**6th module Take-Away Scripts**

1. What is a Control plane?

The components of the Anypoint Platform architecture that you use to design, deploy, and manage APIs and Mule applications. The control plane contains product features and components that are part of the Anypoint Design Center, Anypoint Management Center, and Anypoint Exchange.

2. What is a runtime plane?

The components of the Anypoint Platform architecture to which your APIs and Mule applications are deployed and made available to your users. The runtime plane includes the Mule runtime server and supporting services.

Applications are deployed to runtime plane

All data, payloads are processed in runtime plane.

//CloudHub is included in the runtime.

1. Features of the control plane?

Component, API Designer

Anypoint Studio

Flow Designer

Anypoint Exchange

Anypoint API Community Manager

Anypoint Partner Manager

Access management

Analytics

Anypoint API Manager

Anypoint Runtime Manager

Anypoint Monitoring

//Secrets manager

Anypoint Visualizer

Anypoint Security edge policies

Anypoint Security tokenization

CloudHub runtimes

Runtime Fabric

Customer-hosted runtimes

4. Features of runtime plane?

5. What is network latency?

[Network latency](https://www.keycdn.com/support/what-is-latency), sometimes called lag, is the term used to describe delays in communication over a network. Latency meaning in networking is best thought of as the amount of time it takes for a packet of data to be captured, transmitted, processed through multiple devices, then received at its destination and decoded

6. What is a throughput?

**Throughput tells you how much data was transferred from a source at any given time. Throughput** is the term given to the number of packets that are processed within a specific period of time.

**CloudHub Deployment**

1. What is CloudHub?

Cloudhub is an iPaas. It enables us to deploy and run the application in cloud via runtime manager.

It is a scalable, multi-tenant, elastic, secure and highly available ipaas.

It supports cloud-to-cloud, cloud-to-on premise and on-premise to on-premise and B2B integration.

2. How to deploy an application to CloudHub?

We can deploy an app via runtime manager, via studio, using mule maven plugin, CLI,

Flow Designer, CloudHub API.

1. What happens behind the scenes when you deploy an application to CloudHub?

Whenever we deploy an application to cloudhub a new EC2 instance is created along with a mule runtime and our deployed app inside it. We can achieve application isolation also from it. (By default we have shared load balancer fro cloudhub)

4. Advantages of deploying applications to CloudHub?

Application isolation, 0 downtime.

5. Disadvantages of CloudHub deployment model?

No domain project support, customization point(separate not possible),

(RTF, customized-----,wastage of resources, )

6. //What is a Shared Load Balancer (SLB)?

7. What are the characteristics of a SLB?

SLB url always have xxx.eu.cloudhub.io

- SLB has external IP

- SLB do not allow vanity url

- Do not have custom option for url redirect

- It is shared with multiple instances of application you deploy

- It has a rate limit to make a request. After the limit is exceeded the it throws 503 error

8. What is an EC2 instance?

Short for Amazon Elastic Computer Cloud, Amazon **EC2** is a commercial Web service from Amazon's Web Services (**AWS**) that lets customers "rent" computing resources from the **EC2** cloud. **EC2** provides storage, processing, and Web services to customers

9. What is AWS and how CloudHub is related to AWS?

10. What is zero downtime deployment?

Whenever we deploy/redeploy an app to cloudhub a new EC2 instance gets created, so the previous EC2 instance is deprecated until and only when the new instance/worker is up and running serving every request in between.

11. What is a Worker?

**Worker in cloudhub** is a dedicated instance of Mule that runs your integration application. **Workers** has following characteristics: - Capacity: Each **worker** has a specific amount of capacity to process data. You can select the size of your **workers** when configuring an application..

11. What is an availability zone?

An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region. AZ’s give customers the ability to operate production applications and databases that are more highly available, fault tolerant, and scalable than would be possible from a single data center. All AZ’s in an AWS Region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fiber providing high-throughput, low-latency networking between AZ’s. All traffic between AZ’s is encrypted.

12. What is a region w.r.t AWS data centers?

13. What is a \*\*vCore\*\* while deploying applications to CloudHub?

*A unit of compute capacity for processing on CloudHub, which is equal to one virtual core. Up to ten Mule Applications can be deployed for every VCore purchased.*

14. What is application isolation?

Own mule runtime and resources. Application is deployed to cloudhub and separate EC2 instances are created for individual apps.

15. What is vertical scalability?

To increase the no of workers or to scale out is called as vertical scalability.

16. How do you achieve vertical scalability?

Same as 15th

17. What is horizontal scalability?

To increase the vcore size for performance is called as HS.  
18. How do you achieve horizontal scalability?

We have to increase the vcore size manually to achieve horizontal scalability.

20. What is high availability?

When there are multiple workers assigned for the deployment of the same app, we have multiple apps running at the same time so we can achieve HA.

