

Unlocking Event-Driven Potential with Kafka: A Deep Dive



Marc J. Greenberg
Codemarc Consulting, LLC.

Requirements

expressed as MECE (*mutually exclusive, collectively exhausted*)

Book

- portfolio pdf's, searchable, annotatable
- customizable (cover page, table of contents)
- scalable (to 100s of docs, 1000s of pages)

Member

- update documents (all office doctypes + image, sound and movie)
- role based permissions (redact or remove)
- notification on update (by text, email, tweet, preference based)

System

- all artifacts geographically distributed, secure at rest and in flight (downloadable)
- event triggered - transpile/compile resilient, fault tolerant, scalable (to 100K/cpd)
- distributed log, observable, monitorable, instrumented

NFR

Composability

Useability

Securability

Scalability

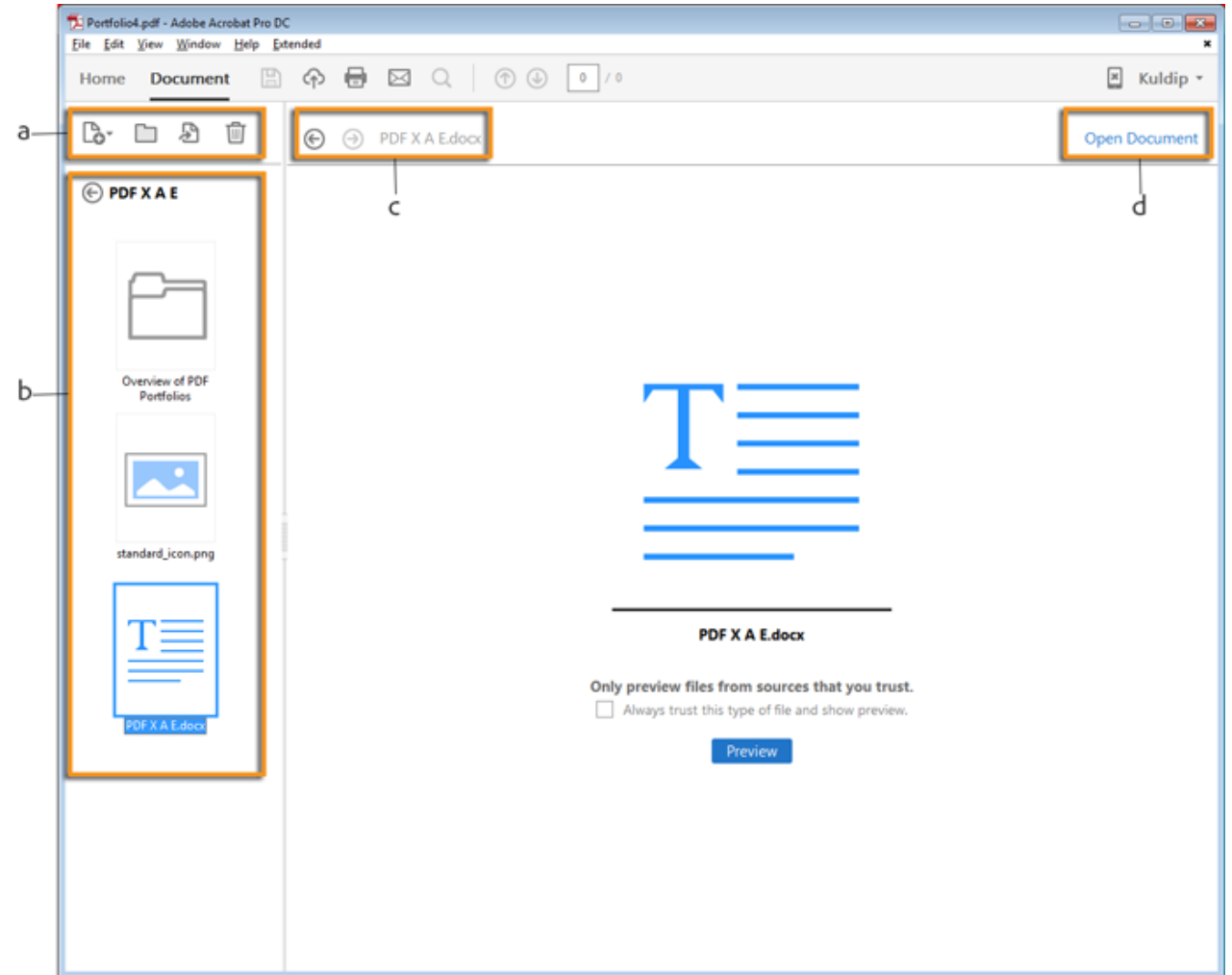
Reliability

Durability

Discovery – What is a Book?

