Scalable IoT-data pipeline with Apache Flink® @ AWS

Mechatronic Circus & Demo Day 2021

Miika Valtonen
CTO, D.Sc. (Tech.)
miika.valtonen@remion.com

REMION

IoT-data pipeline

Deployed with CDK in AWS

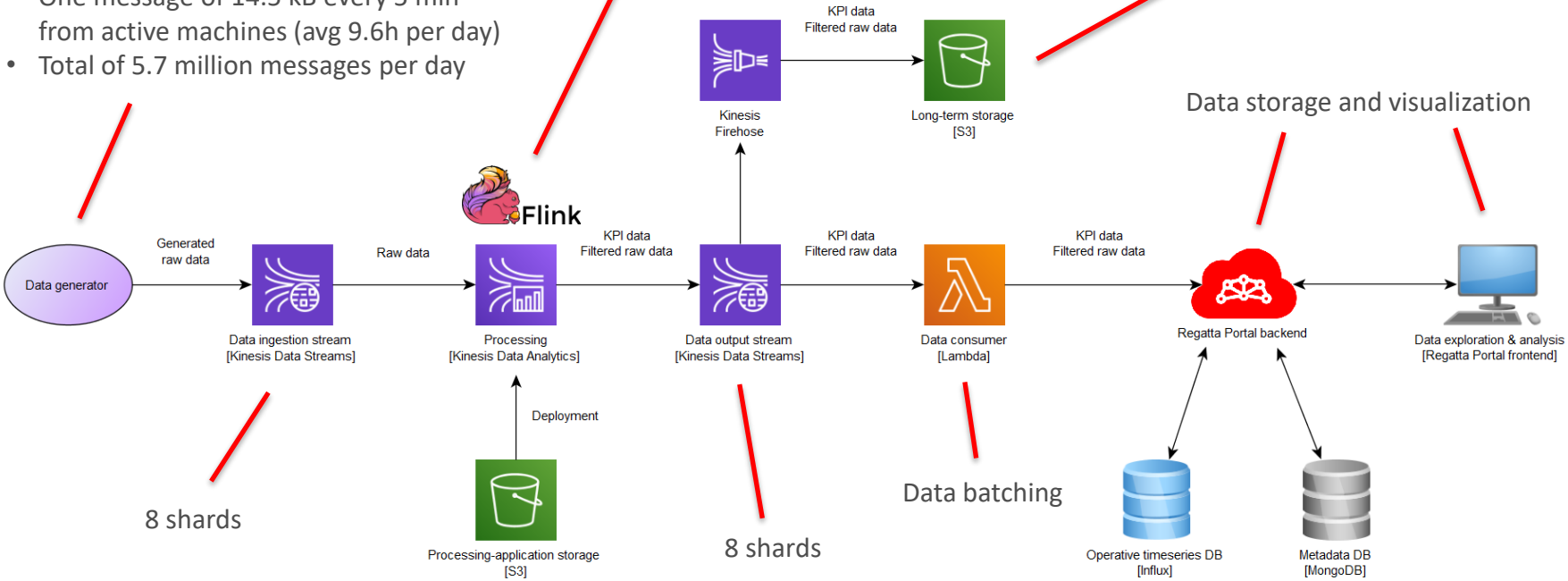
Verbose metrics & logs

- 50000 machines of 3 different types
- 12 realistic signals
- One message of 14.5 kB every 5 min from active machines (avg 9.6h per day)
- Total of 5.7 million messages per day

- Realtime processing
- 32 parallel vCPUs w/ auto-scaling
- Calculates 8 new KPIs

- Long-term storage
- Stream monitoring

Data storage and visualization



8 shards

8 shards

Data batching

For more information on Apache Flink® see https://flink.apache.org/.

REGATTA® A00000001

Signals | Map | Hourly KPIs | Daily data | Realtime

Bulk Actions | + Create signal | Columns

Filtered raw data

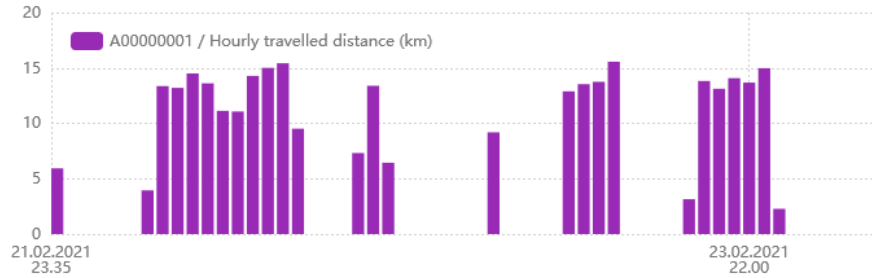
Icon	Description	Variable name	Value	Unit
	GPS latitude	gps_latitude	-34.063687194808935	°
	GPS latitude	gps_longitude	88.18990923650111	°
	Speed	speed_kph	0.00	km/h
	Runtime	runtime_secs	219358.0	s
	Lift count	lift_count	1042	#
	Machine state	machine_state	IDLING	
	Digital input 1	digital_input_1	0.0	
	Digital input 2	digital_input_2	1.0	
	Digital input 3	digital_input_3	1.0	
	Oil temperature	oil_temperature_c	12.00	°C
	Fuel used	fuel_used_litres	12489.3	L
	Derived digital input	derived_digital_input	1.0	

