Data Ingestion with Azure Event Hubs

Data Science Dojo

Typical Event Processing

Applications

Cloud Gateways (WebAPIs)

Scalable Event Broker

External Data Sources

Web/Thick Client Dashboards

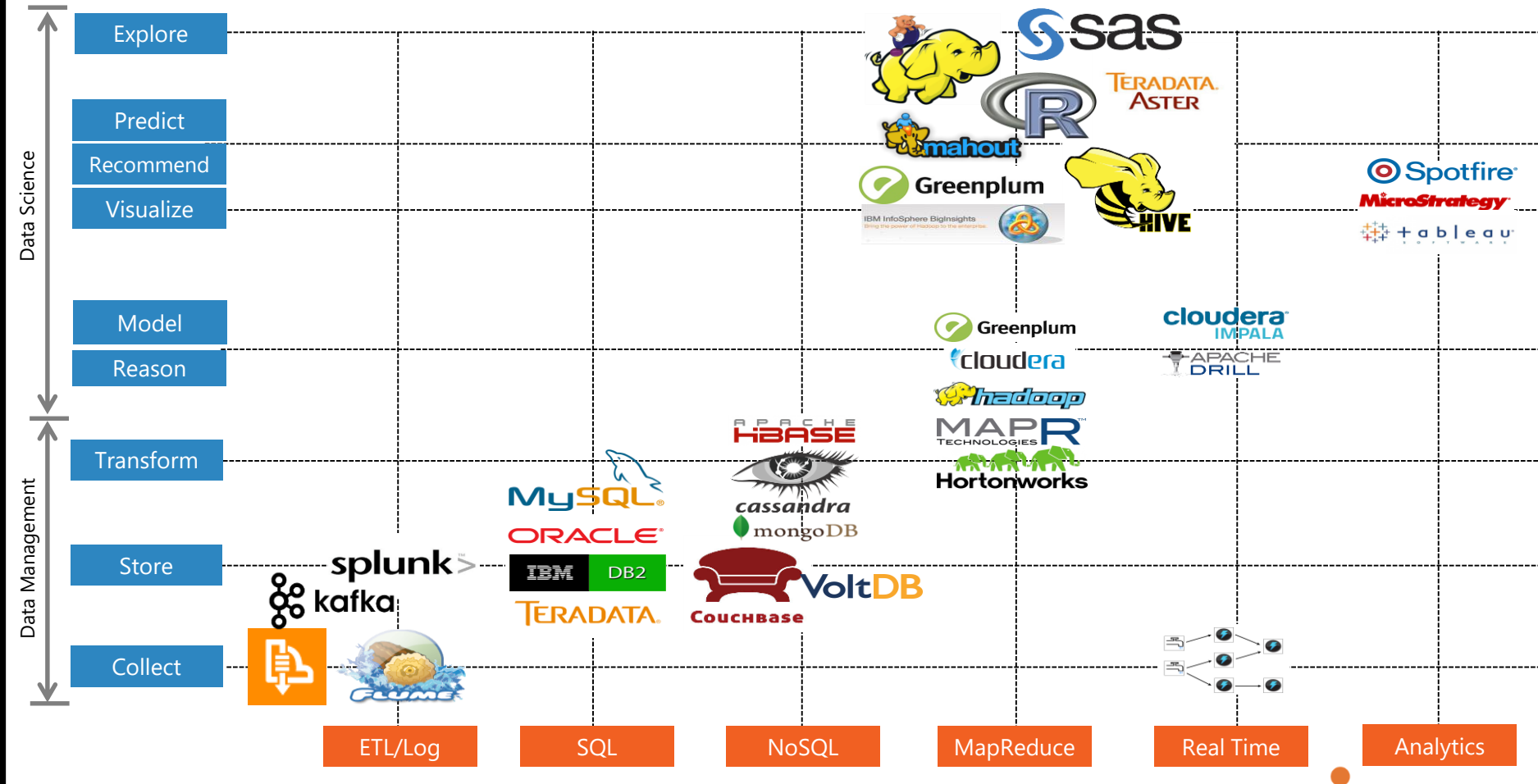
Devices

Field Gateways

Search and Query

Data Analytics

datasciencedojo
data science for everyone

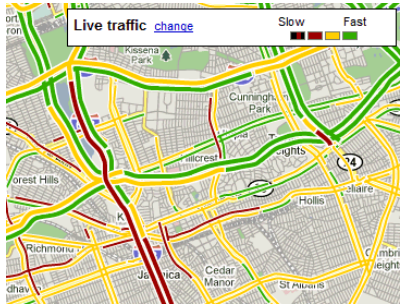


datasciencedojo
data science for everyone

The Post Office & Shipping Centers

- Tracks address changes
- Tries again tomorrow if send failed
- Holds packages in short term
  - Too many failed deliveries
  - Vacations
- Reduces complexity through specialization
- Optimized to send, receive, and temporarily house packages

