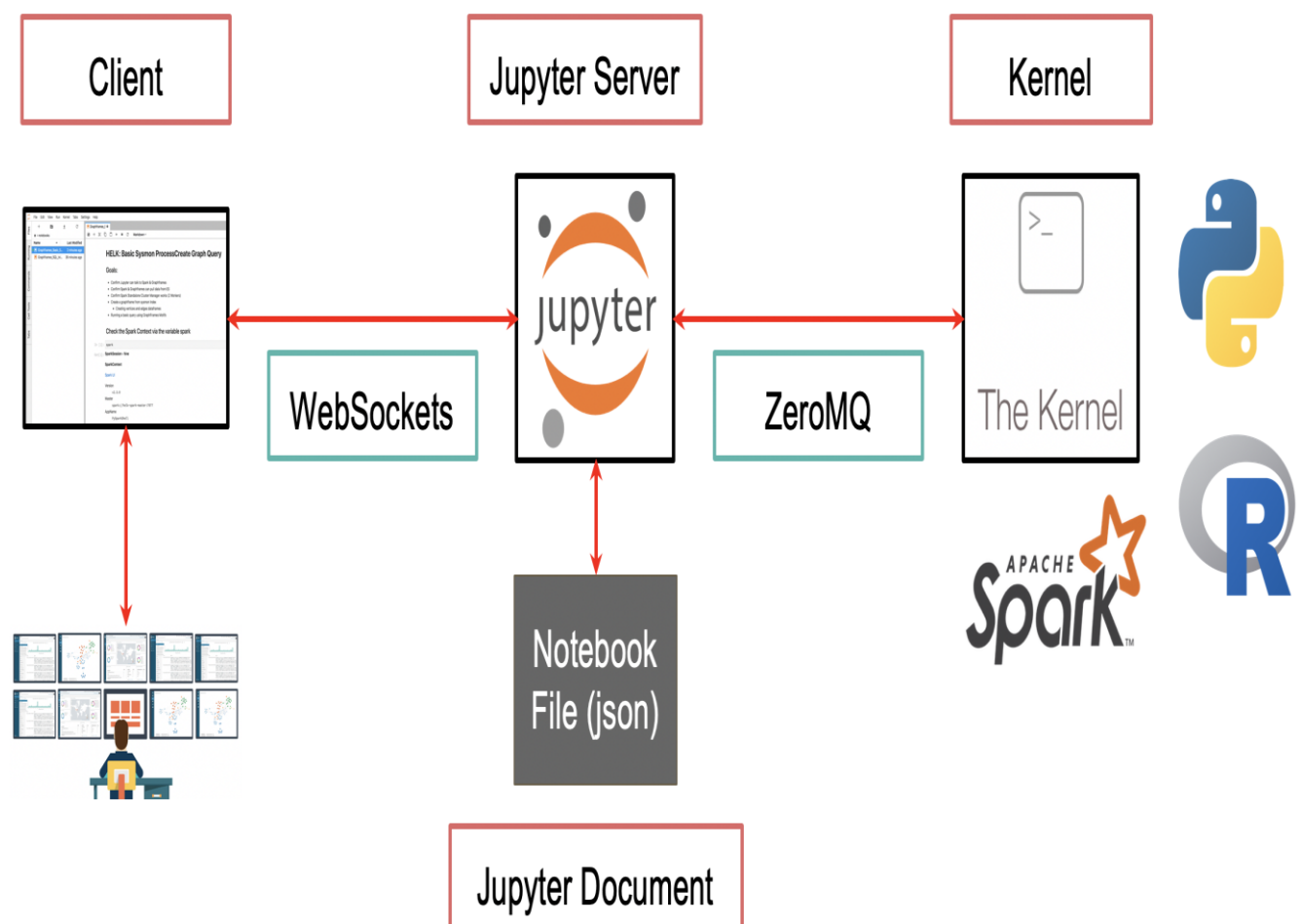


Architecture Design

FIFA World Cup Analysis

Written By	Author 1
Document Version	1.0
Last Revised Date	



STEP-BY-STEP-DATA VISUALIZATION PROCESS

COLLECTING DATA

The first and most important step of data visualization is to gather data in large amounts only then we can apply data visualization techniques on the collected data and we will get some helpful insights from it

CLEAN YOUR DATA

The output received from a data cleaned process is a dataset that is free of errors, missing data, anomalies, etc

