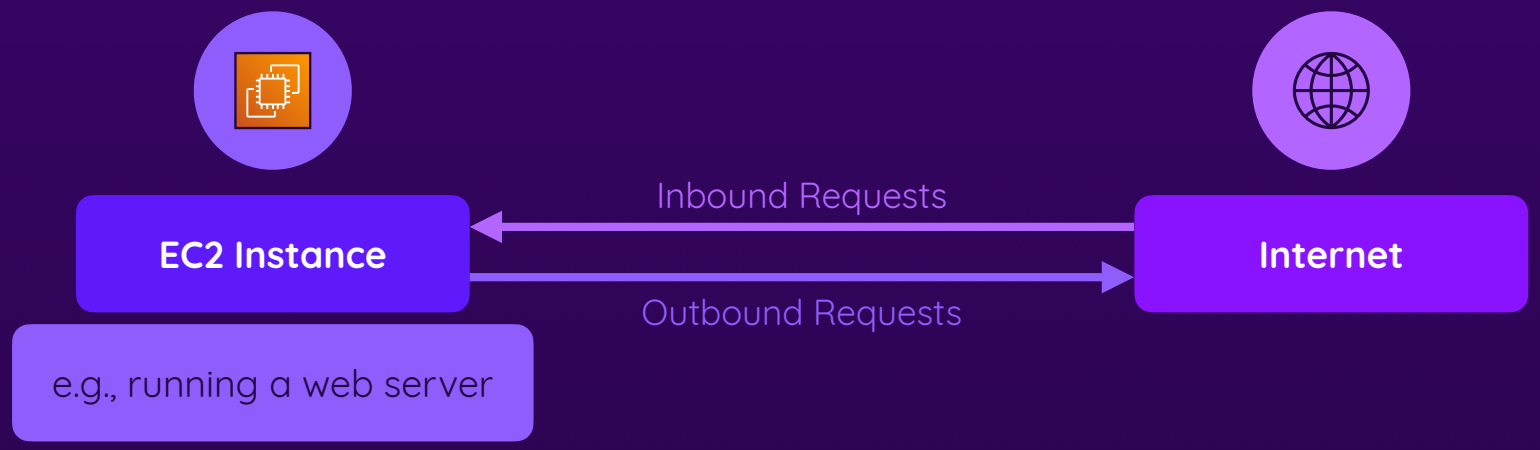
ACADEMIND

VPCs & Multiple EC2 Instances

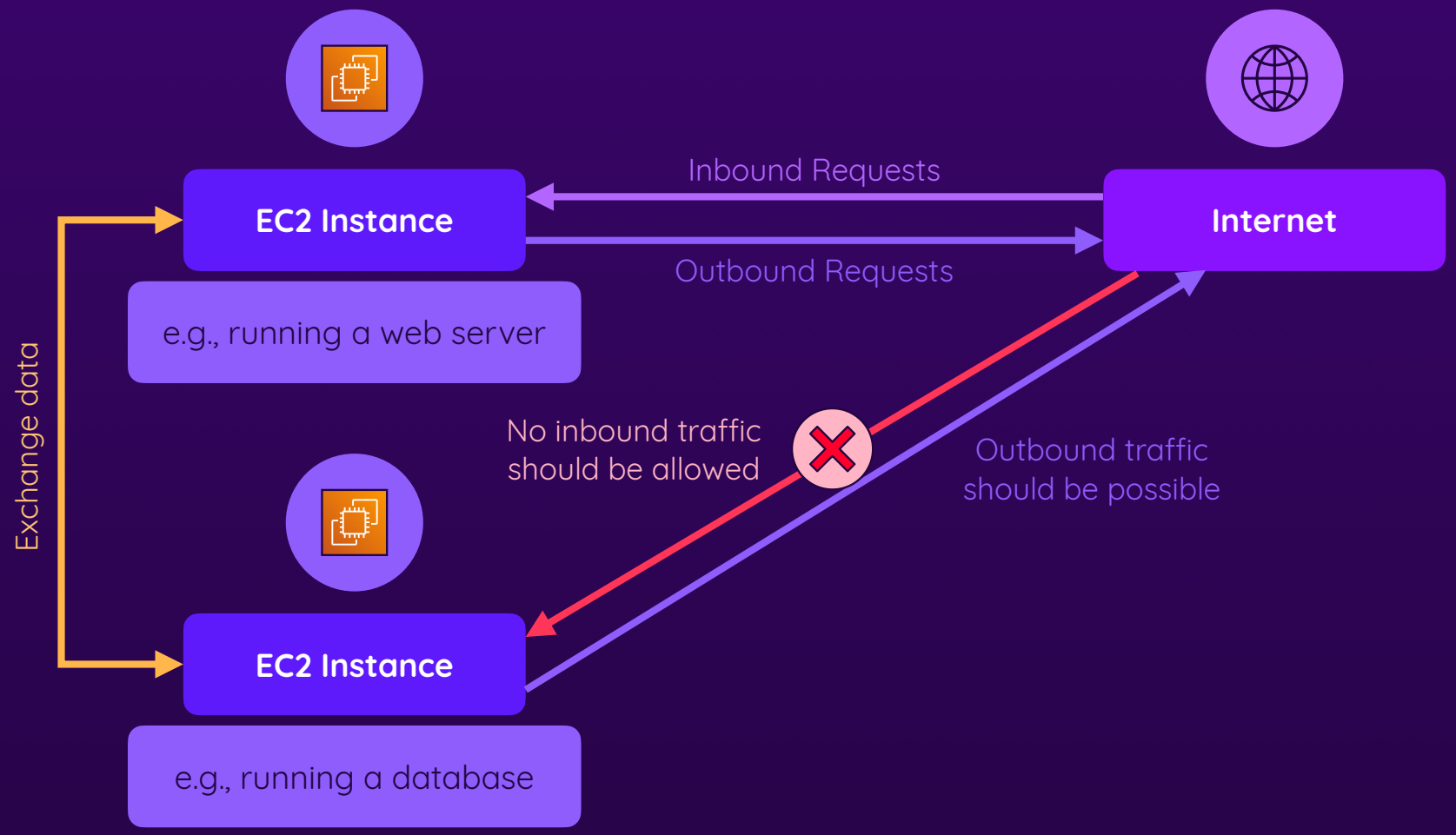
Managing your own network in the cloud

- Understanding VPCs
- Private vs Public Instances
- Managing Network Requests

A Simple Setup



A More Realistic Setup



Public vs Private IP Addresses

Every device in a network has an IP address

Private IPs for internal communication

Multiple servers can send requests between each other via private (internal) IP addresses

Public IPs for internet communication

!

Public IP addresses are a rare resource

Solution 1: IPv6

Solution 2: Limiting IPv4 address usage

Understanding IP (IPv4) Addresses

An IP address is a 32-bit number

172 . 31 . 0 . 0

4 x 8-bit

This is just a notation thing though (for human readability)

