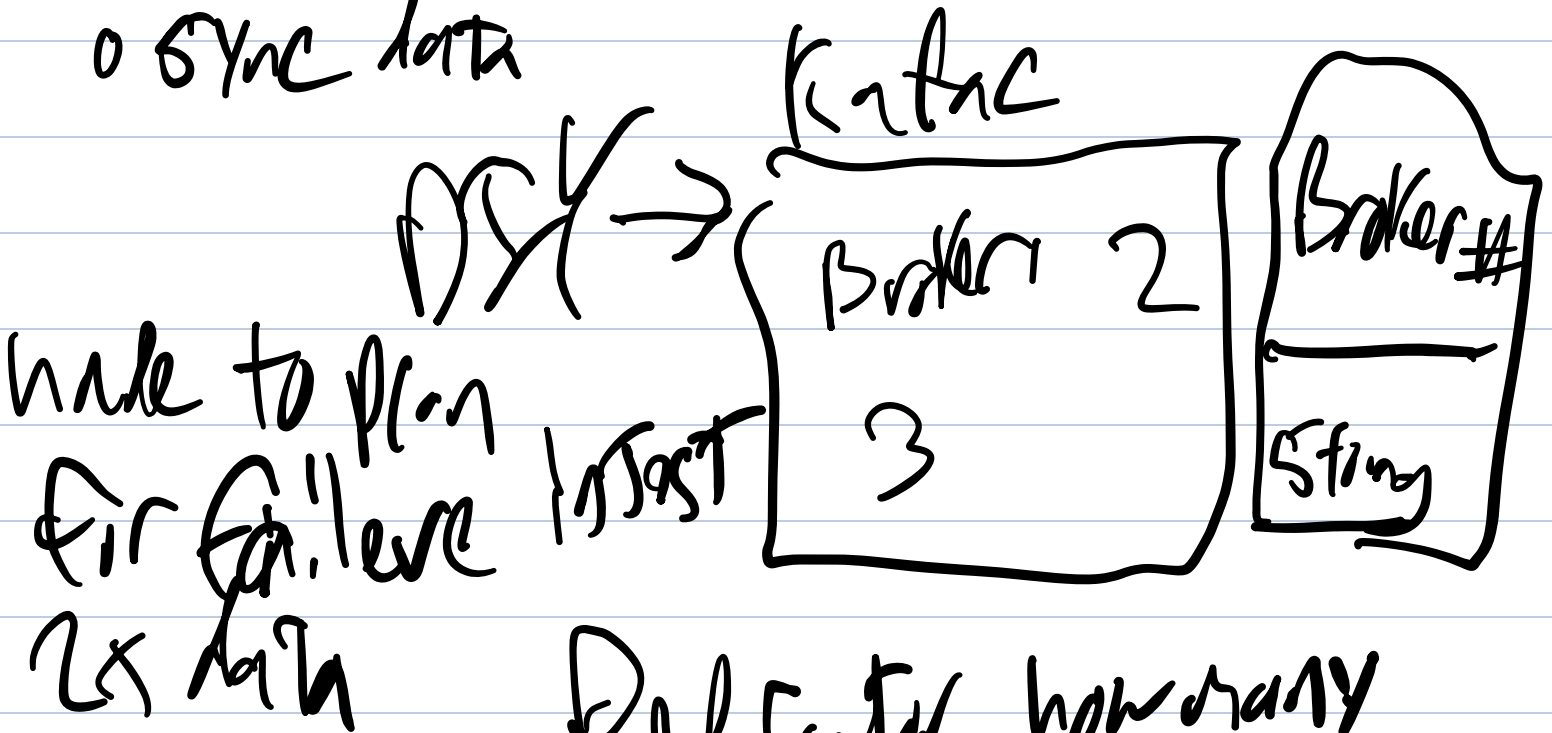


Kafka 12-2-24 ADV 342
flink in seamless twins

- o Capture $\rightarrow$ filter $\rightarrow$ activable
- o sync data



Follow
can't
be leader

Rep factor how many
replicas you have
Default 3 or more
Not great

Test old config

Control by Vol through that
Synth Cluster unstable

Scaling: Change Memory
Rolling restart

Control node last during
restart

Horiz Scale

Rebalancing takes hours

Don't move 20 Partitions
at a time

Use Cruise Control

AU5 H5K

eliminate storage unit
decoated & saved

* Express brokers

3 weeks ago Larch

Pay Fed on Sted

Carve 45MB

3x through Put over Sted
brokers

Scale out +
Scale up

No disruption on scale
up

Rebalancing takes
minutes

100% Faster Recovery from
failures

100% Kafka Compatible

Better Price/Perf at
any scale

Std	Express	Serverless
Manual car	Broker	1 click

fine tuned
control

Best Price
Perf
✱

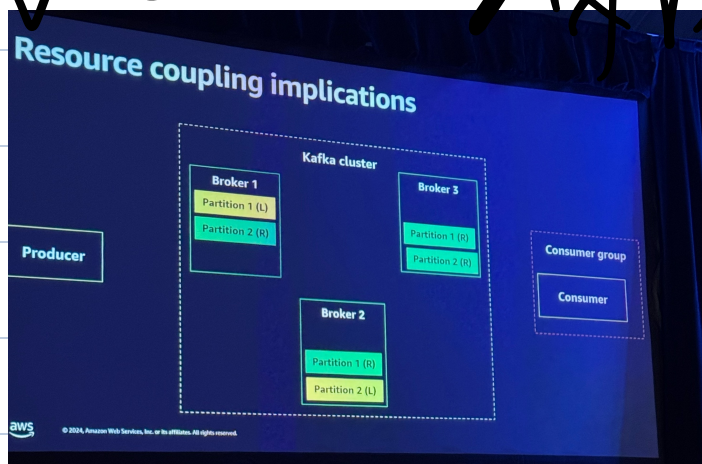
MSK Replicator

Multi vdc connector

Fink

DAO tools for

Streaming ✱ DataStream API



Apache Kafka is distributed; and stateful

Things on the cloud are better
You can now think about compute and storage availability separately with network-attached storage and benefit from the AZ model

But some things you need to still think about

- Network throughput quotas
- Storage throughput
- Storage capacity
- Compute to local storage network throughput

Scale compute and storage together

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitor all key dimensions to know when to scale

Storage volume
EBS
Can my disk sustain the write throughput?

Replication traffic out eats EC2 network bandwidth

Consumer traffic also costs EC2 network bandwidth

Legend:
 → Broker1 data in
 → Broker1 replication out
 → Broker1 replication in
 → Broker1 consumer out

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Achieve increased resiliency with MSK Replicator

RELIABLE AND AUTOMATED DATA REPLICATION FOR MSK CLUSTERS

Build robust streaming applications with cross-region resiliency with just a few clicks

Provide lower latency data access to different geographies

Replicate data on-demand with auto-scaling and pay only for what you use

No need to write code, manage infrastructure, or cross-region networking

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Horizontally scaling

Kafka cluster based on standard brokers

Rebalancing takes hours

- Reserve enough bandwidth for rebalancing
- Carefully orchestrate the movement so that it does not impact applications
- Continuous monitoring and change the plan as needed

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Choice between control and automation

More control ← ----- → More automation

Standard brokers for Amazon MSK	Express brokers for Amazon MSK	Amazon MSK Serverless
- Migrating from an existing Kafka setup - Need fine-grained control of Kafka - Know deeply how to manage Kafka	- Value performance and elasticity - Want/need lower control over Kafka - Operating Kafka estate at scale	- Quickly deploy and scale-up Kafka - Zero Kafka management - New-to-Kafka or streaming

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Express brokers eliminate storage management

Express brokers

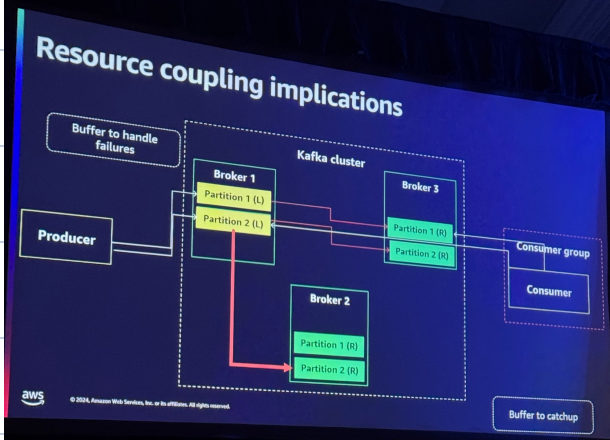
Do not have to manage disks (throughput or capacity)

Unlimited storage capacity per-broker

Pay per GB stored

MSK fully manages storage

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Express brokers deliver ...

Performance

Up to 3x throughput of Apache Kafka Brokers

	Sustained throughput							
M7g (Mbps)	Large	Xlarge	2xlarge	4xlarge	8xlarge	12xlarge	16xlarge	
Standard brokers	12	24	48	96	153.85	153.85	153.85	
Express brokers	15.62	31.25	62.50	125	250	375	500	

© 2014, Amazon Web Services, Inc. or its affiliates. All rights reserved.

FinK Continued