CHOOSE A CHART TYPE

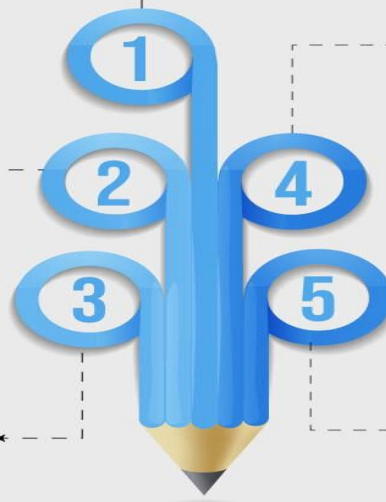
Before choosing a visual chart or graph, it is important to understand your audience or the collected data's domain and then choose a chart or graph accordingly which will best communicate the message

PREPARE DATA

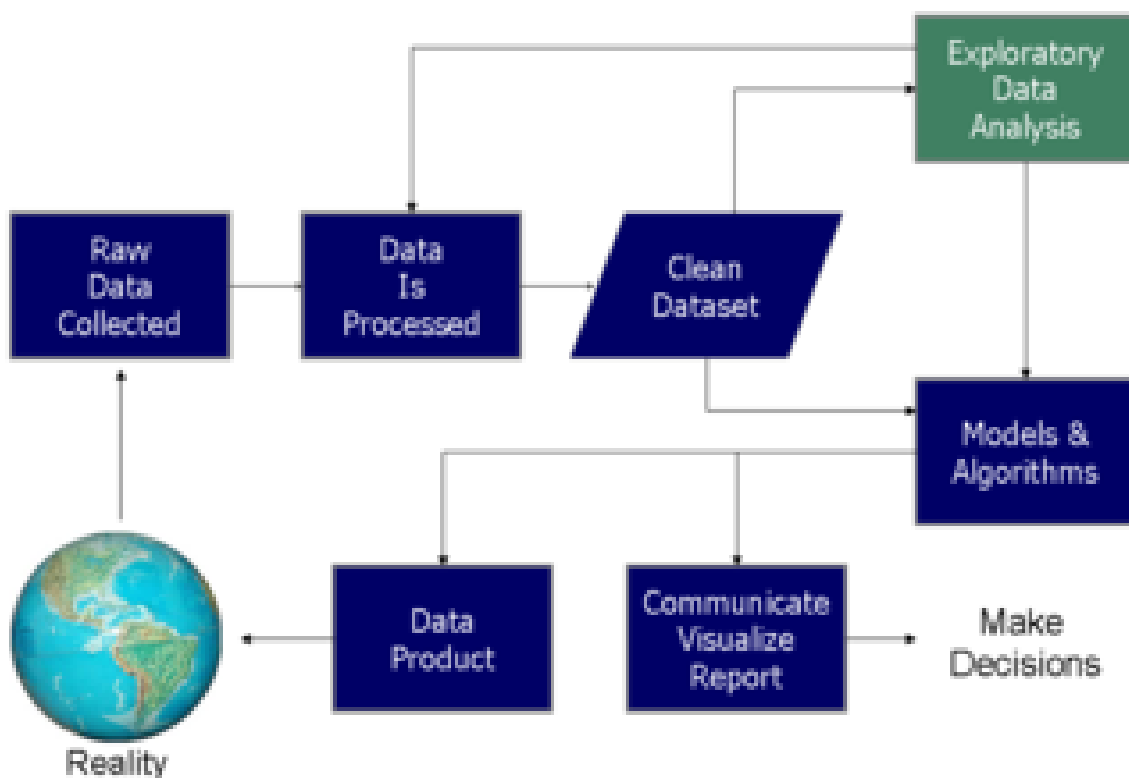
Data preparation tasks include finding data columns that help to make some decisions out of it and which gives some meaningful insights about data, grouping of data creating aggregate values for groups, combining variables to create new columns etc

VISUALIZE DATA

In the final step you'll have the required data you need to create visualizations. Now you can apply all your visualizations skills on the prepared data and represent the data in charts or graphs with meaningful insights



Data Science Process



DOCUMENT CONTROL

Change Record:

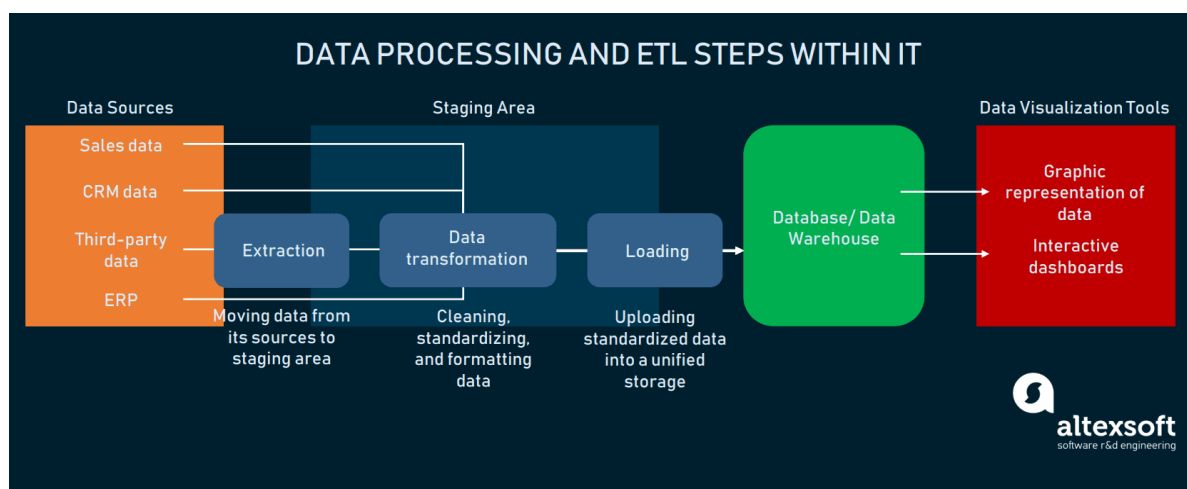
VERSION	DATE	AUTHOR	COMMENTS
1.0	19- Sep - 2021	Saurabh Gupta	Introduction and architecture defined
1.2	20 - Sep - 2021	Saurabh Gupta	Architecture & Architecture description appended and updated.

Reviews:

VERSION	DATE	REVIEWER	COMMENTS
0.2	21- May - 2021	Jyoti	Unit test cases to be added

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS



Contents

1.	Introduction	04
1.1	What is Architecture Design Document?	04
1.2	Scope	04
2.	Architecture	05
2.1	Tableau Architecture.....	05
2.2	Tableau Server Architecture.....	05
2.3	Gateway/Load Balancer	06
2.4	Application Server	06
2.5	VIZQL Server.....	07
2.6	Data Engine.....	07
2.7	Backgrounder	07
2.8	Data Server	07
2.9	Tableau Communication Flow.....	07
3.	Deployment.....	08
3.1	Deployment Options in Tableau	09
3.2	Single Node Architecture.....	10
3.3	Three Node Architecture.....	11
3.4	Five Node Architecture	12

1. Introduction

1.1 What is Architecture design document?

Any software needs the architectural design to represent the design of software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures.

Each style will describe a system category that consists of :

- A set of components (eg: a database, computational modules) that will perform a function required by the system.
- The set of connectors will help in coordination, communication, and cooperation between the components.
- Conditions that how components can be integrated to form the system.
- Semantic models that help the designer to understand the overall properties of the system.

1.2 Scope

Architecture Design Document (ADD) is an architecture design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