Hypothetical Scenario



Sends Traffic Data

Bureau of Transportation Statistics

Server: ABC, Port 123

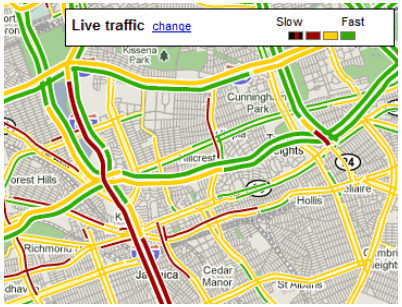
SQL DB

Your Company

datasciencedojo
data science for everyone

Things Change

Bureau of Transportation Statistics



Sends Traffic Data

WAS Server: ABC, Port 123
Now Server: DEF, Port 456

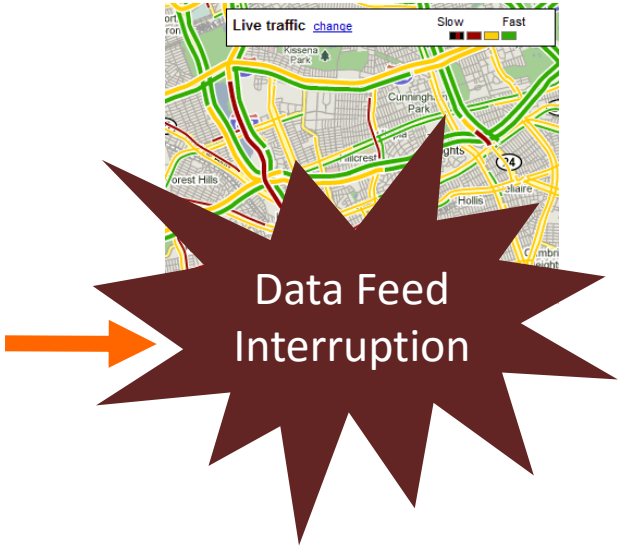
SQL DB

Your Company

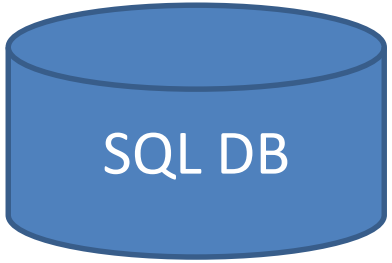
datasciencedojo
data science for everyone

