

The background of the slide features a light blue map-like pattern. It consists of several wavy, solid blue lines that meander across the frame. Interspersed among these solid lines are dashed blue lines, some of which are terminated by small blue 'x' marks. Additionally, there are several small, solid blue dots scattered across the map. A white rectangular box with a thin blue border is centered on the slide, containing the title text.

Stream Processing with Faust

What is Faust?



- Faust was created at Robinhood
- Design goal was to replicate Kafka Streams in Python
- Shares conceptual design patterns with Kafka Streams
- Stream Processing natively written in Python
- Concepts applicable in other frameworks as well

Introduction to Faust

Writing Faust apps is simple and requires no external dependencies

```
import faust

app = faust.App("demo-app", broker="PLAINTEXT://localhost:9092")
topic = app.topic("a-kafka-topic")

@app.agent(topic)
async def purchase(purchases):
    async for purchase in purchases:
        # do something!

if __name__ == "__main__":
    app.main()
```

Hello Faust


Writing Faust apps is simple and requires no external dependencies

```
import faust


app = faust.App("demo-app", broker="PLAINTEXT://localhost:9092")
topic = app.topic("a-kafka-topic")

@app.agent(topic)
async def purchase(purchases):
    async for purchase in purchases:
        # do something!

if __name__ == "__main__":
    app.main()
```

The background of the slide is a light blue map with various wavy lines, some solid and some dashed, and several small blue dots scattered across it. A white rectangular box with a blue border is centered on the slide, containing the title text.

Demo: Your First Faust Application Demonstration

The background of the slide features a light blue map-like pattern. It includes several solid blue lines of varying thicknesses that curve across the frame. Interspersed among these are dashed blue lines, some of which end in small 'x' marks. There are also several solid blue dots scattered across the map. A white rectangular box with a thin blue border is centered on the slide, containing the title text.

Serialization and Deserialization

Python Dataclasses

Python 3.7 Dataclasses are used only for containing data with type hints

```
from dataclasses import dataclass
```

```
@dataclass(frozen=True)
```

```
class Purchase:
```

```
    username: str = ""
```

```
    currency: str = ""
```

```
    amount: int = 0
```

Faust Deserialization

Deserialization is handled by specifying a type to the Faust topic

```
import faust

class Purchase(faust.Record, validation=True, serializer="json"):
    username: str
    currency: str
    amount: int

app = faust.App("demo-app", broker="PLAINTEXT://localhost:9092")
topic = app.topic("purchases", key_type=str, value_type=Purchase)

@app.agent(topic)
async def purchase(purchases):
    async for purchase in purchases:
        # do something!
```


The background of the slide features a light blue map-like pattern. It includes several solid blue lines that curve across the frame, and dashed blue lines that also curve. Small blue dots are placed at various points along these lines. Additionally, there are small 'x' marks scattered across the background, some of which are also connected by dashed lines.

Demo: Deserialization

Faust Serialization

Faust Serialization shares the same semantics as Deserialization

```
import faust


class Purchase(faust.Record, validation=True, serializer="binary|json"):
    username: str
    currency: str
    amount: int

app = faust.App("demo-app", broker="PLAINTEXT://localhost:9092")
topic = app.topic("purchases", key_type=str, value_type=Purchase)

@app.agent(topic)
async def purchase(purchases):
    async for purchase in purchases:
        # do something!
```

The background of the slide features a light blue map-like pattern. It consists of several wavy, solid blue lines that meander across the frame. Interspersed among these solid lines are dashed blue lines, some of which terminate in small 'x' marks. Additionally, there are five solid blue dots scattered across the map: one on the left side, one near the top center, one on the right side, one near the bottom center, and one near the bottom right.

Demo: Serialization

The background of the slide features a light blue map-like pattern. It includes several solid blue lines of varying thicknesses that curve across the frame. Interspersed among these are dashed blue lines, some of which end in small 'x' marks. There are also several solid blue dots scattered across the map. A white rectangular box with a thin blue border is centered on the slide, containing the title text.

