

Monitoring training

How to collect, store and visualize data



Agenda

- **Introduction 9:00**
- **How to collect and store data 9:30 - 11:30**
 - ❖ InfluxDB
 - ❖ Telegraf
 - ❖ OpenSearch
 - ❖ Logstash & Beats

Lunch break

- **How to visualize data 13:00 - 15:00**
 - ❖ OpenSearch Dashboards
 - ❖ Live demo with a demonstration dataset in the OpenSearch Dashboards
 - ❖ Grafana
 - ❖ Example of creating a Grafana dashboard with panels

Why monitoring?

- We want to monitor and control *in real-time* servers, applications, database instances and entire infrastructure:
 - to quickly **find** important **infrastructure and performance metrics** of services/applications on the fly (seeing them on dashboards);
 - to constantly **monitor health and availability** of the services;
 - to quickly **identify problems** (before they occur);
 - to centralize, search, analyze a large volumes of data (es logs).
- Understanding the state of the infrastructure and systems is essential for ensuring the reliability and stability of the services.

Logs and metrics

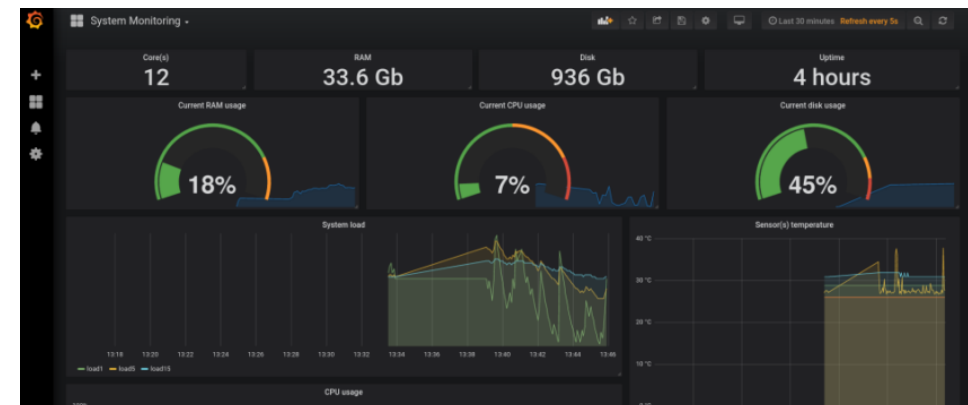
- A *log message* is a system generated set of data when an event has happened to describe the event.

```
[10/Oct/2020:13:55:36 -0700] [error][client 127.0.0.1] client denied by server configuration:  
/export/home/live/test
```

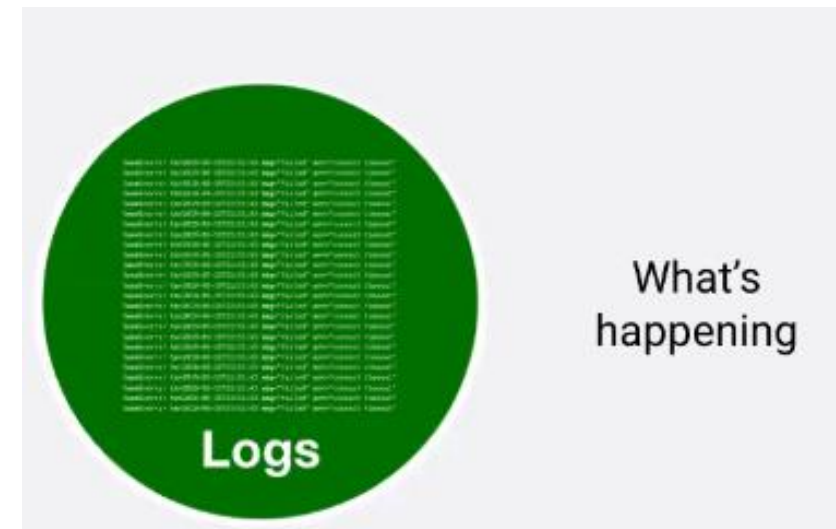
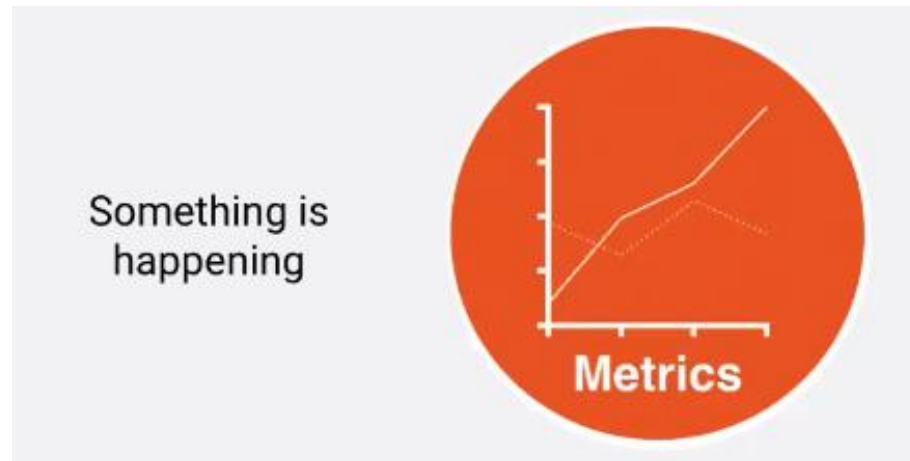
- *Metrics* are a numerical representation of data that can be used to determine a service or component's overall behavior over time.

