ClickHouse as a High-Performance Monitoring System, from zero to hero

DevOpsAI SG 2025

ClickHouse

24.03.26

Welcome!

Agenda

01 Introduction

02 Pre-Requsites

03 ClickHouse Fundamentals

04 Data Modeling - basics

05 Python - data generator

06 Visualization - Grafana

07 Python ML

08 FAQ

09

10

11

12

STAND A CHANCE TO WIN AN R2-D2 LEGO SET!

Scan the QR code and put your name in for the lucky draw!

Meet us at the ClickHouse-sponsored coffee cart during afternoon tea break!

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

01 Introduction

ClickHouse

Who am I ???

ClickHouse

The definition of observability

Definition (software)

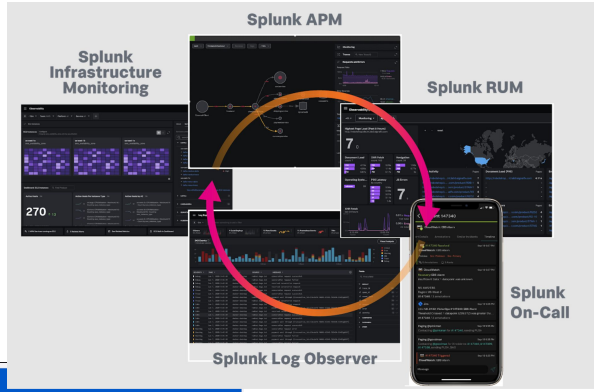
“In software engineering, more specifically in distributed computing, observability is the **ability to collect data** about programs' execution, modules' internal states, and the communication among components. To improve observability, software engineers use a wide range of **logging and tracing techniques** to gather telemetry information, and **tools to analyze and use it**. Observability is **foundational** to **site reliability engineering**, as it is the first step in **triaging a service outage**. One of the goals of observability is to minimize the amount of prior knowledge needed to debug an issue”

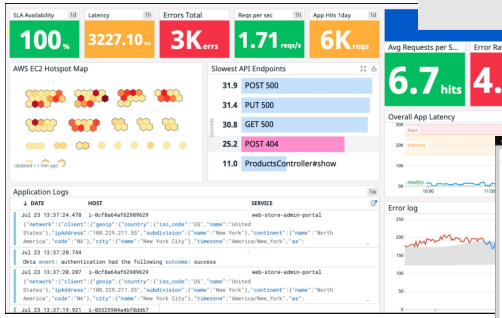
Quoted from *https://en.wikipedia.org/wiki/Observability_(software)*

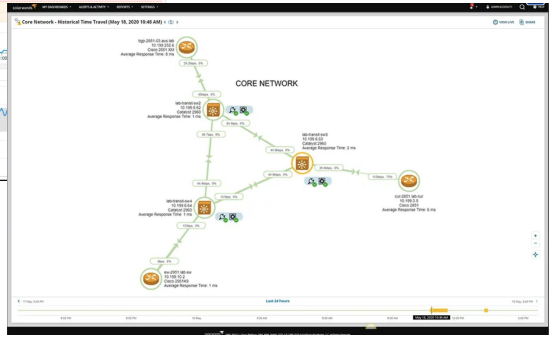
What is happening in the field?

- There are many observability platforms (data dog, splunk, dyntrace, sumologic), aren't they enough for covering everything?

- **Detour**:
  - costs,
  - transparency,
  - speed (performance),
  - complexity vs flexibility





