1. **What are the Cloud services available to store trillions of data**

**1. Object Storage (Best for Huge Volumes / Data Lakes)**

* Designed for unstructured data (logs, images, backups, big data, IoT, analytics).
* Virtually unlimited scaling.

**Services:**

* **AWS S3 (Simple Storage Service)** → Industry standard, trillions of objects, 99.999999999% durability.
* **Azure Blob Storage** → Hot, Cool, and Archive tiers for cost optimization.
* **Google Cloud Storage** → Nearline, Coldline, and Archive options.
* **IBM Cloud Object Storage** → Scalable, resilient for AI/ML and archival workloads.
* **Oracle Cloud Object Storage** → Integrated with Oracle’s data platforms.

**🔹 2. Data Warehousing (Structured / Analytics at Scale)**

* Good for storing & querying trillions of **rows** of structured/semi-structured data.
* Used in analytics, reporting, ML pipelines.

**Services:**

* **Amazon Redshift** → Petabyte-scale data warehouse.
* **Google BigQuery** → Serverless, can handle trillions of rows with SQL.
* **Azure Synapse Analytics** → MPP architecture for enterprise data warehousing.
* **Snowflake (on AWS/Azure/GCP)** → Cloud-native data warehouse, elastic scaling.

**🔹 3. Big Data / Data Lake Services**

* For batch + real-time data ingestion, analytics, ML training.

**Services:**

* **AWS Lake Formation + S3** → Build secure data lakes.
* **Google BigLake + BigQuery** → Unified data lakehouse.
* **Azure Data Lake Storage (ADLS Gen2)** → Optimized for analytics with Hadoop compatibility.
* **Databricks Lakehouse (multi-cloud)** → Combines warehouse + lake.

**🔹 4. Databases at Scale**

* For trillions of **records** (transactions, IoT, telemetry).

**Services:**

* **Amazon DynamoDB** → Scales to trillions of items, fully managed NoSQL.
* **Google Bigtable** → Petabyte-scale NoSQL database (used for YouTube/Maps).
* **Azure Cosmos DB** → Globally distributed, scales elastically.
* **Cassandra (Managed: AstraDB, AWS Keyspaces)** → Open-source distributed NoSQL at scale.

**🔹 5. Cold / Archival Storage (Long-Term, Cheapest)**

* For trillions of records/files you don’t need frequently.

**Services:**

* **AWS Glacier / Glacier Deep Archive**
* **Azure Archive Storage**
* **Google Archive Storage**
* Best suited for backups, compliance, regulatory needs.

**🔹 6. Specialized Data Services**

* **AWS Timestream / InfluxDB Cloud / Azure Data Explorer** → For time-series data (IoT, sensors).
* **AWS Neptune / Azure Cosmos Gremlin API / GCP Graph Engine** → For graph datasets at large scale.

✅ **Recommendation:**

* If you just need to **store raw data at exabyte scale** → go with **Object Storage (S3, Blob, GCS)**.
* If you need **analytics at scale** → combine with **BigQuery, Redshift, or Snowflake**.
* For **transaction-heavy trillions of records** → choose **DynamoDB / Bigtable / CosmosDB**.
* For **cheap long-term archival** → use **Glacier / Archive tiers**.

2. Find out what are the possible parameters used by the retail companies to identify the churn rate

**🔹 1. Customer Demographics**

* Age, gender, income group
* Geographic location (urban/rural, region)
* Household size
* Lifestyle categories

👉 Helps identify if specific customer segments churn more than others.

**🔹 2. Purchase Behavior**

* **Frequency of purchases** (drop in shopping visits)
* **Recency of purchase** (how long since the last order)
* **Basket size / Order value** trends (declining spend)
* **Product category mix** (switching from high-margin to low-margin products)
* **RFM Analysis (Recency, Frequency, Monetary)**

👉 If recency and frequency decline, customer churn likelihood rises.

**🔹 3. Loyalty & Membership Indicators**

* Loyalty program membership status
* Points accumulation vs redemption patterns
* Expired or unused coupons/vouchers
* Cancellation of subscriptions or memberships

👉 Customers disengaging from loyalty programs are high churn risks.

**🔹 4. Customer Engagement**

* Interaction with brand channels (email open rates, app usage, website logins)
* Response to promotions and offers
* Social media engagement (likes, shares, reviews)
* Customer service interactions (complaints, unresolved issues)

👉 Reduced engagement signals weakening brand connection.

**🔹 5. Satisfaction & Experience Metrics**

* Net Promoter Score (NPS) / Customer Satisfaction (CSAT)
* Return/refund frequency
* Cart abandonment rates
* Negative reviews or feedback
* Customer support wait times / dissatisfaction

👉 Poor service experience is one of the biggest churn drivers.

**🔹 6. Competitive Behavior**

* Switching to competitor loyalty programs
* Using competitor apps/websites (if data from third-party trackers is available)
* Price sensitivity (waiting for competitor discounts)

**🔹 7. Financial Parameters**

* Decline in average revenue per user (ARPU)
* Drop in gross margin contribution per customer
* Discounts/coupons dependency (only purchasing during heavy sales)

**🔹 8. External Factors**

* Seasonal buying behavior (churn may appear but is seasonal)
* Economic downturn (customers cutting discretionary spend)
* Lifestyle changes (relocation, job changes)

✅ **How Retailers Use This Data**

* Build **Churn Prediction Models** (usually with ML using logistic regression, decision trees, or neural networks).
* Segment customers into **“at-risk”**, **“loyal”**, and **“churned”** buckets.
* Apply **targeted retention strategies** (special discounts, personalized offers, loyalty boosts).