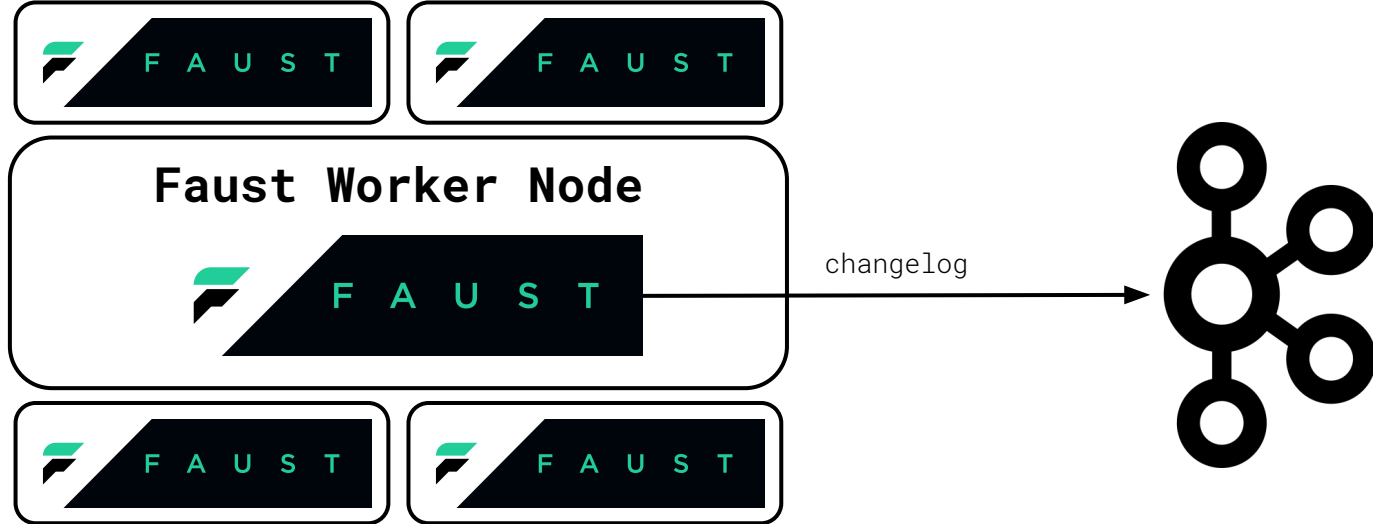
Serialization and Deserialization Summary

The background of the slide features a light blue map-like pattern. It consists of several wavy, solid blue lines and dashed blue lines. Small blue dots are placed at various points along these lines, and small 'x' marks are scattered across the background.

