Introduction to the Cluster Metadata Store (CMS)

Last updated by | Subbu Kandhaswamy | Sep 17, 2020 at 6:27 PM PDT

Contents

Cluster Metadata Store (CMS)

Overview

The Cluster Metadata Store (CMS) provides a single logical point of metadata storage for a SAWA v2 cluster. Located on the Cluster Control Manager node of the control ring, the store provides the source of 'truth' regarding the state of the cluster and its resources. Each resource being managed in the cluster is represented by a row in a table in CMS that reflects the state of that resource. CMS provides the persistence for state-machine driven workflows driven from within the Management Service that control all updates to resources and which ensures that the metadata either reflects a known stable state of a resource or that a workflows is in-progress currently modifying the resource.

Each type of resource that requires management is represented by a table in the store with each instance of the resource represented by a row in that table. Metadata describing the resource is held in columns on the table. Lower level resources that are managed by Windows Fabric exclusively, such as the location and state of replicas is not represented in the store. Some resources are purely logical. For example, Logical Server is an important user abstraction that hides the location of a database on an instance, allowing databases to be distributed across multiple instances but presented to the user as co-located for management purposes.

As a governing principle of the SAWA v2 architecture, all creation, update and deletion of resources and their associated metadata is mediated by managed workflows that are tightly integrated with the metadata store. Fine-grained state machines are defined for each resource type and are instantiated at runtime to manage workflows impacting those resources. State machines describe stable states and unstable or work-in-progress states. State-related information for each resource is also persisted to the resource table.

At runtime a running state machine is instantiated for each resource being operated on based on the data held in the metadata store. This state machine instance mediates creation, all updates and deletion of the resource. All updates to resources and metadata in the store are handled by actions. Actions can be triggered by an event or automatically on entering into an unstable state. An action may transition the resource into a new state based on its outcome. An action on one resource may raise events on other resources allowing workflows to cascade across multiple related resources. The use of unstable states, automatic actions and branching transitions allows complex workflows to be expressed and managed at runtime. Workflows are triggered by an event on a resource in a stable state and continue until all resources affected by the workflow are in a stable state.

Through the use of unstable states and by persisting every state change to the store, a workflow can be designed to retry if resources are temporarily unavailable, and can be restarted and continued in the event of a failure of one or more executing state machines. Workflows can be designed to detect and automatically recover from errors, if necessary, reversing partial changes to resources.

CMS is he source of truth for the state of all SQL-related resources being managed on a cluster, however CMS is not a cache.

CMS Features

- Source of truth for the state of all SQL-Related resources being managed in the cluster.
- Not a cache. The control ring architecture and connectivity model do not require the metadata store to be involved in the user login pipeline or subsequent SQL execution at runtime.
- It is neither updated optimistically before resource updates nor after the fact.
- For any individual resource CMS reflects either the current state of the resource or that a workflow is in progress intending to create, update or delete the resource.
- CMS supports the metadata requirements of all internal cluster management services and user-focused login and provisioning services.
- CMS enables state-machine based workflows through persisting state information. By persisting transient state changes in durable storage, CMS allows workflows to be resilient to unavailable resources, and enables workflow restart in the event of a failure of a workflow controller, the CCM node or the CMS database or the control ring as a whole.
- CMS is not required for or referenced during user login to a database; user login must succeed even if the CMS store is unavailable.
- CMS is query-able from off-cluster for management, monitoring and troubleshooting.
- CMS is secured so that only authorized users or components can update or query data. The store enables and where required enforces adherence to rules governing PII or other sensitive data.
- CMS is monitoring to provide insight into the immediate health and performance of the store (distinct from monitoring the resources represented in the store).
- Full telemetry data is available from all update actions and state changes allowing off-cluster troubleshooting, audit and performance analysis.