21. How can we achieve high availability in CloudHub deployment?

By increasing the no of workers while deployment we can achieve HA.

22. How can we achieve disaster recovery in CloudHub deployment?

23. When should customers choose CloudHub deployment model?

When customer don’t have an infrastructure of their own, Don't have a wider IT presence in my organization to manage the integration platform and need an agile solution to quickly build and deploy

CloudHUb VPC

1. What is a VPC?

A Virtual Private Cloud (VPC) allows you to virtually create a private and isolated network in the cloud. Just as a virtual private network (VPN) provides secure data transfer over the public Internet, a VPC provides secure data transfer between a private enterprise and a public cloud provider. This ensures that each customer's data remains isolated from other customer's data, both in transit and inside the cloud provider's network. This isolation can be accomplished using security policies that require some – or all – of the following elements: private IP addressing, tunneling, encryption, or allocating a unique VLAN to each customer.

1. What is CloudHub VPC?

The Anypoint Virtual Private Cloud (VPC) offering allows you to create a virtual, private, and isolated network segment in the cloud to host your CloudHub workers.

Connecting to your Anypoint VPC extends your corporate network and allows CloudHub workers to access resources behind your corporate firewall. You can connect on-premises data centers through a secured VPN tunnel, or a private AWS VPC through VPC peering, or by using AWS Direct Connect

1. Why do you need a VPC?

When we need direct connection pvt with DC, Vanity url, (VPN comes with VPC),

DLB out of the box.

Specific requirements of controlling incoming traffic to applications

Want to make outbound connection secure using VPN

Want to use dedicated load balancer with Anypoint Platform

1. Advantages of having a VPC?

Vanity url, pvt connection, DLB, Security, Provides ability to prevent some requests(by blocking implemenataion url and directing to proxies)(by implementing firewall rules for implementation..by changing ports for imp imp and proxy)

1. Disadvantages of having a VPC?

Costly,

1. How do you configure a VPC?

<https://apisero.com/anypoint-vpc-simplified/>

1. What is Dedication Load Balancer?

If you don’t want to expose xxx.eu.cloudhub.io to your customer then use DLB

- Used for url redirect

- Provides vanity url capability

- Provide public url

- Enables two way SSL to store certificate

- Used for IP white listing

- It has internal and external IP

1. What are the features or characteristics of DLB?

Same as 7th

1. Differences between SLB and DLB?
2. Advantages of having a DLB?

If you don’t want to expose xxx.eu.cloudhub.io to your customer then use DLB

- Used for url redirect

- Provides vanity url capability

- Provide public url

- Enables two way SSL to store certificate

- Used for IP white listing

- It has internal and external IP

1. What are the ways that we can create a secure connection between CloudHub worker and customers' networks?

By using a VPC configuration in CH

1. What is an IPsec tunnel?

IPSec’s protocol objective is to provide security services for IP packets such as encrypting sensitive data, authentication, protection against replay and data confidentiality.

1. What is direct connect?

Direct Connect is a cloud service from AWS that allows a dedicated network connection to be established between AWS and customers’ premises without using the Internet. Private connectivity can be established between AWS and an external location (e.g., data center, office, colocation environment).

<https://aviatrix.com/learn-center/glossary/direct-connect/>

<https://aws.amazon.com/directconnect/#:~:text=AWS%20Direct%20Connect%20is%20a,from%20your%20premises%20to%20AWS.&text=AWS%20Direct%20Connect%20lets%20you,the%20AWS%20Direct%20Connect%20locations>.

https://www.tutorialspoint.com/amazon\_web\_services/amazon\_web\_services\_direct\_connect.htm

1. What is VPC peering?

A VPC Peering is a network connection between two VPCs that allows traffic routing between the VPCs using private IPv4 or IPv6 addresses

Instances in either VPC can communicate with each other as if they are within the same network

A VPC Peering can be created between VPCs belonging to the same AWS account, between VPCs belonging to different AWS accounts, between VPCs in the same region or between VPCs between different regions

<https://docs.aws.amazon.com/vpc/latest/peering/what-is-vpc-peering.html>

<https://aviatrix.com/learn-center/cloud-networking/aws-vpc-peering/#:~:text=A%20VPC%20Peering%20is%20a,private%20IPv4%20or%20IPv6%20addresses>.

1. What is a vanity URL?

In the Domain Name System, a vanity domain is a domain name whose purpose is to express the individuality of the person on whose behalf it is registered. This contrasts with domain names which resolve to an organisation or a service that organisation offers.

1. How do you implement a two-way SSL using CloudHub VPC?
2. When should a customer choose a CloudHub VPC deployment model?