Storage Overview

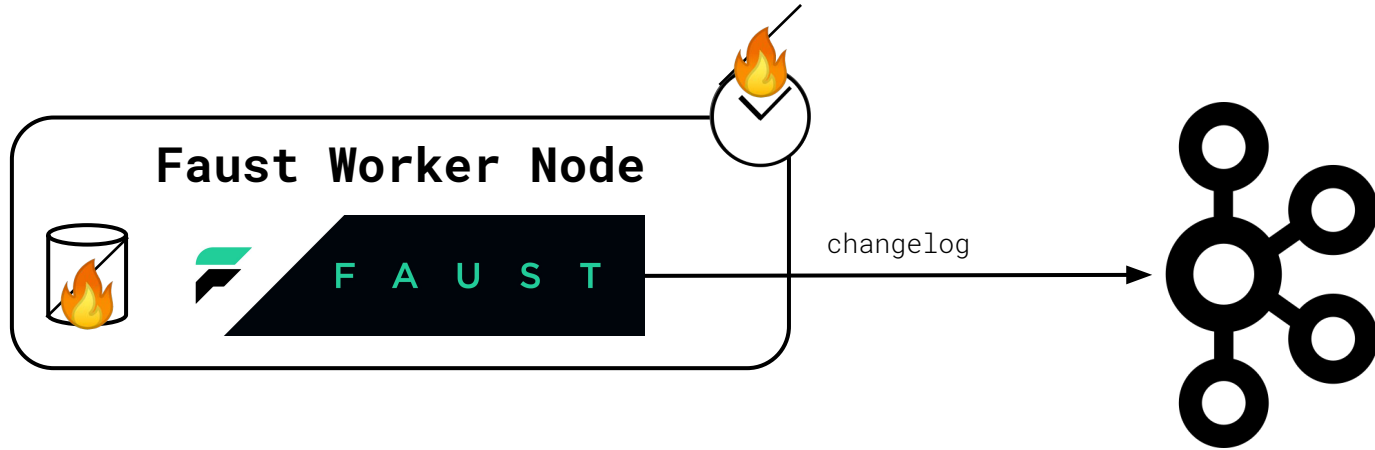
Kafka State

Faust stores its changelog in Kafka, and uses RocksDB for local state



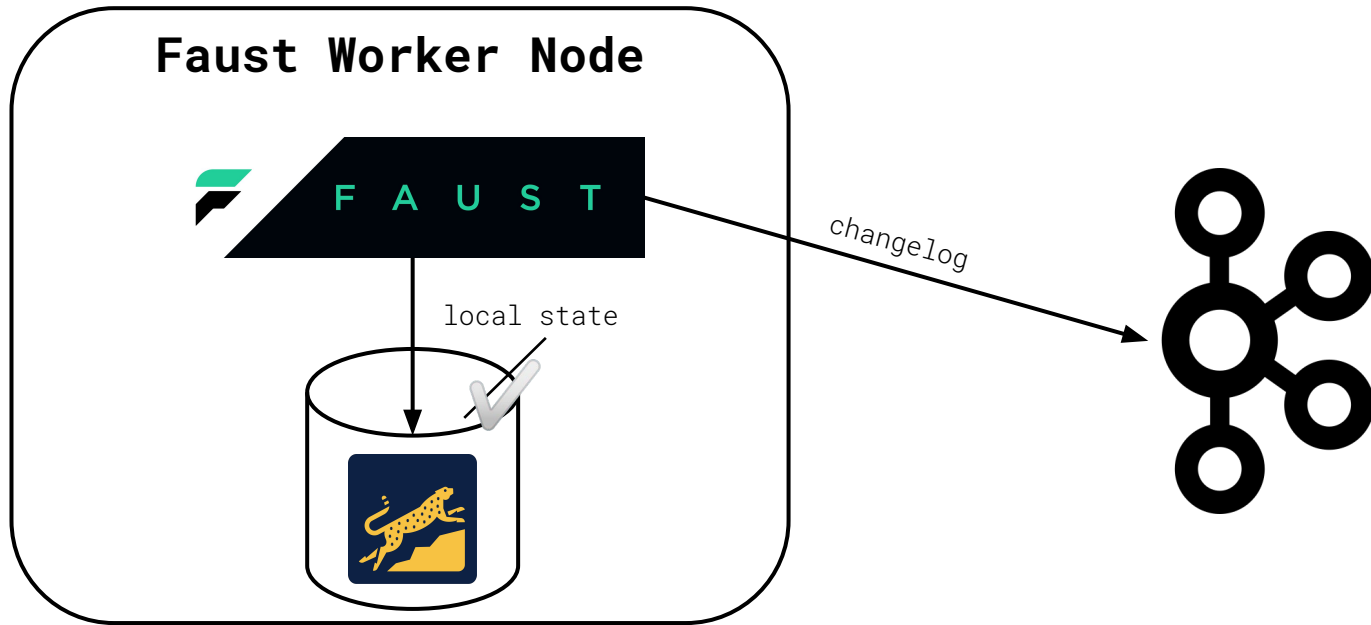
In-Memory

In-Memory storage should only be used for test and local development



RocksDB

Faust uses RocksDB to store local state in Production



The background is a stylized map showing contour lines. Solid blue lines represent contour lines, while dashed blue lines represent stream beds. Five blue dots are placed on the stream beds, indicating specific points of interest or measurement.