PDF Portfolio is a collection of files that are gathered and saved into a PDF container.

1. Like a zip archive or an office doc (structured storage, embedded spreadsheets, images, inception)
2. Api's/Sdk's to create /secure /manipulate
 - [Adobe PDF Tools](#)
 - [jPDFProcess](#)
 - [Aspose.PDF](#)
 - [PDFTron](#)
 - [ITEXT](#)



EXPLORER

OPEN EDITORS

JS index.js

NODE [DEV CONTAINER: NODE.JS]

> .devcontainer

> node_modules

test

> in

> out

.gitignore

JS blank.js

JS index.js

{ } package-lock.json

{ } package.json

> NPM SCRIPTS

> OUTLINE

> TIMELINE

JS index.js

JS index.js > addCoverPage

```

1  const { PDFNet } = require('@pdftron/pdfnet-node')
2  const localizedFormat = require('dayjs/plugin/localizedFormat')
3  const dayjs = require('dayjs')
4  dayjs.extend(localizedFormat)
5
6  // Relative path to the folder containing test files.
7  const input_path = "test/in/";
8  const output_path = "test/out/";
9
10 const addPackage = async (doc, file, desc) => {
11   const files = await PDFNet.NameTree.create(doc, "EmbeddedFiles");
12   const fs = await PDFNet.FileSpec.create(doc, file, true);
13
14   files.put(file, await fs.getSDFObj());
15   fs.setDesc(desc);
16
17   const root = await doc.getRoot();

```

TERMINAL PORTS OUTPUT PROBLEMS

TERMINAL

1: bash

```
node → /workspaces/TheBook/node (master) $ time node index.js
```

```
PDFNet is running in demo mode.
```

```
Permission: write
```

```
Done.
```

```
Permission: read
```

```
Part: test/in/Annotation-Sync-Fafa.jpeg
```

```
Part: test/in/Board-Management-Software-Buyers-Guide.pdf
```

```
Part: test/in/Final-Build-a-board-book.mp4
```

```
Done.
```

```
real    0m1.254s
```

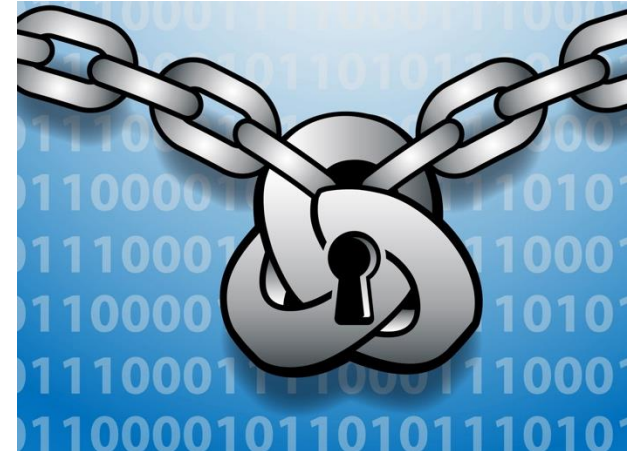
```
user    0m0.403s
```

```
sys     0m0.107s
```

```
node → /workspaces/TheBook/node (master) $
```

Discovery – What is a Member?