2. Architecture

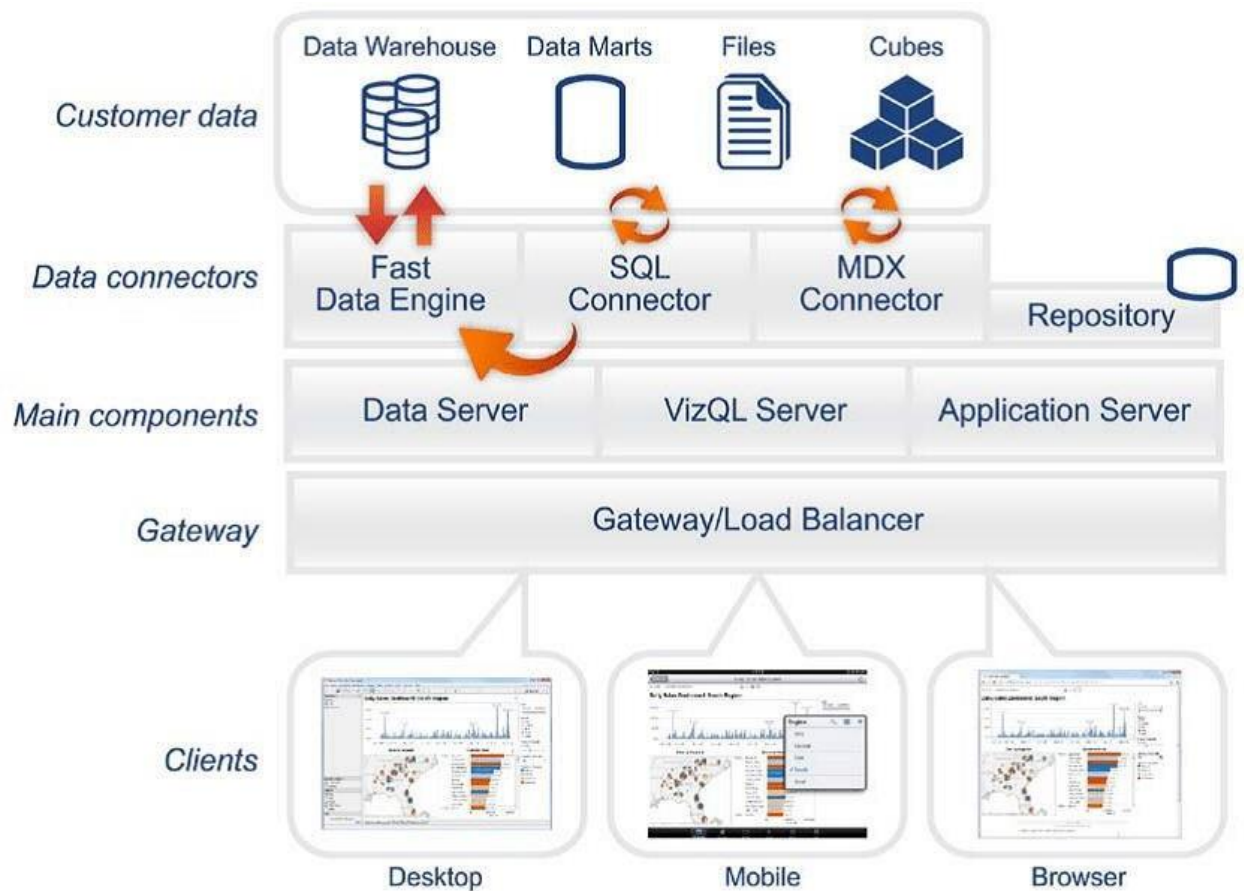


Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

The following diagram shows Tableau Server's

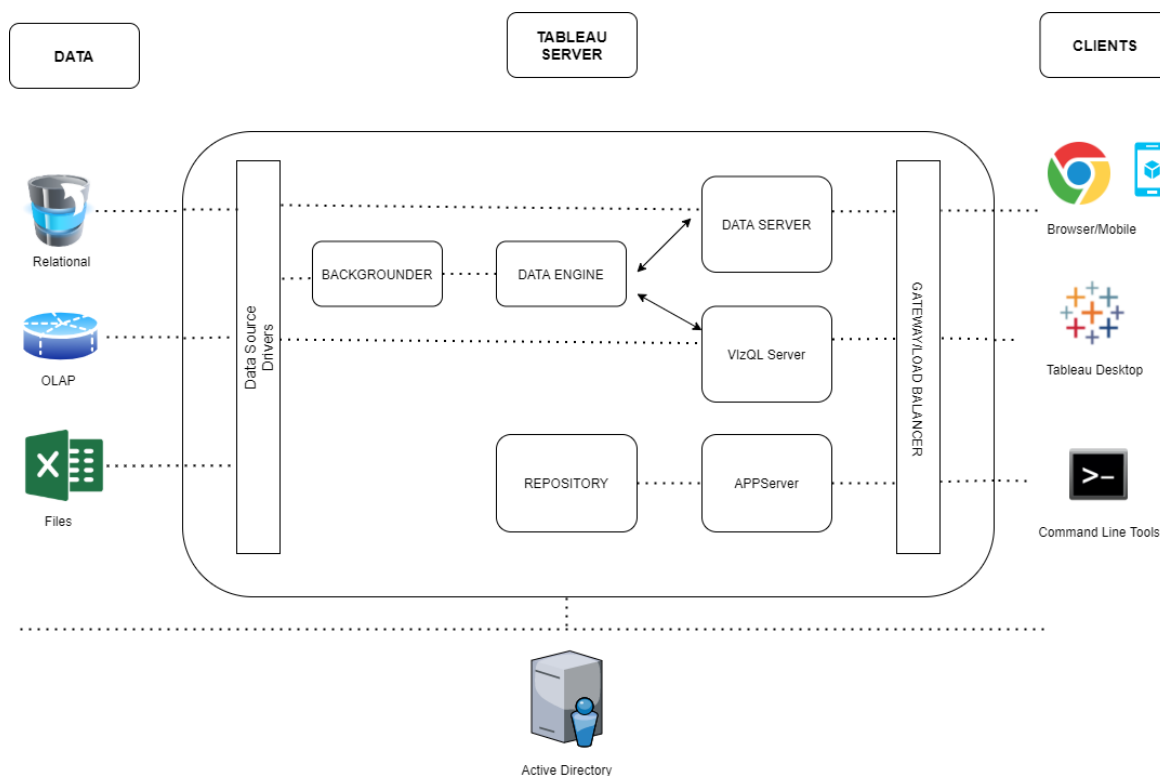


Tableau Server is internally managed by the multiple server processes.