Interruption Occurs

Bureau of Transportation Statistics



WAS Server: ABC, Port 123
Now Server: DEF, Port 456



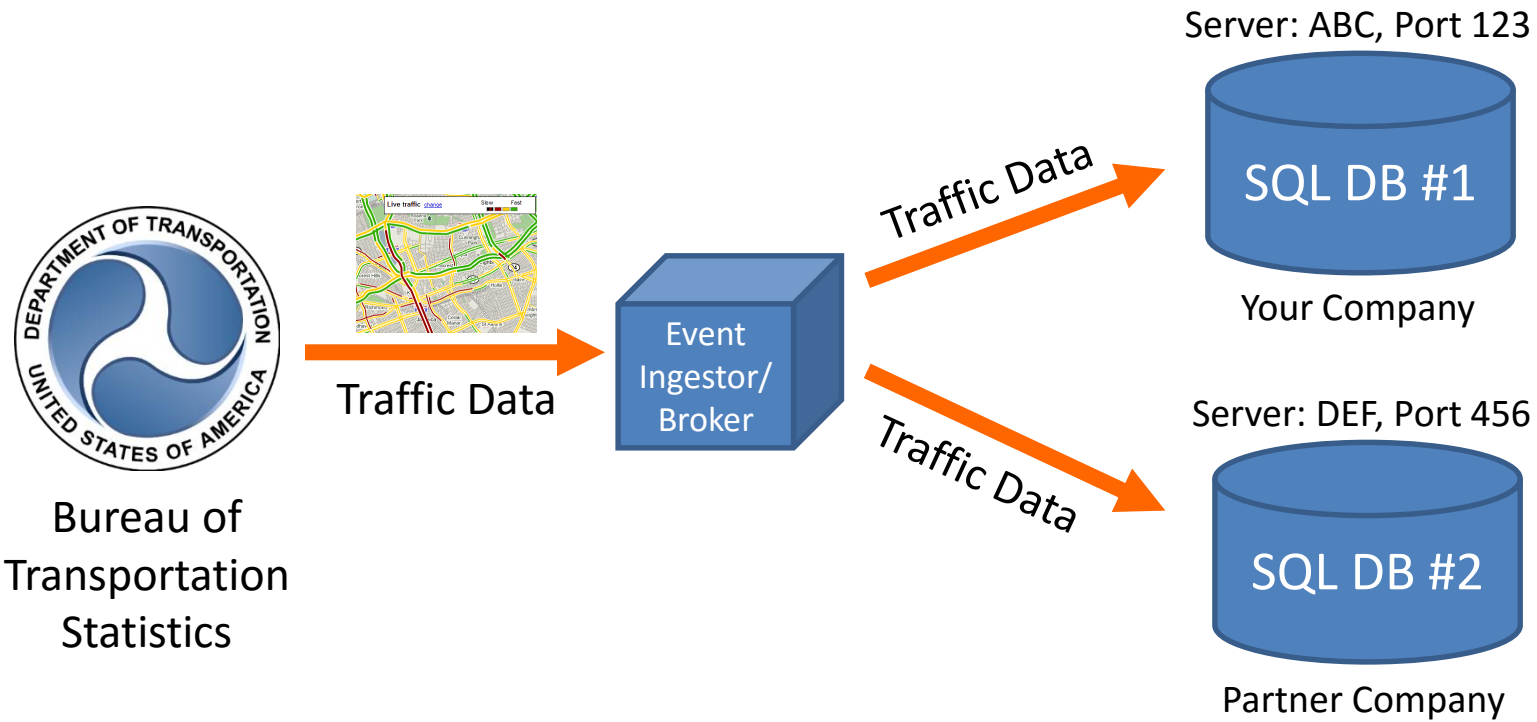
Your Company

datasciencedojo
data science for everyone

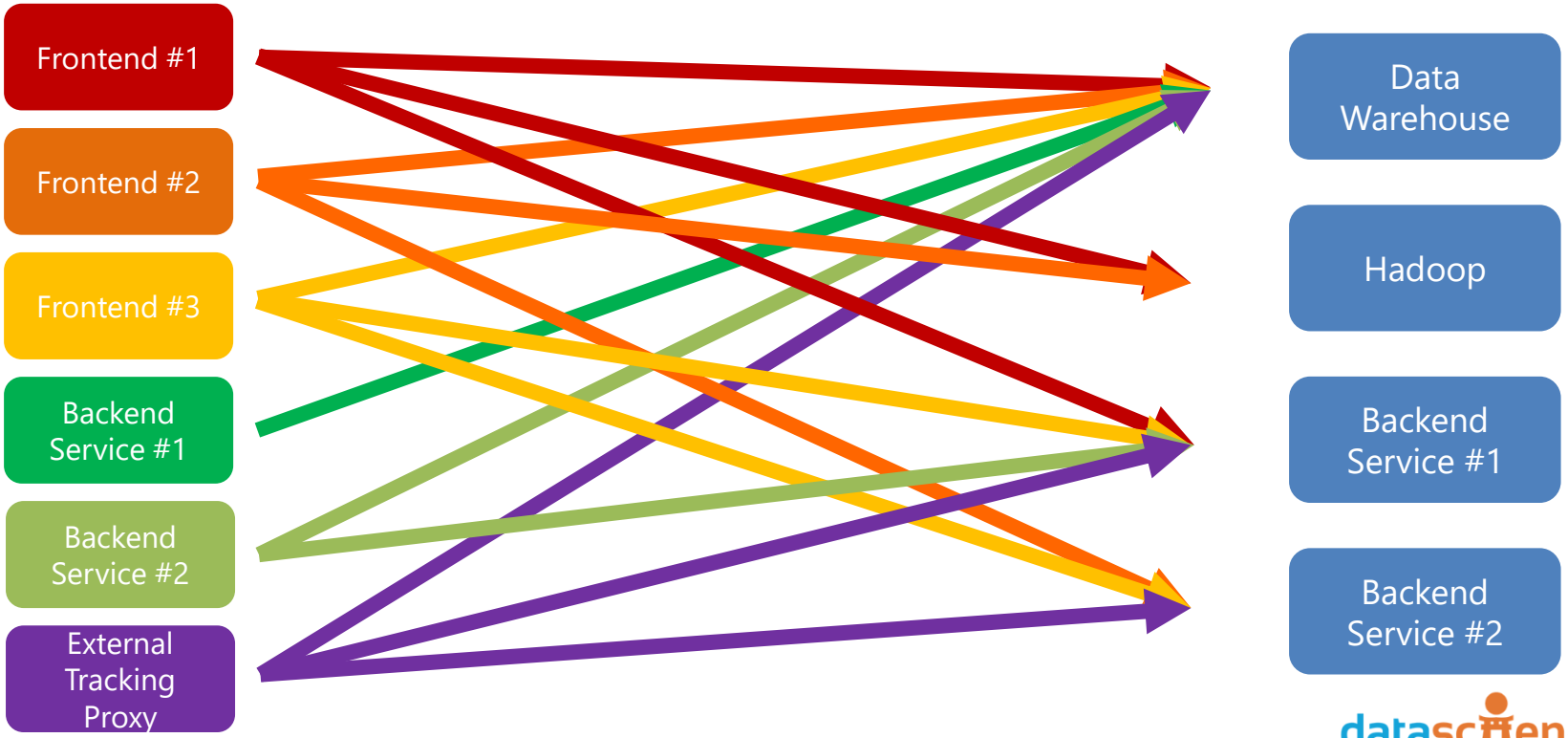
Reactive Remedy

- Call Bureau of Transportation Statistics to change their outbound data to funnel into server DEF and port 456
- Bureau makes support ticket
  - Your request enters their task queue
- 4 days later… ports are changed

