Protecting Your Data from Disaster

Today, every business needs to implement data protection practices to ensure their valuable data is protected from a disaster. Datameer recommends using a combination of industry-standard high availability, replication and backup technologies, and best practices. Together, they can support data protection for any combination of recovery point objectives (RPOs) and recovery time objectives (RTOs).

Data and insight are the lifeblood of business operations today. So it’s essential that disaster recovery services are in place to protect your data from disasters, which can range from disk and node failures to accidental data deletion and even site failures.

To protect your Datameer instance, for example, Datameer recommends that you implement a mix of industry-standard high availability, data replication and backup best practices. You have the flexibility to implement these best practices to meet your specific RPO and RTO objectives. Datameer Services can also help you determine the optimal data protection plan for your Datameer instance, as well as implement the requisite best practices as an additional service.

Proven Data Protection for Critical Components

As shown in FIGURE 1, there are three main components of your Datameer installation that must be protected from potential disasters:

- Configuration files in the Datameer installation directory
- The Oracle® MySQL database instance containing related metadata
- Data files on the Hadoop Distributed File System (HDFS)
**Configuration Files**

To protect from disk and node failures, the configuration files in your Datameer installation directory – including any custom plug-in binaries – can be protected using a RAID volume or SAN attached storage. Data backups should also be performed regularly so that in the event data is accidentally deleted, you can restore it as needed.

**Metadata**

The metadata for a Datameer instance resides in a MySQL database instance. MySQL natively offers high availability and replication configuration options. Additionally, there are native scripts to backup and recover a database instance, as needed.

**Data Files on HDFS**

Datameer uses Hadoop to store imported data and the results of analyses. Because Hadoop is a distributed system designed to tolerate disk and node failures, it’s designed from the ground up to protect your data. Hadoop can be enhanced using Apache™ Falcon to provide enhanced data management capabilities, including replication. In addition, you can configure Hadoop’s file system trash interval to protect your data from accidental deletion.

**Maintaining Harmony Across the Three Components**

Sometimes only one of the three components of an installation fails, and the others do not. For example, suppose a MySQL database instance fails, but the HDFS data and configuration files are healthy. This can result in your system falling into an inconsistent state.

In this scenario, after recovering the MySQL instance, it is possible that the results of a job stored on the HDFS are no longer visible in Datameer – for example, if the metadata for the job was not included in the current backup. In this case, the best practice is to recover the failed component and then restore harmony across all three components to a consistent time period. This minimizes the need to manually synchronize the components after recovery, thus bringing them into a consistent state.

With Datameer, you can quickly and efficiently bring all of the components of your system into a consistent state by simply reprocessing the input data for the job in the inconsistent state (i.e., your Datameer workbook). In this case, the impacted job would need to be rerun to complete the disaster recovery. You can perform these steps using your internal IT resources or engage Datameer Services to assist you.

**Learn More**

Planning for a disaster is a complicated but business-critical endeavor. But with Datameer, you’re always equipped and in control. Take advantage of a wide range of protection services to meet your recovery goals. And at any time, the Datameer Services team can assist you. Our experts can advise you on what components to protect at what levels for different contexts, as well as help implement the appropriate technologies and best practices.

To learn more about data protection, please refer to the following reference documents: