Technical White Paper: April 2022

VMware Sovereign Cloud Technical Guidance

Tanzu Data Services

Table of contents

Table of figures

Introduction

The VMware Sovereign Cloud Technical Guidance for Tanzu Data Services is a technical reference point for VMware Sovereign Cloud Providers which will help in understanding how to augment their Sovereign Cloud with Tanzu Data Services.

Add a tie-in and link here to your other Tech Guidance paper.

VMware's Sovereign Cloud Initiative recognizes VMware Cloud Verified partners that meet VMware's definition of a Sovereign Cloud. Partners who implement the VMware Sovereign Cloud guidance will be better able to protect the sovereignty of their customers workloads and data.

While the definition of what a Sovereign Cloud is, continues to evolve and is truly based on the Sovereign Nation or entity where the workloads/data/people will reside, technical guidance is needed to make our VMware Sovereign Cloud Providers successful in serving their customers. This guidance will be flexible enough to allow different configurations and implementations to comply with the entity that is providing the requirements of their Sovereign Cloud.

VMware Sovereign Cloud Providers benefit from clear guidelines around data sovereignty, data residency, data access, jurisdiction, control, and much more to provide customers with the assurance that their most sensitive data is managed securely. With sovereign cloud capabilities, customers benefit from the scale of a multi-tenant, hybrid cloud environment while maintaining security, access, and control like a traditional on-premises, legacy computing environment.

Legal Disclaimer

This document is intended to provide general guidance for organizations that are considering VMware solutions to help them address compliance requirements. The information contained in this document is for educational and informational purposes only. This document is not intended to provide regulatory advice and is provided "AS IS". VMware makes no claims, promises, or guarantees about the accuracy, completeness, or adequacy of the information contained herein. Organizations should engage appropriate legal, business, technical, and audit expertise within their specific organization for review of regulatory compliance requirements.

Audience

This document is intended for VMware Cloud Provider architects and technical leads responsible for planning and performing the implementation and maintenance of a VMware Sovereign Cloud environment.

Scope

This document addresses the following aspects:

- How can Tanzu Data Services benefit a Sovereign Cloud?
- Which Tanzu Data Services are available for Sovereign Cloud?

Benefits of Using Tanzu Data Services in a VMware Sovereign Cloud

Tanzu Data Services feature a portfolio of on-demand caching, messaging, and database software on VMware Tanzu for development teams building modern applications. Developers can spawn dedicated on-demand instances of data services. Automated provisioning supports rapid development, testing, and continuous delivery practices. Cache legacy data on-platform and expose it to a new world of cloud native applications. Or create new data architectures for applications with resilience, speed, and scale.

When Tanzu Data Services are implemented in a VMware Sovereign Cloud, the result is providing your customers with the ability to develop applications quicker and more efficiently. Managing large amounts of data becomes scalable and easier to manage.

Tanzu Data Services available for VMware Sovereign Clouds

Tanzu SQL (MySQL, Postgres)

High Availability for MySQL

High-availability (HA) MySQL instances offer automatic failover, ensuring that app requests operate continuously and without extended downtime. Because High-Availability is a requirement for data protection and data mobility, you can use the Configuring MySQL Instances for High Availability document to deploy a highly available MySQL cluster.

High availability offers automatic failover, ensuring that app requests operate continuously and without extended downtime. VMware Tanzu™ SQL with MySQL for Kubernetes uses InnoDB Cluster to provide a highly available MySQL instance on Kubernetes. InnoDB Cluster uses Group Replication and other MySQL technologies.

The diagram below shows the architecture of an HA MySQL instance:

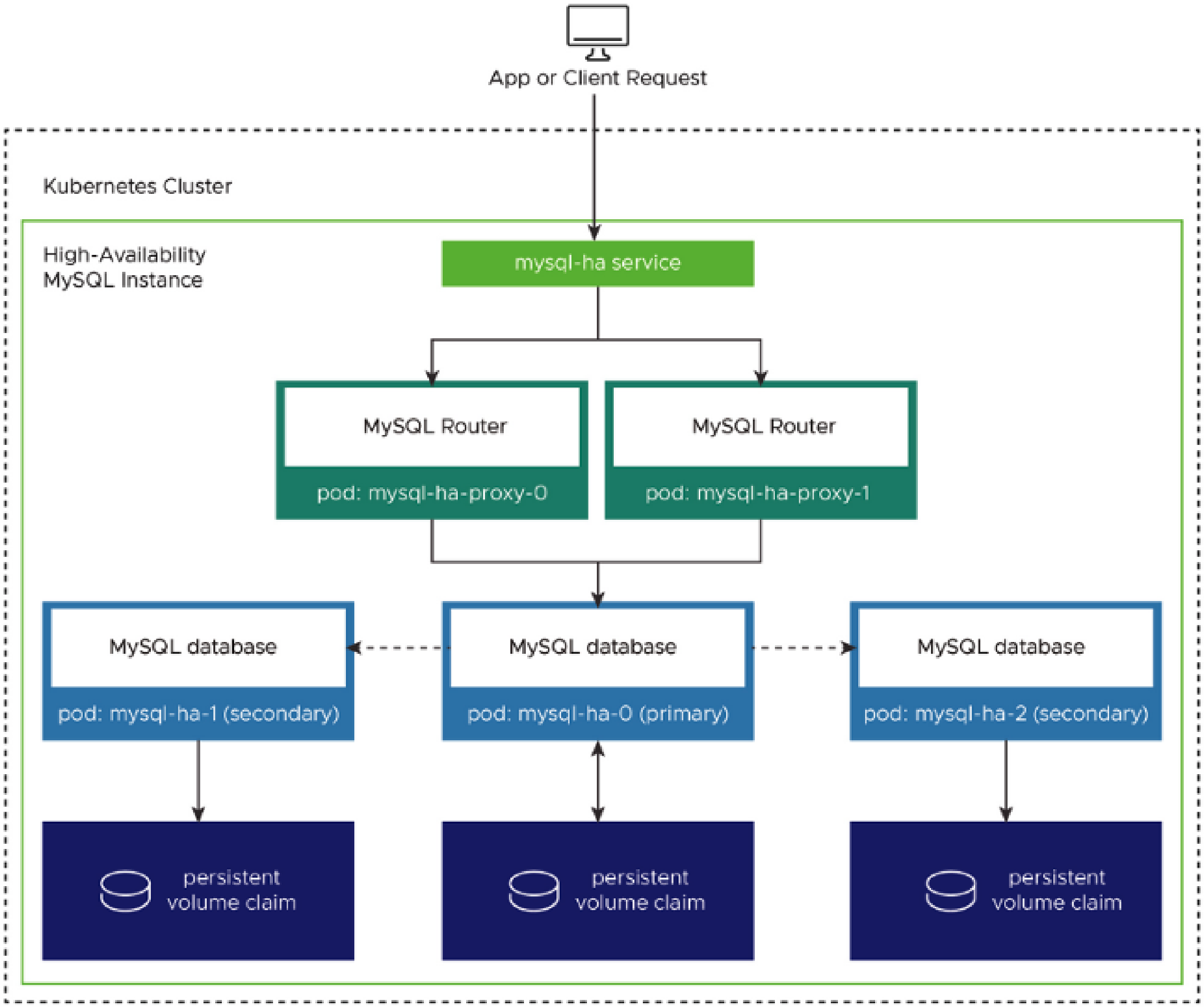

Figure 1: Architecture of an HA MySQL Instance

High Availability for Postgres

In the Tanzu Postgres HA cluster configuration, the topology consists of three pods: one monitor, one primary and one hot standby mirror. The “pg_auto_failover” service ensures that the data is synchronously replicated from the primary to the mirror node. If the primary node is unresponsive, the application requests are re-directed to the mirror node, which gets promoted to the primary. All application requests continue to the promoted primary, while a new Postgres instance is started which becomes the new mirror. If the monitor pod fails, operations continue as normal. The Postgres operator redeploys a monitor pod, and when ready it resumes monitoring of the primary and secondary. Again, it is important to consider how high availability is when using Tanzu Postgres databases in a Sovereign Cloud. Please review the Configuring High Availability in Tanzu Postgres document for further information regarding how to configure HA for Tanzu Postgres.

The diagram below shows how Tanzu Postgres uses the pg_auto_failover extension to provide a highly available Tanzu Postgres cluster on Kubernetes:

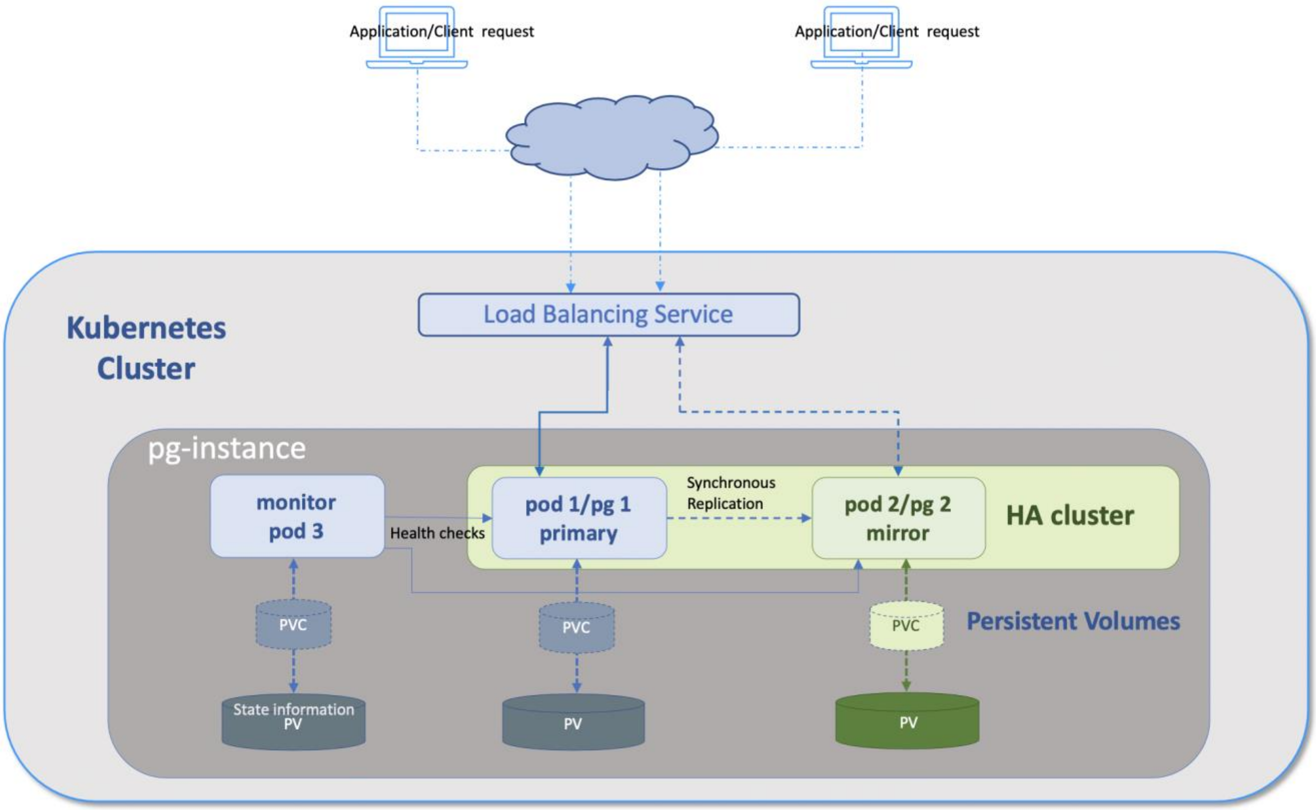

Figure 2: HA Tanzu Postgres Configuration

Encrypting Connections in MySQL

Tanzu MySQL for Kubernetes is configured to require an encrypted connection for client communication. There are various options available in configuring TLS for MySQL. Configuring encryption for all connections is vital in a Sovereign Cloud, as all data-in-flight needs to be secured. Please review the Configuring TLS for MySQL Instances document for all steps required for implementation.

