E2E\_Data\_Engineering\_Project

case class Transaction (

transactionId: String, // UUID, unique identifier : **36 bytes**

amount: Float, // Transaction amount, e.g. 123.45 : **8 bytes**

userId: String, // User identifier, e.g. "user\_12345" : **12 bytes**

transactionTime: BigInt, // Unix timestamp in milliseconds : **8 bytes**

merchantId: String, // Merchant identifier, e.g. "merchant\_1" : **12 bytes**

transactionType: String, // Type of transaction, e.g. "purchase" : **8 bytes**

location: String, // Location identifier, e.g. "location\_1" : **12 bytes**

paymentMethod: String, // Payment method used, e.g. "CreditCard" : **15 bytes**

isInternational: Boolean, // Whether transaction is international : **5 bytes**

currency: String // Currency code, e.g. "NIS”: **5 bytes**

) : **120 bytes**

**Storage calculations per time period:**

**Per Hour:**

* **120 bytes × 1.2 billion = 144 GB per hour (120 × 1,200,000,000 = 144,000,000,000 bytes = 144 GB)**

**Per Day:**

* **144 GB × 24 = 3,456 GB = 3.456 TB per day**

**Per Month (30 days):**

* **3.456 TB × 30 = 103.68 TB per month**

**Per Year:**

* **103.68 TB × 12 = 1,244.16 TB = 1.244 PB per year**

**Storage Calculations with Compression (per hour):**

**Raw Data:**

**1 replica: 144 GB**

**3 replicas: 432 GB (144 GB × 3)**

**With Compression (5x reduction):**

**1 replica: 28.8 GB (144 GB ÷ 5)**

**3 replicas: 86.4 GB (28.8 GB × 3)**

**Daily Storage (24 hours):**

**1 replica: 691.2 GB (28.8 GB × 24)**

**3 replicas: 2,073.6 GB (86.4 GB × 24) = 2.07 TB**

**Monthly Storage (30 days):**

**1 replica: 20.7 TB (691.2 GB × 30)**

**3 replicas: 62.1 TB (2.07 TB × 30)**

**Yearly Storage:**

**1 replica: 248.4 TB (20.7 TB × 12)**

**3 replicas: 745.2 TB (62.1 TB × 12) = 0.745 PB**

**size = Base X (1 + growth\_size) ^n**

**size2 = 1,244.16TB X (1+ 20/100) ^2 = 1,791.59TB (1.792 PB)**

**size5 = 1,244.16TB X (1+ 20/100) ^5 = 3,095.87 TB (1.852 PB)**