Scalability in Cloud computing by using VmWare Vfabric SQLFire

Ms.Shinde Rupali R   Prof.Mr.Maral Vikas B.
rupali.shinde53@gmail.com
Department of Computer Engineering
K.J.COEMR,Pune

May 8, 2013

Abstract

Cloud computing is a computing style in which scalable and flexible IT functionalities are delivered as a service to external customers using Internet technologies. Cloud computing is not a revolutionary idea; Instead, it is an evolutionary concept that integrates various existing technologies to offer a useful new IT provisioning tool. As more and more organizations embrace cloud computing to save money, increase productivity, and to gain the ability to scale their infrastructures up or down at a moments notice, concerns remain around the level of performance, availability, and overall security of their databases once they move to a hosting vendor. Many of the fears organizations have with moving their databases from a dedicated environment to the cloud start with the overall comfort level of managing their data in a virtual environment.However, advancements in both Microsoft SQL Server and cloud infrastructures have changed all that. More specifically, combining SQL Server with a proven hosting service provider offers highly scalable, flexible and cost-effective solutions. Scalability - Cloud computing enables universities to quickly scale up their IT operations as provisioning of new computing resources and software applications can be delivered at a desired pace. Furthermore, constraints on pre-purchasing of resources to meet peak requirement in traditional IT no longer exist.In this paper we are going to use Providing Scalable SQL data in the cloud With VMware vFabric SQL Fire.

Keywords

Cloud computing, Scalabe SQL data, VMware vFabric SQL Fire,OLAP/OLTP.

1 Introduction

With the increasing popularity of Web 2.0 applications, massive amounts of dierent types of data are being generated at an unprecedented scale[8]. Given this rate of continuous growth, coupled with advancement in broadband connectivity, virtualization, and other technologies, the cloud computing model, with its capability to dynamically provide for computation and storage, has emerged as an ideal choice for data-intensive IT infrastructures. The need to provide for capacity both in terms of storage and computation, and to support online transactional processing and online analytical processing in the cloud, has given rise to major challenges in architecting elastic and ecient data servers. The web service applications provided by Internet companies such as emailing, online shopping and social networking, are all based on online transactions that are essentially similar to those in traditional OLTP (online transaction processing) systems. However, in such web applications, system scalability, service response time and service availability demand higher priority than transactional data consistency, which is the foremost requirement of traditional OLTP systems. To better support search and data sharing, large-scale ad-hoc analytical processing of data collected from those web services is becoming increasingly valuable to improving the quality and eciency of existing services, and supporting new functional features. Due to the massive size of web data, traditional OLAP (online analytical processing) solutions (i.e., parallel database systems) fail to scale dynamically with needs. Therefore, both commercial companies and open-source communities have proposed new large-scale data processing systems, such as Hadoop , Hive , Pig and Dryad . Historically, OLTP and OLAP workloads are handled separately by two systems with dierent architectures RDBMS for OLTP and data warehousing system for OLAP. Periodically, data in RDBMS are extracted, transformed and loaded (aka. ETL) into the data warehouse. The system-level separation was motivated by the facts that OLAP is computationally expensive and its execution on a separate system will not compete for resources with the response-critical OLTP operations, and snapshot-based results are gen- erally sucient for decision making. Although this system-level separation provides exibility and eciency, it also results in several inherent limitations, for example, lack of data freshness in OLAP, redundancy of data storage. Several data mining languages purport to implement a cloud computing solution for certain SQL construc-tions, each attempting to reduce the effort needed to launch a MapReduce job. Googles proprietary software, both its distributed file system1 and its Bigtable relational overlay,2 have open source equivalents in the Apache-supported Hadoop3 project. Both use the MapReduce algorithm, and both find writing code for the parallel map and reduce phases to be a tedious and error-prone task. Consequently, both evolved scripting languages to capture query intent and produce MapReduce code automatically.

2 Related Work

VMware vFabric SQLFire is a memory-optimized distributed SQL database delivering dynamic scalability and high performance for data-intensive modern applications[1]. SQLFires memory-optimized architecture minimizes time spent waiting for disk access, the main performance bottleneck in traditional databases. SQLFire achieves dramatic scaling by pooling memory, CPU and network bandwidth across a cluster of machines and can manage data across geographies. Built on the proven vFabric GemFire distributed data management platform, SQLFire is ideal for primary data stores that require high transaction rates, continuous availability, and support by database programming staff without specialized coding assistance. Deployed as the front-end data- management layer for an Oracle or other RDBMS, SQLFire delivers substantially better performance and scalability for modest investments, and retains full compatibility with current DBA skills and tools.

