

User

Qual o ganho de utilizar um job bookmark enabled em um glue job na aws

Bing

O uso de **job bookmarks** em um **AWS Glue job** pode ser benéfico para rastrear dados que já foram processados e evitar o processamento repetido de dados. O bookmark mantém o controle do progresso do trabalho e ajuda a processar dados incrementais .

Ao programar um AWS Glue job com bookmarks, você tem acesso a flexibilidade indisponível em trabalhos visuais. Ao ler de JDBC, você pode especificar a(s) coluna(s) a serem usadas como chaves de bookmark em seu script AWS Glue. Você pode escolher qual `transformation_ctx` aplicar a cada chamada de método. Sempre tenha `job.init()` no início do script e `job.commit()` no final do script. Essas duas funções inicializam o serviço de bookmark e atualizam a alteração de estado para o serviço. Os bookmarks não funcionarão sem chamá-los .

Para trabalhos JDBC, o bookmark mantém o controle das linhas que seu trabalho leu, comparando os valores dos campos-chave com um valor marcado. Isso não é necessário ou aplicável para fluxos de trabalho Amazon S3. Ao escrever um script AWS Glue sem o editor visual, você pode especificar qual coluna rastrear com bookmarks. Você também pode especificar várias colunas. Lacunas na sequência de valores são permitidas ao especificar chaves de bookmark definidas pelo usuário. Se as chaves de bookmark definidas pelo usuário forem usadas, elas devem ser estritamente monotonamente crescentes ou decrescentes. Ao selecionar campos adicionais para uma chave composta, campos para conceitos como “versões menores” ou “números de revisão” não atendem a esse critério, pois seus valores são reutilizados em todo o conjunto de dados. Você pode especificar `jobBookmarkKeys` e `jobBookmarkKeysSortOrder` das seguintes maneiras:

`create_dynamic_frame.from_catalog` - Use `additional_options`. `create_dynamic_frame.from_options` - Use `connection_options` .

Muitos dos métodos de quadro dinâmico AWS Glue PySpark incluem um parâmetro opcional chamado `transformation_ctx`, que é um identificador exclusivo para a instância do operador ETL. O parâmetro `transformation_ctx` é usado para identificar informações de estado dentro de um bookmark de trabalho para o operador fornecido. Especificamente, a AWS Glue usa `transformation_ctx` para indexar a chave para o estado do bookmark. Para que o bookmark funcione corretamente, você deve sempre manter a fonte e o `transformation_ctx` associado consistentes. Alterar a propriedade da fonte ou renomear o `transformation_ctx` pode tornar o bookmark anterior inválido e a filtragem baseada em carimbo de data e hora pode não produzir o resultado correto. Para que os bookmarks de trabalho funcionem corretamente, habilite o parâmetro de bookmark de trabalho e defina o parâmetro `transformation_ctx`. Se você não passar o parâmetro `transformation_ctx`, os bookmarks de trabalho não serão habilitados para um quadro dinâmico ou tabela usada no método. Por exemplo, se você tiver um trabalho ETL que lê e une duas fontes Amazon S3, poderá optar por passar o parâmetro `transformation_ctx` apenas para aqueles métodos que deseja habilitar bookmarks. Se você redefinir o bookmark de trabalho para um trabalho, ele redefine todas as transformações associadas ao trabalho, independentemente do `transformation_ctx` usado .

Espero que isso ajude!

User

POsso usar o AWS ACL para garantir acesso a um bucket s3?

Bing

Sim, você pode usar **AWS ACL** para controlar o acesso a um bucket S3. Cada bucket e objeto tem um ACL anexado a ele como um sub-recurso. Ele define quais contas ou grupos da AWS têm acesso e o tipo de acesso. Quando uma solicitação é recebida em relação a um recurso, o Amazon S3 verifica o ACL correspondente para verificar se o solicitante tem as permissões de acesso necessárias .

No entanto, a maioria dos casos de uso modernos no Amazon S3 não requer mais o uso de ACLs. Recomendamos que você mantenha os ACLs desativados, exceto em circunstâncias incomuns em que você precisa controlar o acesso para cada objeto individualmente. Com ACLs desativados, você pode usar políticas para controlar o acesso a todos os objetos em seu bucket, independentemente de quem os carregou .