Preventative Solution: Middleware

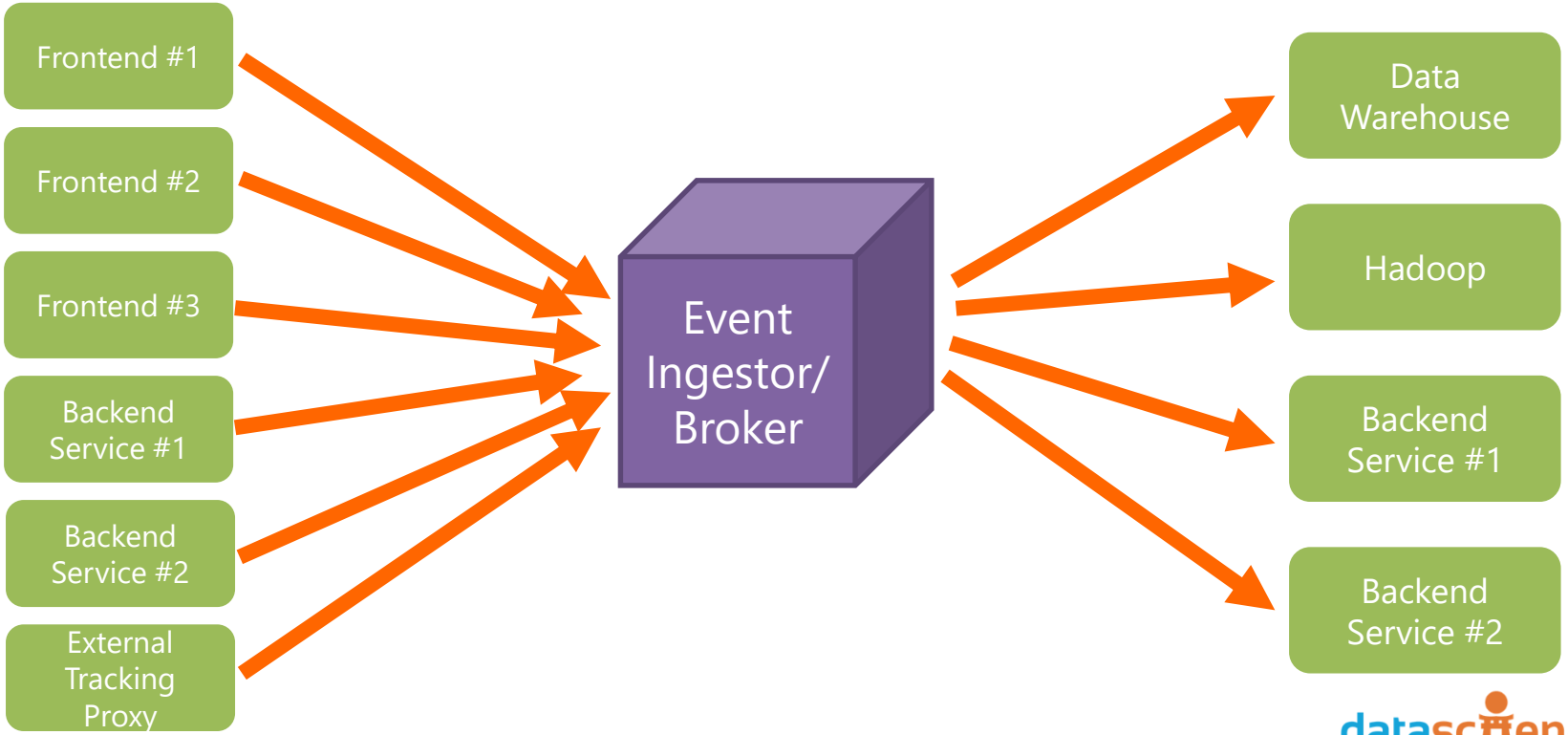


Data Pipeline Complexity at LinkedIn



datasciencedojo
data science for everyone

Data Pipeline Complexity at LinkedIn



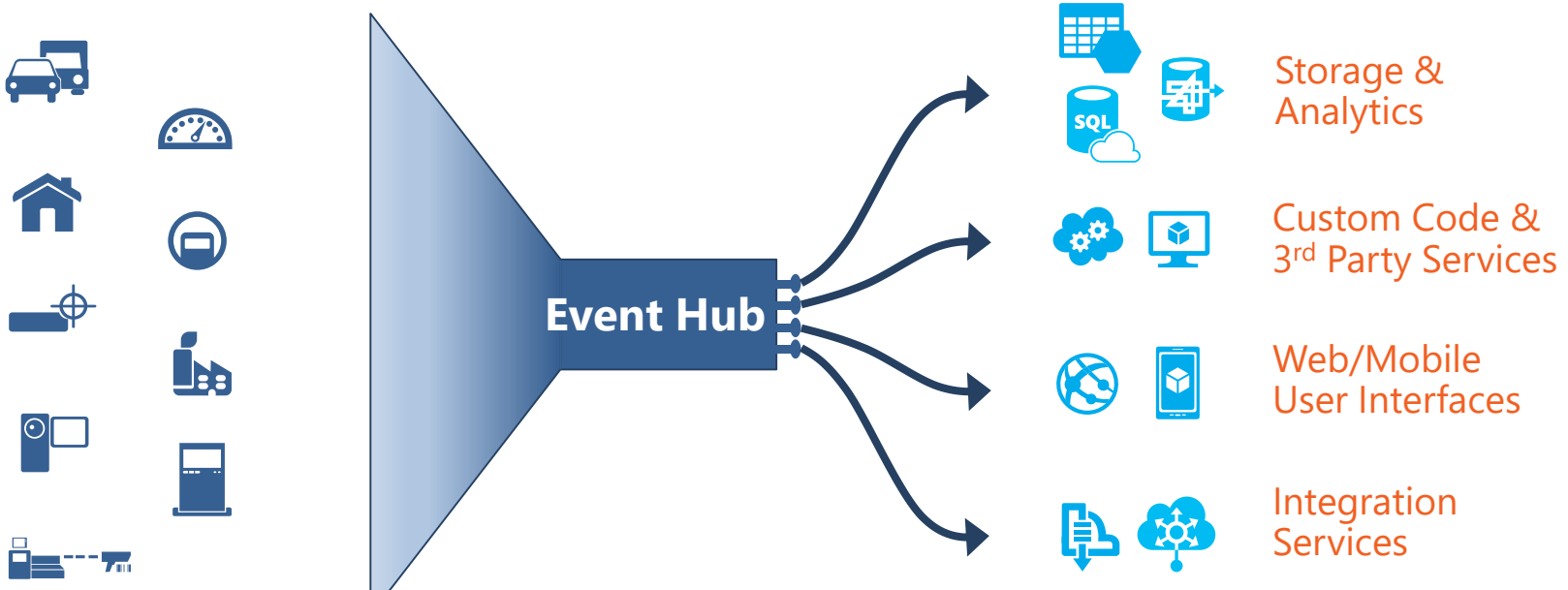
datasciencedojo
data science for everyone