When customer needs a direct connection with pvt data centres, customer needs vanity url,

Security purposes, We can maintain on our own..so flexibility is provided and also when the customer require a DLB

1. What are the firewall rules of VPC?

**Rule #1**: any application running on 8081 can be reached from anywhere.

**Rule #2**: any application running on 8082 can be reached from anywhere.

**Rule #3**: any application running on 8091 can be reached from any application running inside the same Anypoint VPC. If these applications need to be exposed outside the Anypoint VPC they can be exposed using Dedicated Load Balancer.

**Rule #4**: any application running on 8092 can be reached from any application running inside the same Anypoint VPC. If these applications need to be exposed outside the Anypoint VPC, they can be exposed using Dedicated Load Balancer.

<https://apisero.com/anypoint-vpc-simplified/>

Hybrid Deployment

1. What is the hybrid type of deployment model?

With the hybrid deployment option, you deploy your apps from the Runtime Manager cloud console to your Mule servers and use Runtime Manager to manage them. This option provides you with flexibility and control over your on-premises security but requires you to provide the hosting infrastructure.

To use the hybrid option, you first register your Mule servers with the Runtime Manager agent. Then, from Runtime Manager, you can optionally add those servers to server groups or clusters to provide high availability. Finally, you deploy your applications from Runtime Manager to either a server, server group, or cluster.

Here the control plane is managed by mulesoft and runtime plane is managed by customer.

1. What are the types of hybrid deployment model?

Server, Server group and cluster

Server is a single node

1. Advantages of hybrid deployment model?

The hybrid integration model also provides a [secure data gateway](https://www.mulesoft.com/platform/soa/mule-enterprise-security)which allows for protected communication between CloudHub and Mule as an ESB.

High Availibility.

Usage of domain project.

No core limitation

Reduces licensing cost

1. Disadvantages of hybrid deployment model?

Load balancing is not provided for hybrid deployments. You can manage load balancing with the tools connected to your on-premises resources.

The Runtime Manager logging feature is not available for hybrid deployments. You can configure on-premises apps to send data to external analytics tools, such as Splunk or ELK, to manage your logs.

The object store infrastructure is not available for hybrid deployments. To use object stores, you must configure a database to store data.

After an application is deployed and running, you must apply any security updates for the selected runtime version manually.

No domain project support

Reduced out-of-the-box monitoring

Customer has to provide infra

App isolation not there

1. What is a standalone deployment?

It is a method to manually deploy a Mule application to on-premises Mule instance

1. What is a cluster?

A cluster is a set of Mule runtime engines that acts as a unit. In other words, a cluster is a virtual server composed of multiple nodes (Mule runtime engines). The nodes in a cluster communicate and share information through a distributed shared memory grid. This means that the data is replicated across memory in different machines.

Max 8 servers in a cluster.

1. What is a server-group?

A server group is a set of servers that act as a single deployment target for applications so that you don’t have to deploy applications to each server individually.  
Deploying applications to servers in server groups provides redundancy so you can restore applications more seamlessly and quickly, with less downtime.  
Unlike clusters, application instances in a server group run in isolation from the application instances running on the other servers in the group.

1. What is a hazelcast?

//In a Hazelcast grid, data is evenly distributed among the nodes of a [computer cluster](https://en.wikipedia.org/wiki/Computer_cluster), allowing for [horizontal scaling](https://en.wikipedia.org/wiki/Horizontal_scaling) of [processing](https://en.wikipedia.org/wiki/Parallel_computing) and available storage. Backups are also distributed among nodes to protect against failure of any single node. Hazelcast provides central, predictable scaling of applications through in-memory access to frequently used data and across an elastically scalable data grid. These techniques reduce the query load on databases and improve speed.

Hazelcast can run on-premises, in the cloud ([Amazon Web Services](https://en.wikipedia.org/wiki/Amazon_Web_Services), [Microsoft Azure](https://en.wikipedia.org/wiki/Microsoft_Azure), [Cloud Foundry](https://en.wikipedia.org/wiki/Cloud_Foundry), [OpenShift](https://en.wikipedia.org/wiki/OpenShift" \o "OpenShift)), virtually ([VMware](https://en.wikipedia.org/wiki/VMware)), and in [Docker](https://en.wikipedia.org/wiki/Docker_(software)" \o "Docker (software)) containers.

1. How does node replication happen in a hybrid-based deployment model?

//When we go with cluster/server gp type of deployment multiple nodes are created

1. What is the use of node replication?

For HA, it reduces the workload on a single node

1. What is distributed memory in a cluster-based deployment and how it is useful?

In a cluster based env, resources and work load are shared between nodes and replicated for high availability. This is achieved internally by hazelcast implemented by mulesoft.

1. What is unicast in a cluster and when should customers choose unicast?

A unicast cluster requires that you configure the IP addresses of the nodes in the cluster. If a server has multiple interfaces, use the internal IP address that allows the node to communicate directly with other nodes. Clustering across different subnets is not supported.

