



Big Data Fundamentos 3.0

Big Data Fundamentos – Versão 3.0

Principais Ferramentas e Plataformas de DataOps

Abaixo você encontra uma seleção das principais ferramentas e plataformas de DataOps. Acesse os sites e compreenda um pouco melhor como funciona DataOps na prática.

DataKitchen

<https://datakitchen.io/>

Uma das ferramentas DataOps mais populares, DataKitchen é a melhor para automatizar e coordenar pessoas, ambientes e ferramentas em análise de dados de toda a organização. O DataKitchen cuida de tudo - do teste à orquestração, ao desenvolvimento e à implantação. O DataKitchen permite que as organizações criem ambientes de trabalho em questão de minutos para que as equipes possam experimentar sem interromper os ciclos de produção. O pipeline de qualidade do DataKitchen é baseado em três seções principais; dados, produção e valor. É essencial entender que, com esta ferramenta, você pode acessar o pipeline com o código Python, transformá-lo via Linguagem SQL, projetar o modelo em R, visualizar na pasta de trabalho e obter relatórios no Tableau.

Genie

<https://netflix.github.io/genie/>

Desenvolvida pela Netflix, essa ferramenta DataOps é um mecanismo de código aberto que oferece serviços de orquestração de trabalhos distribuídos. Essa ferramenta fornece APIs para desenvolvedores que desejam executar uma ampla variedade de trabalhos com Big Data, usando Hive, Hadoop, Presto e Spark. Genie também fornece APIs para gerenciamento de metadados em clusters de processamento distribuído.

Piper

<https://www.piperr.io/>

Piper é um pacote de ferramentas de DataOps baseadas em aprendizado de máquina que permite que as organizações leiam dados de maneira mais suave e eficiente. Esta solução expõe os dados por meio de um conjunto de APIs que se integram facilmente aos ativos digitais da organização.

Airflow

<https://airflow.apache.org/>

Apache Airflow é uma plataforma de DataOps (e também MLOps) de código aberto que gerencia fluxos de trabalho complexos em qualquer organização, considerando os processos de dados como DAG (Directed Acyclic Graphs). Projetado pelo Airbnb para agendar e monitorar seus fluxos de trabalho, agora as empresas podem utilizar essa ferramenta de código aberto para gerenciar seu processo de dados no MacOS, Linux e Windows.