Example of System metrics:

- Server memory utilization - Used, cached, free
- CPU utilization - Load average and usage
- Number of CPU cores
- Processes - stopped, running, etc..
- Disk Utilization

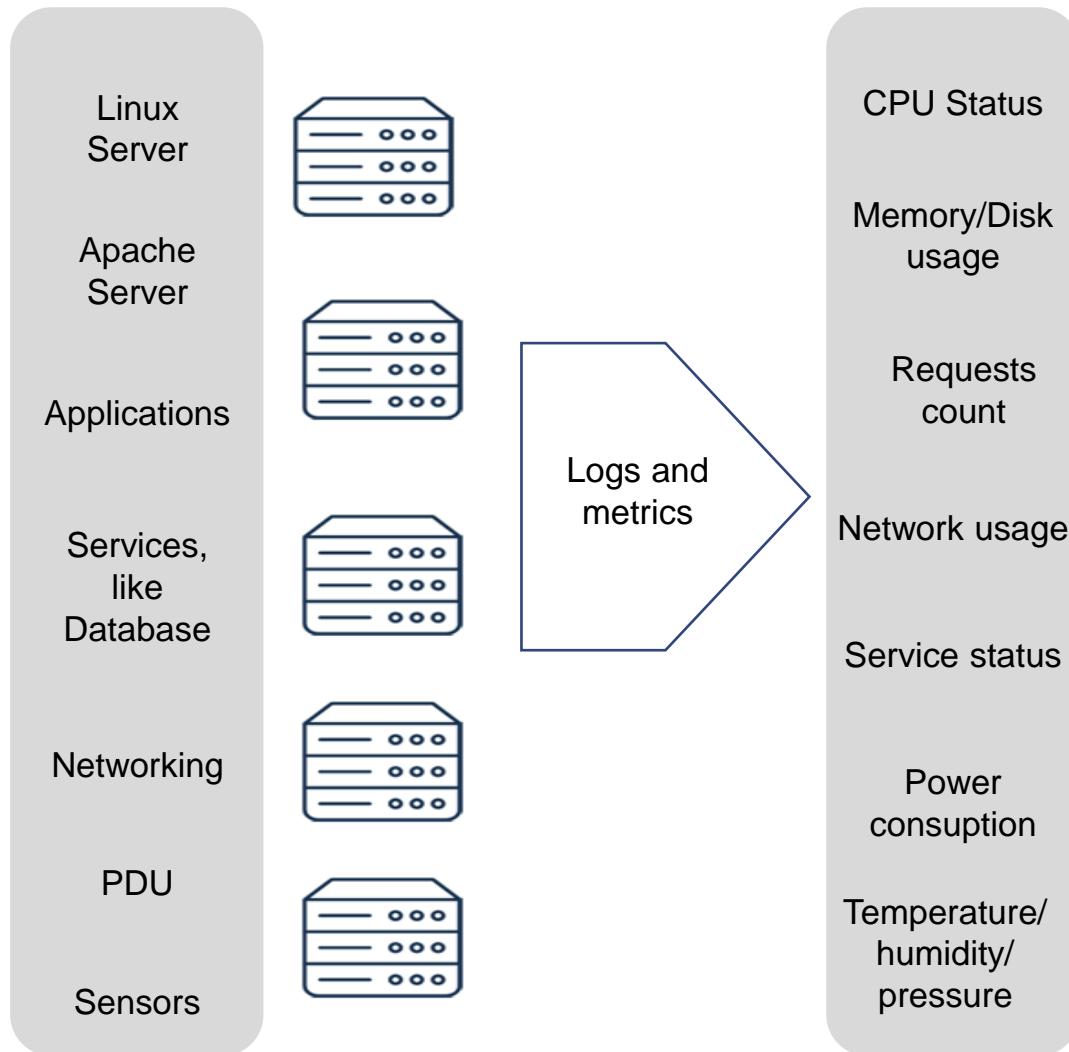


Logs and metrics



You can also extract metrics that are embedded in logs!