1. What is multicast in a cluster and when should customers choose multicast?

A multicast cluster comprises servers that automatically detect each other. Servers that are part of a multicast cluster must be on the same network segment.

1. How do you achieve High availability using a Cluster based deployment model?

We will have to configure multiple nodes for HA in clustering.

1. //When should a customer choose a standalone based hybrid deployment model?

Domain project

1. //When should a customer choose a cluster-based hybrid deployment model?

When HA

1. //When should a customer choose a server-group based hybrid deployment model?
2. How do you balance the load in hybrid?

We have to config ext LB eg: Nginx, apache LB

1. How can we achieve high availability in hybrid?
2. How can we achieve vertical and horizontal scalability in hybrid?

Horizontal 🡪 increasing no of servers/nodes

Vertical 🡪 Increase the java heap size in config file (depending on our local machine) ????

RTF

1. What is runtime fabric type of deployment?

Anypoint Runtime Fabric is a container service that automates the deployment and orchestration of Mule applications and API gateways. Runtime Fabric runs within a customer-managed infrastructure on AWS, Azure, virtual machines (VMs), and bare-metal servers.

Here the control plane is managed by MuleSoft and runtime is managed by the customer.

1. When should a customer choose RTF deployment model?

When the customer needs to work with their own infrastructure and server and need cloudhub features alongside this then they go for RTF.

if your org runs completely on-premise and is looking for an integration solution highly scalable and customizable

1. Advantages of using RTF?

High availability, application isolation, application failover, zero downtime, All cloudhub features in customer hosted environment.

1. Disadvantages of using RTF?

RTF only works with MuleSoft-hosted AnyPoint Platform (US or EU).[[2]](http://infiniters.com/2020/02/26/anypoint-runtime-fabric-overview-part-2/" \l "_ftn2) There is currently no support for customer-hosted control-planes.

There is no native support for ObjectStore or local storage. An external DB or file server should be used for persistent storage of application data.

No support for domain project.

1. Differences between RTF and CloudHub deployment model?

In cloudhub kind of deployment everything is managed by mule, i.e both control plane as well as runtime plane is managed by mule, but in RTF runtime is managed by customer and cloudhub features are can be achieved here easily.

Vcore and size customization

1. Differences between RTF and Hybrid deployment model?

In hybrid deployment updations have to be made manually, we need to setup load balancer manually, we can’t deploy mule 3 and mule 4 apps within the same resources, application share same resource so all app are prone to downtime.

1. What is a Docker container?

A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

1. Differences between a Docker and a VM?
2. What is Kubernetes?
3. //Features of Kubernetes?

<https://cloudacademy.com/blog/docker-vs-virtual-machines-differences-you-should-know/>

1. What is a controller?

The RTF controller is in fact a Kubernetes controller, so the orchestration, distributed database, and internal load-balancing capabilities are still being leveraged here. In addition, most control plane communication with AnyPoint Platform also happens here.

The biggest consideration when deciding on the number and size of controllers is redundancy and load-balancing capacity. Requests to apps running on the worker nodes whether sourced from inside or outside the cluster will always go through the load-balancer on the controller nodes, ensuring the same clustering and zero-downtime upgrades, rollbacks, and scaling that CloudHub customers enjoy. However, keep in mind that if there are multiple controllers, an additional external load-balancer will be required for traffic from outside the cluster.

The current maximum number of supported controller nodes is 5. Due to Kubernetes use of [raft](https://raft.github.io/) algorithm, it’s recommended to have an odd number of controllers.

1. What are the characteristics of a controller?

Same as 11th

1. What is a worker node?

The RTF worker node is where regular Mule applications will get deployed as pods. The more workers you have, the more highly-available and performant your cluster will be. However, the individual Mule application deployment configuration will decide on how much CPU, RAM, and number of replicas will be assigned to any one application.

The current maximum number of supported worker nodes is 16.

1. What is a POD?

<https://kubernetes.io/docs/concepts/workloads/pods/>

1. How do you achieve application isolation in RTF?

Every application that we deploy in RTF is deployed to a separate POD and container

So that application isolation is achieved like cloudhub.

1. How do you achieve vertical and horizontal scalability in RTF?

//Horizontal scalability is achieved by making replicas of the deploying app.

Vertical scalability can be achieved here by increasing the vCore size.

1. How do you achieve high availability in RTF?

By deploying replicas of the mule apps, we can achieve high availability.

Cluster?

1. How do you achieve zero-downtime in RTF?

Whenever we deploy an updated app of an existing app onto RTF, a new POD created for this app by keeping the old one intact. This old one gets deprecated only when the new one is up and running. In this way 0 downtime is achieved by RTF.