3 What is Cloud computing ?

Cloud computing is a significant advancement in the delivery of information technology and services.[5] By providing on demand access to a shared pool of computing resources in a self-service, dynamically scaled and metered manner, cloud computing offers compelling advantages in speed, agility and efficiency. Today, cloud computing is at an early stage in its lifecycle, but it is also the evolution and convergence of several trends that have been driving enterprise data centers and service providers over the last several years. Cloud computing builds off a foundation of technologies such as grid computing, which includes clustering, server virtualization and dynamic provisioning, as well as SOA shared services and large-scale management automation. For the better part of a decade, Oracle has been the leader in these areas with thousands of customer successes and high level of investment. Today, Oracle offers the industrys most complete, open and integrated products and services to enable public, private and hybrid clouds. Oracle aims to make cloud computing fully enterprise-grade and supports both public and private cloud computing to give customers choice. Oracle offers technology that enables organizations to build private clouds, leverage public clouds and provide cloud services to others. Oracle also offers a broad set of horizontal and industry applications that run in a shared services private cloud model as well as a public Software-as-a-Service (SaaS) cloud model. This white paper provides an overview of Oracles cloud computing strategy and how Oracle helps customers and partners plan their evolution and adoption of a cloud computing model.

3.1 Cloud computing service model

Cloud computing encompasses many aspects of computing (from hardware to software) that a single solution is not able to provide all aspects. Generally speaking, cloud computing applications incorporate the combination of the following functional service models[5]:

3.1.1 Infrastructure as a Service (IaaS)

IaaS solutions provide users with physical or virtual resources that satisfy the requirements of the user applications in terms of CPU, memory, operating system and storage. Such Quality of Service (QoS) parameters are established through a Service Level Agreement (SLA) between the customer and the service provider. The end user has full controls over the virtualised computer instance, and can customise the instance accordingly. Unlike purchasing the physical servers, IaaS is usually charged on a utility basis depending on the consumption of the resources. A big name in IaaS space is Amazon.com, which launched Elastic Compute Cloud (EC2) in 2006 to offer a pay-as-you-go hosting service for customers computer applications. In 2008, Fujitsu also opened its London North Data Centre to outsource data storage and computing services with security options covering UK and International regulations, and compliance auditing according to the ISO27001 standard.

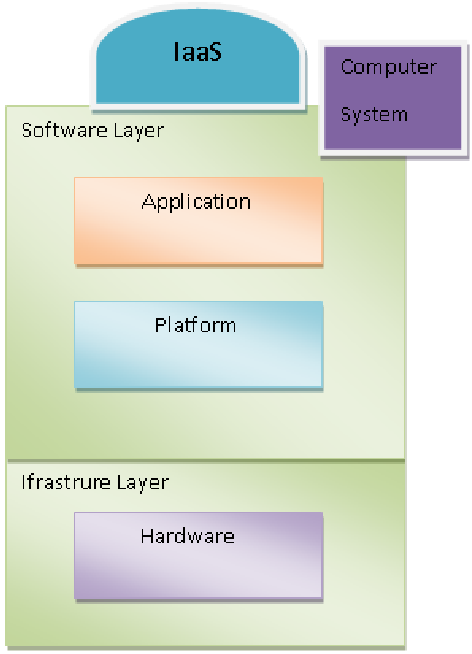

Figure 1: Infrastructure As A service(IaaS)

3.1.2 Platform as a Service (PaaS)

PaaS delivers cloud-based application development tools in addition to services for testing, deploying, collaborating, hosting, and maintaining applications. It hides all the complexity of managing the underlying hardware, provides all the facilities required to support the complete lifecycle of building and deploying web applications and services entirely from the Internet. With PaaS, users can develop new applications or services in the cloud that do not depend on a specific platform to run, and can make them widely available to users through the Internet. The accessibility of PaaS offerings enables any programmer to create enterprise-scale systems that integrate with other web services and databases. Two well-known PaaS solutions, Google AppEngine and Microsoft Azure, provide users with a development platform for creating distributed applications that can automatically scale on demand. Other PaaS service providers are 3Tera, RightScale, DataSynapse, Manjrasoft, Univa UD, Elastra and Enomaly.

