Azure EventHub

August 2021 Camilla Gaardsted

SuperUsers A/S

Azure - Messaging services

Azure Storage Queue

- Simple
- Process one at a time
- Messages can be updated
- Messages are deleted after consumption

Azure Event Hub

- Event stream (millions of messages per second)
- Bulk/batch processing using a cursor for a partition
- Retention period determines deletion of events

Azure IoT Hub

Azure IoT Hub

- Built on the foundation of Azure Event Hub
  - Event stream (millions of messages per second)
  - Bulk/batch processing using a cursor for a partition
  - Retention period determines deletion of events
- Connect and manage billions of IoT devices
- Two way communication

Azure Event Hub - *event ingestor*

- 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.
- Maximum message size allowed for Event Hubs is 1 MB
- Anvendelse:
  - Anomaly detection (fraud/outliers)
  - Application logging
  - Analytics pipelines, such as clickstreams
  - Live dashboarding
  - Archiving data
  - Transaction processing
  - User telemetry processing
  - Device telemetry streaming

Event Hub

Event Hub

- Basic
- Standard (max 40 Tus)

Event Hubs cluster

Dedicated pricing tier
Capacity Units (CUs)

Event Hub – throughput units (TU)

- Throughput in Event Hubs defines the amount of data in mega bytes or the number (in thousands) of 1-KB events that ingress and egress through Event Hubs.

Tus are shared among all event hubs in a namespace

  - 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).

Tus are billed hourly as the maximum number of selected units during a given hour

Event Hub 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.
- Azure region
- Tier
  - Basic (1 consumer group/100 brokered connections
  - Standard (20 consumer groups/ 1000 br connections)
- Throughput units (Tus): 1 til 20 (til 40 TU via support request)
- Auto-inflate (scale-up only) for throughput units with max limit
- Availability zones in one single region (no extra cost)

Event Hub

- Capture your data in near-real time in an Azure Blob storage or Azure Data Lake Storage for long-term retention or micro-batch processing.
- Capture er ikke tilgængelig I basic
- Time to live er 1 dag i basic
- Retention period 1 til 7 dag for standard. Max 90 dage for dedicated.

Event Hubs – key components

- Event Hubs contains the following key components:
- **Event producers**: Any entity that sends data to an event hub. Event publishers can publish events using HTTPS or AMQP 1.0 or Apache Kafka (1.0 and above)
- **Partitions**: Each consumer only reads a specific subset, or partition, of the message stream.
- **Consumer groups**: A view (state, position, or offset) of an entire event hub. Consumer groups enable consuming applications to each have a separate view of the event stream. They read the stream independently at their own pace and with their own offsets.
- **Throughput units**: Pre-purchased units of capacity that control the throughput capacity of Event Hubs.
- **Event receivers**: Any entity that reads event data from an event hub. All Event Hubs consumers connect via the AMQP 1.0 session. The Event Hubs service delivers events through a session as they become available. All Kafka consumers connect via the Kafka protocol 1.0 and later.

Event Hub - protocol

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

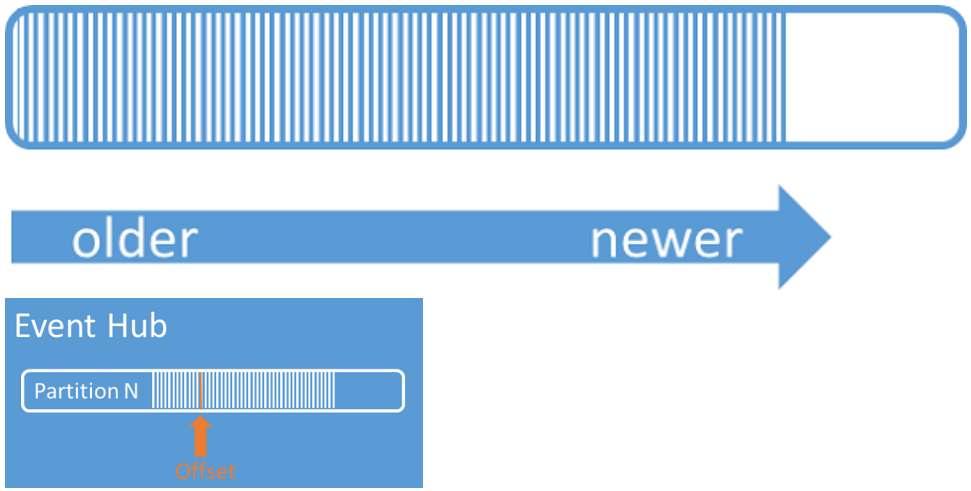
- The choice to use AMQP or HTTPS is specific to the usage scenario. AMQP requires the establishment of a persistent bidirectional socket in addition to transport level security (TLS) or SSL/TLS.
- AMQP has higher network costs when initializing the session, however HTTPS requires additional TLS overhead for every request. AMQP has higher performance for frequent publishers.

Event Hub - partitions

- The number of partitions in an event hub directly relates to the number of concurrent readers you expect to have.
- Partitions are a data organization mechanism that relates to the downstream parallelism required in consuming applications.
- The partition count isn't changeable, so you should consider long-term scale when setting partition count.