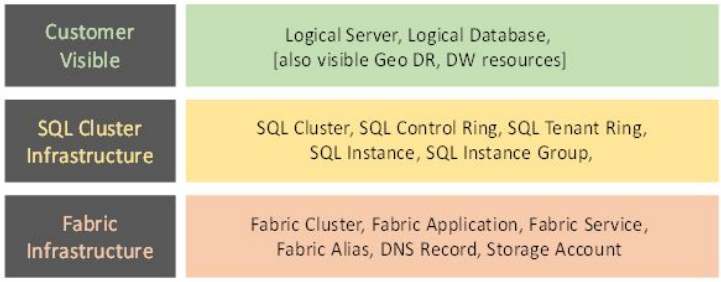
- CMS interactions are designed in a way that it would not affect scenarios requiring performance goals.

Availability

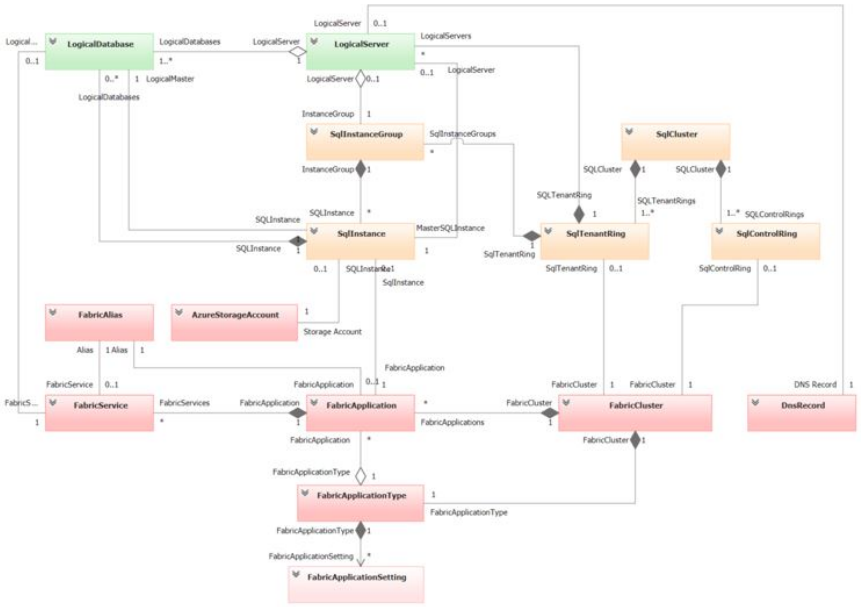
- CMS is a single point of failure in the SAWA v2 architecture. If CMS is unavailable, server and database provisioning and management will be unavailable. It is critical that it achieve maximum reliability.
- As the control ring runs on public compute on SKUs without local database storage, CMS will be implemented on a Remote Storage SQL Database and will be subject to the availability inherent in that storage model.
- As CMS uses Remote Storage, the database files are mastered in Azure Storage and must leverage Azure storage-provided geo-replication.
- CMS is protected from CCM node or Control Ring failure through WinFab.
- CMS is protected from data center disaster and can be recovered from remote backups in the event of failure
- All CMS content must be backed up off-cluster and available for recovery in the event of a cluster failure.

Domain Model

The key concepts that are represented in CMS can usefully be represented in a domain mode. The model can be thought of as organized in layers reflecting the architectural structure of a cluster.



Picture 1



Picture 2

The domain model reflects the structural constraints that apply to a steady-state running cluster after initialization.

CMS v1 Schema

State Machine Data

The following columns are included in all tables managed by a state machine.

Column Name	Data Type	Nullable	Default	Description
state	nvarchar(128)	False		The current state in the FSM state machine.
stable	bit	False		Indicates that the current state is unstable (triggers an automatic action).
error	bit	False		Indicates that the current state is declared as an error state.
request_id	uniqueidentifier	True		The active request id. Populated while the instance is participating in a workflow. This allows correlation of instances of different state machines that are participating in the same request. The value is provided by the client – typically will be provided by RDFE over the RP API. The value is set when an instance joins a workflow, either as the initial state machine or one that joins as a successor through instantiation or an event raised in an action executed on another instance already in the workflow. The value is reset to null when the state machine returns to a stable state.
workflow_position [new]	hierarchyid	True		Hierarchically structured identifier of the position, relative to other state machine instances in the current workflow, at which this instance last received an event or was created. The value is reset to null when the state machine returns to a stable state.

