**M/M/1 system**:

An **M/M/1 queue** is the most basic queueing model. The queuing process is a continuous-time Markov queuing process with the following characteristics of a system:

1. **Poisson arrivals (M)** → Customers arrive randomly at a mean rate λA and the inter arrival time is exponential with mean rate of 1/ λA
2. **Exponential service times (M)** → Each customer is served at a mean rate μ = λS
3. **Single server (1)** → Only one server is available.
4. **Infinite queue capacity** → No limit on the number of customers waiting.
5. **First-Come, First-Served (FCFS) discipline** → Customers are served in the order they arrive.
6. Service times and interarrival times are independent

**Applications of M/M/1 Queue**

* **Banks and ATMs**: Imagine a small branch of bank with one teller/counter (server). Customers arrive randomly to have banking services (arrivals). The time it takes to serve each customer (service time) varies but follows a random pattern with no specific memory. This scenario can be modeled as an M/M/1 queue. Customers arrive randomly and are served by a single teller or ATM.
* **Customer Service Desks**: A single employee assisting customers one by one.
* **Network Routers**: Data packets arriving at a router, where a single server processes them.
* **Public Transport Stops**: People arriving at a bus stop and waiting for a single bus to pick them up.

**Meaning of the letters:**

**M:** The first M refers to arrivals following a Poisson process. This means arrivals occur randomly at a constant average rate (λ), regardless of previous arrivals.

Suppose we have a single server (one teller) in a small branch of bank and customers arrive in the bank with an average of, say, 5 customers every 10 minutes. So, we have Poisson arrival distribution at a mean rate of *λ* = 0.5 customers per minute i.e. on average one customer appears every 1/*λ* = 1/0.5 = 2 minutes. This implies that the interarrival times have an exponential distribution with an average interarrival time of 2 minutes.

**M:** The second M signifies service times that are exponentially distributed. This implies service times are random and have no "memory." Each service time is independent of the previous one.

Think of teller transactions at the bank, where each transaction has a random duration that doesn't depend on how long the previous customer took. In our example, the server has an exponential service time distribution with a mean service rate of 4 customers per minute, i.e., the service rate µ = 4 customers per minute.

**1:** The number 1 indicates there's a single server. In the bank example, this is the teller who assists customers.

We can analyze this system using queuing theory to understand aspects like:

* Average queue length: How many customers are typically waiting in line?
* Average waiting time: How long does a customer typically spend waiting before getting served?
* Server utilization: How much of the time is the barista busy serving customers?

By understanding these metrics, the bank manager can make informed decisions, such as hiring additional staff during peak hours to reduce waiting times or optimizing the service to decrease service time variability.

**Mathematical Formulation**

* λ = Arrival rate (customers per unit time)
* μ = Service rate (customers served per unit time)
* ρ = λ / μ = Traffic intensity (must be ρ < 1 for system stability)