

# DATA SCIENCE FOR BIG DATA

Empower Your Organization to Leverage Hadoop and Spark for Secure and Scalable Data Science



Enables runtime distribution for Hadoop and Spark jobs



Connects to your Spark clusters

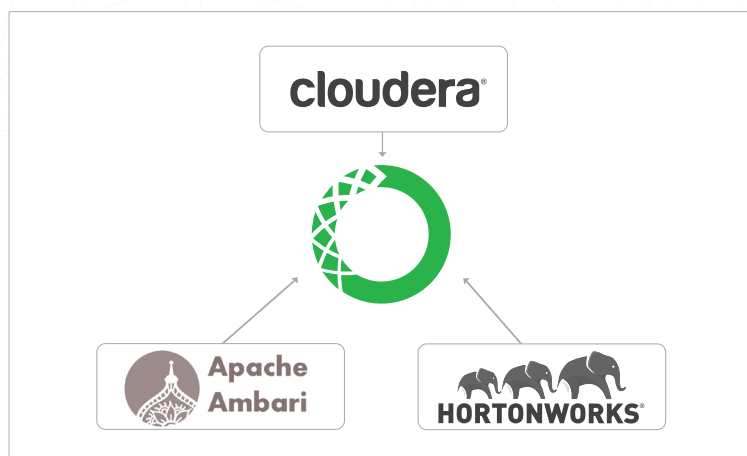


Accesses data from Hadoop clusters



Includes distributed computing with Dask

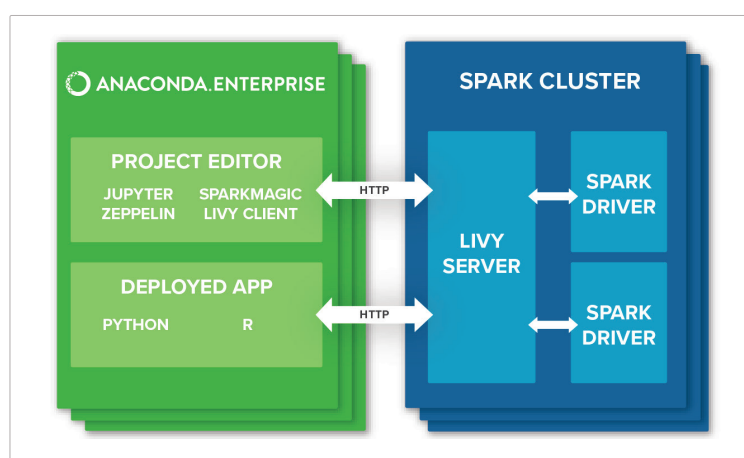
## Runtime Distribution for Hadoop and Spark Jobs



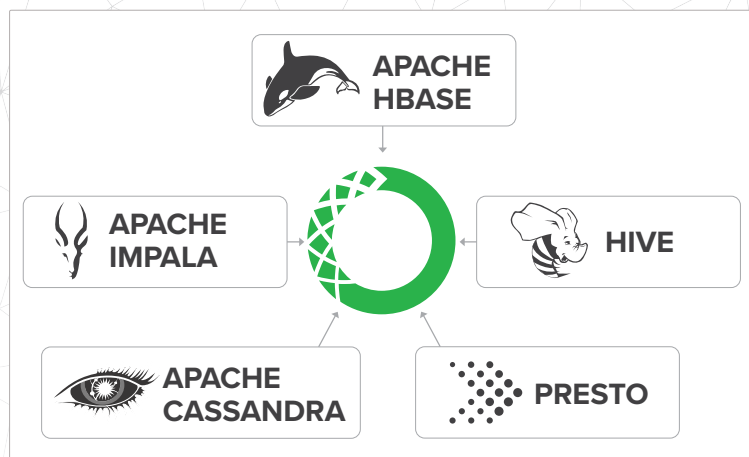
- Distribute Anaconda libraries to your Hadoop and Spark clusters
- Build custom Cloudera Parcels and Ambari Management Packs with Anaconda Enterprise
- knit-conda for on-the-fly runtime package and distribution based on HDFS
- Try free parcel: [docs.continuum.io/anaconda-scale/cloudera-cdh](https://docs.continuum.io/anaconda-scale/cloudera-cdh)

## Anaconda Enterprise Connects to Your Spark Cluster

- Easily connect to one or multiple Spark clusters in JupyterLab or Jupyter Notebooks
- Create Spark projects in your favorite language: Python, Scala or R
- Launch and manage interactive and batch Spark jobs
- Powered by Apache Livy



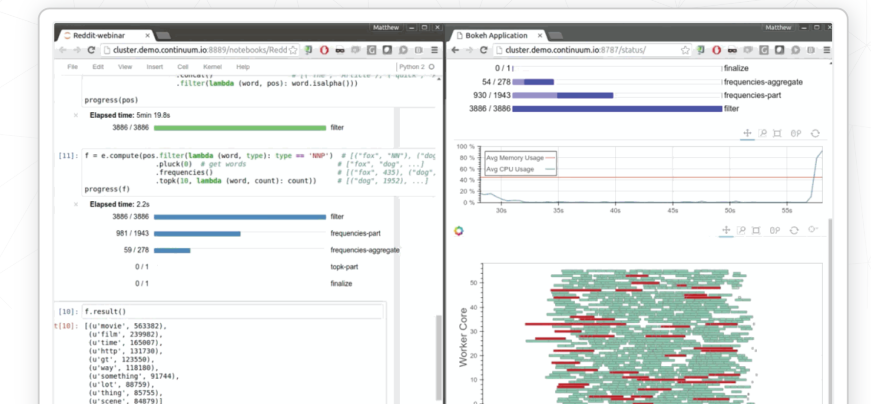
# Query Data from Hadoop in Your Data Science Platform



- Directly access data in secure Hadoop clusters from your choice of SQL or NoSQL datastore through any of the included libraries (Hive, Impala, Drill, Presto and more)
- Build pipelines and dashboards by leveraging any Big Data infrastructure
- Deploy secure and scalable data science projects throughout your organization

## Distributed Computing with Dask

- Open source distributed computing in pure Python with access to the PyData stack
- Provides parallelized NumPy arrays and Pandas DataFrame objects
- Provides rapid feedback and diagnostics to aid humans
- Try Dask and Dask Distributed for free: [dask.pydata.org](http://dask.pydata.org)



## Anaconda Offerings

	Anaconda Distribution	Anaconda Support	Anaconda Enterprise
Distribution	✓	✓	✓
Support	—	✓	✓
Collaboration	—	—	✓
Reproducibility	—	—	✓
Scalability	—	—	✓
Security	—	—	✓
Governance	—	—	✓
Deployment	—	—	✓
Price	FREE	Contact Sales	Contact Sales

Contact Us for a Quote at  
[Sales@anaconda.com](mailto:Sales@anaconda.com) or  
 +1 (512) 776-1066

## About Anaconda, Inc.

With over 4.5 million users, Anaconda is the world's most popular Python data science platform. Anaconda, Inc. continues to lead open source projects like Anaconda, NumPy and SciPy that form the foundation of modern data science. Anaconda's flagship product, Anaconda Enterprise, allows organizations to secure, govern, scale and extend Anaconda to deliver actionable insights that drive businesses and industries forward.