

# CUs in Microsoft Fabric



How does CUs in Microsoft fabric work?

In Microsoft Fabric, **Capacity Units (CUs)** represent the compute power allocated to your Fabric workloads. They define how much processing power you have available and directly impact performance, concurrency, and responsiveness of your workloads.

## How CUs Work in Microsoft Fabric

### 1. Capacity-Based Model

- Microsoft Fabric uses a capacity-based model, where organizations purchase a specific number of CUs.
- These CUs are pooled and shared across different workloads (e.g., Data Engineering, Data Science, Real-Time Analytics, Power BI, etc.).

### 2. Consumption and Scaling

- CUs determine how fast and efficiently your workloads run.
- Different workloads consume CUs at different rates.
- You can scale capacity up or down based on demand.

### 3. Workload Management

- Fabric dynamically allocates CUs to different workloads.
- If a workload reaches the CU limit, it might slow down or queue requests.
- You can assign priorities to workloads to ensure critical tasks get enough resources.

### 4. Fabric Capacity SKUs

- Fabric offers different **capacity SKUs** (e.g., F2, F4, F8, etc.), each with a fixed number of CUs.
- Higher SKUs provide more CUs, allowing for better performance and concurrency.

### 5. Monitoring and Optimization

- You can monitor CU consumption using Fabric's **Capacity Metrics App**.
- Adjust workloads and scheduling to optimize CU usage.
- Autoscale options might be available in some licensing models.

## Example Use Case

If you're running a data pipeline in **Fabric Data Factory**, it consumes CUs based on the complexity and size of the data being processed. If Power BI reports run simultaneously, they also consume CUs, potentially slowing down both workloads if you reach your CU limit.

Would you like details on specific Fabric workloads and their CU consumption?



How much data can be processed per CU?

Microsoft Fabric does not provide a fixed "**X GB per CU**" metric because data processing efficiency depends on several factors:

1. **Workload Type** – Different workloads consume CUs differently (e.g., Data Engineering vs. Power BI).
2. **Query Complexity** – A simple SQL query on a small dataset consumes fewer CUs than a complex ETL pipeline.
3. **Parallelism & Optimization** – Well-optimized workloads process more data per CU.

## General CU Consumption Estimates

While Microsoft doesn't publish exact CU-to-data ratios, we can estimate based on workload types:

Workload	Estimated CU Consumption (Varies by Complexity)
Lakehouse (Data Engineering)	1 CU can process ~1-5 GB of data per hour in a well-optimized pipeline.
Data Factory Pipelines	1 CU can move ~2-10 GB/hour (depends on transformations, connectors).
SQL Analytics	1 CU supports basic queries on ~5-10 GB of structured data.
Real-Time Analytics	1 CU can handle thousands of events/sec, but varies by data size.
Power BI Reports	1 CU can refresh reports handling ~100K-500K rows efficiently.

## Scaling Considerations

- **More CUs = Higher Throughput:** Increasing CUs scales workloads horizontally.
- **Optimization Matters:** Indexing, partitioning, and cache utilization significantly impact how much data a CU can process.
- **Bursting & Queuing:** If CU limits are exceeded, workloads may queue or slow down.

Would you like a more detailed estimate for a specific workload?