No pineapple on pizza! Streaming anomaly detection with Apache Kafka and Apache Flink

Francesco Tisiot - Developer Advocate

@ftisiot

The Anomaly
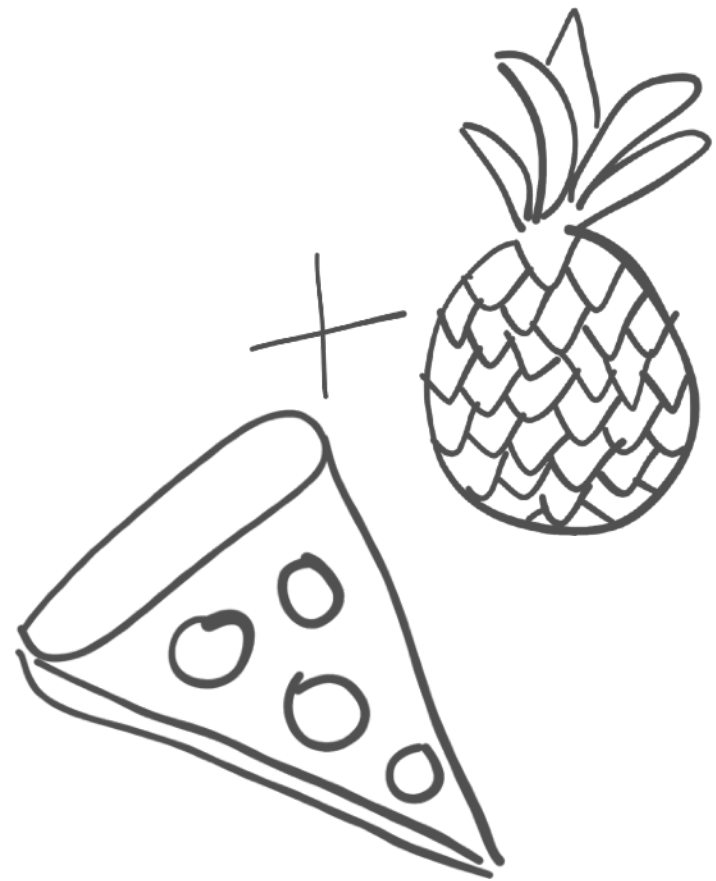
The
Streaming

Apache Flink
+
APACHE kafka®

@ftisiot | @aiven_io

https://github.com/aiven/fake-data-producer-for-apache-kafka-docker

The Data

```
{
  "id": 778,
  "shop": "Luigis Pizza",
  "name": "Edward Olson",
  "phoneNumber": "(935)503-3765x4154",
  "address": "Unit 9398 Box 2056\nDPO AP 24022",
  "pizzas": [
    {
      "pizzaName": "Salami",
      "additionalToppings":
          ["🥓 bacon", "🌶 hot pepper", "🧅 onion"]
    },
    {
      "pizzaName": "Margherita",
      "additionalToppings":
          ["🍌 banana", "🌶 hot pepper", "🍍 pineapple"]
    }
  ]
```