1. Gateway/Load Balancer

It acts as an Entry gate to the Tableau Server and also balances the load to the Server if multiple Processes are configured.

2) Application Server:-

Application Server processes (wgserver.exe) handle browsing and permissions for the Tableau Server web and mobile interfaces. When a user opens a view in a client device, that user starts a session on Tableau Server. This means that an Application Server thread starts and checks the permissions for that user and that view.

3) Repository:-

Tableau Server Repository is a PostgreSQL database that stores server data. This data includes information about Tableau Server users, groups and group assignments, permissions, projects, data sources, and extract metadata and refresh information.

4) VIZQL Server:-

Once a view is opened, the client sends a request to the VizQL process (vizqlserver.exe). The VizQL process then sends queries directly to the data source, returning a result set that is rendered as images and presented to the user. Each VizQL Server has its own cache that can be shared across multiple users

5) Data Engine:-

It Stores data extracts and answers queries.

6) Backgrounder:-

The backgrounder Executes server tasks which includes refreshes scheduled extracts, tasks initiated from tabcmd and manages other background tasks.

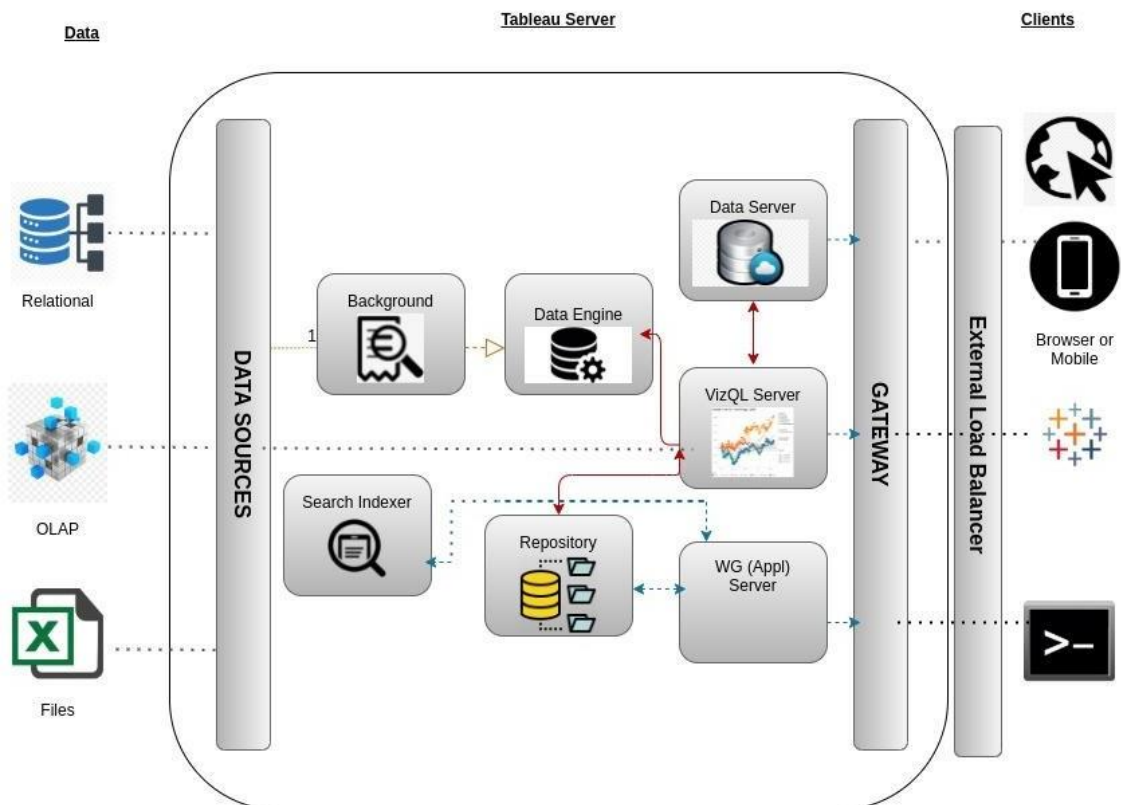
7) Data Server:-

Data Server Manages connections to Tableau Server data sources

It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

8) Tableau Communication Flow

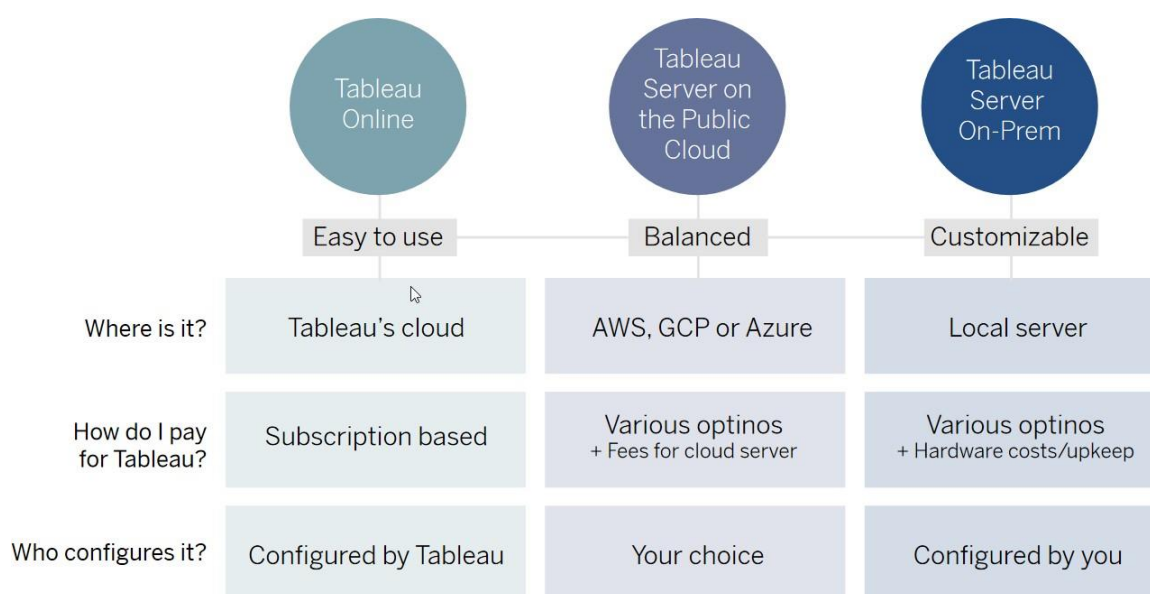
Tableau Communication Flow



3. Deployment Description

3.1 Deployment options in Tableau

Tableau's analytics platform offers three different deployment options depending on your environment and needs. The below graphic shows each option at a glance:



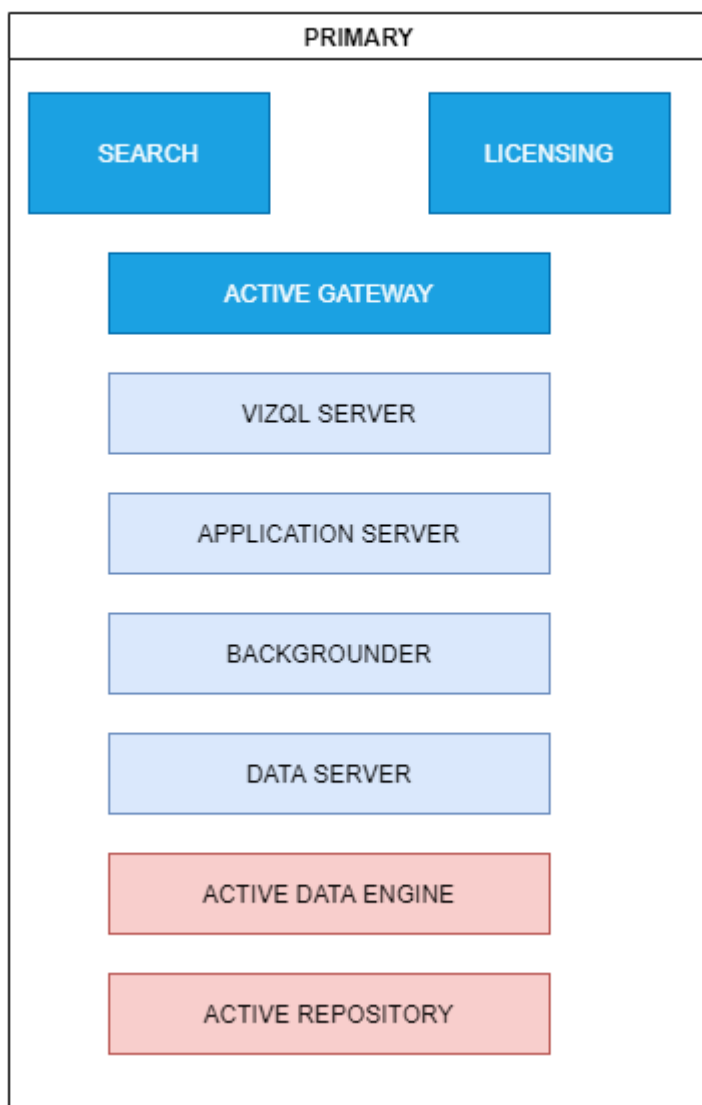
1. **Tableau Online** Get up and running quickly with no hardware required. Tableau Online is fully hosted by Tableau so all upgrades and maintenance are automatically managed for you.

