive data	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	AA Review: LeMans Security Design - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.007	Maintainability	The solution must allow AESM team to perform critical maintenance actions, including make sure renewal of certificates on time, backup recovery are recovered, hot fixes deployment, and to any outstanding issues, SAS license updates.	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	LeMans Operating Model - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.008	Location	Every user must only be able to access from LBG approved locations and methods of connectivity.	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	LeMans User Access Management - Data Migrations - Lloyds Banking Group Confluence	Yes
Non-Functional	HLR.NFR.009	Compatibility	All current interactions with SAS clients, databases, files, external databases, SAS products, and interaction	M	OBJ.01	Mitigation of currency risk	CSF.01	GCP operational model agreed and in place.	LeMans SAS Installation - Data Migrations - Lloyds Banking Group Confluence	Yes

			to external components etc. must be functional within GCP							
Non-Functional	HLR.NFR.010	Certification and Compliance	The solution must be compliant with LBG best practice, standards and policies.	M	OBJ.02	Strategic alignment	CSF.04	Minimal user disruption and business processes / operations		Yes
User Interface	HLR.UIR.001	User interface	The solution should leverage SSO authentication for users, where possible.	S	OBJ.04	Strategic alignment	CSF.04	Minimal user disruption and business processes / operations	LeMans User Access Management - Data Migrations - Lloyds Banking Group Confluence	Yes
User Interface	HLR.UIR.002	User interface	The project will provide the appropriate SAS tools for use on existing user machines	M	OBJ.05	Strategic alignment	CSF.04	Minimal user disruption and business processes / operations	LeMans User Access Management - Data Migrations - Lloyds Banking Group Confluence	Yes

Full list of requirements from requirement traceability matrix is located in LeMans Requirements Definition and Traceability v1.2.xlsx.

2.2 Architecture Decisions

Below is list of key decisions made for Lemans application design:

› Click here to expand...

ID	Decision Points	Details
AK01	IDE Use	The decision was taken to utilize a GCP tenant for the initial build and development. This is encouraged by the LBG team to avoid recurring provisioning issues for IDE and restricted access via LBG devices. Google provisioned the necessary infrastructure to facilitate build including RHEL VMs and Filestore for storage.
AK02	Capacity	The capacity deployment should at least match that of the current LeMans deployment and should support the objective of improving performance during peak processing periods. Additionally, the capacity needs to be capable of flexing to accommodate changing workload requirements.
AK03	SAS 94 Software Stack	The SAS 94 software stack will remain unchanged from the current setup including SPDS. The latest version of SAS software SAS94 M8 will be deployed. LeMans SAS stack will include SAS9.4 M8, SPDS 5.5, LSF 10.1.0.12, PM 10.2.0.12 and licenses to connect to external databases like Oracle, Teradata, and MS SQL.
AK04	Availability on GCP services	GCP services will provide sufficient availability to meet the service level agreement which is required for SAS environments. The performance testing will be done by LBG, SAS and Google.
AK05	SAS 94 Platform	The target platform for SAS 94 will be Linux [RHEL 8] virtual machines [VM's] running on Google Compute Engine.
AK06	Access to internal and cloud systems.	Windows 10 laptops will be provided to enable access to both internal and cloud systems defined by the Bank. MacBook and Dev VDIs required for configuration infrastructure
AK07	Environment Specification	Each Infrastructure Route-To-Live environment will have only one SAS 94 environment and a Filestore Instance per data storage type(3 instances - MDL, Teams and SAS Binaries) . On the Production Infrastructure, we will be having 4 environments [Dev, CIT, UAT and Production]

ID	Decision Points	Details
AK08	RTO and RPO	The RTO and RPO to recover the individual files from snapshots will be <24 hours. However with zonal instances the BIA requirement of (RTO<24hrs and RPO<24hrs) will not be met.. • **RPO <= 1 week** as we do Filestore backup every week. • **RTO (backup) < 2-3 hrs** • **RTO (restore) > 30 hrs**
AK09	High Availability	LeMans GCP key components Compute Engine and Filestore are configured under one **zone** - currently zone A. Regional Filestore is recommended to achieve HA for data storage but is not opted as a solution from bank for cost incurred.
AK10	Scalability	The Filestore Instances can be scaled up to the limit when there is additional storage required or additional Instances will be deployed to meet the requirements. Compute Engine instances can be scaled up to the subnet mask limit. Subnet mask limit increase will be platform decision if we require to increase subnet range.
AK11	Ephemeral Storage	Persistent Disk will be used as a ephemeral storage for storage required for compute from SAS jobs. Zonal Persistent Disk is used to store compute data for SPDSWORK, SASWORK, SASUTIL, and additional configuration of VM such as logs, tmp storage etc. SAS components such process manager, metadata tier configuration, mid-tier configurations, Connect Direct application and other SAS binaries will be stored in Regional PD.
AK12	Logging, Monitoring and Alerting	Dynatrace will be used as the monitoring and alerting solution. It will raise incidents using ServiceNow in production and notifications in Teams channel in non-prod environments. We will engage with Dynatrace team to build new patterns if required.
AK13	User Authentication and Authorization	User authentication and authorization are a combination of SAS IAM, GCP IAM and SPDS IAM
AK14	File Store User Authentication	Filestore Instances are not fully POSIX compliant. ACLs in current version of Filestore(NFSv3) are not supported so alternative design options will be used to control access to the users. When the new version NFSv4.1 which is still in preview state is available , the ACL design can be reconsidered. Also curation of AD in LBG will be required.
AK15	Data Classification	The data classification of LeMans has been confirmed as "Highly Confidential" under LBGs convention. Filestore does not support encryption for data in transit and hence a separate risk and SNC have been raised to support the usage of HC data within LeMans solution. Data at rest: SNC-2023-4414 Risk Associated - RK0045163 Data in transit : SNC-2024-6114 Risk Associated - RSCA Risk - R0520075
AK16	DevOps Process	CI/CD process has been defined by the Cloud Services involving the tools such as Jenkins, Terraform to deploy the infrastructure for Filestore deployment and migration. Features not supported by product module can be deployed using alternate solutions in agreement with product teams.
AK17	GCP DCX RHEL8 Golden Image	GCP DCX RHEL8 Golden images will be customized to include signed binaries used for LeMans Infrastructure.
AK18	Implement temporary storage for SPDSWORK,	Limitation of CMEK support issue on local SSD. Alternative is to use Persistent Disk and accept a performance drop as a result. It will still be

ID	Decision Points	Details
	SASWORK & SASUTIL	expected to be superior to on-premise performance.
AK29	Update pipeline for SAS 9.4 (later M8)	It's decided that there will not be any upgrade pipeline for SAS 9.4 (later version). In case of any requirement of SAS 9.4 in future that will re-deployed and content migrated.
AK20	9.4 Hot Fix pipeline	The decision has taken that SAS 9.4 HF pipeline will be used for updating the HF SAS 9.4 environment.
AK21	SAS and Filestore deployments on Non-Production Route-To-Live Infrastructure will be short lived.	Only the SAS 94 and Filestore environment deployed on the Production Infrastructure RTL Environment will be available 24/7. All Non-Production environments will be short lived and stood up only for testing configuration changes, hotfixes, etc.
AK22	The current physical security model for Users will be retained as the basis of the future design	Filestore limitations led to 46 service accounts being created. ACL cannot be applied to the Filestore NFSv3. Please refer to limitation of Filestore section for more details.
AK23	UTF-8 encoding will replace WLATIN	The double-byte format will lead to data changes, these will be assessed and validated during the migration of data.
AK24	Migration will be undertaken by data layer (RDL > CDL > FDL) and conclude with user migration prior to go-live	LeMans data is hierarchical and requires a comprehensive foundation before the next release
AK26	Migration of source data feeds will be completed and validated by LBG	Needed to test and schedule ongoing batch processes that are reliant on source data (push & pull feeds)
AK27	Route to Live activity will not migrate mid-way through development lifecycle	Migrating during development compromises testing and increases complexity for little benefit
AK28	Dedicated Transfer rates are sufficient for Active data transfers and ongoing Batch	Slow transfer rates will mean the migration estimates are extended
AK29	Dedicated Interconnect sizing has accounted for LeMans activity and will be increased if necessary	Shared resources need to be sized to account for the monthly transfer of 2-3TB of input data, plus user activity
AK30	Data Transfer Appliance to be used for moving the	10 TA40 devices will be used to move the Historical Data, MDL data as part of Release 2 & 3 data migration activities.

ID	Decision Points	Details
	Historical Data	
AK31	Shared storage solution to hold the MDL and Teams data layer	Decision has been made to use Filestore as the underlying shared storage for storing SAS configurations, SPDS data and User data.
AK33	Shared storage Performance	The performance of Filestore is comparatively lesser than the ExaScaler shared storage solution which was investigated as the first storage solution. However, the performance will be equal or better than the current storage solution in On-Premise LeMans. Please refer to R2.1 - EOTR - Data Migrations - Lloyds Banking Group Confluence for details on the performance testing results.
AK34	Regional Filestore is recommended but not used	Regional Filestore was proposed and endorsed as the most resilient solution. However, this option is not a viable in view of the cost involved, and Bank has decided to continue using Zonal Filestore.

2.3 Assumptions

ID	Assumption	Details
AA01	Embedded Process	Embedded Process will be deployed by Teradata to Vantage once supplied by SAS, the TD TTU client will enable the connectivity between SAS and Teradata. The SAS EP (and support functions) enable SAS models to be executed inside Teradata. It is assumed this will be completed by Teradata Vantage and tested during the proof points for GDW.
AA02	Latency concern on access cloud v/s On-prem	Data transfer delay between cloud and on-prem using Interconnect. As the amount of data is very small, it assumed that it's not a potential risk going forward, and the data transfer is not actively used across two locations.
AA03	On-demand vs. reserved pricing	3-year reservation will be assumed for all the environments running on Production Infrastructure. For non-production the on-demand charge rates have been assumed on the basis that these are intended to be short-lived.
AA04	Cron tab is used for scheduling instead of TWS	Crontab is used for scheduling instead of TWS in LeMans . This is noted as a tech debt for the programme.

2.4 Dependencies

ID	Dependency	Details
AD01	Persistent Disks needs to be hardened to support Confidential and Highly Confidential data when associated with GKE and GCE instances.	PDs are required to support the storage of permanent and temporary data by the SAS applications, they need to be hardened to support the data being held on the FS.

ID	Dependency	Details
AD02	Filestore Instances to be hardened to support Confidential and Highly Confidential data	Filestore Instances needs to be hardened to store the HC data by Cloud Services and Tech Optimization.
AD03	Existing LBG data sources will be accessible by SAS from GCP	To support testing and parallel run, current internal data transfers will be re-directed by LBG to GCP staging area
AD04	CODA design and approval	Completion of the Security Design [CODA]
AD05	IMCRA	Completion of the Internally Managed Cloud Risk Assessment
AD06	Network Design	There is dependency on Network Design for LeMans system. Please refer to LeMans - Network Design - R2
AD06	Highly Confidential Data Encryption	Pattern for storing Highly Confidential data in LeMans is required before migrating it to GCP. Required SNCs should be in place to support HC data.
AD07	SOX	SOX compliance to be completed and signed off before the migration of data to GCP. Please refer to Classic | Unified Navigation App | ServiceNow (service-now.com) for complete list of SI tasks and E2E required activities.

2.5 Future Roadmap

ID	Recommendation	Details
AFR01	Use Enterprise scheduling solution TWS	LeMans should use the enterprise solution for scheduling the workloads on compute engines, TWS. Using crontab as solution is an interim state and should be replaced with TWS in due course.
AFR02	Using Regional Filestore for achieving high resiliency requirements	Regional Filestore is proposed and endorsed as the most resilient solution to achieve the required RPO and RTO of <24 hours for LeMans. Bank's has decided to delay implementation of regional Filestore in view of cost and continue to use zonal Filestore.
AFR03	Adoption of latest pattern used in CNE to request data from AP(EDH) consumption layer in LeMans	TO is not part of the VPC-SC perimeter bridge created in CNE to simplify access to AP data as part of SFP-22 | 2024-02-08 | Bridge adoption for cross perimeter access. This pattern is an interim state in the path of collapsing value stream based VPC-SC within the bank. LeMans platform should use these patterns as consumer of EDH data.
AFR04	Design a pattern for GCP inter-project data transfer	Currently project uses two patterns to transfer data across environments, one using SAS connect and another using GCS. We should review the requirements and approve the pattern for Migration and BAU usage.

2.6 Risks

ID	Risk	Impact	Mitigation
AR01	Delays to other Horizontal delivery activities could adversely affect LeMans migration	Delays or de-prioritization of shared components may affect the delivery timescales.	Confirm the estimates with the Platform Horizontal for the delivery of the GCP components
AR02	Slow performance of Dedicated Interconnect from on-premises to GCP will affect data transfers for migration and BAU	If this is less than 50MB/s then the durations estimated will increase	Knowing this as soon as possible allows the plan to be updated accordingly
AR03	Latency concern on access cloud v/s On-premises	Data transfer delay between cloud and on-premises	This is an interim measure in the event of a hybrid solution. In the end state both LeMans and GDW should be co-located in GCP therefore latency should be reduced.
AR04	No Regional Failover	Risk of missing RTO/RPO deadline if there is any regional outage.	There is no requirement for regional failover in the bank for LeMans and the risk is that in the event of a region outage there would be a total loss for LeMans and could only be restored once the region was recovered.
AR05	Storage of Highly Confidential data (Risk Ref: I&S20211021.01)	Users and Batch processes will not able to process the HC data on LeMans	Overarching SNC to handle the HC data at rest for LeMans is raised and is applicable till July 2025. SNC-2023-4414- SNC Submissions App - Power Apps Risk Associated - RK0045163 | Risk | ServiceNow (service-now.com) For lack of encryption for data in transit, SNC SNC-2024-6114 Please refer to AA Review: LeMans Security Design - Data Migrations - Lloyds Banking Group Confluence for complete list of all SNC and risks
AR06	End of Technology Solutions (Risk Ref: R025729)	Currency Risk due to both Operating System and Server Hardware have already reached end of life. Currently running on extended support which will stop Dec 2023	Migrating LeMans solution to GCP platform utilizes the latest hardware and RHEL 8 operating system which is cloud ready with a long lifecycle will mitigate the issue.
AR07	Technical Security (Risk Ref: R046459)	Number of security issues identified on LeMans solution where potential attacks and unauthorized attacks can provide users access to the system which could lead to data damage or theft	Data migrated to GCP will be stored in CMEK encrypted disks which has strong encryption and support for TLS 1.2 protocol for data in motion. Please refer to AA Review: LeMans Security Design - Data Migrations - Lloyds Banking Group Confluence for complete list of all SNC and risks
AR08	User Account Management Risk (Risk Ref: R042023)	Risk Value Stream applications utilise static accounts with weak passwords which is against policy and exposes the Bank to security risk if the password is broken (Martini Risk)	GCP accounts will be defined in Active directory, not locally on the servers and secured in CyberArk safes in alignment with the current Bank standards. One SAS account requires a static password to be retained for the execution of SAS services. This account will still be held in CyberArk – This is a software/application constraint.
AR09	Storage Risk (Risk Ref: R025688)	Le Mans Storage (MDL and Teams folders) and Backup is reaching capacity, the current solution is out of support and can not be extended leading to a risk that without action being taken the application may not be able to complete its backups.	By moving to GCP the technical constraints on expanding team folders and creating new team folders will be removed. There will though be cost constraints on additional storage that need to be determined. Post migration all future backups will be held under a new GCP solution (still to be approved). This will remove future risk of breaching capacity. However all data held on tape at the time of migration will continue to be held under existing policies.

ID	Risk	Impact	Mitigation
AR10	Recovery Time Objective	Le Mans platform is so large that it is impossible to fully re-build the system from scratch within the 24 hour target within the BIA. Whilst highly unlikely, a full re-build of the platform would take upwards of 30 days in the current On-Premise environment.	By implementing the multi tier backup and DR strategy with zonal instances , there is risk for not achieving the RPO <24 hrs and RTO<24hrs. The ongoing backup and DR design will provide the following RPO and RTO. • Restoration from snapshots (individual file/files): RPO and RTO <24 hrs • Restoration of instances(from backup) in case of zonal outage: RPO will be 1 week and RTO>30 hrs • Restoration of instances (form backup) in case of accidental deletion: RPO will be 1 week and RTO (for instance size <=51TB: < 24 hrs and for instance size > 51TB: >30 hrs) • RPO will be 1 week in case of retrieval from archived data and RTO will depends on the size of restoration requested.
AR11	Performance of Shared Storage solution	Delays in implementing DDN ExaScaler Filesystem leads us to select alternative shared solutions such as Filestore which could be implemented as a temporary solution to mitigate the currency risk involved in On-Premise LeMans solution. However, this will have an impact on the performance of the Shared storage as performance of Filestore is lesser than DDN Exascaler. The performance tests captured in 2021 had shown that Filestore struggled under high concurrent load.	By moving to Filestore, we eliminate the currency risk. However, the performance baseline on Filestore should be equal or better than the current storage solution used in On-Premise LeMans solution. Refer R2.1 - EOTR - Data Migrations - Lloyds Banking Group Confluence for performance test. We will be performing rebase lining with larger and multiple Filestore instances across the different data layers and capture them against the current On-Premises. CMT Risk References: LG002 - R0204, LG002 - R0121 CMT Issue Reference: LG002 - I0082
AR12	Filestore running on NFSv3	The current offering of Filestore runs on NFSv3 protocol. This means that there is no sufficient encryption when data is in transit between the LeMans SAS servers and the Filestore Instances.	The risk is low since - Data on GCP is secured encrypted via CMEK keys secured in GCP Cloud KMS or HashiCorp Vault. These keys are managed by Bank approved life cycle management process. Auditing and logging is enforced and presented to the Bank's SIEM. The VPC on GCP cloud is provided as a secured network perimeter. The VPC is not exposed to internet. Network traffic between the Bank's on-premise systems and GCP is secured using a private interconnect and tiered firewalls. Authentication and role based access authorisation is in place using GCP IAM and the Bank's AD. Additional Role-Based Access Controls to be applied to the SPDS database in the form of column-level ACLs, to limit access to sensitive items only for select users with a business justification. Risks and SNC: Data at rest: SNC-2023-4414 Risk Associated - RK0045163 Data in transit : SNC-2024-6114 Risk Associated - RCSA Risk - R0520075

3. LeMans Migration Release Plan

The migration of the LeMans environment will be split into multiple releases to establish the necessary footprint and associated controls within the Banks' GCP environment before then migrating the content and processes to the newly provisioned infrastructure.

As part of this ITSD, we will be covering the technical components involved from Release 0 to Release 3.

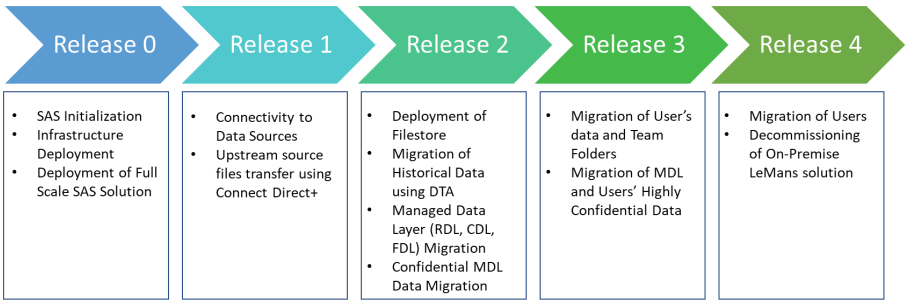
Release 0: SAS Initialization and full scale deployment of SAS Infrastructure on the BLD, INT, PRE and PRD GCP environments.

Release 1: Setting up connections to On-Premise Databases, Setting up parallel feeds from Upstream systems using Connect Direct+

Release 2: Migration of Historical Data using Data Transfer Appliance, Shared File Storage, Managed Data Layer Migration, Confidential Data Migration for MDL

Release 3: Migration of Team Users' folders, Migration of MDL and Team Users' Highly Confidential Data

Release 4: Migrate the users and shutdown On-Premises LeMans solution and start the decommission process.



4. Solution Design

4.1 On Premise High Level Architecture

LeMans is an Advanced Analytics and Modelling SAS Platform initially stood up in 2009. Built around a SAS Grid Architecture, SAS Metadata Server, with Regulatory and Financial reporting capabilities for the Bank. LeMans has continued to grow since it was installed and is currently having :

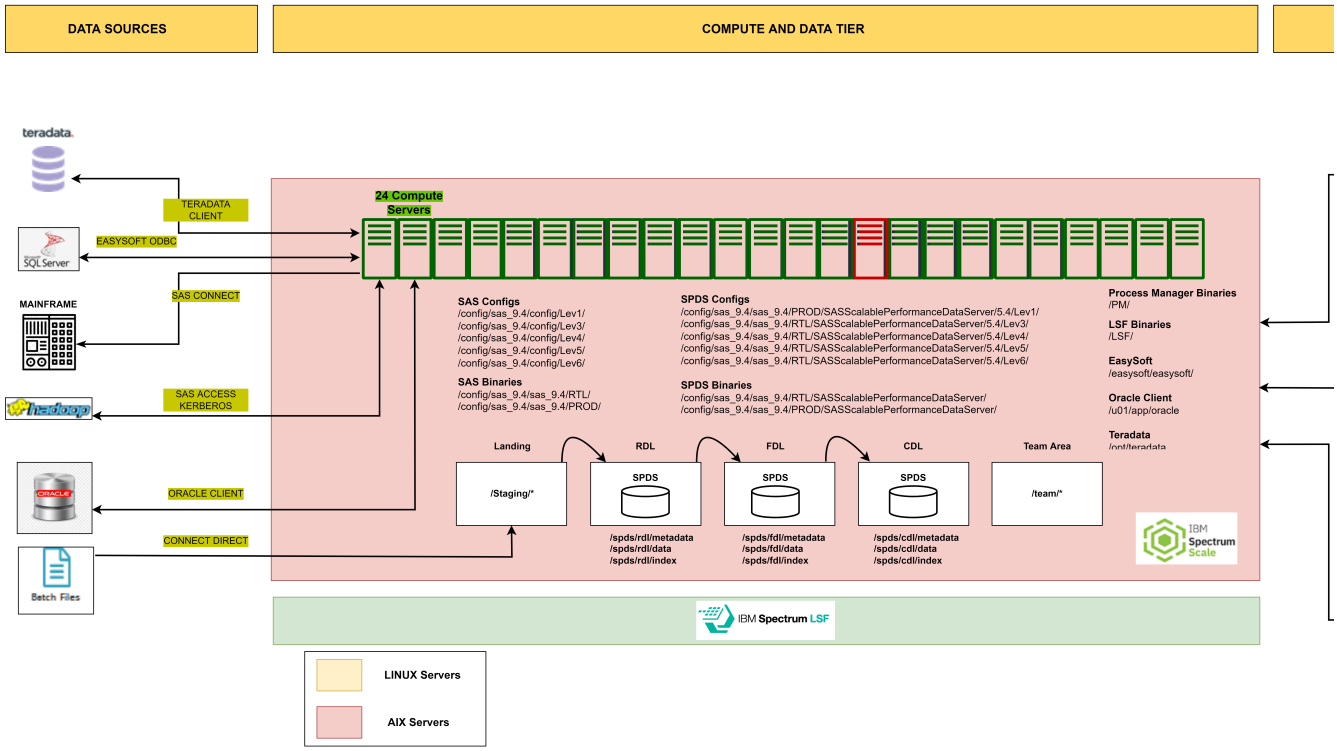
- There are multiple data sources connectors (Teradata, sql server, mainframe, Hadoop, oracle ,different batch files etc) which are processed across the SAS on-prem platform
- The SAS compute and data tier consists of 24 Grid nodes, 400+TB of data spread across SAS SPDS and SAS datasets
- The Metadata tier consist on level1 active/passive set up in production and the web application tier have 5 different server.
- More than 1000+ users are accessing this platform
- The platform supports both ad-hoc users and controlled batch processes, both managed by the Retail Risk AESM business support team

The current on-premises LeMans solutions is running on AIX 7.1 aix, which are the operating systems offered by IBM.