Encrypting Connections in Tanzu Postgres

Tanzu Postgres (from version 1.2) supports Transport Layer Security (TLS) encrypted connections to the Postgres server from clients and applications. The Postgres server by default requires cert-manager self-signed certificates for internal Kubernetes communications. If using Tanzu Postgres in your Sovereign Cloud, you will want to follow the direction in the Configuring TLS for Tanzu Postgres Instances document.

Password Rotation

Rotating passwords in Tanzu SQL with MySQL instances is normally done when a password or user has been compromised. However, in a VMware Sovereign Cloud, it is recommended to rotate passwords on a regular basis. This can be determined by the security team and security protocols put in place by the customer. To understand how to perform a password rotation please visit the Rotating MySQL Credentials document.

NOTE: Tanzu Postgres SQL automatically rotates its “Secrets” for the “appuser”. It is best practice to follow a credential rotation for all applications that are used in a Sovereign Cloud.

Disaster Recovery

Backups and Disaster Recovery in a Sovereign Cloud are critical to maintaining data integrity. Keeping multiple copies may be a design consideration depending on the customers sovereign requirements. Please review the steps to implement this for Tanzu MySQL Instances in the Backing Up and Restoring MySQL Instances document or Backing Up and Restoring Tanzu Postgres document for Tanzu Postgres Instances.

RabbitMQ

VMware Tanzu RabbitMQ for Kubernetes provides the building blocks for a cloud native messaging and streaming service that you can deploy on any Kubernetes cluster. When Tanzu RabbitMQ is deploying a RabbitMQ cluster it also creates a Kubernetes service that allows other pods to use the cluster without the need for additional load balancer. It also provides a disaster recovery solution with a fast replication of messages to a standby cluster.

The diagram below shows the RabbitMQ architecture with Kubernetes Clusters:

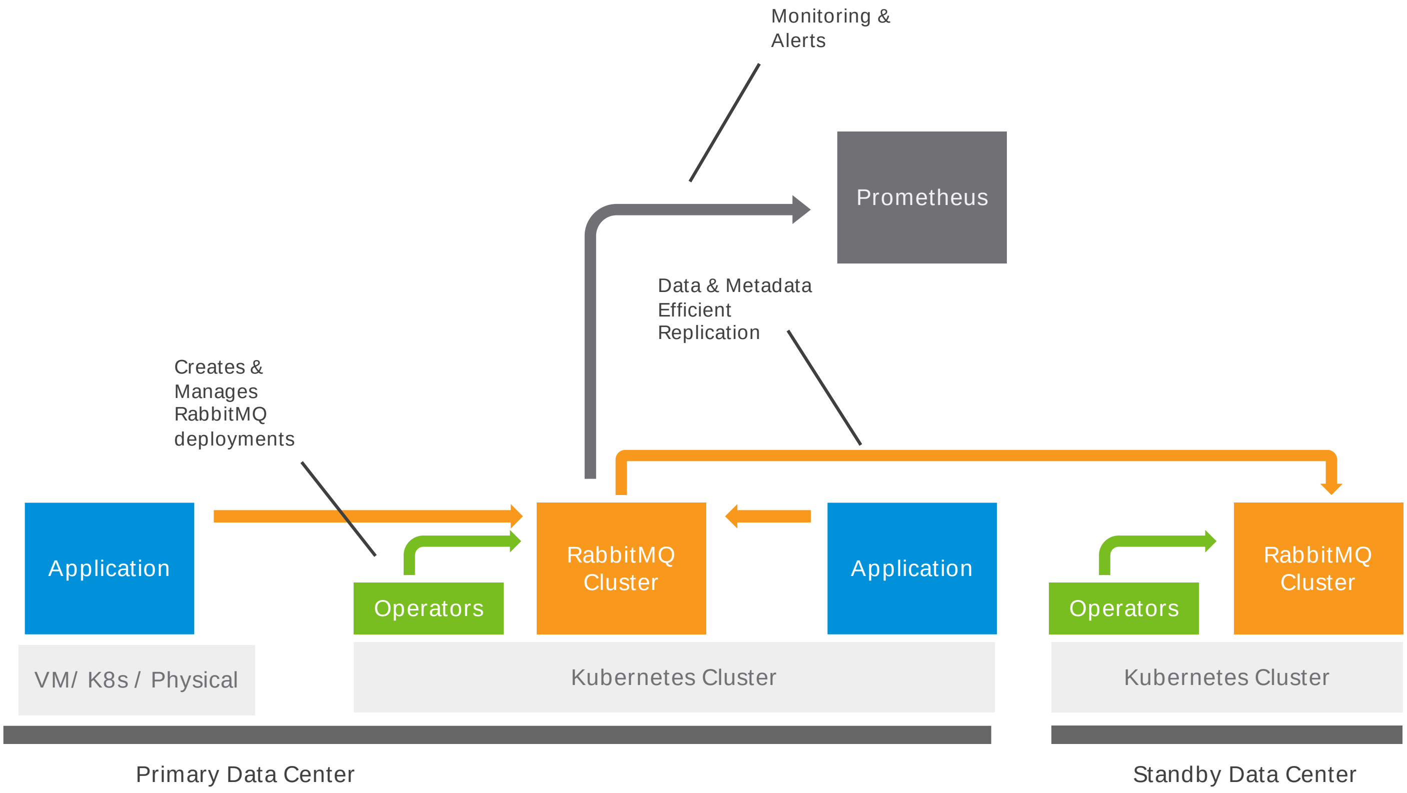

Figure 3: RabbitMQ Architecture

Best Practices

In all areas of a Sovereign Cloud, it is always good to stick with best practices for each application that is implemented. It will help when there is a need to troubleshoot issues that may arise. Please see the following website to learn about the RabbitMQ Best Practices.

Data Safety

When implementing a VMware Sovereign Cloud, data safety is of primary concern. Messaging-based systems are distributed by definition and can fail in different, and sometimes subtle, ways. To avoid losing messages on the RabbitMQ (as opposed to application) side, queues and messages must be able to cope with RabbitMQ node restarts, node and hardware failures. It is important to understand how to make sure your data is safe especially when it is implemented in a Sovereign Cloud. To better understand this please read the Data Safety in RabbitMQ document.

Access Control

Access Control is imperative for the security of a Sovereign Cloud to function and maintain its sovereignty. In RabbitMQ authentication and authorization are often confused or used interchangeably. That's wrong and in RabbitMQ, the two are separated. For the sake of simplicity, we'll define authentication as "identifying who the user is" and authorization as "determining what the user is and isn't allowed to do." Please follow this guide to RabbitMQ Authentication, Authorization, Access Control.

Encryption

Encryption of all data at all times is absolutely necessary when managing a Sovereign Cloud. TLS has two primary purposes: encrypting connection traffic and providing a way to authenticate (verify) the peer to mitigate against Man-in-the-Middle attacks. Both are accomplished using a set of roles, policies and procedures known as Public Key Infrastructure (PKI). Learn more about using RabbitMQ Encryption.