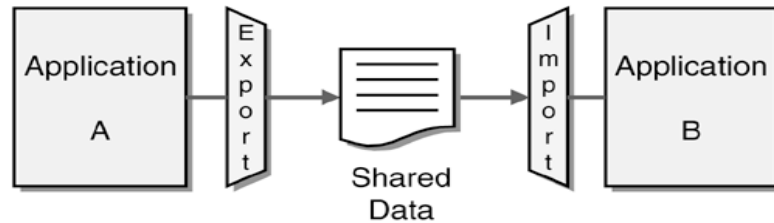
- A member of a collaboration team
- Identity
 - Authentication / Authorization
 - Privileges (granularity)
- Role Basis
 - Content Based Roles
 - Author
 - Editor
 - Reviewer
 - Calculated
 - Path based
 - Bitmapped (level1, level2, level3)
- Data Policies
- At Rest vs Inflight
- Retention
- Redact vs Remove
 - By page?
 - By item?



System - Integration Architectures

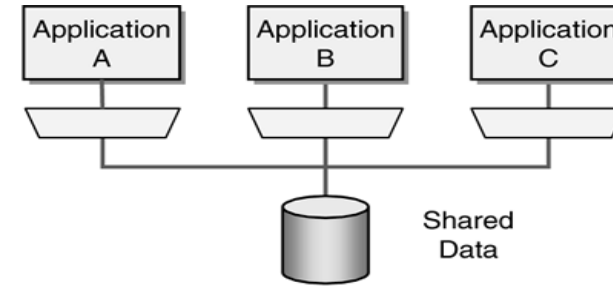
File Transfer

Applications share data by exchanging files. They use the export (producer) or import (consumer) patterns