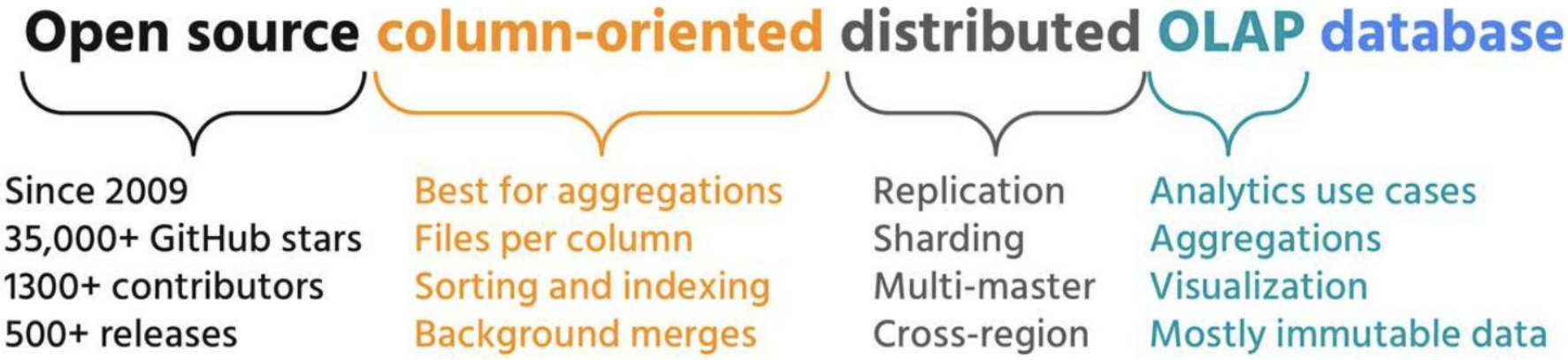


Yet another product / platform in the field...

ClickHouse - new kid on the block

- Costs - we might not be the cheapest but definitely not expensive (price calculator)
- Transparency - our pricing is transparent, the data stored is totally visible in table format (SQL), no middleware in between, can fit into workflows through connectors
- Performance - we are fast (ClickBench) in both ingestion and retrieval; storage is efficient (lz4 by default, also available for zstd)
- Complexity - check how Splunk and others' UI were, ClickHouse is simply SQL and tables which are developer friendly (previous slide)

What is ClickHouse actually?



Objective of the workshop

Illustrate what a typical (simplified) observability workflow looks like and how ClickHouse fits into the flow.

Throughout the journey:

- Understand the basics of ClickHouse (e.g. Schema, Materialized Views)
- Extensibility through integrations (e.g. with Python client and Grafana)

02 Pre Requisites

ClickHouse

Tools before you start (15 min)

Instructions (for Mac only) available at https://shorturl.at/BO0HO

- Docker Desktop
- Docker Compose (for older version of Docker Desktop that does not come with `compose`)
- (optional) Git
- (optional) Python3
- A modern web browser (chrome, firefox)

And...

We are all GOOD !

ClickHouse fundamentals

ClickHouse

Understanding Columnar Storage

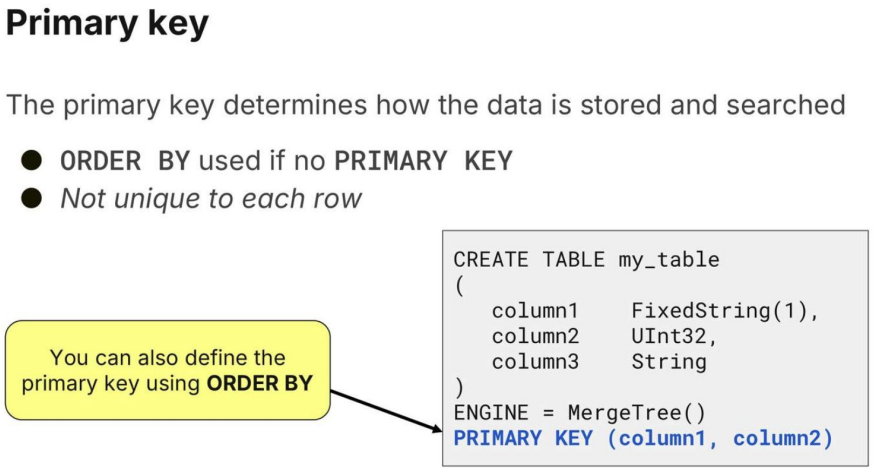
- One reason ClickHouse is so fast is because it is *column-oriented*
- But you need to understand columnar storage to get the benefit

So what does it mean?

Concepts

- In ClickHouse, per table required a table engine; typically we will be using MergeTree(s) which is optimized for high performance inserts and queries (should be good enough for most cases)

- Primary key is different...



Concepts (cont’)

- Data are retrieved in Granule(s)

  - **granule** – a logical breakdown of rows inside an uncompressed block; default is 8,192 rows
  - **primary key** – the sort order of a table
  - **primary index** – an in-memory index containing the values of the primary keys of the first row of each granule
  - **part** – a folder of files consisting of the column files and index file of a subset of a table’s data

A *granule* is the smallest indivisible amount of data that ClickHouse reads when searching rows

Data Modeling - Basics

ClickHouse

Schema

- In ClickHouse, table Schema (aka DDL) is important.
- It is a big-data oriented SQL db (not **noSQL** db) check it out on Elasticsearch too!
- It is possible to alter the existing columns but have to pay a price; it is done through `mutation`
- Instead of updating the schema, we could consider Materialized views