KPIs

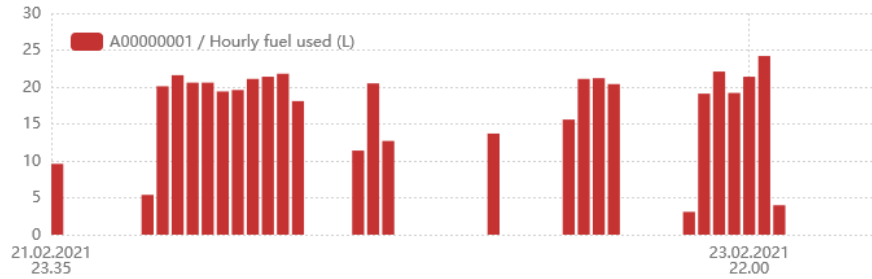
	Hourly travelled distance	hourly_travelled_distance_km	0.81	km
	Hourly fuel used	hourly_fuel_used_litres	0.90	L
	Hourly runtime	hourly_runtime_secs	00:04	h
	Hourly runtime moving laden	hourly_runtime_moving_laden_secs	00:01	h
	Hourly runtime moving empty	hourly_runtime_moving_empty_secs	00:00	h
	Hourly runtime idling	hourly_runtime_idling_secs	00:03	h
	Hourly derived digital input active time	hourly_derived_digital_input_active_time_secs	00:04	h
	Hourly lift count	hourly_lift_count	1	#

Hourly runtime by state (6+d*)

Hourly travelled distance (6+d*)



Hourly fuel used (6+d*)



Scalability

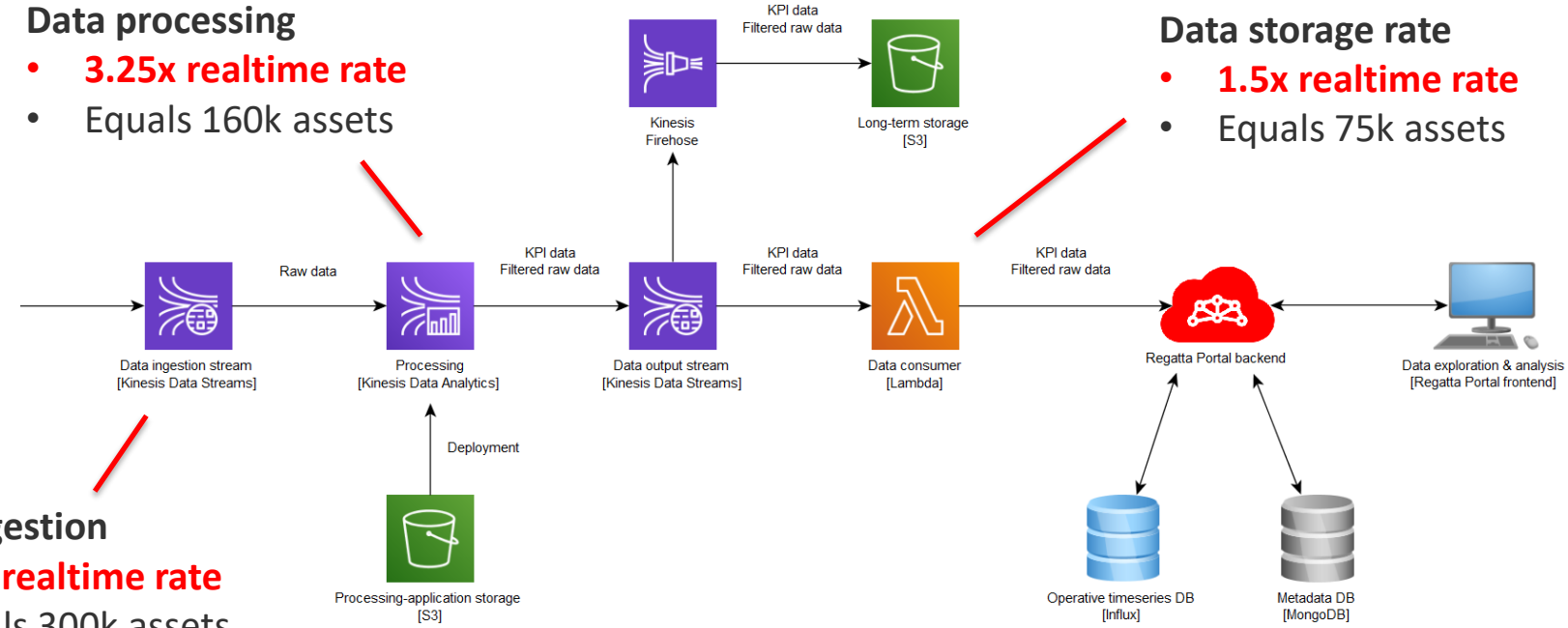
24h continuous stress test with 50000 assets and 20 signals