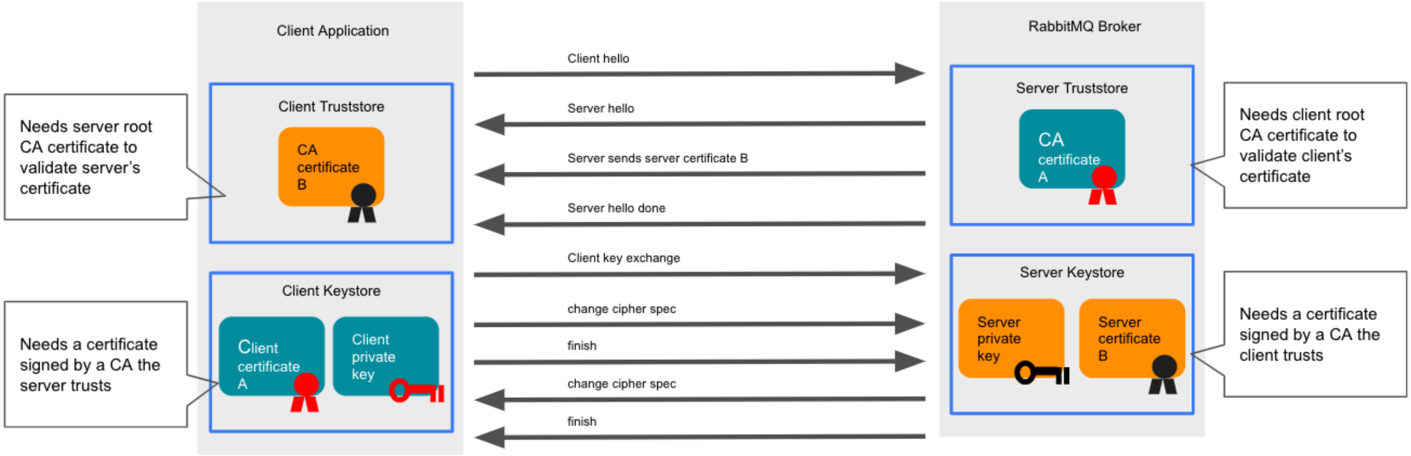

Figure 4: Example of encrypted communication in RabbitMQ

High Availability

A RabbitMQ cluster is a logical grouping of one or several nodes, each sharing users, virtual hosts, queues, exchanges, bindings, runtime parameters and other distributed state. The composition of a cluster can be altered dynamically. All RabbitMQ brokers start out as running on a single node. These nodes can be joined into clusters, and subsequently turned back into individual brokers again. There are many ways to form a cluster with RabbitMQ and the guide called RabbitMQ Clustering Guide should be used to properly setup high availability for your sovereign cloud environment.

Federation

The high-level goal of the Federation plugin is to transmit messages between brokers without requiring clustering. This is useful for a number of reasons. The federation plugin allows you to make exchanges and queues federated. A federated exchange or queue can receive messages from one or more upstreams (remote exchanges and queues on other brokers). A federated exchange can route messages published upstream to a local queue. A federated queue lets a local consumer receive messages from an upstream queue. Federation can be bidirectional as shown below. Learn more from the RabbitMQ Federation Plugin document.

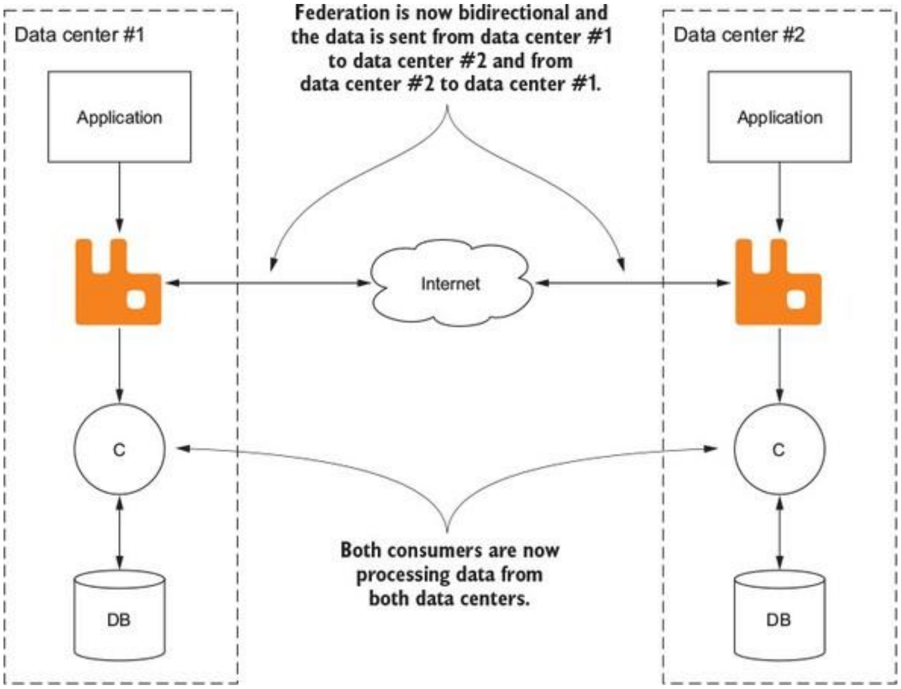
Data center #1
Federation is now bidirectional and the data is sent from data center #1 to data center #2 and from data center #2 to data center #1.
Data center #2
Application
Application
Internet
C
C
Both consumers are now processing data from both data centers.
DB
DB

