COMPARISON OF 3 STORAGE METHODS

### **1. Raw Storage (As-Is Storage)**

**Description:**

* Stores the raw SEC financial data in its original format (CSV, TSV, or TXT) without any transformation.
* Data is loaded directly into Snowflake tables without normalization or restructuring.

### **Advantages:**

✔ **Fast Ingestion:** Data is loaded quickly without transformation.  
✔ **Full Data Retention:** Retains all raw information, which allows for flexible downstream transformations.  
✔ **Easy Debugging:** Since it mirrors the original source, debugging and tracing errors is straightforward.

### **Disadvantages:**

✖ **Complex Querying:** Queries require additional processing (e.g., joins, filters, and transformations) to extract meaningful insights.  
✖ **Inefficient Storage:** Raw data often includes unnecessary columns, leading to high storage costs.  
✖ **Performance Issues:** Since data isn't optimized for analytics, queries can be slow and resource-intensive.

### **2. JSON Transformation (Semi-Structured Storage)**

**Description:**

* Transforms SEC financial data into JSON format before storing it in Snowflake.
* Uses a schema-on-read approach where the data structure remains flexible, and parsing occurs during query execution.

### **Advantages:**

✔ **Flexible Schema:** Easily accommodates schema changes without affecting existing records.  
✔ **Nested Data Support:** Can store hierarchical data structures efficiently (e.g., financial statements, company details).  
✔ **Better Read Performance:** Reduces the need for complex joins compared to raw storage.

### **Disadvantages:**

✖ **Slower Write Performance:** JSON parsing and transformation can add overhead during ingestion.  
✖ **Complex Querying:** Requires specialized SQL functions (e.g., FLATTEN() in Snowflake) to extract nested values.  
✖ **Higher Compute Costs:** Querying semi-structured data can be more expensive than querying structured tables.

### **3. RDBMS Approach (Denormalized Fact Tables)**

**Description:**

* Transforms and structures SEC financial data into relational tables optimized for analytical queries.
* Creates three key fact tables (Balance Sheet, Income Statement, and Cash Flow) with company and period identifiers.

### **Advantages:**

✔ **Optimized for Analytics:** Structured tables provide faster and more efficient queries.  
✔ **Data Integrity & Consistency:** Uses relational integrity to ensure consistency.  
✔ **Easier Business Insights:** Pre-aggregated financial metrics reduce the need for complex calculations during analysis.

### **Disadvantages:**

✖ **More ETL Complexity:** Requires transformation logic before ingestion, increasing development effort.  
✖ **Rigid Schema:** Schema changes require data migrations and modifications.  
✖ **Longer Load Times:** Data must be processed and structured before being inserted into tables.