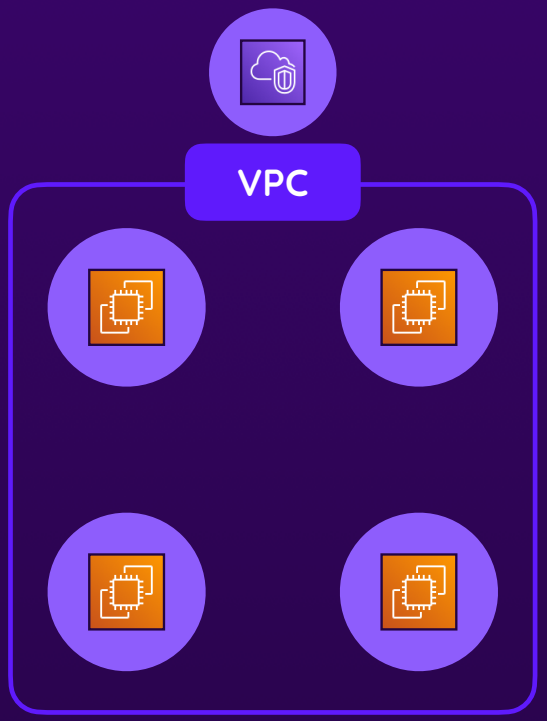
IPv4 Addresses Are A Rare Resource

Less than 4.3bn available IPv4 addresses

↓

Not enough for all the devices (with internet access) we have around the world

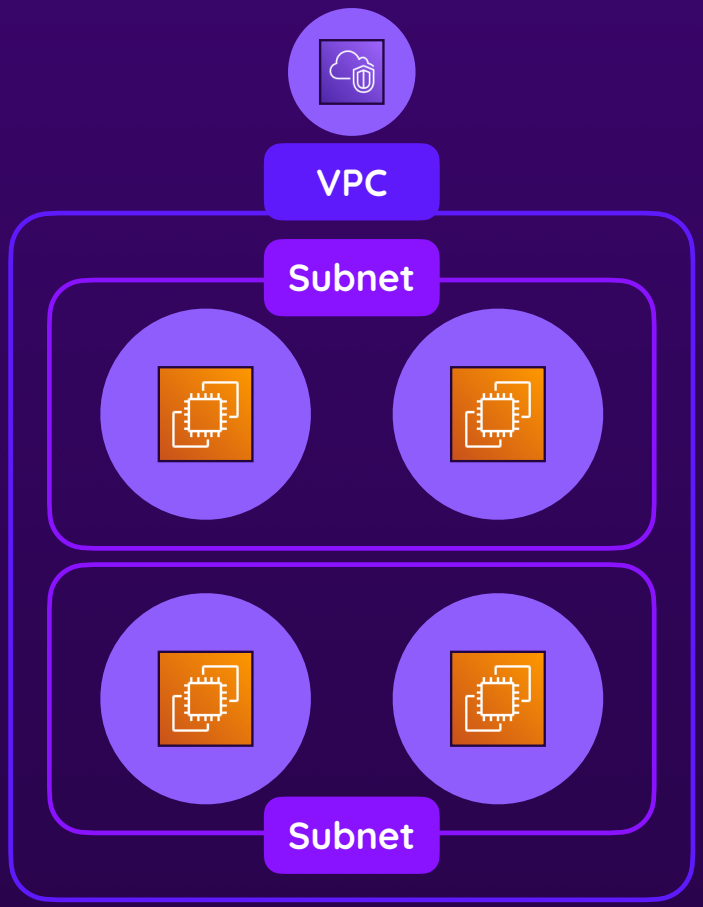
Introducing Virtual Private Clouds (VPCs)



Resources (e.g., EC2 Instances) are grouped & organized in a customer-managed network

