Archival and Restoration using GCS - LLD

1. Overview

The objective of this document is to provide low level design details required for implementation of archival and restoration of Filestore data used in LeMans.

Archival Process for LeMans - HLD - Data Migrations - Lloyds Banking Group Confluence and Filestore Backup, Snapshots and Archival Design - Data Migrations - Lloyds Banking Group Confluence should be read before the low-level design document. To understand the role of archival process in the LeMans backup workflow, please refer to section LeMans Weekly Backup Workflow in LeMans backup workflow

2. Archival and Restoration Configuration

2.1 Archival Configuration

Archival process runs from Filestore snapshots so the transfer time does not cause any interference with the applications. We will schedule the archival processes for instances at different times and distribute the workload on client VMs in terms of size of transfer.

> **Note**
> Single VM can be used in pre-dev, prod-dev, prod-cit, and prod-uat environments inline with the cost and requirements to restore data in production only.

Below section describes the detailed configuration of compute engine instances, schedule, bucket and Filestore instance details:

> **Note**
> The time to run the scripts is still in discussion with business, and may be updated

2.1.1 PRE Environment

2.1.1.1 DEV Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Hostname	Schedule	Day	Timing(UK Time)
pre-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group	WEEKLY	SUNDAY	6:05 PM
pre-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group	WEEKLY	SUNDAY	6:05 PM
pre-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group	WEEKLY	SUNDAY	6:05 PM

2.1.2 PRD Environment

2.1.2.1 DEV Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Hostname	Schedule	Day	Timing(UK Time)
prd-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-dev-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-dev-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM

2.1.2.2 CIT Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Hostname	Schedule	Day	Timing(UK Time)
prd-cit-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-cit-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-cit-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-cit-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-cit-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM

2.1.2.3 UAT Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Hostname	Schedule	Day	Timing(UK Time)
prd-uat-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-uat-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-uat-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-uat-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-uat-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM

2.1.2.4 PRD Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Instance	Schedule	Day	Timing Time)
prd-prod-lm94-stage-fs1	Google Cloud Filestore for Staging	10 TB	/staging	to-lem94-prd-01-stb-eu-sas-lemans-staging-e6f0eff2	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:05 PM
prd-prod-lm94-spds-fs2	Google Cloud Filestore for MDL 1	67.5 TB	/spds1	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-prod-lm94-spds-fs3	Google Cloud Filestore for MDL 2	67.5 TB	/spds2	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-prod-lm94-spds-fs4	Google Cloud Filestore for MDL 3	67.5 TB	/spds3	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-prod-lm94-spds-fs5	Google Cloud Filestore for MDL 4	67.5 TB	/spds4	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	6:05 PM
prd-prod-lm94-eusr-fs1	Google Cloud Filestore for ADM	1 TB	/team/adm	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs2	Google Cloud Filestore for Asset Finance - Credit Policy	1 TB	/team/af_credit_policy_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs3	Google Cloud Filestore for Asset Finance - Customer Experience	1 TB	/team/af_customer_experience_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs4	Google Cloud Filestore for Asset Finance - Capital & Impairment	6.75 TB	/team/af_impairment_and_capital_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs5	Google Cloud Filestore for Asset Finance - Lex Autolease	1 TB	/team/af_lex_autolease_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs6	Google Cloud Filestore for Asset Finance - Modelling	4 TB	/team/af_modelling_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs7	Google Cloud Filestore for Asset Finance - Portfolio Management	1 TB	/team/af_portfolio_management_insight	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs8	Google Cloud Filestore for Asset Finance - Portfolio Performance	1 TB	/team/af_portfolio_performance_team	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs9	Google Cloud Filestore for Analytics Modelling	8.5 TB	/team/analytics_mod	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs10	Google Cloud Filestore for Banking	7.75 TB	/team/banking	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs11	Google Cloud Filestore for BASEL	3 TB	/team/basel	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs12	Google Cloud Filestore for BICC Admin	2 TB	/team/bicc_admin	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs13	Google Cloud Filestore for BICC Developers	3.5 TB	/team/bicc_developers	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs14	Google Cloud Filestore for	2 TB	/team/caprep	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Instance	Schedule	Day	Timing Time)
	Capital Reporting							
prd-prod-lm94-eusr-fs15	Google Cloud Filestore for Cards	17.5 TB	/team/cards	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs16	Google Cloud Filestore for Cards Repricing	2 TB	/team/cards_repricing	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs17	Google Cloud Filestore for CBR Gen	1.5 TB	/team/cbrgen	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs18	Google Cloud Filestore for CBR Reporting	1.75 TB	/team/cbrrep	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs19	Google Cloud Filestore for CCFD	1 TB	/team/CCFD	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs20	Google Cloud Filestore for CF Analytics	6 TB	/team/cf_analytics	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs21	Google Cloud Filestore for COO	1 TB	/team/chief_operating_office	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs22	Google Cloud Filestore for Collections & Recoveries	2.5 TB	/team/collections_recoveries	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs23	Google Cloud Filestore for Consumer Cards	7 TB	/team/consumer_cards_business	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs24	Google Cloud Filestore for CPM	7 TB	/team/cpm	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs25	Google Cloud Filestore for CRDiv	80 TB	/team/crdiv	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs26	Google Cloud Filestore for Customer	12.5 TB	/team/customer	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs27	Google Cloud Filestore for Decision Science	65 TB	/team/decision_science	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs28	Google Cloud Filestore for Fraud Modelling	4 TB	/team/fmod	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs29	Google Cloud Filestore for Forecasting	9 TB	/team/forecast	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs30	Google Cloud Filestore for Fraud	1 TB	/team/fraud	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs31	Google Cloud Filestore for Fraud Reporting	4 TB	/team/frep	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs32	Google Cloud Filestore for Loans	7 TB	/team/loans	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs33	Google Cloud Filestore for Macros	1 TB	/team/macro	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs34	Google Cloud Filestore for Modelling	6.5 TB	/team/modelling	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs35	Google Cloud Filestore for Modelling Data	2 TB	/team/modelling_data	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllc84c6dcf6.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs36	Google Cloud Filestore for Mon Dev	1.5 TB	/team/mon_dev	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs37	Google Cloud Filestore for Mortgages	7.75 TB	/team/mortgages	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs38	Google Cloud Filestore for Performance Monitoing	2 TB	/team/perf_mon	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs39	Google Cloud Filestore for Portfolio Analytics	5 TB	/team/portfolio_analytics	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs40	Google Cloud Filestore for Private Banking	1 TB	/team/private_banking	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs41	Google Cloud Filestore for RBB CR	3 TB	/team/rbbcr	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	glli59d42206.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs42	Google Cloud Filestore for RICI	2.25 TB	/team/rici	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs43	Google Cloud Filestore for Risk Reporting	4.5 TB	/team/riskrep	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs44	Google Cloud Filestore for Stress Testing	1 TB	/team/stress_test	to-lem94-prd-01-stb-eu-sas-lemans-spds-f97ffe6f	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-eusr-fs55	Google Cloud Filestore for Basel	10 TB	/team/basel2	to-lem94-prd-01-stb-eu-sas-lemans-team-09e91984	gllt118c6e2b.gcp.cloud.group	WEEKLY	SUNDAY	8:30 PM
prd-prod-lm94-sas-fs1	Google Cloud Filestore for SAS Application	10 TB	/sasshare01	to-lem94-prd-01-stb-eu-sas-lemans-9bc03f90	gllu2c607a7b.gcp.cloud.group	WEEKLY	SUNDAY	8:05 PM

2.2 Restoration Configuration

2.2.1 PRE Environment

2.2.1.1 DEV Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Hostname
pre-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group
pre-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group
pre-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-pre-04-stb-eu-sas-lemans-e87e6bac	gtl3672f3b59.gcp.cloud.test.group

2.2.2 PRD Environment

2.2.2.1 DEV Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Instance
prd-dev-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group
prd-dev-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group
prd-dev-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group
prd-dev-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group
prd-dev-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-04-stb-eu-sas-lemans-9008722f	gllh622fc29d.gcp.cloud.group

