

lement des termes supplémentaires reconnus par l'industrie qui ne sont pas utilisés dans les vidéos de cours. Ces termes sont essentiels pour que vous puissiez les on professionnelle.

grégées. Les fonctions d'agrégation couramment utilisées telles que count(), avg(), max() et d'autres sont intégrées aux Dataframes. Les utilisateurs peuvent table.
DataFrame à l'aide de la méthode du schéma d'impression. Il est impératif de noter les types de données dans chaque colonne. Les utilisateurs peuvent appliquer
es et l'analyse itérative d'énormes volumes de données.
un plan logique dans la phase d'analyse. Le plan logique évolue vers un plan logique optimisé dans la phase d'optimisation logique. Il s'agit de l'étape nt et l'élagage sont applicables ici. Catalyst génère plusieurs plans physiques basés sur le plan logique dans la phase de planification physique. Un plan physique nent effectuer le calcul. Un modèle de coût sélectionne ensuite le plan physique ayant le coût le plus faible. Cela explique l'étape d'optimisation basée sur les ysiq ue sélectionné et génère le bytecode Java à exécuter sur les nœuds.
gles d'arborescence de données en arrière-plan. Catalyst effectue les quatre tâches de haut niveau suivantes pour optimiser une requête : analyse, optimisation
en, de manière séquentielle pour l'amélioration et l'exécution des requêtes. La fonction principale de Catalyst consiste à dériver un plan de requête physique en œuvre d'une gamme de transformations telles que le refoulement des prédicats, l'élagage des colonnes et le repliement constant sur le plan logique.
ête. L'optimiseur Catalyst sélectionne un chemin de requête qui entraîne une consommation minimale de temps et de mémoire. Comme les requêtes peuvent embles de données font partie du calcul.
le temporaire utilisée pour exécuter des requêtes SQL. Les vues temporaires et globales sont prises en charge par Spark SQL. Une vue temporaire a une portée ctuel. Une vue temporaire globale existe dans l'application Spark générale. Cette vue peut être partagée entre différentes sessions Spark.