SQL Server Blog

Official News from Microsoft's Information Platform

Architecture of the Microsoft Analytics Platform System

SQL Server Team 20 May 2014 10:00 AM 3

In today's world of interconnected devices and broad access to more and more data, the ability to glean ambient insight from the variety of data sources has been made quite hard by the variety and speed with which data is being delivered. Think about it for a minute, your servers continue to provide interesting data to you about the operations happening in your business but now you have data coming to you from the temperature sensors in the A/C units, the power supplies and networking equipment in the data center that can be combined to show that spikes in temperature and traffic have a dramatic effect on the life of a server. This type of contextual data is growing to include larger and with more detailed insights into the operations and management of your business. As we look to the future, Pew Research has released a report that predicts 50 billion connected devices by 2025. That is 5 devices for every person expected to be alive. With data coming from sources like manufacturing equipment to jet airliners, from mobile phones to your scale, and to things we haven't even imaged yet the question really because how do you take advantage of all of these data sources to provide insight into the current and future trends in your business.

In April 2014, Microsoft announced the Analytics Platform System (APS) as Microsoft's "Big Data in a Box" solution for addressing this question. APS is an appliance solution with hardware and software that is purpose built and pre-integrated to address the overwhelming variety of data while providing customers the opportunity to access this vast trove of data. The primary goal of APS is to enable the loading and querying of terabytes and even petabytes of data in a performant way using a Massively Parallel Processing version of Microsoft SQL Server (SQL Server PDW) and Microsoft's Hadoop distribution, HDInsight, which is based off of the Hortonworks Data Platform.

Basic Design

An APS solution is comprised of three basic components:

1. The hardware – the servers, storage, networking and racks.
2. The fabric – the base software layer for operations within the appliance.
3. The workloads – the individual workload types offering structured and unstructured data warehousing.

The Hardware

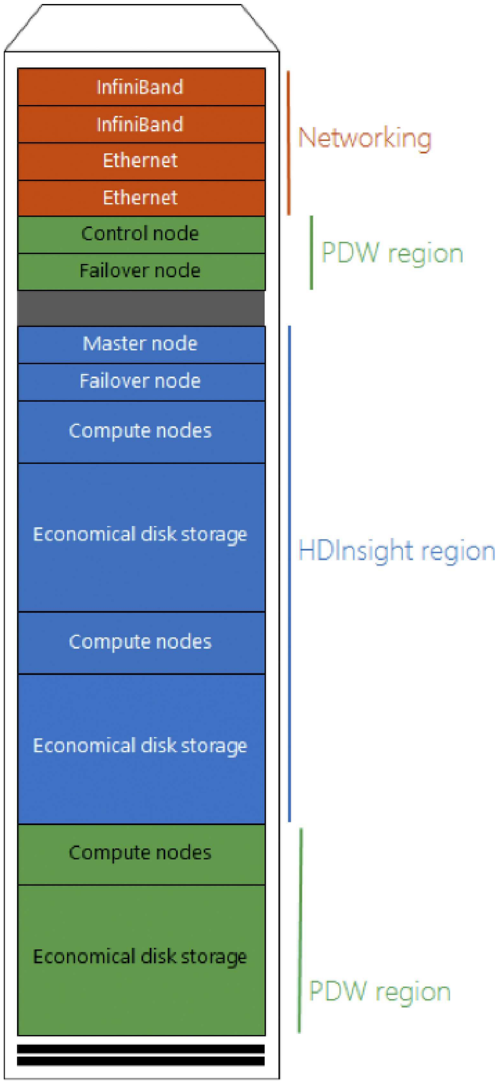
Utilizing commodity servers, storage, drives and networking devices from our three hardware partners (Dell, HP, and Quanta), Microsoft is able to offer a high performance scale out data warehouse solution that can grow to very large data sets while providing redundancy of each component to ensure high availability. Starting with standard servers and JBOD (Just a Bunch Of Disks) storage arrays, APS can grow from a simple 2 node and storage solution to 60 nodes. At scale, that means a warehouse that houses 720 cores, 14 TB of RAM, 6PB of raw storage and ultra-high speed networking using Ethernet and InfiniBand networks while offering the lowest price per terabyte of any data warehouse appliance on the market (Value Prism Consulting).