What monitoring?




How many users online?

Enough resources?

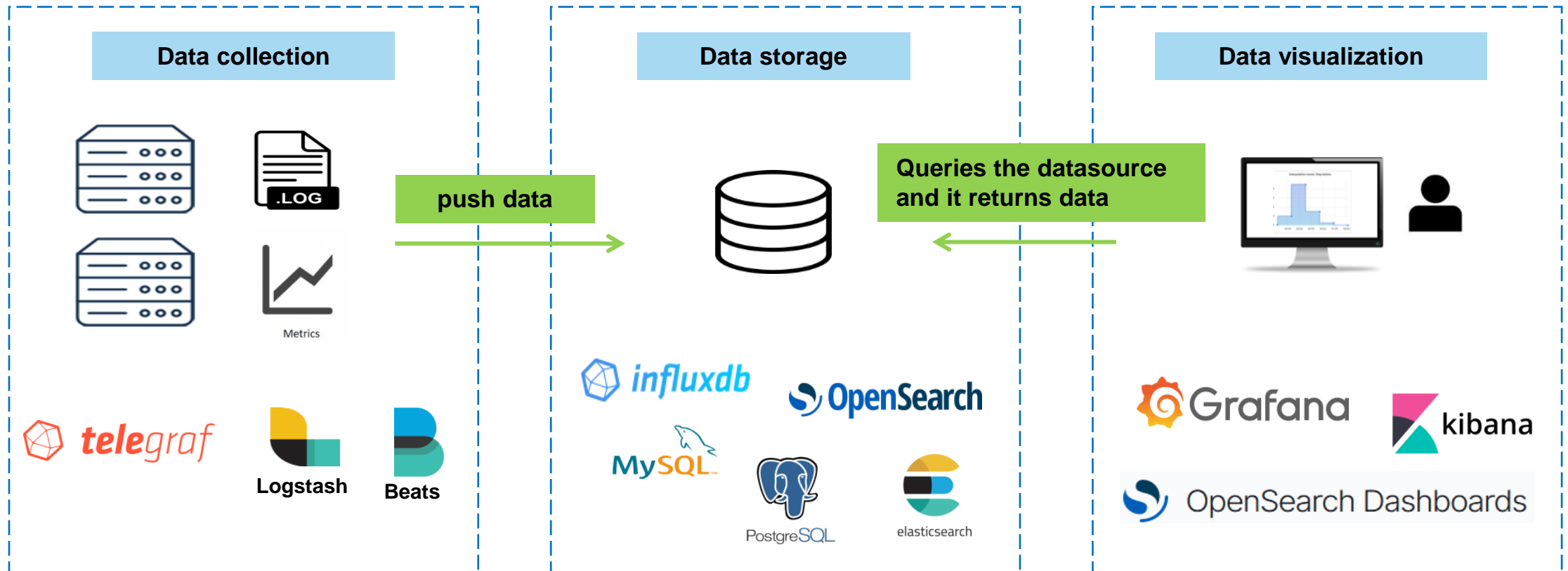
Errors?

Temperature in a room?



Service is running?