The LeMans application is run through 24 on-premises servers as stated. These servers are shared between 5 environments. The environments are Deployment, CIT, SIT, UAT and Production, all of which are SAS environments. 3 test environments are designated to CIT, UAT and Production, 2/3 for the test environment overall and the rest for production, all of which are spread across the Route-to-Live environment.

Storage is on-premises. LeMans currently has a storage solution called Spectrum Scale which is an IBM product where all the scalable performance data server (SPDS) SAS data is held. It provides shared storage allowing all servers within that environment to access the same data. As Spectrum Scale is not an offering by Google a comparable alternative is required.

LeMans PROD/RTL Architectur



4.2 Solution Design Target (GCP) Architecture

4.2.1 Google Cloud Environment Hierarchy

To ensure consistency with the DCX environment hierarchy, the proposed environment structure will follow the approach below as defined by CCoE following the Hub and Spoke model for project hierarchies. Each environment will be setup in a uniform manner and share the same approach for GCP projects. LeMans is onboarded as a new workstream in TechOptimization(TO) value stream. Detailed steps to create a new workstream can be found in CNE workstream onboarding guide - Workstream Onboarding 1 - Cloud Knowledge Base - Lloyds Banking Group Confluence

The below diagram shows the GCP organization, folder and project structure used in LeMans:

The deployment of SAS environments across the different LBG domains mentioned below:

- BLD is not used in the TO LZ as it is a disconnected environment with no access to other environments or on-premises therefore cannot provide integration with AD, etc
- INT provides the connectivity into the on-premise and other GCP solutions (i.e. EDH). Any upgrade, Hotfix related activity will be first deployed on lower environments starting INT, and then deployed to PRE and PRD.
- PRE is part of RTL, and will provision on-demand for specific purposes
- PRD will contain Data RTL for full-scale solution and will be persistent
    - There are four Production projects under Data RTL, prd-01, prd-02, prd-03 and prd-04. Similar structure is recommended in the initial designs of environment for PRE as well but project is using only PRE-DEV environment.

4.2.2 Google Cloud Project Design

Multiple projects are required to support LeMans on GCP, and will be deployed within the LeMans perimeter.

This follows the conventions of other projects which have a loosely-coupled architecture, with multiple projects being brought together to server a particular application such as EDH, MDL, etc. On that basis LeMans will also follow the same approach having several components as outlined below.

	Project	Description	Naming convention	Purpose	Duration
1	LeMans	SAS 9.4 Project	to-lem94-<env>-01	Hosting Compute Engine instances	This project are long-lived
2	LeMans	SAS Historical Data Project	to-lemhis-<env>-01	Transfer Appliance	This project will be short-term until the historical data migration completes

These projects will provide a setup that enables the data transfer from the current on-premises solution and the ongoing processing of the SAS environment.

- The SAS 9.4 project will be the long-standing that persist for the lifetime of LeMans solution on GCP.
- The historical data project is intended to be short-lived and will serve the historical data migration via Transfer Appliance from the on-premise LeMans to GCP. Data stored in historical project will be moved to Filestore instances in -lem94 project. Buckets in lemhis projects will be provided remote project permissions against the service account. Please find details of remote permissions provided to LeMans history project buckets in LeMans - Google Cloud Storage Bucket Details - Data Migrations - Lloyds Banking Group Confluence

To support the RTL process of LeMans the proposed projects need to be created per domain and support the provisioning of the LeMans business route to live and production environments:

	Description	SAS 9.4 Project	SAS Historical Data Project	Status
INT	**Gamma**	to-lem94-int-01-96e3	to-lemhis-int-01-c5a3	ACTIVE
PRE	**Development**	to-lem94-pre-04-a8e1	to-lemhis-pre-01-98a8	ACTIVE
PRD	**Development**	to-lem94-prd-04-8f60	N/A	ACTIVE
	CIT	to-lem94-prd-03-9677	N/A	ACTIVE
	UAT	to-lem94-prd-02-43ef	N/A	ACTIVE
	Production	to-lem94-prd-01-5006	to-lemhis-prd-02-d1d2	ACTIVE

4.2.3 LeMans GCP - High Level Platform Architecture

LeMans migration to GCP will follow a lift-and-shift approach to delivery and therefore will retain the SAS 9.4 software and SPDS data at the core of the delivery. The migration to GCP will require adoption of Linux in place of the current AIX operating system and will also introduce changes to the system which will be accounted for in the migration approach. To replicate the SAS9.4 software, the LeMans GCP solution will require a set of VMS to operate on which will be satisfied by Google Compute Engine (GCE). DCX platform will be hosting these VMs.

A single hardened image of RHEL 8 will be used as the basis for the installation of the SAS 9.4 M8 application software to support the migration of current data and processes. The operating system requirement is consistent across all of the SAS contexts that are required (Metadata, Compute, Grid and Mid-Tier), therefore a single hardened image can host SAS 9.4 M8. Any upgrade, Hotfix related activity will be first deployed on lower environments starting INT, and then deployed to PRE and PRD.

Within the GCE instances the configuration and application state for 9.4 is held on PDs which are made regionally available to support failover activities. Attached to the GCE instances, these disks provide support for the transient and temporary workloads required by SAS end-user processing. For processing of huge data volumes as part of the analytical and modelling processes supported by the SAS application, Google Filestore is chosen by undertaking multiple options review between Lloyds, Google and SAS in Q4 2021.

The following diagram presents an E2E GCP platform architecture used in LeMans. The key components and flows are described below:

- Multiple VPC Service perimeters used and how they are connected to each other. Access to another perimeter allows secure access of data, e.g. LeMans can access EDH data if the access allowed from AP service perimeter
- Central host and LeMans service projects used to allow VPC sharing in Production service perimeter. There are four projects in production service perimeter. These projects need data transfer mechanism. Currently two patterns are used, SAS connect and transfer via GCS buckets. A tech debt is noted in Future Roadmap section(AFR4) in this page to analyse the requirements and approve the pattern for inter-project data transfer.
- Connectivity with on-premise data sources and LBG network.
- Google managed tenant project where Filestore instances are located
- Compute Engine instances created in a single zone and connected to Filestore via Private Service Access. We have four compute engine instances in PRD-PRD, one in PRD-DEV, PRD-CIT, PRD-UAT and PRE-DEV, required to execute gcloud commands to create and manage snapshots, backups, archiving Filestore data and restoring archived data from Cloud Storage.
- Connection to Insights project to store GCP operation logs and KMS project to store KMS keys used with GCP resources

4.2.4. GCP Project Resource Management

LeMans GCP projects will contain resources such as virtual machines clusters (Metadata Server, Grid controller , Grid worker and Mid-Tier), Persistent Disks, Filestore, IAM service accounts, Subnets, GCS buckets etc. All infrastructure resources should be created as per CNE guidelines for the specific resource.

4.2.4.1 GCP Projects and APIs

The following document contains steps followed to create GCP projects and enable APIs in LeMans infrastructure setup:

LeMans GCP Project and Enabling APIs (environment-to-provision) - Data Migrations - Lloyds Banking Group Confluence

Details of APIs enabled in LeMans - LeMans GCP Project Structure - Data Migrations - Lloyds Banking Group Confluence

4.2.4.2 Service Accounts Details

The following document contains steps followed to create GCP projects and enable APIs in LeMans infrastructure setup:

LeMans gcp service accounts creation - Data Migrations - Lloyds Banking Group Confluence

Details of Service Accounts created in LeMans - LeMans GCP Project Structure - Data Migrations - Lloyds Banking Group Confluence

4.2.4.3 GCP Resource creation and management

Resource creation and management for resources such as **Compute Engine, BigQuery, Storage bucket, Filestore, IAM service accounts** used within LeMans can be found in Deployment and modification of Infrastructure. More detailed steps for child resources are linked in the same document. To understand the end to end GCP Project structure and resource creation and management engineering processes , please refer to LeMans GCP Project Structure.

Details of GCS buckets created and service accounts having permissions against them are documented in LeMans - Google Cloud Storage Bucket Details - Data Migrations - Lloyds Banking Group Confluence

4.2.4.4 GCP KMS and Secrets Manager

In LBG GCP environments, we use KMS service to create and manage CMEK keys and use them to encrypt data at rest with resource creation. KMS is a service that allows to create , import and manage encryption keys and perform the operations in a single cloud service. Cloud KMS service enables the data owner to be the ultimate custodian of the data similar to a the traditional on-prem system. There is a central key management project within each value stream and resources created in workstream projects use the keys created in central KMS project.

Secrets Manager is used to store credentials used by application layer on GCP. In LeMans, As part of VM provisioning, multiple start-up services are run such as Ad integration, CyberArk onboarding, DNS registration, etc. They use credentials stored in Secrets Manager, for e.g. At the time of bootstrap, the VM will retrieve Chef PEM certificate (using RHEL service account) from Google secret manager to connect with On-Prem Chef server securely and gets cookbooks run list to be executed on VM based on Policy Group defined in On-Prem GitHub. VM retrieves the Jenkins token from Google secret manager to get authenticated with Jenkins server and execute the cookbooks.

Please refer to LeMans | KMS and Secrets Manager Details - Data Migrations - Lloyds Banking Group Confluence for details of KMS keys created for resources and Secrets-Service Accounts integration for VM bootstrap.

4.3 Infrastructure Architecture

LeMans SAS platform is a 9.4 Grid deployment solution deployed on-premises which needs to be migrated to Google Cloud Platform. The LeMans solution was originally implemented in 2010 and supports a large SAS workload made up of Batch processing of around 5000 jobs and Interactive workloads with 1000 users and around 600 of them accessing the platform concurrently.

SAS platform will be hosted on GCP DCX(**D**ata**c**entre E**x**tension) platform in bank. The GCP DCX platform offers Virtual Machine (Server) hosting on GCP Compute Engine (GCE) that is connected to the bank. The platform is a Secure Landing Zone within GCP offering LBG Services for you to consume from the Virtual Machines you create to run your applications on.

The platform also contains a large SPDS data warehouse known as Managed Data Layer containing over 320 TB of data which is sourced through golden source systems and Internal Databases such as Hadoop environment "EDH" and Teradata environment "GDW". The platform also contains end-user storage area of around 280 TB containing user created SAS programs and datasets. Filestore(NFSv3) will be used to store the historical and MDL/Teams data on GCP as a storage solution until an alternative strategic approach that addresses the risks and limitations is available.

The migration of the solution from on-premises to GCP will be lift & shift and run as-is with some changes required to reflect the new platform. Wherever available, new working ways will be adopted. Public cloud product catalogue has the curation details of all products allowed and used within the bank. For any product feature updates and queries, we should contact the product team.

4.3.1 GCP DCX Architecture Overview

All Virtual Machines built on DCX are:

- Built from LBG Security assured Golden images.
- Automatically joined to LBG domains (GLOBAL, TEST01GLOBAL or IAGLOBAL).
- Built to LBG config workbooks standards with security, monitoring and alerting agents installed by default.
- Charged back to your cost code based on your usage of the VMs, with no up-front provisioning costs.
- Integrated with a centralised patching solution, with the ability to define schedules based on the operational requirements.
- Automatically backed up, with the ability to define schedules based on the operational requirements.
- Capable of being powered down and up on a defined schedule to optimise cost. By default this is applied to all non-production servers.

When you onboard into GCP DCX:

- You will be provided with a Google workstream which will allow the provisioning of projects that enable route to live environments. This provides the ability to deploy Virtual Machines (built from golden images) and supporting components such as Cloud Storage and Load Balancers within these projects.
- A Networking Hub allows resources in your projects to communicate with required services located in Google Cloud, On-Premise and external SaaS. This is known as a hub and spoke model. Application specific firewall rules can be implemented for system and user access to your virtual machines.
- GCP DCX shared services (as detailed in the diagram below) help to provision your virtual machines, maintain currency, patching and enforce compliance.
- GCP operations services (as detailed in the diagram below) help to monitor and manage your virtual machines.
- On-Premise and SaaS based services (as detailed in the diagram below) help to secure and alert upon your virtual machines and provide human and system access patterns to the platform.

4.3.2 LeMans Infrastructure on DCX

4.3.2.1 Workstream and Resource Curation

In LeMans, we are creating the Compute resources using DCX curated infrastructure modules. There are list of Git templates used for a workstream creation and management of resources. Compute resources are managed under the `workstream-compute-template`.

- `disk`: This module is used for creating Regional and Zonal Persistent Disk.
- `ilb`: The Internal Load Balancer (ILB) module is used to deploy an internal load balancer.
- `ospatch`: The `ospatch` module is used to patch RHEL servers using the `osconfig` agent.
- `rhel`: This module is responsible for creating and managing RHEL8 servers
- `UMIG`: This module is used for creating Unmanaged Instance Groups (UMIGs)

Detailed implementation steps for compute and other types of resources(IAM, storage, network) are noted in Deployment and modification of Infrastructure - Data Migrations - Lloyds Banking Group Confluence

4.3.2.2 Agents Installed on RHEL VMs in LeMans

RHEL is used for all installations. RHEL supports up to< 64 CPU sockets> and up to 24TB RAM. A number of other agents and updates are installed onto the operating system.

The below table covers all the Chef installed products on the server when a pattern is successfully deployed:

4.3.2.2.1 Installed Products

The below table covers all the Chef installed products on the server when a new VM is successfully deployed:

DEV	PRE-PROD	PROD	LBG	Template (T), Chef (C)

Chef Client	✅	✅	✅	LBG	C
Sentinel One	✅	✅	✅	LBG	C
Dynatrace	✅	✅	✅	LBG	C
Urban Code	✅	✅	✅	LBG	C
Google OS Config	✅	✅	✅	LBG	C
Google logging agent	✅	✅	✅	LBG	C
UKM	✅	✅	✅	LBG	C

4.3.2.2.2 Agentless Products

The below table covers all the agentless products used to access or vulnerability scan the VM once successfully deployed:

	DEV	PRE-PROD	PROD
QUALYS	✅	✅	✅
CyberArk	✅	✅	✅

4.3.2.2.3 Detailed Description

Agent	Description
QUALYS	LBG Scanning tool Vulnerability Manager for scanning of O/S vulnerabilities.
Chef Client	Continuous Compliance
SentinelOne	Endpoint Security
Dynatrace	Infrastructure Monitoring
Urban Code	UrbanCode software can automate builds, deployments and releases
CyberArk	CyberArk Password Manager used for secure password rotation and privileged access control
UKM	SSH Key Management

4.3.2.2.3.1 Qualys Vulnerability Report

DCX team sends us monthly reports via mail. A sample email is attached(FW Qualys Vulnerability Report for IT.GCPMA-IAAS. CLOUD ENGINEERING - GCP Tech Ops - April 2025.msg).The list provides all servers across TO, so we have to filter the servers belonging to LeMans. Vulnerability cause and solution is provided in the attached report, and we have to take actions as per the solution provided. For example, If the issue reported is an outdated package, we wait for next patching schedule to upgrade the packages.

4.3.2.2.3.2 Dynatrace Configuration

The Dynatrace agent is deployed as part of the standard image provided by DCX. The agent is configured with the user "**dtuser**" which is setup as a local account. Since this is configured by DCX as a local user and not as a Active Directory based account, the logs which are required to be ingested to Dynatrace such as "Backup and Archive logs" needs to be opened up for the **dtuser** and provided with read and execute permissions (2775) for folders and read permissions (664) for files within those folders respectively.

On the logs which are present on the /var/log location, ACL's can be applied to allow only the "**dtuser**" to have Read and Execute permissions and not others who are not part of the LeMans SAS Users Group.

On the logs which are present on the /user_logs location, the ACL's cannot be applied due to limitations on Filestore NFS protocol which means the other users who are not part of LeMans SAS Users Group such as **"dtuser"** will have Read and Execute permissions. This has been reviewed with Security Arch team and they would be detailing the issue with the help of an SNC.

#	Server	Folder	Owner	Group	Permission	ACL User	ACL Permission on Folder
1	All Servers	/user_logs/SAS94/prod/LeMans_Backup	srvapplemsas01	gg_lm_sas_sasusers01	drwxrwsr-x.	NA	NA
2	Backup Servers	/var/log/nonutf8	srvapplemsas01	gg_lm_sas_sasusers01	drwxrwsr--.	dtuser	u:dtuser:rX
3		/var/log/fst-snap	srvapplemsas01	gg_lm_sas_sasusers01	drwxrwsr--.	dtuser	u:dtuser:rX
4		/var/log/fst-backup	srvapplemsas01	gg_lm_sas_sasusers01	drwxrwsr--.	dtuser	u:dtuser:rX
5		/var/log/archival	root	gg_lm_sas_sasusers01	drwxrwsr--.	dtuser	u:dtuser:rX

> **Note**
> The same principle is followed on other Business RTL environments such as PRD-UAT, PRD-CIT, PRD-DEV and Infrastructure RTL environment PRE-DEV. The only difference would be the ownership and group that would be

defined according to the corresponding environments.

4.3.2.3 OS Patching

We are following the OS patching process defined by DCX in GCP DCX - Red Hat Enterprise Linux 8 - Product Description - Cloud Knowledge Base - Lloyds Banking Group Confluence.

RHEL updates will be delivered to the server following in the agreed maintenance window, patching will be delivered using the Google OS Config agent.

	Detailed Description	**RHEL** **INT Environment**	**RHEL** **Pre-P**
Maintenance RHEL Options for Automated Patching	The maintenance RHEL defines the time where maintenance can be performed on the deployed server for that pattern type. The maintenance RHEL will automatically inherit the IaC configured schedule maintenance RHEL and support automated OS patching. Each service project will have its own job.	Customizable via VM metadata key/pairs	Custo key/p
Support Model	Who will support VM i.e OS and components running on the OS Who will Support platform? i.e. Underlying infrastructure like Terraform, Jenkins and Chef	VM would be supported by Business platform team (Earlier know as value stream) including patching, Backup and overall administration GCP DCX Platform is supported by Cloud Services SRE.	VM w Busin know includ overa GCP by Cl

We use OS patch management to apply operating system patches across a set of Compute Engine VM instances (VMs). Long running VMs require periodic system updates to protect against defects and vulnerabilities.

In order to use the OS patch management feature, you must set up the OS Config API and install the OS Config agent. The OS Config service enables patch management in your environment while the OS Config agent uses the update mechanism for each operating system to apply patches. Updates are pulled from the package repositories (otherwise called the *distribution source package*) or a local repository for the operating system. Details of OS patching for LeMans is documented in LeMans RHEL Patching Linux Servers - Data Migrations - Lloyds Banking Group Confluence

4.3.2.4 Additional RHEL Configurations

In LeMans, we are modifying the `/etc/fstab` file to add additional disk and Filestore mount point details for permanent mount.

We have also installed Connect: Direct agents on Grid controller servers in installation directory `/cdirect`. Additionally, In `/etc/systemd/system`, we have created the `cdirect.service` file. This file is required to start and stop the Connect Direct service.

4.3.2.4 Chef Overrides

Chef overrides refers to custom configurations that are applied during provisioning of a VM. Using cookbooks and recipes, Chef defines the setup and configurations of the servers. Overrides are a way to modify the default attributes defined in cookbooks and direct to use the override attributes instead of default ones. For e.g. SAS Le Mans application requires the nftables (OS firewall) rules to be overridden to enable connectivity to connect direct application, so recipe *include_recipe '::firewall'* is added.

In LeMans, following override recipes are added by DCX on our request. Sample JIRA ticket is attached for reference: [GCPMRO-32] cic_overrides_rhel8 cookbook enhancement - Lloyds Banking Group JIRA

```
include_recipe '::audit'
include_recipe '::firewall'
include_recipe '::selinux_permissive'
include_recipe '::rpcbind'
include_recipe '::script'
include_recipe '::sssd_ldap_filter'
include_recipe '::ulimits'
```

The following section contains individual link and scripts for each recipe.

