WEEK 2 GRADED ACTIVITY

SYSTEM THINKING

TIAGO ROBERTO MUTSCHALL KREBS

DATA EVENT PIPELINE

DATA EVENT PIPELINE

Scenario: what is the issue and what role do you play?

I am the technical software engineering manager responsible for R&D on the Data squads (engineering, analyst, and science) at an Edge Computing company. A new high throughput distributed application will be delivered globally in the next months and collecting raw data from it is crucial to monitoring health, troubleshooting, measuring performance, and, creating insights to support further decisions. Currently, there is no such of system available in the company.

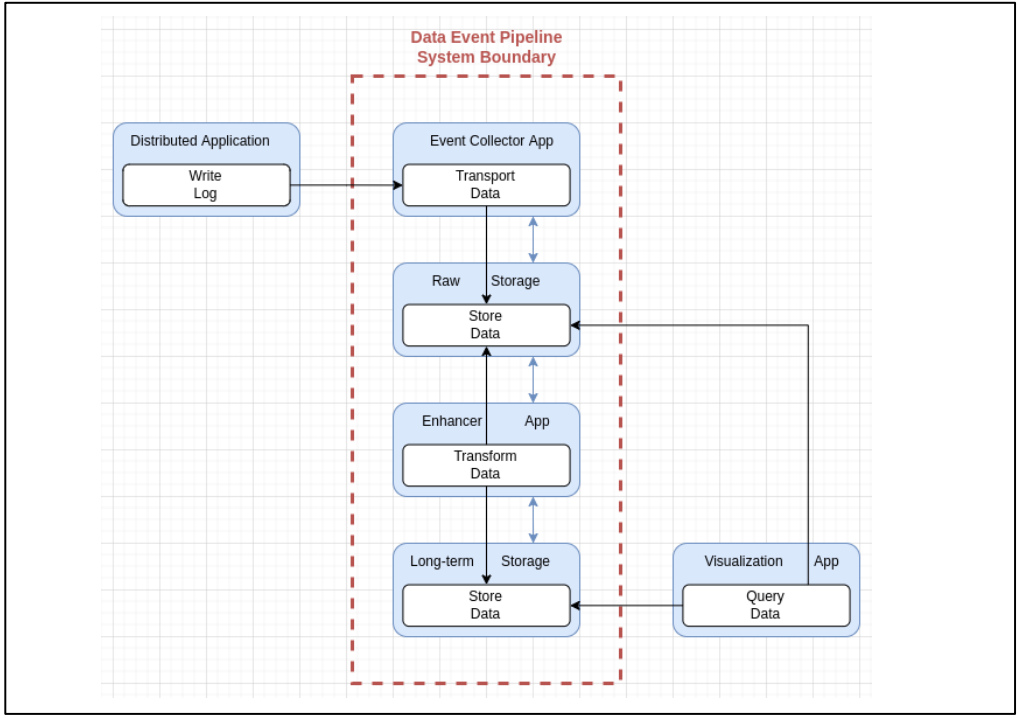
What is the expected emergent function and benefit?

The expected emergence is to store near real-time, enhanced and structured data in a single point of observation for a long period of time. The benefit is to provide useful information and convenience of use to the users and the company.

What are the problems associated with emergence of function, performance, ilities and emergencies:

The problem is to keep data throughput (performance) and consistency (reliability) at the same time.

What are the important entities in the system and context? Describe using the visual representation, modify as needed (you can draw it in a paper, take a picture, and paste below). Read the tutorial for representing systems.



Do not include any information that would violate security, confidentiality, or intellectual property regulation.

DATA EVENT PIPELINE

What are the pairs of entities that are related, what are the interaction and structure? Describe using the grid.

Pairs of components	Structure (connection, location or sequence)	Interaction (processes interacting directly, passing or sharing an operand)
Distributed Application & Event Log Collector	Distributed Application ships Logs of events to Event Log Collector	Share granular raw data
Event Log Collector & Raw Storage	Event log collector send data to Raw Storage	Share blocks of raw data
Enhancer App & Raw Storage	Ehancer App push data from Raw Storage	Share blocks of raw data
Enhancer App & Long-Term Storage	Enhancer App send data to Long-term storage	Share strucured data
Visualization App & Long-term Storage	Visualization App query data from Long-Term Storage	Share structured data
Visualization App & Raw Storage	Visualization App query data from Raw Storage	Share blocks of raw data

DATA EVENT PIPELINE

Q1. How can you reason through entities and relations to better understand the problem?

The raw data created by the distributed application is sent locally to the event collector, which creates blocks of raw data and sends it to the centralized raw storage. Then the enhancer app gets the data from the raw storage, transforms it in a structured way, and sends to the long-term storage. Finally, the Visualization App can search the structured data and raw data through queries.

Q2. Is there one or two system solution that you would propose?

The system itself is the proposal to create the emergence wanted and solve the problems of keeping data throughput (performance) and consistency (reliability) at the same time. More precisely, create a dedicated data pipeline based on two-layer storage and constant stream and transformation of data with the intent to have a single point of observation and availability concerns.