@ftisiot | @aiven_io

Sleeping for...0.0s
Sending: {'id': 283575, 'shop': 'Mammamia Pizza', 'name': 'Latoya Reynolds', 'phoneNumber': '2972657096', 'address': '555 Kerri Expressway Apt. 873\nSouth Michael, KY 02071', 'pizzas': [{'pizzaName': 'Marinara', 'additionalToppings': ['🍓 strawberry', '🌶 hot pepper', '🧀 blue cheese', '🧀 blue cheese', '🍅 tomato']}, {'pizzaName': 'Marinara', 'additionalToppings': ['🍍 pineapple', '🐟 tuna', '🍅 tomato', '🍍 pineapple', '🌶 hot pepper']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🫑 green peppers', '🍍 pineapple', '🍅 tomato']}, {'pizzaName': 'Margherita', 'additionalToppings': ['🐟 tuna', '🐟 tuna', '🥓 bacon']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🥚 egg', '🧅 onion']}, {'pizzaName': 'Mari & Monti', 'additionalToppings': []}, {'pizzaName': 'Mari & Monti', 'additionalToppings': []}, {'pizzaName': 'Salami', 'additionalToppings': ['🧄 garlic']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🧄 garlic', '🧅 onion', '🐟 tuna', '🥓 bacon']}], 'timestamp': 1659426822541}
Sleeping for...0.0s
Sending: {'id': 283576, 'shop': 'Circular Pi Pizzeria', 'name': 'Evelyn Mendez', 'phoneNumber': '001-751-498-2255x68922', 'address': '844 Andrea Neck Apt. 697\nNorth Amberton, KS 67696', 'pizzas': [{'pizzaName': 'Margherita', 'additionalToppings': ['🍓 strawberry', '🍅 tomato', '🍍 pineapple', '🐟 tuna', '🍓 strawberry']}], 'timestamp': 1659426822546}
Sleeping for...0.0s
Sending: {'id': 283577, 'shop': 'Its-a me! Mario Pizza!', 'name': 'Ian Aguilar', 'phoneNumber': '868.746.5974', 'address': '7425 Brown Valleys\nLake Nicholasland, WY 76186', 'pizzas': [{'pizzaName': 'Diavola', 'additionalToppings': ['🌶 hot pepper', '🫒 olives', '🥓 bacon', '🐟 tuna', '🍅 tomato']}, {'pizzaName': 'Margherita', 'additionalToppings': ['🥓 bacon', '🧄 garlic', '🍅 tomato', '🐟 tuna']}], 'timestamp': 1659426822551}
Sleeping for...0.0s
Sending: {'id': 283578, 'shop': 'Ill Make You a Pizza You Cant Refuse', 'name': 'Anthony Underwood', 'phoneNumber': '(607)408-5327x38049', 'address': '25123 Katherine Ramp\nHarrisontown, VA 96063', 'pizzas': [{'pizzaName': 'Mari & Monti', 'additionalToppings': ['🌶 hot pepper', '🧅 onion']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🐟 tuna', '🥚 egg']}, {'pizzaName': 'Peperoni', 'additionalToppings': ['🫑 green peppers', '🌶 hot pepper', '🫑 green peppers', '🍅 tomato']}, {'pizzaName': 'Marinara', 'additionalToppings': ['🐟 tuna', '🍓 strawberry', '🍍 pineapple']}, {'pizzaName': 'Marinara', 'additionalToppings': ['🌶 hot pepper', '🫑 green peppers']}], 'timestamp': 1659426822555}
