

ANALYSE DE DATA AVEC HIVE ET AUTOMATISATION DU WORKFLOW OOZIE

Présenté par:

El Mahdi JAMRANI

Abdessamad AHCHOUC

Achraf MZIMIZ

Naoufal CHABAA



PLAN

Introduction

Analyse des données avec Hive

Automatisation des workflows avec Oozie

Demonstration

Conclusion



INTRODUCTION

Dans l'ère du Big Data, Hadoop est essentiel pour stocker et traiter de grandes quantités de données. Hive, un outil clé de l'écosystème Hadoop, facilite l'analyse de ces données grâce à son interface de type SQL. Pour automatiser et orchestrer ces processus, Oozie intervient en tant que gestionnaire de workflows. En combinant Hadoop, Hive et Oozie, les entreprises peuvent non seulement analyser leurs données efficacement, mais aussi automatiser leurs workflows pour une gestion optimale des processus.





QU'EST-CE QUE HIVE ?

- **Hive est une infrastructure de data warehouse construite sur Hadoop, permettant d'interroger et de gérer de grands ensembles de données résidant dans le stockage distribué.**
- **Utilise un langage semblable à SQL, appelé HiveQL.**



AVANTAGES DE HIVE

- Facilité d'utilisation pour les utilisateurs familiers avec SQL.
- Idéal pour les analyses ad-hoc et la création de rapports.
- Supporte les formats de données structurées et semi-structurées.



CAS D'UTILISATION

- Agrégation de données clients pour des analyses marketing.
- Rapports financiers et analyses de performances.



QU'EST-CE QU'OOZIE ?

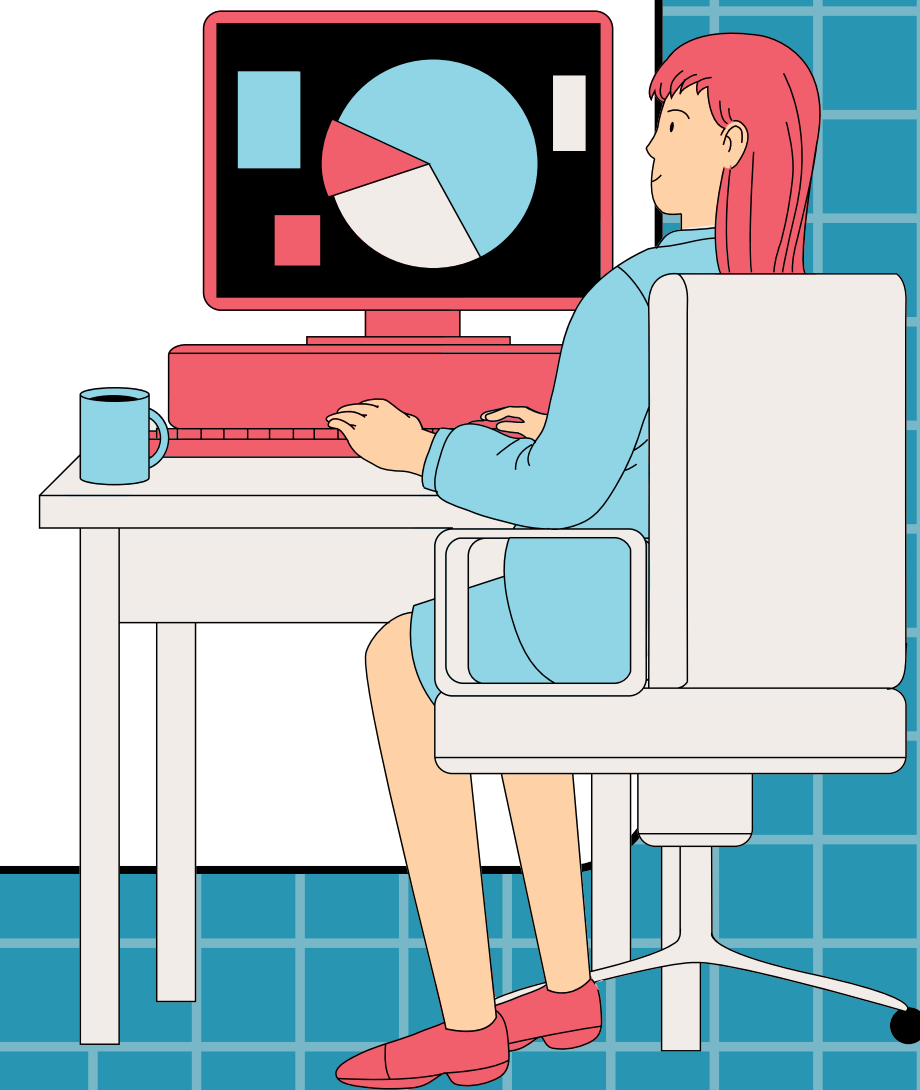
Oozie est un système de gestion de workflow pour Hadoop qui permet de définir, planifier et gérer des workflows composés de tâches de traitement de données.