2. **Tableau Server** deployed on public cloud: Leverage the flexibility and scalability of cloud infrastructure without giving up control. Deploy to Amazon Web Services, Google Cloud Platform, or Microsoft Azure infrastructure to quickly get started with Tableau Server (on your choice of Windows or Linux). Bring your own license or purchase on your preferred marketplace.

3. **Tableau Server deployed on-premises:** Manage and scale your own hardware and software (whether Windows or Linux) as needed. Customize your deployment as you see fit.

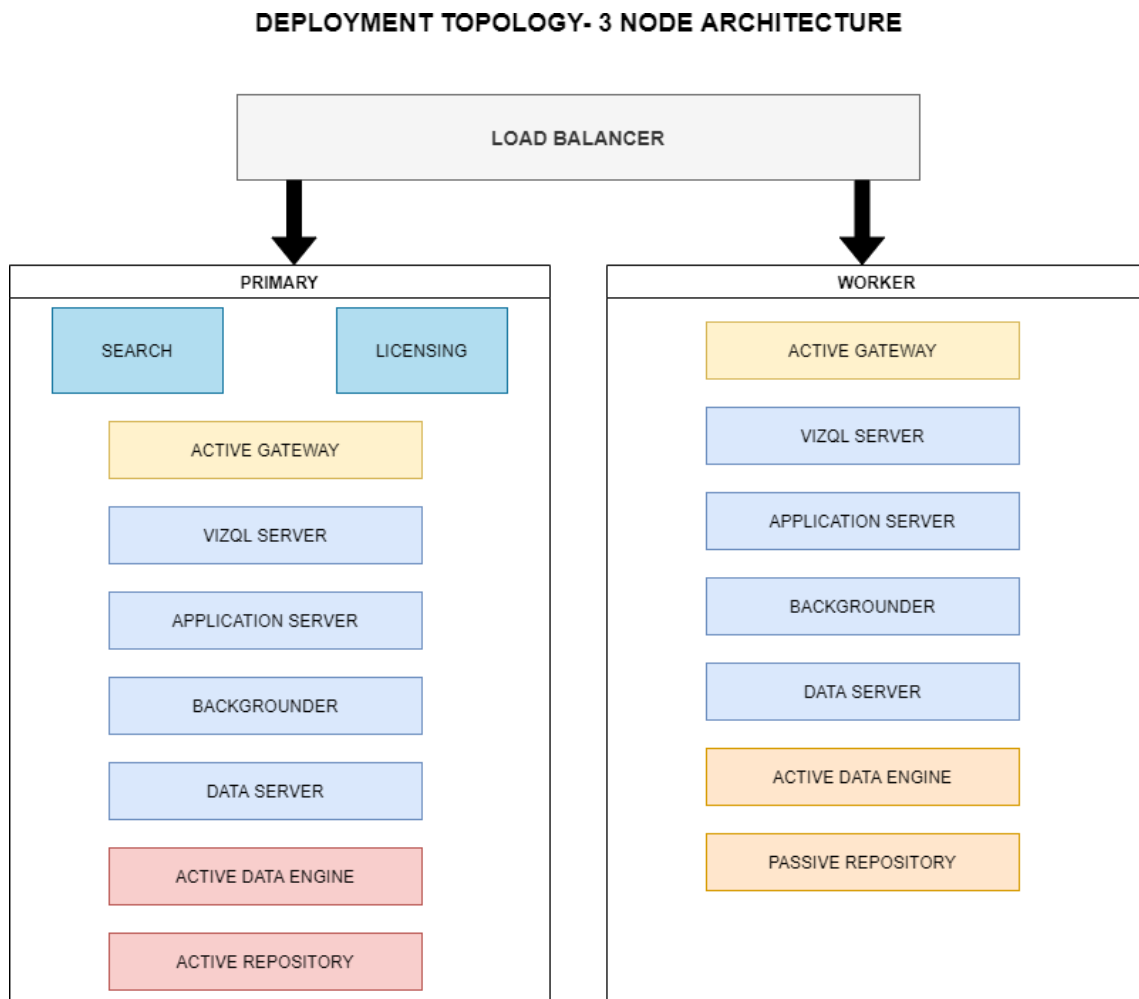
3.2 Single Node Architecture

DEPLOYMENT TOPOLOGY - SINGLE NODE ARCHITECTURE



This architecture is a single node architecture. This is the most simple deployment topology.