1. What are the minimum requirements of controller and worker nodes?

Production – 3 controller nodes and 3 worker nodes

Development – 1 controller node and 2 worker node

1. How do you balance the load in RTF?

RTF have internal load balancers within the controller nodes that directs the requests to the worker nodes. However with increase in number of controller nodes an external LB has to be configures for efficient performance.

1. What control plane capabilities are available for RTF?

PCE

1. What is PCE?

Using Anypoint Platform Private Cloud Edition (PCE) one can run and manage Mule applications on their local servers using their required security policies

 It enables to maintain control over data storage and processing of data. The control plane appliance can be installed onto the organizations own data centers, such as on-prem or Azure or AWS. Thus allows to configure and run Anypoint Platform software and maintain all data storage, processing, and transmission

Anypoint Platform PCE uses Docker and Kubernetes to provide built-in high availability and scalability.

In this the control plane components like design center and management are released into docker containers. Those docker containers are managed by the Kubernetes. The Runtime plane has the bare metal infrastructure. On top of it are the VMs and all the applications are deployed in those VM. The control plane uses the Docker and Kubernetes while the runtime plane is the collection of the clusters.

1. When should a customer choose a PCE deployment model?

Organizations in industries such as banks, insurance corps, healthcare have scenarios where they want to more tightly manage their data.

Have Strict compliance laws on data processing within our own data centers.

1. Advantages of PCE?

  Control Plane and Runtime plane are secured by the organization, No flow of metadata to the Mulesoft hosted cloud, All the exchange assets are strictly remain on premise

1. Disadvantages of PCE?

Requires expertise to manage the control plane, difficult to mange issues, cost of infra

1. What are the differences between RTF and PCE?

In RTF both control plane is managed by mulesoft and then runtime is managed by the customer however in PCE control as well as runtime plane is managed by customer.

Rtf has anypoint security, in pce we have to configure ourselves.

(AS comes with RTF by def?)

1. What control plane capabilities are available for PCE?

Slide aayasha

1. What are the basic configurations of a PCE?

There are two basic configuration 3 node and 6 node configuration.

Db and services are difff in 6 node and same in 3 node

1. How do you install the runtime plane in PCE?
2. How do you install the control plane in PCE?
3. Does PCE support application isolation and why?

PCE supports application isolation. All the applications are deployed in a docker which is managed by Kubernetes.

1. How do you achieve vertical and horizontal scalability in PCE?
2. How do you achieve zero-downtime in PCE?
3. How do you balance the load in PCE?

In PCE user have to configure load balancer separately. The load balancer is places before the nodes so that all the incoming traffic are routed properly. Generally PCE uses round robin strategy for load balancing

---------------------------------------------------

**Pivotal Cloud Foundry (PCF)**

* 1. What is PCF?
  2. PCF is a cloud native platform for deploying and operating modern applications.
  3. PCF can be deployed on-premises and on many cloud providers to give enterprises a hybrid and multi-cloud platform.
  4. It provides a uniform way for you to launch, and quickly iterate on ideas in the language of your choice.

* 1. When should a customer choose a PCF deployment model?

If customer organization has strict regulatory or compliance requirements

If customer has PCF or cloud foundry in his landscape and want to have mule control plane/runtime hosted in the same cloud platform for better latency

* 1. Advantages of PCF?

Latency: when your core systems are in one data center and you want your runtime to be in same data center

Compliance requirements: Government specific projects

GDPR and Data residency requirements

* 1. Disadvantages of PCF?

 High Availability: Customer need to configure

Load Balancing: Customer need to configure

Logging: No , need to leverage third party logging services like Splunk

Monitoring: We can only trigger Alerts , can’t see utilization info like in Cloud Hub control plane

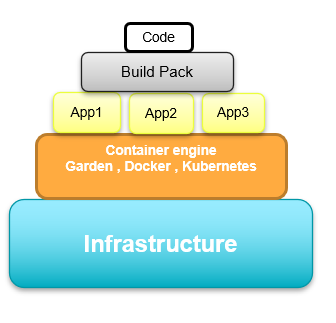
Object Store: Customer need to configure

Scheduling: No

Security Updates: Manual

Explain the basic architecture of PCF.