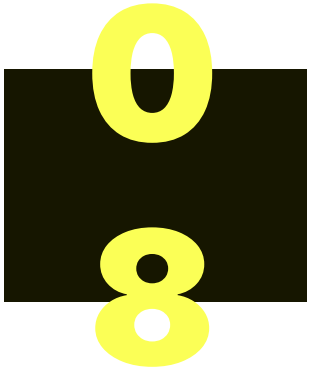
Materialized View

Use cases:

- If your source / raw data required to be processed (e.g. extraction of certain fields from a json data structure)
- If the existing table's Index (sorting-key / primary-key) does not meet a use case's requirements
- To filter out data from the source table into dedicated tables (e.g. forward a data row based on the `region` field's value, each region would have its own table)
- To consolidate various source tables' data into a single target table

Materialized View

- General link : https://shorturl.at/o94Oy
- Refreshable MV : https://shorturl.at/SbbPw
- Incremental MV : https://shorturl.at/IxJRy



FAQ

ClickHouse

Who am I ???

ClickHouse

STAND A CHANCE TO WIN AN R2-D2 LEGO SET!

Scan the QR code and put your name in for the lucky draw!

Meet us at the ClickHouse-sponsored coffee cart during afternoon tea break!

Thank you!

A single image on a slide

- Slide with a single line of code

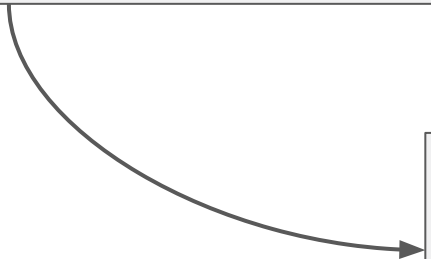
```
code_block
```

- Slide with a multi-line of code

```
multiline codeblock
```

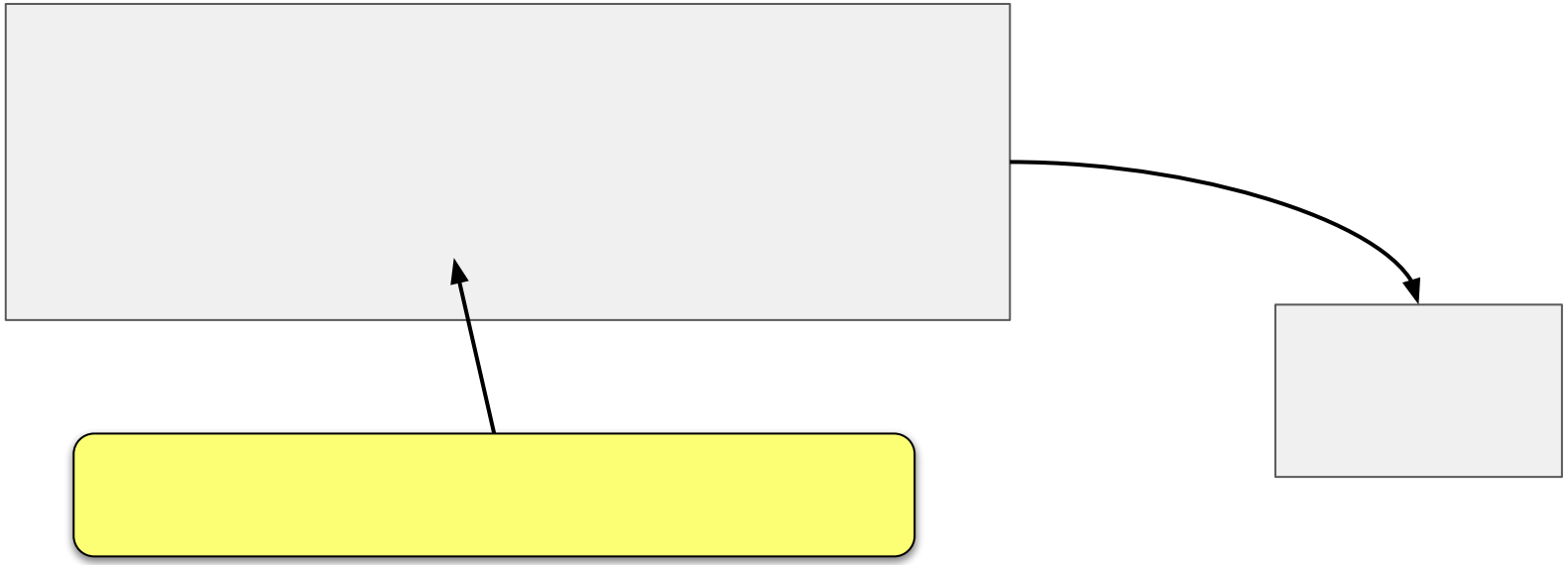
- Slide with a multi-line of code and an result

```
multiline codeblock
```



```
multiline codeblock
```