Components of monitoring architecture



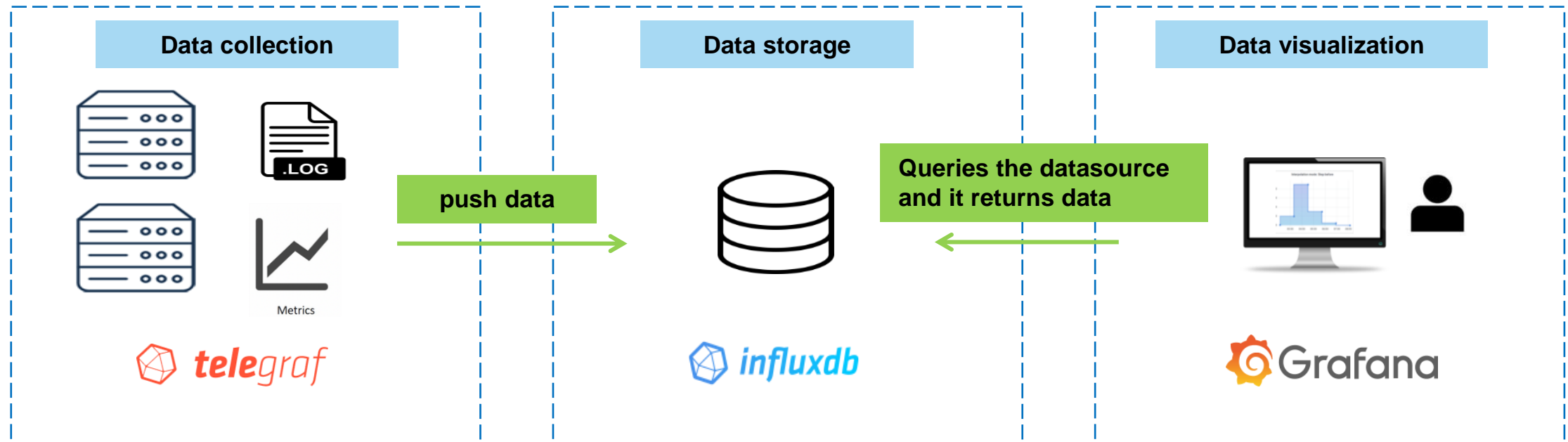
How does a monitoring system work?

A monitoring system typically:

- organizes and correlates data from various inputs
- accepts and store incoming and historical data (logs and metrics)
- provides visualizations of data
- optionally initiates automated responses when the values meet specific requirements (alerts)

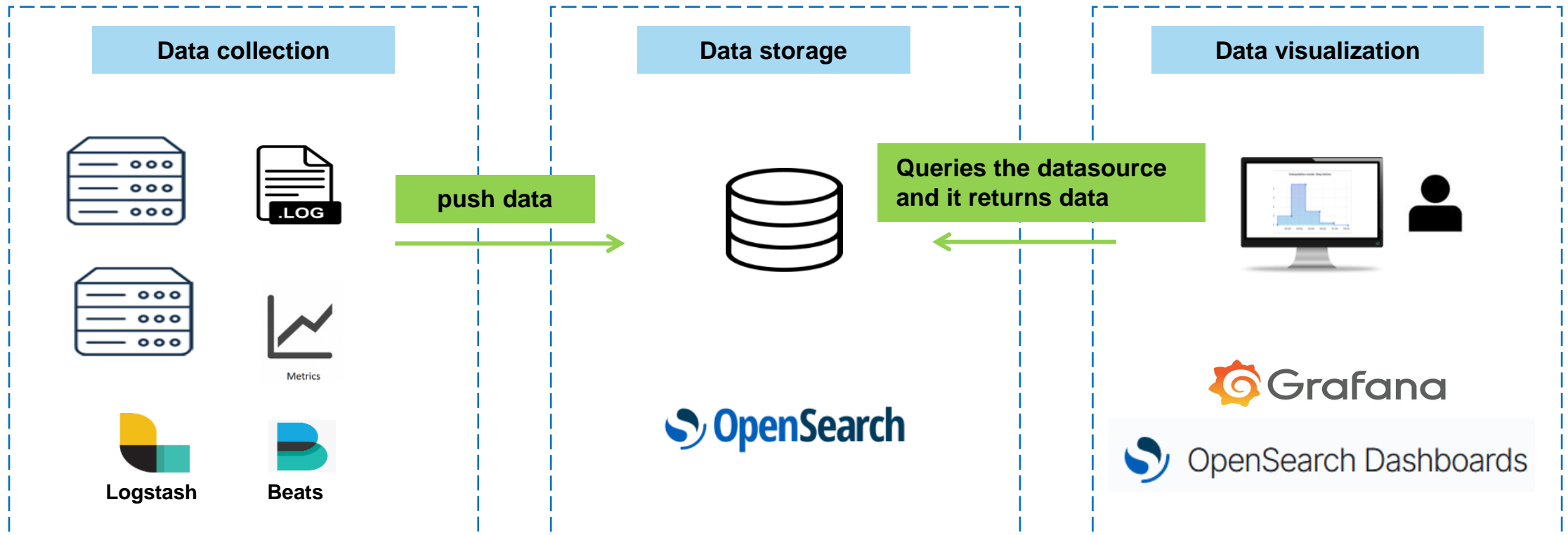
Scenario 1: Monitoring with Telegraf & InfluxDB & Grafana

