



PDW Region

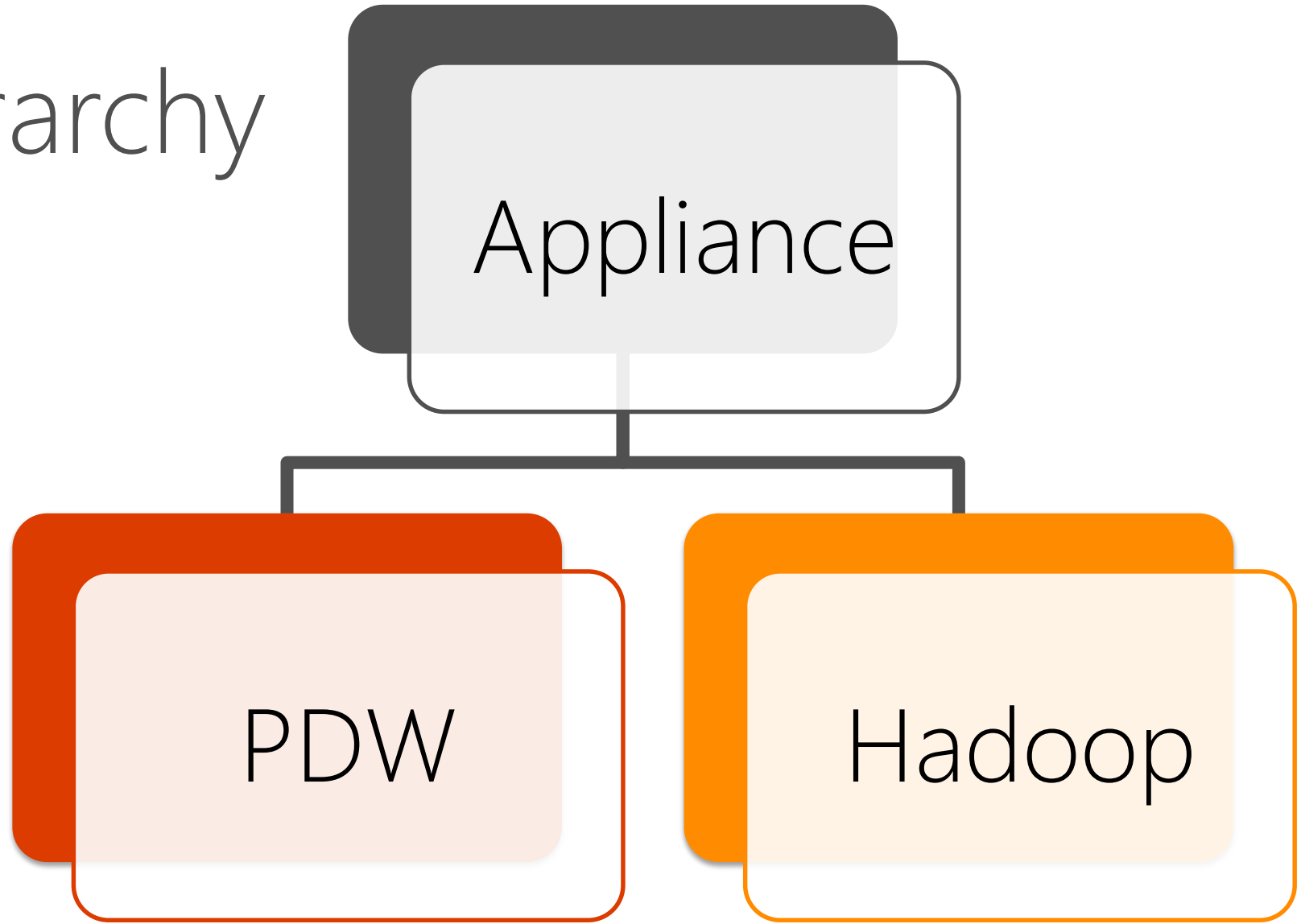
Parallel Data Warehouse

Agenda

- PDW Explained
- PDW Region Nodes
- Tools of the Trade
- Extending the Region
- Lab: PDW Region Anatomy