IP address assignment & network traffic can be controlled via VPCs

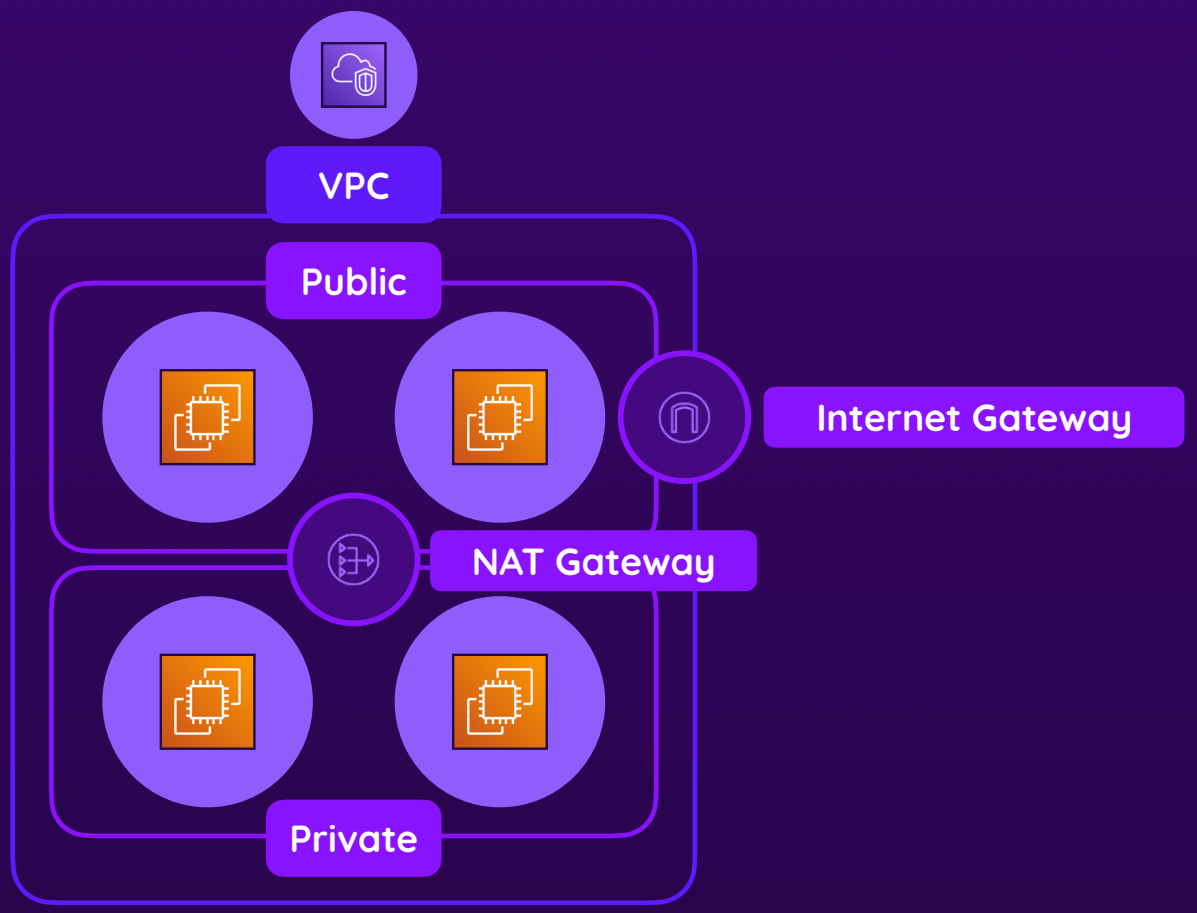
VPCs & Subnets



You actually control network request settings on subnet-level

This allows you to make subnets "**private**" (only internal requests) or "**public**" (internet requests are possible)

Public vs Private Subnets



What About Security Groups?

Security Group	Private Subnets
Firewalls, directly attached to EC2 instances	Not directly attached to instances
Technically, requests still reach the instances	Instead: Contain multiple EC2 instances
Security groups just allow which requests to allow or block	Technical isolation from the internet: Requests technically don't reach the instances

Public vs Private Subnets

Public Subnet

Associated with a route table that has an internet gateway route

Instances can communicate with each other **AND the internet**

Private Subnet

Associated with a route table that has **no internet gateway route**

Instances can communicate with each other only

To still allow for outgoing internet access, a **NAT gateway** can be used

Understanding CIDR IP Ranges

An IP address is a 32-bit number

172 . 31 . 0 . 0 / 16

Defines how many bits are fixed

4 x 8-bit

This is just a notation thing though (for human readability)

Understanding CIDR IP Ranges

172 . 31 . 0 . 0 / 16	172 . 31 . 0 . 0 / 24	0 . 0 . 0 . 0 / 0
16 bits are fixed	24 bits are fixed	0 bits are fixed
Range	**Range**	**Range**
172.31.0.0 ↕ 172.31.255.255	172.31.0.0 ↕ 172.31.0.255	Unlimited
65,536 available addresses	256 available addresses	All possible IP addresses

A <u>higher</u> /X number implies <u>less</u> available IP addresses in the range

Elastic IPs

Automatically assigned IPs (by subnet) will change when instances are stopped / restarted

You can't control which public IP address gets assigned to an instance

Elastic IPs are managed & assigned by you

Elastic IPs don't change and can be re-assigned

Always Use Elastic IPs?

Automatically assigned IPs (by subnet) will change when instances are stopped / restarted

