Keep F64: summarise all history, retrieve policy detail only when needed

Architecture working-session comparison | 12 November 2026 | Public evidence only; no tenant inspection. One-page A3 landscape print layout. |

Recommendation: test Option 1a first. Use Direct Lake on OneLake monthly/daily summary tables, small **Import dimensions**, and explicit policy-level DirectQuery to Snowflake. This retains all 15 years at both grains without another detail estate. **Option 2 is the F64 alternative** if automatic routing or Excel usability wins and its refresh/memory/security tests pass. Neither option has been proven on customer data. Scaling is not the starting recommendation.

Criterion	1a - Lead DL summary / Snowflake DQ	1b - Variant DL summary / OneLake DQ	2 - F64 alternative Import aggregation / DQ	3 - Last resort Scale unchanged DL fact	0 - Current reference Full detail DL on F64?
F64 at 1.5B / 4B; headroom	**Feasible at both, conditional.** Detail DQ avoids DL row cap. Illustrative summary 36M: **1.464B / 97.6%** row headroom.	Same row feasibility. 4B Delta storage is not a DL table in the model. SQL endpoint adds F64 compute.	**Feasible at both, conditional.** DL 1.5B cap not applicable. Import constrained by 25 GB/model incl. operations.	**Not F64.** F128: 3B cap, fails at 4B by 1B. F256: 6B cap; 2B headroom at 4B. Memory still tested.	1.5B is at cap: zero margin. At 4B: **2.5B over**. Approximate count cannot diagnose the actual failure.
All 15 years	Retain all 15y at summary grain **and** policy detail. Monthly 36M example is an assumption, not measured. No history truncation. Daily grain offered if requirements justify it.			All 15y at policy grain, if complete source history exists and capacity tests pass.	History coverage not established from the reported 1.5B rows.
Summary / policy query path	Pivot/report -> named summary measure -> DL aggregate. Named policy measure / filtered detail pivot -> DQ Snowflake.	Summary -> DL aggregate. Policy -> DQ SQL endpoint -> copied OneLake detail.	Same detail-table business measures -> automatic Import hit; unsupported grain/measure -> DQ Snowflake (preferred).	Both questions -> same policy-grain Direct Lake fact on larger SKU.	Both -> full fact DL; OneLake over-limit fails, SQL may fall back to DQ if enabled.
Dimensions / relationships / tooling	**Import dims preferred** for small caches and Excel hierarchies. Regular 1:* to DL aggregate; **limited** cross-source 1:* to DQ detail, even for OneLake SQL. DL+Import+DQ supported only with DL on OneLake; combine DQ using **XMLA**. No automatic DL aggregation routing.		Same-source **Dual** dimensions: regular relationships to Import aggregate / DQ detail. Desktop Manage aggregations mappings.	No model grain change. Existing DL dimensions / Excel restrictions remain.	Direct Lake type, relationships and fallback configuration unknown.
Excel versus SSAS cube	Named Summary / Policy measures; use separate filtered policy pivot. Power BI drillthrough is not an Excel pivot button. Import dim hierarchies supported; DL hierarchies and session-scoped MDX not supported. Workbook compatibility gate.		Most transparent summary/detail switch if MDX queries hit aggregates. Not full SSAS parity; **SSO-enabled sources ignore Import aggregates**.	More capacity does not restore unsupported MDX or Direct Lake hierarchies.	Excel share and cube-only features unconfirmed; test actual workbooks.
Copies / duplication	**1 detail estate:** Snowflake. Plus dbt summary, OneLake summary, dim cache and transient DL pages.	**2 detail estates:** Snowflake + OneLake. Plus Delta summaries and dim cache; justify separately.	**1 detail estate** with Snowflake DQ; plus source summary, Import summary and Dual dim cache. OneLake variant creates second detail copy.	**2 detail estates:** Snowflake + OneLake; engine cache extra, no durable Import detail.	**2 detail estates:** Snowflake + OneLake, plus query cache.
Load / refresh / F64 load	dbt aggregate -> daily small publication -> Delta -> dim refresh + reframe. Detail bypasses copy; source freshness watermark required. DQ still consumes F64 orchestration and Snowflake compute/egress.	Daily full-grain incremental copy -> SQL visibility -> Spark Delta summary -> dim refresh + reframe. Copy reliability still critical; Spark + SQL + PBI contend on F64.	Refresh source aggregate + Dual cache directly; incremental partitions if needed. No copy dependency in Snowflake variant. Refresh memory competes with queries.	Same daily copy and framing path; bigger SKU does not repair stuck transfers.	Reported daily copy variable minutes/hours; verify root cause separately from row guard.
Change / owner / effort	**L overall:** data team/dbt designs grain and publication; BI team/XMLA rebuilds facts/measures; capacity/security owner configures identity.	**L:** same BI work plus data engineer copy/Delta processing and operational ownership.	**M-L:** BI team maps aggregations and Dual dimensions; data team source summaries; identity owner approves non-SSO/RLS.	**S provisioning, M validation:** capacity admin; no model-change benefit. Do not execute without approval.	**S diagnosis:** data/BI/capacity owners capture error, row groups, SKU and freshness.
Risk / preview / first test	Limited join key cardinality, AWS latency, snapshot mismatch; no preview dependency. **First:** XMLA + Excel filter test.	Copy stalls, SQL/Spark contention; optional MLV core GA, some refresh/authoring features **preview**; Spark Delta baseline. **First:** watermark test.	Import memory, missed rewrites, SSO/security and staleness. No preview required. **First:** compress candidate + trace Excel cache hits.	Only after both F64 patterns fail agreed tests for a capacity-solvable reason. **First:** approved synthetic benchmark.	**First:** exact error, actual SKU, DL type, fallback, rows and row groups. 158 files alone is not the cause.
List capacity cost / month	**F64: $8,409.60 PAYG at 730h; $5,002.67 reservation monthly equivalent.** Existing F64 is not free; no SKU increase. Excludes Snowflake, egress, storage, licenses and engineering.			**F128 $16,819.20 / $10,005.33**; **F256 $33,638.40 / $20,010.67** (PAYG / reserved equivalent). Costs below apply to either listed region.	F64 cost only if still assigned; existing F256 availability and spare headroom not confirmed.

Public list prices: East US AND South Central US (USD)

SKU	PAYG / hour	1y reservation total	3y reservation total
F64	$11.52	$60,032	$180,096
F128	$23.04	$120,064	$360,192
F256	$46.08	$240,128	$720,384

Cost increase avoided: F128 +$8,409.60 PAYG / +$5,002.67 reserved equivalent per month; F256 +$25,228.80 / +$15,008.00. Azure MCP retrieved $0.18/CU-hour and $938/CU-year ($2,814/CU/3y); derived SKU totals, not negotiated quotes. Both reservation terms have the same monthly equivalent in returned records.

Other routes: Plain full-detail Import + incremental refresh is not ruled out by row count, but needs an all-history memory/refresh proof. Snowflake mirroring is documented without a preview marker (explicit GA wording not found on Learn); it still replicates detail. S3 shortcuts can expose eligible external Iceberg/Delta, not native Snowflake tables. See evidence appendix.

One-week go/no-go and open questions

Proposed, not product SLAs: summary p95 <=3s warm / <=8s cold; filtered policy p95 <=15s; correctness and required Excel interactions 100%; end-to-end daily availability <=60m; peak model memory <=20 GB target (25 GB SKU upper bound); no throttling or errors at agreed concurrency. Begin with 10 summary + 2 detail users, then actual peak. Both F64 patterns must fail after tuning before considering Option 3.

Confirm: still F64? F256s available? DL on OneLake or SQL, fallback enabled? Does 4B represent complete 15y? Required grain/dimensions, daily vs monthly, nonadditive measures? Excel share, session MDX, drillthrough, policy usage and concurrency? Source SSO/RLS requirements? Where are source/compute regions and publication cutoffs? Can a second detail copy be avoided?

Evidence: DL limits, UDA, Excel | source groups / XMLA | DL + Import regular joins | aggregation mappings / Dual / SSO | memory / DQ limits | reservation units. All operational counts, workload shares, grain cardinalities, performance, memory budgets and effort are unverified assumptions/estimates, not Learn claims.