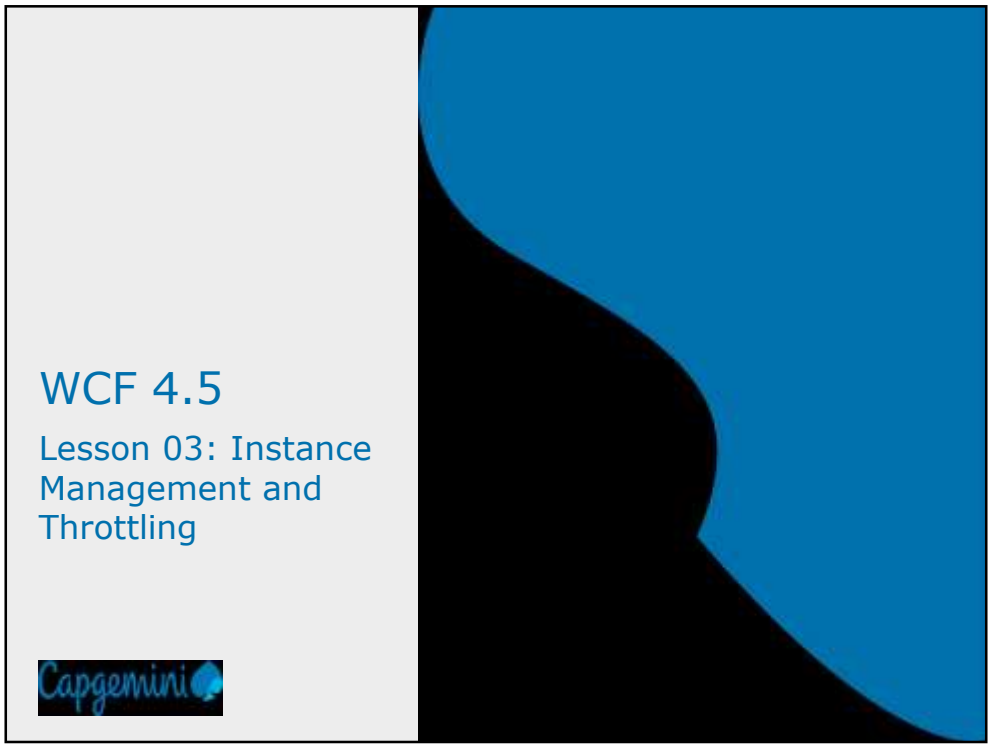
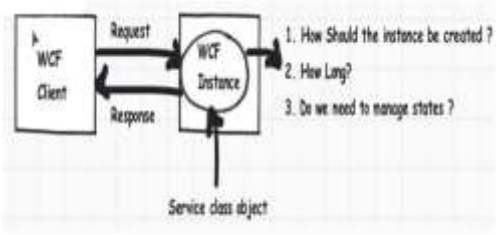
Instructor Notes:



Instructor Notes:

Instance Management

- Instance management refers to the way a service handles a request from a client.
- Instance management is set of techniques WCF uses to bind client request to service instance, governing which service instance handles which client request.
- In a normal WCF request and response communication following sequence of actions takes place:-
  - WCF client makes a request to WCF service object.
  - WCF service object is instantiated.
  - WCF service instance serves the request and sends the response to the WCF client.

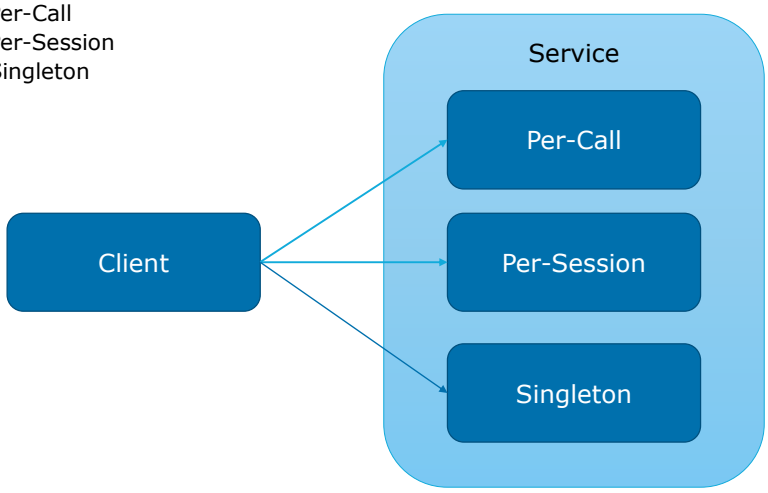


Many times we would like to control the way WCF service objects are instantiated on WCF server. You would like to control how long the WCF instances should be residing on the server.
We need Instance Management because applications differ too much in their needs for scalability, performance, throughput, transactions and queued calls. So one solution for all applications doesn't fit.
In a normal WCF request and response communication following sequence of actions takes place:-
• WCF client makes a request to WCF service object.
• WCF service object is instantiated.
• WCF service instance serves the request and sends the response to the WCF client.
Following are different ways by which you would like to create WCF instances:-
• You would like to create new WCF service instance on every WCF client method call.
• Only one WCF service instance should be created for every WCF client session.
• Only one global WCF service instance should be created for all WCF clients.