Sleeping for...0.0s
Sending: {'id': 283579, 'shop': 'Marios Pizza', 'name': 'Dr. Phillip Turner', 'phoneNumber': '(690)657-0659x099', 'address': 'Unit 6876 Box 6282\nDPO AP 47591', 'pizzas': [{'pizzaName': 'Salami', 'additionalToppings': ['🌶 hot pepper', '🧀 blue cheese', '🫑 green peppers']}], 'timestamp': 1659426822558}
Sleeping for...0.0s
Sending: {'id': 283580, 'shop': 'Circular Pi Pizzeria', 'name': 'Leonard Park', 'phoneNumber': '371.695.3646x7547', 'address': '5586 Henry Keys\nStephenport, HI 05662', 'pizzas': [{'pizzaName': 'Margherita', 'additionalToppings': ['🍓 strawberry', '🍅 tomato', '🫒 olives', '🥓 bacon']}, {'pizzaName': 'Mari & Monti', 'additionalToppings': ['🥚 egg']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🍓 strawberry']}, {'pizzaName': 'Margherita', 'additionalToppings': ['🥚 egg', '🫑 green peppers', '🥚 egg', '🍍 pineapple', '🍅 tomato']}], 'timestamp': 1659426822560}
Sleeping for...0.0s
Sending: {'id': 283581, 'shop': 'Its-a me! Mario Pizza!', 'name': 'Amy Odonnell', 'phoneNumber': '642.793.1573x6271', 'address': '4099 James Underpass Suite 705\nPooleland, DE 15590', 'pizzas': [{'pizzaName': 'Diavola', 'additionalToppings': []}], 'timestamp': 1659426822561}
Sleeping for...0.0s
Sending: {'id': 283582, 'shop': 'Marios Pizza', 'name': 'Frank Donaldson', 'phoneNumber': '775.144.8297x539', 'address': '98536 Jason Skyway Apt. 594\nVancemouth, AK 57565', 'pizzas': [{'pizzaName': 'Salami', 'additionalToppings': ['🫑 green peppers', '🥚 egg']}, {'pizzaName': 'Peperoni', 'additionalToppings': ['🐟 tuna', '🧀 blue cheese', '🌶 hot pepper']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🧄 garlic', '🍓 strawberry', '🧅 onion', '🐟 tuna']}, {'pizzaName': 'Diavola', 'additionalToppings': []}], 'timestamp': 1659426822563}
Sleeping for...0.0s
Sending: {'id': 283583, 'shop': 'Luigis Pizza', 'name': 'Beverly Hooper', 'phoneNumber': '161.099.6606', 'address': '7476 Jackson Row Suite 681\nStephensonport, SD 62010', 'pizzas': [{'pizzaName': 'Diavola', 'additionalToppings': ['🫑 green peppers', '🧅 onion', '🫒 olives', '🫑 green peppers']}, {'pizzaName': 'Salami', 'additionalToppings': ['🧀 blue cheese', '🌶 hot pepper', '🧀 blue cheese']}, {'pizzaName': 'Diavola', 'additionalToppings': ['🍍 pineapple', '🍓 strawberry', '🐟 tuna', '🧀 blue cheese', '🫑 green peppers']}, {'pizzaName': 'Peperoni', 'add