PDW Region Overview

Region Hierarchy



PDW Region Overview

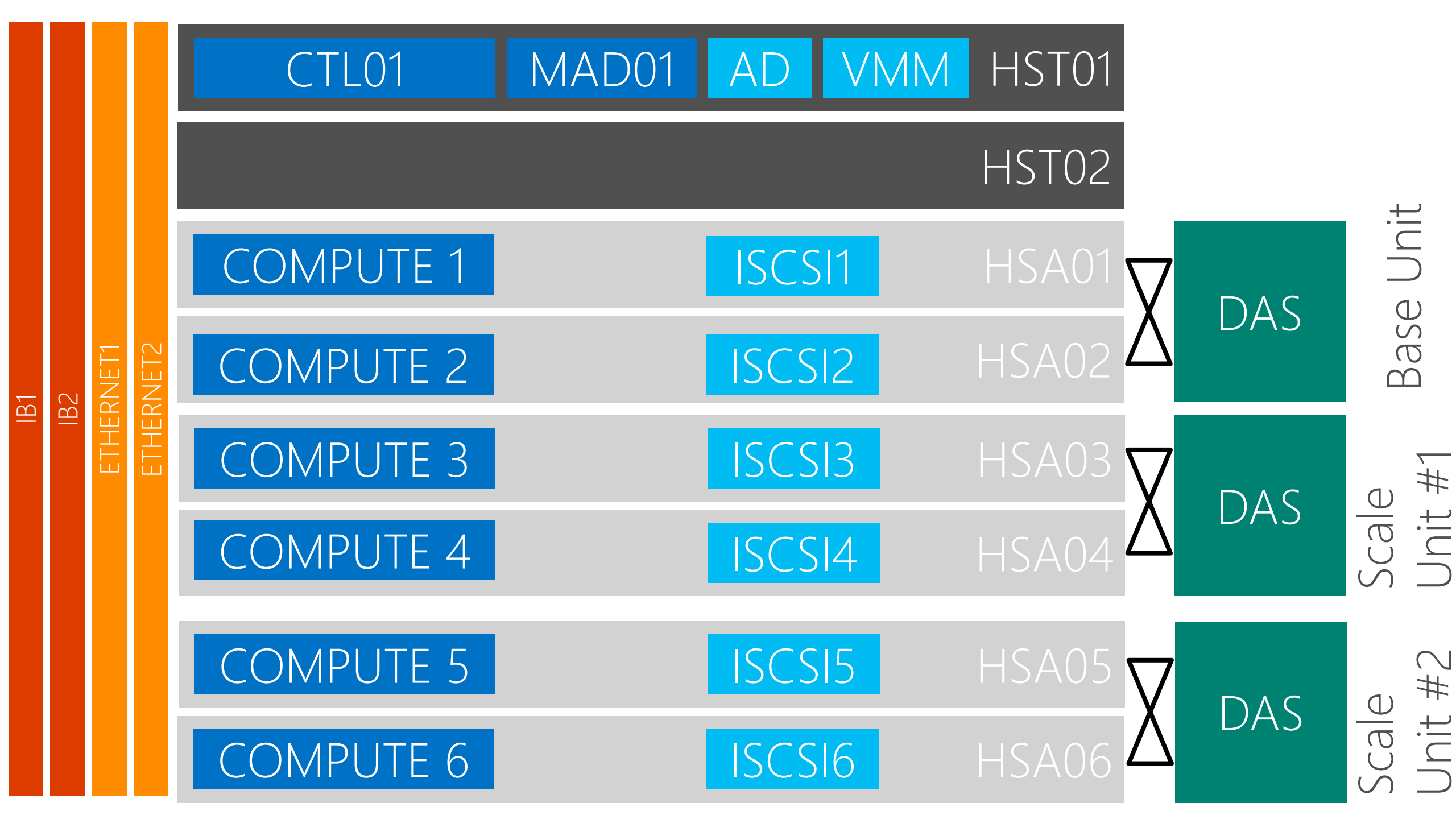
PDW is designed for...

- Data Warehouse Workloads
- Agnostic Big Data Integration
- Ease of Use
- Ease of Management

PDW Region Overview

PDW Is not designed for...