Popular Event Brokers

Azure Event Hubs

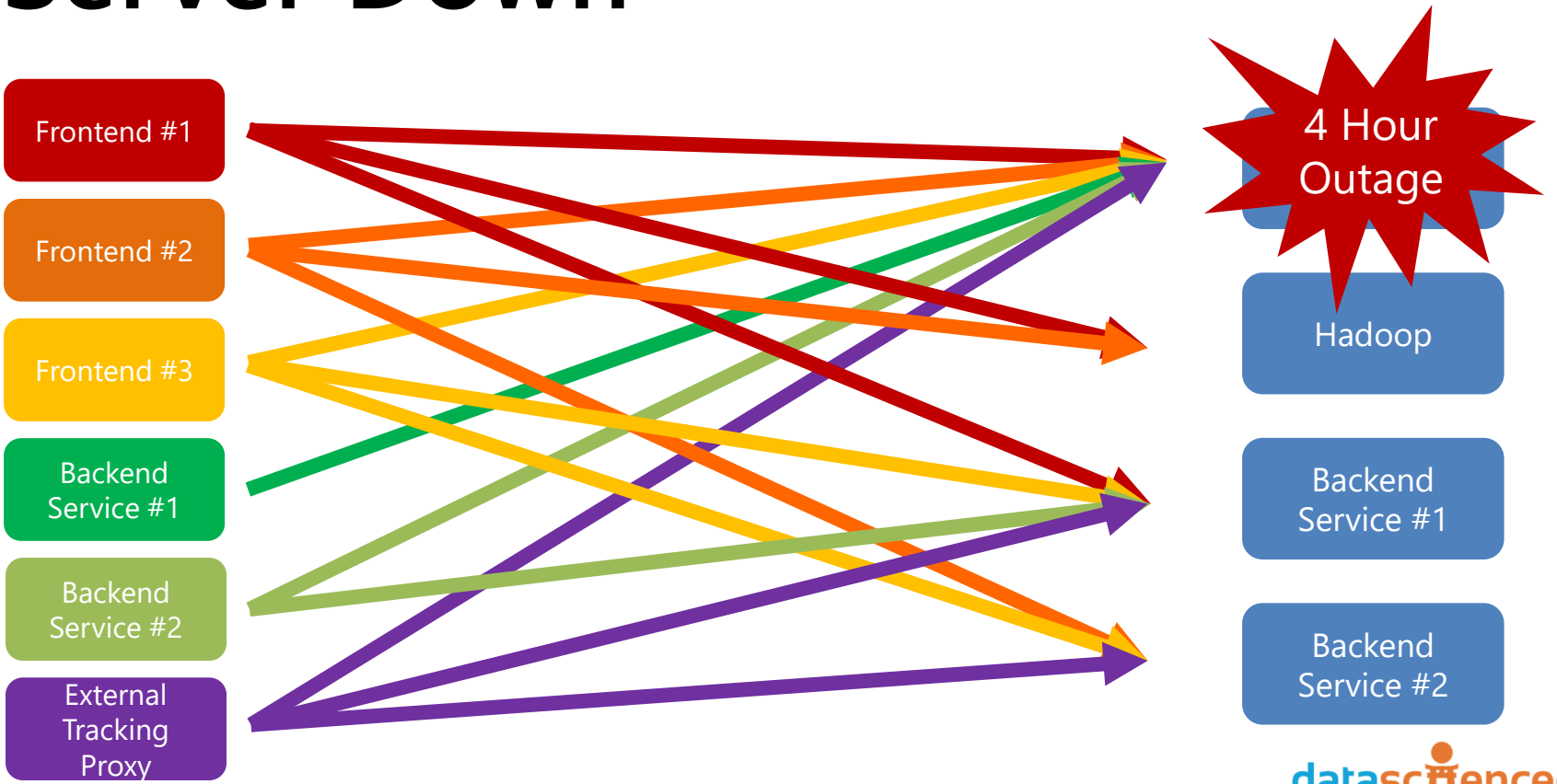
datasciencedojo
data science for everyone

Event Hub for IoT: Big Data Ingestion



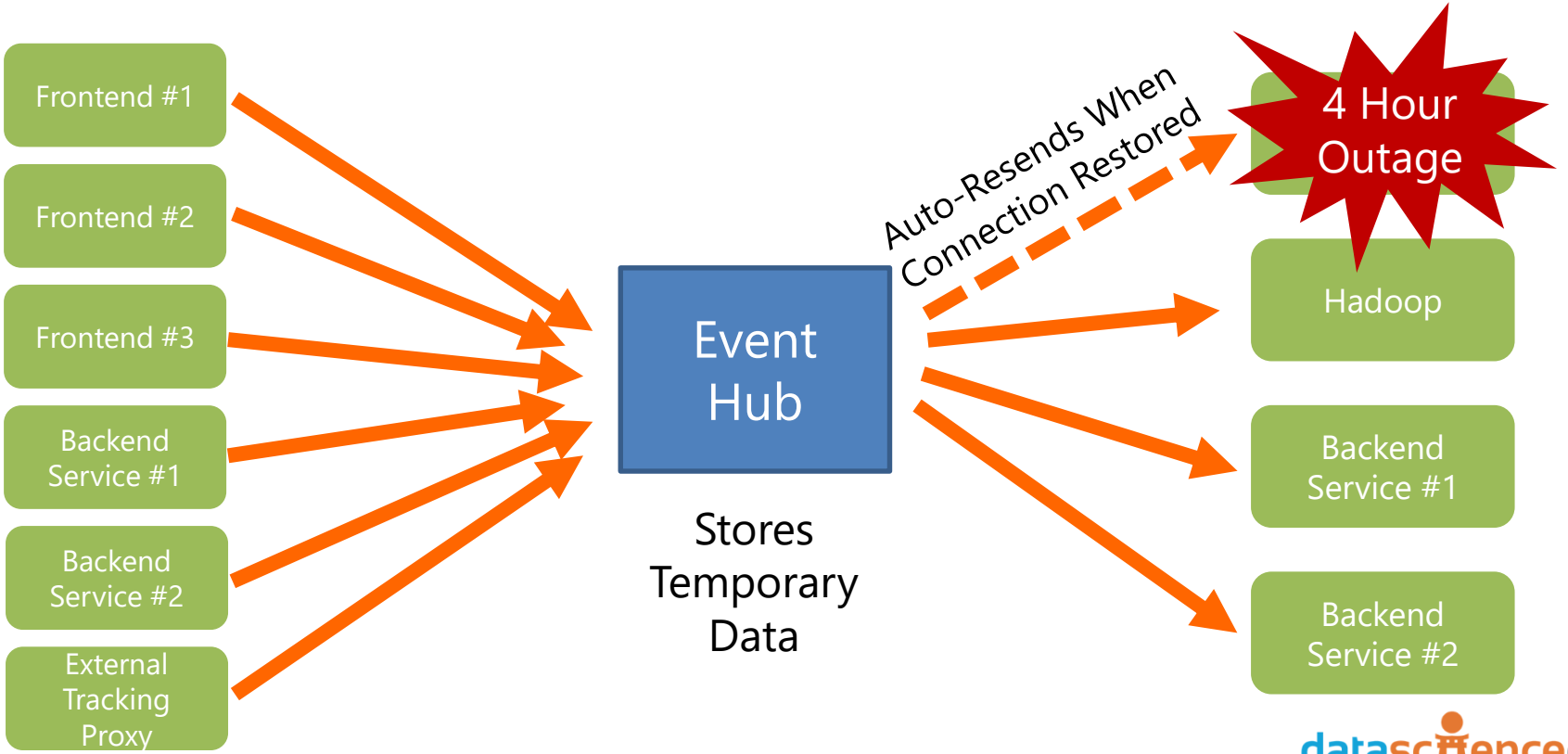
datascience dojo
data science for everyone