What is an Anomaly?
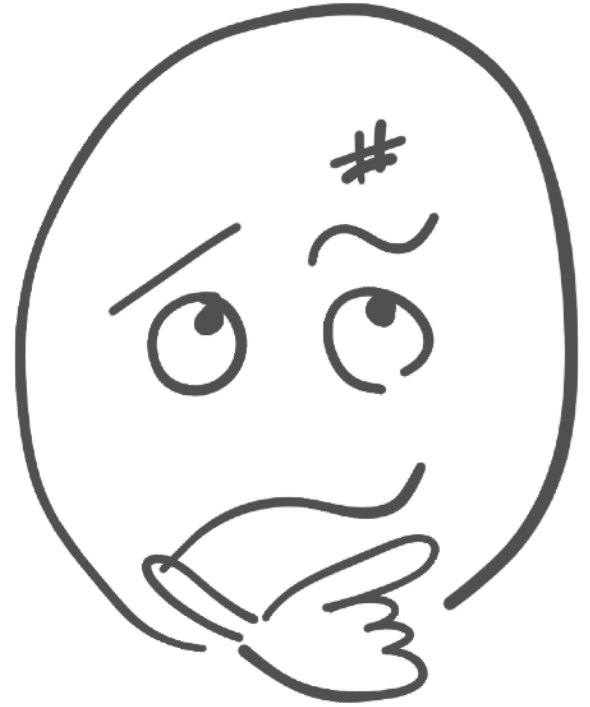
Get Those Bad Toppings
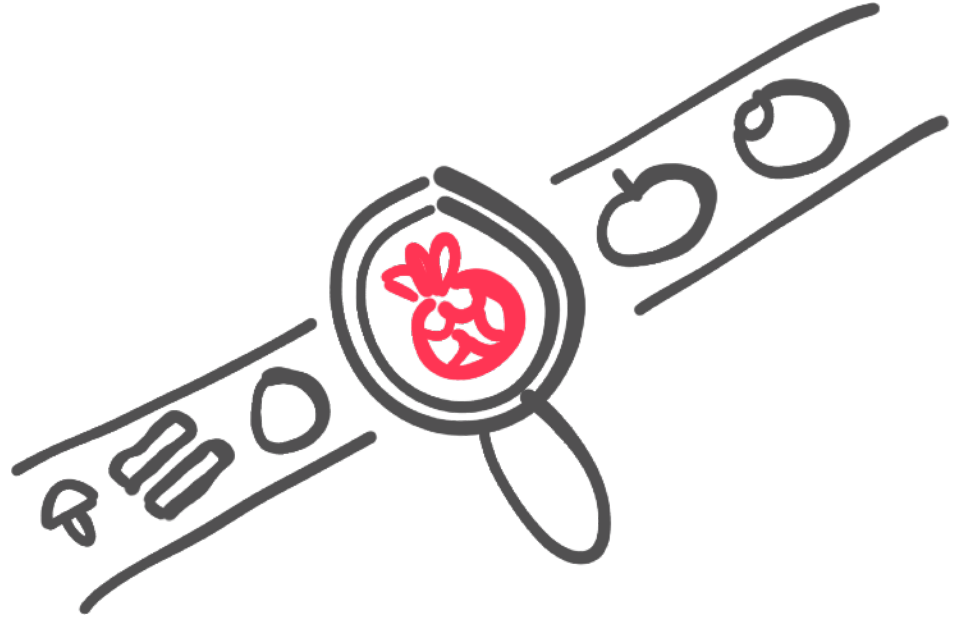
```sql
CREATE TABLE pizza_orders (
 id INT,
 shop VARCHAR,
 name VARCHAR,
 phoneNumber VARCHAR,
 address VARCHAR,
 pizzas ARRAY <ROW (pizzaName VARCHAR, additionalToppings ARRAY <VARCHAR>)>,
 orderTimestamp TIMESTAMP(3) METADATA FROM 'timestamp',
 WATERMARK FOR orderTimestamp AS orderTimestamp - INTERVAL '5' SECOND
) WITH (
  'connector' = 'kafka',
  'properties.bootstrap.servers' = '<HOST>:<PORT>',
  'topic' = 'pizza-orders',
  'scan.startup.mode' = 'earliest-offset',
  'value.format' = 'json',
  ...
);
```
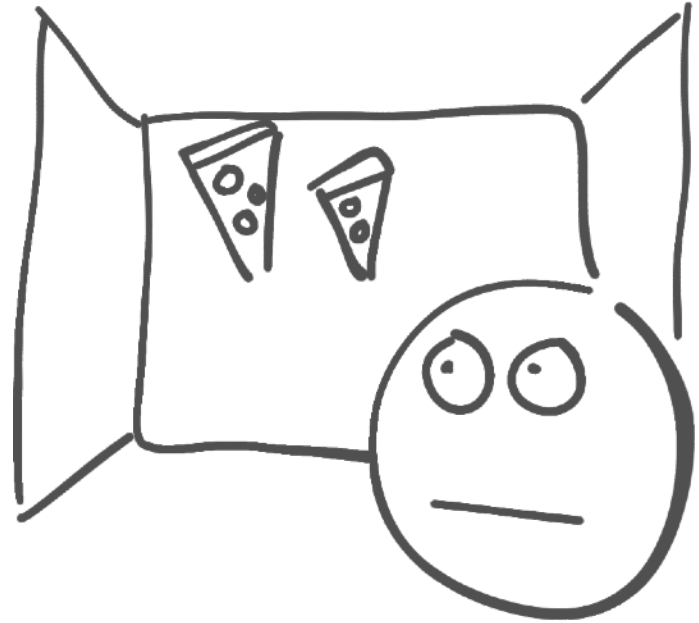
```
select
    id,
    name,
    c.topping,
    orderTimestamp
from pizza_orders
cross join UNNEST(pizzas) b
cross join UNNEST(b.additionalToppings) as c(topping)
where c.topping in
    ('🍍 pineapple', '🍓 strawberry','🍌 banana')
```