- OLTP applications
- Very High Concurrency
- Data Warehouse Hosting Consolidation
- Infrastructure as a Service (IaaS)



PDW Region Nodes

PDW Region Nodes

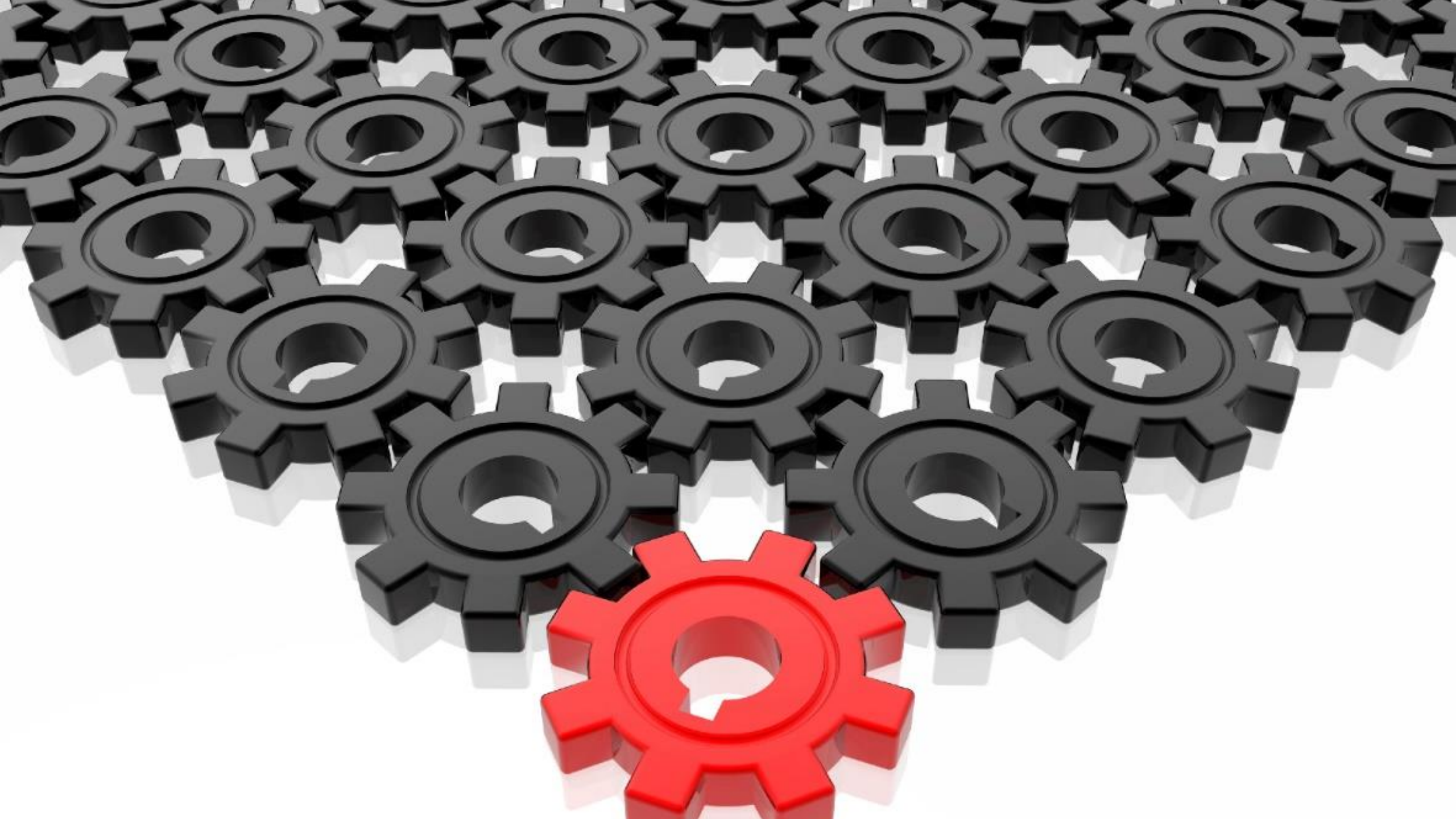
PDW Processing Nodes

- Control
- Compute (>1)