Server Down



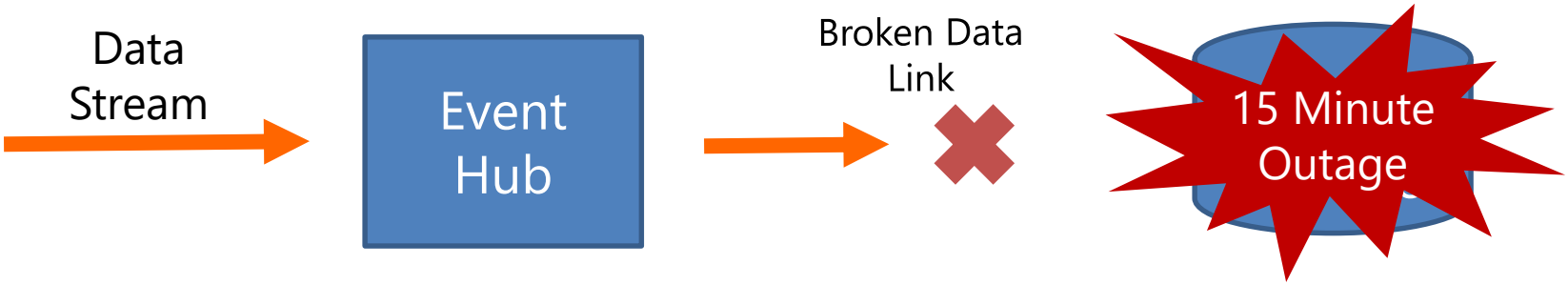
datasciencedojo
data science for everyone