Data processing

- **3.25x realtime rate**
- Equals 160k assets

Data storage rate

- **1.5x realtime rate**
- Equals 75k assets



Data ingestion

- **6.2x realtime rate**
- Equals 300k assets

⇒ The demo architecture is suitable for **processing realistic and realtime IIoT data**

⇒ It could be **optimized** and **scaled up** in AWS

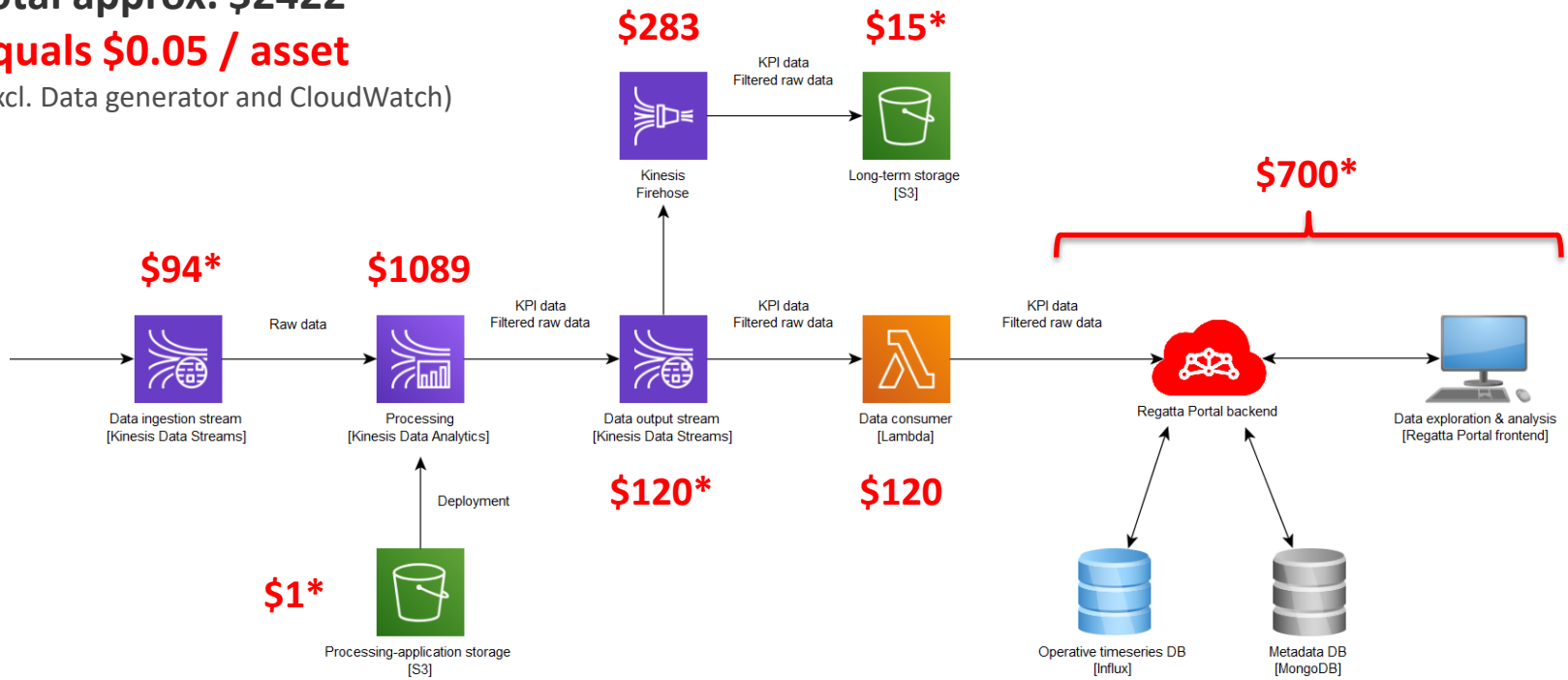
Operating costs

Monthly costs per service (excl. tax)

Total approx. $2422

Equals $0.05 / asset

(excl. Data generator and CloudWatch)



* Approximation (cost class split between multiple services)

Conclusion

Strengths

- Flink provides a **strong Java/Scala –based framework and platform** for processing realtime data at scale
- Kinesis Data Analytics (KDA) **eases Flink comissioning and deployment**
- KDA **operating-costs seem reasonable** and many optimizations can be done

Weaknesses ✓

- **An analytical mindset with strong coding skills** are required to master IIoT Flink applications
- **Best practices** for industrial IoT-data processing are not generally available

➡ We are developing **Regatta® Framework for Flink** to ease its adoption**:**
Data processing and storage libraries, templates, schemas, data formats, reprocessing management, metrics, integration testing, IaC deployments in AWS, ...

REMION

Trusted Industrial IoT Ecosystem Partner

150+
IOT PROJECTS SINCE 2003

BILLIONS OF EUR
WORTH MACHINES AND DEVICES MONITORED

Selected Customers

KONGSBERG

SANDVIK

KONECRANES®

wind controller

MANTSINEN