4.3.2.4.1 Audit Log Recipe

› Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast

----------------------------audit.rb-------------------------
#
```

```
# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: audit

# Author:: Anistan, Amala (Cloud Platform)

#

# Copyright:: 2024, The Authors, All Rights Reserved.

#

# Audit configuration
# Override the Audit configuration in the lbg_cic_rhel8_hardening cookbook
# Audit

filrule = node['lbg']['cic']['hardenlinux']['audit']['rulsfile']
# Creation of immutable file for audit

file "#{filrule}" do

  action :create

  mode '0740'

  owner 'root'

  group 'root'

end
# Update of module rules on auditd

test_rules = [

  '-w /team/adm -p w -k file_del',

  '-w /team/af_credit_policy_team -p w -k file_del',

  '-w /team/af_customer_experience_team -p w -k file_del',

  '-w /team/af_impairment_and_capital_team -p w -k file_del',

  '-w /team/af_lex_autolease_team -p w -k file_del',

  '-w /team/af_modelling_team -p w -k file_del',

  '-w /team/af_portfolio_management_insight -p w -k file_del',

  '-w /team/af_portfolio_performance_team -p w -k file_del',

  '-w /team/af_pricing_team -p w -k file_del',

  '-w /team/analytics -p w -k file_del',

  '-w /team/analytics_mod -p w -k file_del',

  '-w /team/banking -p w -k file_del',

  '-w /team/basel -p w -k file_del',

  '-w /team/bicc_admin -p w -k file_del',

  '-w /team/bicc_developers -p w -k file_del',

  '-w /team/caprep -p w -k file_del',
```

```
'-w /team/decision_science -p w -k file_del',
'-w /team/mon_dev -p w -k file_del',
'-w /team/cards -p w -k file_del',
'-w /team/mortgages -p w -k file_del',
'-w /team/cards_repricing -p w -k file_del',
'-w /team/fca_market_strategy -p w -k file_del',
'-w /team/perf_mon -p w -k file_del',
'-w /team/cbrgen -p w -k file_del',
'-w /team/portfolio_analytics -p w -k file_del',
'-w /team/cbrrep -p w -k file_del',
'-w /team/fmod -p w -k file_del',
'-w /team/private_banking -p w -k file_del',
'-w /team/CCFD -p w -k file_del',
'-w /team/forecast -p w -k file_del',
'-w /team/rbbcad -p w -k file_del',
'-w /team/cf_analytics -p w -k file_del',
'-w /team/fraud -p w -k file_del',
'-w /team/rbbcd -p w -k file_del',
'-w /team/cf_monitoring_and_reporting_team -p w -k file_del',
'-w /team/frep -p w -k file_del',
'-w /team/rbbcr -p w -k file_del',
'-w /team/chief_operating_office -p w -k file_del',
'-w /team/rbbfraud -p w -k file_del',
'-w /team/collections_recoveries -p w -k file_del',
'-w /team/loans -p w -k file_del',
'-w /team/rici -p w -k file_del',
'-w /team/consumer_cards_business -p w -k file_del',
'-w /team/macro -p w -k file_del',
'-w /team/riskrep -p w -k file_del',
'-w /team/cpm -p w -k file_del',
'-w /team/mbna -p w -k file_del',
'-w /team/verde -p w -k file_del',
'-w /team/crdiv -p w -k file_del',
'-w /team/stress_test -p w -k file_del',
```

```
  '-w /team/crp -p w -k file_del',

  '-w /team/modelling -p w -k file_del',

  '-w /team/customer -p w -k file_del',

  '-w /team/modelling_data -p w -k file_del',

  '-w /spds1 -p w -k file_del',

  '-w /spds2 -p w -k file_del',

  '-w /spds3 -p w -k file_del',

  '-w /spds4 -p w -k file_del',

  '-w /staging -p w -k file_del',

  '-w /sasshare01 -p w -k file_del',

  '-w /cdirect -p w -k file_del',

]
execute 'augenrules_add' do

  command '/sbin/augenrules'

  command '/usr/sbin/auditctl -R /etc/audit/audit.rules'

  action :nothing

end
execute 'augenrules' do

  command '/sbin/augenrules'

  action :nothing

end
test_rules.each do |r|

  append_if_no_line "ensure #{r} in audit.rules" do

    path '/etc/audit/rules.d/lemans_customs_logs.rules'

    line r

    notifies :run, 'execute[augenrules_add]', :delayed

  end

end
```

4.3.2.4.2 Firewall Recipe

› Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
--------------------firewall.rb------------------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: firewall
```

```
#

# Copyright:: 2022, The Authors, All Rights Reserved.

#

# Firewall configuration
# Override the nftables configuration in the lbg_cic_rhel8_hardening cookbook
# Don't automatically run the NFTables install script, as ordering needs to be adjusted

edit_resource(:execute, 'Install the NFTables on host') do

  action :nothing

end
# replace the cookbook file with a template

delete_resource(:cookbook_file, '/etc/nftables/nftables.rules')
# Always allow ssh

firewall_ports = %w(22)

# Add the specified extra TCP port(s) to the list of ports to be opened in the firewall

firewall_ports += node['lbg']['cic']['overrides_rhel8']['extra_firewall_ports']

ports_list = firewall_ports.join(', ')
template '/etc/nftables/nftables.rules' do

  source 'nftables.rules.erb'

  variables(ports: ports_list)

  # The install script doesn't reapply the rules if the nftables.rules file is updated, force

  notifies :run, 'execute[nft -f /etc/nftables/nftables.rules]', :immediately

  notifies :run, 'execute[Install the NFTables on host]', :immediately # run the install scrip

end
execute 'nft -f /etc/nftables/nftables.rules' do

  action :nothing

end
```

4.3.2.4.3 SElinux Recipe

› Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
-------------------selinux_permissive.rb---------------------------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: selinux

#

# Copyright:: 2022, The Authors, All Rights Reserved.

#

# Override the SELinux settings in lbg_cic_rhel8_hardening to run in permissive mode
delete_resource(:template, '/etc/selinux/config')
```

```
delete_resource(:execute, 'setenforce0')
ruby_block 'Set SELinux to permissive mode on reboot' do

  block do

    f = Chef::Util::FileEdit.new('/etc/selinux/config')

    f.search_file_replace_line(/^SELINUX=enforcing/, "SELINUX=permissive\n")

    f.write_file

  end

  not_if "grep 'SELINUX=permissive' /etc/selinux/config"

end
# Set SELinux to permissive mode temporarily, to avoid having to reboot

execute 'setenforce 0' do

  only_if "[ `getenforce` != 'Permissive' ]"

end
```

4.3.2.4.4 rpcbind Recipe

> Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
--------rpcbind.rb---------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: rpcbind

#

# Copyright:: 2022, The Authors, All Rights Reserved.

#

# rpcbind service configuration
# Check if rbcbind service is installed

rpcbind_check = shell_out('rpm -q rpcbind')

rpcbind_installed = rpcbind_check.exitstatus == 0
# Install the rpcbind package if it's not installed

package 'rpcbind' do

  action :install

  only_if { !rpcbind_installed }

end
# Enable and start the rpcbind service if it's installed

service 'rpcbind' do

  action [:enable, :start]

  only_if { rpcbind_installed }
```

```
end
```

4.3.2.4.5 selinux Recipe

Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
----------script.rb-------------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: script

#

# Copyright:: 2022, The Authors, All Rights Reserved.

#

# Linux script configuration
# Override the linux-script configuration in the lbg_cic_rhel8_hardening cookbook

#

# Linux file Permissions / User and Group Settings override

#

cookbook_file '/usr/local/bin/linux-script.sh' do

  source 'linux-script.sh'

  action :create

  mode '0700'

  owner  'root'

  group  'root'

end
```

4.3.2.4.6 SSSD_LDAP Recipe

Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
---------------sssd_ldap_filter.rb-------------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: sssd_ldap_filter

# Author:: Anistan, Amala (Cloud Platform)

#

# Copyright:: 2024, The Authors, All Rights Reserved.

#

# sssd_ldap_filter addition
lbg_cic_banner 'Start of Override SSSD LDAP filter recipe'

Chef::Log.info('')
# default path of the sssd.conf file
```

```
sssd_conf_path = '/etc/sssd/sssd.conf'
# Define required domain based on the 2nd character of the hostname

nodename = node['hostname'].downcase

env = nodename[1, 1]
# Non-prod identifiers

prod = %w(L l)

preprod = %w(T t)

nonprod = %w(D I A O d i a o)
case env

when *prod

  # additional variables PRD

  additional_config = node['lbg']['cic']['overrides_rhel8']['ldap_filter_prd']
when *preprod

  # additional variables PRE-PROD

  additional_config = node['lbg']['cic']['overrides_rhel8']['ldap_filter_pre']
when *nonprod

  # additional variables IAGLOBAL

  additional_config = node['lbg']['cic']['overrides_rhel8']['ldap_filter_int']
else

  Chef::Log.info('Valid domain not available for this host - hostname is not GLOBAL, TEST01GLO

end
current_config = {}

current_section = nil
# Check if the file exists

if File.exist?(sssd_conf_path)

  # Read exisiting sssd.conf

  File.readlines(sssd_conf_path).each do |line|

    line.strip!
    # Identify section headers

    if line.start_with?('[') && line.end_with?(']')

      current_section = line

      current_config[current_section] ||= {}

    elsif current_section && line.include?('=')

      # Assuming format is key = value

      key, value = line.split('=', 2).map(&:strip)

      current_config[current_section][key] = value unless key.nil? || value.nil?

    end

  end
```

```
  # Find and update the appropriate domain section

  domain_sections = current_config.keys.select { |section| section.start_with?('[domain/') }
  if domain_sections.any? && !additional_config.empty?

    # Checking domain section

    domain_section = domain_sections.first

    changes_made = false
    # Merge new variables into the existing configuration

    additional_config.each do |key, value|

      # Check if the varibales already exist

      unless current_config[domain_section].key?(key)

        current_config[domain_section][key] = value

        changes_made = true

      end

    end
    # Write back to sssd.conf

    if changes_made

      File.open(sssd_conf_path, 'w') do |file|

        current_config.each do |section, settings|

          file.puts section

          settings.each do |key, value|

            file.puts "#{key} = #{value}"

          end

          file.puts ''

        end

      end
      # Restart the sssd service

      system('systemctl restart sssd')

    else

      Chef::Log.info('No changes were made to sssd.conf file, service restart not required')

    end

  else

    Chef::Log.info('Domain section is not found in sssd.conf')

  end

else

  Chef::Log.info('SSSD.Conf file is not found')
```

```
end
```

4.3.2.4.7 Ulimit Recipe

Click here to expand...

```
https://ghe.service.group/SOAR-SDDT-Chef-Enterprise/cookbook_lbg_cic_overrides_rhel8/blob/mast
-------------unlimits.rb----------------
#

# Cookbook:: lbg_cic_overrides_rhel8

# Recipe:: unlimits

# Author:: Anistan, Amala (Cloud Platform)

#

# Copyright:: 2024, The Authors, All Rights Reserved.

#

# Limits.conf file update
lbg_cic_banner 'Start of Override limits.conf file'

Chef::Log.info('')
# Define the path for the limits.conf file

filperm = node['lbg']['cic']['hardenlinux']['core']['fileperm']
# Stopping the automatic updates to the limit.conf file from hardening process to allow adjust

edit_resource(:template, "#{filperm}") do

  action :nothing

end
# File creation with permissions and ownership of the limit.conf

file "#{filperm}" do

  owner 'root'

  group 'root'

  mode  '0755'

  action :create_if_missing

end
# Deploy the limits.conf file using a template

template "#{filperm}" do

  source 'limits.conf.erb'

  mode '0400'

  owner 'root'

  group 'root'

  variables(

    core_hard: node['lbg']['cic']['overrides_rhel8']['limits']['core_hard'],

    nofile: node['lbg']['cic']['overrides_rhel8']['limits']['nofile'],
```

```
    nproc_soft: node['lbg']['cic']['overrides_rhel8']['limits']['nproc_soft'],
    nproc_hard: node['lbg']['cic']['overrides_rhel8']['limits']['nproc_hard'],
    stack: node['lbg']['cic']['overrides_rhel8']['limits']['stack']
  )
  not_if { ::File.exist?("#{filperm}") && ::File.read("#{filperm}").include?('*    soft     np
  action :create
end
```

4.3.2.5 Image Maintenance

Maintenance for Value Stream servers across all environments involves several key tasks to ensure optimal performance and security. These tasks include ensuring that the server images used are no more than two versions behind the latest release, which helps in maintaining compatibility and security. All images should be released quarterly, ensuring that there are four images available per year for each Windows and RHEL flavour. While a monthly image release would be ideal, the overhead of manual Qualys scans has led to a proposal for quarterly image creation after discussions with the SRE team.

It is the responsibility of the Value Stream to keep the latest images for the application servers. Please note that this process adds an overhead for application rebuilds during the image refresh. This means that all application reconfigurations are required whenever the image is refreshed. Additionally, it is the Value Stream's responsibility to keep the application documentation up-to-date until the applications are automated.

Value Streams need to agree internally with the Business Owners on whether refreshing the images is required or not.

Please refer to GCP DCX - Servers Regular maintenance - Tech Optimisation - Lloyds Banking Group Confluence for details on other maintenance activities and RACI.

[GCP DCX SRE] - IaaS -Images & Agent versions - Tech Optimisation - Lloyds Banking Group Confluence documents the release schedule for Google IAAS images that are referenced across TOs

Hostname retention during image upgrade is tested in PRE environment, and the following documents are providing the details of the steps and testing documents using the options provided by DCX for retaining the hostname while updating the Golden Image on the SAS LeMans VMs.

Hostname retention testing on Application servers - Data Migrations - Lloyds Banking Group Confluence

hostname retention testing on Lemans test server - Data Migrations - Lloyds Banking Group Confluence

LeMans - DCX Hostname Change Testing Plan - Data Migrations - Lloyds Banking Group Confluence

4.3.3 Connect:Direct Agents on GCP Infrastructure

IBM Connect:Direct is point-to-point (peer-to-peer) file-based integration middleware meant which provides assured delivery, high-volume, and secure data exchange within and between enterprises. The On-Premises Le Mans solution currently receives around 300+ batch feeds from various upstream systems using Connect Direct. These feeds arrive daily and will be used by the Managed Data Layer batch processes. As LeMans solution is migrated to GCP, the connect direct feeds arriving from the source systems is pointed to the GCP environment.

LeMans Connect: Direct Implementation describes how the Connect:Direct Agent will be deployed and configured on the LeMans solution running in GCP.

4.4 Network Architecture

LBG operates a hub and spoke network topology, with a central hub ("Mgmt") terminating the connectivity back to the LBG on-premises network (via Co-Lo) and providing centralized platform tooling for provisioning environments for Value Streams. Each Value Stream within GCP at LBG represents a 'spoke' and is allocated a dedicated Shared VPC network by default. This Value Stream Shared VPC is connected to the Mgmt hub by means of Google Cloud HA VPN, providing highly-available connectivity over redundant IPsec tunnels capable of supporting 3Gbps throughput. Each Value Stream Shared VPC might support multiple workloads/verticals.

LeMans vertical/workstream is part of TO Value Stream and LeMans subnets will be allocated within TO Shared VPCs according to the following principles:

- LeMans will be isolated on the network, e.g. LeMans subnets will be separated from other subnets
- Each LeMans workload shall have it's own subnet, e.g. SAS 9.4 instances shall be isolated from Filestore File System
- Each environment (BLD, INT, PRE, PRD) provides its own VPC for separation across environments.

However, since Filestore is a Google Cloud native solution, it will be deployed within a Google Managed Private VPC & Subnet (PSA) which will be accessible from the LeMans SAS 94 subnet using Private VPC Peering connection setup between the two resources.

Complete information on the network design including Subnets, Firewall Rules and Ports are documented in the Confluence page mentioned - LeMans - Network Design - R2

4.4.1 IP Usage by Environment

Subnet used in our design has /26 subnet mask for the PRD and PRE environments providing 62 usable primary IP addresses. Breakdown of the IP usage across all the environments are provided below. For a full DR scenario, we will need to rebuild all VMs in another zone so the IPs consumed would be double of what is currently used (except for the ILB which is static). The below numbers are subjected to change with infrastructure change and should be updated in LeMans - Network Design - R2

Services	PRD-PRD	PRD-DEV	PRD-CIT	PRD-UAT	PRE-DEV
Grid worker nodes	16	2	2	4	2
Grid controller nodes	2	2	2	2	2
Mid-Tier Servers	2	2	2	2	2
Metadata Servers	3	3	3	3	3
Housekeeping Servers	4	1	1	1	1
Mid-tier Load Balancer	1	1	1	1	1
Connect Direct Load Balancer	1	N/A	N/A	N/A	1

4.4.2 Load Balancers

4.4.2.1 Middle-Tier Load Balancer

Google's Network Load Balancer is used to distribute the traffic and provide HA capability.

The detailed pattern on the Network ILB in LeMans is defined in the Confluence pages LeMans SAS Load Balancer - Data Migrations - Lloyds Banking Group Confluence and the implementation with test scenarios are defined in Lemans SAS TCP ILB - Data Migrations - Lloyds Banking Group Confluence

Load Balancer Details

Environment	Load Balancer Details
PRD-PRD	to-lem94-prd-01-clb-euwe2-backend-cb5e
PRD-UAT	to-lem94-prd-01-clb-euwe2-backend-bf35
PRD-CIT	to-lem94-prd-01-clb-euwe2-backend-8050
PRD-DEV	to-lem94-prd-01-clb-euwe2-backend-1e46

4.4.2.2 Connect Direct Load Balancer

Connect Direct allows the feeds files to be sent and received by the source and target systems within LBG. This is then consumed by the batch processes. In order to provide High Availability in case of host failure, CD agents are to be deployed on the Grid Primary and Secondary Controller Servers and Google's Network Load Balancer is configured to distribute the traffic to the CD nodes.

The detailed steps on the implementation steps for the CD agents is defined in the Confluence page LeMans Connect:Direct Implementation - Data Migrations - Lloyds Banking Group Confluence

Load Balancer Details

Environment	Load Balancer Details
PRD-PRD	to-lem94-prd-01-clb-euwe2-backend-5902
PRE-DEV	to-lem94-pre-01-clb-euwe2-backend-4e7b

4.4.3 Firewalls

we have 3 types of firewalls:

1. CoLo Firewall: This is owned the by LBG Network team and it is applied at the colocation facilities where the GCP Network and the LBG network are joined. Details of setup are available in LeMans Network | CoLo firewall updates - Data Migrations - Lloyds Banking Group Confluence
2. GCP Firewall: Every value stream comes with default ingress deny and default egress deny firewall hence the LeMans platform has to open all the required firewall for various distributed communication requirements. For every connection

that is opened the data in transit has to be encrypted using TLS 1.2 else appropriate SNC must be in place. Details of Firewalls are covered in SAS 9.4 - Network Requirements - Data Migrations - Lloyds Banking Group Confluence

3. RHEL nftables firewall (Host firewall): Every Rhel VM is in the DCX comes with default deny for both ingress and egress through nftables firewall.

4.4.4 CIDR Range

Following sections present the CIDR range for subnets across environments including the IPs allocated to Filestore Private Service Access in host project. Private Address Space IP Range is allocated per VPC

4.4.4.1 Production

Project	Name	Purpose	Region	Size	Routable	Comments
to-hst-prd	google-managed-service-to-hst-<prj>-cnw-<n>	Filestore PRD	europe-west2	/19	No	Allocated IP range for all Filestore instances in Production (Private Address Space IP Range is allocated per VPC)
to-lem94-prd-01	lem94-csn-euwe2-lem94-sas-01-prd-01	SAS 9.4 PRD	europe-west2	/26	Yes	Allocated IP range for LeMans GCE instances in PROD - PRD Environment
to-lem94-prd-02	lem94-csn-euwe2-lem94-sas-02-prd-02	SAS 9.4 UAT	europe-west2	/26	Yes	Allocated IP range for LeMans GCE instances in PROD - UAT Environment
to-lem94-prd-03	lem94-csn-euwe2-lem94-sas-03-prd-03	SAS 9.4 CIT	europe-west2	/26	Yes	Allocated IP range for LeMans GCE instances in PROD - CIT Environment
to-lem94-prd-04	lem94-csn-euwe2-lem94-sas-04-prd-04	SAS 9.4 DEV	europe-west2	/26	Yes	Allocated IP range for LeMans GCE instances in PROD - DEV Environment

4.4.4.2 PRE

Project	Name	Purpose	Region	Size	Routable	Comments
to-hst-prd	google-managed-service-to-hst-<prj>-cnw-<n>	Filestore PRE	europe-west2	/23	No	Allocated IP range for all Filestore instances in PRE (Private Address Space IP Range is allocated per VPC)
to-lem94-pre-04	lem94-csn-euwe2-lem94-sas-01-pre-04	SAS 9.4 DEV	europe-west2	/26	Yes	Allocated IP range for LeMans GCE instances in PRE- DEV Environment

4.5 SAS Application Architecture

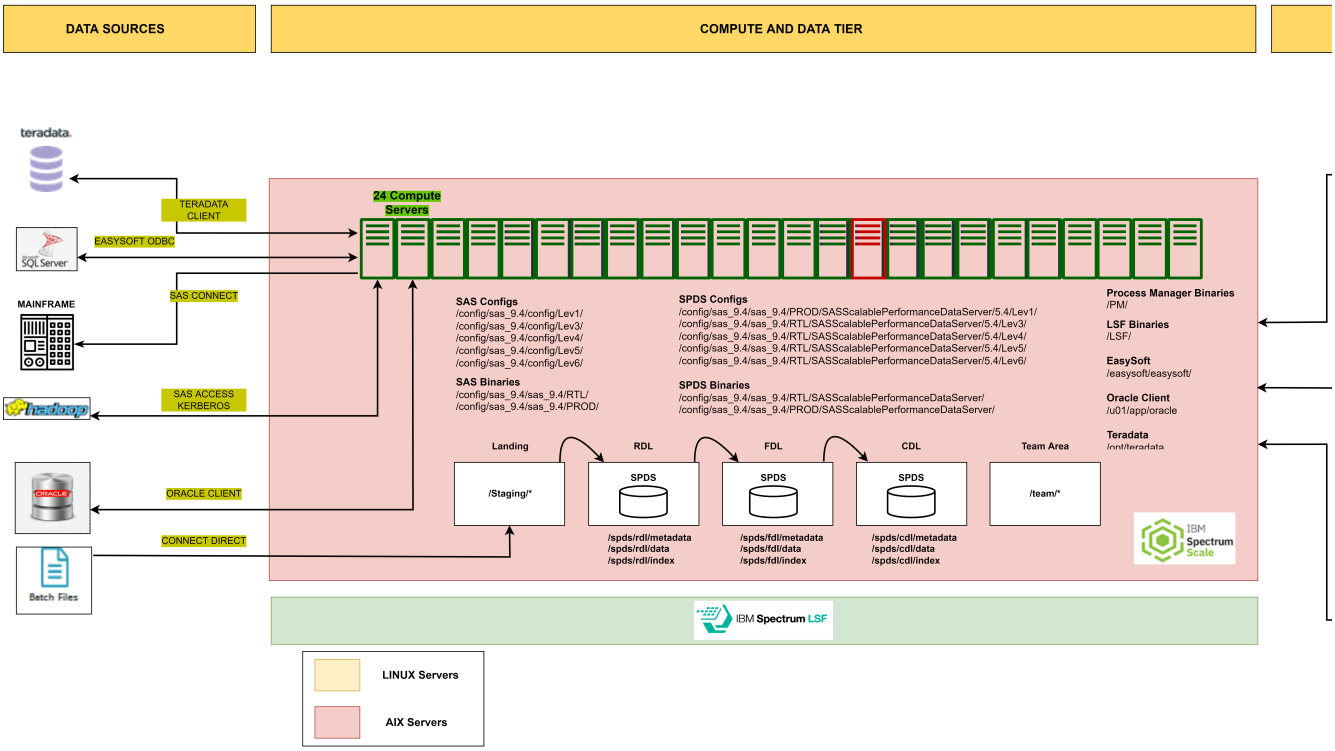
The SAS 94 solution will be deployed based on multi-tiered architecture similar to on-premises deployment which contains Metadata tier, Grid Compute Tier and Middle Tier.

4.5.1 Current Architecture on On-Prem

The current production environment provides shared infrastructure to host multiple SAS logical instances (Levs) to provide business route to live capabilities.

- 3 Metadata Servers running on AIX for the 5 Business Lev's configured with Active/Passive setup.
- 24 Compute Servers running on AIX deployed using Grid Architecture for all the Business Lev's.
- 6 types of data sources are configured to be used with the LeMans application.
- 5 Middle Tier servers running on RHEL are setup for the Business Lev's with F5 Load Balancer to provide the HA and resiliency.
- IBM Spectrum Scale (GPFS) is used as the distributed/shared storage solution to host the MDL data and Team's data.