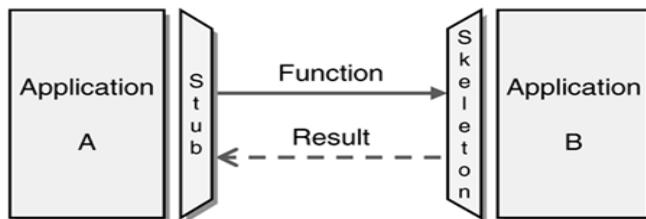
Shared Database

Applications store the data they wish to share in a common database



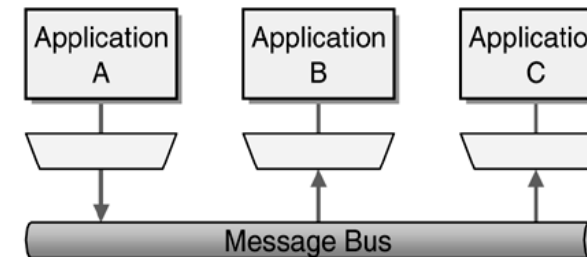
Remote Procedure

Applications share data by expose functionality that can be invoked remotely



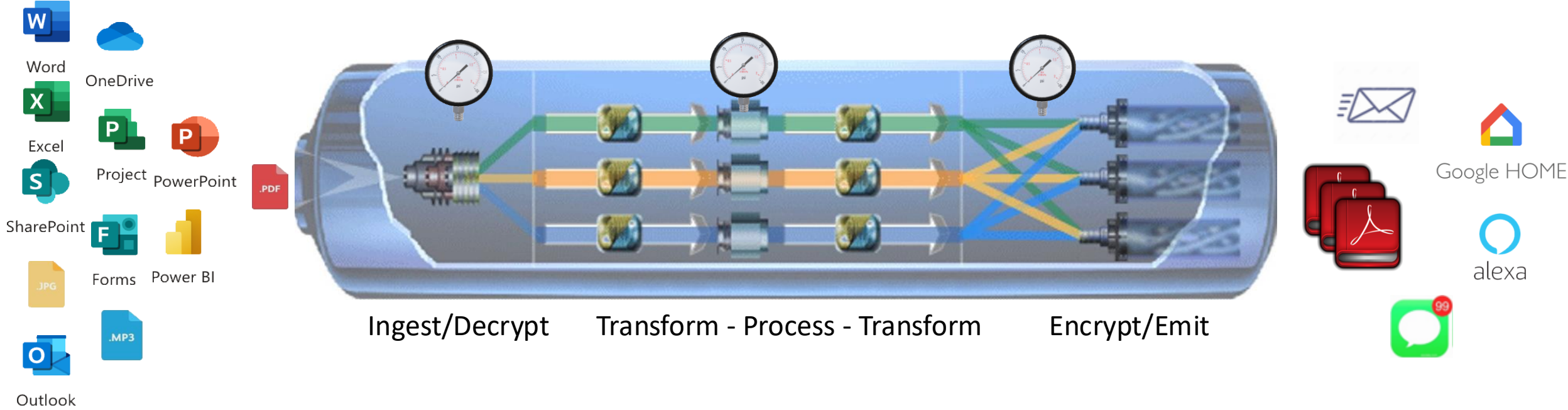
Message Passing

