**Azure Event Hubs**

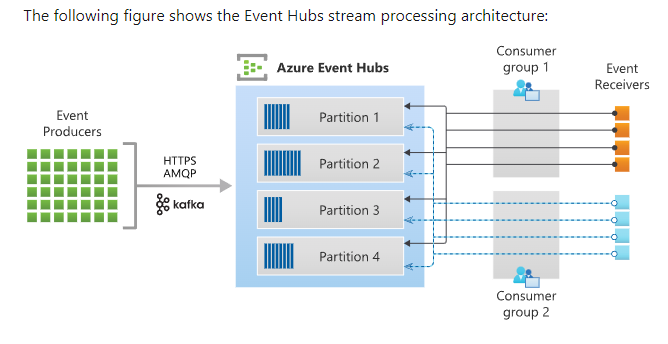
**a fully managed Platform-as-a-Service (PaaS)**

Azure Event Hubs is a big data streaming platform and event ingestion service. It can receive and process millions of events per second. Data sent to an event hub can be transformed and stored by using any real-time analytics provider or batching/storage adapters.

The following scenarios are some of the scenarios where you can use Event Hubs:

* Anomaly detection (fraud/outliers)
* Application logging
* Analytics pipelines, such as clickstreams
* Live dashboarding
* Archiving data
* Transaction processing
* User telemetry processing
* Device telemetry streaming

Event Hubs represents the "front door" for an event pipeline, often called an event ingestor in solution architectures. An event ingestor is a component or service that sits between event publishers and event consumers to decouple the production of an event stream from the consumption of those events. Event Hubs provides a unified streaming platform with time retention buffer, decoupling event producers from event consumers.



| **Limit** | **Basic** | **Standard** | **Premium** | **Dedicated** |
| --- | --- | --- | --- | --- |
| Maximum size of Event Hubs publication | 256 KB | 1 MB | 1 MB | 1 MB |
| Number of consumer groups per event hub | 1 | 20 | 100 | 1000 No limit per CU |
| Number of brokered connections per namespace | 100 | 5,000 | 10000 per PU  For example, if the namespace is assigned 3 PUs, the limit is 30000. | 100, 000 per CU |
| Maximum retention period of event data | 1 day | 7 days | 90 days | 90 days |
| Maximum TUs or PUs or CUs | 40 TUs | 40 TUs | 16 PUs | 20 CUs |
| Number of partitions per event hub | 32 | 32 | 100 per event hub, but there is a limit of 200 per PU at the namespace level.  For example, if a namespace is assigned 2 PUs, the limit for total number of partitions in all event hubs in the namespace is 2 \* 200 = 400. | 1024 per event hub 2000 per CU |
| Number of namespaces per subscription | 1000 | 1000 | 1000 | 1000 (50 per CU) |
| Number of event hubs per namespace | 10 | 10 | 100 per PU | 1000 |
| Capture | N/A | Pay per hour | Included | Included |
| Size of the schema registry (namespace) in mega bytes | N/A | 25 | 100 | 1024 |
| Number of schema groups in a schema registry or namespace | N/A | 1 - excluding the default group | 100 1 MB per schema | 1000 1 MB per schema |
| Number of schema versions across all schema groups | N/A | 25 | 1000 | 10000 |
| Throughput per unit | Ingress - 1 MB/s or 1000 events per second Egress – 2 MB/s or 4096 events per second | Ingress - 1 MB/s or 1000 events per second Egress – 2 MB/s or 4096 events per second | No limits per PU \* | No limits per CU \* |

**There are two factors which influence scaling with Event Hubs.**

* Throughput units (standard tier) or processing units (premium tier)
* Partitions

## **Throughput units**

The throughput capacity of Event Hubs is controlled by throughput units. Throughput units are pre-purchased units of capacity. A single throughput lets you:

* Ingress: Up to 1 MB per second or 1000 events per second (whichever comes first).
* Egress: Up to 2 MB per second or 4096 events per second.

Throughput units are pre-purchased and are billed per hour. Once purchased, throughput units are billed for a minimum of one hour. Up to 40 throughput units can be purchased for an Event Hubs namespace and are shared across all event hubs in that namespace.

The **Auto-inflate** feature of Event Hubs automatically scales up by increasing the number of throughput units, to meet usage needs. Increasing throughput units prevents throttling scenarios, in which:

* Data ingress rates exceed set throughput units.
* Data egress request rates exceed set throughput units.

The Event Hubs service increases the throughput when load increases beyond the minimum threshold, without any requests failing with ServerBusy errors.