- Code, result, callout

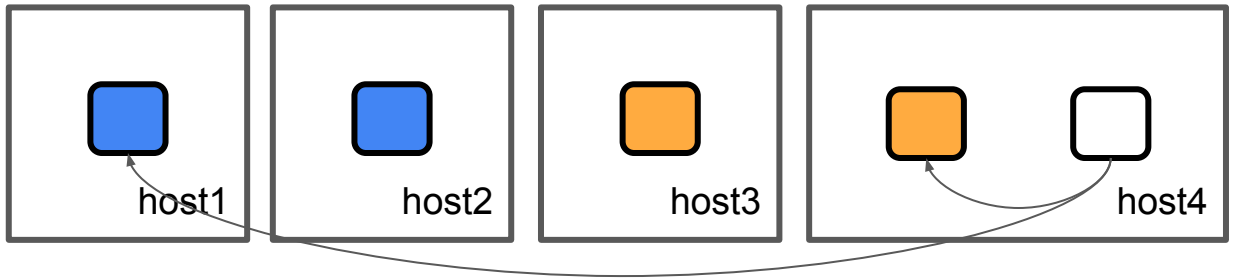


A nice transition slide for asking a question or making a point

- Slide with a table:

- Table with row and column headers

- shards and replicas



Let's not use camel case for this type of slide

Assets

centered callout

```
codeblock
```

```
multiline codeblock
```

Callout with a strong contrast

Fonts and Colors

- Normal font is "**Inter**"
- Code font is "**Roboto Mono**" size 14
  - unless it must be shrunk to fit on the slide
-

```
CREATE TABLE my_table
(
   column1    FixedString(1),
   column2    UInt32,
   column3    String
)
ENGINE = MergeTree
ORDER BY (column1, column2)
```

Highlight code in this blue color

Other resources

- Template for datasheets:
  - Google doc: https://docs.google.com/document/d/1tngw0LeTd8dOYamf92wWccgenC6z5Ooaq2ES6hQo8C0/edit
  - PDF: https://acrobat.adobe.com/id/urn:aaid:sc:VA6C2:ca1dfb61-384b-489b-9651-d7c8f1d7aa78

Lab 1.1: Module_1

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

02 Module_2

ClickHouse

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

Lab 2.1: Module_2

Objective: In this lab, you will

Module_3

ClickHouse

Lab 3.1: Module_3

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

Module_4

ClickHouse

Lab 4.1: Module_4

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

05 Module_5

ClickHouse

Lab 5.1: Module_5

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

Module_6

ClickHouse

Lab 6.1: Module_6

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

07

Module_7

ClickHouse

Lab 7.1: Module_7

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

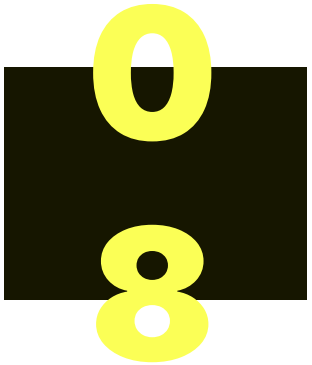
08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12



Module_8

ClickHouse

Lab 8.1: Module_8

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

09 Module_9

ClickHouse

Lab 9.1: Module_9

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

10

Module_10

ClickHouse

Lab 10.1: Module_10

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

Module_11

ClickHouse

Lab 11.1: Module_11

Objective: In this lab, you will

Agenda

01 Module_1

02 Module_2

03 Module_3

04 Module_4

05 Module_5

06 Module_6

07 Module_7

08 Module_8

09 Module_9

10 Module_10

11 Module_11

12 Module_12

12 Module_12

ClickHouse

Lab 12.1: Module_12

Objective: In this lab, you will

Thank you!

ClickHouse Community and other resources:

- Visit learn.clickhouse.com or clickhouse.com/learn for free, official ClickHouse Training
- ClickHouse Docs: https://clickhouse.com/docs
- ClickHouse Community Slack (free): https://clickhouse.com/slack
- ClickHouse Support help: https://clickhouse.cloud/support
- ClickHouse Cloud (free 30-day trial): https://clickhouse.cloud
- Agenda: Survey
- Certification: https://clickhouse.com/learn/certification

ClickHouse Certified Developer

Share your digital badge!

Thank you!