Figure 5: RabbitMQ Bidirectional Federation configuration

Disaster Recovery

Disaster Recovery typically refers to a response of a major incident (a disaster) such as the loss of an entire data center, massive data corruption or any other kind of failure that could cause a total loss of service and/or data. Disaster Recovery attempts to avoid permanent partial or total failure or loss of a system and usually involves building a redundant system that is geographically separated from the main site. When using RabbitMQ in a sovereign cloud a good DR plan is critical to business continuity. The following blog called Disaster Recover and High Availability 101 will guide you in setting up Disaster Recovery for your RabbitMQ environment.

Greenplum

Tanzu Greenplum is a massively parallel processing (MPP) database server that supports next generation data warehousing and large-scale analytics processing.

By automatically partitioning data and running parallel queries, it allows a cluster of servers to operate as a single database supercomputer performing tens or hundreds of times faster than a traditional database. It supports SQL, MapReduce parallel processing, and data volumes ranging from hundreds of gigabytes to hundreds of terabytes.

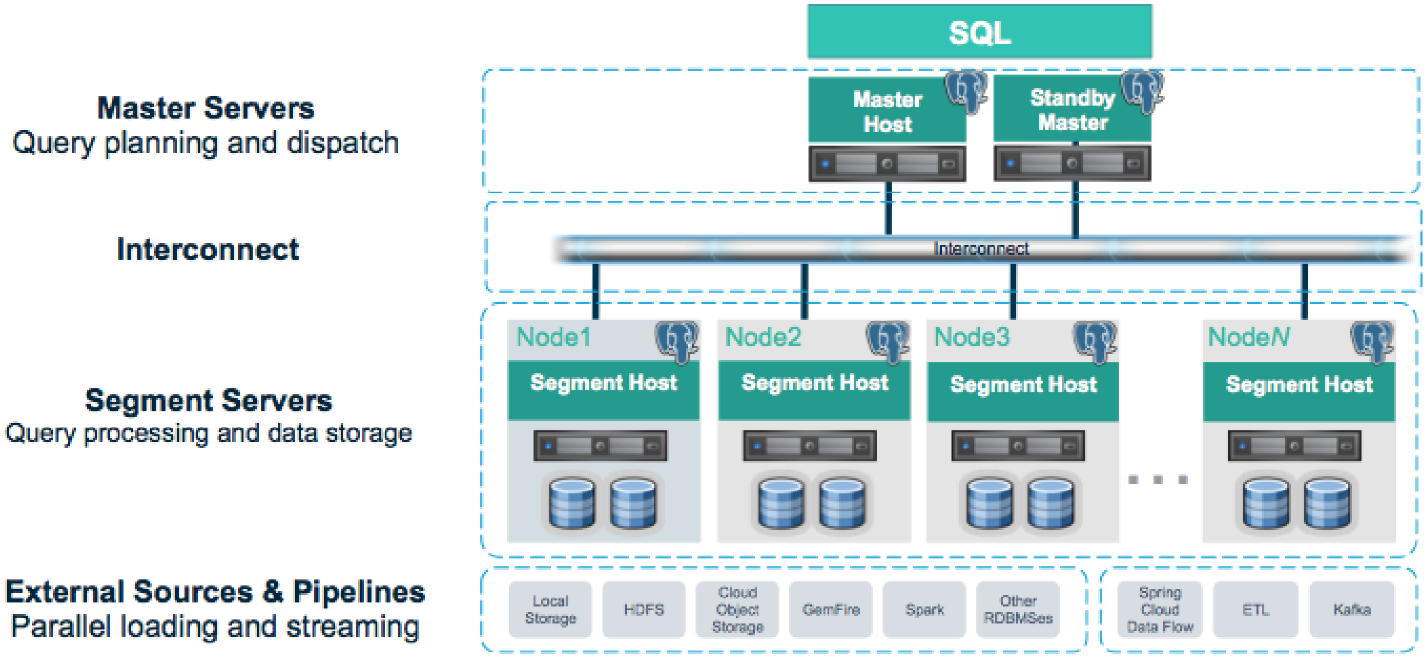

Figure 6: Greenplum Architecture

Best Practices

Greenplum Database is an analytical MPP shared-nothing database. This model is significantly different from a highly normalized/transactional SMP database. Because of this, the following best practices are recommended.