3.1.3 Software as a Service (SaaS)

SaaS is a model of software deployment where a provider delivers its software as a service to be used by customers on demand. Under the traditional SaaS model, an application resides at an offsite data centre where the service provider maintains data, servers and related hardware. End-users access the application remotely via an Internet browser. The SaaS model is predicated on a one-to-many or multi-client

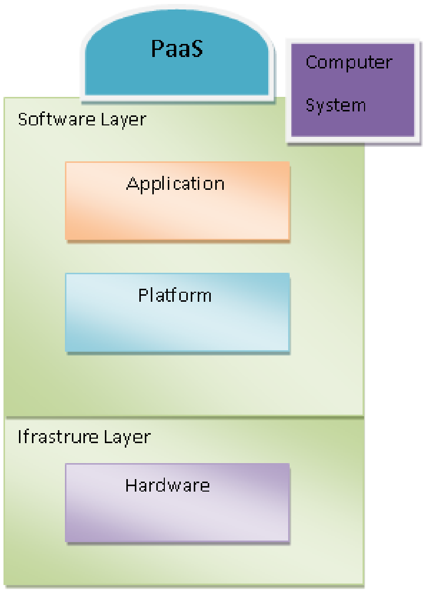

Figure 2: Platform As A service(PaaS)

delivery model whereby an application is shared across clients, providing opportunities to customise the system to the needs of each customer. Examples of SaaS are Salesforce.com and Clarizen.com, which respectively provide online Customer Relationship Management (CRM) and project management services. Google Apps provides desktop applications which are hosted in the cloud and replaces traditional desktop based Microsoft Office software. Northgatearinso offers on demand Human Resources (HR) solutions based on SAP Human Capital Management (HCM) module.

3.2 Key Benefits Achieved through Cloud Computing

- Flexibility - Cloud computing allows universities to expand or contract com-

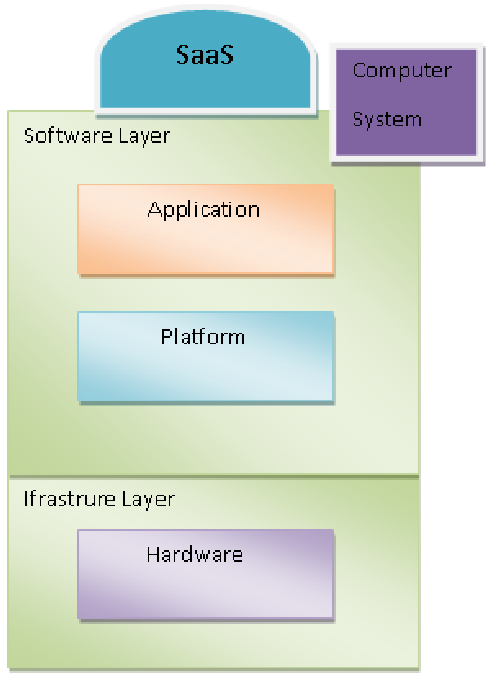

Figure 3: Softare As A service(SaaS)

puting power as required and allows bursts of computing power to be utilised on an on-demand basis. This flexibility helps ensure resource-intensive processes will not slow down other business processes and computing services are always operating at optimal cost.

- Scalability - Cloud computing enables universities to quickly scale up their IT operations as provisioning of new computing resources and software applications can be delivered at a desired pace. Furthermore, constraints on pre-purchasing of resources to meet peak requirement in traditional IT no longer exist.

- Economics - Traditional IT has multiple fixed and variable cost elements. In order to fulfil business requirements and sustain day-to-day business operations, universities must invest a large fixed amount for initial IT infrastructure establishment and continue to spend variably for software and hardware maintenance. By outsourcing IT functions to the cloud, universities can leverage the features of a lean IT structure to reduce the overall IT expenditures involved in software licensing, infrastructure development, on-going support and upgrades.

