Documentação Sync

Requisitos 🔗

Requisito	Descrição técnica
O script deve realizar comandos SQL no ambiente CDP com configurações de qual ambiente executar as queries centralizadas em um arquivo de configuração	Adicionar arquivo config.py com as configurações necessárias para execução do script pela função main
O script deve controlar as execuções com base nos cenários de teste destacados na documentação, executando-os de forma inteligente e garantindo a integridade dos testes dos cenários em cada etapa	 É preciso separar use cases / test scenarios para poder executar de forma sequencial, com controle da execução efetuada pelo usuário, informando quando deve executar o próximo: Declarar uma classe de Log personalizado para informar sempre ao final de cada use case o que deve-se fazer; Adicionar novos README em cada diretório de classe para facilitar leitura e interpretação do que representam
O script deve ser capaz de realizar limpeza dos processos executados automaticamente	Criar arquivo de cleanup.py desfazendo o que foi realizado no comando e fazê-lo parte da evolução do script: • Destacar no arquivo README a necessidade de atualizar este arquivo sempre que ocorrer evoluções no script também • Adicionar um template de PR para checar alterações como essas na documentação e no script. • Arquivo deve deletar tabelas do Glue (avaliar se é factível e não impactará o Sync) • OBS: Em prod não será possível sem BREAKING GLASS

#	Cenário	Detalhamento	Como validar	Resultado esperado
1	Sincronismo de uma nova tabela e partições simples	O script realizará os seguintes comandos: 1. CREATE EXTERNAL TABLE para a tabela tb_sync_scenario_si mple_partitions; 2. ALTER TABLE ADD PARTITION adicionando 10 novas partições para a tabela tb_sync_scenario_si mple_partitions; 3. INSERT INTO com 10 novas linhas para cada partição nova na tabela.	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_scenario_si mple_partitions foi criada e avaliar as partições existentes com o comando: SHOW PARTITIONS ng2.tb_sync_scenario_simple_partition s	No CDP 1. Tabela criada com 10 novas partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso; 2. Dados retornados na consulta da tabela

			Na AWS	
			Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_scenario_si mple_partitions e abrir os detalhes; 4. Validar a criação das 10 partições 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional: SELECT * FROM ng2.tb_sync_scenar io_simple_partition s	
2	Sincronismo de uma nova tabela e partições compostas	O script realizará os seguintes comandos: 1. CREATE EXTERNAL TABLE para a tabela tb_sync_scenario_co mposed_partitions; 2. ALTER TABLE ADD PARTITION adicionando 10 novas partições para a tabela tb_sync_scenario_co mposed_partitions; 3. INSERT INTO com 10 novas linhas para cada partição nova na tabela.	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_scenario_co mposed_partitions foi criada e avaliar as partições existentes com o comando: SHOW PARTITIONS ng2.tb_sync_scenar io_composed_partit ions	No CDP 1. Tabela criada com 10 novas partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso; 2. Dados retornados na consulta da tabela

			Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_scenario_co mposed_partitionse abrir os detalhes; 4. Validar a criação das 10 partições 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional: SELECT * FROM ng2.tb_sync_scenario_composed_partitions	
3	Teste de estresse massivo do sincronismo de 10 mil partições simples para uma mesma tabela	O script realizará os seguintes comandos: 1. CREATE EXTERNAL TABLE para a tabela tb_sync_massive_sc enario_simple_partiti ons; 2. ALTER TABLE ADD PARTITION adicionando 10 novas partições para a tabela tb_sync_massive_sc enario_simple_partiti ons;	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_massive_sc enario_simple_partiti ons foi criada e avaliar as partições existentes com o comando: SHOW PARTITIONS ng2.tb_sync_massi	No CDP 1. Tabela criada com 10 mil novas partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso

			ve_scenario_simple _partitions	
			Na AWS	
			Após execução do Metadata Sync:	
			1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_massive_sc enario_simple_partiti ons abrir os detalhes; 4. Validar a criação das partições 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional: SELECT * FROM ng2.tb_sync_scenar io_composed_partit ions	
4	Teste de estresse massivo do sincronismo de 10 mil partições compostas para uma mesma tabela	O script realizará os seguintes comandos: 1. CREATE EXTERNAL TABLE para a tabela tb_sync_massive_sc enario_composed_pa rtitions; 2. ALTER TABLE ADD PARTITION adicionando 10 novas partições para a tabela tb_sync_massive_sc	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_massive_sc enario_composed_pa rtitions foi criada e avaliar as partições existentes com o comando: SHOW	No CDP 1. Tabela criada com 10 mil novas partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso;