Fabric

The fabric layer is built using technologies from the Microsoft portfolio that enable rock solid reliability, management and monitoring without having to learn anything new. Starting with Microsoft Windows Server 2012, the appliance builds a solid foundation for each workload by providing a virtual environment based on Hyper-V that also offers high availability via Failover Clustering all managed by Active Directory. Combining this base technology with Clustered Shared Volumes (CSV) and Windows Storage Spaces, the appliance is able to offer a large and expandable base fabric for each of the workloads while reducing the cost of the appliance by not requiring specialized or proprietary hardware. Each of the components offers full redundancy to ensure high-availability in failure cases.

Workloads

Building upon the fabric layer, the current release of APS offers two distinct workload types – structure data through SQL Server Parallel Data Warehouse (PDW) and unstructured data through HDInsight (Hadoop). These workloads can be mixed within a single appliance offering flexibility to customers to tailor the appliance to the needs of their business.

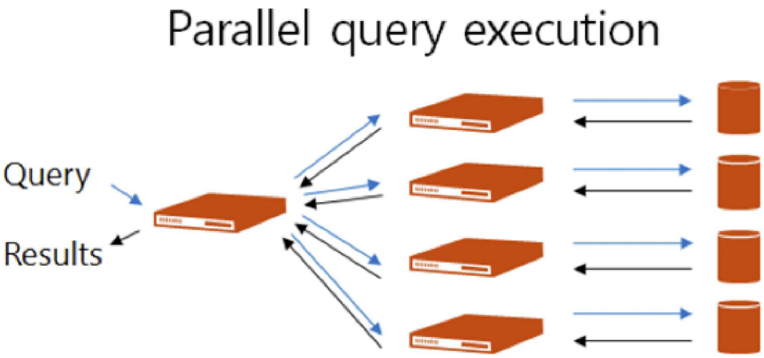


SQL Server Parallel Data Warehouse is a massively parallel processing, shared nothing scale-out solution for Microsoft SQL Server that eliminates the need to 'forklift' additional very large and very expensive hardware into your datacenter to grow as the volume of data exhaust into your warehouse increases. Instead of having to expand from a large multi-processor and connected storage system to a massive multi-processor and SAN based solution, PDW uses the commodity hardware model with distributed execution to scale out to a wide footprint. This scale wide model for execution has been proven as a very effective and economical way to grow your workload.

HDInsight is Microsoft's offering of Hadoop for Windows based on the Hortonworks Data Platform from Hortonworks. See the HDInsight portal for details on this technology. HDInsight is now offered as a workload on APS to allow for on premise Hadoop that is optimized for data warehouse workloads. By offering HDInsight as a workload on the appliance, the pressure to define, construct and manage a Hadoop cluster has been minimized. Any by using PolyBase, Microsoft's SQL Server to HDFS bridge technology, customers can not only manage and monitor Hadoop through tools they are familiar with but they can for the first time use Active Directory to manage security into the data stored within Hadoop – offering the same ease of use for user management offered in SQL Server.

Massively-Parallel Processing (MPP) in SQL Server

Now that we've laid the groundwork for APS, let's dive into how we load and process data at such high performance and scale. The PDW region of APS is a scale-out version of SQL Server that enables parallel query execution to occur across multiple nodes simultaneously. The effect is the ability to run what appears to be a very large operation into tasks that can be managed at a smaller scale. For example, a query against 100 billion rows in a SQL Server SMP environment would require the processing of all of the data in a single execution space. With MPP, the work is spread across many nodes to break the problem into more manageable and easier ways to execute tasks. In a four node appliance (see the picture below), each node is only asked to process roughly 25 billion rows – a much quicker task.



