Ec2(Elastic Compute Cloud)

Instance Types Or Families

1. Micro Instances
2. General Purpose Instances
3. Compute Optimized Instances
4. Accelerated Computing
5. HPC Optimized
6. Memory Optimized
7. Storage Optimized Instances

https://aws.amazon.com/ec2/instance-types/

Ec2(Micro Instances)

- Low Cost Instant And Small Amount Of CPU
- Designed For Lower Throughput Applications And Web Sites(Less Demanding Applications And Websites)
- But These Applications And Websites Require Additional Compute Cycles Periodically.
- When It Requires Additional CPU Cycle That Time It Will Burst Its CPU Performance
- Micro Instances Provide 613 Mb Of Memory
- Support 32-bit And 64-bit Platforms On Both Linux And Windows

Ec2(Micro Instances)

- T1 Micro: (Default) 613 Mib Of Memory, Up To 2 Ecus (For Short Periodic Bursts), Ebs Storage Only
- Ec2 Compute Unit (ECU) – One Ec2 Compute Unit (ECU) Provides The Equivalent Cpu Capacity Of A 1.0-1.2 Ghz
- Micro Instance Pricing For On-demand Instances Starts At $0.02 Per Hour For Linux And $0.03 Per Hour For Windows

Ec2 Instances

Instance CPU Performance Types

1. fixed performance instances
2. Burstable Performance instances

Fixed Performance Instances

- it provides consistent CPU performance

Burstable Performance instances

- Provides a baseline CPU performance under normal workload.

- When The Workload Increases,It Will Burst(Increase The CPU Performance)

