

User

Estilos de distribuição disponíveis no Amazon Redshift

Bing

O Amazon Redshift oferece quatro estilos de distribuição para tabelas: AUTO, EVEN, KEY e ALL . Quando você cria uma tabela, pode especificar um estilo de distribuição ou deixar que o Amazon Redshift escolha um estilo de distribuição ideal com base no tamanho dos dados da tabela .

- **Distribuição AUTO:** O Amazon Redshift atribui um estilo de distribuição ideal com base no tamanho dos dados da tabela. Por exemplo, se o estilo de distribuição AUTO for especificado, o Amazon Redshift inicialmente vai atribuir o estilo de distribuição ALL a uma tabela pequena. Quando a tabela crescer, o Amazon Redshift poderá alterar o estilo de distribuição para KEY, escolhendo a chave primária (ou uma coluna da chave primária composta) como chave de distribuição. Se a tabela crescer e nenhuma das colunas for adequada para ser a chave de distribuição, o Amazon Redshift vai alterar o estilo de distribuição para EVEN .
- **Distribuição EVEN:** O nó de liderança distribui as linhas ao longo das fatias de modo round-robin, independente dos valores de qualquer coluna específica. A distribuição EVEN é apropriada quando uma tabela não participa de junções. Também é apropriado quando não há uma opção clara entre a distribuição KEY e ALL .
- **Distribuição KEY:** As linhas são distribuídas de acordo com os valores em uma coluna. O nó líder coloca os valores correspondentes na mesma fatia do nó. Se você distribuir um par de tabelas nas chaves de união, o nó líder coloca as linhas nas fatias de acordo com os valores nas colunas de união. Desta forma, os valores correspondentes das colunas comuns são armazenados fisicamente juntos .
- **Distribuição ALL:** Uma cópia de toda a tabela é distribuída para cada nó. Onde a distribuição EVEN ou a distribuição KEY coloca apenas uma porção das linhas da tabela em cada nó, a distribuição ALL garante que todas as linhas sejam dispostas para cada junção na qual a tabela participa. A distribuição ALL multiplica o armazenamento exigido pelo número de nós no cluster e, portanto, ela demora muito mais tempo para carregar, atualizar ou inserir dados em várias tabelas. A distribuição ALL é apropriada somente para tabelas relativamente lentas; ou seja, tabelas que não são atualizadas frequentemente ou extensivamente .

Espero ter ajudado! 😊 .

User

Joinstep é uma operação de transformação no spark?

Bing

Não encontrei nenhuma referência a “Joinstep” como uma operação de transformação no Apache Spark. No entanto, o Apache Spark oferece várias operações de transformação, como map, filter, flatMap, reduceByKey, join, union, distinct, groupByKey, sortByKey, cogroup, subtractByKey, sample, pipe, coalesce, repartition, repartitionAndSortWithinPartitions, zip, zipWithIndex, zipWithUniqueId e outras .

Espero ter ajudado! 😊 .

User

Qual é a denominação do comando para registrar mudanças locais em um repositório, usando sistemas de controle de versão svn ou cvs

Bing

Não encontrei nenhuma referência a “Joinstep” como uma operação de transformação no Apache Spark. No entanto, o Apache Spark oferece várias operações de transformação, como map, filter, flatMap, reduceByKey,

join, union, distinct, groupByKey, sortByKey, cogroup, subtractByKey, sample, pipe, coalesce, repartition, repartitionAndSortWithinPartitions, zip, zipWithIndex, zipWithUniqueId e outras .

O comando para registrar mudanças locais em um repositório usando sistemas de controle de versão

Subversion (SVN) é `svn commit` . Já em sistemas de controle de versão **Concurrent Versions System (CVS)**, o comando é `cvs commit` .

Espero ter ajudado! 😊 .