You can't control which public IP address gets assigned to an instance

Elastic IPs are managed & assigned by you

Elastic IPs don't change and can be re-assigned

Elastic IPs should be used with care

Scarce resource: You can only have a few EIPs per region & account

Unused EIPs incur charges

There often are better alternatives

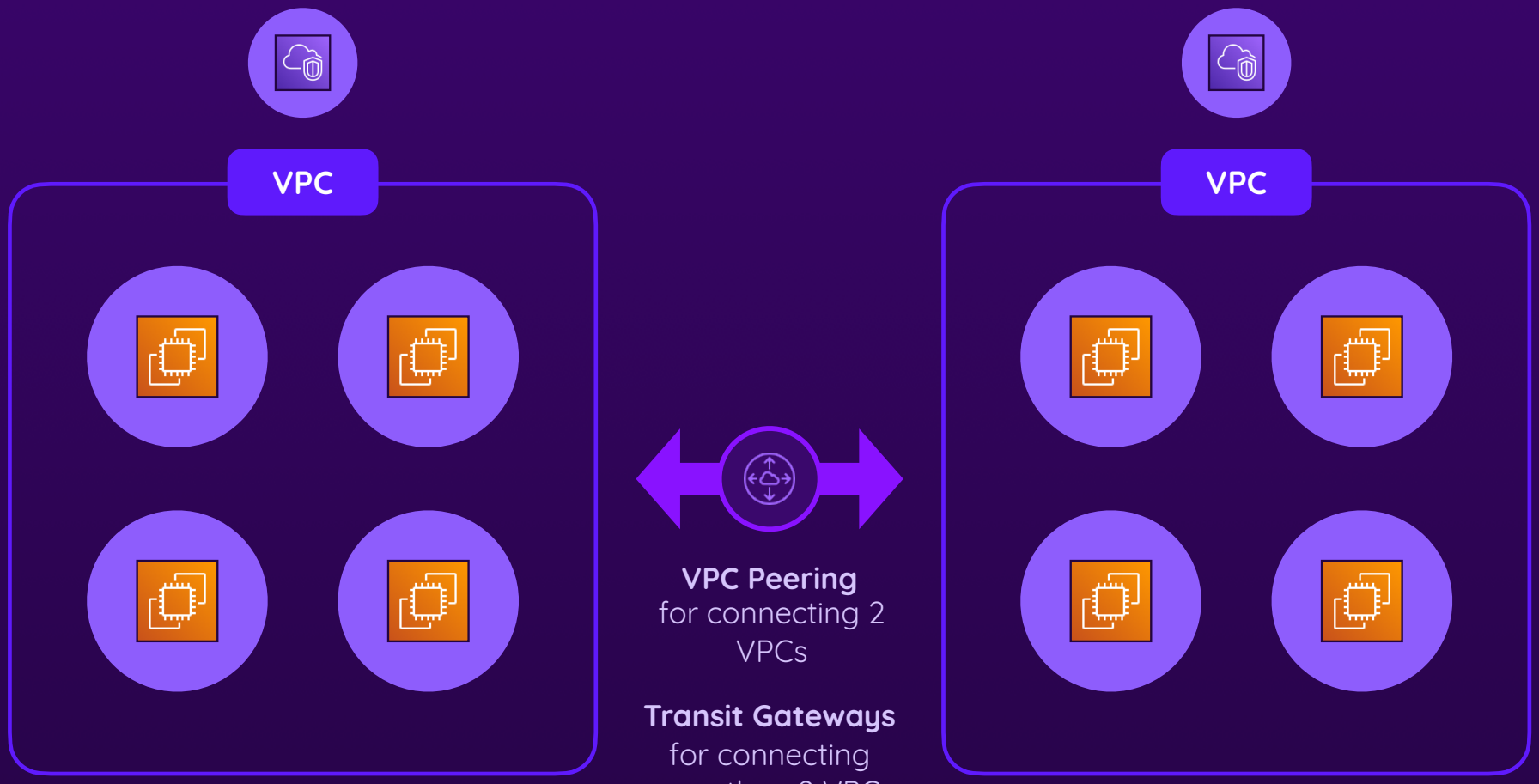
e.g., use DNS for exposing applications / websites to the world

e.g., use application integration services (like SQS) for connecting workloads

Security Groups & Network ACLs

Security Group Preferred	Network ACL (Access Control List)	Private / Public Subnets
Firewall for a single EC2 instance	Firewall for entire subnets	Defines technical connectivity
Checks incoming / outgoing requests and conditionally blocks or allows them	Checks incoming / outgoing requests and conditionally blocks or allows them	No internet access without internet gateway (incoming + outgoing) or NAT gateway (outgoing)
Stateful: Responses are always allowed (if the request passed)	Stateless: Requests & responses are decoupled	Does not control any requests or responses
Multiple instances can have different security groups	One NACL can be associated with multiple subnets	Multiple instances can be in the same subnet
Security groups can be re-used for multiple instances		