- Inherited Resiliency - Cloud computing removes single points of failure since the Internet is a highly resilient computing environment. Some competitive service providers also add extra functionalities to enhance resiliency. For example, the Availability Zones and Elastic IP Address features of Amazon.com EC2 allow multi-location of application software and dynamic IP address re-mapping mechanism in an event of service interruption.

- Highly Automated Cloud computing services are maintained by dedicated IT professionals of cloud service providers. As a result, universities IT staff no longer need to worry about complex details behind the delivered computing services, such as hardware maintenance, constant software update, etc.

4 What is vFabric SQLFire?

- vFabric SQLFire enables dynamic horizontal scaling by creating a shared pool of memory across multiple standalone physical devices or x86 rack/blade hardware, and then expanding or shrinking the pool as demand changes.[3]

- Replicated or partitioned tables can be managed in memory alone, or in memory and on disk. Application logic can be routed to the data location, for increased performance from parallel execution.

- Used as either the primary datastore or a front-end data-management layer for one or more existing databases, vFabric SQLFire assures continuous availability for data within and across datacenters. Any table can be configured to be replicated or partitioned into one or more redundant copies.

- Built on a foundation of VMware vFabric GemFire technology, vFabric SQLFire benefits from years of testing and commercial production reliability of its distribution subsystem. vFabric SQLFire includes a fast SQL query engine that compiles a query plan into byte codes, and a sophisticated cost-based optimizer. Anyone with relational database experience will find the vFabric SQLFire configuration and deployment model simple and intuitive to use and adapt. Unlike many popular data grids.

- vFabric SQLFire offers native persistence and recovery capabilities, and can be used as a distributed datastore. Its use of established standards such as SQL, JDBC and ADO.NET makes vFabric SQLFire easy to adopt by traditional database developers writing custom applications. Configuration and deployment is straightforward, and the product works effectively with a large ecosystem of compatible products and frameworks, including object-relational mapping tools (Hibernate, NHibernate, etc.), schema-editing and database management tools, Spring JDBC, and others.

- Applications using the standard SQL syntax supported by vFabric SQLFire can easily migrate to and from other relational databases, for flexibility and future-proofing as well as unparalleled performance.

5 vFabric SQLFire Architecture

5.1 Overview

vFabric SQLFire is an in-memory distributed data management platform that can be spread across many virtual machines, JVMs, and vFabric SQLFire servers to manage application data. Using dynamic replication and partitioning, vFabric SQLFire offers the following features in the platform: data durability, trigger-based event notification, parallel execution, high throughput, low latency, high scalability, continuous availability, and WAN distribution. The following figure1 shows vFabric SQLFire as the middle tier data layer that orchestrates data delivery to enterprise data consuming applications. As demand from consuming applications increases, the middle tier data layer expands to appropriately meet seasonal workloads. Further, vFabric SQLFire is a full data management system capable of managing transactions and data consistency so that enterprise data consuming application can rely on vFabric SQLFire as the system of record. [Fig.4.] For additional persistence,

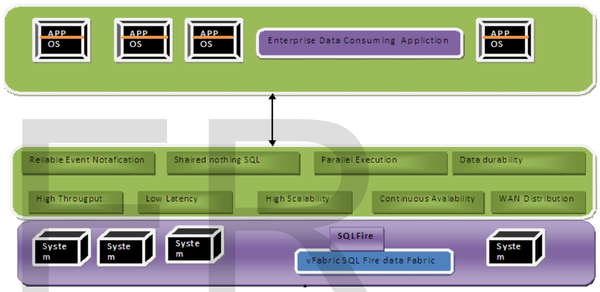

Figure 4: Enterprise Data Management with vFabric SQLFire

data can be written behind to a backup store like a relational database, or other disk stores for archival purposes. vFabric SQLFire provides full persistence durability using its own native shared-nothing persistence mechanism. Figure 5 illustrates how vFabric SQLFire can write and read synchronously and/or asynchronously to external datastores. vFabric SQLFire is based on the VMware vFabric GemFire distributed data management product and the Apache Derby project the Apache Derby project is used for its RDBMS components, JDBC driver, query engine, and network server.

