

Key Terms

08:41 AM

Latency

Latency is a measure of delay. In a network, latency measures the time it takes for some data to get to its destination across the network.

Firewall

A Firewall is a network security device that monitors and filters incoming and outgoing network traffic

ARN

Amazon Resource Names

Amazon Resource Names (ARNs) uniquely identify AWS resources.

Policies

A policy is an JSON object in AWS that defines permissions on AWS resources

IOPS

IOPS measures the **number of read and write operations per second**, ---> number of operation you doing per sec

Throughput

throughput measures the **number of bits read or written per second**. ----> number of operation you doing per sec

Durability

99.999999999%

Durability is a measurement of the **tiny errors that occur in files due to these underlying media**. When you write, read, and rewrite gigabytes, terabytes, and petabytes of information to the same drive, one or more individual bytes can **get corrupted or lost**.

Fault tolerance

Fault tolerance is the property that enables a system to continue operating properly in the event of the failure of one or more faults within some of its components.

Structured

Structured data is highly specific and is stored in a predefined format,

Feedback -->

Title, date Desc

<String> <Date> <String>

Unstructured

where unstructured data is a conglomeration of many **varied types of data** that are stored in their native formats

Feedback

Desc

ETL -

ETL stands for “**extract, transform, load**,” the three processes that, in combination, move data from one database, multiple databases, or other sources to a unified repository—typically a data warehouse.

Serverless

Serverless is a cloud computing execution model in **which the cloud provider allocates machine resources on demand, taking care of the servers on behalf of their customers.**

"Serverless" is a misnomer/mislead in the sense that servers are still used by cloud service providers to execute code for developers.

Data science --> ML / AL / ->

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. **Wikipedia**

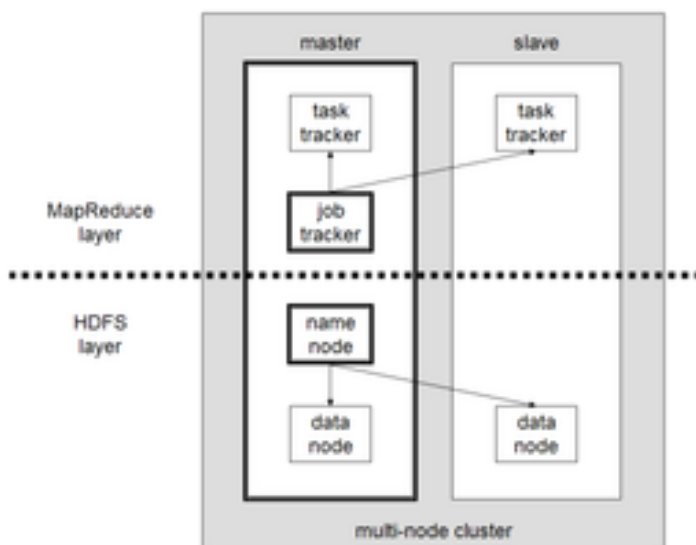
Data Analytics -> E, T, L --> Vistal

Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. **Wikipedia**

Goal to find insights from data and get customer behaviour ---> Can take business decision

Apache Hadoop / Hadoop

Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. **Wikipedia**



Data lake

A data lake is a vast pool of raw data, the purpose for which is not yet defined.

Data warehouse

A data warehouse is a repository for structured, filtered data that has already been processed for a specific purpose.

Redshift Scheme

Each schema in a database contains tables and other kinds of named objects. By default, a database has a single schema, which is named PUBLIC.

create schema if not exists ps_store_schema;

You can use schemas to group database objects under a common name. Schemas are similar to file system directories, except that schemas cannot be nested.

/parent/child
/thbs -> table,

Subquery

A Subquery or Inner query or a Nested query is a query within another SQL query and embedded within the WHERE clause. A subquery is used to return data that will be used in the main query as a condition to **further restrict the data to be retrieved**.

Join --> Not return // slow execution of query will be high

Throttling will delay executing a function. It will reduce the notifications of an event that fires multiple times.

Debouncing will bunch a series of sequential calls to a function into a single call to that function. It ensures that one notification is made for an event that fires multiple times.

The Information Technology Infrastructure Library (ITIL)-

ITIL provides a set of best practices for change management that makes it easier for IT professionals to roll out and prioritize changes efficiently, without negatively impacting customers or agreed-upon service levels.

To stay competitive and avoid the stress of implementing changes without direction, it's important to understand these guidelines.

Incident Management ---> aims to manage the lifecycle of all Incidents (unplanned interruptions or reductions in quality of IT services). The primary objective of this ITIL process is to return the IT service to users as quickly as possible.

Change management --->>> is an IT service management discipline.

The objective CM in this context is to ensure that standardized methods and procedures are used for efficient and prompt handling of all changes to control IT infrastructure in order to minimize the number and impact of any related incidents upon service.

Problem Investigation:

As ITIL defines it, a problem is "a cause or potential cause of one or more incidents."

Incident

And an incident is a single unplanned event that causes a service disruption. In other words, incidents are the nasty episodes on-call employees are typically scrambling to resolve as quickly and completely as possible.