Temporary Storage

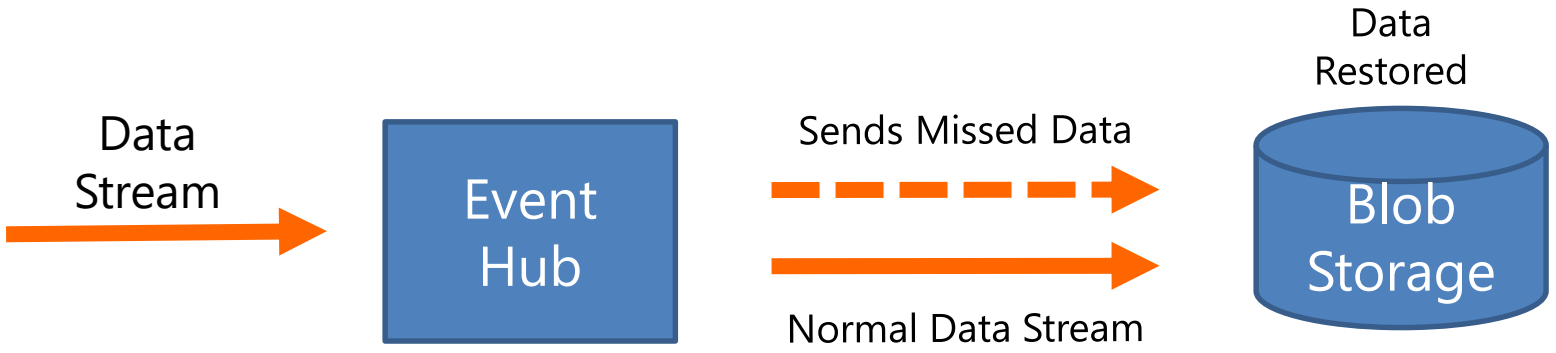


Demo: Normal Scenario

Demo: Output Downage



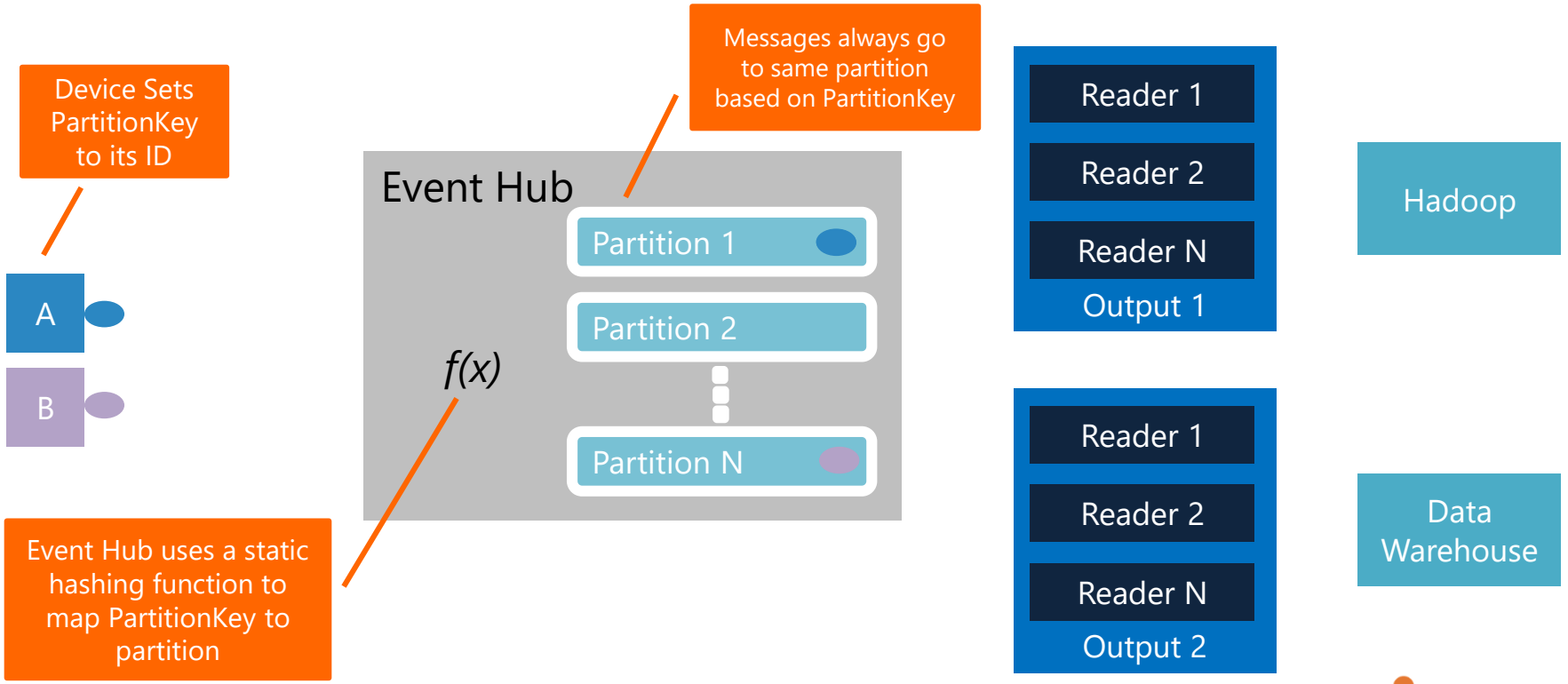
Demo: Output Restored



The Post Office

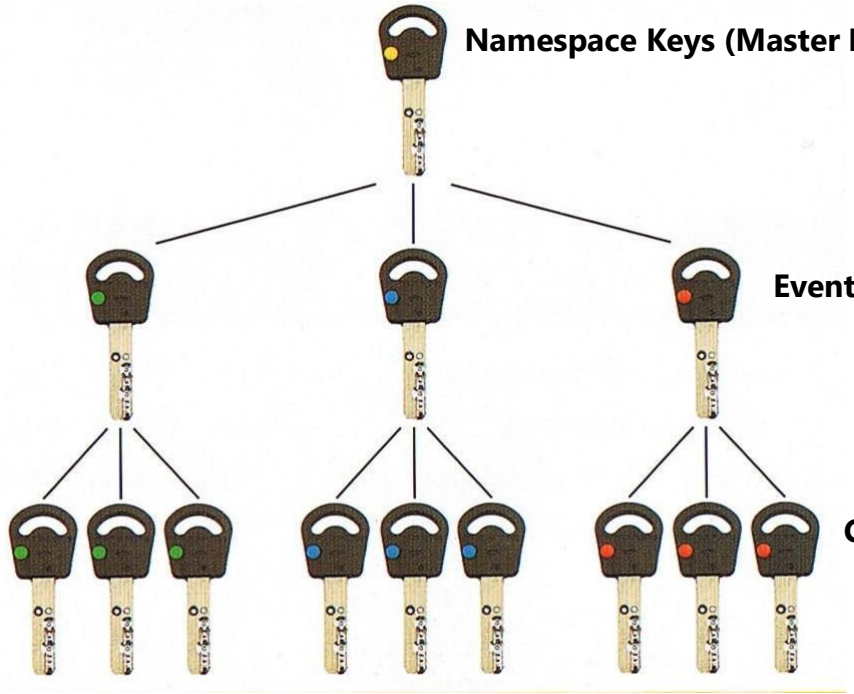
- Tracks address changes
- Tries again tomorrow if send failed
- Holds packages in short term
  - Too many failed deliveries
  - Vacations
- Reduces complexity through specialization