Applications share data by exchanging messages thru a common messaging system



Anatomy of a Message Flow

Observe-ability / Monitor-ability

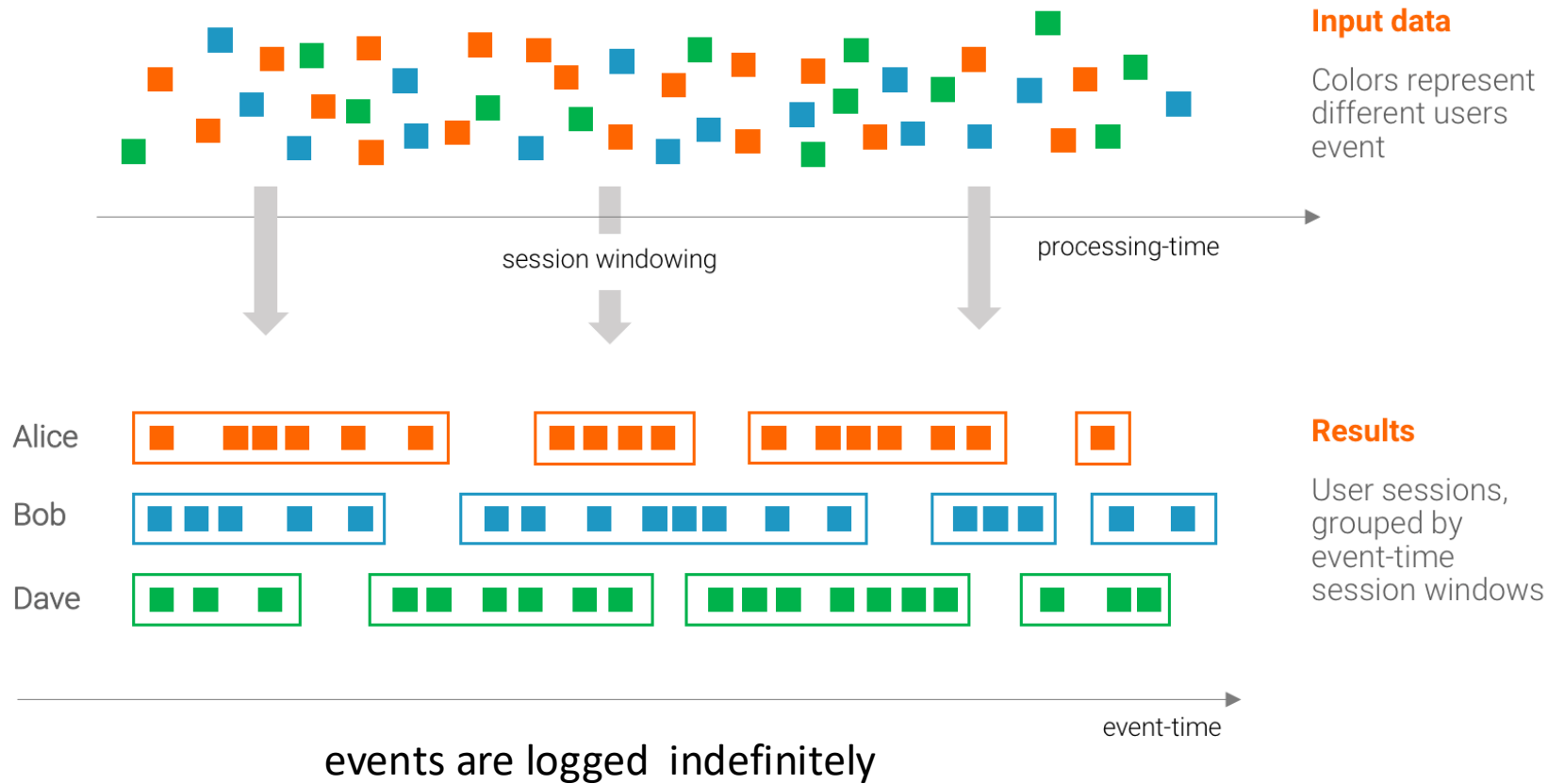


Enter Kafka



A single scalable platform to connect everyone to every event

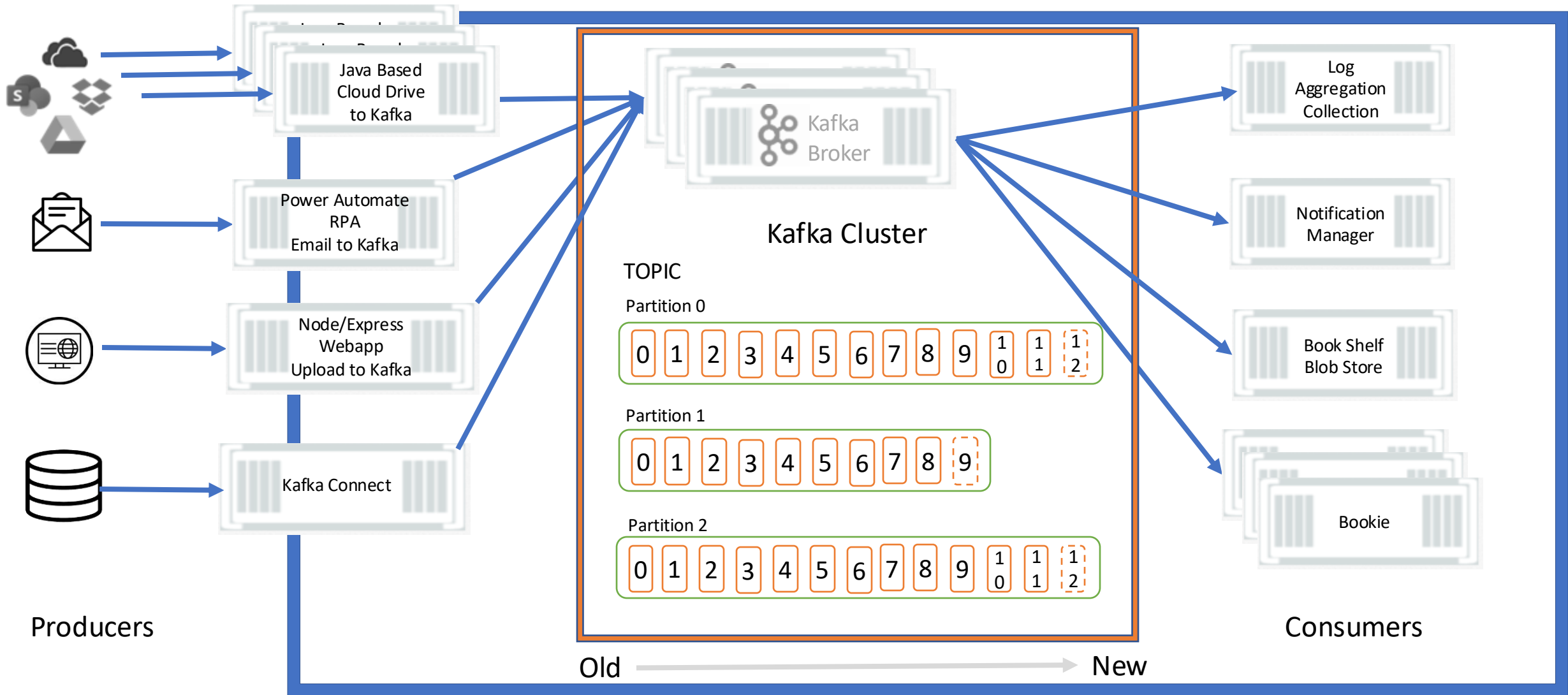
Real-time stream of instantly (read / action) –able events