QU'EST CE QUE HUE



Hue est une interface utilisateur Web open-source pour interagir avec l'écosystème Hadoop. Développé initialement par Cloudera, Hue simplifie l'utilisation de plusieurs services Hadoop, tels que Hive, Pig, HBase, et d'autres outils de gestion de données, en fournissant une interface graphique conviviale.



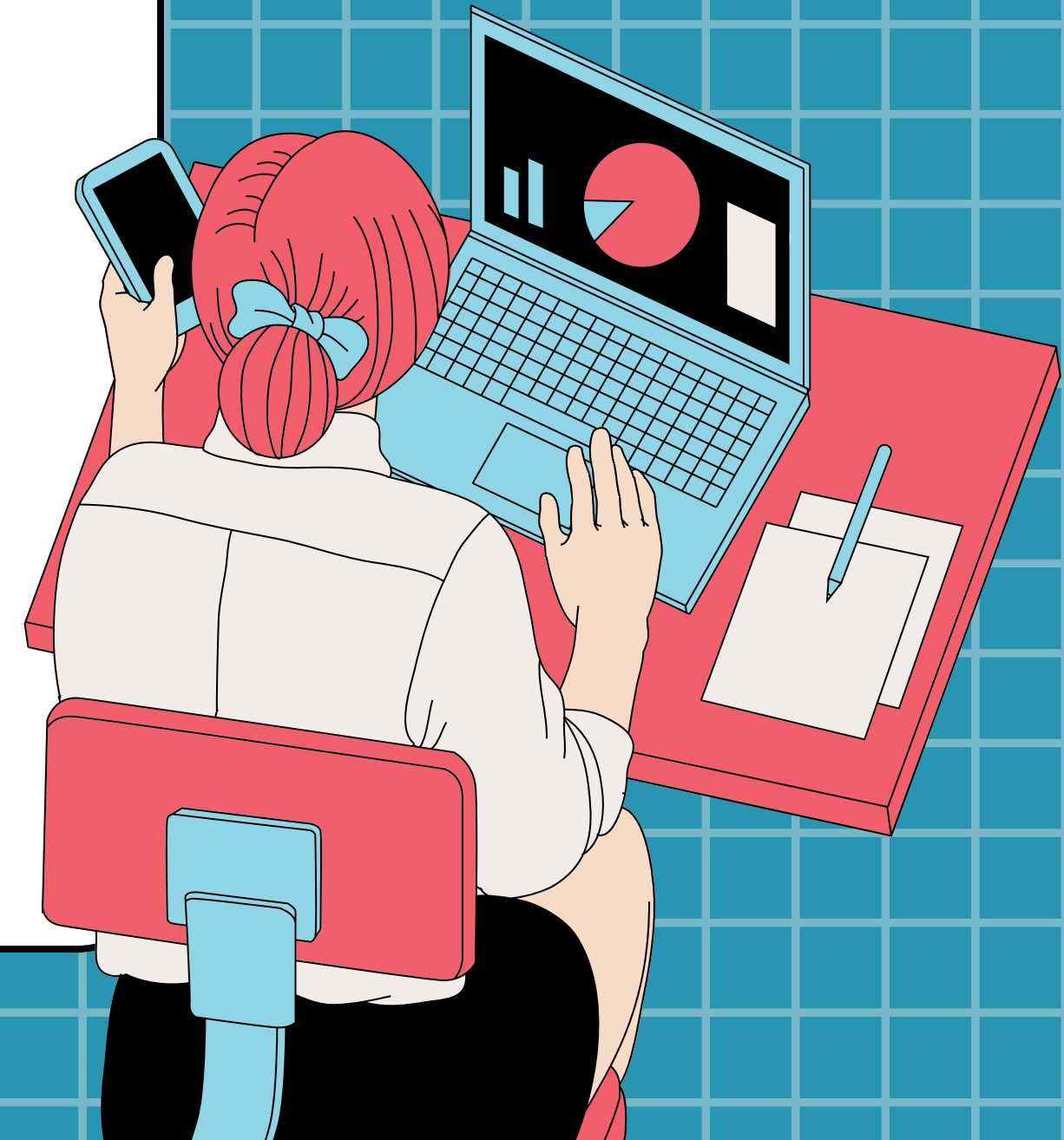
AVANTAGES D'OOZIE

- Coordination de tâches complexes dans un workflow.
- Support pour différents types de tâches : MapReduce, Hive, Pig, Java, etc.
- Intégration avec les outils de planification de tâches existants.