**Pivotal Cloud Foundry (PCF)**



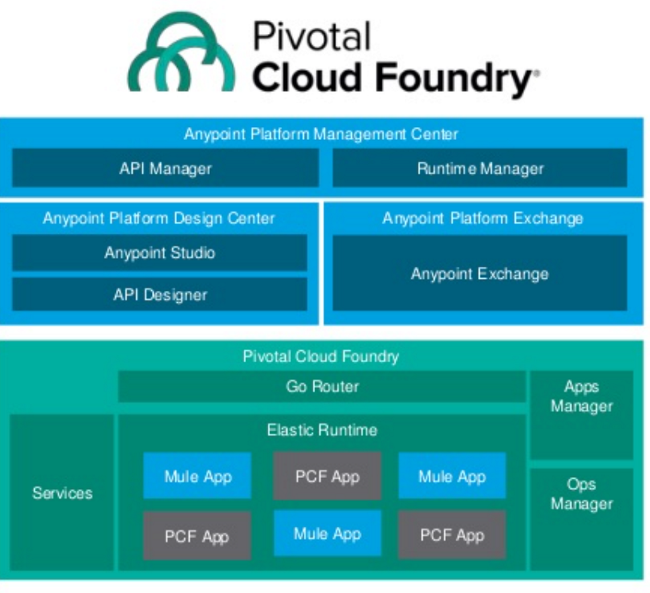
* Pivotal cloud foundary hosting a control plane and Runtime plane
* Pivotal cloud foundary can be hosted on customer Data center or AWS or GCP or on any cloud providers
* If we have an infrastuctutre On top of ‘Infrastructure’ ,PCF provides own container engine which is like Garden or Docker or Kubernetes
* On top of container engine ,we deploy our containerized applications.
* PCF determines mule applications with the help of ‘Build Pack’
* What usaually happens is when ever you create a code ,the code has some thing called as build pack
* Job of ‘Build Pack’ is to determine what kind of application it is , mulesoft has created its PCF build packs , using those build packs PCF determines those are mule applications and they would require a mule runtime and accordingly it will help to deploy the application.
* Like PCE ,Control Plane and Runtime Plane are hosted in customer’s data center
* Unlike manual runtime installation in PCE, runtimes are provisioned automatically using PCF technology .
* Unlike PCE, every application will get separate runtime in PCF

**Buildpacks:**

* Buildpacks provide framework and runtime support for apps.
* Buildpacks typically examine your apps to determine what dependencies to download and how to configure the apps to communicate with bound services.

When you push an app, Cloud Foundry automatically detects an appropriate buildpack for it. This buildpack is used to compile or prepare your app for launch.

**Integration Architecture:**



 Anypoint Platform for PCF uses the following services:

* **Runtime Manager Service**: exposes PCF as a deployment target within the Anypoint Runtime Manager and lets you deploy Mule applications to PCF, alongside other deployment targets.
* **API Gateway Service**: enables you to create and deploy a Mule API proxy that intercepts traffic to a non-Mule application running on PCF.
  + MuleSoft interface to apply policies over the API and
  + the PCF interface to view usage and policy metrics.
* **Anypoint Metering Service**: ---provides an API
* that enables you to view usage metrics for Mule applications deployed to PCF.
* You can view usage metrics for
  + production, pre-production, staging, and development environments.
  1. What are the differences between RTF and PCF?

Customer manages both control plane and runtime plane

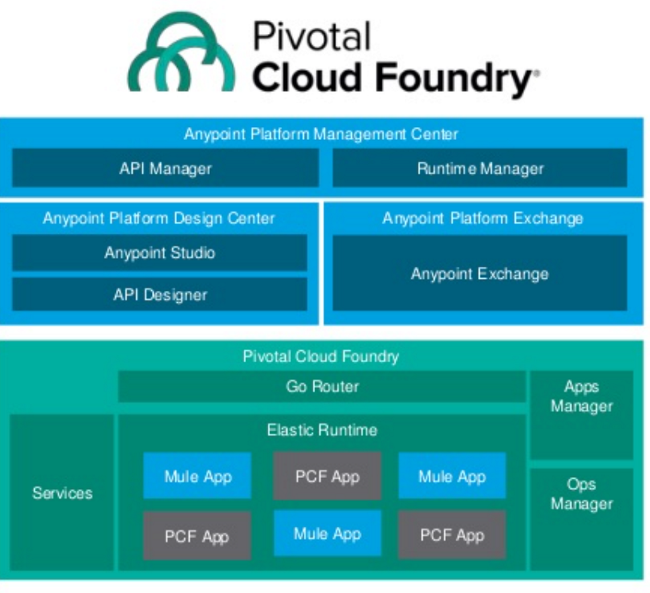
* 1. What are the differences between PCE and PCF?

Instead of customer manually installing mule runtime plane ,the way runtimes are provisioned is automatic using PCF build packs technology .It uses PCF droplets , so for each application PCF droplet gets created which will have a mule runtime .Applications will be deployed on to PCF droplets.

Will get application isolation here as well

* 1. **What control plane capabilities are available for PCF?**

**Integration Architecture:**



 Anypoint Platform for PCF uses the following services:

**Runtime Manager Service**: exposes PCF as a deployment target within the Anypoint Runtime Manager and lets you deploy Mule applications to PCF, alongside other deployment targets.