To accomplish such a feat, APS relies on a couple of key components to manage and move data within the appliance – a table distribution model and the Data Movement Service (DMS).

The first is the table distribution model that allows for a table to be either replicated to all nodes (used for smaller tables such as language, countries, etc.) or to be distributed across the nodes (such as a large fact table for sales orders or web clicks). By replicating small tables to each node, the appliance is able to perform join operations very quickly on a single node without having to pull all of the data to the control node for processing. By distributing large tables across the appliance, each node can process and return a smaller set of data returning only the relevant data to the control node for aggregation.

To create a table in APS that is distributed across the appliance, the user simply needs to add the key to which the table is distributed on:

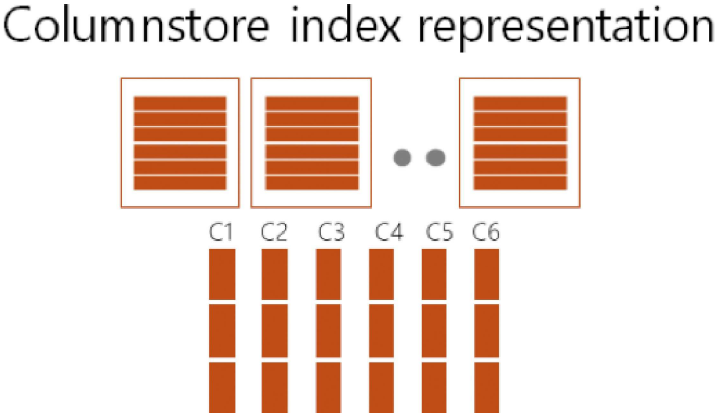
```
CREATE TABLE [dbo].[Orders]
(
   [OrderId] ...
)
WITH
(
   DISTRIBUTION = HASH([OrderId])
)
```

This allows the appliance to split the data and place incoming data onto the appropriate node onto the appropriate node in the appliance.

The second component is the Data Movement Service (DMS) that manages the routing of data within the appliance. DMS is used in partnership with the SQL Server query (which creates the execution plan) to distribute the execution plan to each node. DMS then aggregates the results back to the control node of the appliance which can perform any final execution before returning the results to the caller. DMS is essentially the traffic cop within APS that enables queries to be executed and data moved within the appliance across 2-60 nodes.

Performance

With the introduction of Clustered Column Indexes (CCI) in SQL Server, APS is able to take advantage of the performance gains to better process and store data within the appliance. In typical data warehouse workloads, we commonly see very wide table designs to eliminate the need to join tables at scale (to improve performance). The use of Clustered Column Indexes allows SQL Server to store data in columnar format versus row format. This approach enables queries that don't utilize all of the columns of a table to more efficiently retrieve the data from memory or disk for processing – increasing performance.



By combining CCI tables with parallel processing and the fast processing power and storage systems of the appliance, customers are able to improve overall query performance and data compression quite significantly versus a traditional single server data warehouse. Often times, this means reductions in query execution times from many hours to a few minutes or even seconds. The net results is that companies are able to take advantage of the exhaust of structured or non-structured data at real or near real-time to empower better business decisions.

To learn more about the Microsoft Analytics Platform System, please visit us on the web at http://www.microsoft.com/aps.

Tweet 54 Like 11 Share 57 Save this on Delicious

Comments

#
21 May 2014 6:22 AM

yep

Great

#
21 May 2014 2:30 PM

Lucas

Can you alter existing tables to be distributed? In other words, can I use
with
{
DISTRIBUTION = HASH([OrderId])
}
with alter table command?

#
21 May 2014 3:47 PM

James H

Lucas - you would likely want to use a CREATE TABLE as SELECT (CTAS) operation to change a replicated table to a distributed table.