LeMans PROD/RTL Architectur

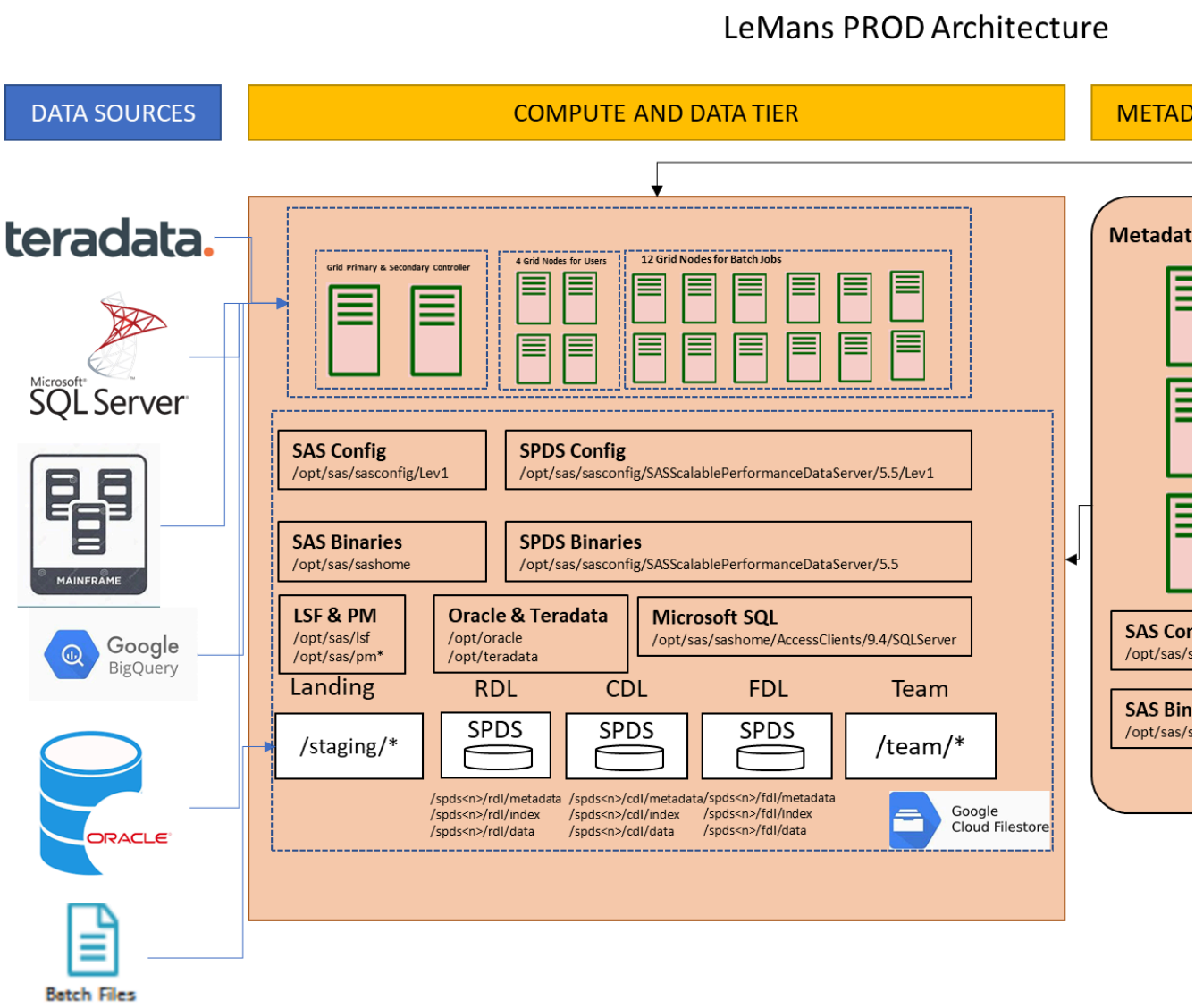


4.5.2 Target Architecture

4.5.2.1 LeMans PROD Architecture

The below Architecture provides distributed infrastructure to host the LeMans Production environment on GCP.

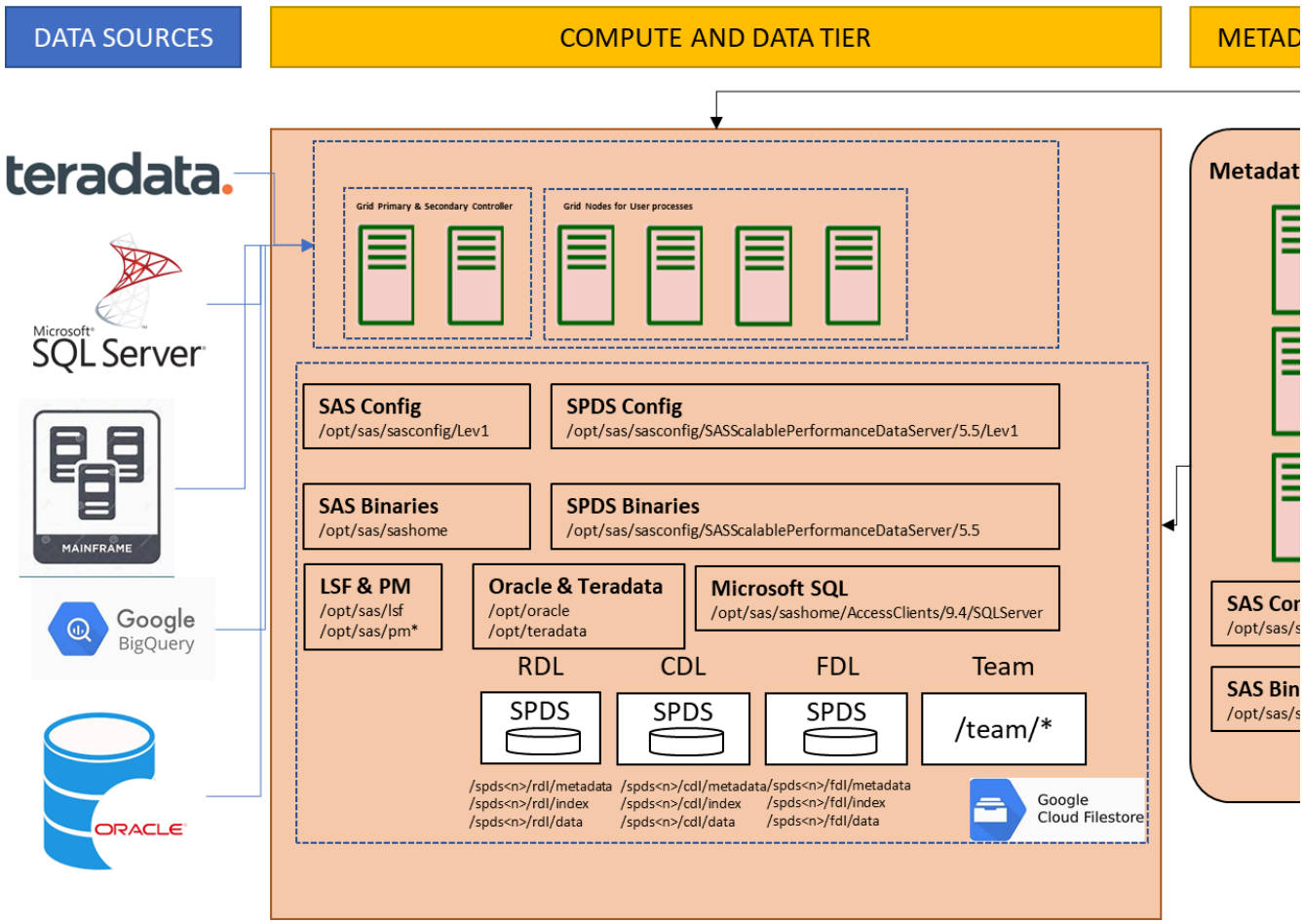
- 3 Metadata Servers deployed in clustered mode which provides high availability.
- 2 Grid controller servers where the Core SAS Services such as WIP DB, LSF Master daemons, Platform Process Manager, Connect Direct agents are deployed.
- 4 Grid nodes are provided for the user's to execute their compute workload.
- 12 Grid nodes are provided for the batch workload.
- 6 types of data sources are configured to be used with the LeMans application.
- 2 Middle Tier servers are provided for the Web Application services. They are configured with GCP Network Load Balancer to provide HA.
- Google Cloud Filestore is configured as the distributed storage solution to host the MDL and team's data layer.



4.5.2.2 LeMans PROD-RTL Architecture

The below Architecture provides distributed infrastructure to host the LeMans Business RTL environments.

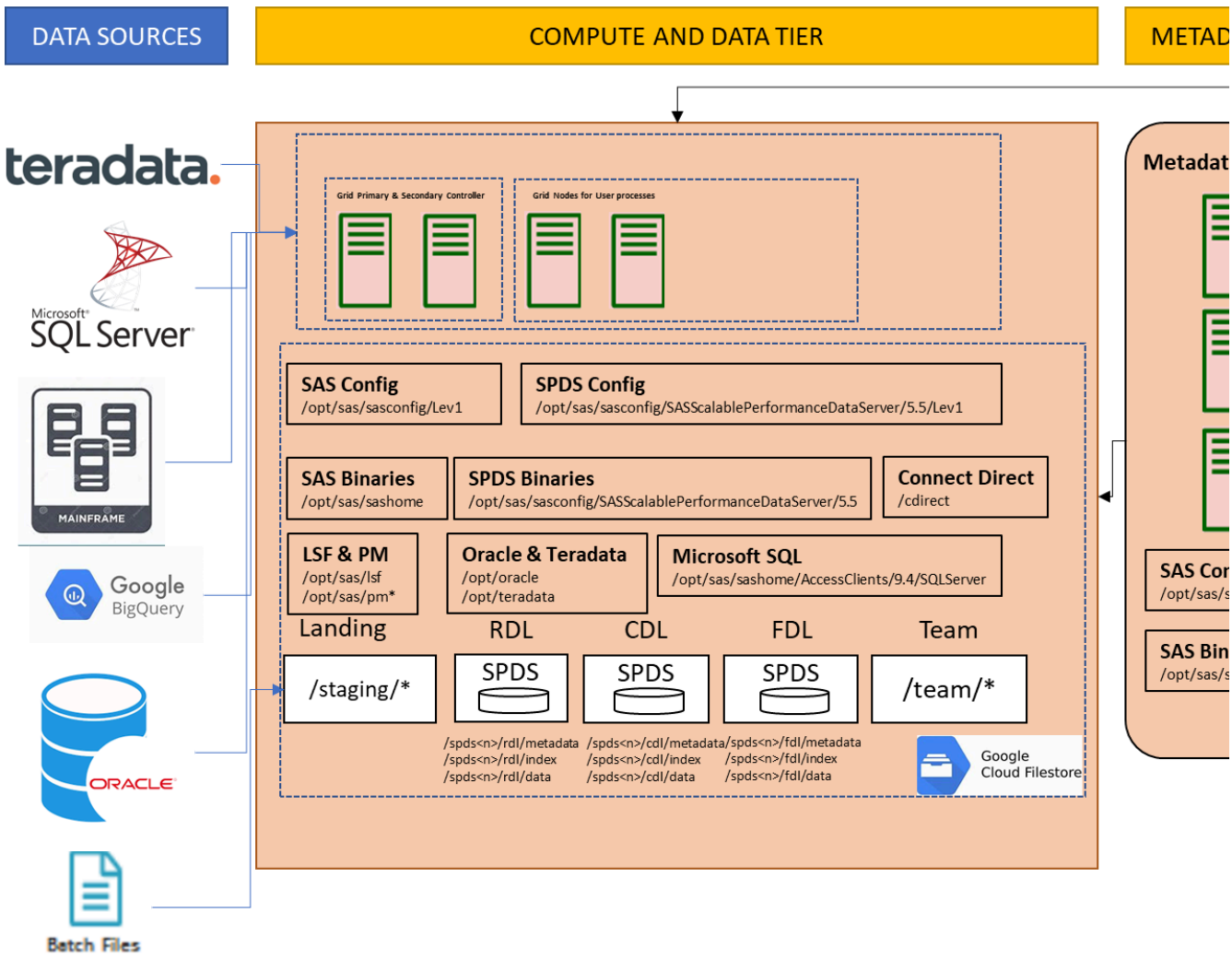
- 3 Metadata Servers deployed in clustered mode which provides high availability.
- 2 Grid controller servers where the Core SAS Services such as WIP DB, LSF Master daemons, Platform Process Manager are deployed.
- 4 Grid nodes for PRD-UAT and 2 Grid Nodes for PRD-DEV and PRD-CIT environments are provided for the user's to execute their compute workload.
- 6 types of data sources are configured to be used with the LeMans application.
- 2 Middle Tier servers are provided for the Web Application services. They are configured with GCP Network Load Balancer to provide HA.
- Google Cloud Filestore is configured as the distributed storage solution to host the MDL and team's data layer.



4.5.2.3 LeMans PRE Architecture

The below Architecture provides distributed infrastructure to host the LeMans Infrastructure RTL environment.

- 3 Metadata Servers deployed in clustered mode which provides high availability.
- 2 Grid controller servers where the Core SAS Services such as WIP DB, LSF Master daemons, Platform Process Manager, Connect Direct agents are deployed.
- 2 Grid Nodes are provided for the user's to execute their compute workload.
- 6 types of data sources are configured to be used with the LeMans application.
- 2 Middle Tier servers are provided for the Web Application services. They are configured with GCP Network Load Balancer to provide HA.
- Google Cloud Filestore is configured as the distributed storage solution to host the MDL and team's data layer.



4.5.3 SAS 94 Product Specification

SAS 9.4 product will be deployed on the LeMans GCP platform to continue supporting the existing 9.4 code and models that is running on the LeMans On-Premises solution. Since we are taking the lift and shift migration approach, the same set of core products will be deployed on the LeMans GCP platform. The version of the SAS 9.4 software which is currently deployed on On-Premise is SAS 94 M5 whereas on the LeMans GCP platform we will deploy the latest release of the SAS 94 solution which is SAS 94 M8 which provides additional support for products such as Google BigQuery, SAS/ACCESS to Microsoft SQL Server.

	LeMans On-Prem	LeMans on GCP
SAS 94 Version	SAS 94 M5	SAS 94 M8
LSF Version	10.1.0.3	10.1.0.12
PM Version	10.1.0.0	10.2.0.12

4.5.4 SAS 94 Product Stack

The below SAS 94 products are included in the SAS 94 license for the LeMans GCP Platform.

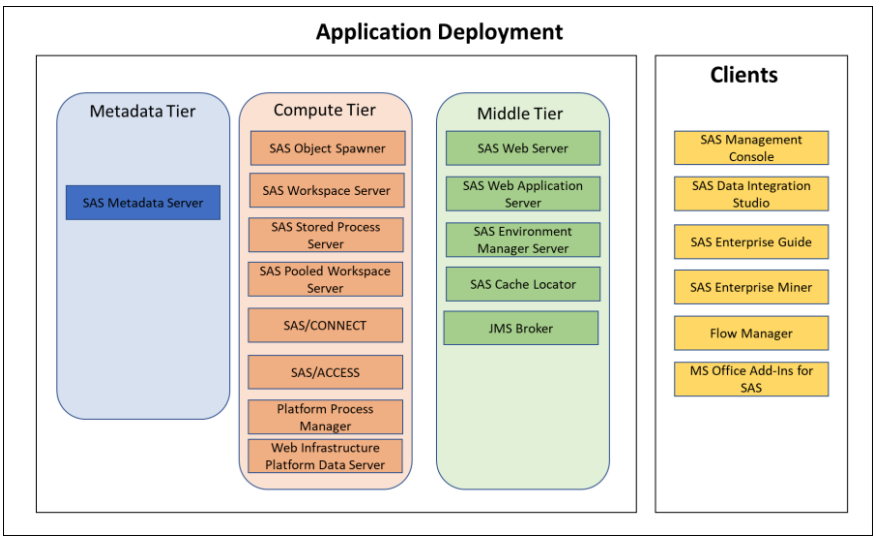
	Enabled in LeMans	Documentation/Evidence
Grid Manager for Platform	Y	LeMans Product Stack
High Performance Suite	N	

MDDB Server common products	Y	LeMans Product Stack
SAS Add-in for Microsoft Excel	Y	LeMans Product Stack
SAS Credit Scoring	Y	LeMans Product Stack
SAS Enterprise Guide	Y	LeMans Product Stack
SAS Enterprise Miner Client	Y	LeMans Product Stack
SAS Enterprise Miner Server	Y	LeMans Product Stack
SAS Integration Technologies	Y	LeMans Product Stack
SAS OLAP Server	Y	LeMans Product Stack
SAS Workspace Server for Enterprise Access	Y	LeMans Product Stack
SAS Workspace Server for Local Access	N	
SAS/ACCESS Interface to DB2	N	LeMans SAS/Access Configurations LeMans - External Connections - Data Migrations - Lloyds Banking Group Confluence SAS/Access to Google BigQuery
SAS/ACCESS Interface to Microsoft SQL Server	Y	
SAS/ACCESS Interface to ODBC	Y	
SAS/ACCESS Interface to Oracle	Y	
SAS/ACCESS Interface to PC Files	N	
SAS/ACCESS Interface to Teradata	Y	
SAS/ACCESS to Google BigQuery	Y	
SAS/ACCESS to Hadoop	Y	
SAS/ACCESS to Spark	N	
SAS/CONNECT	Y	
SAS/ETS	Y	LeMans Product Stack
SAS/GRAPH	Y	LeMans Product Stack
SAS/OR	Y	LeMans Product Stack
SAS/Secure 168-bit	Y	LeMans Product Stack
SAS/STAT	Y	LeMans Product Stack

4.5.5 SAS Software Layout

The SAS 94 software can be customized based on specifications required by the environment. Due to the complexity and size of the existing LeMans on-premise solution, the same approach is followed when deploying the SAS software on the LeMans GCP Platform. The SAS 94 application will be a multi-machine deployment where different services are grouped together in tiers based on the type of service. SAS Grid consists of the typical 3-tier SAS implementation, but includes additional compute nodes and a special Grid Controller node in the Compute Tier which runs the LSF processes. Specifically, the SAS Grid requires the following logical tiers:

- Metadata Server
- Mid-tier Server
- Grid Controller (GC) Server
- Grid Compute Node(s)



4.5.5.1 Metadata Tier

The Metadata Tier manages information shared by all components in a deployment; controls authentication and authorization; enables communication among server processes. The Metadata tier will be deployed as a clustered service across 3 RHEL VM's where they act together as a single deployment called metadata quorum. This approach provides redundancy and high availability of the metadata server in a **single zone**. Clustering ensures that the server continues to operate if any one of the server host machine fails, however this doesn't provide resiliency against zonal failure.

The below services will be deployed on the Metadata Tier.

- SAS Metadata Server
- SAS Environment Manager Agent

4.5.5.2 Compute Tier

The Compute Tier provides SAS analytic and business processing to requests initiated by the client and middle tiers. The below activities are performed on the Compute Tier.

- The Compute Tier is distributed using Grid Manager provided by IBM LSF where the computing load is distributed to be executed across multiple machines.
- LSF requires the SAS configuration used by the Compute Tier to be present on a shared drive. Filestore is used as the shared FS for the SAS configuration.
- Consists of Grid controller nodes (primary and secondary) and Grid worker node(s).
- The data sources consumed by the SAS processes are configured on the Compute Tier.
- The Data Layer (MDL, Staging & Teams) is mounted on all nodes.

4.5.5.2.1 SAS Services

The below services will be configured and started on the Grid Primary Controller node.

- SAS Web Infrastructure Platform Data Server
- SAS OLAP Server
- SAS Object Spawner
- SAS CONNECT Spawner
- SAS DIP Job Runner Server
- SAS Cache Locator
- SAS Environment Manager Agent

The remaining Grid nodes and the Grid secondary nodes will be running only the below services.

- SAS Object Spawner
- SAS Environment Manager Agent.

4.5.5.2.2 SPDS Services

The SPDS Services are configured to run on the Grid Controller Nodes (Primary & Secondary) and all the Grid Worker Nodes. The SPDS software has been configured as an individual service on all the Grid nodes where the configuration is local to each node instead of a shared configuration as it is currently setup on On-Premises. This is due to a limitation on the Filestore storage service where the SPDS on a shared configuration is not working as expected and the recommendation from SAS is to have the SPDS configured locally for each node.

4.5.5.2.3 IBM LSF

LSF is deployed on the Shared FS where the shared SAS configuration is deployed as this is a pre-requisite.

The below LSF service deamons will be running on the Grid Controller Nodes.

- mbatchd (Master Batch Daemon)
- LIM (Load Information Manager)
- PEM (Process Execution Manager)
- sbatchd (Slave Batch Daemon)
- RES (Remote Execution Service)
- VEMKD (Virtual Execution Manager Kernel Daemon)

- EGO (Enterprise Grid Orchestrator)

The LSF service daemons will be running on all the Grid Worker Nodes.

- sbatchd (Slave Batch Daemon)
- RES (Remote Execution Service)

4.5.5.2.3 IBM Process Manager

The Process Manager server is installed and configured on the Grid Primary Controller Server on a Regional Persistent Disk except for PRD-Prod environment where the Process Manager software is installed on the Grid Secondary Controller Server on a Regional Persistent Disk.

4.5.5.3 Middle Tier

The Middle Tier enables users to access intelligence data and functionality with a web browser. This tier provides web-based interfaces for report creation and information distribution, while passing analysis and processing requests to the SAS servers.

SAS Middle-Tier Application will be deployed on 2 nodes to provide load balancing and high availability. This is achieved using Google Cloud Native Load Balancer.

The below services will be deployed on the Middle Tier nodes.

- SAS Web Server
- SAS JMS Broker
- SAS Cache Locator Server
- SAS Environment Manager Server
- SAS Environment Manager Agent
- SAS Server1
    - SAS BI Web Services
    - SAS Web Infrastructure Service
    - SAS Preferences
    - SAS Shared Services
    - SAS Logon Service
    - SAS Stored Process Service
    - SAS Workflow
    - SAS Theme Designer
    - SAS Identity Services
    - SAS Authorization Service
    - SAS Theme Content Service
- SAS Server2
    - SAS Studio
    - SAS Environment Manager
    - SAS Web Documentation Service
- SAS Server11
    - SAS Enterprise Miner
- SASServer14
    - SAS Platform Web Services

4.5.5.4 Client Tier

The Client Tier is where the SAS tools are install

ed on the users' machine which they use to connect to the SAS Platform to perform various activities such as administering the SAS platform, submitting processes to execute on the Compute Tier, viewing the reports, etc.

The below client tools will be deployed on the user's devices bases on the requirements.