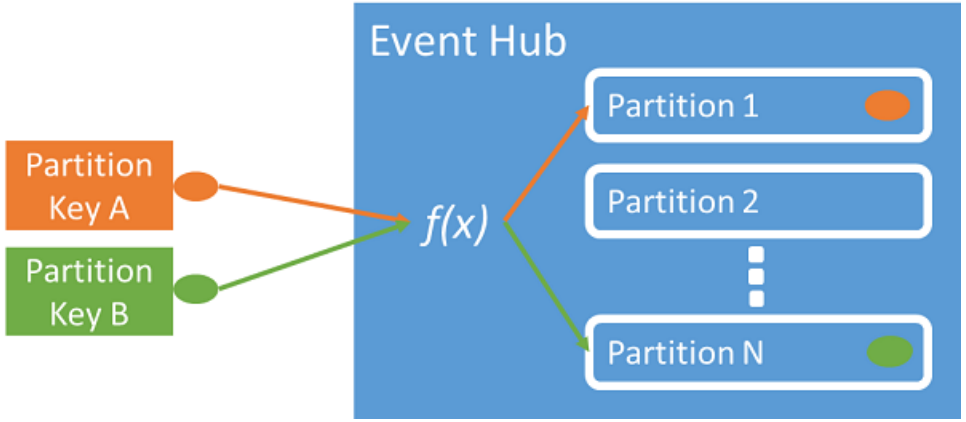
Event Hub - partitions

- Event Hubs retains data for a configured retention time that applies across all partitions in the event hub. Events expire on a time basis; you cannot explicitly delete them. Because partitions are independent and contain their own sequence of data, they often grow at different rates.

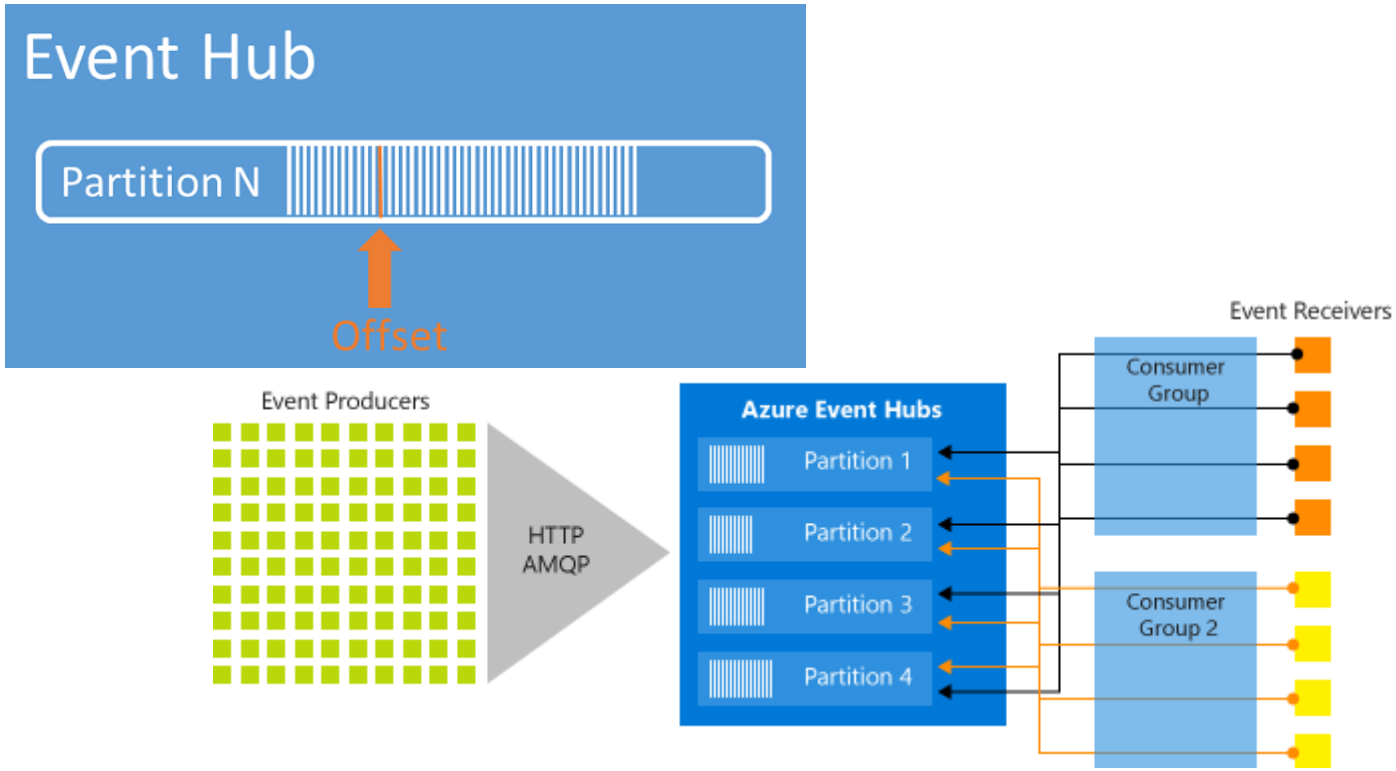


Event Hub – partition keys

- Event Hubs ensures that all events sharing a partition key value are delivered in order, and to the same partition. If partition keys are used with publisher policies, then the identity of the publisher and the value of the partition key must match. Otherwise, an error occurs.



Event Hub


Event Hub
Partition N
Offset
Event Receivers
Event Producers
HTTP
AMQP
Azure Event Hubs
Partition 1
Partition 2
Partition 3
Partition 4
Consumer Group
Consumer Group 2

Apache Hadoop

- A collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation.
- Provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

Azure HDInsights

Azure HDInsight is a managed Apache Hadoop service that lets you run Apache Spark, Apache Hive, Apache Kafka, Apache HBase, and more in the cloud.