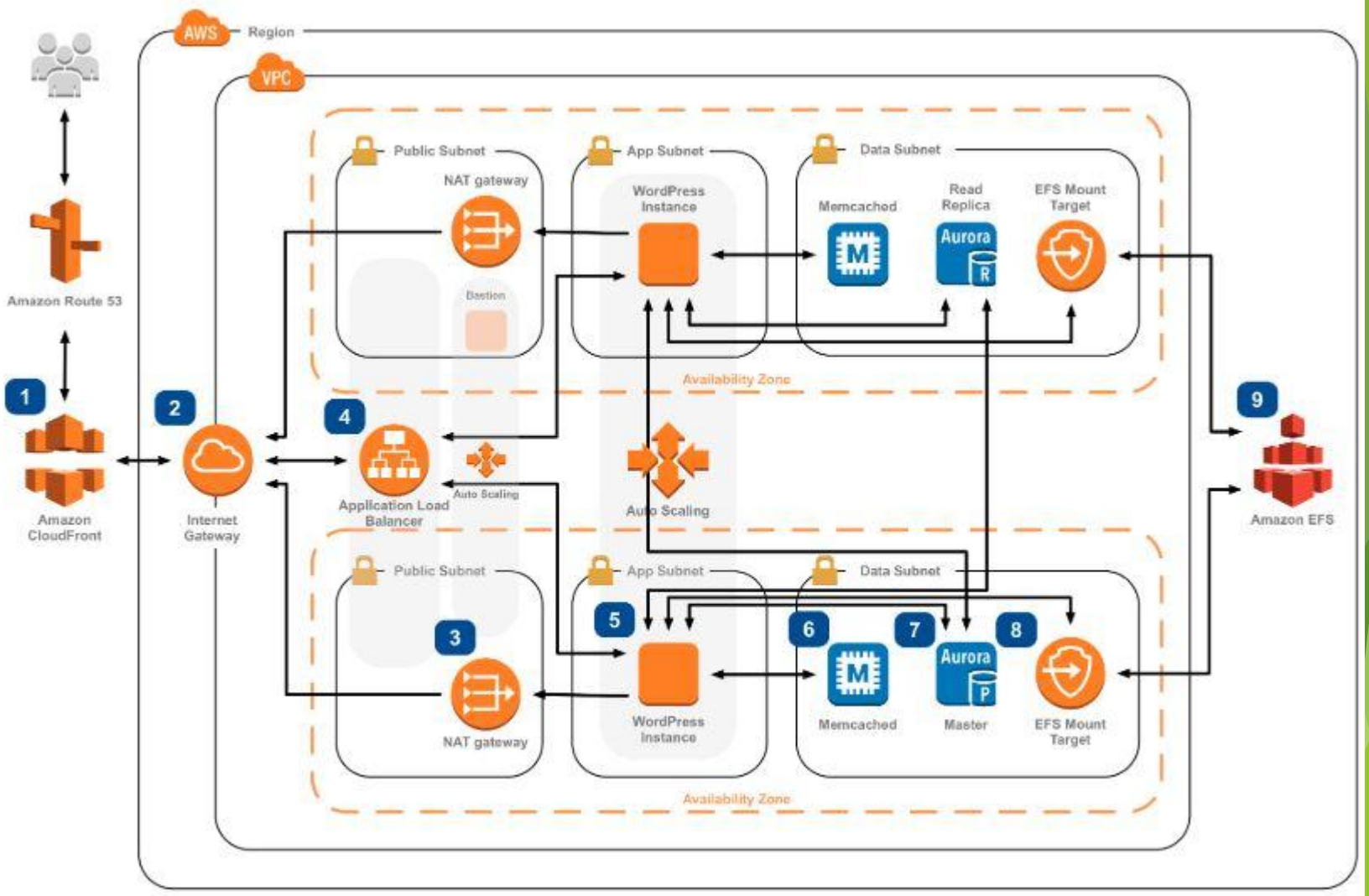
CPU Credit

- It Regulate The Cpu Burst Of An Instance.
- Suppose You Are Operating An Instance At 100% Of CPU Performance For 5 Minutes,Then You Are Spending 5 CPU Credits.
- If You Are Running An Instance At 50% Of CPU Performance For 5 Minutes Then CPU Credit Is 2.5

- When You Create An Instance ,You Will Get An Initial CPU Credit
- And In Every Hour ,It Will Automatically Get Certain Amount Of CPU Credit(Depends On Your Instance Type)
- If You Are Not Bursting The CPU,The CPU Credit Will Be Added To The CPU Credit Balance
- If The CPU Credit Turns Zero,Your Instance Will Work Only On Baseline

FAMILY	Type	Speciality	Use Case
F	F1	Field Programmable Gate Arrays	Genomics research, financial analytics, real-time video processing, big data search and analysis, and security.
I	I3	Non-Volatile Memory Express (NVMe) SSD-backed instance storage	NoSQL databases (e.g. Cassandra, MongoDB, Redis), in-memory databases (e.g. Aerospike), scale-out transactional databases, data warehousing, Elasticsearch, analytics workloads.
G	G3	optimized for graphics-intensive applications	3D visualizations, graphics-intensive remote workstation, 3D rendering, application streaming, video encoding, and other server-side graphics workloads.
H	H1	H1 instances feature up to 16 TB of HDD-based local storage	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS, network file systems, log or data processing applications such as Apache Kafka, and big data workload clusters.
T	T3	burstable general-purpose instance type	Micro-services, low-latency interactive applications, small and medium databases, virtual desktops, development environments, code repositories, and business-critical applications
D	D2	instances feature 48 TB of HDD-based local storage	Massively Parallel Processing (MPP) data warehousing, MapReduce and Hadoop distributed computing, distributed file systems, network file systems, log or data-processing applications.
R	R5	memory-intensive applications	High performance databases, data mining & analysis, in-memory databases, distributed web scale in-memory caches, applications performing real-time processing of unstructured big data, Hadoop/Spark clusters, and other enterprise applications.
M	M5	General Purpose Instances	Small and mid-size databases, data processing tasks that require additional memory, caching fleets, and for running backend servers for SAP, Microsoft SharePoint, cluster computing, and other enterprise applications
C	C5	compute-intensive workloads	High performance web servers, scientific modelling, batch processing, distributed analytics, high-performance computing (HPC), machine/deep learning inference, ad serving, highly scalable multiplayer gaming, and video encoding.
P	P3	general purpose GPU instances	Machine learning, high performance databases, computational fluid dynamics, computational finance, seismic analysis, molecular modeling, genomics, rendering, and other server-side GPU compute workloads.

Running A Web App On AWS



AWS Reference Architectures