DAG 1: Monitoramento do Elasticsearch

Esta DAG realiza um monitoramento simples do Elasticsearch usando um HttpSensor.

Pontos importantes para finetuning:

* Utiliza um sensor HTTP para verificar a disponibilidade do Elasticsearch.
* Configurado para executar a cada 10 minutos.
* Implementa retry e timeout para lidar com falhas temporárias.

Relevância para IA:

* Monitoramento de infraestrutura é crucial para garantir a disponibilidade de sistemas de IA.
* Pode ser adaptado para monitorar endpoints de modelos de IA implantados.

DAG 2: Análise de Kafka

Esta DAG consome mensagens do Kafka e as processa.

Pontos importantes para finetuning:

* Consome mensagens de um tópico Kafka específico.
* Limita o consumo a 100 mensagens por execução.
* Implementa tratamento de erros e logging.

Relevância para IA:

* Pode ser usado para ingerir dados de streaming para treinamento ou inferência de modelos.
* O processamento de mensagens pode incluir pré-processamento de dados para modelos de IA.

DAG 3: Análise de Saída do Spark

Similar à DAG 2, mas focada em consumir saídas do Spark.

Pontos importantes para finetuning:

* Consome mensagens de um tópico Kafka específico para saídas do Spark.
* Tenta analisar as mensagens como JSON.

Relevância para IA:

* Útil para coletar resultados de processamento em larga escala do Spark, que podem ser usados para treinamento ou avaliação de modelos.

DAG 4: Verificação do MongoDB

Esta DAG verifica a conexão com o MongoDB e coleta estatísticas.

Pontos importantes para finetuning:

* Verifica a conexão com o MongoDB.
* Coleta estatísticas do banco de dados.
* Lista as coleções disponíveis.

Relevância para IA:

* Monitoramento de bancos de dados é importante para garantir a disponibilidade de dados para modelos de IA.
* Pode ser adaptado para verificar a integridade de conjuntos de dados usados no treinamento.

DAG 5: Monitoramento de DAGs

Esta DAG monitora a saúde de outras DAGs no Airflow.

Pontos importantes para finetuning:

* Verifica DAGs falhas.
* Envia notificações por e-mail em caso de falhas.

Relevância para IA:

* Crucial para manter a integridade de pipelines de IA complexos.
* Permite resposta rápida a falhas em processos de treinamento ou inferência.

DAG 6: Monitoramento do PostgreSQL

Similar à DAG 1, mas focada no PostgreSQL.

Pontos importantes para finetuning:

* Usa um SqlSensor para verificar a disponibilidade do PostgreSQL.

Relevância para IA:

* Importante para garantir a disponibilidade de bancos de dados relacionais que podem armazenar metadados de modelos ou resultados de inferência.

DAG 7: Notificação por E-mail

Esta DAG implementa um sistema de notificação por e-mail para falhas.

Pontos importantes para finetuning:

* Usa um callback de falha para enviar e-mails.
* Inclui informações detalhadas sobre a falha no e-mail.

Relevância para IA:

* Essencial para manter os stakeholders informados sobre problemas em pipelines de IA.

DAG 8: Execução de Notebooks Jupyter

Esta DAG executa notebooks Jupyter programados.

Pontos importantes para finetuning:

* Usa o PapermillOperator para executar notebooks.
* Permite passagem de parâmetros para os notebooks.

Relevância para IA:

* Extremamente útil para automatizar experimentos de ML/DL.
* Facilita a reprodutibilidade de experimentos e análises.

Conclusão

Estas DAGs fornecem uma base sólida para construir pipelines de IA robustos e automatizados. Elas cobrem aspectos cruciais como ingestão de dados, monitoramento de infraestrutura, notificações e execução automatizada de experimentos. Para finetuning de IA, seria benéfico adicionar DAGs específicas para:

1. Pré-processamento de dados em larga escala.
2. Treinamento automatizado de modelos.
3. Avaliação e comparação de modelos.
4. Implantação de modelos em produção.
5. Monitoramento de desempenho de modelos em produção.