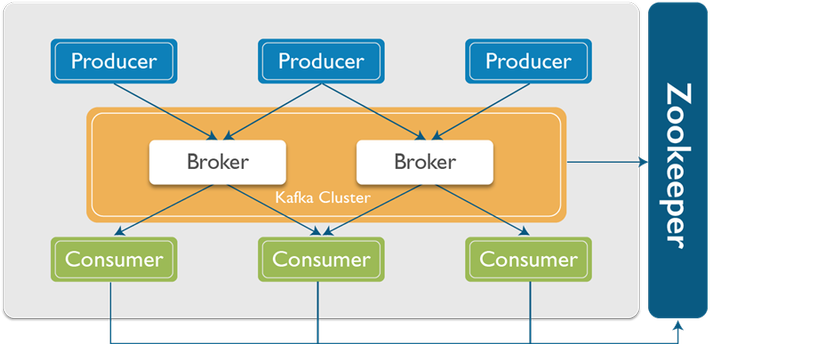
**BÁO CÁO MESSAGE QUEUE**

1. **KAFKA**

**Kafka** là gì? – Đó là hệ thống message pub/sub phân tán (distributed messaging system). Bên pulbic dữ liệu được gọi là producer, bên subscribe nhận dữ liệu theo topic được gọi là consumer. Kafka có khả năng truyền một lượng lớn message theo thời gian thực, trong trường hợp bên nhận chưa nhận message vẫn được lưu trữ sao lưu trên một hàng đợi và cả trên ổ đĩa bảo đảm an toàn. Đồng thời nó cũng được replicate trong cluster giúp phòng tránh mất dữ liệu.



* **PRODUCER**: Kafka lưu, phân loại message theo topic, sử dụng producer để publish message vào các topic. Dữ liệu được gửi đển partition của topic lưu trữ trên Broker.
* **CONSUMER**: Kafka sử dụng consumer để subscribe vào topic, các consumer được định danh bằng các group name. Nhiều consumer có thể cùng đọc một topic.
* **TOPIC**: Dữ liệu truyền trong Kafka theo topic, khi cần truyền dữ liệu cho các ứng dụng khác nhau thì sẽ tạo ra cá topic khác nhau.
* **PARTITION**: Đây là nơi dữ liệu cho một topic được lưu trữ. Một topic có thể có một hay nhiều partition. Trên mỗi partition thì dữ liệu lưu trữ cố định và được gán cho một ID gọi là offset. Trong một Kafka cluster thì một partition có thể replicate (sao chép) ra nhiều bản. Trong đó có một bản leader chịu trách nhiệm đọc ghi dữ liệu và các bản còn lại gọi là follower. Khi bản leader bị lỗi thì sẽ có một bản follower lên làm leader thay thế. Nếu muốn dùng nhiều consumer đọc song song dữ liệu của một topic thì topic đó cần phải có nhiều partition.
* **BROKER**: Kafka cluster là một set các server, mỗi một set này được gọi là 1 broker
* **ZOOKEEPER**: được dùng để quản lý và bố trí các broker.

Tại sao nên sử dụng Apache Kafka?

Kafka là dự án opensoure, đã được đóng gói hoàn chỉnh, khả năng chịu lỗi cao, hiệu năng rất tốt và dễ dàng mở rộng mà không cần dừng hệ thống.

### **Một vài use case cho kafka:**

* Sử dụng như một hệ thống message queue thay thế cho ActiveMQ hay [RabbitMQ](https://topdev.vn/blog/rabbitmq-la-gi/)
* Website Activity Monitoring: theo dõi hoạt động của website
* [Stream Processing](https://topdev.vn/blog/stream-processing-la-gi/): Kafka là một hệ thống rất thích hợp cho việc xử lý dòng dữ liệu trong thời gian thực. Khi dữ liệu của một topic được thêm mới ngay lập tức được ghi vào hệ thống và truyền đến cho bên nhận. Ngoài ra Kafka  còn là một hệ thống có đặc tính **duribility** dữ liệu có thể được lưu trữ an toàn cho đến khi bên nhận sẵn sàng nhận nó.
* Log Aggregation: tổng hợp log
* Metrics Collection: thu thập dữ liệu, tracking hành động người dùng như các thông số như page view, search action của user sẽ được publish vào một topic và sẽ được xử lý sau
* Event-Sourcing: Lưu lại **trạng thái** của hệ thống để có thể tái hiện trong trường hợp system bị down.

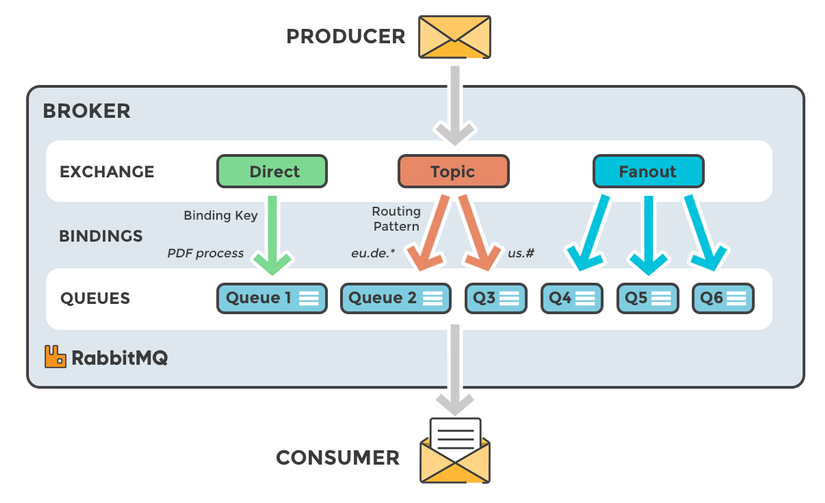
1. **RabbitMQ**

**Khái niệm:** RabbitMQ là một message-queuing software có thể được biết đến như là một người vận chuyển message trung gian hoặc một người quản lí các queue. Nói một cách đơn giản, nó là một phần mềm nơi các queue được định nghĩa, phục vụ cho ứng dụng với mục đích vận chuyển một hoặc nhiều message.