enario_composed_pa **PARTITIONS** rtitions; ng2.tb_sync_massi ve_scenario_compo sed_partitions Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_massive_sc enario_composed_pa rtitions abrir os detalhes; 4. Validar a criação das partições 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional: **SELECT * FROM** ng2.tb_sync_massi ve_scenario_compo sed_partitions No CDP No CDP 5 Sincronismo do Drop de partições em uma O script realizará o tabela com partições simples seguinte comando: 1. Tabela deve estar Após execução do 1. ALTER TABLE com apenas 5 comando: **DROP PARTITION** partições; 1. Acessar o DataHub dropando as 5 das cujo teste foi 10 partições da executado; Na AWS tabela 2. Acessar o database tb_sync_scenario_si 1. Sincronismo dos ng2; mple_partitions; metadados efetuado 3. Validar que a tabela com sucesso e as 5 tb_sync_scenario_si partições devem ter mple_partitions não sido deletadas do

possui mais

			partições: SHOW PARTITIONS ng2.tb_sync_scenar io_simple_partition s Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_scenario_si mple_partitions abrir os detalhes; 4. Validar a deleção das tabelas; 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional:	Glue Catalog para a tabela; 2. Dados retornados na consulta da tabela.
			seguinte query para avaliar que os dados estão ok e a tabela	
6	Sincronismo do Drop de de partições em uma tabela com partições compostas	O script realizará o seguinte comando: 1. ALTER TABLE DROP PARTITION dropando as 5 das 10 partições da tabela tb_sync_scenario_co mposed_partitions;	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_scenario_co mposed_partitions não possui mais	No CDP 1. Tabela deve estar com apenas 5 partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso e as 5 partições devem ter sido deletadas do

			partições: SHOW PARTITIONS ng2.tb_sync_scenar io_composed_partit ions Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_scenario_co mposed_partitions abrir os detalhes; 4. Validar a deleção das tabelas; 5. Acessar o serviço Athena e realizar a	Glue Catalog para a tabela; 2. Dados retornados na consulta da tabela.
			seguinte query para avaliar que os dados estão ok e a tabela está funcional: SELECT * FROM ng2.tb_sync_scenar io_composed_partit ions	
7	Teste de estresse massivo do sincronismo do drop de 10 mil partições simples para uma mesma tabela	O script realizará o seguinte comando: 1. DROP TABLE IF EXISTS dropando a tabela tb_sync_massive_sc enario_simple_partiti ons e suas 10 mil partições simples;	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_massive_sc enario_simple_partiti ons não possui mais	No CDP 1. Tabela deve estar com apenas 5 partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso e as 5 partições devem ter sido deletadas do

			partições: SHOW PARTITIONS ng2.tb_sync_massi ve_scenario_simple _partitions Na AWS Após execução do Metadata Sync: 1. Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol) 2. Acessar o Glue Catalog e clicar em Tables; 3. Buscar pela tabela tb_sync_massive_sc enario_simple_partiti ons abrir os detalhes; 4. Validar a deleção das tabelas; 5. Acessar o serviço Athena e realizar a seguinte query para avaliar que os dados estão ok e a tabela está funcional:	Glue Catalog para a tabela; 2. Dados retornados na consulta da tabela.
8	Teste de estresse massivo do sincronismo do drop de 10 mil partições compostas para uma mesma tabela	O script realizará o seguinte comando: 1. DROP TABLE IF EXISTS dropando a tabela tb_sync_massive_sc enario_composed_pa rtitions e suas 10 mil partições compostas;	No CDP Após execução do comando: 1. Acessar o DataHub cujo teste foi executado; 2. Acessar o database ng2; 3. Validar que a tabela tb_sync_massive_sc enario_composed_pa rtitions não possui	No CDP 1. Tabela deve estar com apenas 5 partições; Na AWS 1. Sincronismo dos metadados efetuado com sucesso e as 5 partições devem ter sido deletadas do

mais partições:
SHOW PARTITIONS
ng2.tb_sync_massi
ve_scenario_compo
sed_partitions

- Glue Catalog para a tabela;
- 2. Dados retornados na consulta da tabela.

Na AWS

Após execução do Metadata Sync:

- Acessar a conta de controle do ambiente referente ao teste (dev, hom ou pro das contas Analyticscontroltest ou Analyticsdatalakecont rol)
- Acessar o Glue
 Catalog e clicar em
 Tables;
- Buscar pela tabela tb_sync_massive_sc enario_composed_pa rtitions abrir os detalhes;
- Validar a deleção das tabelas;
- Acessar o serviço
 Athena e realizar a
 seguinte query para
 avaliar que os dados
 estão ok e a tabela
 está funcional:

SELECT * FROM ng2.tb_sync_massi ve_scenario_compo sed_partitions