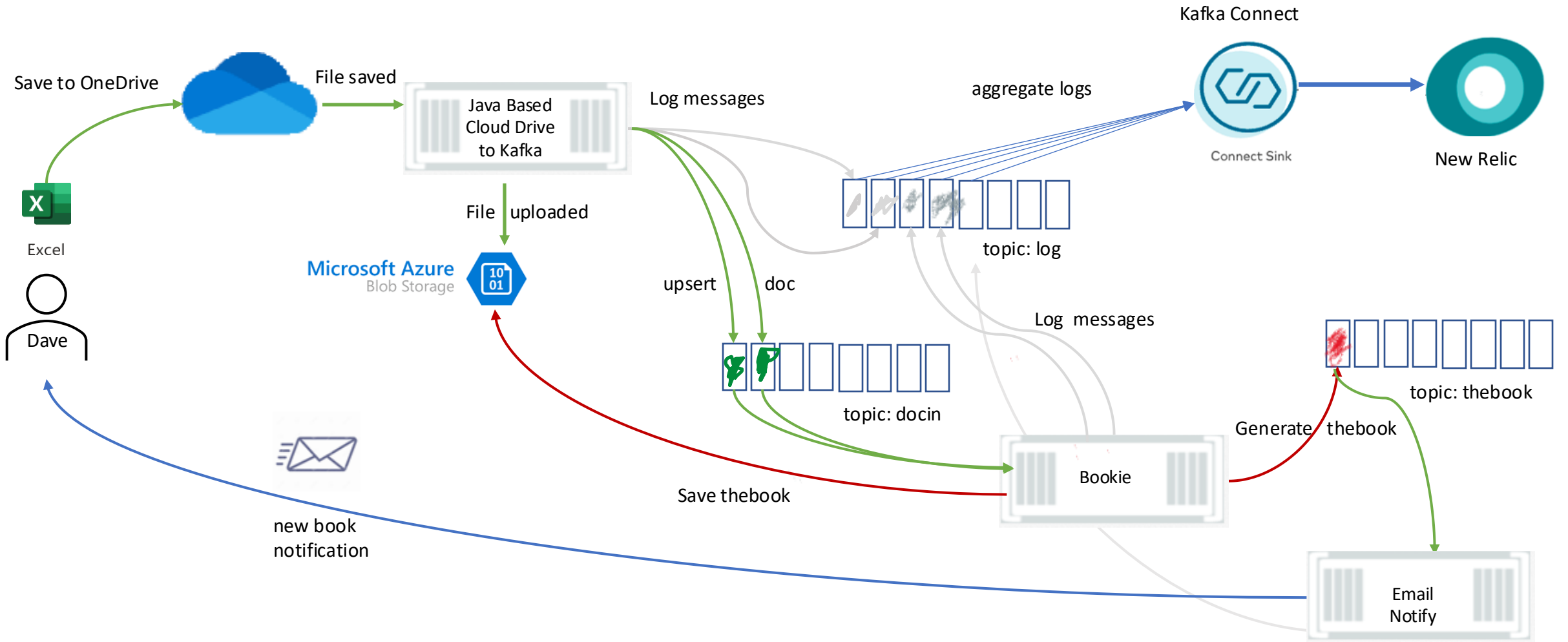
Kafka i

- events streaming platform
- open-source distributed
- publish-subscribe
- events stored in a log (topics)
- geographically distributed
- key/value pairs partitioned
- persistent record of events
- durable (written/replicated)