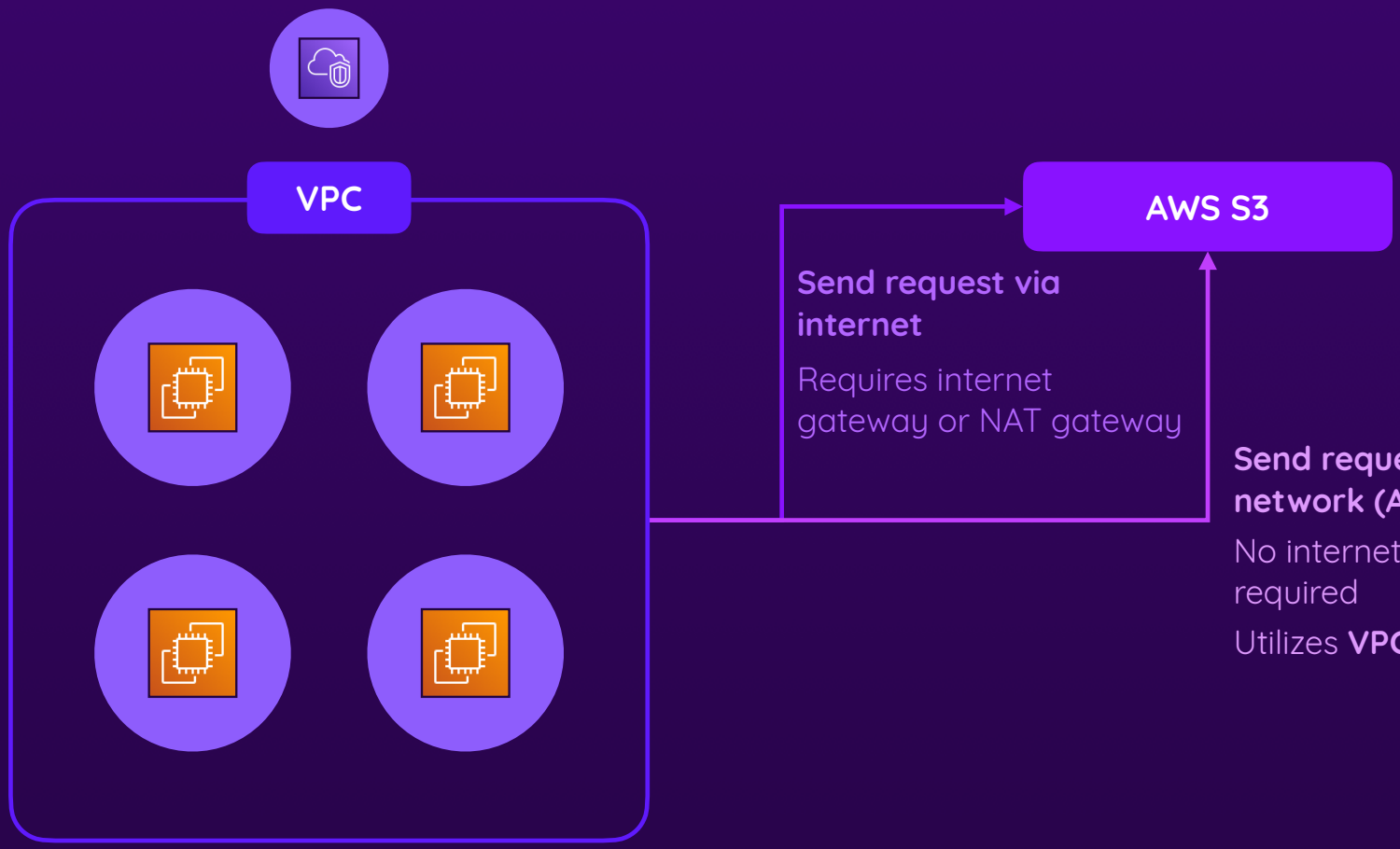
VPC Peering & Transit Gateways



VPC Peering
for connecting 2 VPCs

Transit Gateways
for connecting more than 2 VPCs

VPC Endpoints & AWS PrivateLink



Send request via internet

Requires internet gateway or NAT gateway

Send request via AWS network (AWS PrivateLink)

No internet connectivity required

Utilizes **VPC endpoints**

Summary

VPCs (Virtual Private Cloud)	Network Management	Request Management
Your own network in the cloud (for grouping EC2 instances)	Every VPC has an IP CIDR block assigned (range of IPs)	Internet gateways allow for two-way internet access
A VPC contains at least two subnets & one route table	Subnets get parts of the VPC CIDR block assigned	NAT gateways enable outgoing internet requests
Subnets can be "**public**" or "**private**"	EC2 instances receive auto-assigned public and private IPs	NACLs allow or deny requests on subnet-level
Route table settings control subnet "visibility"	Elastic IPs can be used for fixed IP addresses	Endpoints (PrivateLink) connect AWS services to VPCs
	VPC peering or transit gateways can connect VPCs	