- SAS Enterprise Guide (All Users)
- SAS Management Console
- SAS Data Integration Studio
- SAS Enterprise Miner
- SAS OLAP Studio
- SAS Add-In for Microsoft Office
- Flow Manager
- Calendar Editor

4.5.5.5 SAS Installation

Below table provides SAS installation with environment wise installation documents, along with the start-up activities(manual and automatic both):

Environment	Document
PRD-PRD	SAS Prod-Prod Deployment Documentation - Data Migrations - Lloyds Banking Group Confluence
PRD-CIT	SAS Prod-CIT Deployment Documentation-2025 - Data Migrations - Lloyds Banking Group Confluence

Environment	Document
PRD-UAT	SAS Prod-UAT Deployment Documentation-2025 - Data Migrations - Lloyds Banking Group Confluence
PRD-DEV	SAS Prod-Dev Deployment Documentation-2025
PRE-DEV	Lemans - Pre-Dev SAS 9.4 - Data Migrations - Lloyds Banking Group Confluence

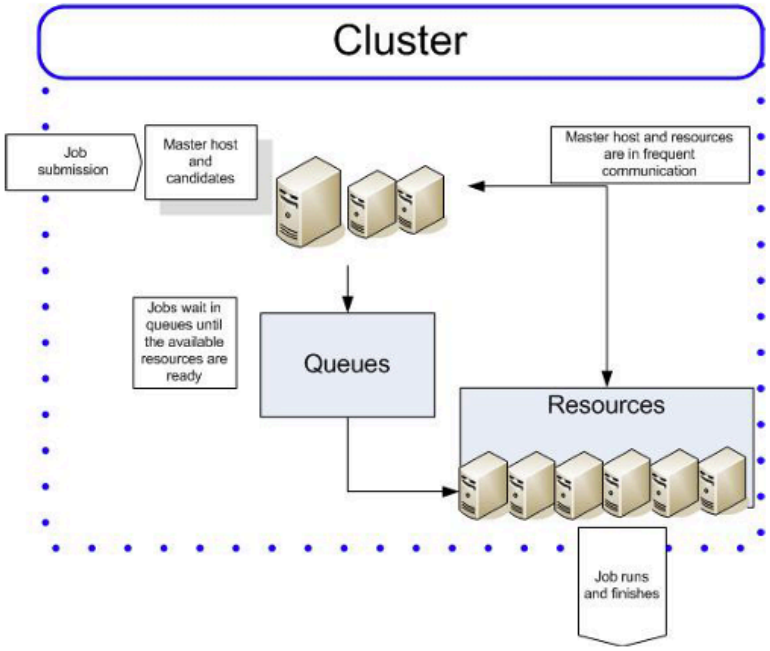
Generally, if we have to start and stop SAS services, the steps are documented in Prod_Prod_SAS_Services start/stop Document: - Data Migrations - Lloyds Banking Group Confluence for PRD-PRD environment.

4.5.6 Platform LSF and Process Manager

4.5.6.1 Platform LSF

The Platform LSF ("LSF", short for load sharing facility) bundled with SAS license is a software that distributes work across existing heterogeneous IT resources to create a shared, scalable, and fault-tolerant infrastructure, that delivers faster, more reliable workload performance and reduces cost. LSF balances load and allocates resources, and provides access to those resources. LSF provides a resource management framework that takes your job requirements, finds the best resources to run the job, and monitors its progress. Jobs always run according to host load and site policies.

The LSF software is deployed as a cluster on the Compute Tier where the jobs are submitted and needs to be distributed to the nodes based on the load.



The LSF cluster consists of the following components.

- **Master Batch Daemon** - The master batch daemon is the service which accepts the incoming job submission requests and directs them to one of the worker nodes which are available and less utilized.
- **Batch Daemon** - This daemon will accept the submitted job and execute it on the corresponding host. This service is configured to run on all the Grid Worker nodes.
- **Queues** - Any jobs submitted to the Master daemon will be assigned to the corresponding queue based on the configuration and the queue will assign the job to the worker node when it is ready and available.
- **Resources** - The Grid Nodes are grouped into resources where the jobs can be submitted for execution. More than one resource group can be created to handle different types of job executions.

The below list provides high level summary of the configurations to be setup on the LeMans GCP platform.

- LSF software should be deployed on Google Cloud Filestore which is a shared storage solution. This is a requirement for LSF to be able to distribute the jobs across all the Grid nodes.
- Master Batch Daemon should be configured to run on the Grid Primary and Secondary Controller server. In case of Grid Primary host failure, the Daemon will start running on Grid Secondary Controller node thus providing High Availability to continue to accept incoming jobs and submitting to the slave batch daemons.
- Slave Batch Daemon should be configured to run on all the Grid Compute Nodes.
- Queues for each type of activity & user to be created and assigned the corresponding Grid Nodes.

The LSF configuration on the LeMans GCP Prod environments will be replicated where possible from the existing LeMans On-Premise configurations. The configurations that are currently setup in On-Prem is captured in the Confluence page LeMans Custom Configuration Changes - Data Migrations - Lloyds Banking Group Confluence

Custom configurations for the LeMans GCP Platform to be setup is captured in the Confluence page LeMans LSF Configuration

4.5.6.2 Platform Process Manager

Process Manager is a workload management tool that allows users to automate their business processes in UNIX and Windows environments. Process Manager provides flexible scheduling capabilities and load balancing in an extensible, robust execution environment. Using the Process Manager Client, users can create and submit complex flow definitions to Process Manager Server, which manages the dependencies within a flow and controls the submission of jobs to the IBM Platform LSF(LSF®) master host. LSF provides resource management and load balancing, and runs the jobs and returns job status to the Process Manager Server. From Process Manager Client, users can also monitor and control their workflows within Process Manager.

The system is made up of the following components:

- Process Manager Server - The Process Manager server is where the flows are scheduled for execution.
- Process Manager Client - It consists of the Process Manager clients (Flow Manager & Calendar Editor) which will be installed on the user's device such as Desktop, Laptop or VDI.
    - Flow Manager - This client is used to manage the flows which are scheduled for execution such as monitoring, rerunning failed flows, killing flows/jobs, ad-hoc execution, etc.
    - Calendar Editor - This client tool is used for creating and managing the system calendars and custom calendars which can be provided as in put for time based triggers when scheduling flows.

The Process Manager server is installed and configured on the Grid Primary Controller Server on a Regional Persistent Disk except for PRD-Prod environment where the Process Manager software is installed on the Grid Secondary Controller Server on a Regional Persistent Disk. There is no fail over option enabled for this service as this cannot be stored on the Filestore which is the shared storage solution available for use in LeMans application. This is due to the security limitations on Filestore where root squash is enabled. And for Process Manager to work properly, root squash should be disabled as per SAS recommendation. In case of host failure, the Process Manager configuration stored on the RPD needs to be mounted on another Grid node and the configuration needs to be updated across PM and SAS to point to the new Grid nodes before the service can be brought up again for scheduling flows and executing them. A testing done for failover and recovery of PM is documented in Process Manager Failover Testing & Fallback Testing - Data Migrations - Lloyds Banking Group Confluence

Deployment and Custom configurations specific for the LeMans on GCP is detailed in the Confluence page LeMans GCP - Platform Process Manager - Data Migrations - Lloyds Banking Group Confluence

4.5.7 SAS 94 Deployment

The high-level diagram for each environment represents the key components that make up the SAS 9.4 infrastructure deployed into the LBG tenant on GCP.

> **Note**
> The individual components and their sizing information for each of the environment including the VM type, size, disks, Filestore components, etc. are captured in the Environment Sizing document. Please refer to this Confluence page for detailed information on all the environments LeMans Environment Compute Engine & Disk Details - Data Migrations - Lloyds Banking Group Confluence

4.5.7.1 SAS94 – Dev Environment

The above diagram for Dev environment represents the key components that make up the SAS 9.4 infrastructure deployed into the LBG tenant on GCP which are summarized below.

- 3 Metadata Nodes which are clustered to provide High Availability within one zone
- Grid Primary Controller Server where SAS Services, LSF Master Daemon, Process Manager Services are running
- Grid Secondary Controller Server acts as a failover host for the LSF Master Daemon services
- 2 Grid Worker Nodes where the LSF services are running the submitted jobs.
- 2 Mid-Tier nodes containing the SAS Web Applications.
- Network Load Balancer to distribute the load between the Mid-Tier servers. This also acts as High Availability service
- Filestore for SAS Binaries and Configuration for Compute Tier.
- Filestore to host the MDL Data & Teams' data
- GCS Storage to store the archived data

4.5.7.2 SAS94 – CIT Environment

The above diagram for CIT environment represents the key components that make up the SAS 9.4 infrastructure deployed into the LBG tenant on GCP which are summarized below.

- 3 Metadata Nodes which are clustered to provide High Availability within one zone
- Grid Primary Controller Server where SAS Services, LSF Master Daemon, Process Manager Services are running
- Grid Secondary Controller Server acts as a failover host for the LSF Master Daemon services
- 2 Grid Worker Nodes where the LSF services are running the submitted jobs
- 2 Mid-Tier nodes containing the SAS Web Applications
- Network Load Balancer to distribute the load between the Mid-Tier servers. This also acts as High Availability service
- Filestore for SAS Binaries and Configuration for Compute Tier
- Filestore to host the MDL Data & Teams' data
- GCS Storage to store the archived data

4.5.7.3 SAS94 – UAT Environment

The above diagram for UAT environment represents the key components that make up the SAS 9.4 infrastructure deployed into the LBG tenant on GCP which are summarized below.

- 3 Metadata Nodes which are clustered to provide High Availability within one zone
- Grid Primary Controller Server where SAS Services, LSF Master Daemon, Process Manager Services are running
- Grid Secondary Controller Server acts as a failover host for the LSF Master Daemon services
- 2 Grid Worker Nodes where the LSF services are running the submitted jobs.
- 2 Mid-Tier nodes containing the SAS Web Applications.
- Network Load Balancer to distribute the load between the Mid-Tier servers. This also acts as High Availability service
- Filestore for SAS Binaries and Configuration for Compute Tier.
- Filestore to host the MDL Data & Teams' data
- GCS Storage to store the archived data

4.5.7.4 SAS94 – Prod Environment

The above diagram for Production environment represents the key components that make up the SAS 9.4 infrastructure deployed into the LBG tenant on GCP which are summarized below.

- 3 Metadata Nodes which are clustered to provide High Availability within one zone
- Grid Primary Controller Server where SAS Services, LSF Master Daemon, Process Manager Services are running
- Grid Secondary Controller Server acts as a failover host for the LSF Master Daemon services
- 4 Grid Worker Nodes where the LSF services are running the jobs submitted by users.
- 16 Grid Worker Nodes which executes the jobs submitted via batch.
- 2 Mid-Tier nodes containing the SAS Web Applications.
- Network Load Balancer to distribute the load between the Mid-Tier servers. This also acts as High Availability service
- Filestore for SAS Binaries and Configuration for Compute Tier.
- Filestore to host the MDL Data & Teams' data
- GCS Storage to store the archived data

4.5.8 SAS Configuration

Build Configuration
SAS binaries and configurations implemented in LeMans are documented in LeMans SAS Environment Build Details - Data Migrations - Lloyds Banking Group Confluence

4.5.8.1 SAS Metadata Configuration

Several aspects of the SAS 94 configuration will be setup to reflect the existing on-premises environment.

The configurations to be applied on the target environment is defined in the Confluence pattern below.

LeMans Custom Configuration Changes

4.5.8.2 SAS Middle Tier Server Configuration

The Middle Tier servers contains the Web Server and Web Application Servers. While creating the plan for the Middle-Tier deployment, multi-machine option should be selected so that the middle-tier can be deployed on 2 servers.

The Web Server and Web App Server should be configured with TLS 1.2 while performing the deployment. Necessary certificates should be created using LBG's Internal Certificate Authority and provided during the Installation.

Additionally, the Network Load Balancer should be created and configured to point to the web servers running on the middle-tier. Load balancer is configured with two mid-tier servers. Each server contains Web Server and Web App Server. The request from load balancer goes to Web Application Servers running on both nodes behind load balancer. Web Server has internal routing to Web App Server so it can send request to Web App Server running on either nodes,(mid tier 1 or mid tier 2). If one of the Web Servers are inaccessible, the load balancer will redirect the request to another.

The Middle Tier servers contains the Web Server and Web Application Servers. While creating the plan for the Middle-Tier deployment, multi-machine option should be selected so that the middle-tier can be deployed on 2 servers.

The Web Server and Web App Server should be configured with TLS 1.2 while performing the deployment. Necessary certificates should be created using LBG's Internal Certificate Authority and provided during the Installation.

Additionally, the Network Load Balancer should be created and configured to point to the web servers running on the middle-tier. Load balancer is configured with two mid-tier servers. Each server contains Web Server and Web App Server. The request from load balancer goes to Web Application Servers running on both nodes behind load balancer. Web Server has internal routing to Web App Server so it can send request to Web App Server running on either nodes,(mid tier 1 or mid tier 2). If one of the Web Servers are inaccessible, the load balancer will redirect the request to another.

4.5.8.2.1 Flow of Communication

1. Client Request User sends a request(.e.g. via browser) to the internal-facing URL of load balancer
2. Load Balancer: It distributes the request to server A or server B, using health checks and routing algorithm of load balancer.
3. Web Server: On the selected server by load balancer, the HTTP server(Apache Web Server) receives the request. It acts as reverse proxy, forwarding the requests to internal SAS Web Application(e.g. SASStudio, Viya UI etc.)
4. Web App Layer: The SAS web applications handle the business login and responds to the client via the same path - Web App → Web Server → Load Balancer → Client
5. Session affinity is enabled on the internal load balancer level as per SAS recommendation.
6. Sticky sessions are enabled by default when we build SAS mid-tier with load balancing mode.

4.5.8.2.2 Two levels of Load Balancing

4.5.8.2.2.1 Web Server Load Balancing

Clients(e.g. browsers) send requests to a load balancer in front of web servers. Each web server is identical and has reverse proxies setup to web applications.

4.5.8.2.2.2 Load Balancing between Multiple App Servers

Each web server forwards the requests to backend web application servers(WAS). If both server A and Server B are running web applications, web server can balance requests among them too.

The configuration for the Middle-Tier servers on the LeMans GCP platform is defined in the Confluence page below.

LeMans SAS Environment Build Details - Data Migrations - Lloyds Banking Group Confluence

There is an ongoing investigation for the expected functionality of load balancer, as part of which session affinity is enabled and configurations are updated. Please refer to Mid-Tier Internal Load Balancer SAS Investigation - Data Migrations - Lloyds Banking Group Confluence

ILB configuration steps are part of installation documents for each environment, e.g. SAS Prod-UAT Deployment Documentation-2025 - Data Migrations - Lloyds Banking Group Confluence

4.5.8.3 SAS Compute Server Configuration

4.5.8.3.1 SAS Batch Server Configuration

The SAS 94 configuration for the Batch Server is defined in the Confluence pattern below.

LeMans Server Connections - Batch Server

4.5.8.3.2 SAS Workspace Server Configuration

The SAS 94 configuration for the Workspace Server is defined in the Confluence pattern below.

LeMans Server Connections - Workspace Server

4.5.8.3.3 SAS Night Queue Configuration

The schedule queues that is required for the ad-hoc users to submit their processes/jobs overnight is detailed in the Confluence pattern below.

LeMans Server Connections - Night Queue

4.5.8.3.4 SAS Schedule Queue Configuration

The LeMans schedule queue is a batch process for users of LeMans. It provides the users with the functionality to schedule their scripts (.sas) to run for a designated amount of time and frequency. It has the functionality to include dependency checks against any required source tables (in team folders and the MDL).

The detailed steps are provided in the Confluence pattern below.

LeMans Server Connections - Schedule Queue

4.5.8.4 External Databases

SAS Compute resources requires access to the external databases listed below to process the data from these data sources. . All the connections to and from the data sources to the LeMans application needs to be encrypted with TLS1.2 protocol.

External Database
Oracle DB
Teradata
Microsoft SQL DB
Google BigQuery

LeMans - External Connections capture the status of the different Database Connections that is configured on the LeMans Production deployment on the GCP Platform.

4.5.8.5 Internal Databases

The SAS Web Infrastructure Platform Data Server will be used to store application data:

Component	Detail
Administration	SAS internal Postgres (default)
Environment Manager	SAS internal Postgres (default)

The internal databases are necessary to hold the SAS application state.

4.5.8.6 Ports

The below table contains the base SAS ports will be used for the 9.4 deployment.

Server	Port Number	Description
Metadata Server	8561	Metadata Service
Metadata Server	5660	Deployment Agent Service
Metadata Server	2144	Environment Agent Service
Grid Controller Node	9432	WIP Database Service
Grid Controller Node	1966	Process Manager Service
Grid Controller Node	41415	SAS Cache Locator Service
Grid Controller Node	7551	SAS Connect Service
Grid Node	8591	Object Spawner Service
Grid Node	8581	Workspace Server Service
Grid Node	5400	SPDS Service
Grid Node	8601	Stored Process Service
Mid-Tier Node	8443	Web Application Service
Mid-Tier Node	7343	Environment Manager Service

These ports are necessary to ensure network traffic and access is unimpeded. Amendments can be made if port conflicts arise, however as each server is dedicated for SAS use, this is unlikely to be required. The detailed list of ports that is

configured for SAS 94 and Filestore application is documented in the Network Design Confluence page below. This includes the firewall rules that needs to be defined on each of the RTL environments.
SAS 9.4 - Network Requirements - Data Migrations - Lloyds Banking Group Confluence

Note: The Port number defined above are for all the Business RTL environments which will be configured as Lev1. If the environments are configured as Lev1, Lev2, Lev3 & Lev4 for each of the Business RTL environment, then the port numbers will reflect the corresponding Lev's.

4.5.8.7 Folder Structures

The folder structures defined in the LeMans SAS 94 environment will be different for the SAS Binaries and Configurations compared to the On-Premises environment whereas the folder structures for the data layer and team's folder area will remain the same to make sure the paths referenced in metadata libraries and codes are not impacted.

The following table defines the folder structures for the SAS Binaries & Configurations.

SAS component	Path
SAS Home	/opt/sas/sashome
SAS Config	/opt/sas/sasconfig
LSF	/opt/sas/lsf
Process Manager	/opt/pm
Workspace Server	/opt/sas/sasconfig/Lev1/SASApp/WorkspaceServer
Stored Process Server	/opt/sas/sasconfig/Lev1/SASApp/StoredProcessServer
Object Spawner	/opt/sas/sasconfig/Lev1/ObjectSpawner
Backup Folder	/opt/sas/backup
SAS Work & Util	/saswork
SPDS Work & Util	/spdswork
SAS Grid work	/opt/sas/gridwork
SAS Logs	The Locations of each SAS server Component are under the following location: • Metadata Logs Location: /user_logs/SAS94/prod/MetadataServer • ObjectSpawner Logs: /user_logs/SAS94/prod/Objectspawner • ConnectSpawner Logs : /user_logs/SAS94/prod/ • PooledWorkspaceServer Logs: /user_logs/SAS94/prod/PooledWorkspaceServer • BatchServer Logs : /user_logs/SAS94/prod/BatchServer • WorkspaceServer Logs :/user_logs/SAS94/prod/WorkspaceServer • StoredProcessServer Logs: /user_logs/SAS94/prod/StoredProcessServer • SPDS Logs :/user_logs/SAS94/prod/SPDS • SAS EVM Agent logs: /user_logs/SAS94/prod/SASEnvironmentManager/agent • GridServer Logs : /user_logs/SAS94/prod/GridServer • SAS Metabacakups:/backup/Metabackups/ Mid Application log locations are under the default directories only. SAS Help Center: Middle-Tier Logs and Log Locations

Note: The Lev number can be set to Lev1 for all Business RTL environments or can be set similar to the configuration on On-Premises environments. Regardless of the definition, the resources will be isolated for each environment.

4.5.9 SAS Client Tools

Desktop Applications or Thick clients will be used are SAS Enterprise Guide, SAS Management Console, SAS DI Studio, SAS Enterprise Miner, SAS OLAP Studio and SAS Add-in for Office. Packaging team will be creating packages for each of the client application which will replace the older version of the clients on user's devices. Thin Clients on the browser will provide access to SAS Studio, SAS Environment Manager.

Details of Thick and Thin Clients used in LeMans are available in LeMans SAS Environment Thin and Thick client

4.5.10 Scheduling

Platform Suite for SAS can be used as an add-on for SAS DI Studio to provide enterprise scheduling capabilities on a single-server deployment, such as LeMans. This will replace Platform Process Manager shipped with the Grid environments. This provides the ability to continue to schedule the execution of the existing DI processes within the new 9.4 deployment.

4.5.11 Storage

The SAS Server storage will be hosted on Regional Persistent Disks which will present each file system to the SAS Servers.

Best practices for Storage Configuration will be implemented based on SAS recommendations.
Source: http://support.sas.com/resources/papers/proceedings16/SAS6761-2016.pdf paper which covers “Best Practices for Configuring Your I/O Subsystem for SAS9 Applications”. In addition, SAS provides scripts that can be used to measure the IO throughput obtained across various file systems in an existing platform. This can be of use to capture a baseline note http://support.sas.com/kb/53/876.html

The following table provides the details of storage required for the SAS environment.

Storage Volume	Storage Device Type	Size	Mount Point	File System Type
SAS binaries & config (Metadata and Mid Tier)	Regional Persistent Disk	100 GB	/opt/sas	xfs
Process Manager	Regional Persistent Disk	100 GB	/opt/pm	xfs
LSF, SAS binaries & config (Compute Tier)	Filestore Shared FS	10 TB	/opt/sas /opt/sas/logs /opt/sas/backup	nfs
SASWORK & UTIL Location	Persistent Disk	3.7 TB	/saswork	xfs
MDL Data	Filestore Shared FS	280 TB	/sasdata	nfs
Team's Folders	Filestore Shared FS	310.75 TB	/teams	nfs
staging	Filestore Shared FS	10 TB	/staging	nfs
SAS Depot	GCS Buckets	NA		

There are 4 groups of file systems in use by SAS, each with their specific characteristics. More details on environment sizing and project detail are available on LeMans - Project Details - Data Migrations - Lloyds Banking Group Confluence and LeMans Environment Compute Engine & Disk Details - Data Migrations - Lloyds Banking Group Confluence

4.5.12 Patching

SAS provides updates that addresses the software issues identified on the application called HotFixes. These HotFixes provides solution to the problems identified on the SAS software and is fully supported by SAS. Technical Support's hot fix focus is on solving critical and frequently recurring problems. Defects that are assigned an "alert" or "high" priority are considered for hot fixes. If the fix can be created with a reasonably low impact on the source code and testing coverage can confirm accurate results, the hot fix is pursued.

4.5.12.1 Schedule

It is recommended to be proactive by keeping your SAS installation updated with all of the latest fixes every 6 months. However, in the case of any critical alert that needs addressed immediately, it is recommended to apply it as soon as possible. The hotfixes should follow the Route-To-Live process by deploying the fixes on the PRE-DEV environment and then progressing on to the PRD-DEV, PRD-CIT, PRD-UAT and finally on PRD-PROD environment to make sure the patches are tested thoroughly before being deployed on Prod.

If any issues are encountered when deploying the HotFixes, a case needs to be raised with SAS Technical Support to address the issue.

4.5.12.2 Deployment of HotFixes

The process to deploy the hotfixes on the SAS application remains the same as the current process that is followed on LeMans On-Prem environment. Additionally the instructions for deploying the hotfixes will be provided as part of the SAS notes provided with the hotfix.

4.6 Data Architecture

4.6.1 Filestore Storage Architecture

Filestore instances are fully managed file servers on Google Cloud that can be connected to a number of client types, Compute Engine being the client used in LeMans. As a type of persistent file storage, Filestore supports multiple concurrent application instances accessing the same file system simultaneously. It is a Google managed high performance storage solution which is deployed as Platform-as-a-Service. The Filestore Storage Solution is based on NFS v3 protocol.