**API Gateway Service**: enables you to create and deploy a Mule API proxy that intercepts traffic to a non-Mule application running on PCF.

* + MuleSoft interface to apply policies over the API and
  + the PCF interface to view usage and policy metrics.

**Anypoint Metering Service**: ---provides an API

* + that enables you to view usage metrics for Mule applications deployed to PCF.
  + You can view usage metrics for
  + production, pre-production, staging, and development environments.
  1. What are droplets in PCF and how is it used?

Job of ‘Build Pack’ is to determine what kind of application it is , mulesoft has created its PCF build packs , using those build packs PCF determines those are mule applications and they would require a mule runtime and accordingly it will help to deploy the application.

* 1. How do you install the runtime plane in PCF?

Runtime plane gets generated automatically using PCF build packs technology

* 1. How do you install the control plane in PCF?

Install the Any point platform Service Brokers for PCF tile.

* 1. Does PCF support application isolation and why?

Yes , It supports application isolation , because for each application PCF droplet gets created which will have a mule runtime .Applications will be deployed on to PCF droplets.

* 1. How do you achieve vertical and horizontal scalability in PCF?
  2. How do you achieve zero-downtime in PCF?

Customer need to configure

* 1. How do you balance the load in PCF?
* Load Balancing: Customer need to configure

**Runtime Behaviour/ Thread Pool**

1. What is a thread?

A thread is a path of execution within a process. A process can contain multiple threads

1. What is a thread pool?

Server Programs such as database and web servers repeatedly execute requests from multiple clients and these are oriented around processing a large number of short tasks. An approach for building a server application would be to create a new thread each time a request arrives and service this new request in the newly created thread. While this approach seems simple to implement, it has significant disadvantages. A server that creates a new thread for every request would spend more time and consume more system resources in creating and destroying threads than processing actual requests.

**A thread pool reuses previously created threads to execute current tasks and offers a solution to the problem of thread cycle overhead and resource thrashing.** Since the thread is already existing when the request arrives, the delay introduced by thread creation is eliminated, making the application more responsive.

1. What is staged event-driven architecture (SEDA)?

<https://www.slideshare.net/anir37/using-seda-in-mule#:~:text=1.,of%20stages%20connected%20by%20queues>.

Mule 3 uses this

1. //What is reactive programming?

Mule 4 uses this.

1. //How reactive programming is used in mule?
2. //What are the features of reactive programming?
3. What is backpressure?

·       Responsive: The system responds in a timely manner if at all possible. Responsive systems focus on providing rapid and consistent response times.·       Resilient: The system stays responsive in the face of failure. This applies not only to highly-available, mission critical systems. Any system that is not resilient will be unresponsive after a failure. Resilient is achieved by containment, isolation and replication.·       Elastic: The system stays responsive under varying workload and bottleneck. Reactive Systems can react to changes in the input rate by increasing or decreasing the resources allocated to service these inputs·       Message Driven: Reactive Systems rely on [asynchronous](https://www.reactivemanifesto.org/glossary#Asynchronous) [message-passing](https://www.reactivemanifesto.org/glossary#Message-Driven) to establish a boundary between components that ensures loose coupling, isolation and location transparency.

1. How backpressure is implemented in mule?
2. What is automatic back-pressure?

[Mule 4 applications](https://www.mulesoft.com/webinars/api/best-practices-testing-mule-applications) are automatically configured so that the event source receives a back pressure signal when all threads are currently executing and no free threads remain in a required thread pool. In practical terms this will trigger the HTTP Listener, for example, to respond with a 503 – “Server busy”, and the JMS Listener will not acknowledge receipt of a message. OutOfMemory errors are avoided as a result of this configuration

1. What is manual backpressure?

Mule developers can also configure each event processor to signal back pressure to the event source through the “maxConcurrency” attribute. This configuration affects the number of events that can pass through the event processor per second.

1. How to configure a manual backpressure?

Same as 10th.

Eg: in batch we can set up max concurrency

1. What is auto tuning in mule and how does it work?
2. What are the different processing types in Mule?

* CPU\_LITE

o   For tasks that take up to 10ms to execute.

o   No blocking IO operations should be executed here

o   Default size is 2 \* cores

o   Logger, HTTP Requester(why not IO?)

* CPU\_INTENSIVE
  + For tasks that take more than 10ms to execute (duration is not enforced, but misclassifying tasks has bad consequences).
  + Typically for transformations, encrypt/decrypt, heavy computation, etc.
  + Non-blocking IO operations should be executed here.
  + Default size is 2 \* cores.
  + Dataweave, Scripting

* BLOCKING\_IO
  + All blocking IO operations should happen here.
  + Significantly larger than the other pools, as most threads here are expected to be in a blocked state.
  + Default size comes from a formula that considers the available memory, the default size of the streaming buffers and other concepts.
  + Database, IO

1. What is CPU\_LITE processing type?
2. Examples of connectors/processors which use CPU\_LITE processing type?

Logger, Set payload, set variable

1. What is CPU\_INTENSIVE processing type?
2. Examples of connectors/processors which use CPU\_INTENSIVE processing type?