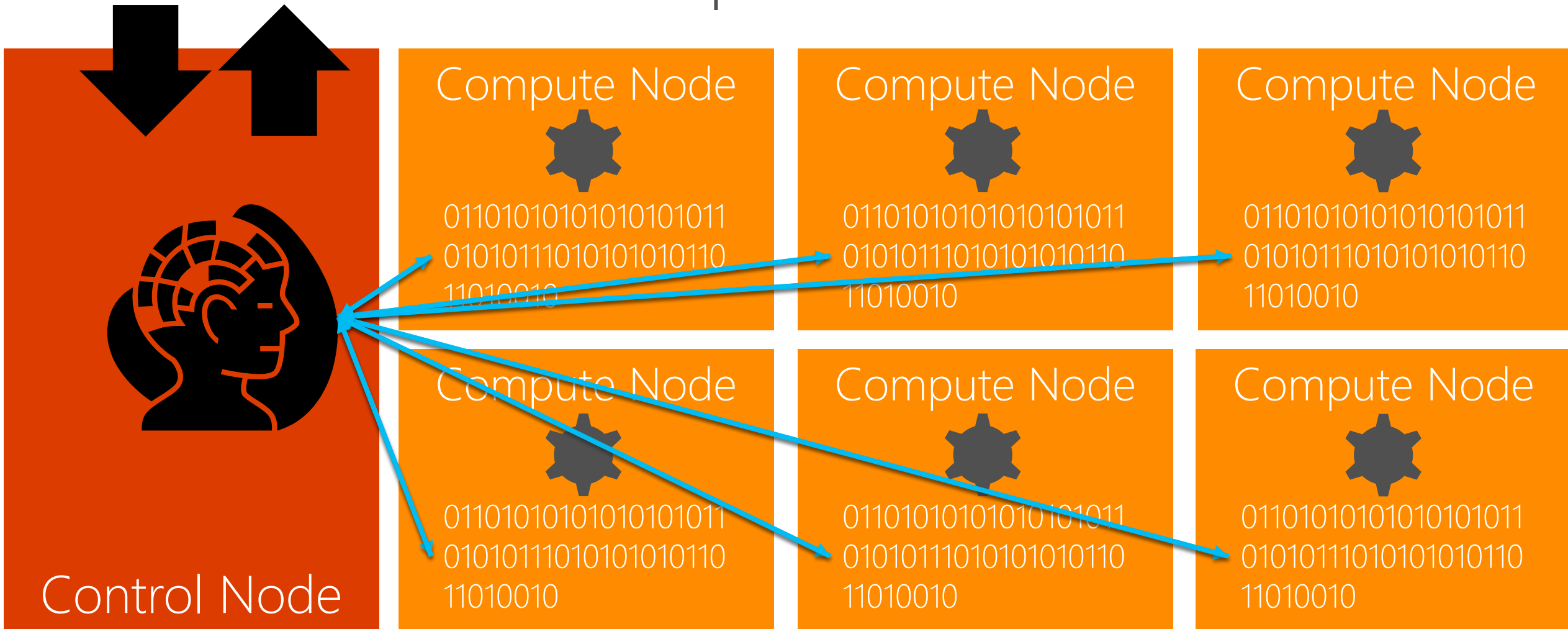
Infrastructure Nodes

- Virtual Machine Manager
- Fabric Active Directory
- Management
- ISCSI (>1)

Processing Nodes



Control and Compute Workload Nodes



Services Inside the Control Node

PDW Services

- PDW Engine
- DMS Core
- PDW Agent
- SQL Server
- Management Console (IIS)

Responsibilities

- Parse SQL / syntax check
- Validate & authorize
- Generate D-SQL plan
- Orchestrate D-SQL plan
- Centralised data processing
- Collate diagnostic Info
- Mgmt Console web app

Control Node Log Files

C:\ProgramData\Microsoft\Microsoft Sql Server
Parallel Data Warehouse\100\Logs

- <Domain>-CTL01_DWEngine_movement
- <Domain>-CTL01_DWEngine_errors
- <Domain>-CTL01_DWEngine_server
- <Domain>-CTL01_sqldwagent.exe_activity

Resources in Control Node

Virtual Machine

- 32 cores
- 230GB RAM mapped to VM
- 90GB OS volume
- 1.2TB data volume

SQL Server

- 160GB MAX server memory
- MAXDOP 0
- 2% memory grant

Services Inside the Compute Node

PDW Services

- DMS Core
- PDW Agent
- SQL Server

Responsibilities

- Hold user data
- Process queries
- Move data
- Load data

Resources in Compute Node

Virtual Machine

- 32 cores
- 238GB RAM mapped to VM
- 90GB OS volume
- 1.8TB data volume

SQL Server

- 190GB MAX server memory
- MAXDOP 4
- Resource Governor provides workload management

Infrastructure Nodes

Management (MAD01)

Active Directory for PDW Region

- Trusted with Customer Domain for Integrated Authentication

Configuration of Appliance

- Global settings
- Region specific settings

Management Tools

- dwconfig.exe
- PDW Appliance Validator (PAV)

ISCSI Node

- Part of the availability infrastructure
- Hosts quorum for Storage Spaces
- Sits in the fabric domain

Active Directory Node (AD)