CAS D'UTILISATION

- Automatisation des processus ETL (Extract, Transform, Load).
- Orchestration de pipelines de données pour le traitement batch.

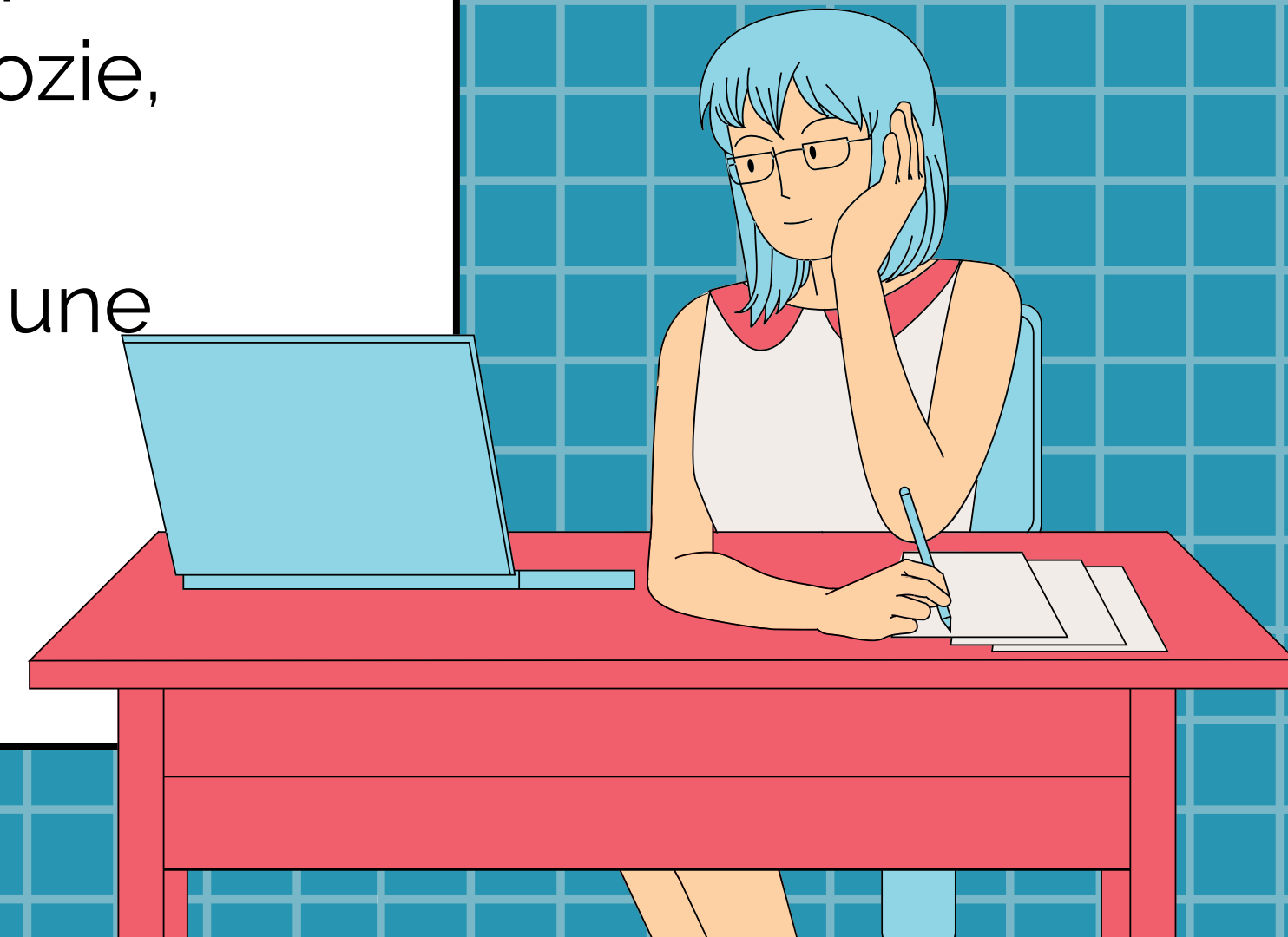




DEMONSTRATION

CONCLUSION

Hive fournit des outils puissants pour l'analyse des données d'entreprise, chacun ayant ses propres avantages et cas d'utilisation spécifiques. Oozie, quant à lui, offre une solution robuste pour l'automatisation des workflows, permettant une gestion efficace des processus de données complexes.



**MERCI DE VOTRE
ATTENTION**