Transform Message, PGP Connector, JCE connector ()

1. What is the BLOCKING\_IO processing type?cryptions
2. Examples of connectors/processors which use BLOCKING\_IO processing type?

Database connector, sfdc connectors

1. //What is the Proactor pattern?

Proactor is a design pattern for asynchronous execution. It segregates all the task that  
will be executed in the flow into respective categories and then assigns required thread  
pools to each of those categories

1. What is Grizzly thread pool?

Http connector uses separate thread pool called selector pool for its processing. It uses

Java Grizzly libraries internally. These thread pools are called Grizzly thread pool.

//The minimum size of the shared Grizzly pool for the HTTP Listener is determined upon the deployment of the first app to the Mule runtime that uses an HTTP Listener. The size of the dedicated Grizzly for the HTTP Requester pool is determined upon deployment of each app that uses an HTTP Requester.

1. What is Grizzly Shared thread pool?

Thread pool used by Http Listener component is Grizzly shared thread pool, this has to be shared among all the applications running on the same mule runtime

1. What is Grizzly Dedicated thread pool?

Thread pool used by Http Request component is Grizzly dedicated thread pool, there will be dedicated thread pool for each application running in the same mule runtime

1. What is Java NIO and how is it used in mule?

Java NIO and NIO2 libraries provides helps so that threads do not block waiting for IO  
intensive operations.  
NIO Selector thread pool is a separate reserved custom thread pool which deals with  
blocking IO operations. In case there is a blocking IO call, the thread which was executing  
the process is released immediately back to its own pool so that it can perform other  
operations, and the blocking operation is delegated to these selector pools (These  
threads belong to the OS kernel and schedule management is taken care of by the  
underlying OS, more the number of base cores better will be the processing).

1. What is the selector thread and how is it used in mule?
2. What is thread switching?
3. Why is thread switching expensive?

Switching the CPU from one thread to another involves suspending the current thread,  
saving its state (e.g., registers), and then restoring the state of the thread being  
switched to. The thread switch actually completes at the moment a new program  
counter is loaded into PC; at that point, the CPU is no longer executing the thread  
switching code, it is executing code associated with the new thread.

1. How does mule try to avoid thread switching?

Due to optimizations regarding latency, thread switches are omitted when an IO or  
CPU\_INTENSIVE task is followed by a CPU\_LIGHT one. Reasoning behind this  
optimization is that executing said CPU\_LIGHT task is most likely cheaper than the  
thread switch.

1. What is the UBER thread pool?

Mule 4.3 contains one unique thread pool, called the UBER pool. This thread pool is managed by Mule and shared across all apps in the same Mule instance. At startup, Mule introspects the available resources (such as memory and CPU cores) in the system and tunes automatically for the environment where Mule is running. This algorithm was established through performance testing and found optimal values for most scenarios.

The single thread pool allows Mule to be efficient, requiring significantly fewer threads (and their inherent memory footprint) to handle a given workload when compared to Mule 3.

1. What is a pooling strategy?

Pooling strategy indicate the type of thread pool used in Mule runtime. In mule 4.3  
default pooling strategy is UBER in which there is a only one thread pool  
Another option for pooling strategy is DEDICATED thread pool in which there will be  
separate thread pool for each type of processing types  
We can configure pooling strategy in scheduler-conf.properties file

1. How does mule increase t
2. What is JVM tuning?

**JVM tuning** mainly involves optimizing the garbage collector for better collection performance so that applications running on VMs can have a larger throughput while using less memory and experiencing lower latency.

1. What is Heap memory?
2. What is metaspace memory?

Metaspace memory is the memory allocated to store metadata about the application the JVM is running. It contains class definitions, method definitions, and other information about the program. Metaspace size can be limited by using the MaxMetaspaceSize option inside wrapper.config file

Module 5

1. Mule project is a MuleSoft project which have all the dependencies defined in itself, it doesn’t             share its resources or use shared resources.

2. Mule supports the ability to define selected connectors as common resources and expose them to all applications deployed under a same domain. These resources are known as shared resources, to host these you must create a Mule Domain Project

3. Logging helps you to monitor and troubleshoot your application and server.  
If you are running you application in your local machine, then you can see the logs in your console but debugging in the production doesn’t seems to be same. So there we can use Logging

We can log message Synchronously as well as Asynchronously.

In sync logging, exec of flow stops until logger logs the message. Same thread is used for logging.

In async, separate thread is used for logging and mule uses this by def, it is mostly like fire and forget kind of method.

4.