Active directory for the physical infrastructure

- Also known as the Fabric

All servers in the appliance included

- PDW Region
- Hadoop Region
- Pre-configured trusts
 - Fabric AD 2-way trust PDW Region domain
 - Fabric AD 2-way trust Hadoop Region domain

Virtual Machine Manager (VMM)

- Deployment of virtual machines
- Configuration of virtual machines
- Hosts Windows Update Services (WSUS)
- Sits in the Fabric domain

PDW Tools of the Trade

Tools of the Trade

PDW Development

- SSDT
- SQLCMD
- SSIS / SSAS / SSRS
- Power BI
- SSIS adapters
- dwloader.exe

Management

- Management Console
- dwconfig.exe
- pav.exe
- PowerShell
- System Center

Tool Distribution with PDW

Location

- Control Node
 - C:\PDWINST\ClientTools\

Top Tip!

Join the PDW Advisors group on Yammer.

The documentation team keep the chm file up to date there so you are assured of the latest version

Tools Distributed

- SSIS Destinations
- Clienttools.msi
 - dwloader
 - Dwsql
- Adventureworks
- Help File

Solution Development

- Power Pivot
- Power Map
- Power Query
- Power View

Connecting To PDW

Management Console

<https://Control Node IP Address/>

Development tools

- TCP Port 17001
- Security

Connecting
from SQL
2012?

Use SNAC 11

Connecting
from SQL
2008 R2?

Use SNAC 10

File Home Insert View

Apply & Refresh Close Query

Remove Top Rows Reduce

Remove Columns

Sort

Split Column Group By Transform

Replace Values Data Type: Fill Down Use First Row As Headers Unpivot

Merge Queries Append Queries Combine

Help

Navigator

	Name	Data	Schema	Item	Kind
1	call_center	Table	dbo	call_center	Table
2	catalog_page	Table	dbo	catalog_page	Table
3	catalog_returns	Table	dbo	catalog_returns	Table
4	catalog_sales	Table	dbo	catalog_sales	Table
5	customer	Table	dbo	customer	Table
6	customer_address	Table	dbo	customer_address	Table
7	customer_demographics	Table	dbo	customer_demographics	Table
8	date_dim	Table	dbo	date_dim	Table
9	household_demographics	Table	dbo	household_demographics	Table
10	income_band	Table	dbo	income_band	Table
11	inventory	Table	dbo	inventory	Table
12	item	Table	dbo	item	Table
13	promotion	Table	dbo	promotion	Table
14	reason	Table	dbo	reason	Table
15	ship_mode	Table	dbo	ship_mode	Table
16	store	Table	dbo	store	Table
17	store_returns	Table	dbo	store_returns	Table
18	store_sales	Table	dbo	store_sales	Table
19	time_dim	Table	dbo	time_dim	Table
20	vBackups	Table	dbo	vBackups	View
21	vDatabaseRoleMembers	Table	dbo	vDatabaseRoleMembers	View
22	vDatabaseSize_SpaceUsed	Table	dbo	vDatabaseSize_SpaceUsed	View
23	vDB_SpaceUsed	Table	dbo	vDB_SpaceUsed	View