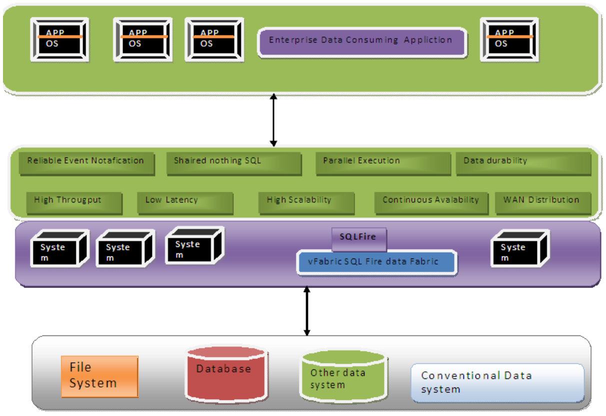

Figure 5: vFabric SQLFire Architecture with Capability to Write to Traditional Storage Systems

The partitioning technology of GemFire is used to implement horizontal partitioning features of vFabric SQLFire. vFabric SQLFire specifically enhances the Apache Derby components,such as the query engine, the SQL interface, data persistence, and data eviction, and adds additional components like SQLcommands, stored procedures, system tables,functions, persistence disk stores, listeners, and locators, to operate a highly distributed and fault tolerant data management cluster.[1]

6 VMware vFabric SQLFire Highlights

- Higher throughput and lower latency than traditional databases SQLFire manages data and indexes in memory, and is optimized for distributed data access rather than for disk I/O. The SQLFire architecture introduces no singlepoints of contention, either with disk seeks, CPU cycles, or network access.[2]

- Scales easily to meet the highest demand - SQLFire is scalable because it partitions data dynamically across many member nodes, distributing the load uniformly across servers, and making it easy to scale in or out at any time

- Provides built-in fault tolerance and horizontal scalability - In SQLFire, each member node can persist data in disk files independently of other members. Failures in disks or cache failures in one node do not affect other instances being able to safely operate on their disk files. This "shared nothing" persistence architecture allows applications to be configured such that tables can be persisted on different nodes across the cluster, reducing the impact of disk latencies.

- High availability around the globe and Disaster Recovery options SQLFire lets you synchronously replicate data at memory speed within one datacenter which makes scalability and fault tolerance simple. SQLFire can also asynchronously replicate data across the WAN to remote datacenters, which makes it easy to run the same application in different locations and meeting disaster recovery and/or regulatory requirements.

- Ideal for high transaction rates SQLFire memory-speed write performance is ideal for large-scale databases with high transaction volumes and demanding Service Level Agreements (SLAs). It is architected to manage complex transaction-oriented web portals where disk and network overheads choke optimum service delivery to customers.

- Caching options for existing data SQLFire can be used as a primary database or as a middle-tier cache for an existing database, offering read-through, write-through and write-behind caching.

- Accommodates application demand with near-linear scale SQLFire can be used as a primary database or as a middle-tier cache for an existing database, offering read-through, write-through and write-behind caching.

- Standard SQL Syntax and tools Database administrators can use and adapt SQLFire using their current skill sets and tools.

7 vFabric SQLFire Design and Sizing

Designing and sizing is a three step process shown in Fig.6 Step 1, determine entity groups, step 2, determine memory size of the data fabric, and step 3, establish building block virtual machine and JVM size, and how many vFabric SQLFire members are needed. After these steps have been concluded, a load test should be used to verify the calculation and sizing assumptions.

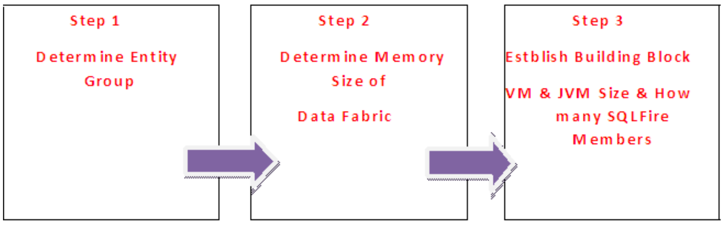

Figure 6: vFabric SQLFire Three Steps Design and Sizing Process

8 Partitioning

For custom enterprise applications backed by RDBMS schema tables that are highly accessible during the course of a regular business day, a partitioning strategy improves performance to more easily meet SLA requirements[1]. These types of frequently accessed tables are often referred to as hot tables due to their high rate of data insert, update, delete, and read operations. In addition to the high rate of data change, it is common that these hot tables also become unmanageable within a single node due to their size. In such cases, a vFabric SQLFire horizontal partitioning strategy can be used to split the data volume into smaller, more manageable data partitions. With vFabric SQLFire horizontal partitioning, an entire row is stored in the same hash indexed bucket. Buckets are the containers for data that determine its storage site, redundant sites, and the unit of migration for rebalancing. You can hash-partition a table based on its primary key or on an internally-generated unique row ID if the table has no primary key. Other partitioning strategies can be specified in the PAR-