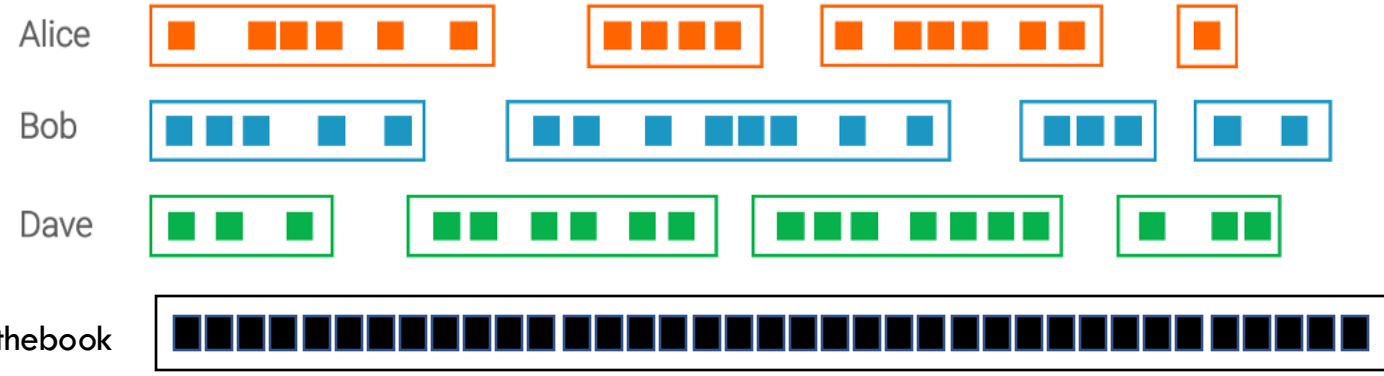




Now Apps - Topics



New Apps - Topics



ask onboard
about the book



Send me the
latest copy

The book for this month was updated 54 times, The last updated today at 11:01 am, 1 images was added by Bob. Other recent changes include "the marketing pdf" was updated by Alice and the "sales xls" was updated by Dave.