Query Settings

PROPERTIES

Name

Query1

Description

APPLIED STEPS

Source

LOAD SETTINGS

☒ Load to worksheet

SSDT

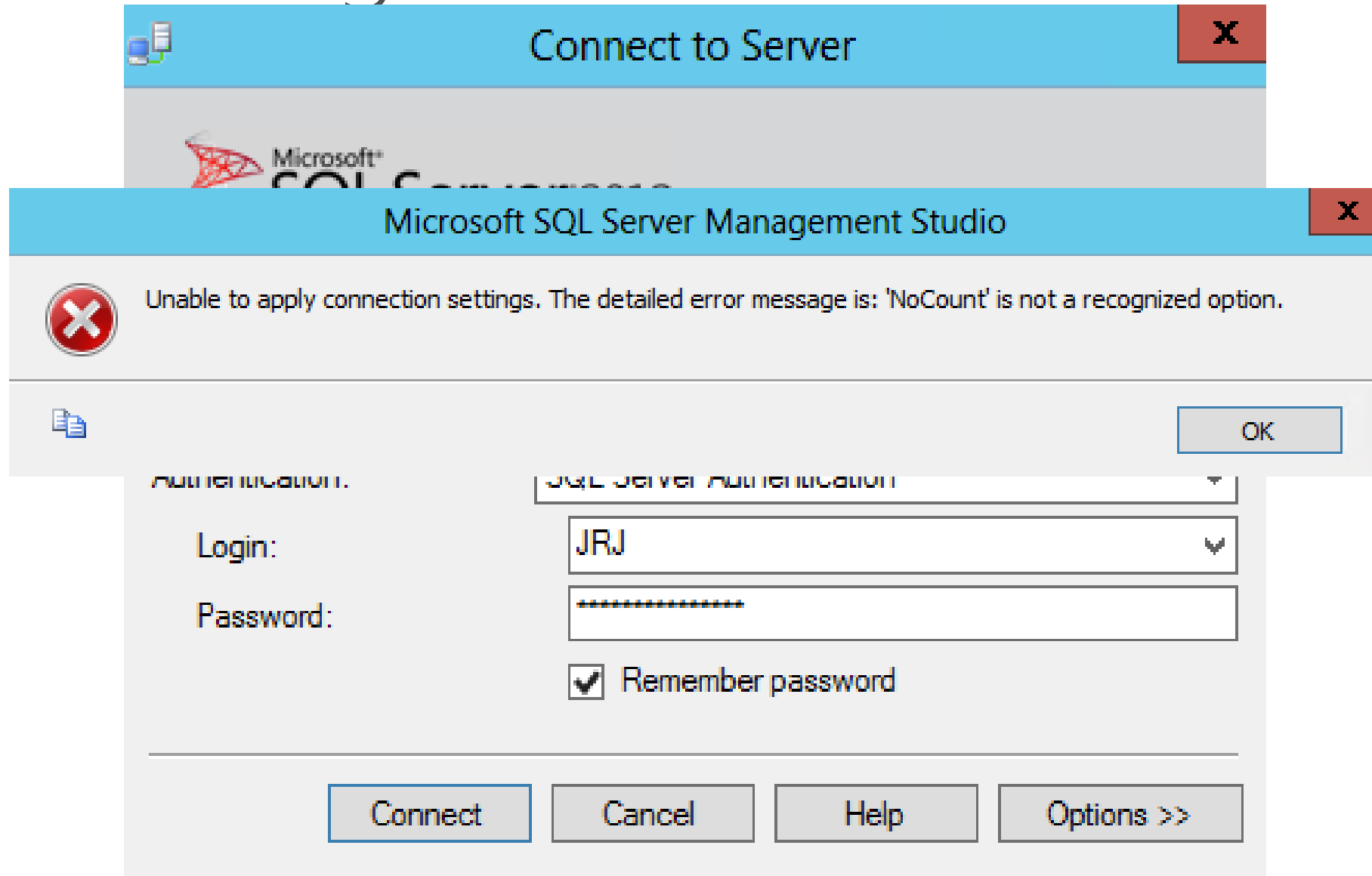
- SQL Server Data Tools
- Used for writing queries against PDW
- SSMS is not a supported tool
- SSDT for PDW part of standard SSDT deployment model

SSMS: How to use it if you need it

Here's how you use it...

- Object Explorer Tree Control does not work
- You *can* run a basic query using SSMS

Connecting via SSMS



SSDT Support

PDW Version	VS 2012	VS 2013	SSDT Update
RTM			Dec 2012
AU1			Jan 2014

.dsql

Your SQL Scripts must end in .dsql for SSDT to recognise them as PDW scripts!

Feature	Supported?
Tree Control	
Query Window	
Graphical Execution Plan	
Database Project System	
Distributed SQL Execution Plan	
Object Script Generation	
Object Script Wizard	
Unit Test	
IntelliSense	
Localdb	
Code Analysis	
Schema / Data Comparisons	

SQLCMD Support

- ..\Microsoft SQL Server\110\Tools\Binn\
- SNAC 10 2008 R2
- SNAC 11 for 2012
- QUOTED IDENTIFIER ON ← Mandatory
- Must be set at SQLCMD Invocation

```
SQLCMD.exe -I
```

Extending the PDW Region

Extending the PDW Region

Extend PDW Region by adding Data Scale Units

VMM solely used for deployment

- Host provisioning
- Creation of virtual machines

Supported Extensions

All actions performed by CSS

Replace Node

- Hardware failure

Replace VM

- VM corruption

Add Unit

- Data Scale Unit

Add Region

- Hadoop Region only

AU1 Enhancement to PDW Add Unit

- PDW Region now online during hardware installation and domain configuration
- Only offline during data re-distribution phase

PDW Region must have enough free space to re-distribute the largest table.