## **Processing units**

Event Hubs Premium provides superior performance and better isolation with in a managed multitenant PaaS environment. The resources in a Premium tier are isolated at the CPU and memory level so that each tenant workload runs in isolation. This resource container is called a Processing Unit(PU). You can purchase 1, 2, 4, 8 or 16 processing Units for each Event Hubs Premium namespace.

## **Partitions**

Event Hubs organizes sequences of events sent to an event hub into one or more partitions. As newer events arrive, they're added to the end of this sequence.

A partition can be thought of as a "commit log". Partitions hold event data that contains body of the event, a user-defined property bag describing the event, metadata such as its offset in the partition, its number in the stream sequence, and service-side timestamp at which it was accepted.

**Event Hubs is designed to help with processing of large volumes of events, and partitioning helps with that in two ways:**

* Even though Event Hubs is a PaaS service, there's a physical reality underneath, and maintaining a log that preserves the order of events requires that these events are being kept together in the underlying storage and its replicas and that results in a throughput ceiling for such a log. Partitioning allows for multiple parallel logs to be used for the same event hub and therefore multiplying the available raw IO throughput capacity.
* Your own applications must be able to keep up with processing the volume of events that are being sent into an event hub. It may be complex and requires substantial, scaled-out, parallel processing capacity. The capacity of a single process to handle events is limited, so you need several processes. Partitions are how your solution feeds those processes and yet ensures that each event has a clear processing owner.

### **What is an Event Hubs namespace?**

A namespace is a scoping container for Event Hub/Kafka Topics. It gives you a unique FQDN. A namespace serves as an application container that can house multiple Event Hub/Kafka Topics.

### **When do I create a new namespace vs. use an existing namespace?**

Capacity allocations throughput units (TUs) or processing units (PUs)) are billed at the namespace level. A namespace is also associated with a region.

You may want to create a new namespace instead of using an existing one in one of the following scenarios:

* You need an Event Hub associated with a new region.
* You need an Event Hub associated with a different subscription.
* You need an Event Hub with a distinct capacity allocation (that is, the capacity need for the namespace with the added event hub would exceed the 40 TU threshold and you don't want to go for the dedicated cluster)

### **What ports do I need to open on the firewall?**

You can use the following protocols with Azure Event Hubs to send and receive events:

* Advanced Message Queuing Protocol 1.0 (AMQP)
* Hypertext Transfer Protocol 1.1 with TLS (HTTPS)
* Apache Kafka

See the following table for the outbound ports you need to open to use these protocols to communicate with Azure Event Hubs.

| **WHAT PORTS DO I NEED TO OPEN ON THE FIREWALL?** | |
| --- | --- |
| **Protocol** | **Ports** |
| AMQP | 5671 and 5672 |
| HTTPS | 443 |
| Kafka | 9093 |

The HTTPS port is required for outbound communication also when AMQP is used over port 5671, because several management operations performed by the client SDKs and the acquisition of tokens from Azure Active Directory (when used) run over HTTPS.

The official Azure SDKs generally use the AMQP protocol for sending and receiving events from Event Hubs. The AMQP-over-WebSockets protocol option runs over port TCP 443 just like the HTTP API, but is otherwise functionally identical with plain AMQP. This option has higher initial connection latency because of extra handshake round trips and slightly more overhead as tradeoff for sharing the HTTPS port. If this mode is selected, TCP port 443 is sufficient for communication.

### **What is the message/event size for Event Hubs?**

The maximum message size allowed for Event Hubs is 1 MB.

### **Do throughput units apply to all event hubs in a namespace?**

Yes, throughput units (TUs) apply to all event hubs in an Event Hubs namespace. It means that you purchase TUs at the namespace level and are shared among the event hubs under that namespace. Each TU entitles the namespace to the following capabilities:

* Up to 1 MB per second of ingress events (events sent into an event hub), but no more than 1000 ingress events, management operations, or control API calls per second.
* Up to 2 MB per second of egress events (events consumed from an event hub), but no more than 4096 egress events.
* Up to 84 GB of event storage (enough for the default 24-hour retention period).

### **How are throughput units billed?**

Throughput units (TUs) are billed on an hourly basis. The billing is based on the maximum number of units that was selected during the given hour.

### **How many partitions do I need?**

The number of partitions is specified at creation and must be between 1 and 32. The partition count isn't changeable in all tiers except the [dedicated tier](https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-dedicated-overview), so you should consider long-term scale when setting partition count. Partitions are a data organization mechanism that relates to the downstream parallelism required in consuming applications. The number of partitions in an event hub directly relates to the number of concurrent readers you expect to have.

**Pricing:**

https://azure.microsoft.com/en-in/pricing/details/event-hubs/