Harvesting Big Data

Supporting scalable telemetry data management for autonomous machines in agriculture

Angus Ross

4 million
acres farmed

200,000
Field hours

100+
Commercially running Machines

Lots of Data

Weather
Temperature
Wind
Location

Usage
Engine hours
Fuel consumption
Hectares

Operation
Working
Idle
Sleep

Job
Progress
Spray records
Weed maps

Logging
Events
State changes
Journal logs

Lots of Data

Weather
200,000
Rows a day

Events
250,000
Rows a day

Total
~ 2.5 billion
Rows

Usage metrics
60,000
Rows a day

Logging
40 million
Rows a day

What, When, Where, Why

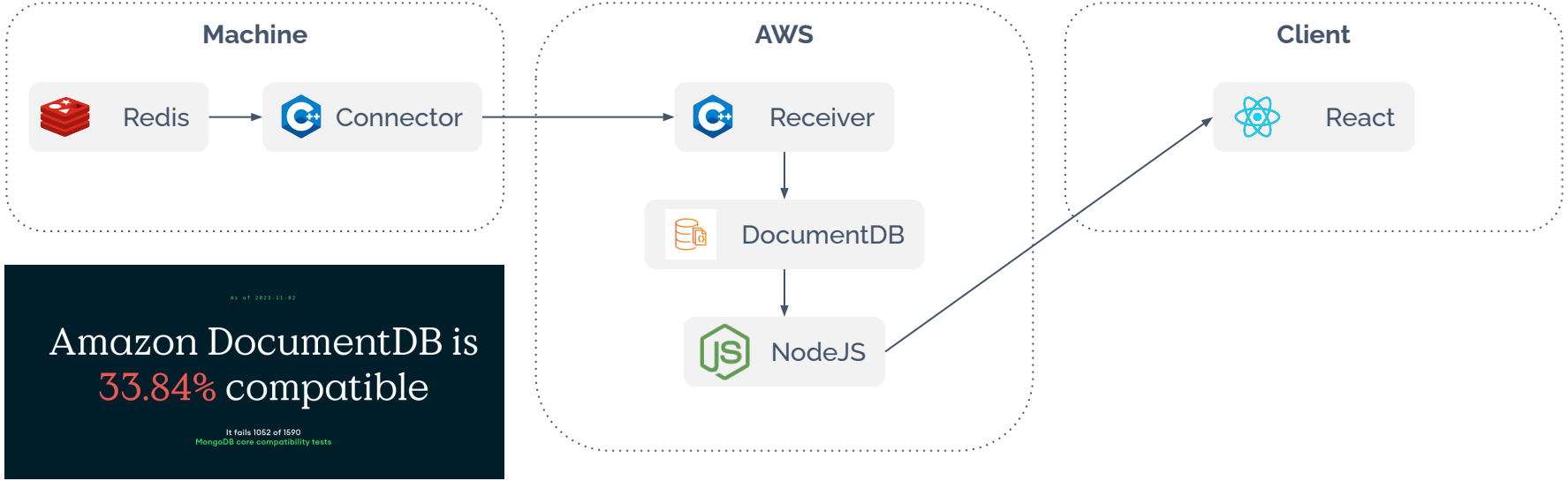
Customers

- Accurate application
- Spray Records
- Efficiency
- Uptime reports

Internal

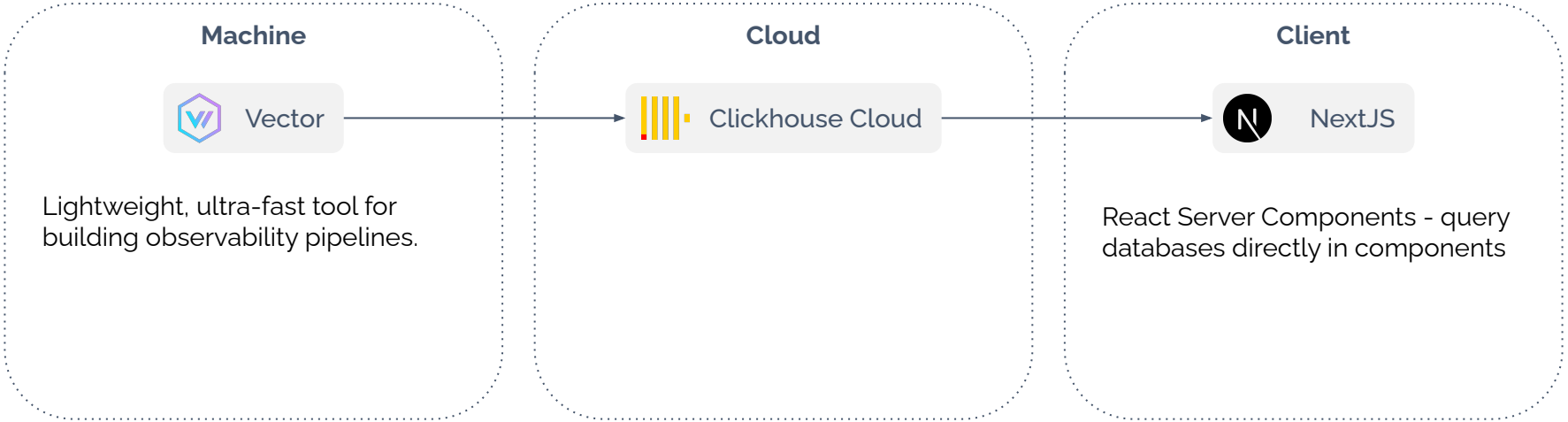
- What went wrong
- Software updates producing increase in productivity