Google Filestore is to be provisioned to meet the requirements of LeMans for storing SAS data (both SPDS and SAS Datasets) in GCP. This reduces the complexity of support and meets the performance requirements in I/O testing completed on GCP instances replicating the Production size. Filestore will provide replacement of the current shared filesystem used on LeMans - IBM Spectrum Scale. The current LeMans application leverages a filesystem of roughly 400TB holding the SPDS and Team Folder data across all Production, and RTL environments. This solution is mirrored between two datacentres and is configured to run active-active, providing the LPARs with a consistent view of the SAS data within.

On-Premises LeMans solution uses IBM Spectrum Scale as a high performance storage solution where the data is stored for the MDL Layer and User Data folders. Each MDL Layer has different mount points created within Spectrum Scale and the corresponding library definitions were configured in SAS application. Each Library definition will contain 3 folders namely metadata, index and data. This is based on how SPDS libraries store the dataset.

Example Library definition:

IBM Spectrum Scale - SPDS Library Definition

```
libname=FDAFREP pathname=/spds/fdl/metadata/fdafrep
```

```
roptions="datapath=('/spds/fdl/data/fdafrep')
          indexpath=('/spds/fdl/index/fdafrep')" OWNER=bicbatch ;
```

In GCP, LeMans will be using Filestore Instances as a high performance shared storage solution to store the MDL data layer which consists of around 250 TB of data. Since Filestore has a limit of 100 TB per instance, 4 100 TB instances will be deployed to store the MDL data layer. Each Library definition is split in to 3 areas namely metadata, index and data and separate paths have been configured on On-Premise LeMans solution.

In order to improve performance, all the MDL library definitions will be modified to include all the 4 filestore instances which will be used as a round-robin storage that will be used by the SPDS libraries.

Example Library definition for Production:

Filestore - SPDS Library Definition

```
libname=FDAFREP pathname=/spds1/fdl/metadata/fdafrep
roptions="datapath=('/spds1/fdl/data/fdafrep' '/spds2/fdl/data/fdafrep' '/spds3/fdl/data/fdafrep
          indexpath=('/spds1/fdl/index/fdafrep' '/spds2/fdl/index/fdafrep' '/spds3/fdl/index/fda
```

Please refer to Filestore - MDL Library Definitions for more details.

4.6.1.1. Limitations

Although Filestore Instance is a high performance storage, there are still limitations on how the Filestore can be used within the LeMans solution. The Limitations are described below.

4.6.1.1.1 Restriction on Sizing

The Filestore Instances has a limited capacity for each tier as mentioned below. Also, the scalability of Filestore Instances is fixed in size. Due to this limitation, we need to uplift the storage allocated to some of the teams to meet the minimum storage requirement which will increase the total storage allocated to the LeMans solution.

Filestore Tier	Min Size	Max Size	Scalability
Zonal Low-Capacity	1 TB	9.75 TB	Up or down in 256 GB units
Zonal High-Capacity	10 TB	100 TB	Up or down in 2.5 TB units

4.6.1.1.2 Performance Limitation

The performance of the Filestore Instances varies based on which Tier is being used to store the data. In LeMans, we will be using High-Capacity tier exclusively for MDL Data layer, but for Teams folders a combination of Zonal High-Capacity and Low-Capacity tier based on the size required for each team. This has an impact on performance and each team might get different read/write throughput based on which tier their Filestore Instance is based on.

The performance throughput for the Filestore Tiers is provided below.

Filestore Tier	Read Throughput	Write Throughput	READ IOPS	WRITE IOPS
Zonal Low Scale	260 MiB/s per 1 TiB	88 MiB/s per 1 TiB	9,200 per 1 TiB	2,600 per 1 TiB
Zonal High Scale	650 MiB/s per 2.5 TiBs	220 MiB/s per 2.5 TiBs	23,000 per 2.5 TiBs	6,500 per 2.5 TiBs

4.6.1.1.3 Group Limitation

NFS server has a limitation on how many groups a user is member of. The problem occurs when a user, who is a member of more than 16 groups, tries to access a file or directory on an NFS mount that depends on his group rights in order to be authorized to see it. The default authorization mechanism for NFS (auth_sys) will take only a subset of your groups and send it to the NFS server to check if you have rights to read a file. This leads to unpredictable and intermittent permission problems when it looks like you *should* have permission.

Since Filestore uses NFSv3 as a storage mechanism, a ticket has been raised with Google to check if Filestore might be affected by this limitation. Google responded that the current version of Filestore will be affected by this limitation. Google confirmed that this issue should be resolved when Filestore is upgraded to NFSv4.1 by Google however there is no ETA at the moment on when Filestore NFSv4.1 will be available for General use.

4.6.1.1.4 Quotas

One of the requirements from the LeMans Teams Data Layer is to have individual spaces allocated to the 55 teams who is currently using the LeMans solution.

With Filestore however, we will not be able to deploy a single Filestore Instance and dedicate ring fenced storage to the teams. So a decision has been taken to deploy individual Filestore Instances to each of teams with the required size allocated to them.

4.6.1.1.5 Large IP Address range

Although each Filestore Instance will be accessed from the client machines with a single IP address, Filestore requires a minimum of 64 IP address to be allocated to it to manage one single instance. This is a limitation as the requirement is to have 64 Filestore Instances deployed on the Production environment which means 4096 IP addresses needs to be allocated to the Filestore Instances. This is a concern and needs to be addressed with the help of Network team.

4.6.2 Block Storage for Compute Engine

We are using durable block storage. Durable, or persistent, block storage is for data that you want to preserve after you stop, suspend or delete the VM, or even if the VM crashes or fails.

Zonal Persistent Disk is used to store compute data for SPDSWORK, SASWORK, SASUTIL, and additional configuration of VM such as logs, tmp storage etc.

Regional Persistent Disk(RPD) provides high availability and can be used for disaster recovery if an entire data center is lost and can't be recovered. SAS components such process manager, metadata tier configuration, mid-tier configurations, Connect Direct application and other SAS binaries will be stored in Regional PD.

While provisioning the VM, we are enabling disk snapshot for boot disk and additional disks which are zonal in nature. By enabling this feature, snapshots can be taken at regular intervals and used to restore a VM if needed. In our environment, we have set different retention policies for each environment. For the integration environment, the retention policy is 7 days, while for the pre-production and production environments, it is 7 days. This snapshot policy will take snapshots every 12 hours, with the first snapshot taken at 11:00 PM. There is no need to have more than 7 days on all environments as the password for the AD account which binds the VMS to Active Directory is changed every 6 days and the backups which are older than 6 days cannot be used.

```
workstream-compute--template repo


## Snapshot Variables
add_snapshot_flag           = true  #Flag to indicate if snapshot has to be added on boot and ad
snapshot_hours_cycle        = "12"
snapshot_start_time         = "23:00"
snapshot_max_retention_days = "7"
snapshot_vss_aware          = false
```

Details of Persistent disks and their mount points for PRE, and PRD environments are documented in LeMans Environment Compute Engine & Disk Details - Data Migrations - Lloyds Banking Group Confluence. The details of mount point and disk can be found in Terraform code and Google Cloud Console.

4.6.3 Google Cloud Storage for Archived Data

Cloud Storage is a service for storing objects in Google Cloud. In LeMans, it is used to store historical data migrated through Data Transfer Appliance(covered in later sections) and archived from Filestore as per archival and retention process defined later in the Resiliency section of this paper. Steps to create and maintain GCS buckets are given in Bucket creation for Lemans. Features like CMEK encryption, object versioning, IAM access control are enabled against the buckets. Please refer to LeMans - Retention Policy for lifecycle rules enabled on the buckets.

4.6.4 Data Transfer Appliance

LeMans migration team will be transferring the data from the On-Premises LeMans servers to the GCP LeMans servers. This includes various data use cases, migration of historical data, processes across RDL-CDL-FDL layer

Due to the volume of data that is transferred as part of this work, Google's Data Transfer Appliance will be used as the network bandwidth does not support transferring large amounts data over the Interconnect.

4.6.4.1 Component Design

The Component Design for DTA has been carried out as a separate activity and has been approved to use for Release 2.

Document type	Confluence details
DTA Design Document	LeMans Data Transfer using DTA
Approved CODA Documents	/wiki/spaces/CODAVLT/pages/165094245 /wiki/spaces/CODAVLT/pages/165094268 Please note , CODA documents are stored in a Vault and need to raise a separate request to access this. The document is locked for view.
Rsync Patterns	LeMans - SSH & RSYNC access for Migration

4.6.5 Data Classification

LeMans is classified as storing Highly Confidential data include Teams and SPDS(MDL) data. As Filestore is a distributed file system, it is necessary to secure all information to the highest level, therefore all storage will be hardened to hold Confidential and Highly Confidential information. By default, Filestore automatically encrypts the data at rest. Although NFSv3 does not encrypt data in transit, all in-transit data to and within Google Cloud is private.

The SNC which support the risk of handling highly confidential data for Filestore for LeMans solution are-

Data at rest: SNC-2023-4414

Data in transit : SNC-2024-6114

4.6.5.3 Future considerations for handling HC data

Data in Transit - Upgrade to Filestore NFSv4.1 (still in preview state)	**Data at - Integrate SAS with CyberArk for Data Encryption**
NFSv4.1 provides support for encryption in transit, which is crucial for handling highly confidential data. By upgrading from NFSv3 to NFSv4.1, the data will be encrypted as it travel over the network.	CyberArk can manage and automate encryption for sensitive datasets. By integrating SAS with CyberArk, we can ensure that data is encrypted using AES standard both at rest and in transit, regardless of the underlying storage solution. Refer to SAS Help Center: ENCRYPT= Data Set Option for more detail on AES
Steps to Upgrade: • Ensure that the existing applications, including SAS, are compatible with NFSv4.1 • Migrate the existing Filestore data to a new Filestore instance that supports NFSv4.1. This might involve creating a new Filestore instance and transferring the data over. • Update the mount configurations on Compute Engine instances to use NFSv4.1	Integration steps: • Setup CyberArk for Encryption Management: Configure CyberArk to manage AES encryption keys and handle encryption/decryption processes. • SAS Integration: Modify the SAS application to interface with CyberArk via APIs to encrypt data before writing it to storage (Filestore) and decrypt data when reading. • End-to-End Encryption: Ensure that encryption is applied to all data exchanges, including those in transit.
Benefits: • Encryption in transit is natively supported, improving security without requiring additional changes to the data handling processes.	Benefits: • Centralized management of encryption policies. • Seamless integration with SAS, providing strong encryption without significant performance overhead.
Challenges: • Complexity : Application and environments currently using NFSv3 may face compatibility issues when transitioning to NFSv4.1 • Set up complexity: Setting up and managing Kerberos for secure NFSv4.1 can be complex and may require additional setup for maintenance	Challenges: • Complexity: The integration process can be complex, requiring custom development, particularly if SAS does not natively support direct integration with CyberArk. • Performance: Real-time key retrieval from CyberArk may introduce some latency, although this can be minimized with appropriate optimization.

4.6.6 Inter-Environment Data Transfer

GCP projects need data transfer mechanism for use cases like moving data from historical project to data projects, or to copy data from PRD-PRD to PRD-UAT. Currently two patterns are used, SAS connect and transfer via GCS buckets. A tech debt is noted in Future Roadmap section(AFR4) in this page to analyse the requirements and approve the pattern for inter-project data transfer.

LeMans GCP - Data Transfer to RTL Environments - Data Migrations - Lloyds Banking Group Confluence

4.7 Security Architecture

4.7.1 Security Design

The security design spans both the Google Infrastructure and the SAS Application and describes the inter-play between the two. Detailed security design created for the LeMans application is captured in the Confluence page below.

AA Review: LeMans Security Design - Data Migrations - Lloyds Banking Group Confluence

4.7.2 Authentication

4.7.2.1 Virtual Machines

4.7.2.1.1 End-User Authentication

The RHEL virtual machines deployed on GCP uses PAM based authentication against active directory and is integrated with the corresponding directory domain based on the type of environment as mentioned below.

GCP Environment	Active Directory Domain
INT	IAGLOBAL.LLOYDSTSB.COM
PRE	TEST01GLOBAL.LLOYDSTSB.COM
PRD	GLOBAL.LLOYDSTSB.COM

The virtual machines are built using the golden images provided by the DCX platform which will be integrated to the AD using the terraform tags provided. This will allow the VM's to register and integrate with the Active Directory domain by the way of chef cookbooks which executes during the build process.

The detailed process on how the VM's are joined with the Active Directory is provided in the Confluence page below.

TO - Domain Join - Tech Optimisation - Lloyds Banking Group Confluence

4.7.2.1.2 Service Account Authentication

The service accounts are non human accounts which are created to manage the application services, batch processes, etc. These are privileged accounts which has elevated permissions to control the application.

The service accounts are created in Active Directory and onboarded to the CyberArk vaults as per LBG policies which then can be used to access the VM's using SSH connection via the jumphost "psmproxy.service.group". Policies are applied against the different service accounts which dictate the password management such as password complexity, rotation, approval, etc..

The below personas will have access to the corresponding service accounts which can be used by the members to access the VM's using SSH.

Account	Persona	Account Description	Example(s)	Account Managed by
System	ADM	Several accounts are defined to install and run the SAS software across the server estate. • These account s are used to support spwaned sessions for the stored processes under the "sassrv" account. • Others are responsible for the application state, the SAS services are run as the "sas" account. • The first user account is "sasdemo" which has no authority but is used in post-install validation and health-checks. • The LSF Admin account is "lsfadmin" and "lsfuser" is for installing and configuring LSF and PM services	SRVAPPLEMSAS01 SRVAPPLEMSASSRV01 SRVAPPLEMLSFADMIN01 SRVAPPLEMLSFUSER01	Active Directory
Batch	AESM	Non-human accounts are used to execute the batch process run by the SAS scheduling solution. Circa 30,000 processes run per month for data ingestion and reporting within LeMans. Production and Route-to-Live accounts are separated	SRVAPPLEMBICBATCH01 SRVAPPLEMRTLBATCH02	Active Directory

Further details on the authentication is captured in the Confluence page User Access Management

4.7.2.2 SAS Application

4.7.2.2.1 End-User Authentication

The SAS application is configured to use PAM to authenticate users against Active Directory domain. SAS provided a supporting feature that extends UNIX host authentication to recognize an additional provider such as Active Directory. When a SAS server asks its UNIX host to validate a user's credentials, the host sends the user's ID and password to the configured additional provider for verification. PAM extends the host's authentication process to recognize an additional provider; PAM does not modify the Metadata Server's behaviour.

Along with the AD integration, additional configurations are setup to allow only members of specific Active Directory groups to access each of the SAS environments which operates based on the LPAU[Least Privileged Access] approach to followed across the GCP platform within LBG. More details on the specific AD groups that are configured for each SAS environment will be covered in the LLD design.

PAM configurations for LeMans are documented in LeMans SAS Environment Build Details - Data Migrations - Lloyds Banking Group Confluence

4.7.2.2.2 SAS Internal Accounts Authentication

The SAS Application comes with Internal accounts intended for only metadata administrators and some service identities. These internal accounts exist only in the metadata and can be created and managed in SAS Management Console.

Account	Persona	Account Description	Example(s)	Scope
Internal	AESM, ADM	Internal accounts exist within the SAS application and are used to administer functions within the metadata of the environment, provided to individuals with unique identifiers. They have no authority or access to data providing segregation of responsibility between SAS content and SAS metadata. These accounts are used in the promotion of content through the route-to-live, the configuration of security, and server functions, amongst other SAS administration tasks.	sasadm@saspw JDoe@saspw	Metadata

4.7.2.2.3 Service Identities Authentication

The service accounts are non-human accounts created to manage the application services, batch processes, etc. These accounts are created on the Active Directory and onboarded to CyberArk where the credentials for these accounts are managed. Authorized users can extract the credentials for the service accounts from the CyberArk portal and then can connect to the SAS application using the SAS Client software or SAS Web Portal.

Further details on the authentication is captured in the Confluence page User Access Management

4.7.2.3 Filestore

For Filestore, the existing authentication mechanism on RHEL is used to control access to the data area and Identity and Access Management (IAM) to control access to instance operations, such as creating, editing, viewing, and deleting instances.

4.7.2.4 Google Console

Access to the Google Cloud Console is enabled through the use of Federated accounts which can allow for Single Sign On (SSO) using the user's GLOBAL Domain Active Directory account.

4.7.3 Authorization

There are various levels of authorization to the LeMans platform based on the persona and the client that is used to connect and the levels of security at each layer is provided below.

4.7.3.1 Physical Security

To access the LeMans application on GCP, additional network security has been applied by the use of Forescout where only members of the corresponding AD groups are allowed to access the application. Forescout AD groups are created for each GCP environment and user's needs to be member of the corresponding groups.

GCP Environment	Forescout AD Group
BLD	NA
INT	GG_GCP_DEV_Developer_AL05559
PRE	GG_GCP_PRD_DEVELOPER_AL05559
PRD	GG_GCP_PRE_DEVELOPER_AL05559

Unless the users are part of this group, the LeMans application or the Infrastructure cannot be accessed from User's devices.

4.7.3.2 SAS Application Security

To access the SAS Application, the users needs to be part of the corresponding "SAS Users" Active Directory group which is created for each environment and based on the access requirements, the user's identity is setup on the SAS Metadata Repository and provided with the right access. This below list will only allow the users to access the SAS Application. Any additional access to the data layer requires the users be part of the corresponding AD groups. The AD groups setup for the data layer is captured in the sections 3.1.4 and 3.1.5 of the User Access Management Confluence page.

GCP Environment	SAS Users AD Group
BLD	NA
INT	iaglobal\GG_SAS_LEM_SASUSERS01
PRE	test01global\GG_LM_SAS_SASUSERS01, test01global\GG_LM_SAS_SASUSERS02, test01global\GG_LM_SAS_SASUSERS03, test01global\GG_LM_SAS_SASUSERS04
PRD	global\GG_LM_SAS_SASUSERS01, global\GG_LM_SAS_SASUSERS02, global\GG_LM_SAS_SASUSERS03, global\GG_LM_SAS_SASUSERS04

4.7.3.2 SPDS Security

All data within the /spds mount is owned by a single **sas** account, which is assigned to the **sasuser** group.

All physical directories and data is assigned a read-only access to the members of the sasuser group, which all LeMans users are members by default to access and start SAS sessions.

Directory Example	Business Owner	ID	Group	Pattern	Octal
/spds*N*		SRVLEMSAS01	GG_LEM_SASUSERS01	drwxr-s---	2750
/spds*N*/data/		SRVLEMSAS01	GG_LEM_SASUSERS01	drwxr-s---	2750
/spds*N*/data/rdl/rdcds/		SRVLEMSAS01	GG_LEM_SASUSERS01	drwxr-s---	2750

Note:

> **Note**
> All SPDS folders and libraries are owned by the same account and have the same permissions applied. The access to the data is governed by the SPDS ACLs, which are applied.
>
> Content is protected via this permission from being accessed or amended inadvertently as the **sas** account is restricted.
>
> There is an automated process which runs every night that creates the SPDS accounts for each of the user that is setup in the Metadata repository and the users gain access to the SPDS data via the SPDS permissions granted and maintained in the SPDS passmgr internal database.

SPDS ACLs

In production the SPDS ACLs are applied per domain to enforce BICBATCH as the owner of all data, with read-only permissions applied to the user community.

DEFAULT	*None*	*None*	R, W, A, C	-, -, -, -
Non-Sensitive	*None*	*None*	R, W, A, C	R, -, -, -
Confidential	SENSITIV	PII	R, W, A, C	R, -, -, -
Very Sensitive	VSENS	*None*	R, W, A, C	R, -, -, -
Highly Confidential	*None*	PSI	R, W, A, C	R, -, -, -
HCC	HCC	HCC	R, W, A, C	R, -, -, -

4.7.3.3 Google Cloud Console Access

Users' need to raise IAM requests to the individual Google projects to access the resources created within the project. The standard access provides read access to the cloud resources deployed within the project. If any elevated privileges are required to managed the cloud resources, JIT[Just in Time] access method is provided by the GCP IAM team to request the required roles to be added to the user's account for a specified duration with the right Justification on the back of a SNOW Change Request/Incident.

More details on requesting the JIT access and the allowed and denied roles are provided in the Confluence page 2.3 Privileged Identity Provisioning & Management - Tech Optimisation - Lloyds Banking Group Confluence

4.7.4 Users, Groups & Identities

The different personas required for the LeMans application and the Active Directory accounts required for the LeMans SAS 94 solution is detailed in the Confluence page LeMans User Access Management

4.7.5 Data in Transit Encryption

LBG provided CA signed certificates will be used to secure the SAS 94 environment. SAS supports TLS 1.2+ for Web Server and Web Application Server configuration. The certificates will be provided by LBG prior to the Deployment phase.

Environment	Connection	Secure (Y/N)	Details
All Environments	SAS Servers - Filestore	N	All data transferred between SAS and Filestore is accessed through private service. The Filestore is deployed within the Private Service Access subnet which is controlled by GCP and VPC peering b/w SAS Subnet and Filestore subnet. SNC raised for *Data in transit* : SNC-2024-6114 ; Associated risk is logged in - RK0045040
	SAS/SECURE – Client/Server Data Transfer	Y	AES Encryption Level will be chosen during SAS Configuration
	SAS Thin Client – SAS Web Server	Y	All data over the wire must be encrypted using TLS1.2 Site-signed certificates
	SAS EV Agents	Y	Self-signed generated by SAS Deployment Wizard
	SAS Web Server – SAS Web Application Server	Y	
	SAS Environment Manager Client – Web Server	Y	

4.7.6 Data at Rest Encryption

For all keys, these will be customer managed and secured in an appropriate customer vault.

Environment	Data Type	Algorithm/Encryption Level	Details
All Environments	Credentials	SAS005	Default SAS Encryption level will be used to

Environment	Data Type	Algorithm/Encryption Level	Details
			encrypt all credentials in SAS Configuration.
	Data at Rest	Customer Managed Encryption Key	Persistent Disks and Data stored in Filestore will be encrypted using Customer (LBG) provided encryption key. SNC raised for *Data in rest*: SNC-2023-4414 ; Associated risk is logged in - RK0045163

4.8 Resiliency

To design resiliency in the solution, we should ensure the solution can withstand failures, and recover quickly when needed. Below are the key resiliency principles including backup, disaster recovery(DR) and high availability(HA) aspects of the solution:

4.8.1 High Availability

High Availability can be achieved in the SAS 94 and Filestore solution using a combination of cloud native services and features provided within the application itself.

4.8.1.1 HA in SAS Application Infrastructure

SAS Solution provides High Availability by providing multiple VMs for each tier, replicating data or keeping the services underneath in sync. This provides ability for SAS application to be up and running if some nodes are affected. This however, does not provide resilience against zonal failure as all nodes are created under one zone.

Metadata Tier: The Metadata Cluster consists of minimum 3 machines where one machine is the master and the other two act a worker nodes. All nodes can server the requests to store metadata, including the master node. In case of failure on any one of the machines, the cluster continues to run without any impact to the users. However, if more than one machine goes down, then the cluster will not be able to accept client connections as it expect a minimum of 2 machines running at any given time.