      Vậy, tại sao chúng ta phải cần đến RabbitMQ? Hãy tưởng tượng, hiện bạn đang có một web service, phải nhận rất rất nhiều request mỗi giây, mà lại phải đảm bảo rằng không có bất cứ một request nào bị mất. Và web service của bạn luôn luôn sẵn sàng tiếp nhận request mới thay vì locked bởi đang xử lí request trước đó. Vậy ý tưởng ở đây là đặt chúng vào một queue giữa web service và processing service. Lúc này sẽ đảm bảo rằng 2 process sẽ hoàn toàn tách rời nhau. Ngoài ra, queue sẽ lưu trữ những request, không bị thiếu sót request nào khi số lượng của chúng trở nên vô cùng lớn.

### **Những khái niệm cơ bản trong RabbitMQ**

* **Producer**: Ứng dụng gửi message.
* **Consumer**: Ứng dụng nhận message.
* **Queue**: Lưu trữ messages.
* **Message**: Thông tin truyền từ Producer đến Consumer qua RabbitMQ.
* **Connection**: Một kết nối TCP giữa ứng dụng và RabbitMQ broker.
* **Channel**: Một kết nối ảo trong một Connection. Việc publishing hoặc consuming từ một queue đều được thực hiện trên channel.
* **Exchange**: Là nơi nhận message được publish từ Producer và đẩy chúng vào queue dựa vào quy tắc của từng loại Exchange. Để nhận được message, queue phải được nằm trong ít nhất 1 Exchange.
* **Binding**: Đảm nhận nhiệm vụ liên kết giữa Exchange và Queue.
* **Routing key**: Một key mà Exchange dựa vào đó để quyết định cách để định tuyến message đến queue. Có thể hiểu nôm na, Routing key là địa chỉ dành cho message.
* **AMQP**: Giao thức Advance Message Queuing Protocol, là giao thức truyền message trong RabbitMQ.
* **User**: Để có thể truy cập vào RabbitMQ, chúng ta phải có username và password. Trong RabbitMQ, mỗi user được chỉ định với một quyền hạn nào đó. User có thể được phân quyền đặc biệt cho một Vhost nào đó.
* **Virtual host/Vhost**: Cung cấp những cách riêng biệt để các ứng dụng dùng chung một RabbitMQ instance. Những user khác nhau có thể có các quyền khác nhau đối với vhost khác nhau. Queue và Exchange có thể được tạo, vì vậy chúng chỉ tồn tại trong một vhost.



Hình 1. Sơ đồ vận chuyển message trong RabbitMQ

### 1.2. Các loại Exchange

#### a. Direct Exchange

      Direct Exchange vận chuyển message đến queue dựa vào routing key. Thường được sử dụng cho việc định tuyến tin nhắn unicast-đơn hướng (mặc dù nó có thể sử dụng cho định tuyến multicast-đa hướng). Các bước định tuyến message:

* Một queue được ràng buộc với một direct exchange bởi một routing key K.
* Khi có một message mới với routing key R đến direct exchange. Message sẽ được chuyển tới queue đó nếu R=K.

#### b. Default Exchange

      Mỗi một exchange đều được đặt một tên không trùng nhau, default exchange bản chất là một direct exchange nhưng không có tên (string rỗng). Nó có một thuộc tính đặc biệt làm cho nó rất hữu ích cho các ứng dụng đơn giản: mọi queue được tạo sẽ tự động được liên kết với nó bằng một routing key giống như tên queue.

      Ví dụ, nếu bạn tạo ra 1 queue với tên "hello-world", RabbitMQ broker sẽ tự động binding default exchange đến queue "hello-word" với routing key "hello-world".

#### b. Fanout Exchange

      Fanout exchange định tuyến message tới tất cả queue mà nó bao quanh, routing key bị bỏ qua. Giả sử, nếu nó N queue được bao quanh bởi một Fanout exchange, khi một message mới published, exchange sẽ vận chuyển message đó tới tất cả N queues. Fanout exchange được sử dụng cho định tuyến thông điệp broadcast - quảng bá.

#### c. Topic Exchange

      Topic exchange định tuyến message tới một hoặc nhiều queue dựa trên sự trùng khớp giữa routing key và pattern. Topic exchange thường sử dụng để thực hiện định tuyến thông điệp multicast. Ví dụ một vài trường hợp sử dụng:

* Phân phối dữ liệu liên quan đến vị trí địa lý cụ thể.
* Xử lý tác vụ nền được thực hiện bởi nhiều workers, mỗi công việc có khả năng xử lý các nhóm tác vụ cụ thể.
* Cập nhật tin tức liên quan đến phân loại hoặc gắn thẻ (ví dụ: chỉ dành cho một môn thể thao hoặc đội cụ thể).
* Điều phối các dịch vụ của các loại khác nhau trong cloud

#### d. Headers Exchange

      Header exchange được thiết kế để định tuyến với nhiều thuộc tính, đễ dàng thực hiện dưới dạng tiêu đề của message hơn là routing key. Header exchange bỏ đi routing key mà thay vào đó định tuyến dựa trên header của message. Trường hợp này, broker cần một hoặc nhiều thông tin từ application developer, cụ thể là, nên quan tâm đến những tin nhắn với tiêu đề nào phù hợp hoặc tất cả chúng.