Instructor Notes:

Instance Management (Cont…)

- WCF framework has provided 3 techniques by which we can manage WCF service instance creation.
  - Per-Call
  - Per-Session
  - Singleton



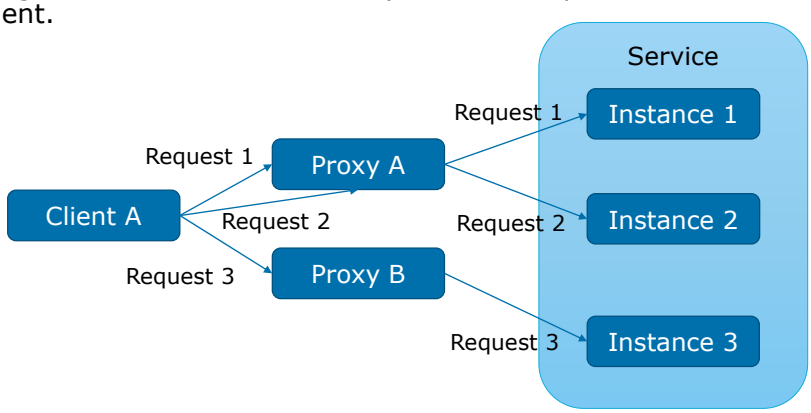
To meet the above scenarios WCF has provided 3 ways by which you can control WCF service instances:-

- Per Call
- Per session
- Single instance

Instructor Notes:

Per-Call Instance Mode

- When we configure WCF service as per call, new service instances are created for every method call you make via WCF proxy client.
- A service instance is created only when the client call is in progress and it will be destroyed after response is sent back to client.

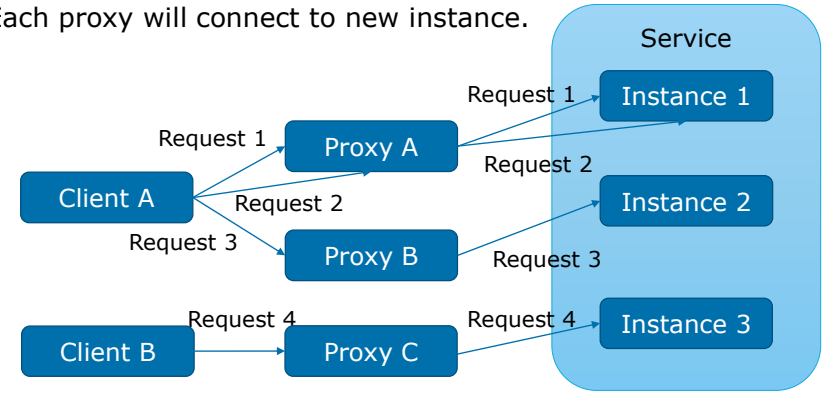


When should you use per call
• You want a stateless services
• Your service hold intensive resources like connection object and huge memory objects.
• Scalability is a prime requirement. You would like to have scale out architecture.
• Your WCF functions are called in a single threaded model.

Instructor Notes:

Per-Session Instance Mode

- One instance of the service class is created per client session.
- A logical session is maintained between client and service.
- When the client creates new proxy to particular service instance, a dedicated service instance will be provided to the client.
- Each proxy will connect to new instance.

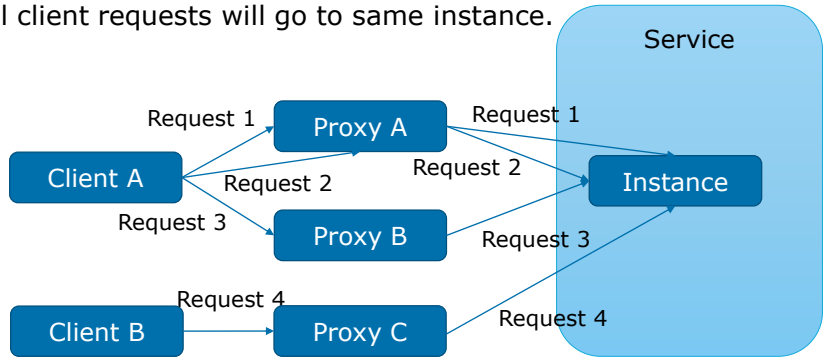


When should you use per session
- You want to maintain states between WCF calls.
- You want ok with a Scale up architecture.
- Light resource references

Instructor Notes:

Singleton Instance Mode

- One instance of the service class handles all incoming requests. This implements a singleton.
- When a service is configured as a singleton, all clients connect to the same shared service instance; it is not a factor, to which endpoint of the service they are connected.
- All client requests will go to same instance.



When should you use singleton mode
• You want share global data through your WCF service.
• Scalability is not a concern.

Instructor Notes:

Throttling

- WCF throttling provides some properties that you can use to limit how many instances or sessions are created at the application level.
- Performance of the WCF service can be improved by creating proper instance.
- Service Throttling can be configured either Administrative or Programmatically.

Attribute	Description
maxConcurrentCalls	Limits the total number of calls that can currently be in progress across all service instances. The default is 16.
maxConcurrentInstances	The number of InstanceContext objects that execute at one time across a ServiceHost. The default is Int32.MaxValue.
maxConcurrentSessions	A positive integer that limits the number of sessions a ServiceHost object can accept. The default is 10.

Throttling in WCF is used to limit the service throughput so that the resource consumption (memory, processor, disk, network, etc.) in the system is at an acceptable level, i.e., ensure the service doesn't consume resources beyond acceptable limits. The ServiceThrottlingBehavior class can be used to control the performance of WCF services.

The word throttling is mainly derived from the throughput keyword. Throughput means that the work performs in a given specific time interval. The throughput of any WCF Service mainly depends on the number, for instance, the number of sessions you create. Based on the number of instance and session, you can increase and decrease the performance for WCF Service.

Controlling the number of instance and session for WCF Service, we have to use throttling. In throttling, you have to specify the number for the concurrent call and concurrent session, and the concurrent instance allows for the client to communicate with WCF Service.