Detailed Notes – Historical Risk and P&L Data Storage Initiative

# 1. Core Objective

To enable structured querying of historical risk and P&L data for analysis, reporting, and audit purposes. The initiative is not aimed at primary trading or real-time risk feeds, but instead focuses on enabling strategists, risk analysts, and finance users to access deep historical datasets in a robust, performant, and user-friendly manner.

# 2. Scope & Use Case Breakdown

## Use Case A – Historical Risk & P&L Data Analysis (Strats)

- Data required includes netting set-level risks (delta, vega, cross-gamma, etc.), trade-level metadata, and explained P&L.  
- Tools used include CortexScript/EUC querying `.dat` files on a network drive.  
- Current data is fragmented across thousands of files and prone to failure.  
- Kumar noted issues with compressed files particularly affecting P&L explain reliability.  
- Daily year-to-date P&L queries are performed, requiring access to up to a year of data.  
- Users need near-instant responses to business queries, e.g., total premiums by year.

## Use Case B – Full Risk Stack Access (Collins)

- Requires access to all historical risks including IR delta, IR vega, FX vega, and cross-gamma.  
- Collins emphasized the inclusion of cross-gamma as essential.  
- Use case involves visualizing risk evolution across various risk dimensions.  
- Must preserve data granularity and completeness.

## Use Case C – Metadata & Auditability

- Trade-level attributes such as booking information and sanity attributes are crucial.  
- Needed for audit trails, sanity checks, and reconciliations.

## Use Case D – Dual Versions / Discontinuity Management

- Batch reruns and releases create discontinuities in the time series.  
- System must support retaining multiple versions for the same date.  
- Views must allow filtering and selection logic to carve out these discontinuities.

# 3. Data Characteristics

- ~3 TB historical data in compressed `.dat` format.  
- ~250 MB daily incremental data.  
- Netting set-level legal form.  
- Coverage includes data back to 2020 and some from earlier years.

# 4. Data Storage & Query Architecture

- Primary Storage: ValueStore (object storage layer).  
- Viewing databases using ClickHouse: RISC View, Market Risk View, CCR View.  
- Proposed creation of a dedicated XVA Historical View for custom use cases.

# 5. Current vs. Target State

Current:  
- EUC scripts loading `.dat` files manually.  
- Fragile and complex logic to extract data.  
  
Target:  
- Centralized, queryable database view.  
- Fast, structured queries using SQL or equivalent APIs.  
- Support for overrides, dual versions, metadata tagging.

# 6. Performance & Availability

- Online, full-history access is required.  
- Expect sub-minute query times for most use cases.  
- Must avoid slow, offline data recovery processes.

# 7. Business Continuity Classification

- ValueStore currently not BC5-rated.  
- BC4 acceptable for historical data.  
- Live data use cases to be handled separately.  
- BC ratings determine support level, downtime tolerance, and infrastructure cost.

# 8. NAS Storage Limitations

- NAS has poor reliability despite BC5 rating.  
- Limited space, slow access, fragmented control.  
- ValueStore seen as a more scalable and robust alternative.

# 9. Override and Change Management

- Users often need to manually fix or override historical data.  
- Require ability to update or patch values with minimal friction.  
- View/query model should support version tracking and correction logic.

# 10. Next Steps

- Draft high-level use cases and query patterns.  
- Define and confirm viewing database architecture.  
- Engage IT for BC rating discussions.  
- Begin historical data ingestion into ValueStore.  
- Plan a proof-of-concept/demo session.  
- Design APIs and transition plan from EUC.