Column Name	Data Type	Nullable	Default	Description
next_successor	int	False	0	The id used for the next successor in a workflow. When an event is raised in an action on another state machine or a state machine is initialized, this value is appended to the current workflow_position to create the workflow position of the successor. The value is then incremented by the FSM framework. The value is reset to 1 when the state machine returns to a stable state.
create_time [new]	datetime2	False		The time the record was created.
last_update_time	datetime2 [was datetime]	False		The time the record was last updated. Has the same value as CreateTime when first created.
last_state_change_time [new]	datetime2	False		The time the state was last changed. May be different from last_update_time. The time is not changed if a transition flows recursively to the same state.
concurrency_token	bigint	True		A value that is changed whenever the row is changed.
process_id [new]		True		The fabric id of the process containing an FSM instance that is currently actively managing the state machine on this row. Is populated when a long running action is in progress that is not-interruptible. Allows a running state machine to be picked up by the same process if the

Column Name	Data Type	Nullable	Default	Description
				process fails and is restarted by WinFab.
last_exception [was exception]	xml	True		Details of the last exception raised during execution of the state machine on this instance. Cleared on next successful action or state transition on the instance.

Customer Visible Data

Logical Databases

Table: tbl_logical_databases

View: logical_databases

FSM: LogicalDatabaseStateMachine

Contains one row per database. Includes Logical Master and User databases.

	Column Name	Data Type	Nullable	Default	Description
PK	logical_server_name [new]	nvarchar(128)	True		The logical serve that contains the database. [persisted computed colum based on followi -> SQL Instance- SQlL InstanceGrc ->LogicalServer]
PK	logical_database_id [new]	uniqueidentifier	False	newsequentialid()	Globally-scoped database Id. Allo tracking cross-cluster database moves or relationships, for example.
	sql_instance_name was InstanceName]	nvarchar(128)	False		Identifies the SQ instance in which the database is hosted. This value may change if the database edition changed betwee Premium and Standard which v cause the databa to be moved between instance Will also change the database is moved between instances to load balance instance
	logical_database_name	nvarchar(128)	False		The name of the database.
	dropped_time	datetime2	True		The time this database was dropped or

	Column Name	Data Type	Nullable	Default	Description
					deferred droppe This is included i an unique index with logical_server_na database_name t enforce name uniqueness but allow database names to be reus once dropped.
	edition	nvarchar(25)	False	Standard	Standard | Premi
	max_size_bytes	bigint	False	1048576 (1 MB)	Database size quota. If the database exceed this size it is capp and allows only read and delete queries. Permitted values defined by a set client side busine rules that current allow 20MB, 100 1GB, 5GB, 10GB, 20GB, 30GB, 40G 50GB, 100GB, 150GB. These rul are not enforced the database.
	service_level_objective	nvarchar(25)	False	Shared	Shared|P1|P2|P2| Set of values change over time P11-P14 under consideration.
	fabric_cluster_name [was ClusterName]	nvarchar(128)	True		The fabric cluste containing the database service

	Column Name	Data Type	Nullable	Default	Description
					Part of the datab service key.
	fabric_service_uri	nvarchar(256)	True		The URI of the database service Combined with fabric cluster provides the fore key to the databa service.

Constraints:

Unique (SQLInstanceName & Name)

Logical Database Features

Table: tbl_sql_database_features

View: sql_database_features

FSM: DatabaseFeatureStateMachine

Contains one row per feature enabled on a database. Used for allow_listing features on a specific database.

Name	Data Type	Nullable	Default	Description
database_id	bigint	False		The id of the database having the feature.
feature_name	nvarchar(50)	False		The name of the feature being described.
feature_value	sql_variant	False		The feature value for this database.

