IBM Databand hands-on Workshop

Welcome to our hands-on workshop where you will learn to deploy databand on an OpenShift cluster. Additionally, we will explore the integration capabilities with Apache Airflow, PostgreSQL, python and IBM Datastage.

Note that there is a full documented set of instructions for deploying databand on a Kubernetes / OpenShift production environment. For educational purposes, some shortcuts have been made during this workshop.

Contents

Video Playlist (in German)

I recorded a quick guide of each chapter in this YouTube playlist

Part 1: Deployment and Setup

Chapter	Content	Video (in German)
0	Prerequisites	Abschnitt 0
1	Hardware provisioning	Abschnitt 1
2	Databand deployment	Abschnitt 2
3	Airflow deployment	Abschnitt 3
4	Airflow integration	Abschnitt 4
5	DataStage integration	Abschnitt 5
6	Postgres deployment	Abschnitt 6

Part 2: Development and Observability

Video Playlist (in German)

I recorded a quick guide of each chapter in this YouTube playlist

Chapter	Content	Video (in German)
7	Preparation	Abschnitt 7
8	SQL Airflow pipelines	Abschnitt 8
9	Python pipelines	Abschnitt 9
10	Python on Airflow pipelines	Abschnitt 10
11	DataStage pipelines	Abschnitt 11

Chapter	Content	Video (in German)
12	Alerts and exceptions	Abschnitt 12
13	Customized user metrics example: **Quantum Computing**	Abschnitt 13

How and why was this workshop developed?

All contents were developed and tested on a MacBook using Microsoft Visual Studio Code with popular extensions for Markdown, Jupyter, Python and bash. Future enhancements will include Windows and Linux to adapt the commands when necessary.

The intention of this workshop was to provide IBMers and its business parterns with a single set of intructions for self-education that can be updated and enhanced by other collaborators using the functions of git repositories. A secondary goal was to minimize or even reduce to zero the ammount of keystrokes necessary to deliver this workshop to a wider audience, either remotely or face to face.

It would be appreciated if the typos, suggestions, improvements, etc. would be reported by opening issues on the git repository. Alternatively, you can send your feedback directly to angelito@de.ibm.com.

Disclaimer

The contents of this workshop have been fully tested and verified but the normal evolution of the many software pieces (most of them open source) will cause disruptions at some point in time. Please apologyze for the inconveniences if that ocurrs. Just open an issue in this repository to let us know. Anyway, we are looking forward to receive your feedback and perhaps your willingness to collaborate and improve this effort.

Acknowledgements

Thanks to Matt Koscak, Adrian Houselander and Elena Lowery for having written the documentation I adapted and used as the starting point to compile this workshop.

Repository mirrors

If you notice formatting failures while displaying the webpages, it may be caused by the differences in the rendering engines hosting the git repositories. Just clone the repository on your local machine and display it with Jupyter. Alternatively, you may try one of these mirrors:

- GitLab: https://gitlab.com/angel-ibm/deployment-databand
- GitHub: https://github.com/angel-ibm/deployment-databand

- Enterprise github (IBM internal): https://github.ibm.com/angelito/databand-workshop

If you need a written documentation, all the contents are also available in PDF and Word format: - Documentation in gitlab - Documentation in github