Se você deseja implementar um controle granular sobre objetos individuais em seu bucket, use ACLs. Se você deseja implementar um controle global, como tornar um bucket inteiro público, use políticas .

Espero ter ajudado!

User

O que o comando sql truncate faz?

Bing

O comando SQL `TRUNCATE TABLE` é usado para remover todos os registros de uma tabela, mas não a tabela em si . É mais rápido e usa menos recursos do que o comando `DELETE` . O `TRUNCATE TABLE` é uma instrução DDL (Linguagem de Definição de Dados) que remove todos os dados de uma tabela, mas mantém a estrutura da tabela intacta .

Espero ter ajudado!

User

Ao criar uma tabela no sql serve, o que significa definir um atributo como “int identity”

Bing

Quando você define um atributo como “**int identity**” em uma tabela SQL Server, você está criando uma coluna que é preenchida automaticamente com valores incrementais à medida que novas linhas são adicionadas à tabela .

O valor inicial da coluna é definido pelo parâmetro **seed** e o incremento é definido pelo parâmetro **increment**. Por exemplo, se você definir a coluna como “**int identity(1,1)**”, o primeiro valor será 1 e cada nova linha adicionada à tabela terá um valor incrementado em 1 .

Espero ter ajudado!

User

Utilizando try except e else para gerenciar erros, quando o bloco else será executado?

Bing

O bloco `else` em um bloco `try-except-else` é executado **somente** se nenhuma exceção for lançada no bloco `try`. O bloco `else` é opcional e é colocado após o bloco `try` e antes do bloco `except`. Se uma exceção for lançada no bloco `try`, o bloco `else` não será executado .

Espero ter ajudado!

User

O Spark é escrito em sacala?

Bing

Sim, o **Apache Spark** é escrito em **Scala** e é executado em uma máquina virtual Java . O Spark é uma plataforma de computação em cluster com paralelismo que distribui dados entre os clusters para processá-los de maneira paralela . Atualmente, o Spark suporta as seguintes linguagens para o desenvolvimento de aplicativos: R, Python, Java e Scala .

Espero ter ajudado!

User

O Spark roda em uma máquina virtual java porque sua fonte é escrita em java?

Bing

O **Apache Spark** é escrito em **Scala** e é executado em uma máquina virtual Java . O Spark é uma plataforma de computação em cluster com paralelismo que distribui dados entre os clusters para processá-los de maneira paralela . Atualmente, o Spark suporta as seguintes linguagens para o desenvolvimento de aplicativos: R, Python, Java e Scala .

Espero ter ajudado!

User

O spark é uma engine?

Bing

Sim, o **Apache Spark** é uma engine de processamento de dados de software livre para grandes conjuntos de dados . O Spark foi desenvolvido em 2009 na UC Berkeley e é escrito em **Scala** . Ele é executado em uma máquina virtual Java e é projetado para fornecer velocidade computacional, escalabilidade e programabilidade necessárias para Big Data, especificamente para aplicativos de streaming de dados, dados de gráficos, machine learning e inteligência artificial (IA) . O Spark processa dados de 10 a 100 vezes mais rápido que as alternativas e escala distribuindo o trabalho de processamento em grandes clusters de computadores, com paralelismo integrado e tolerância a falhas . O Spark suporta as seguintes linguagens para o desenvolvimento de aplicativos: R, Python, Java e Scala .

Espero ter ajudado!

User

Qual o melhor recurso para manipular serviços da aws através de um shell?

Bing

O **AWS CloudShell** é uma ferramenta que fornece acesso fácil a uma shell Linux predefinida em uma janela do navegador. Ele permite que você execute comandos da AWS CLI, AWS SDKs e outras ferramentas de linha de comando para trabalhar com recursos da AWS diretamente no console do navegador .

O AWS CloudShell é uma opção conveniente para gerenciar serviços da AWS através de um shell, pois você não precisa baixar ou instalar nenhum recurso adicional. Além disso, como você já está autenticado no shell, não precisará configurar as credenciais antes de fazer chamadas .

Espero ter ajudado!