GeoDR Relationships [was GeoDR Relations]

Table: tbl_geodr_relationships

View: geodr_relationships

FSM: GeodrRelationshipStateMachine

Contains one row per GeoDR relationship. Each row describes a replication relationship between a source database defined in the SQL Cluster and another database that may be in the current SQL cluster or some other SQL Cluster.

Note: *Placeholder table schema is based on a modified version of the SAWA v1 Gateway Metadata schema and is subject to review.*

	Column Name	Data Type	Nullable	Default	Description
PK	source_database_id	bigint	False		
PK	replica_sql_cluster_name	nvarchar(128)	False		The SQL cluster in which the replica is located. May be a reference to the current or different SQL cluster.
PK	replica_database_guid	bigint	False		
	link_id	uniqueidentifier	False		
	geo_replica_state	nvarchar(64)	True		
	geo_replica_substate	nvarchar(64)	True		
	copy_complete_subscription_id	uniqueidentifier	False		
	terminate_complete_subscription_id	uniqueidentifier	False		
	is_created_with_database	bit	False	0	

Logical Servers

Table: tbl_logical_servers

View: logical_servers

FSM: LogicalServerStateMachine

Contains one row per customer (logical) server.

	Column Name	Data Type	Nullable	Default	Description
PK	name	nvarchar(128)	False		Logical server name.
	logical_server_id	uniqueidentifier	False	newsequentialid()	A guid assigned to the
	sql_instance_group [was InstanceGroupName]	nvarchar(128)	True		The SQL instance group that contains the SQL instances that host the databases in this logical server.
	control_ring_dns_name	nvarchar(256)	True		<verify usage, nullable?>
	customer_subscription_id [new]	uniqueidentifier	False		The SQL customer subscription Id under which this logical server was created.
	admin_login_name [was AdminName]	nvarchar(128)	True		The admin user login name assigned by the customer. A temporarily stored value, only

	Column Name	Data Type	Nullable	Default	Description
					stored while being applied to the database.
	admin_login_password [was AdminPassword]	varbinary(1024)	True		The admin password. Encrypted. A temporarily stored value, only stored while being applied to the database.

SQL Cluster Infrastructure

SQL Clusters [was Global Metadata]

Table: tbl_sql_clusters

View: sql_clusters

FSM: SqlClusterStateMachine

Contains one row representing the SQL cluster itself. Contains metadata that is either specific to the SQL cluster or is global to the SQL Azure environment.

	Column Name	Data Type	Nullable	Description
PK	name	nvarchar(128)	False	The SQL cluster name.
	primary_control_ring_name [was PrimaryControlClusterName]	nvarchar(128)	False	
	add_remove_control_ring_sequence	bigint	False	
	security_principal_sequence	Bigint	False	

SQL Control Rings [was Control Hosted Services]

Table: tbl_sql_control_rings extends tbl_fabric_clusters

View: sql_control_rings

FSM: SqlControlRingStateMachine

Contains one row per control ring in the SQL cluster.

Initially only one control ring will exist per SQL Cluster thus only one per row per CMS database. The North Star architecture allows for multiple control rings to exist in a single SQL cluster. A primary ring contains the active CMS which is the only stateful service on the control ring. Other control ring services can be accessed at either ring. This provides protection in the case that a stateless service in the ring is unavailable or overloaded.

	Column Name	Data Type	Nullable	Description
PK	name [was ClusterName]	nvarchar(128)	False	Control ring name. This is the DNS name of the fabric cluster that forms the control ring. <example>
	add_remove_sequence	bigint	True	
	processed_alias_uri	nvarchar(256)	True	
	dns_record_id	uniqueidentifier	True	
	is_bootstrap	bit	False	
	ipv4_address	nvarchar(15)	True	

SQL Instances [was Instances]

Table: tbl_sql_instances

View: sql_instances

FSM: SqlInstanceStateMachine

Contains one row per SQL Server instance running in a tenant ring and composed within a SQL Instance Group. Includes both Remote Storage and Local Storage instances.

	Column Name	Data Type	Nullable	Description
PK	name	nvarchar(128)	False	The name of the SQL Instance, constructed from the logical server name + a suffix indicating either the hosted premium database or a count of SQL instances in the SQL Instance group associated with the logical server, where 0 indicates this is the master instance containing the logical master database for a logical server. Examples H1679ab12\0 – the master SQL instance for a logical server H1679ab12<database guid> - an instance containing a premium database
	sql_instance_id [was UniqueId]	uniqueidentifier	False	
	Type	nvarchar(128)	False	The type of SQL instance CMS SQL.LocalStorage SQL.RemoteStorage FQS.Compute FQS.Control
	service_level_objective	nvarchar(128)	False	The service level objective assigned to the instance which acts as a size for the instance and determines the way the Windows Fabric

	Column Name	Data Type	Nullable	Description
				allocates instances to a node. Values are: Shared|P1|P2|P2|P4
	target_service_level_objective	nvarchar(128)	True	The target service level objective, used when changing the size of the instance. Values are:Shared|P1|P2|P2|P4
	worker_service_affinity_tag	nvarchar(128)	True	<Is this a general purpose affinity tag that should be in fabric_applications?>
	storage_account_affinity_tag	nvarchar(128)	True	
	sql_instance_group_name [was InstanceGroupName]	nvarchar(128)	True	The SQL Instance Group in which this SQL instance participates.
	security_principal_id	bigint	True	
	collation	nvarchar(128)	True	
	trace_flags	xml	True	
	reset_admin_login_name [was ResetAdminName]	nvarchar(128)	True	
	reset_admin_login_password [was ResetAdminPasswordHash]	varbinary(1024)	True	
	reset_admin_login_time [was ResetAdminDate]	datetime2	True	
	tenant_ring_name [was ClusterName]	nvarchar(128)	True	Local fabric cluster DNS name (bound to a Worker service) <clarify> <verify

	Column Name	Data Type	Nullable	Description
				column length vs. DNS name>
	fabric_application_uri	nvarchar(256)	True	With tenant_ring_name identifies the application that implements this SQL Instance.
	dns_name	nvarchar(256)	True	Global DNS (bound to Control services) <clarify>
	fabric_cluster_dns_name [was ClusterDnsName]	nvarchar(256)	True	<clarify wrt FabricClusterName, DNS name>
	fabric_service_uri	nvarchar(256)	True	Current Winfab service Uri
	fabric_service_sequence	bigint	True	
	storage_account_dns_name	nvarchar(128)	True	Current placement
	storage_container_name	nvarchar(64)	True	
	storage_policy_name	nvarchar(128)	True	
	master_key_alg	nvarchar(32)	True	
	master_key_length	int	True	
	master_key	[was MasterKeyBlob]	varbinary(MAX)	True
	processed_database_name	nvarchar(128)	True	Transient reference used during drop or other ordered workflow processing over multiple databases.

	Column Name	Data Type	Nullable	Description
	extension_data	xml	True	DW sourced <not here?>

SQL Instance Groups [was Instance Groups]

Table: tbl_sql_instance_groups

View: sql_instance_groups

FSM: SqlInstanceStateMachine

Contains one row per SQL Instance Group. Includes a row for CMS in the control ring to support holding storage group and keys.

Column Name	Data Type	Nullable	Description
name	nvarchar(128)	False	The name of the SQL Instance Group. For SQL database is either 'CMS' or the corresponding logical server name.
members [delete]	xml	True	Used to manage locking the set of databases in the absence of an efficient query and lock behavior for a set of state machines in FSM
security_principal_id	bigint	True	
security_principal_password [was Password]	varbinary(1024)	True	
extension_data	xml	True	Reserved for FQS use only. May be revised later.

SQL Tenant Ring [was Worker Hosted Services]

Table: tbl_sql_tenant_rings

View: sql_tenant_rings

FSM: SqlTenantRingStateMachine

Contains one row per tenant ring in the SQL cluster. Extends tbl_fabric_hosted_services.

There may be multiple tenant rings per SQL cluster.

Column Name	Data Type	Nullable	Description
name [was ClusterName]	nvarchar(128)	False	The tenant ring name. This is the name of the fabric cluster that forms the control ring.<example>
certificate	varbinary(MAX)	True	Certificate used for ring-wide encryption of data. Not used in OneBox implementations.
initial_counts	xml	True	Initial allocation of nodes per node role in the tenant ring. Will be used to assist in allocation of nodes to VMs during placement. Initially, SQL DDC will use the initial values as the max.
current_counts	xml	True	Current count of nodes of each role.
used_counts	xml	True	
target_counts	xml	True	Target count. Used to allow buffering for expansion. The tenant ring state machine will grow the ring or shrink the ring as required if the target counts differ from the current counts.
placement_base_weight	int	False	
placement_affinity_tag	nvarchar(128)	True	A tag used in placement expressions to ensure only control ring applications are placed on the control ring.

Constraints:

Secrets

Table: tbl_secrets

View: secrets

FSM: SecretsStateMachine

Contains one row per individual secret.

Column Name	Data Type	Nullable	Description
consumer_name	nvarchar(128)	False	Name of the consumer, could be instance or instance group
consumer_type	nvarchar(128)	False	Consumer type
secret_type	nvarchar(128)	False	Type of Secret. Example: Cert, SAS key, Symmetric key, etc.
generation_id	int	False	Generation or version of the secret
state	nvarchar(128)	False	Indicates state of the FSM
thumbprint	nvarchar(128)	True	Primary generation number. Used during rollover of secret
public_blob	nvarchar(MAX)	True	Public key of the blob (could be null for symmetric keys)
private_blob	Nvarchar(MAX)	False	Private key blob
issuer_name	nvarchar(128)	False	Who issued this key – to track changes to issuer
issuer_type	int	False	Type of issuer
properties	XML	True	Customer properties for this secret (e.g. key size, expiration, policy or container name for SAS keys, etc.)
marked_as_deleted	Bit	False	A bit is set if the secret should be deleted from clients

Instance Group Secrets

Table: tbl_instance group_secrets

View: instance group_secrets

FSM: InstanceGroupSecretsStateMachine

Contains one row per instance group secret.

Column Name	Data Type	Nullable	Description
name	nvarchar(128)	False	Name of the instance group
state	nvarchar(128)	False	Indicates state of the FSM
previous_generation	int	True	Previous generation number. Used during rollover of secret
current_generation	int	False	Current generation number.
next_generation	int	True	Next generation number. Used during rollover of secret
primary_generation	int	False	Primary generation number. Used during rollover of secret

Instance Secrets

Table: tbl_instance secrets

View: instance secrets

FSM: InstanceSecretsStateMachine

Contains one row per instance secret.

Column Name	Data Type	Nullable	Description
name	nvarchar(16)	False	Name of the instance
state	nvarchar (128)	False	Indicates state of the FSM
previous_generation	int	True	Previous generation number. Used during rollover of secret
current_generation	int	False	Current generation number.
next_generation	int	True	Next generation number. Used during rollover of secret
primary_generation	int	False	Primary generation number. Used during rollover of secret
current_app_property	nvarchar(128)	True	Current App property name

Infrastructure

Azure Storage Accounts

Table: tbl_azure_storage_accounts

View: azure_storage_accounts

FSM: StorageAccountStateMachine

Contains one row per storage account.

