Final Takeaway

Data is the new oil

Large data repositories
- Data lakes

Azure Data Lake Storage Gen2
- Blob Storage + HDFS
- Best of both worlds

Final Takeaway

Create ADLS Gen2

Many ways of ingesting data
- Distcp
- Azure Data Factory
- AzCopy

REST API

Final Takeaway

Hadoop commands

Spark job
- Hive, Impala, ...

ABFS driver

HDInsight cluster
- Cloudera and other vendors

Security comes first

Takeaway

Spin up clusters as needed

Run Big Data jobs
- Spark
- Hive, Impala, ...

Deprovision when no longer needed

Never lose your data
- Possible to store metadata

Implementing Microsoft Azure Data Lake Storage Gen2

THANK YOU!

Xavier Morera

PASSIONATE ABOUT ENTERPRISE SEARCH AND BIG DATA

@xmorera   www.xaviermorera.com