

Nell'Hadoop Distributed File System (HDFS), i termini "namespace", "NameNode" e "DataNode" sono strettamente collegati ma si riferiscono a concetti leggermente diversi. Ecco una spiegazione di ciascun termine:

#### Namespace:

Il namespace si riferisce alla struttura gerarchica dei file e delle directory all'interno di HDFS. Funziona come un albero di directory virtuale che organizza e identifica i file e le directory nel sistema di file distribuito. Il namespace nel suo insieme rappresenta la struttura e i metadati dei file e delle directory in HDFS.

#### NameNode:

Il NameNode è uno dei componenti chiave di HDFS. È il server centrale che gestisce il metadata e il controllo del sistema di file. Alcune delle sue funzioni principali includono:

**Gestione del namespace:** Il NameNode mantiene un registro di tutto il metadata dei file e delle directory all'interno del sistema HDFS. Questo include informazioni come le posizioni dei blocchi di dati, le autorizzazioni di accesso e le dimensioni dei file. Il NameNode è responsabile della gestione delle operazioni sul file system, come la creazione di file, la creazione di directory e la gestione delle autorizzazioni.

**Gestione dei blocchi di dati:** Il NameNode tiene traccia dei blocchi di dati che costituiscono i file all'interno di HDFS. Mappa i blocchi di dati ai DataNode che li archiviano fisicamente. Il NameNode coordina anche il recupero dei blocchi di dati dai DataNode quando richiesto dai client.

**Monitoraggio dello stato dei DataNode:** Il NameNode riceve report periodici dai DataNode per monitorare lo stato di ogni DataNode nel cluster HDFS. Questo aiuta a rilevare guasti o errori hardware dei DataNode e a garantire l'affidabilità del sistema.

#### DataNode:

I DataNode sono i nodi di archiviazione nel cluster HDFS. Ogni DataNode è responsabile dell'archiviazione fisica dei blocchi di dati sui propri dischi locali. Le loro funzioni principali sono:

**Archiviazione dei dati:** I DataNode memorizzano i blocchi di dati effettivi dei file. Ricevono i blocchi di dati dal NameNode e li scrivono sui dischi locali. I DataNode mantengono anche una replica dei blocchi di dati per garantire la tolleranza ai guasti.

**Lettura e scrittura dei dati:** Quando un client richiede un file o parte di esso, i DataNode forniscono i blocchi di dati corrispondenti per la lettura o la scrittura. I DataNode sono responsabili dell'effettiva trasmissione dei dati tra i client e il disco locale.

**Report periodici al NameNode:** I DataNode inviano report periodici al NameNode per informarlo sullo stato dei blocchi di dati che archiviano. Questo aiuta il NameNode a tenere traccia dell'integrità dei blocchi di dati e a rilevare eventuali guasti o errori.

In sintesi, il namespace si riferisce alla struttura e ai metadati dei file e delle directory in HDFS. Il NameNode è il server centrale che gestisce il metadata e il controllo del sistema di file, mentre i DataNode sono i nodi