	Column Name	Data Type	Nullable	Description
PK	dns_name	nvarchar(128)	False	account name, (e.g. tomtalbay.blob.core.windows.net)
	subscription_id	nvarchar(128)	False	
	storage_account_name	nvarchar(128)	False	Account name of subscription
	containers_used	int	False	Number of containers used.
	primary_key	varbinary(4000)	True	Primary account key
	secondary_key	varbinary(4000)	True	Secondary account key
	use_primary_key	bit	False	Indicates if the primary key should be used for SAS signatures.
	placement_base_weight	int	False	Weight used for allocation.
	placement_affinity_tag	nvarchar(128)	True	Affinity tag.

Constraints:

Unique: (subscription_id & storage_account_name)

Azure DNS Records

Table: tbl_fabric_aliases

View: fabric_aliases

FSM: FabricAliasStateMachine

Contains one row per DNS Record.

TBD

Fabric Aliases

Table: tbl_fabric_aliases

View: fabric_aliases

FSM: FabricAliasStateMachine

Contains one row per Windows Fabric alias.

	Column Name	Data Type	Nullable	Description
PK	fabric_service_uri	nvarchar(256)	False	The exposed alias URI.
	add_remove_control_service_sequence	bigint	False	Sequence number for add remove events
	target_fabric_cluster_name	nvarchar(128)	False	The cluster name of the target fabric service.
	target_fabric_service_uri	nvarchar(256)	False	The service URI component of the underlying fabric service identifier.

Fabric Application

Table: tbl_fabric_applications

View: fabric_applications

FSM: FabricApplicationMachine

Contains a row for each Fabric Application.

	Column Name	Data Type	Nullable	Description
PK	fabric_cluster_name [was ClusterName]	nvarchar(128)	False	The name of the fabric clus which this application is loc
PK	fabric_application_uri	nvarchar(256)	False	
	deployment_id_sequence	bigint	True	Sequence number used to t progress of a redeployment of interruption.
	application_type_name	nvarchar(256)	False	Name of the fabric applicat
	current_application_type_version	nvarchar(256)	True	The current fabric applicatic version of this application. I null if the application has n been created.
	target_application_type_version	nvarchar(256)	False	The target fabric applicatio version. This is different fro current version if there is ar upgrade in progress, or if th an upgrade in progress tha
	deployment_parameters	xml	True	Parameter values that are substituted into the applica manifest during windows fa application provisioning. Parameters used vary by ap type.
	upgrade_policy	xml	True	Transient state during upgr Upgrade policy to use wher WinFab to kick off the upgr Essentially XML-serialized UpgradePolicyDescription.
	rollout_key	nvarchar(256)	True	Transient state during upgr of the currently in-progress for this app instance.
	last_upgrade_start_time	datetime2	True	Start time of the last upgrac the application instance.

	Column Name	Data Type	Nullable	Description
	last_upgrade_progress_change_time	Datetime2	True	Last time the upgrade prog changed for this applicatior instance.
	upgrade_progress	xml	True	Xml serialization of WinFabApplicationUpgrade for this application instance

Fabric Application Settings

Table: tbl_fabric_application_settings [TBD]

View: fabric_application_settings [TBD]

FSM: FabricApplicationTypeMachine

Contains deployment configuration settings for a specific application type.
Schema TBD, likely to be structured as multiple tables that are collectively related to tbl_application_type and **not** managed by a state machine.

Fabric Application Type

Table: tbl_fabric_application_types

View: fabric_application_types

FSM: FabricApplicationTypeMachine

Contains a row for each Fabric Application Type for each fabric cluster and additional state to drive/track the progress of an application type upgrade.