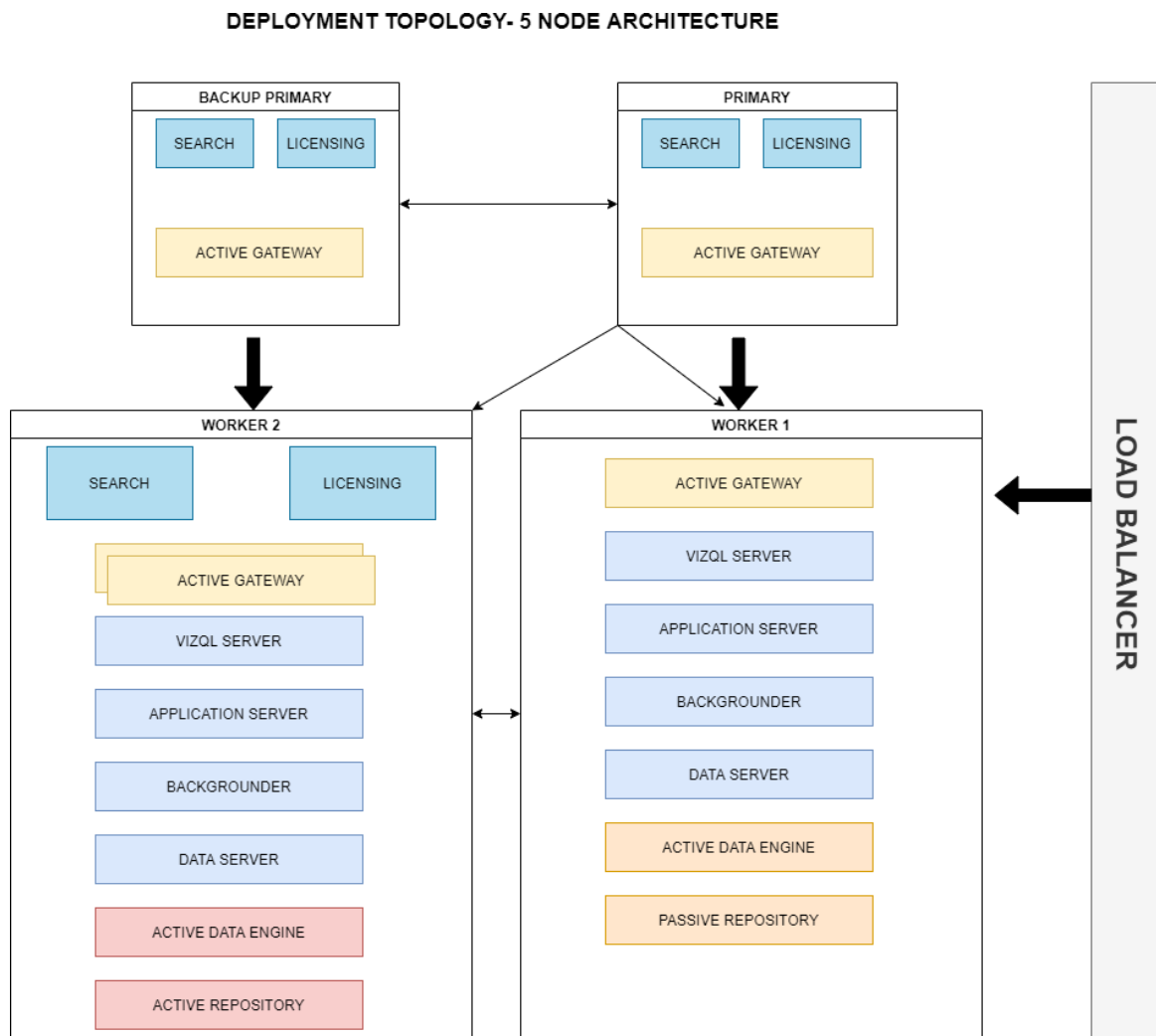
3.3)3 Node Architecture



This architecture is a 3 Node Architecture which is more capable to handle concurrent requests.

If we need failover or high availability, or want a second instance of the repository, we must install Tableau Server on a cluster of at least three computers. In a cluster that includes at least three nodes, you can configure two instances of the repository, which gives our cluster failover capability.

3.4)5 Node Architecture



When we install Tableau Server on a Five-node cluster, we can install server processes on one or both nodes. A five-node cluster can improve the performance of Tableau Server, because the work is spread across multiple machines.

Note the following about five-node clusters:

- A five-node cluster does not provide failover or support for high availability.
- You can't install more than one instance of the repository on a two-node cluster, and the repository must be on the initial node.