Grid Compute Tier: The Grid Compute Tier consists of both the Grid Controller Nodes and Grid Worker nodes which can have high availability using LSF EGO (Enterprise Grid Orchestrator). There are two Grid controller nodes that keep polling each other, and if one of them goes down the other one is available to orchestrate the requests to worker nodes. Among worker nodes, managed instance groups should be created to auto-scale the VMs horizontally. In current configuration, 16 worker nodes are running and if some of them become unhealthy, the other can keep the serving the solution as per SAS Grid architecture. The nodes in this tier use Filestore as distributed storage, HA solution for which is described later.

Please note that managed instance groups curated with DCX are not fit for LeMans usage, reasons including the high start up time and inability for VMs to automatically detach from resources(e.g. Filestore) after shutdown.

Middle Tier: The Middle Tier consists of Web Application Services which is load balanced using a combination of internal configuration and also using the Cloud Internal Load Balancer provided by Google. This is already covered in section 4.5.8.2.

In addition to the Web Application services, we have configured HA on the Cache Locator and the JMS services. Documentation for Cache Locator and JMS broker below.

SAS Prod-UAT Deployment Documentation-2025 - Data Migrations - Lloyds Banking Group Confluence

SAS Prod-UAT Deployment Documentation-2025 - Data Migrations - Lloyds Banking Group Confluence

4.8.1.2 HA in Backup Solution

VMs used for the backup solution do not receive any user traffic. Recommendation is to use Managed Instance Groups(MIGs) to automatically scale automatically. Please note that usage of MIGs is rejected due to high start-up time of VMs, instead additional VMs are included in the standby configuration.

4.8.1.3 HA in GCP Native Services

Persistent Disks: We are using regional persistent disks for SAS binaries, and Process Manager. Regular snapshots of persistent disks are taken.

Load Balancer: We are using pass-through network load balancer to distribute incoming network traffic across multiple instances(GCE), ensuring that if one VM in unhealthy, traffic is routed to healthy instances.

Filestore: Recommended solution to achieve HA in Filestore is to use Regional Filestore that provides data replication across two zones. However, Bank has decided to use zonal Filestore interim(see Architecture Decision log AK35). With zonal Filestore, we are implementing data replication and backup strategies described later .

Cloud Storage: While cloud storage provides inbuild HA, we are using object versioning to restore previous versions.

4.8.2 Backup Strategy

4.8.2.1 Application Backup - Metadata Server Backup and Recovery Facility

Several aspects of the SAS application can be backed up to ensure business continuity and resilience. Metadata Server Backup Facility backs up all registered metadata repositories , the repository manager, and the metadata server's

configuration directory. The metadata backup will happen every night at 1 am and on Sundays, it will take backup and also reclaim the unused disk space to optimise storage.

We can use the SAS Management Console to recover the SAS Metadata Server. The following options are available:

- We can recover just the metadata repositories and the repository manager, or We can choose to recover the configuration files as well.
- We can recover from a backup that is listed in the backup history panel, or We can recover from backup files that are stored in an alternate network-accessible location.
- We can use the roll-forward recovery to apply metadata updates that are stored in the journal file.

The recovery facility provides safeguards to ensure the integrity of the backup files from which We are recovering. The recovery operation checks to make sure that the backup directory contains all of the correct files and that the files have the correct name and file sizes. In addition, each backup file contains a universal unique identifier that is used to make sure that We are recovering files for the correct metadata server. If any problems exist, the recovery is not started and a warning message is displayed.
During recovery operations, the metadata server is automatically paused to a RECOVERY state. This state is similar to an OFFLINE state but more restrictive. External systems such as Windows Services Manager report the server as paused. After the recovery, the metadata server automatically takes a new backup. If the recovery was successful, the server is automatically returned to the state that it was in before the recovery process.

4.8.2.2 Compute Engine Instance

While provisioning the VM, we are enabling disk snapshot for boot disk and additional disks. By enabling this feature, snapshots can be taken at regular intervals and used to restore a VM if needed. In our environment, we have set different retention policies for each environment. For the integration environment, the retention policy is 7 days, while for the pre-production and production environments, it is 30 days. This snapshot policy will take snapshots every 12 hours, with the first snapshot taken at 11:00 PM.

VMs are created using DCX provided custom image that comes with OS configuration and additional VM configuration is stored in Terraform state as the source of truth. There is no auto-healing and automatic failover enable for VMs including SAS and backup VMs, we need to create the new VMs manually if the instances become unhealthy

We are also creating backup of other data volume attached to VM such as Filestore in addition to persistent disk.

4.8.2.3 Data Backup

The data layer in LeMans solution consists of Managed Data Layer, Staging Data Layer, Team Data Layer, Application Backups & Logs. This data layer is stored in the Filestore File System. Given nature of data and access requirements, multiple services are used as part of data backup strategy.

Filestore Snapshots: Filestore snapshots will be created on regular basis to provide RTO and RPO of less than 24 hours, in case of data loss. It is also suitable for faster ad-hoc file restoration but comes additional cost overhead.

Filestore Managed Backups: Filestore comes with **managed backup** feature which can backup the metadata and data stored on the Filestore instances to the regional buckets where the fileshare can be recovered across zones in case of zonal failure. The backups taken by Filestore is incremental and differential stored in a backup chains.

Cloud Storage Archival: Required for archiving data as per the retention policies. It ensures secure preservation of data over the long term, while helping us optimize storage cost. It can also be used for ad-hoc file restoration. It is suitable for long term archiving. It can also be used for full restoration, but it is slower than Filestore Backups.

In addition to the above, **Regional Filestore*** is required to use for HA.

Achieving RTO and RPO goals depend on tiers of backup, disaster recovery and failover design. We need all above defined services along to achieve RTO and RPO goals for Filestore.

Overall RPO of the solution is tied to the times backups are secured offsite, in this case using managed backups and cloud storage as Snapshots are stored in Filestore instances. We are maintaining weekly backups, therefore the RPO for data storage is >= 24 hours <= 1 week.

To calculate the RTO of the solution, we look at the time taken to restore data from the point of failure. Given our data volume and restoration time from managed backups, the expected time to recover all instances with full data can be up to 36 hours.

The below documents cover option paper, HLD and LLDs for different patterns used to define backup and DR strategies for Filestore:

Document Type	Confluence Link	Details

4.8.2.3.1 Backup strategy options paper	LeMans Filestore LeMans Backup, Archival & Recovery - Options Paper	Covers the two options in combination of patterns to achieve resiliency requirements.
4.8.2.3.2 HLD for snapshots, backup and archival	Filestore Backup, Snapshots and Archival Design	Covers the details of patterns approved in option paper
4.8.2.3.3 LLD for snapshots creation and management	Filestore Snapshots LLD	Details of compute engine client, IAM roles, schedule and increase in size of Filestore instances
4.8.2.3.4 LLD for managed backups	Filestore Managed Backup LLD	Details of compute engine client, IAM roles, schedule, full scale restoration
4.8.2.3.5 LLD for archival and restoration from GCS	Archival and Restoration using GCS - LLD	Details of compute engine client, IAM roles, schedule, restoration of SAS, SPDS, and Teams data
4.8.2.3.6 IAM Design for backup solution orchestration	IAM - Service Account Design for LeMans - Data Migrations	Details of how the backup solution is run using service accounts and access to these service accounts from Run team
4.8.2.3.7 Restoration of data from Archival	POC - Filestore Data Restoration from Snapshots and Archival	Details of restoration scenarios and scripts required for run team to restore the data from GCS buckets and Snapshots
4.8.2.3.8 Build document for Filestore Backup and Snapshots	Filestore Backup and snapshot build document	Build scripts and explanation for backups and snapshots
	LeMans Weekly Backup Suspending and Resuming SAS Jobs - Data Migrations - Lloyds Banking Group Confluence	Details of suspending/ resuming SAS jobs and closing/reopening GRID queues
4.8.2.3.9 Build document for Filestore Archival	Filestore Archival | Build and Testing Documentation - Data Migrations - Lloyds Banking Group Confluence	Build scripts and explanation for archival process
	Archival Non-UTF8 Filename Sanitization Script - Data Migrations - Lloyds Banking Group Confluence	Build scripts and explanation for renaming non-UTF8 scripts
4.8.2.3.10 Archival Retention Policies	LeMans - Cloud Storage Buckets Retention Policy - Data Migrations - Lloyds Banking Group Confluence	Details of archival policies set on GCS buckets
4.8.2.3.11 Archival Restoration Documents	Archival Restoration for SPDS Full Table Recovery | Build and Testing Document - Data Migrations - Lloyds Banking Group Confluence Archival Restoration for SPDS Single Table Recovery | Build and Testing Document - Data Migrations - Lloyds Banking Group Confluence Archival Restoration for Files | SAS, Staging, and Team Folders | Build and Testing Document - Data Migrations - Lloyds Banking Group Confluence	Details of restoration process from GCS buckets
4.8.2.3.12 Filestore Backup, Snapshots, & Archival Proof of Concepts	Filestore Backup, Snapshots, & Archival Proof of Concepts - Data Migrations - Lloyds Banking Group Confluence	Details of POCs conducted for various use cases in Backup solution

> **Note**
> Please note - Regional Filestore was proposed and endorsed as the required option for resilient solution. However, this option is not a viable in view of the cost involved, and Bank has decided to continue using Zonal Filestore. This is noted in decision log#AK35

4.8.3 Disaster Recovery(DR)

LeMans applications on Compute Engine and data on Filestore are stored in a **single zone** - europe-west2a(Zone A). As part of the entire solution should failover to (europe-west2b) Zone B with minimum downtime. Failover process includes launching VMs in Zone B, restoring Filestore data and reconnecting the services.

As part of backup and data replication strategy, the zonal Filestore instances are backed up using managed backup and zonal Compute Engine instance disks are backed up using disk snapshots. Please refer to section 4.8.2 Backup Strategy for more details.

4.8.3.1 Failover Strategy

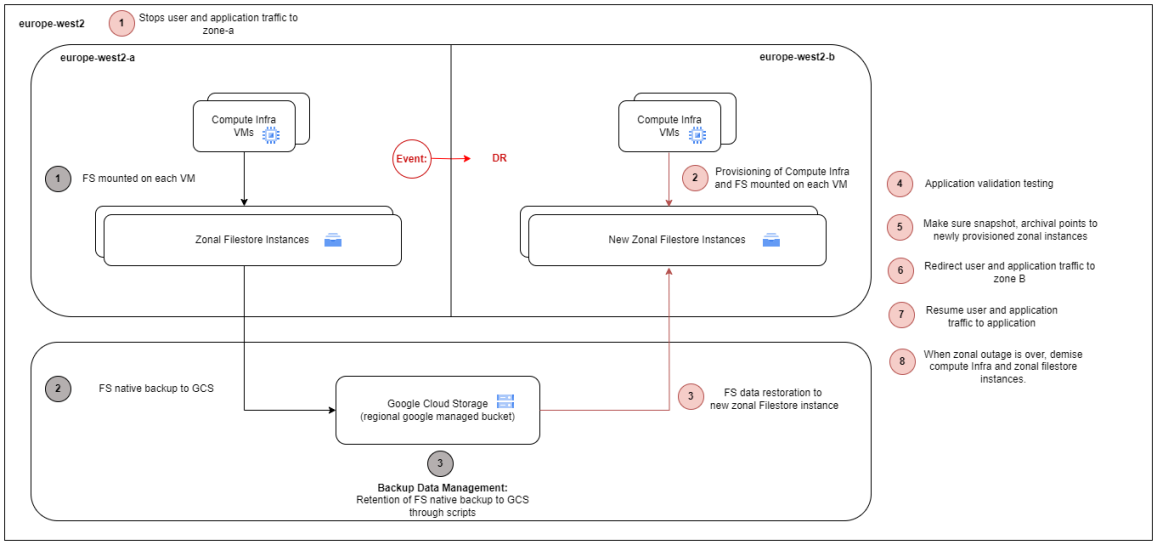
Design pattern Filestore - Disaster Recovery - Data Migrations outlines the disaster recovery solution for application Compute Engine and Filestore instances.

The following steps should include all VMs(SAS environment and backup VMs) and Filestore instances used in the platform.

1. When disaster strikes in Zone A, all user and application traffic coming in that zone will get failed responses.
2. Provision new compute engine instances in Zone B through terraform scripts. This should also include the VMs that can provide orchestration to Backup, Snapshots and Archival processes.
3. Provision new Filestore instances in Zone B using same instance name and mount point. Mount newly provisioned zonal Filestore instances to Virtual Machine's in Zone B. Mechanism to restore the managed backups to instances is described in LLDs.
4. Validate the backup processes are up and running in the new VMs and have the updated references, if any, to the new VMs.
5. Update the load balancer receiving user traffic to point to new VMs in Zone B.
6. Application teams should resume applications, such as start the SAS and Task Scheduler servers on the VM's. They should validate the SAS Application services, SPDS services across all the SAS VM's and validate batch process by executing the flows on Platform Process Manager. SRE team should if all new VMs are

accessible from CyberArk, check the infrastructure health in Dynatrace.

7. As the load balancer health checks are successful, user traffic is resumed. SRE team should monitor the health of platform.
8. When the zonal outage in Zone A is over, we should delete all infrastructure created in that zone. Zone B now becomes the primary zone.



DR SARM document should be used to carry out scheduled DR testing. Detailed design steps to recover data from Filestore in case of small scale and full disaster recovery scenarios is documented in Disaster Recovery Scenarios as part of Filestore Managed Backup LLD - Data Migrations - Lloyds Banking Group Confluence

4.8.3.2 External Connections

4.8.3.2.1 Connect:Direct

Connect:Direct agents are installed on Zonal Persistent Disks on both the Grid controller nodes. In order to provide High Availability in case of host failure, CD agents are to be deployed on the Grid Primary and Secondary Controller Servers and Google's Network Load Balancer is configured to distribute the traffic to the CD nodes. CD service is configured to start automatically when the VM is created. In case of zonal failure, we need to create new UMIG for zone-b and then update the Load Balancer Configuration to point to the new UMIG. After that the CD service will be back up and running in the VM created in zone-B using the snapshot from VM in zone-A.

Connection from IBM Connect:Direct uses GCP load balancer URL as destination. As the LB URL doesn't change during DR, there is no change required to establish connectivity. We can request source team to send new files to test the C:D connectivity with new VMs in zone-B.

LeMans Connect:Direct Implementation - Data Migrations - Lloyds Banking Group Confluence

4.8.3.2.2 Teradata, MS SQL, Oracle

The connection from GCP compute engine instances to on-premise databases are initiate from GCP. In case of DR on GCP, the external database connections are not impacted as the firewalls are setup at subnet level, which do not change when we failover to another zone. In case of zonal failure, we will use the same Firewall rules and database connections, hence no impact. We are testing the connections with the database in PRD-UAT DR, a test case is documented below -

Testing scenario for External databases:

DB & External Connections	Telnet Testing for the DBs		Run the below commands for the given IPs and Port: ssh -v -p <IP Address> <Port> **Oracle:** • DLIA CLOUD - lbgorepr-pdb001-app-sc1.service.group 2484 • FDM - LBGBAPPR-PDB001-APP-sc1.lloydsbanking.cloud 2484 • MIP - LBGMIPPR-PDB001-APP-sc1.service.group 2484 • RDW - p27606prw883.machine.group 2484 **SQL Server:** • CRM36 MODEL DATA - SQLLGV0A36.GLOBAL.LLOYDSTSB.COM 14331 • CRAS - SQLLGV09DE.GLOBAL.LLOYDSTSB.COM 14331 • RDM DATA HUB - SQLLGV0865.GLOBAL.LLOYDSTSB.COM 14331 • PEGA FRAUD MI - SQLLGV0AAA.GLOBAL.LLOYDSTSB.COM 14331 **Teradata:** • GDW: gdwprodcop1.lloydsbanking.cloud 1025
	DB Libname connection		Test Oracle FDM Connection: LIBNAME FDMBAPRD ORACLE DEFER=YES PATH=<path> SCHEMA=<schema> USER=<userid> PASSWORD=<password>; Test GDW Teradata Connection:

			LIBNAME GDWVIEWS TERADATA SERVER SCHEMA=<schema> USER=<userid PASSWORD=<password>;

4.8.3.3 DR Planning

DR planning document ensures that procedures are resources to recover application and data are detailed. It should cover SOE for SRE team to follow and should use the failover strategy. LeMans DR Planning - PRD UAT - Data Migrations - Lloyds Banking Group Confluence is a live document and will contain the detailed DR steps in review with the SRE team.

4.9 Component Design

All the components and services that are required to support the deployment of Filestore for LeMans are documented. Public cloud product catalogue has up to date curation details of all products allowed and used within the bank. For any product feature updates and queries, we should contact the product team.

4.9.1 GCP Components

GCP Services	Description
Filestore Backup	Filestore backup is a copy of a file share that includes all file data and metadata of the fi backup is created.
Firewall	Firewalls are software-based, cloud deployed network devices, built to stop or mitigate
Google Cloud Filestore	Google Filestore is a high performance storage solution to store the LeMans data layer
Google Cloud Storage (GCS)	Google Cloud Storage is a RESTful online file storage web service for storing and acces infrastructure.
Google Data Transfer Appliance	Google Data Transfer Appliance is a high-capacity storage device that enables you to tr Google upload facility, where the data is uploaded to Cloud storage for ingestion
Google Platform Projects / Folders	A project organizes all Google Cloud Platform resources. A project consists of a set of u and monitoring settings for those APIs.
Google Private Services Access	Private services access lets you reach the internal IP addresses of these Google and LE be hosted by Google and accessed via PSA
Identity and Access Management (IAM)	Identity and Access Management (IAM), offers to grant granular access to specific GCI other resources.
Virtual Private Cloud (VPC)	Virtual Private Cloud (VPC) gives the flexibility to scale and control how workloads conn

4.9.2 SAS Environment Sizing

Based on the inputs from LeMans and fed into the SAS EEC sizing model for hosted systems within GCP, the recommendation is to provide the target SAS 94 architecture which is defined in the Confluence pattern - LeMans Environment Compute Engine & Disk Details - Data Migrations - Lloyds Banking Group Confluence

The following documents were created for sizing analysis in collaboration with SAS and Google. The sizing of SAS environment were baselined as per these:

Le Mans - X-Ray Assessment Oct 2020.pdf

Le Mans on Google Cloud - High Level Architecture.pdf

4.9.3 Filestore Environment Sizing

Based on the analysis performed between LBG, SAS Institute and Google within GCP, the recommendation is to provide the target Filestore architecture as follows:

4.9.3.1 INT Environment

Since the Integration [INT] environment is Infrastructure Route-To-Live environment, we will be creating only three Filestore deployment compared to Pre-Prod and Prod Infrastructure environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
int-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
int-lm94-spds-fs2	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds2	
int-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
int-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £9,340

Testing of the throughput against VM types and SAS scenarios will be run in INT, requiring larger Filestore instances in the MDL to support testing:

Component ID	Component Name	Component Type	Storage Layout	Mount Point
int-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	100 TB	/spds1
int-lm94-spds-fs2	Google Cloud Filestore for MDL	High-Capacity	100 TB	/spds2
int-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team
int-lm94-sas-fs1	Google Cloud Filestore for SAS Application	High-Capacity	1 TB	/opt/sas

Total Cost per month: £61,883

4.9.3.2 PRE Environment

The Pre-Prod Infrastructure environment should be setup like Production Environment, hence all the Filestore environments [Dev, CIT, UAT & Prod] will be deployed.

4.9.3.2.1 DEV Environment

The below table provides the sizing information for the Dev Filestore solution deployed on the Pre-Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
pre-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
pre-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
pre-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £6,557

4.9.3.2.2 CIT Environment

The below table provides the sizing information for the CIT Lustre solution deployed on the Pre-Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
pre-cit-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
pre-cit-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
pre-cit-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £6,557

4.9.3.2.3 UAT Environment

The below table provides the sizing information for the UAT SAS 94 solution deployed on the Pre-Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
pre-uat-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
pre-uat-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
pre-uat-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup,

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
					/user_logs

Total Cost per month: £6,557

4.9.3.2.4 PROD Environment

The below table provides the sizing information for the Prod SAS 94 solution deployed on the Pre-Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
pre-prod-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
pre-prod-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
pre-prod-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £6,557

4.9.3.3 PRD Environment

4.9.3.3.1 DEV Environment

The below table provides the sizing information for the Dev SAS 94 solution deployed on the Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
prd-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
prd-dev-lm94-spds-fs2	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds2	
prd-dev-lm94-spds-fs3	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds3	
prd-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
prd-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Enterprise	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £12,550

4.9.3.3.2 CIT Environment

The below table provides the sizing information for the CIT SAS 94 solution deployed on the Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
prd-cit-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
prd-cit-lm94-spds-fs2	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds2	
prd-cit-lm94-spds-fs3	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds3	
prd-cit-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
prd-cit-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Low-Capacity	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £12,550

4.9.3.3.3 UAT Environment

The below table provides the sizing information for the UAT SAS 94 solution deployed on the Prod Infrastructure Environment.

Component ID	Component Name	Component Type	Storage Layout	Mount Point	Consolidated Folders
prd-uat-lm94-spds-fs1	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds1	
prd-uat-lm94-spds-fs2	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds2	
prd-uat-lm94-spds-fs3	Google Cloud Filestore for MDL	High-Capacity	10 TB	/spds3	
prd-uat-lm94-eusr-fs1	Google Cloud Filestore for Teams	High-Capacity	10 TB	/team	
prd-uat-lm94-sas-fs1	Google Cloud Filestore for SAS Application	Low-Capacity	1 TB	/sasshare01	/opt/sas, /backup, /user_logs

Total Cost per month: £12,550

4.9.3.3.4 PROD Environment

The below table provides the sizing information for the Prod SAS 94 solution deployed on the Prod Infrastructure Environment.

› Click here to expand...

