

**Investment Domain Modernization on AWS**

**Deep-Dive Reference & Word-Document Blueprint**

**Executive Summary**

Modernizing the Investment Domain on AWS demands a secure, cost-effective and governable data-lake / warehouse architecture. Storing raw data in Amazon S3 (Iceberg format) and curated data in Amazon Redshift unlocks serverless querying (Athena), scalable warehousing, and cross-account data sharing. Robust ingress / egress pathways, coupled with CloudWatch-centric observability, ensure regulatory compliance and operational excellence[1][2][3].

**Recommended Word-Document Structure**

|  |  |  |
| --- | --- | --- |
| Section | Purpose | Key Content |
| Cover Page | Branding & confidentiality banner | Title, classification, author, version |
| Revision History | Transparent change log | Table: Version, date, editor, summary[4][5] |
| Table of Contents | Auto-generated | Use **References ➜ Table of Contents** |
| 1. Overview | Scope & guiding principles | Cloud adoption rationale |
| 2. Storage | S3 Iceberg, Redshift, S3 Tables[2][6] |
| 3. Ingress Patterns | Pull & Push methods (deep-dive below) |
| 4. Egress Patterns | Push & Pull methods |
| 5. Consumption | BI/ML paths (Athena, Redshift, Snowflake, JDBC, SharePoint, etc.) |
| 6. Metrics & Logging | CloudWatch, CloudTrail, log filters[7][8] |
| 7. Security & Governance | IAM, SCP/RCP data perimeter[9][10] |
| 8. Cost & Performance | Partitioning, compression, concurrency scaling[11][12][13] |
| 9. Risks & Mitigations | Technical, operational, compliance |
| 10. Appendices | Glossary, architecture diagrams, IAM policies |

**Versioning tip:** include version number in both file name *and* footer (e.g., *Investment\_Domain\_Modernization\_v1.0.docx* and “Version 1.0” in footer)[14]. Store drafts in SharePoint or OneDrive to leverage built-in version history[15][16].

**1 — Data Storage Architecture**

|  |  |  |  |
| --- | --- | --- | --- |
| Option | Pros | Cons / Limits | Security |
| **Amazon S3 + Iceberg** | Open table format, time-travel, schema evolution[2][17] | Requires catalog (Glue, AWS S3 Tables), eventual consistency | SSE-KMS, bucket policies, VPC endpoint |
| **Amazon Redshift (RA3)** | Massively parallel, auto-materialized views, concurrency scaling[3][18] | Higher $/TB vs. S3; cluster management | VPC only, SG/NACL least-privilege[19] |
| **Amazon S3 Tables (Preview)** | Fully-managed Iceberg warehouse; 3× query throughput[2] | Regional availability, preview SLA | Table-level IAM, automatic maintenance |

**2 — Data Ingress Patterns (Deep Research)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pattern | Flow | Typical Throughput | Pros | Cons / Limits | Security / Best Practice |
| **API Pull (Amazon API Gateway + Lambda → S3)** | Pull | Up to 10 K rps | Near real-time, serverless scaling | API limits, pagination | WAF for OWASP Top-10[20]; SigV4 auth |
| **Database Pull (DMS / Glue JDBC)** | Pull | MB–GB/hr | CDC, minimal code | Source load, network ACLs[21] | TLS, IAM Secrets Manager rotation[22] |
| **Transfer Family (SFTP)** | Push | Up to 600 MB/s[23] | Managed keys, workflows with PGP[24] | Cost per endpoint, port 22 scanning | VPC-internal endpoint; IP allowlist[25] |
| **AWS DataSync (NAS/SMB/HDFS)** | Push/Pull | 10 Gb/s per task[26] | Compression, incremental, checksums[27] | Agent deployment | TLS-encrypted; Direct Connect bypass option[28] |
| **SharePoint / Confluence (AppFlow)** | Pull | API-limited | Low-code SaaS sync[29][30] | SaaS rate limits; HTML cleanup | OAuth 2.0 tokens; limit PII export[31] |
| **AWS CLI / SDK (manual)** | Push | Human-speed | Simple PoC | Error-prone, no auditable workflow | Deny-all SCP; MFA-protected IAM |

**3 — Data Egress Patterns**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pattern | Flow | Pros | Cons / Limits | Security |
| **Athena / ODBC/JDBC** | Pull | Serverless, pay-per-scan | Requires partition/format tuning[11][32] | Glue policies; SSL-JDBC[33] |
| **Redshift Data API** | Pull | IAM-based; returns JSON | 100 MB result cap | KMS auth; rotate secrets |
| **S3 Shortcut (MS Fabric)** | Pull | Familiar to analysts | Read-only ACL, no Iceberg writes | IAM user with scoped policy[34] |
| **DataSync → On-Prem NAS** | Push | Verify checksums, throttling[35] | Direct Connect fees | Private VIF; AES-256 TLS |
| **Transfer Family SFTP** | Push | PGP encryption workflows[24] | Endpoint cost | IP-based SG; CloudTrail logs |

**4 — Data Consumption & Query Performance**

1. **Athena** – enable cost-based optimizer, partition pruning, columnar formats[36][12].
2. **Redshift** – RA3 with managed storage; short-query acceleration; auto-table-optimization[37].
3. **Snowflake External Tables** – cross-account IAM role granting READ on S3 (least privilege)[38][39].
4. **BI Tools (JDBC/ODBC)** – enforce SSL & IAM database authentication; rotate secrets via Secrets Manager[22][40].

**5 — Metrics, Logging & Observability**

|  |  |  |
| --- | --- | --- |
| Layer | Key Metrics | Source |
| **Ingress** | DataSync BytesTransferred, Transfer Family FilesUploaded | CloudWatch[41] |
| **Storage** | S3 RequestCount, S3 4xxErrorRate, Iceberg snapshot age | CloudWatch, Glue metrics |
| **Warehouse** | Redshift ConcurrencyScalingActiveClusters, WLM Queue Wait[42] | CloudWatch |
| **Queries** | Athena EngineExecutionTime, DataScannedInBytes[11] | CloudWatch Logs Insights |
| **Security** | CloudTrail EventCount by API, IAM AssumeRole failures | CloudTrail Lake |
| **Cost** | AWS Cost Explorer – S3 GET/PUT, Athena $ per TB | CUR + Athena |

Use CloudWatch **metric filters** to turn log lines (e.g., BytesTransferred) into custom KPIs and alarms[8][43].

**6 — Security & Governance Highlights**

1. **Least-Privilege IAM** – Deny \*:\* except needed actions; use cross-account roles for S3 sharing, not hard-coded keys[38][44][45].
2. **Network Controls** – SGs for instance-level, NACLs for subnet-level (stateless)[21][19]. Deny 0.0.0.0/0 ingress where possible[46].
3. **Encryption** – TLS 1.2+ for JDBC/ODBC[40][47]; SSE-KMS or client-side for PII.
4. **Data Perimeter** – SCPs/RCPs restricting S3 actions to corporate IP ranges prevent exfiltration[9][10].
5. **Supply-Chain Patching** – Keep Redshift drivers ≥ 2.1.0.32 to avoid CVE-2024-12744[48].

**7 — Cost & Performance Optimization**

|  |  |  |
| --- | --- | --- |
| Component | Action | Benefit |
| S3 Iceberg | Parquet + Snappy; 128 MB objects; partition by date-bucket | 30-70% scan reduction[32] |
| Athena | Use CBO, LIMIT+ORDER, correct join ordering[11][49] | 2× faster queries[36] |
| Redshift | Auto-materialized views, RA3 elastic storage, concurrency scaling[13] | 35× throughput burst |
| DataSync | Parallel tasks, exclude unchanged files[26][35] | Shorter migration windows |
| Transfer Family | Batch into workflows to minimize per-GB costs[50] | Lower transfer bill |

**8 — Reference Architecture Snapshots**

* **Hybrid Pull** – On-prem Oracle → DMS → S3 (Iceberg) → Athena / Redshift Spectrum.
* **Partner Push** – Vendor SFTP → Transfer Family → S3 staging → Glue ETL → Curated Redshift.
* **Cross-Account Sharing** – Producer bucket with resource policy granting arn:aws:iam::<consumer>:role/DataLakeReader plus Lake Formation tag-based access.

**9 — Authoring the Word File**

1. **Create Template**: *File → New → Custom Office Templates → “Investment\_Modernization.dotx”*[51][52].
2. **Insert Quick Parts** for **Document Property** fields (Title, Author) so cover-page metadata auto-flows[53].
3. **Automatic Revision Table**: Use 3-column footer (Date | Page # | Version) linked to StyleRef “Version” field[4][14].
4. **Navigation**: Apply “Heading 1/2/3” styles; Word will auto-generate Table of Contents and keep hyperlinks intact[54].
5. **Co-Authoring & Version History**: Store the template in SharePoint; turn on **Require Check-Out** and **Major/Minor versions** to track d0.1…v1.0 lifecycle[15][16].

**10 — Appendices**

A. Glossary (SCP, RCP, DataSync, Iceberg, CBO, RA3, etc.)  
B. Sample IAM Inline Policies (Least privilege S3 access)  
C. Cost Model Spreadsheet links (CUR query examples)  
D. Links to AWS Prescriptive Guidance papers & whitepapers used

**Next Steps**

1. Copy the content sections above into your Word template following Section 9.
2. Embed architecture diagrams ([draw.io](http://draw.io) or PNG exports) under Appendix D.
3. Route draft **v0.5** for peer review; after approvals, update revision table and promote to **v1.0**.

With this blueprint and the in-depth pattern analysis, your Word deliverable will satisfy auditors, architects, and engineering teams alike—while remaining maintainable through future iterations.