	Column Name	Data Type	Nullable	Description
PK	fabric_cluster_name	nvarchar(128)	False	The name of the fabric cluster on which this application type is located.
PK	application_type_name	nvarchar(256)	False	Name of fabric application type.
	application_build_path	nvarchar(256)	True	Relative path in the image store of the package we're currently upgrading to.
	current_application_type_version	nvarchar(256)	False	The current version of this application type on this fabric cluster. This is the version new applications will get instantiated with.
	target_application_type_version	nvarchar(256)	False	Target version for this application type on this cluster. This is different from the current version only when there is an upgrade going on.
	applications_upgrading	nvarchar(max)	True	Pipe (|) separated list of applications currently undergoing upgrade that are being tracked. [will be upgrade to an xml column in a later change]
	last_processed_application	nvarchar(256)	True	Uri of the last application upgrade was kicked off for. Used to save iteration state when kicking off upgrade for a large number of applications.
	deployment_tag	nvarchar(256)	True	Textual tag for the current deployment going on.
	rollout_key	nvarchar(256)	True	Transient state during upgrade. Key of the

	Column Name	Data Type	Nullable	Description
				currently in-progress rollout for this app type.
	upgrade_policy	xml	True	Transient state during upgrade. Upgrade policy to use when calling WinFab to kick off the upgrade. Essentially XML-serialized UpgradePolicyDescription.
	last_rollout_outcome	bit	True	Indicates the overall success or failure for the whole rollout. Starts out as true, flips to false as soon as an app instance fails upgrade.
	last_rollout_start_time	datetime2	True	Start time of the last rollout for this application type.

Fabric Cluster [was Fabric Hosted Service]

Table: tbl_fabric_clusters

View: fabric_clusters

FSM: FabricClusterStateMachine

Contains a row for each Fabric Cluster within the SQL Cluster. Fabric Cluster contains base data common to SQL Control Rings and SQL Tenant Rings.

	Column Name	Data Type	Nullable	Description
PK	name [was ClusterName]	nvarchar(128)	False	
	hosted_service_subscription_id	nvarchar(128)	False	The internal Microsoft-owned subscription id under which the fabric cluster is created.
	hosted_service_name	nvarchar(128)	False	The exposed Azure name of the fabric cluster.
	hosted_service_dns_name	nvarchar(128)	True	<verify usage, name; is this different to the above>
	deployment_id	nvarchar(128)	True	
	deployment_id_sequence	bigint	True	Sequence number used to track the progress of a deployment in case of interruption. <verify usage>
	deployment_id_replaced_sequence	bigint	True	Sequence number used to track the progress of a redeployment in case of interruption.
	processed_fabric_uri	nvarchar(256)	True	<verify usage>
	ipv4_address	nvarchar(128)	True	

Fabric Services

Table: tbl_fabric_services

View: fabric_services

FSM: FabricServiceStateMachine

Contains a row for each Fabric Service deployed in the SQL Cluster. Includes database services and SQL instance application services deployed on the tenant rings, as well as infrastructural services deployed to both the control ring and tenant rings.

Column Name	Data Type	Nullable	Description
fabric_cluster_name [was ClusterName]	nvarchar(128)	False	The fabric cluster name.
fabric_service_uri	nvarchar(256)	False	
deployment_id_sequence	bigint	True	
application_uri	nvarchar(256)	False	
service_type_name	nvarchar(256)	False	
placement_constraints	nvarchar(2000)	True	
initialization_data	varbinary(MAX)	True	
kind	nvarchar(128)	False	Stateless or Stateful
target_replica_set_size	int	False	
has_persisted_state	bit	True	An additional qualifier required by Windows Fabric when kind == Stateful
min_replica_set_size	int	True	
metrics	xml	True	
correlations	xml	True	

Fabric Nodes

Table: tbl_fabric_nodes

View: fabric_nodes

FSM: NodeStateMachine

Contains a row for each Fabric Node in the Fabric Clusters. Key to uniquely identify a node is the combination of node name and the fabric cluster name it belongs to.

Column Name	Data Type	Nullable	Description
name	nvarchar(128)	False	The node name.
fabric_cluster_name	nvarchar(128)	False	Name of the fabric cluster the node belongs to.
fabric_state	nvarchar(128)	False	Textual representation of the state reported by WinFabric. This gets refreshed periodically (currently every 1 minute).
deactivation_intent	nvarchar(64)	True	The intent expressed for deactivation, if the state machine is in the process of deactivating the node. NULL otherwise.

How good have you found this content?