2.2.1.2 CIT Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Instance
prd-cit-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group
prd-cit-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group
prd-cit-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group
prd-cit-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group
prd-cit-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-03-stb-eu-sas-lemans-5b67e04e	glla4bc72fe0.gcp.cloud.group

2.2.2.3 UAT Environment

Component ID	Component Name	Storage Layout	Mount Point	Bucket	Client Compute Engine Instance
prd-uat-lm94-spds-fs1	Google Cloud Filestore for MDL	10 TB	/spds1	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group
prd-uat-lm94-spds-fs2	Google Cloud Filestore for MDL	10 TB	/spds2	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group
prd-uat-lm94-spds-fs3	Google Cloud Filestore for MDL	10 TB	/spds3	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group
prd-uat-lm94-eusr-fs1	Google Cloud Filestore for Teams	10 TB	/team	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group
prd-uat-lm94-sas-fs1	Google Cloud Filestore for SAS Application	1 TB	/sasshare01	to-lem94-prd-02-stb-eu-sas-lemans-129fa8f7	gllv819c6b75.gcp.cloud.group

2.2.2.4 PRD Environment

2.3 Execution Workflow of Archival Sub-Processes

The archival process to backup data from Filestore to GCS comprise of sub-processes to handle known-issues described later in this page. The following diagram branches from the E2E execution flow from the LeMans backup workflow.

2.3.1 Running housekeeping job to handle non-utf8 issue

A housekeeping job will run before the weekly backup job runs on Sundays to handle non-utf8 character issues with filenames. If users upload the files with non-utf8 characters after this housekeeping job has run, then the archival job cannot process them. So, the users should be communicated about this and risk associated. As per the current development testing, we should allow a window of 6 hours for this process to complete.

3. Steps to Restore data from Archival

3.1 Querying the GCS bucket for requested files

Run team will use the client VM where all Filestore instances are mounted to list the files available in GCS bucket. Script to list the files will provide pitdate or query date range as input. A range of query date will filter the files copied in that time frame. Weekly archival will run every Sunday, so a range between Sunday and closest date of user request will provide the relevant results. If you provide a single date, the script should fetch the files backed up on that date.

gcloud commands to list files

```
Clustered table
gsutil ls -la gs://<bucket</<filestore-path>/spds*/<filestore-table-path>/<table-name-with member>*
example - gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/data/rddmuser/experianbureau_hd_202409*


Non-Clustered table or other files

gsutil ls -la gs://<bucket>/<filestore-path>/<file-name>
example - gsutil ls gs://to-lem94-prd-01-stb-eu-sas-lemans-team-09e91984/prd01/team_data/team/bicc_developers/bicc_developers_inventory.txt
```

3.2 Identifying the version of the file to be restored

Up to three versions of a file are stored in buckets. When we query the file from above step, all live and noncurrent versions of files will show in the response. Run team will identify the backup date to use as query date and run the scripts to filter the output based on query date as explained above. Specific version can be retrieved in the same way as retrieving the live file using copy command. Refer to LeMans - Retention Policy for more details on how versioning has been setup.

3.3 Active and Inactive Files

TSM backup utility provides a flag Active/Inactive, the most recent backup version is the *active* version. Any other backup version is an *inactive* version. Every time Tivoli Storage Manager backs up your files, it marks the new backup version as the active backup, and the last active backup becomes an inactive backup. On GCP, we achieve this by using versioning and backup date. The latest version of the file stored in Archival becomes the active file. In case of SPDS cluster, the files can be distributed in multiple folders. In the following example, two versions of a partial file are available in **location** and one live version is available in another **location**. In this case, based on latest timestamp, the highlighted version becomes active backup as this version is live on Filestore.

3.4 Restoration of Archived data to Filestore Instances

Once the files to be restored are identified, we restore them using gcloud storage cp command. First we copy the files to a temp location on Filestore and then these can be copied back to original locations or any other location provided by users.

3.4.1 Restoration from Clustered SPDS Tables

SPDS clustered tables are created with partitions called members. The members are created using hundreds of small partial files. When application writes data to SPDS tables, load balancer can direct the write or update to any of the four SPDS instances. So, If a given partial file PARTIAL_FILE_X is written to SPDS1 instance, the next time the linked record is updated, the updated PARTIAL_FILE_X can go to SPDS3 instance. Due to this distributed nature, while restoring the files from archival, we will query all SPDS folders for a given member in the table.

3.4.1.1 Restoration Order

Restoration order is first restoring metadata files, then data files, and index files in the end. The restoration order is an SPDS internal working requirement. There is only metadata file per member, but there can be multiple versions of the metadata files.
Example: Restore a given member for 26-09-2024

SPDS order of restoration

```
gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/data/rddmuser/experianbureau_hd_202409* | grep "2024-09-26"
gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/metadata/rddmuser/experianbureau_hd_202409* | grep "2024-09-26"
gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/index/rddmuser/experianbureau_hd_202409* | grep "2024-09-26"
```

3.4.1.2 Restoring .cdf file

In addition to .mdf files in metadata folder, we have to restore .cdf file. There is one .cdf file available per table

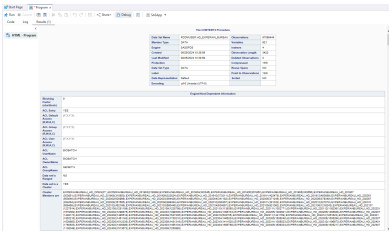
SPDS all commands

```
gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds/rdl/metadat/rddmuser/*.cdf*
```

3.4.1.3 Example SPDS Cluster Restoration

Click here to expand...

1. We have a table EXPERIANBEAUREU_HD with members from 2019 till date copied to GCP Filestore and archived to GCS. User requests to restore member 202409 from table.



2. We query the GCS bucket for this table across all four SPDS folder
3. We identify the restoration date, and run the restoration script with date as query date. In this example 26-09-2024 is used as query date. Only one version of each partial file is returned as a result of providing query date.

```
[srvapplemsas04@gll1831575789 soft_link]$ gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/data/rddmuser/experianbureau_hd_202409* | grep "2024-09-26T09:*"
  121618542  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.0.1.spds9#172
  121537663  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.12.1.spds9#17
  115313540  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.15.1.spds9#17
  121687604  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.3.1.spds9#172
  121744789  2024-09-26T09:28:18Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.6.1.spds9#172
  121552599  2024-09-26T09:28:18Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.9.1.spds9#172
  122006774  2024-09-26T09:28:36Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.1.1.spds9#172
  121887386  2024-09-26T09:28:36Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.10.1.spds9#17
  121467967  2024-09-26T09:28:36Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.13.1.spds9#17
  121610787  2024-09-26T09:28:35Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.4.1.spds9#172
  122108409  2024-09-26T09:28:35Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.7.1.spds9#172
  121732332  2024-09-26T09:28:48Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.11.1.spds9#17
  121738303  2024-09-26T09:28:48Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.14.1.spds9#17
  121704998  2024-09-26T09:28:48Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.2.1.spds9#172
  121514035  2024-09-26T09:28:48Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.5.1.spds9#172
  121711109  2024-09-26T09:28:49Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/data/rddmuser/experianbureau_hd_202409.dpf._spds1_rdl_metadata_rddmuser.8.1.spds9#172
[srvapplemsas04@gll1831575789 soft_link]$ gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/index/rddmuser/experianbureau_hd_202409* | grep "2024-09-26T09:*
    3874816  2024-09-26T09:28:18Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/index/rddmuser/experianbureau_hd_202409.hbxmatch_key._spds1_rdl_metadata_rddmuser.0.1
    1556480  2024-09-26T09:28:18Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/index/rddmuser/experianbureau_hd_202409.idxhd_urn_key._spds1_rdl_metadata_rddmuser.0.
    1466368  2024-09-26T09:28:18Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/index/rddmuser/experianbureau_hd_202409.idxmatch_key._spds1_rdl_metadata_rddmuser.0.1
   39239680  2024-09-26T09:28:36Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/index/rddmuser/experianbureau_hd_202409.hbxhd_customer_no._spds1_rdl_metadata_rddmuse
   39272448  2024-09-26T09:28:36Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds2/rdl/index/rddmuser/experianbureau_hd_202409.hbxhd_urn_key._spds1_rdl_metadata_rddmuser.0.
      24576  2024-09-26T09:28:47Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/index/rddmuser/experianbureau_hd_202409.hbxsource_id._spds1_rdl_metadata_rddmuser.0.1
    1564672  2024-09-26T09:28:47Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/index/rddmuser/experianbureau_hd_202409.idxhd_customer_no._spds1_rdl_metadata_rddmuse
    1228800  2024-09-26T09:28:47Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds3/rdl/index/rddmuser/experianbureau_hd_202409.idxsource_id._spds1_rdl_metadata_rddmuser.0.1
[srvapplemsas04@gll1831575789 soft_link]$ gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/metadata/rddmuser/experianbureau_hd_202409* | grep "2024-09-26T0
    1581308  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/metadata/rddmuser/experianbureau_hd_202409.mdf.0.0.0.spds9#1727342899854010  metagene
[srvapplemsas04@gll1831575789 soft_link]$ gsutil ls -la gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds*/rdl/metadata/rddmuser/hd_experian_bureau.cdf.0.0.0.spds* | grep "202
      82900  2024-09-26T09:28:19Z  gs://to-lem94-prd-04-stb-eu-sas-lemans-9008722f/arch-poc/spds1/rdl/metadata/rddmuser/hd_experian_bureau.cdf.0.0.0.spds9#1727342899805112  metageneration
[srvapplemsas04@gll1831575789 soft_link]$
```

3.4.2 Ad-hoc File Restoration Process

Run team restores the Filestore data archived in cloud storage to Filestore instances. The requests can include restoring multiple files, single cluster table member, full table restoration etc. To allow segregation of duties and read-only access to GCS buckets, we are creating a new compute engine instance in prod-prod for restoration activities only. RHEL service account of this VM will have limited permissions, different than other housekeeping VMs.

4. Build Script Guidelines & Known Issues

4.1 Guidelines for Build Scripts

1. Use --continue-on-error - If any operations are unsuccessful, the command will exit with a non-zero exit status after completing the remaining operations. This is useful to prevent the complete batch archival job from failing if the underlying files have any errors preventing them from copying. Exiting with non-zero code will result in error log and alerting via incident. Note, this flag only applies to the actual copying or rsync operation. If an error occurs while iterating over the files, an internal operation in gcloud rsync and copy commands, gcloud will abort the entire rsync operation. Non-utf 8 encoding issue with filenames is an example where this flag will not prevent the entire archival operation from failing, this issue is explained in the later sections of this page.
2. Use --preserve-posix - Enables POSIX attributes to be preserved when objects are copied. With this feature enabled, gcloud storage will copy several fields provided by the stat command: access time, modification time, owner UID, owner group GID, and the mode (permissions) of the file.
3. Use --ignore-symlinks - Ignore file symlinks instead of copying what they point to. Enabled by default, Usage explained in the later section.
4. Restoration script should provide user parameters in form of environment variables or configuration file.

5. Separate scripts for archival of each instance or code logic to avoid impact of failure of archival of one Filestore instance should not effect the other.
6. Write the logs to a log file, this log file is used by Dynatrace to raise alerts. Error scenarios should be tested in testing for fatal errors, some error scenarios can be -
    a. query date is missing, incorrect, format issues
    b. file name of paths are incorrect, or missing
    c. gcloud command crashes
    d. permission issues on the files or destination bucket operations
    e. errors returned by gcloud commands - error codes
7. Use automatic retries if the archival script fails. `gcloud storage` commands retry the errors without requiring us to take additional action but we can customise the retry parameters as per our testing outcomes. Retry strategy | Cloud Storage | Google Cloud
8. Testing should include causing the gcloud command to crash, and testing the alerting and error handling works.
9. Testing should include multiple data types in tables, such as SCD2 table in SPDS clusters.

4.2 Known Issues and Remediation

4.2.1 Non-utf8 Issue

gcloud storage rsync command fails when a non-utf8 character is encountered in the file name. You might see error messages like *"'utf-8' codec can't encode characters in position...: surrogates not allowed"* if you are trying to sync files that have non-utf8 characters.

This issue is observed in Teams folder files created by users and does not impact any batch jobs, as confirmed by AESM. This issues doesn't get picked up by flags(e.g. --continue-on-error) setup in the gcloud commands to avoid failure of the archival job. To avoid failure of archival process , we will run a housekeeping job every week before archival job, that can rename such files and provide logging for run team to identify the files. Solution should log the original and renamed files in a location accessible to run team and AESM team. It is the responsibility of AESM or Run team to inform users for changes made to the files as part of this process.

Details of the solution implementation is available in Filename Sanitization Script Overview - Data Migrations - Lloyds Banking Group Confluence

4.2.2 Handling Symbolic Links

Google cloud rsync tool will follow symbolic links and sync the target of the symbolic link to Google cloud storage, not the symbolic link itself. This means that if you run the gcloud storage rsync command on a directory containing symbolic links, the tool will copy the content of the symbolic link points to. Running gcloud storage rsync or gsutil rsync over a directory containing operating system-specific file types(e.g. symlinks, device files, sockets etc) can cause various problems such as:

1. Cannot preserve symlinks while restoration using cp commands when the files were originally transferred using rsync.
2. Running gcloud rsync over directories with broken symlink can abort the archival job
3. Skips symlinks that point to directories

Key challenge with symlinks in cloud storage is that it does not inherently maintain the symlink relationship the same way local filesystems will do.

We will take an additional backup of the symbolic link metadata and create symbolic links manually after restoring the content from GCS. The original files symlink point to will be copied using rsync operation. We will do this by:

1. Maintaining a list of symlinks and their targets as part of archival process

```
find /path/to/directory -type l -exec ls -l {} \; > /filestore/home/Symlink/<filestore_instance>_symlinks_and_targets_<yyyy-mm-dd>.txt
```

2. Copy(via cp command) symlink metadata along with the data during archival process, create a folder for symlink metadata in the bucket

```
gcloud storage rsync symlinks_and_targets_<yyyy-mm-dd>.txt gs://<archival-bucket>/
```

3. Download the data files and then symlink files on local. Using scripts to recreate symlinks on Filestore after restoration of files from GCS.

```
gcloud storage cp gs://<archival-bucket>/symlinks_and_targets_<yyyy-mm-dd>.txt /path/to/directory/
```

4. Recreate the symlinks from downloaded files, build team can reference the code below. Script should run as the owner of the respective folder similar to restoration of archived data. We noticed some symlinks are created as root owners in Team folders. This needs rectification in data and discussion with AESM.

Sample script

```
LOCAL_DIR

#Define the output files for symlinks and targets

SYMLINK_FILE="symlinks.txt"
TARGET_FILE="targets.txt"

# Extract symbolic link paths
awk '{print $9}' symlinks_and_targets_<yyyy-mm-dd>.txt > $SYMLINK_FILE

# Extract target file paths
awk '{print $11}' symlinks_and_targets_<yyyy-mm-dd>.txt > $TARGET_FILE

# check if both symlink and target file exists
if [[ ! -f "$SYMLINK_FILE" ]] || [[ ! -f "$TARGET_FILE" ]]; then
    echo "Error: symlinks and target files don't exist"
    exit 1
fi

# restore symlinks based on symlinks.txt and targets.txt
paste $SYMLINK_FILE $TARGET_FILE | while IFS=$ '\t' read -r symlink target; do
    if [[ -e "$symlink" ]]; then
        echo "skipping symlink: $symlink"
    else
        ln -s "$target" "$symlink"
        echo "restored symlink: $symlink -> $target"
    fi
done
```

4.2.2.1 Build Script Details

Symlink creation script

Path of symlink script -/opt/backup_script

```
#!/bin/bash
set -x

# List of source directories
source_dirs=("/sasshare01/sas" "/spds1" "/spds2" "/spds3" "/spds4" "/sasshare01/user_logs" "/sasshare01/backup" "/team/mbna" "/staging" "/team/basel"
# Log directory and file
log_dir="/var/log/archival/symlink"
timestamp=$(date '+%Y-%m-%d')
LOG_FILE="$log_dir/success_${timestamp}.log"
LOG_LEVEL="INFO"
# Define the environment variable
env="prd" # Replace with your actual environment variable
# New log directory on filestore instance
hostname=$(hostname)
new_log_dir="/user_logs/SAS94/$env/LeMansBackup/FSArchival/SymlinkInventory/$hostname"
```

```
# ========================================================================
# Script to manage symbolic links
# ========================================================================

# Step 1: Create log directories if they don't exist
mkdir -p "$log_dir"
# Check if log directory was created successfully
if [ ! -d "$log_dir" ]; then
    echo "Failed to create log directory: $log_dir"
    exit 1
fi

# Logging function with configuration options
function f_logger {
    local l_loglevel=$1 # ERROR, WARNING, INFO, DEBUG
    local l_logsource=$2
    local l_logmsg=$3
    local l_timestamp=$(date "+%Y-%m-%d %H:%M:%S")
    local l_logfile=${4:-$LOG_FILE} # Default to LOG_FILE if not provided
    local l_console=${5:-false} # Default to false if not provided
    # Log level filtering (only log messages with level >= configured level)
    declare -A log_levels=( ["ERROR"]=1 ["WARNING"]=2 ["INFO"]=3 ["DEBUG"]=4 )
    local configured_level=${LOG_LEVEL:-"INFO"}
    if [ ${log_levels[$l_loglevel]} -le ${log_levels[$configured_level]} ]; then
        echo -e "$l_timestamp $HOSTNAME $l_loglevel $l_logsource: $l_logmsg" >> $l_logfile
        if [ "$l_console" = true ]; then
            echo -e "$l_timestamp $HOSTNAME $l_loglevel $l_logsource: $l_logmsg"
        fi
    fi
}

# Function to log symbolic links and their targets with detailed information
log_symlinks() {
    local source_dir=$1
    local owner=$(stat -c '%U' "$source_dir")
    local timestamp=$(date '+%Y-%m-%d')
    local log_dir="${source_dir}/SymlinkInventory"
    local log_file="${log_dir}/symlink_${timestamp}.txt"

    # Switch to the owner of the source directory and create the inventory directory
    if id "$owner" &>/dev/null; then
        su - "$owner" -c "mkdir -p '$log_dir'"
        if [ $? -eq 0 ]; then
            if [[ "$source_dir" == "/team/cpm" ]]; then
                su - "$owner" -c "find '$source_dir' -type d \( -name 'cf_monitoring_and_reporting_team' -o -name 'mg' -o -name 'md' -o -name 'ma' -o
            else
                su - "$owner" -c "find '$source_dir' -type l -exec ls -l {} \; > '$log_file'"
            fi
            if [ $? -eq 0 ]; then
                f_logger "INFO" "log_symlinks" "Symbolic links for $source_dir have been logged in $log_file"
            else
                f_logger "ERROR" "log_symlinks" "Failed to log symbolic links for $source_dir due to permission issues"
            fi
        else
            f_logger "ERROR" "log_symlinks" "Failed to create SymlinkInventory directory for $source_dir"
        fi
    else
        f_logger "ERROR" "log_symlinks" "User $owner does not exist for $source_dir"
    fi
}

# Iterate over the list of source directories
for source_dir in "${source_dirs[@]}"; do
    f_logger "INFO" "main" "Starting processing for directory $source_dir"
    if [ -d "$source_dir" ]; then
        log_symlinks "$source_dir"
    else
        f_logger "ERROR" "main" "Directory $source_dir does not exist"
    fi
    f_logger "INFO" "main" "Completed processing for directory $source_dir"
done

# Function to copy log file with retry mechanism
copy_log_file() {
    local attempt=0
    local max_attempts=5
    local backoff=1

    while [ $attempt -lt $max_attempts ]; do
        su - "$log_owner" -c "mkdir -p '$new_log_dir' && rsync -av '$LOG_FILE' '$new_log_dir'"
        if [ $? -eq 0 ]; then
            f_logger "INFO" "main" "Log file has been successfully copied to $new_log_dir using rsync"
            return 0
        else
            f_logger "ERROR" "main" "Failed to copy log file to $new_log_dir using rsync, attempt $((attempt + 1))"
            sleep $backoff
            backoff=$((backoff * 2))
            attempt=$((attempt + 1))
        fi
    done

    f_logger "ERROR" "main" "Failed to copy log file to $new_log_dir after $max_attempts attempts"
    return 1
}

# Create new log directory and copy log file as the owner of /user_logs
log_owner=$(stat -c '%U' /user_logs)
if id "$log_owner" &>/dev/null; then
    copy_log_file
else
    f_logger "ERROR" "main" "User $log_owner does not exist for /user_logs"
```

```
fi

f_logger "INFO" "main" "Symbolic link inventory information has been collected for all directories. Logs have been moved to $new_log_dir. Script comp
```

Symlink restoration script

```
#!/bin/bash

# Prompt the user for the symlink file
read -p "Enter the symlink file name: " SYMLINK_FILE

LOG_DIR="/var/log/archival/symlink"
LOG_FILE="$LOG_DIR/restore_symlink.log"
LOG_LEVEL="INFO"

# Ensure the log directory exists
mkdir -p "$LOG_DIR"

# Check if log directory was created successfully
if [ ! -d "$LOG_DIR" ]; then
    echo "Failed to create log directory: $LOG_DIR"
    exit 1
fi

# Logging function with configuration options
function f_logger {
  l_loglevel=$1 # ERROR, WARNING, INFO, DEBUG
  l_logsource=$2
  l_logmsg=$3
  l_timestamp=$(date "+%Y-%m-%d %H:%M:%S")
  l_logfile=${4:-$LOG_FILE} # Default to LOG_FILE if not provided
  l_console=${5:-false} # Default to false if not provided

  # Log level filtering (only log messages with level >= configured level)
  declare -A log_levels=( ["ERROR"]=1 ["WARNING"]=2 ["INFO"]=3 ["DEBUG"]=4 )
  configured_level=${LOG_LEVEL:-"INFO"}
  if [ ${log_levels[$l_loglevel]} -le ${log_levels[$configured_level]} ]; then
    echo -e "$l_timestamp $HOSTNAME $l_loglevel $l_logsource: $l_logmsg" >> $l_logfile
    if [ "$l_console" = true ]; then
      echo -e "$l_timestamp $HOSTNAME $l_loglevel $l_logsource: $l_logmsg"
    fi
  fi
}

# Start logging
f_logger "INFO" "restore_symlinks" "Starting symlink restoration process..."

# Read the file line by line
while IFS= read -r line; do
  # Extract the source, target, owner, and group from the line
  source=$(echo $line | awk '{print $9}')
  target=$(echo $line | awk '{print $11}')

  # Create the symlink and log the action
  if ln -s "$target" "$source"; then
    f_logger "INFO" "restore_symlinks" "Successfully created symlink: $source -> $target"
  else
    f_logger "ERROR" "restore_symlinks" "Failed to create symlink: $source -> $target"
  fi
done < "$SYMLINK_FILE"

f_logger "INFO" "restore_symlinks" "Symlinks restored successfully."
```

5. Proof of Concept

Details of proof of concepts done for various data types in archival and restoration are provided in Archival Proof of Concept - Data Migrations - Lloyds Banking Group Confluence

6. References

Archival and Restoration using GCS - LLD - Data Migrations - Lloyds Banking Group Confluence
Filestore Backup, Snapshots and Archival Design - Data Migrations - Lloyds Banking Group Confluence