Allow Mistakes

```
select
    id,
    name,
    count(*) nr_bad_items,
    LISTAGG(c.topping) list_bad_items,
    orderTimestamp
from pizza_orders
cross join UNNEST(pizzas) b
cross join UNNEST(b.additionalToppings) as c(topping)
where c.topping in
    ('🍍 pineapple', '🍓 strawberry','🍌 banana')
group by id, name, orderTimestamp
having count(*) > 3
```

Calculate Over Windows

```sql
SELECT window_time,
    window_start,
    window_end,
    topping,
    count(*) nr_orders
FROM TABLE(TUMBLE(TABLE pizza_orders,
DESCRIPTOR(orderTimestamp), interval '5' seconds))
    cross join UNNEST(pizzas) b
    cross join UNNEST(b.additionalToppings) as c(topping)
where c.topping in ('🍍', '🍓','🍌')
group by window_time,
    topping,
    window_start,
    window_end
having count(*) > 10
```

Detect Trends
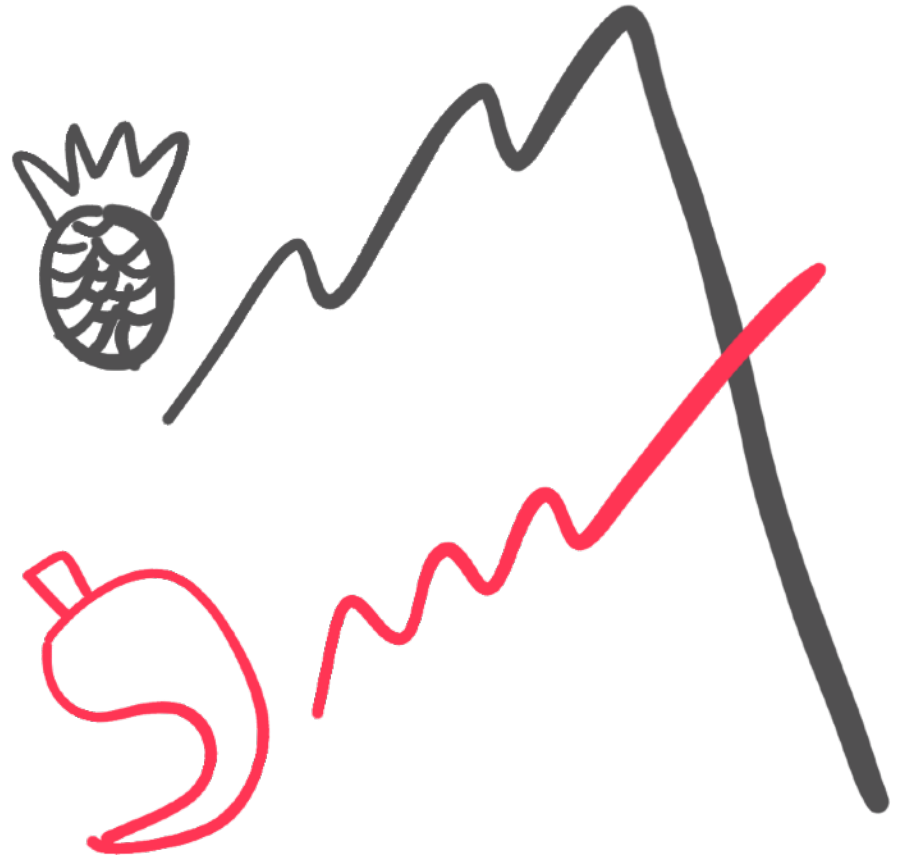
```
MATCH_RECOGNIZE (
    PARTITION BY topping
    ORDER BY window_time
    MEASURES

        ...
    PATTERN (START_ROW NR_UP+ NR_DOWN{2})
    DEFINE
        NR_UP AS

            ...
        NR_DOWN AS

            ...
)
```

https://ftisiot.net/talks/anomaly-detection/

https://github.com/ftisiot/anomaly-detection-with-apache-flink-and-kafka

@ftisiot | @aiven_io