Event Hub, Stream Management

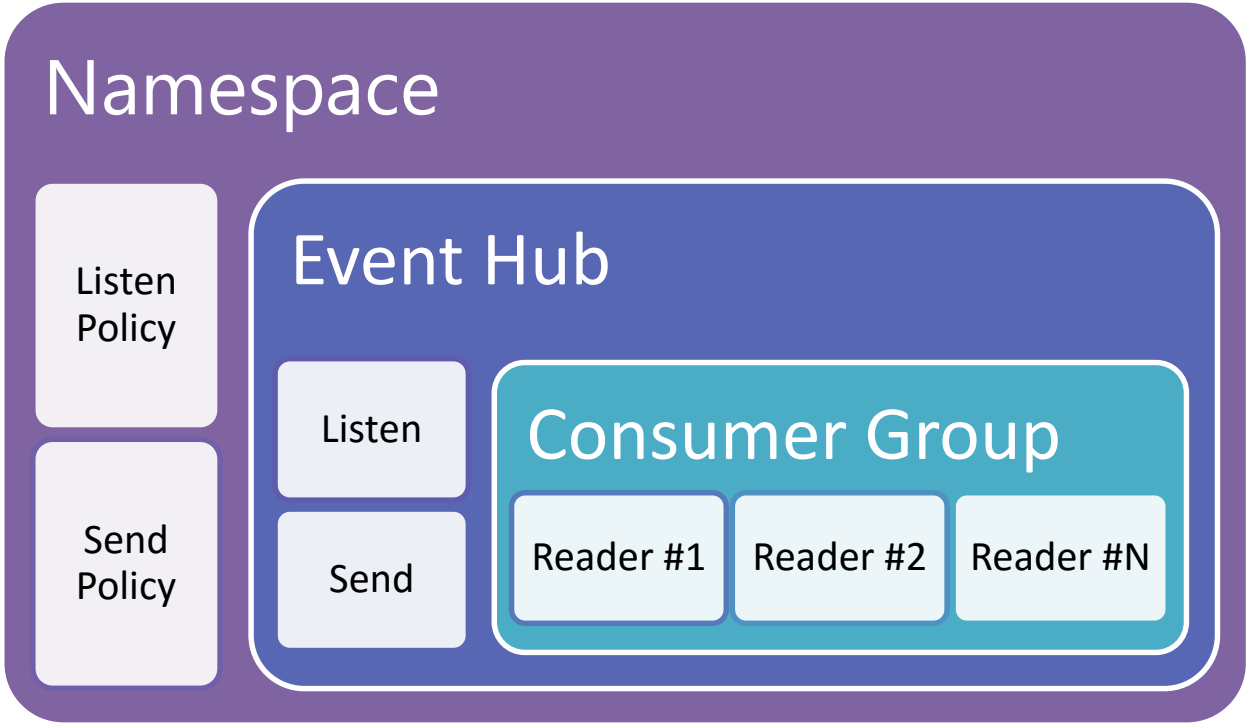


Service Bus Namespace

Access Rights, Policy, Keys



Access Rights



Access Rights

Hands-On Lab

Credit Card Transactions (swipes)

- Credit card transactions are usually done in batch as an end-of-the-day send.
- Stream process for insights now.
- US mainland transactions

Streaming to Event Hub

Swipes

Credit Card Reader (Synthetic)

Message Broker (DataScienceDojo's Webpage)

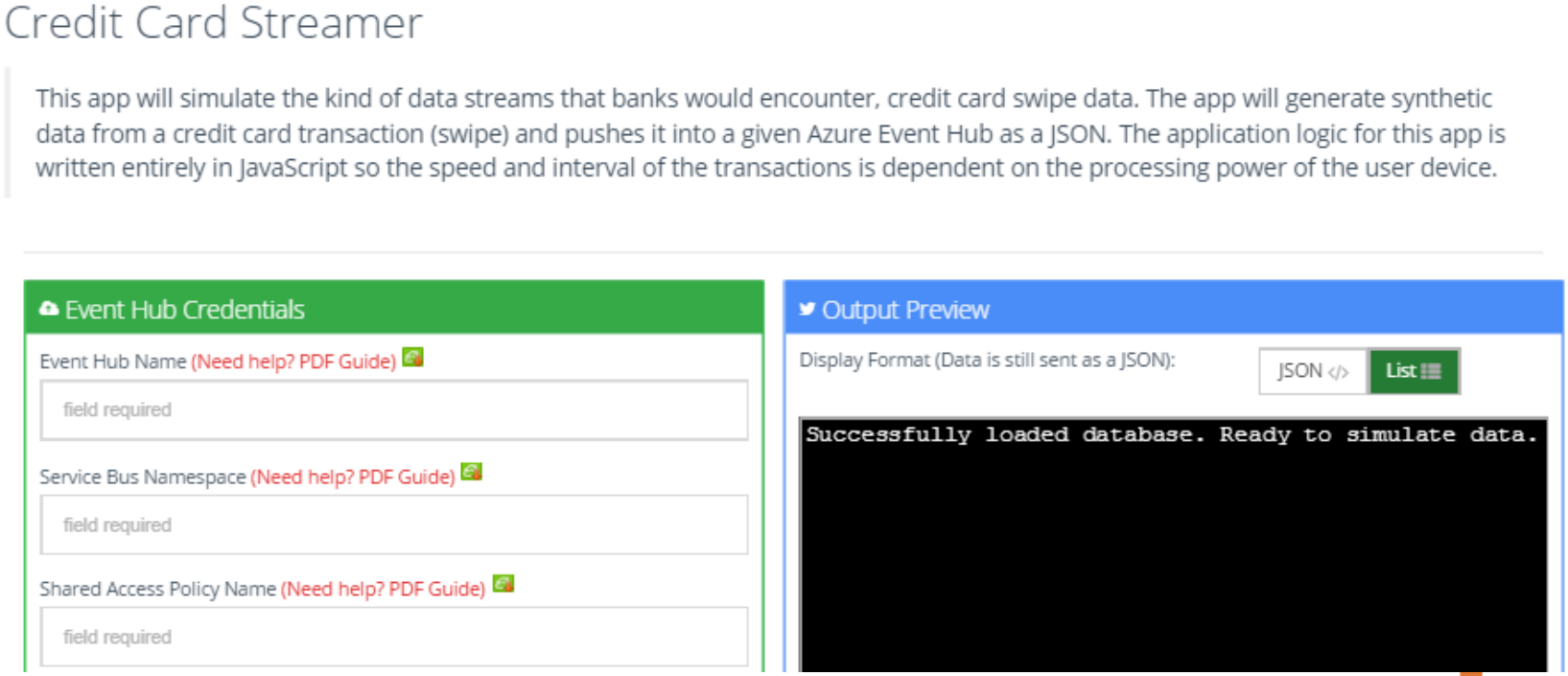
Data Ingestor (Azure Event Hub)

The Data


```
{
  "swipe_date":"2015-05-22T20:16:27.122Z",
  "transaction_id":3127484,
  "card_type":"VISA",
  "card_number":"4913419738164560",
  "expiration_month":"02",
  "expiration_year":"18",
  "cvv_code":"520",
  "user_id":"972288",
  "user_gender":"male",
  "user_first_name":“Alexander",
  "user_last_name":“Hamilton",
  "merchant":“McDonald’s",
  "transaction_amount":13.64,
  "balance":336.48,
  "merchant_fee":.5,
  "swipe_city":"New York",
  "swipe_state":"New York",
  "swip_city_state":"New York, NY",
  "InstanceNo":1
}
```

The Streamer

- http://demos.datasciencedojo.com/app/credit-card-streamer/



Inside the Event Hub

Swipes

data science dojo

Service Bus Namespace:
YourNameSpace

Event Hub 1:

- Credit Card Swipes

Credit Card Reader (Synthetic)

Message Broker (DataScienceDojo's Webpage)

Setting Policies

Message Broker (DataScienceDojo's Webpage)

Data Ingestor (Azure Event Hub)

Stream Processor (Azure Stream Analytics)

QUESTIONS