Streams

Streams Basics

Streams are automatically created for agent functions

```
source_topic = app.topic('source-topic')
```

```
@app.agent(source_topic)
```

```
async def process(stream):
```

```
    async for value in stream:
```

```
        # do something
```

Streams Basics

Results of stream operations are placed in a defined output topic

```
source_topic = app.topic('source-topic')
```

```
dest_topic = app.topic('dest-topic')
```

```
@app.agent(source_topic)
```

```
async def process(stream):
```

```
    async for num_msg, demo_value in stream.enumerate():
```

```
        await dest_topic.send(key=num_msg, value=demo_value)
```

Faust Streams

Message Life Cycle and Acknowledgment

- Faust manages consumer groups and offsets
- Faust uses the `aiokafka` to interact with Kafka
- `aiokafka` manages the consumer and offsets
- Uses one underlying subscription to topics for all agents

The background of the slide is a light blue map. It features several solid blue lines representing rivers or streams, which are irregular and wavy. Interspersed among these are dashed blue lines, some of which end in a small 'x' mark. There are also several small, solid blue dots scattered across the map, some positioned near the solid lines and others in open areas.

Demo: Streams in Action

Processors

Processor functions can be chained into streams to transform data

```
def transform(input_model):  
    return transform(input_model)  
  
async def delete_field(input_model):  
    return delete_pii(input_model)  
  
demo_stream = app.stream(  
    demo_topic,  
    processors=[transform, delete_field]  
)
```

Operations

Faust Streams

- Operations are helpers for processing streams
- `group_by` re-partitions the input stream on a new key
- `filter` removes unwanted data

The background of the slide features a light blue map-like pattern. It includes several solid blue lines of varying thicknesses that curve across the frame. Interspersed among these are dashed blue lines, some of which end in small 'x' marks. There are also several solid blue dots scattered across the map. A white rectangular box with a thin blue border and a subtle drop shadow is centered on the slide, containing the title text.

Demo: Processors and Operations

The background of the slide is a light blue map. It features several solid blue contour lines that curve across the frame. Interspersed among these are dashed blue lines, some of which end in a small 'x' mark. There are also five solid blue dots scattered across the map, likely representing specific locations or data points.

Streams Summary

The background of the slide is a light gray map. It features several blue lines, some solid and some dashed, which represent geographical features like rivers or roads. There are also several small blue dots scattered across the map, possibly representing locations or points of interest. The overall style is clean and modern.

Tables

Tables

Tables are the result of aggregations and have dict-like syntax

```
purchase_stopici = app.t.topic(c('source-topic',) value_type=Purchase)
```

```
currency = app.Table("my-first-table", default=istr)
```

```
@app.page(t(purchase_stopici))
```

```
async def process_stream(currency(purchases):
```

```
    async for purchase in purchases.group_by(Purchase.currency):
```

```
        currency.setdefault(currency, 0)
        currency[currency] += purchase.amount
```

The background of the slide is a light blue map with various wavy lines, some solid and some dashed, and several small blue dots scattered across it. A white rectangular box with a blue border is centered on the slide.

Demo: Tables in Action

Windowing in Faust

Faust supports Tumbling and Hopping Windows

```
table = app.Table("currency_total", int)
# Tumbling Window
tumbling_table = table.Tumbling(size=timedelta(minutes=5))
# Hopping Window
hopping_table = table.Hopping(
    size=timedelta(minutes=5),
    step=timedelta(minutes=1),
    expires=timedelta(minutes=60),
)
```

The background of the slide is a light blue map. It features several solid blue lines representing coastlines or major roads, and several dashed blue lines representing smaller paths or rivers. There are also several small blue dots scattered across the map, possibly representing cities or specific locations. The overall style is clean and modern.

Demo: Tumbling Windows

The background of the slide is a light blue map with various wavy lines, some solid and some dashed, and several small blue dots scattered across it. A white rectangular box with a blue border is centered on the slide.

Demo: Hopping Windows

The background of the slide features a light blue map-like pattern. It includes several solid blue lines that curve across the frame, and several dashed blue lines, some of which end in small 'x' marks. There are also five solid blue dots scattered across the map. A white rectangular box with a thin blue border is centered on the slide, containing the title text.

Faust Summary