Telegraf, InfluxDB, and Grafana is a popular combination for monitoring system



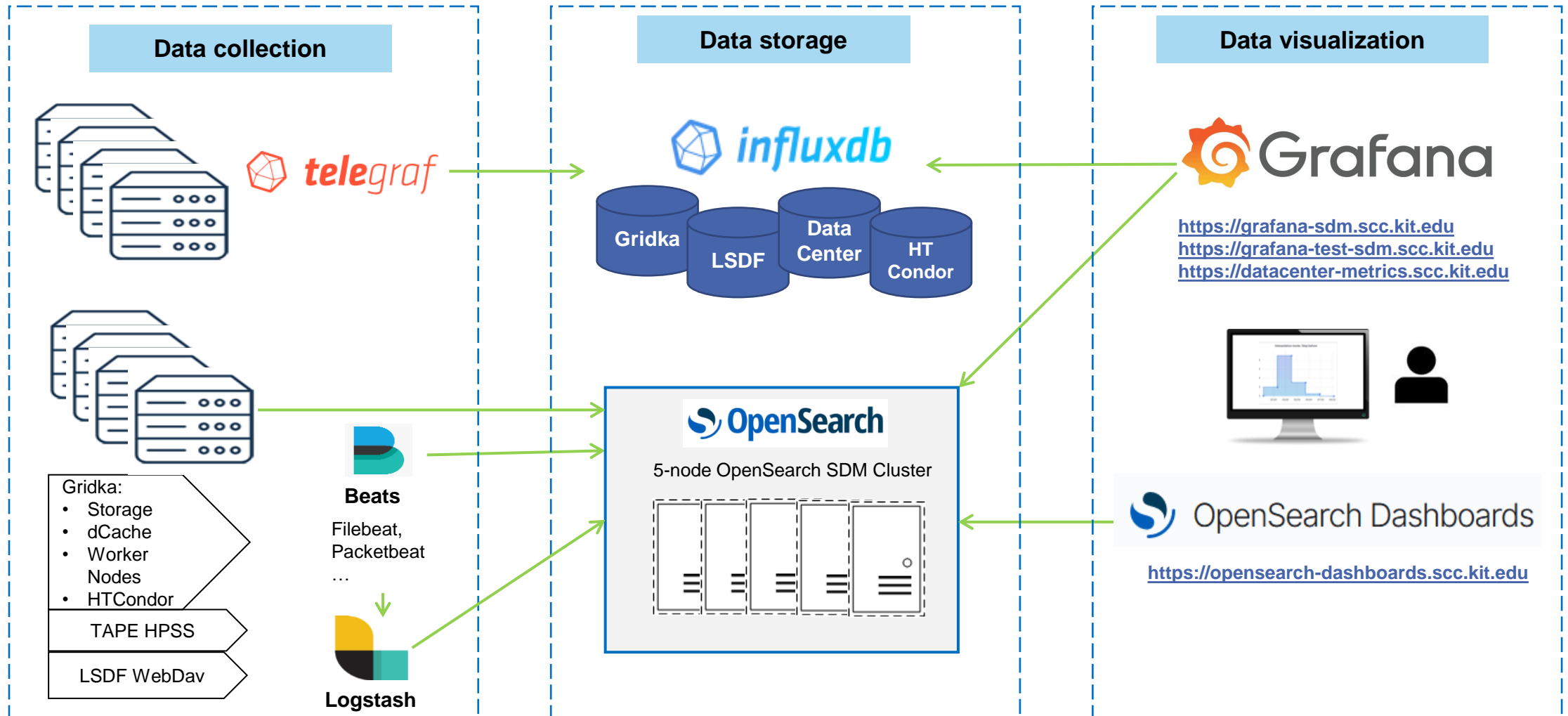
- Telegraf is used to collect the data and send it to InfluxDB
- InfluxDB stores the data
- Grafana reads from the database and present the data as customizable dashboards

Scenario 2: Monitoring using OpenSearch



- Logstash/Beats is used to collect and transform the data
- OpenSearch stores the data
- Grafana/OpenSearch Dashboards read from OpenSearch and visualize data in dashboards and graphs

Monitoring at SDM



Monitoring at SDM

- Monitoring of logs and metrics from various resources (LSDF, GridKa, TAPE etc..) for many use cases. For instance:
 - Server metrics
 - Storage operations
 - GridKa experiments
 - Error debugging
 - Temperature, humidity, pressure in campus server rooms
- Service availability expectation: employees responsible for the monitoring infrastructure are expected to be available for work only inside of their regular hours.
- No 'On-call' service.