TITION BY clause in a CREATE TABLE statement. Strategies supported by vFabric SQLFire include hash-partitioning on columns other than the primary key, range-partitioning, and list-partitioning. In Fig. 7 an example Flights table is partitioned by FLIGHTID (primary key for Flights table) into three buckets, where the first bucket for rows 1 to 3 resides on vFabric SQLFire server 1, the second bucket for rows 4 to 6 resides on vFabric SQLFire server 2, and the third bucket for rows 7 to 9 resides on vFabric SQLFire server 3. vFabric SQLFire directs that all access to flights data by the FLIGHTID primary key for rows 1 to 3 are executed on vFabric SQLFire Server 1, and similarly for the other rows. vFabric SQLFire automatically manages all partitioning through this bucketing system so long as the designer has provided the correct PARTITION BY COLUMN(FLIGHTID) clause in the table definition.

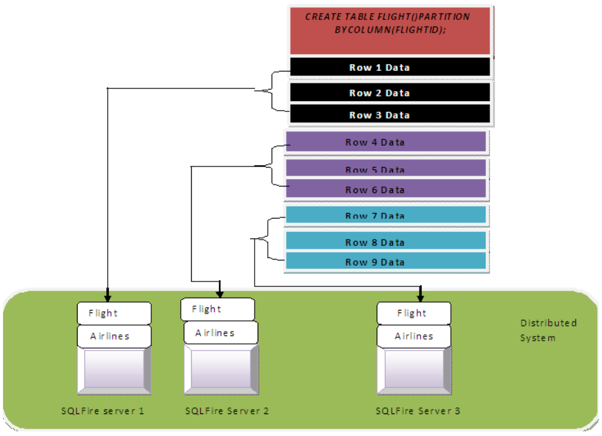

Figure 7: Partitioning Flights Table Using vFabric SQLFire Bucket System

9 CONCLUSION AND FUTURE WORK

ThrivOn's VMware vFabric Consulting Services can help accelerate your journey to Platform as a Service based on VMware vFabric solutions. Our expert vFabric consultants will help you leverage best practices for planning migrations, expand your knowledge base around vFabric solutions, develop a PaaS Plan and vFabric PaaS Roadmap, and help transform your cloud application development environment.

- ThrivOns SQL Fire Planning Service : Provides an efficient transition plan to Platform as a Service for application development. With guidance from ThrivOns vFabric consultants, you'll gain a through in-depth understanding of successful PaaS deployment factors and related process, steps, implementation and production scenarios.
- ThrivOns SQL Fire Accelerator: Provides core knowledge, expert guidance, and a non-production installation, configuration of Data Director in your environment; including a high-level architectural review.

References

[1] Emad Benjamin,Lyndon Adams, Vincent Ford, Jonathan Fullam, Mark Johnson, Catherine Johnson, Gideon Low, Jeff Markham, Sudhir Menon, Jags Ramnarayan, Mike Stolz, Lise Storc, Guillermo Tantachuco, Sumedh Wale, and James Williams. " *VMware*

vFabric SQLFire Best Practices Guide" 2012.

[2] *" VMware vFabric sqlFireFast, Scalable SQL Data in The Cloud"* 2011.

[3] *" VMware vFabric SQLFire Memory-Optimized Distributed SQL Database"*2011.

[4] *"Providing Scalable SQL data in the cloud With VMware vFabric SQL Fire"* 2011.

[5] *" Cloud Computing A newsletter for IT Professionals Issue 3 "*.

[6] James l. Johnson *" SQL in the Clouds "*2009.

[7] *" Cloud Computing Basics "*.

[8] Yu Cao , Chun Chen , Fei Guo , Dawei Jiang , Yuting Lin ,Beng Chin Ooi , Hoang Tam Vo , Sai Wu , Quanqing Xu*"A Cloud Data Storage System for Supporting Both OLTP and OLAP"*2010.