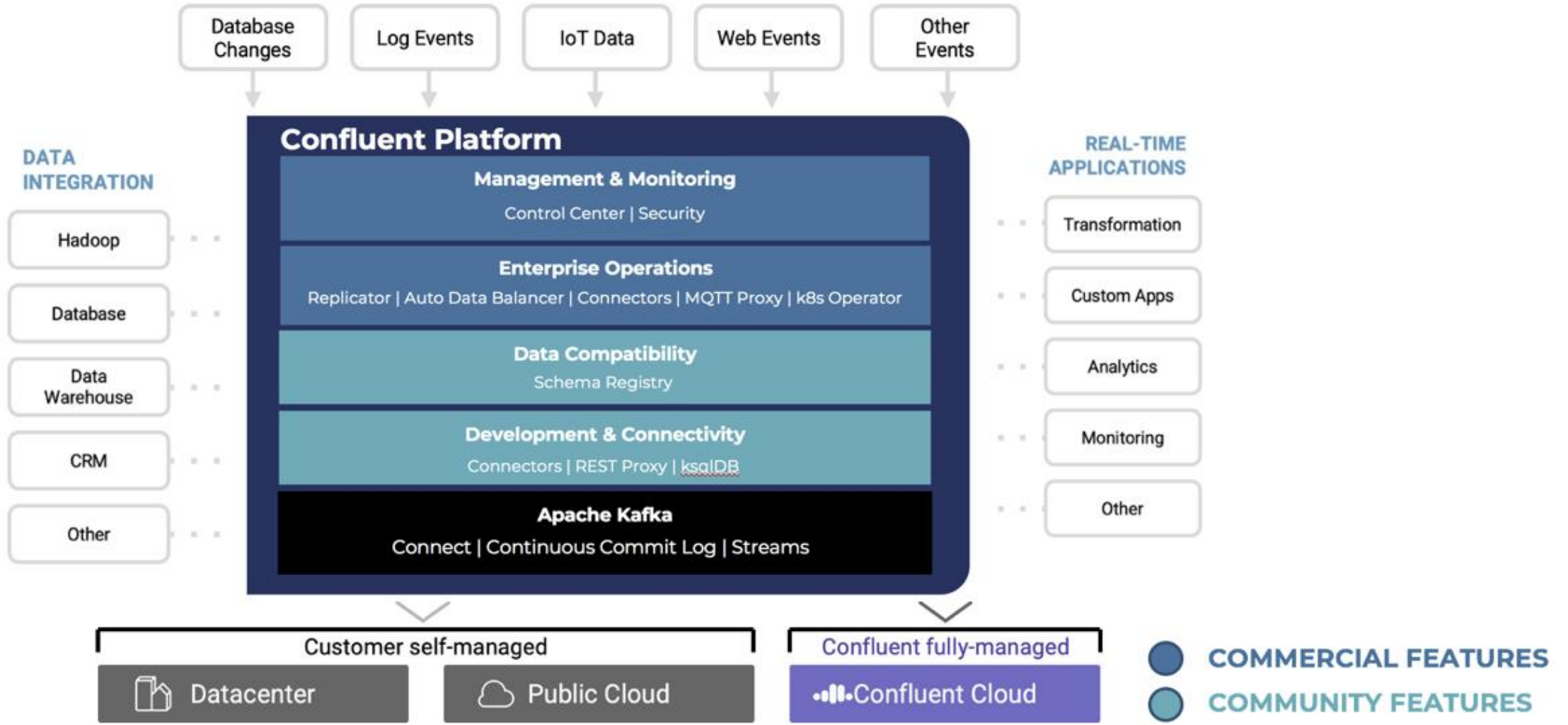


I am emailing you the latest version of "The Book"



Confluent Platform

A Complete Set of development, operations, and management capabilities to run Kafka at scale



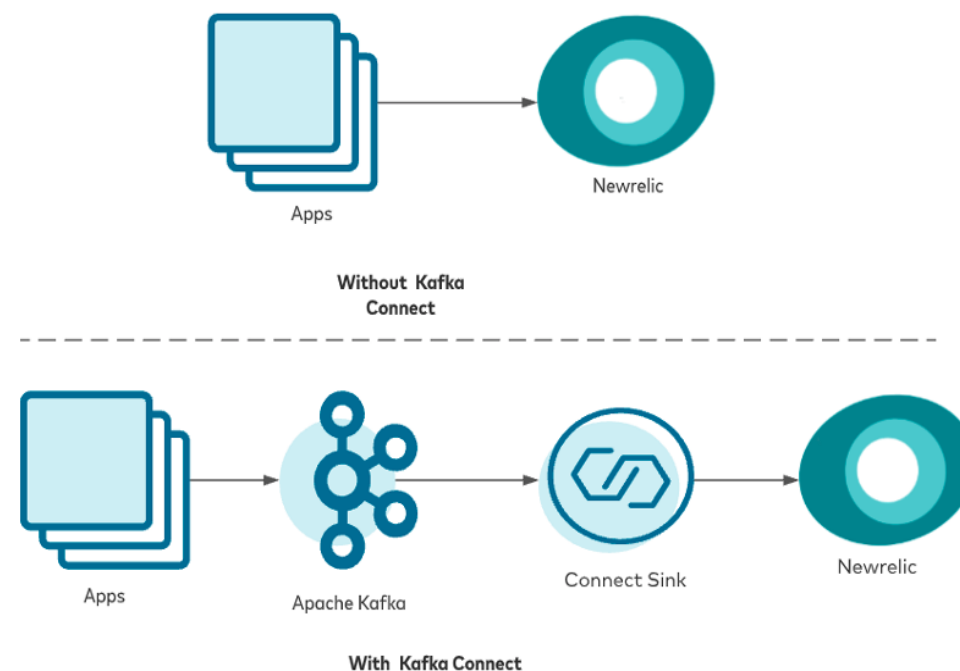
RabbitMQ vs Apache Kafka

Characteristic	Apache Kafka	RabbitMQ
Features	Distributed with guaranteed durability and availability, the data is shared and replicated.	It offers relatively less support for these features
Performance	100,000 messages/second	20,000 messages/second
Processing	reliable log distributed processing. Also, stream processing semantics built into the Kafka Streams	FIFO based, reading from the HEAD and processing 1 by 1.

Distributed Log, Observable, Monitored

New Relic APM

- SaaS focused on performance and availability monitoring
- Connects app and infrastructure perf
- Rich, detailed transaction data
- Realtime error analysis



References

- [What is Apache Kafka?](#) *Tim Berglund*
 - [Kafka 101 Events](#)
 - [Kafka 101 Topics](#)
 - [Kafka 101 Partitioning](#)
- [Designing Event Driven Systems](#) *Ben Stopford*
- [Enterprise Integration Patterns](#) *Hope and Bobby Woolf*
- [Technology Strategy Patterns](#) *Eben Hewitt*
- [PDFTRON samples](#)
- Links
 - [infoslack/awesome-kafka](#)
 - [Apache Confluence](#)

Imagine

- There is no beginning
- There is no ending
- Above us only Sky

