

# SPECIFIC TOOLS YOU'LL NEED TO GET YOUR DATABASE READY FOR

# AI



Based on all the AI work we have accomplished over the past few years we developed the following checklist to help you prepare your data using private cloud or on-premise systems and software ...which is a critical first step. Don't hesitate to contact us with any questions.

## 1. DATA INTEGRATION

- Integration tools like Talend, Informatica, or Apache NiFi consolidate data from multiple sources into a single, unified view.

## 2. DATA CLEANING AND PREPARATION

- Use a private cloud or on-premise data cleaning tool like OpenRefine, Excel, or SQL to identify and correct errors, inconsistencies, and missing values in the data.

## 3. DATA TRANSFORMATION

- Data transformation tools like Apache Beam, Apache Spark, or AWS Glue convert data into a format suitable for AI models, such as structured or semi-structured data.

## 4. DATA LABELING

- Use a private cloud or on-premise data labeling tool like Labelbox, Hive, or Amazon SageMaker to identify and label the data that will be used to train AI models consistently and efficiently.

## 5. DATA STORAGE

- Distributed file systems (DFS) like Hadoop Distributed File System (HDFS), Amazon S3, or Google Cloud Storage store the data in a scalable and durable manner.

## 6. DATA SECURITY

- Implement appropriate security measures to protect the data from unauthorized access or misuse using tools like Apache Hadoop, AWS Key Management Service (KMS), or Google Cloud Key Management Service (KMS) during storage and transmission.

## 7. DATA GOVERNANCE

- Establish clear policies and procedures for data management and use, utilizing tools like Apache Atlas, AWS Lake Formation, or Google Cloud Data Fusion to manage data access and usage.

## 8. AI MODEL DEVELOPMENT

- Learning frameworks like TensorFlow, PyTorch, or Scikit-learn develop and train AI models using the prepared data.

## 9. DEPLOYMENT

- Deploy the trained AI models into production environments using tools like Kubernetes, Docker, or AWS Elastic Beanstalk in a scalable and efficient manner.

## 10. MONITORING AND MAINTENANCE

- Continuously monitor the performance of the AI models in production with tools like Prometheus, Grafana, or New Relic to monitor the models' performance and make necessary adjustments.

“

By using private cloud or on-premise systems and software only, you can ensure that your data is stored and processed securely and efficiently within your infrastructure, without relying on any external services or platforms.

”

**TALK TO OUR TECH EXPERT TODAY!**