Tech Exercise

Eero Tuomikoski

EVALUATION & KEY SOLUTIONS

Customer has multiple data sources and consumers. Data modifed on the fly

- Recommendation to use Kafka streaming platform, e.g. for short retention time (7d)

Customer's data sources and consumers are located in two separate public clouds

- Recommendation to connect Kafka's with Kafka Mirrormaker 2.0 for disaster recovery & scaling
- Use VPC peering to connect two public clouds

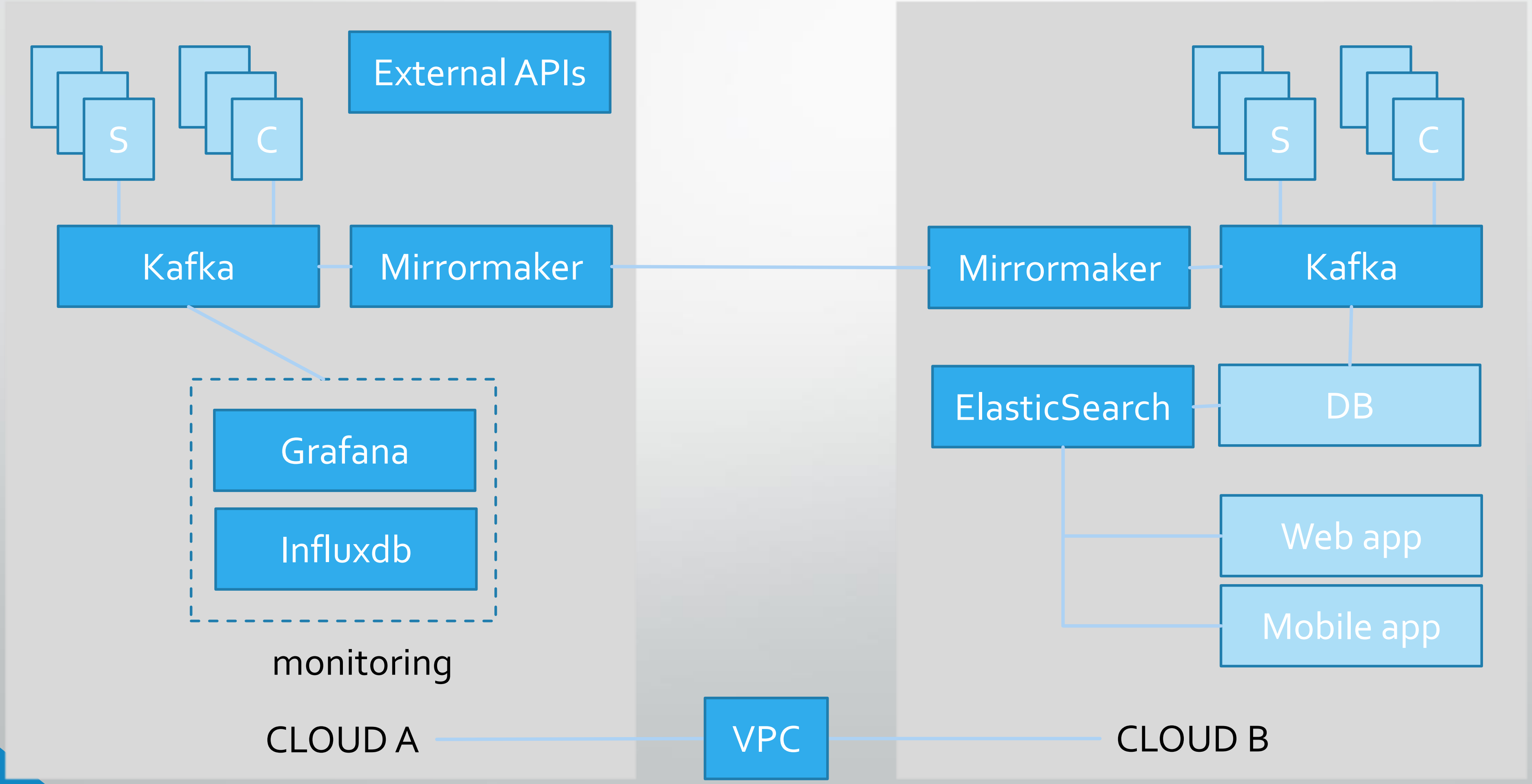
Customer needs insight to system ops & metrics

- Recommendation to use Grafana for monitoring & trending with some data store e.g. influxdb

Customer's mobile & web app backend database is struggling under load

- Recommendation to use Elasticsearch to provide near-realtime search capabilies and utilize it as mobile & web app backend

ARCHITECTURE



SCALING & COST ESTIMATE

3 MB/s peak rate data production -> 7d = 1,8 TB -> estimated max. storage 1 TB.

Kafka (business-8, one Google, one Azure):	2500 $/m
Kafka Mirrormaker 2.0 (business-8, one Google, one Azure)	1500 $/m
Elasticsearch for web & mobile app backend (business-16)	1640 $/m
Grafana monitoring (startup-8)	180 $/m
Influx monitoring (startup-28)	490 $/m

	6310 $/m
Support tier: Business	1000 $/m

CYBERSEC

- Ensure your development process has cybersecurity checkpoints/testing.
- Estimate your cyber security posture regularly e.g. with 3rd party pen-testing.
- Monitor security events in your system.
- Use encryption to protect your data at rest/in transit.
- Use VPC peering to connect public clouds. With VPC peering, virtual machines in different VPCs can connect to each other without going over public internet
- Protect public internet facing interfaces (e.g. REST APIs) with strong security. With API gateway, or e.g. OAuth for 3rd party access, using data encryption e.g. TLS, using quotas & throttling. Monitor your interfaces.