Pipeline - Previous



https://www.isdocumentdbreallymongodb.com

Pipeline - Current



Pipeline - Current

Machine

```
sources:
  clickhouse_source:
    address: ...
    encoding: json
    strict_path: false
    type: http_server
transforms:
  modify_path:
    type: remap
    inputs:
    - clickhouse_source
    source: .path = replace(.path,"/","") ?? ""
sinks:
  clickhouse:
    auth: ...
    date_time_best_effort: true
    skip_unknown_fields: true
    database: default
    endpoint: https://clickhouse.cloud
    inputs:
    - modify_path
    table: "{{.path}}"
    type: clickhouse
    batch: ...
    buffer: ...
```

Cloud

```
CREATE TABLE default.metrics
(
    `id` UUID,
    `version` UInt8,
    `swarmbotID` String,
    `date` DateTime64(3),
    `insertedAtDate` DateTime DEFAULT now(),
    `location` Array(Float64),
    `fuelUsed` Decimal(18, 12) DEFAULT '0',
    `engineLoadAvg` Decimal(12, 0) DEFAULT '0'
    ...
)
ENGINE MergeTree()
ORDER BY date
```

Client

```
import { createClient } from "@clickhouse/client";

const client = createClient({
    url: process.env.CLICKHOUSE_HOST,
    username: process.env.CLICKHOUSE_USER,
    password: process.env.CLICKHOUSE_PASSWORD,
    database: process.env.CLICKHOUSE_DATABASE,
    compression: {
        request: true,
        response: true,
    },
});
export async function getMetrics(swarmbotID: string) {
    const response = await client.query({
        query: `SELECT * FROM "metrics"
            WHERE swarmbotID = {swarmbotID:String}
            GROUP BY date
            ORDER BY date;`,
        format: "JSONEachRow",
        query_params: {
            swarmbotID,
        },
    });
    const result = await response.json();
    return result;
}
```

✓ Weather checks enabled

Start Aug 26, 2024 11:37 AM

End Aug 26, 2024 4:21 PM

4.74hr 17 hours ago

Code Versions `10.4.8`

Product

Name	2,4D/Gly/DRA
Rate of Product	10L/100L
Concentration	100L/ha

Application

Boom/Nozzle Type	Hayes 24m / Test
Crop Type	Fallow
Target Weeds	Feather top rhodes grass/ milk thistle

Operator

Name	
Contact	
Qualifications	

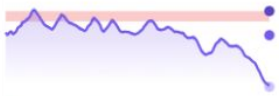
Temperature

Max 34 °C
Min -1 °C

Spray Window 100%

Delta T

Max 11 °C
Min 0 °C



Spray Window 84%

Wind

Max 22 km/h
Min 0 km/h

Spray Window 100%

Wind Gusts

Max 29 km/h

Spray Window 100%

Combined Spray Window — In window — Outside window

Operating Hours — Working — Travel — Weather — Sleep

Acheived Spray Window: 90%

Engine

Engine Hours	4.37 hrs
Fuel Used	33.10 L
Average Fuel Rate	8.38 L/hr

Job

Hectares	60.40 Ha
Hectares per Hour	16.68 Ha/hr
Average Speed	7.89 km/hr

Speed

- Weather observations
- Metrics
- Spray Records

- 3 tables
- 3 Inner joins
- Lots of arithmetic

```
Elapsed: 1.087s    Read: 39,045,316 rows (3.85 GB)
```

Speed

Finding what events are stopping our swarmbots.

- Window Functions - leadInFrame
- Inner joins

Search results... Elapsed: 0.810s Read: 6,361,702 rows (418.88 MB)

#	eventMessage	total
1	The SwarmBot is at the GoTo location and paused	732.4709002777778
2	A boom position error has occured - Tap for details	499.3484625000001
3	All jobs are completed	457.15310305555556
4	Internal SwarmBot error - SwarmBot location has timed out	450.7798213888889
5	Waiting for the green area to load	422.8725825000001
6	The fuel tank is empty	409.4012333333334
7	Could not load the override areas	396.8370525
8	Waiting to determine the type of attachment	329.7893319444444
9	Another SwarmBot is too close	303.19906888888886
10	Warming hydraulic oil by sending oil over the relief valve on rear manifold	276.4738325

Compression

compressed_size	uncompressed_size	compression_ratio	compression_percentage
26.27 GiB	417.84 GiB	0.063	93.712
8.85 GiB	207.93 GiB	0.043	95.745
5.83 GiB	58.89 GiB	0.099	90.097
4.42 GiB	113.95 GiB	0.039	96.122
3.67 GiB	65.78 GiB	0.056	94.428
3.58 GiB	75.46 GiB	0.047	95.261
2.02 GiB	9.87 GiB	0.205	79.513
2.06 GiB	106.67 GiB	0.019	98.073
1.65 GiB	25.84 GiB	0.064	93.606
1.39 GiB	31.73 GiB	0.044	95.615
1.41 GiB	43.48 GiB	0.032	96.757

Pros

- Documentation, Articles, Blog posts
  - Lots of examples
- Support
  - Very knowledgeable
  - Quick to respond
- Community
  - Events, Training
- Console
  - Very easy to test queries
- Simple and fast by default

Cons

- Console AI
  - Not very helpful debugging
- Identifying slow queries
  - Mongo has lots of analytic dashboards around identifying problem queries

What Next

- Open Telemetry
- Intensive location data - Weedmaps
- See if chdb is useful for us
- Vector - More observability (size of buffer, avg time in buffer etc)

Questions

www.swarmfarm.com

SWARM FARM ROBOTIC AGRICULTURE