Component ID	Filesystem	Component Name	Owner	Component Type	Storage Layout	New Mount
prd-prod-lm94-stage-fs01	100.67.193.2:/stageshare01	Google Cloud Filestore for Staging	Babatunde Akinkugbe	High-Capacity	13 TB	/staging
prd-prod-lm94-spds-fs01	100.67.193.66:/spdsshare01	Google Cloud Filestore for MDL 1	Babatunde Akinkugbe	High-Capacity	75 TB	/spds1
prd-prod-lm94-spds-fs02	100.67.192.194:/spdsshare02	Google Cloud Filestore for MDL 2	Babatunde Akinkugbe	High-Capacity	75 TB	/spds2
prd-prod-lm94-spds-fs03	100.67.192.66:/spdsshare03	Google Cloud Filestore for MDL 3	Babatunde Akinkugbe	High-Capacity	75 TB	/spds3
prd-prod-lm94-spds-fs04	100.67.193.130:/spdsshare04	Google Cloud Filestore for MDL 4	Babatunde Akinkugbe	High-Capacity	75 TB	/spds4
prd-prod-lm94-eusr-fs11	100.67.195.2:/eusrshare11	Google Cloud Filestore for ADM	Lee McGinty	Low-Capacity	1 TB	/team/adm
prd-prod-lm94-eusr-fs12	100.67.195.66:/eusrshare12	Google Cloud Filestore for Asset Finance - Credit Policy	Babatunde Akinkugbe	Low-Capacity	1 TB	/team/af_credit_
prd-prod-lm94-eusr-fs13	100.67.196.2:/eusrshare13	Google Cloud Filestore for Asset Finance - Customer Expericence	Roselyne Renel	Low-Capacity	1 TB	/team/af_custon
prd-prod-lm94-eusr-fs14	100.67.195.130:/eusrshare14	Google Cloud Filestore for Asset Finance - Capital & Impairment	Lee McGinty	Low-Capacity	6.75 TB	/team/af_impair
prd-prod-lm94-eusr-fs15	100.67.195.194:/eusrshare15	Google Cloud	Steve Byron	Low-Capacity	1 TB	/team/af_lex_au

Component ID	Filesystem	Component Name	Owner	Component Type	Storage Layout	New Mount
		Filestore for Asset Finance - Lex Autolease				
prd-prod-lm94-eusr-fs16	100.67.198.2:/eusrshare16	Google Cloud Filestore for Asset Finance - Modelling	Roselyne Renel	Low-Capacity	4 TB	/team/af_modell
prd-prod-lm94-eusr-fs17	100.67.198.66:/eusrshare17	Google Cloud Filestore for Asset Finance - Portfolio Management	Lee McGinty	Low-Capacity	1 TB	/team/af_portfol
prd-prod-lm94-eusr-fs18	100.67.198.130:/eusrshare18	Google Cloud Filestore for Asset Finance - Portfolio Performance	Lee McGinty	Low-Capacity	1 TB	/team/af_portfol
prd-prod-lm94-eusr-fs19	100.67.198.194:/eusrshare19	Google Cloud Filestore for Analytics Modelling	Jasjyot Singh	Low-Capacity	8.5 TB	/team/analytics_
prd-prod-lm94-eusr-fs20	100.67.199.2:/eusrshare20	Google Cloud Filestore for Banking	Ian Goodchild	Low-Capacity	8.25 TB	/team/banking
prd-prod-lm94-eusr-fs21	100.67.199.66:/eusrshare21	Google Cloud Filestore for BASEL	Roselyne Renel	Low-Capacity	3 TB	/team/basel
prd-prod-lm94-eusr-fs22	100.67.199.130:/eusrshare22	Google Cloud Filestore for BICC Admin	Lee McGinty	Low-Capacity	2 TB	/team/bicc_adm
prd-prod-lm94-eusr-fs23	100.67.199.194:/eusrshare23	Google Cloud Filestore for BICC Developers	Babatunde Akinkugbe	Low-Capacity	4.5 TB	/team/bicc_deve
prd-prod-lm94-eusr-fs24	100.67.200.2:/eusrshare24	Google Cloud Filestore for Capital Reporting	Babatunde Akinkugbe	Low-Capacity	2 TB	/team/caprep
prd-prod-lm94-eusr-fs25	100.67.200.66:/eusrshare25	Google Cloud Filestore for Cards	Steve Byron	High-Capacity	20 TB	/team/cards
prd-prod-lm94-eusr-fs26	100.67.200.130:/eusrshare26	Google Cloud Filestore for Cards Repricing	Jasjyot Singh	Low-Capacity	2 TB	/team/cards_rep
prd-prod-lm94-eusr-fs27	100.67.200.194:/eusrshare27	Google Cloud Filestore for CBR Gen	Jasjyot Singh	Low-Capacity	1.5 TB	/team/cbrgen
prd-prod-lm94-eusr-fs28	100.67.201.2:/eusrshare28	Google Cloud Filestore for CBR Reporting	Chintain Pindoria	Low-Capacity	2 TB	/team/cbrrep

Component ID	Filesystem	Component Name	Owner	Component Type	Storage Layout	New Mount
prd-prod-lm94-eusr-fs29	100.67.201.66:/eusrshare29	Google Cloud Filestore for CCFD	Chintain Pindoria	High-Capacity	1.25 TB	/team/CCFD
prd-prod-lm94-eusr-fs30	100.67.201.130:/eusrshare30	Google Cloud Filestore for CF Analytics	Roselyne Renel	Low-Capacity	6 TB	/team/cf_analyti
prd-prod-lm94-eusr-fs31	100.67.201.194:/eusrshare31	Google Cloud Filestore for COO	Jayne Opperman	Low-Capacity	1 TB	/team/chief_ope
prd-prod-lm94-eusr-fs32	100.67.202.2:/eusrshare32	Google Cloud Filestore for Collections & Recoveries	Jayne Opperman	Low-Capacity	2.5 TB	/team/collection
prd-prod-lm94-eusr-fs33	100.67.202.66:/eusrshare33	Google Cloud Filestore for Consumer Cards	Roselyne Renel	Low-Capacity	7.5 TB	/team/consume
prd-prod-lm94-eusr-fs34	100.67.203.66:/eusrshare34	Google Cloud Filestore for CPM	Jasjyot Singh	Low-Capacity	7.5 TB	/team/cpm
prd-prod-lm94-eusr-fs35	100.67.202.130:/eusrshare35	Google Cloud Filestore for CRDiv	Lee McGinty	High-Capacity	80 TB	/team/crdiv
prd-prod-lm94-eusr-fs36	100.67.202.194:/eusrshare36	Google Cloud Filestore for Customer	Chintain Pindoria	High-Capacity	15 TB	/team/customer
prd-prod-lm94-eusr-fs37	100.67.203.2:/eusrshare37	Google Cloud Filestore for Decision Science	Lee McGinty	High-Capacity	65 TB	/team/decision_
prd-prod-lm94-eusr-fs38	100.67.203.130:/eusrshare38	Google Cloud Filestore for Fraud Modelling	Lee McGinty	Low-Capacity	4 TB	/team/fmod
prd-prod-lm94-eusr-fs39	100.67.203.194:/eusrshare39	Google Cloud Filestore for Forecasting	Jayne Opperman	Low-Capacity	9 TB	/team/forecast
prd-prod-lm94-eusr-fs40	100.67.204.66:/eusrshare40	Google Cloud Filestore for Fraud	Lee McGinty	Low-Capacity	1 TB	/team/fraud
prd-prod-lm94-eusr-fs41	100.67.204.2:/eusrshare41	Google Cloud Filestore for Fraud Reporting	Steve Byron	Low-Capacity	4 TB	/team/frep
prd-prod-lm94-eusr-fs42	100.67.204.130:/eusrshare42	Google Cloud Filestore for Loans	Roselyne Renel	Low-Capacity	7 TB	/team/loans
prd-prod-lm94-eusr-fs43	100.67.204.194:/eusrshare43	Google Cloud Filestore for Macros	Babatunde Akinkugbe	Low-Capacity	1 TB	/team/macro
prd-prod-lm94-eusr-fs44	100.67.205.2:/eusrshare44	Google Cloud	Steve Byron	Low-Capacity	8 TB	/team/modelling

Component ID	Filesystem	Component Name	Owner	Component Type	Storage Layout	New Mount
		Filestore for Modelling				
prd-prod-lm94-eusr-fs45	100.67.205.66:/eusrshare45	Google Cloud Filestore for Modelling Data	Jasjyot Singh	Low-Capacity	2.25 TB	/team/modelling
prd-prod-lm94-eusr-fs46	100.67.205.130:/eusrshare46	Google Cloud Filestore for Mon Dev	Chintain Pindoria	Low-Capacity	1.5 TB	/team/mon_dev
prd-prod-lm94-eusr-fs47	100.67.205.194:/eusrshare47	Google Cloud Filestore for Mortgages	Roselyne Renel	Low-Capacity	9.5 TB	/team/mortgage
prd-prod-lm94-eusr-fs48	100.67.206.2:/eusrshare48	Google Cloud Filestore for Performance Monitoing	Chintain Pindoria	Low-Capacity	2 TB	/team/perf_mon
prd-prod-lm94-eusr-fs49	100.67.206.66:/eusrshare49	Google Cloud Filestore for Portfolio Analytics	Lee McGinty	Low-Capacity	6.25 TB	/team/portfolio_
prd-prod-lm94-eusr-fs50	100.67.206.130:/eusrshare50	Google Cloud Filestore for Private Banking	Roselyne Renel	Low-Capacity	1 TB	/team/private_ba
prd-prod-lm94-eusr-fs51	100.67.206.194:/eusrshare51	Google Cloud Filestore for RBB CR	Roselyne Renel	Low-Capacity	3.25 TB	/team/rbbcr
prd-prod-lm94-eusr-fs52	100.67.207.2:/eusrshare52	Google Cloud Filestore for RICI	Chintain Pindoria	Low-Capacity	2.25 TB	/team/rici
prd-prod-lm94-eusr-fs53	100.67.207.66:/eusrshare53	Google Cloud Filestore for Risk Reporting	Jonathan Burgess	Low-Capacity	4.5 TB	/team/riskrep
prd-prod-lm94-eusr-fs54	100.67.207.130:/eusrshare54	Google Cloud Filestore for Stress Testing	Chintain Pindoria	Low-Capacity	1 TB	/team/stress_tes
prd-prod-lm94-eusr-fs55	100.67.209.194:/eusrshare55	Google Cloud Filestore for Basel	Babatunde Akinkugbe	Low-Capacity	10 TB	/team/basel2
prd-prod-lm94-sas-fs01	100.67.192.130:/sasshare01	Google Cloud Filestore for SAS Application	Babatunde Akinkugbe	High-Capacity	13 TB	/sasshare01

Total Cost per month: £171,133

> **Note**
> BLD, INT, PRE-UAT, PRE-CIT, and PRE-PRD environments are not available in LeMans Route-to-live

Summary:

- 280.0TB allocation for Managed Data Layer
- 310.75TB allocation for Team Folders
- 10.0TB allocation for SAS Services (install)
- 10.0TB allocation for Staging

Notes:

- *CPM folders have been consolidated
- ` Fraud folders have been consolidated
- ^Decision Science Folders have been consolidated
- /team/analytics has been removed
- CRDiv allocation has been formally recognised:
  - /data/fdl is currently used by DS, might not be essential in the to-be
  - /verde areas need to be considered (~10TB) - with this there is a long likelihood that the /team/ds will need to be split out into 2 areas.
  - Do DS/CRDiv need to move all the /verde data? Question to be asked
- New CRDiv area will be mapped to the Decision Science AD groups.

4.9.4 Filestore Configuration

Several aspects of Filestore configuration will be setup to reflect the best-practices recommended for the file system.

4.9.4.1 Filestore Instance Options

The below options will need to be set when creating the Filestore instances along with the default options that is provided by Cloud Services.

S.No	Option	Value	Description
1	access_mode	READ_WRITE	Allows the clients to read and write data into the Filestore Instances
2	squash_mode	NO_ROOT_SQUASH	Allows the data stored on the Filestore Instances to be owned by the required users and groups instead of root.
3	ip_ranges	<SAS 94 Subnet range>	Allows the GCE instances within this IP range to access the Filestore Instances
4	capacity	<Size of the Filestore>	Allows the project to define the size of the Filestore
5	name	<name of the Filestore>	Allows the project to define the name of the Filestore
6	tier	<Filestore Tier>	Allows the project to define the service tier of the Filestore

> **Note**
> The squash_mode for the Filestore Instance should be "**no_root_squash**". However, root_squash is required to setup the Filestore Instance for the first time as the root account is required to create the required folders and applying ACLs on them. So, there has been an exception made in the guardrail to allow for the "root_squash" mode to be made available for use within 48 hours of launching the Filestore Instance after which the squash_mode needs to be reverted back to "no_root_squash".

4.9.4.2 Mount points

Filestore for MDL and Teams area are created and mounted as separate devices. Please refer to Lemans Filestore Process - Data Migrations - Lloyds Banking Group Confluence for details on individual mount points.

4.9.5 Filestore Performance

4.9.5.1 IO Throughput

The below throughput values calculated using the baseline throughput supported by Filestore tiers used. The actual performance throughput testing has been performed by the Testing team on the **lower** environments and the results are captured in the Confluence page R2.1 - EOTR - Data Migrations - Lloyds Banking Group Confluence. Performance testing in **Production** environment is in progress and results will be captured in R3.4 EOTR - Data Migrations - Lloyds Banking Group Confluence

Click here to expand...

INT Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	10 - 200	2600 - 52000	
Team Layer	10	2600	
Total	**20**	**5200**	**260 MB/s/core**

PRE-DEV Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	10	2600	
Team Layer	10	2600	

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
Total	**20**	**5200**	**162 MB/s/core**

PRE-CIT Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	10	2600	
Team Layer	10	2600	
Total	**20**	**5200**	**162 MB/s/core**

PRE-UAT Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	10	2600	
Team Layer	10	2600	
Total	**20**	**5200**	**162 MB/s/core**

PRE-PRD Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	10	2600	
Team Layer	10	2600	
Total	**20**	**5200**	**162 MB/s/core**

PRD-DEV Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	30	7800	
Team Layer	10	2600	
Total	**40**	**10400**	**325 MB/s/core**

PRD-CIT Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	30	7800	
Team Layer	10	2600	
Total	**40**	**10400**	**325 MB/s/core**

PRD-UAT Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	30	7800	
Team Layer	10	2600	
Total	**40**	**10400**	**162 MB/s/core**

PRD-PRD Environment

Data Layer	Total Size in TB	Filestore Total Throughput (MBps)	SAS Throughput per core
MDL Layer	436	113360	
Team Layer	328	76460	
Total	**764**	**189820**	**370 MB/s/core**

> **Note**
> Filestore IO throughput is taken from Google documentation at the time of initial design. The numbers may vary as Google improves the performance of Filestore.

4.9.5.2 Performance Considerations

4.9.5.2.1 MDL Library Definitions

The MDL Data Layer consists of SAS SPDS Libraries which can be configured to be stored across multiple folders. This will help distribute the data across all the Filestore Instances which will increase the SPDS dataset read/write performance.

Detailed pattern for the Library definition changes is captured on the confluence pattern Filestore - MDL Library Definitions

4.9.5.2.2 Filestore Client Tuning

The below options for Filestore client mounts are considered to improve reliability and performance.

Filestore Mount	Option	Value
/opt/sasshare	Consistent View of NFS Filesystem across all client machines	actimeo=0
/opt/sasshare	Disable update on when the file was last accessed	noatime

5. FinOps

LeMans FinOps process is documented in LeMans FinOps - Data Migrations - Lloyds Banking Group Confluence

6. Logging, Monitoring and Alerting

6.1 Logging

6.1.1 SAS Application Logging

The default logging levels for metadata & compute will be applied, and logging will be configured to reference a unique /opt/sas/logs directory as part of the configuration. This change avoids the default logs being written to the Configuration directory and mitigates the associated risks of the mount becoming 100% full, leading to a SAS outage.

The centralized log location /opt/sas/logs will have the following logs.

- Metadata Server
- Object Spawner
- Workspace Server
- PooledWorkspace Server
- StoredProcess Server
- SPDS Server
- Batch Server

The Mid Tier logs will be stored in the default location. Housekeeping jobs should be setup to archive the older logs.

SAS Logs will be stored on Filestore application for various SAS Services and Batch Jobs. SAS uses a log4j to configure log output.

Logging Policies:

Event	Online Retention	Backup Retention(GCS)
Standard Events	6 months	12 months
SOX Events	6 months	15 months
Enhanced Oversight Model Events	6 months	15 months

6.1.2 LeMans Filestore Backup Logging

LeMans Filestore backup solution has components such as managed backups, snapshots, archiving data to GCS and restoration to compute engine instances. These are orchestrated using crontab and Unix scripts. The scripts generate custom logs, required for run team for investigating issues, validating the status of jobs, and raising alerts in ServiceNow for any incidents.

LeMans Backup Logging, and Alerting describes the patterns used for LeMans backup logging.

6.2 Monitoring & Alerting

Dynatrace is a observability tool for applications and their underlying infrastructure. The LBG bank has identified Dynatrace as tool of choice for monitoring applications and infrastructure. Here at LBG, Dynatrace implementations are requested by

value-stream teams and implemented by EMAS team. Dynatrace is required as part of your Service Introduction. Dynatrace provides capability for monitoring , Dashboard and Alerting.

We are monitoring metrices generated by Infrastructure, SAS processes, and custom logs. We will use the patterns created by Dynatrace to monitor and send alerts. Detailed process and metrics are defined in Lemans Monitoring & Alerting through Dynatrace - Data Migrations - Lloyds Banking Group Confluence

Dynatrace supports Alerting using Teams channels and ServiceNow incident. Alerts generated in non-prod environments are notified via **Teams** channels. Alerts generated in production environments create **ServiceNow incidents**. Project teams are responsible to raise engagement with Dynatrace and provide requirements.

Dynatrace Alerts & Default Thresholds - EMAS - Public - Lloyds Banking Group Confluence explains the incident creation workflow

How To Request a Teams Connector Or Web Hook. - Platform Lab Initiated - Lloyds Banking Group Confluence covers the instructions to create a Teams connector for non-prod alerts.

ServiceNow incident queue, priority, alert scenarios and engagement raised with Dynatrace is detailed in Lemans Monitoring & Alerting through Dynatrace - Data Migrations - Lloyds Banking Group Confluence

6.2.1 Infrastructure Monitoring

Dynatrace is the standard morning tool used across GCP for Infrastructure Monitoring and Alerting. The Infrastructure monitoring includes default monitoring of CPU load, Processes running, Memory usage, Disk Usage, Network Usage, System events such as start-up/shutdown etc., The Infrastructure alerts are generated by Dynatrace for servers monitored by the OneAgent.. Dynatrace agent gathers metrics every 10 seconds for infrastructure metrics.

EMAS team is point of contact for initial configuration, troubleshooting. Please use EMAS confluence space to raise engagement depending on the type of resource - EMAS - Public Home - EMAS - Public - Lloyds Banking Group Confluence

Dynatrace provides separate solutions for:

6.2.1.1 VM host

Infrastructure default monitoring includes metrics from Compute Engine instances . We need to work with Dynatrace team to enable IAM roles for Dynatrace agent to be able to read metrics from Compute Engine instances.

Dynatrace monitoring for VMs is documented in DT Host Alerting - EMAS - Public - Lloyds Banking Group Confluence

6.2.1.2 GCP PaaS Services

Monitoring data from GCP PaaS services transferred to Dynatrace using 'Dynatrace GCP Function' component which is running from mgmt GKE cluster.

The metrics are queried from the google monitoring APIs and posted to Dynatrace via a REST call.

Reference - EMAS architecture page - Dynatrace GCP Function Design - EMAS - Public - Lloyds Banking Group Confluence

Dynatrace on boarding is done using DT GCP Function Onboarding - EMAS - Public - Lloyds Banking Group Confluence

6.2.1 Application Monitoring

SAS Environment Manager will be configured to monitor directory space usage, log file events, and configuration file changes and to report alerts. SAS Alerts will be configured in SAS EVM to send mail alerts based on the triggers configured at different threshold levels.

SAS Management Console can also be used to monitor the activity of SAS servers, including the metadata server, object spawner, pooled workspace server, stored process server, and workspace server. Following types of information can be viewed on SAS Management Console:

- currently connected clients
- graphs showing the number of spawned server sessions over time
- sessions that are active, inactive, or terminated
- server options and properties
- performance counters
- loggers and associated logging levels
- logging messages, at the level specified

SAS processes running on Compute Engine instances such as SAS Object Spawner, SAS Connect Spawner, SAS Process Manager etc are monitored by Dynatrace agent.

6.2.3 Custom Log Monitoring

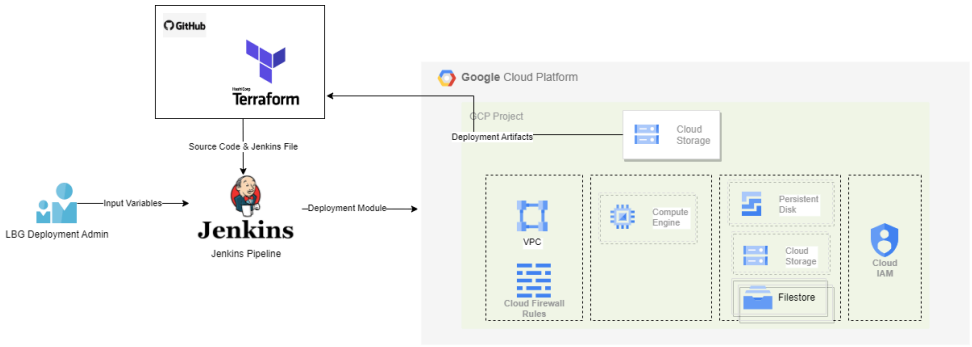
We have shell scripts scheduled via cron to create and manage Filestore snapshots, backups, and archival procedures. These scripts create success and error logs. Dynatrace agent can read the log files and generate alerts. Custom log configuration allows us to add log files that have not been autodetected by Dynatrace. Please refer to Lemans Monitoring & Alerting through Dynatrace - Data Migrations - Lloyds Banking Group Confluence for more details on the alerts setup

7. Infrastructure and Application CI/CD

7.1 Automation Toolset

Component	Description	Remarks
Jenkins	CI Tool will be used a starting point for the Deployment Automation. The deployment pipeline will be run on Jenkins by the Deployment Admin.	The Jenkins instance existing in LBG can be used to host this pipeline. Else a new instance for running Jenkins on this GCP project should be created. Spinnaker is used for CD in the upper environments (RTL, INT, PRE-PRD, and PRD).
GitHub Enterprise	This will be enterprise repository where all code artefacts and scripts will be stored.	Jenkins will be configured to clone this repo during pipeline execution.
GCP Cloud Storage	SAS Depot for SAS94 and other Deployment related artefacts will be stored in a Cloud Storage bucket.	Details of this GCS bucket will be passed as input to the Jenkins pipeline.
Terraform	Terraform is the agreed method for provisioning GCP Infrastructure. HashiCorp Configuration Language (HCL) allows for concise descriptions of resources using blocks, arguments, and expressions. Terraform will used to create GCE, Disks, VPC, Subnets, Firewall rules, Cloud Storage, Filestore and other relevant GCP resources.	There are established patterns in LBG for using Terraform.
Ansible	Ansible is a configuration automation tool which will be used for the deployment of the SAS Application and perform the post deployment configurations.	Ansible is currently being evaluated by Cloud Engineering for use with the LBG Tenant.
Bash Script	Automation in Unix Servers will be done through Ansible scripts which internally employ Bash scripts. These will be pulled from the GitHub repo and be used to perform the following – Pre-Reqs, Deployment, Post Install Steps for SAS94 environment.	Established pattern in LBG for using Bash script should be identified.
Chef InSpec	GCP Infrastructure testing is being done with Chef InSpec in LBG. We need to integrate our Pipelines with existing Chef process.	

7.2 CI/CD Pipeline



7.3 Automation Stages

Stage	Description
Initialization	Terraform Initialization and necessary backend environment requisites will be performed in this step.
Infrastructure Build	Terraform apply will be executed and declared GCP resources will be created.
SAS Pre-Requisites	SAS Pre-Requisites on the Linux server will be performed.
SAS Deployment	SAS Software Deployment including installation of Binaries and Configuration will be performed in this stage.
SAS Post Install Steps	Post install Steps for SAS 94 environment will include Securing environment with SSL, adding Data source connection, etc.
Testing	Testing Automation pack will be executed to ensure the provisioned SAS 94 platform is up and running as expected. End to end testing will be performed.