- Greenplum Database performs best with a denormalized schema design suited for MPP analytical processing for example, Star or Snowflake schema, with large fact tables and smaller dimension tables.
- Use the same data types for columns used in joins between tables.

Find out more in the Greenplum Best Practices Guide.

Database Security Configuration

It cannot be overstated how important database security is for any customer, especially those who operate in a sovereign cloud. The configuration that may be required by your customer may differ depending on the sovereignty requirements. The following guide called Greenplum Database Security Configuration Guide will help you as a Cloud Provider to start with a good foundation of security for your database.

Encrypting Data & Connections

Securing your database is not the only activity needed to make sure that your customer’s data remains safe and secure. Encryption is also needed at the data level and even in the connections that are allowed to the database itself. Here is an excerpt from the Encrypting Data and Database Connections Guide:

Encryption can be used to protect data in a Greenplum Database system in the following ways:

Connections between clients and the master database can be encrypted with SSL. This is enabled by setting the ssl server configuration parameter to on and editing the ‘pg_hba.conf’ file. See "Encrypting Client/Server Connections" in the Greenplum Database Administrator Guide for information about enabling SSL in Greenplum Database.

Greenplum Database 4.2.1 and above allow SSL encryption of data in transit between the Greenplum parallel file distribution server, ‘gpfdist’, and segment hosts. See Encrypting gpfdist Connections for more information.

Network communications between hosts in the Greenplum Database cluster can be encrypted using IPsec. An authenticated, encrypted VPN is established between every pair of hosts in the cluster. Check your operating system documentation for IPsec support or consider a third-party solution such as that provided by Zettaset.

The pgcrypto module of encryption/decryption functions protects data at rest in the database. Encryption at the column level protects sensitive information, such as passwords, Social Security numbers, or credit card numbers. See Encrypting Data in Tables using PGP for an example.

High Availability

Making sure your Greenplum Database is highly available in a sovereign cloud environment is very important. All applications and infrastructure should be highly available where possible. Greenplum Database supports highly available, fault-tolerant database services when you enable and properly configure Greenplum high availability features. To guarantee a required level of service, each component must have a standby ready to take its place if it should fail. Please refer to the High Availability document for Greenplum Databases in order to configure this in your environment.

Enhanced Data Security & Access Control

As enterprises seek to become more analytically driven, they face a balancing act: Capitalize on the proliferation of data science throughout the company, while protecting sensitive data from loss, misuse, or unauthorized disclosure. However, the continued increase of the regulation of data privacy is complicating how companies make data available to analysts. This white paper Enhanced Data Security and Access Control with Pivotal Greenplum discusses common vulnerabilities to data in motion and at rest, and the controls available to Greenplum users both natively and via Pivotal partner solutions. With this portfolio of controls, enterprises can protect sensitive data while preserving access to the users who need to make use of it.

Summary

The ability to augment your sovereign cloud with Tanzu Data Services is a great way to help your sovereign cloud customers to move towards a developer ready environment. This will continue to support a flexible, scalable, and secure approach to data management at a highly efficient and fast pace.

Revision History		
Date	Changes	Modified By
April 2022	Initial Document Creation	Cory Allen

Copyright © 2021 VMware, Inc. All rights reserved. VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001
VMware and the VMware logo are registered trademarks or trademarks of VMware, Inc. and its subsidiaries in the United States and other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. VMware products are covered by one or more patents listed at vmware.com/go/patents.
Item No: vmw-wp-tech-temp-uslet-word-2021 8/21