**Product Brief**

# Neo4j Connector for Apache Spark

**The Power of Connections Fueled by All the Data at Your Spark Fingertips**

Neo4j Connector for Apache Spark is an integration tool that moves and reshapes data bi-directionally between the Neo4j graph database platform and Apache Spark™.
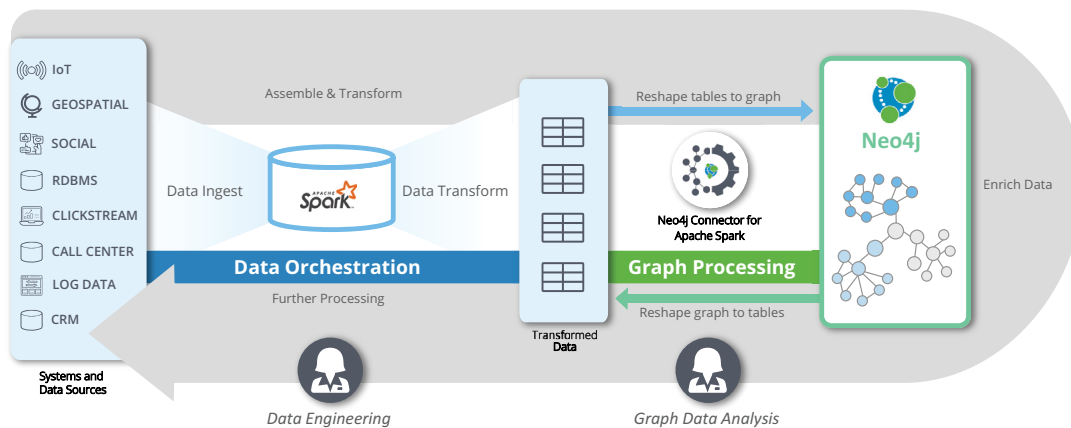
With Neo4j Connector for Apache Spark, users can consolidate their data pipelines and supercharge their Neo4j Graphs with access to the massive Spark ecosystem. The connector allows data scientists and application developers to easily meld Neo4j graph data and Spark data to gain new insights, answer more questions and create new solutions.

## Connect to Myriad Data Sources

Spark is one of the most widely used tools in Big Data. It is the data orchestration layer of choice for many organizations, bringing together data from siloed sources and systems, and readying it for processing and analysis either within or outside of Spark.

As a unified data processing platform, Spark combines many data sources and operational systems such as Teradata, Oracle, Snowflake, MongoDB and more. Spark ingests and extracts data from these various sources, delivers cleansed data to storage systems, such as data lakes and data warehouses, and further transforms data to a structure ready for analysis.

## How Neo4j Connector for Apache Spark Works



Neo4j Connector for Apache Spark is a seamless integration between Spark and Neo4j. Neo4j plugs into the Spark DataSource API to read and write to Spark DataFrames (tables with rows and columns). The connector processes nodes, relationships and queries in both directions using Neo4j's Bolt driver for Java.

## Neo4j Connector for Apache Spark allows you to take advantage of:

- **Rapid access to any data connected to Apache Spark:** Seamless integration between Neo4j and Apache Spark means you easily incorporate the most up-to-date data – regardless of where it lives.

- **Fast and efficient upload of data between Neo4j and Apache Spark:** Parallel and partitioned data loading allows you to upload data in the most performant manner.

- **Flexibility to shape and load only data that is needed – and when it is needed:** Perform on-the-fly data manipulation with graphs utilizing user-defined functions (UDFs).

- **Ease-of-use for developers and data scientists:** Intuitive to use and rapid to deploy, the connecter can be leveraged by data scientists and developers without requiring new skill sets.
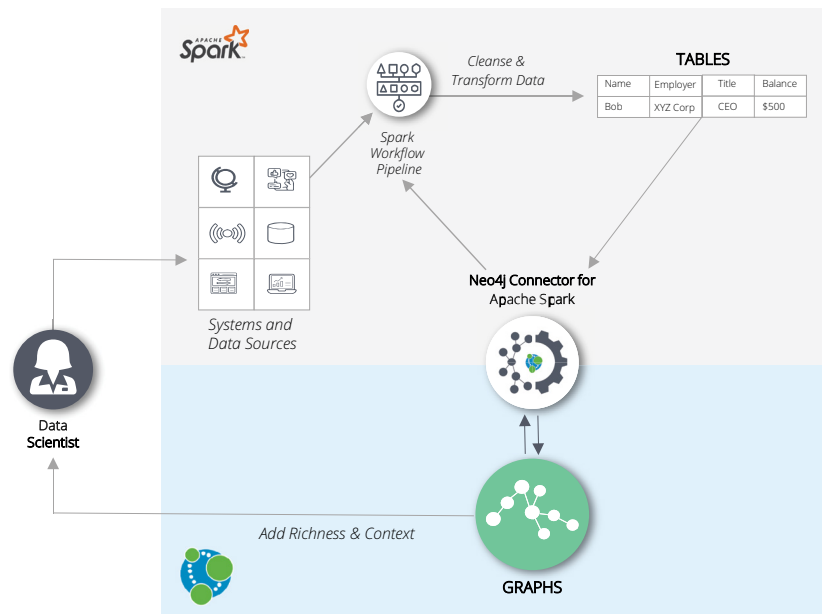
## Product Brief

"Spark has evolved into a viable production platform to meet enterprise needs. It is easy for developers to learn and use to develop solutions. Spark has also cultivated a vibrant community of committers and solutions. Spark's architecture and its applicability to ingest, process and analyze both operational and analytical workloads allow it to reduce the time between obtaining data and delivering insights."*

*Gartner, "Introduction to and Evaluation of Apache Spark for Modern Data Architectures", Sanjeev Mohan and Sumit Pal, January 14, 2020.*

## The Power of Neo4j Connector for Apache Spark

**For Neo4j Customers:** With this connector, Neo4j graphs can connect to any other system or data source via Spark. The Spark Connector transforms tabular data sources to graph data to reveal more context and insight inside Neo4j.

The bidirectional integration means that Spark cleans and transforms data that drives Neo4j graph applications and feeds graph data into any Spark workflow.



**For Spark Users:** The Neo4j Connector for Apache Spark brings advanced graph capabilities to the Spark ecosystem so businesses can use contextual information to improve forecasting, analytics and predictions. This connector enables teams to easily add Neo4j graph-enriched data to improve high-value processes, like machine learning, without reworking existing pipelines.

**Neo4j Connector for Apache Spark is available for free with full support provided to Neo4j Enterprise Edition customers. Visit** neo4j.com/product/connectors/apache-spark-connector/ **for more information.**

---

Neo4j is the leading graph database platform that drives innovation and competitive advantage at Airbus, Comcast, eBay, NASA, UBS, and more.

Thousands of community deployments and more than 400 customers harness connected data with Neo4j to reveal how people, processes, locations and systems are interrelated. Using this relationships-first approach, applications built using Neo4j tackle connected data challenges including artificial intelligence, fraud detection, real-time recommendations and master data. Find out more at Neo4j.com.

Questions about Neo4j?

Contact us around the globe:
**info@neo4j.com**
**neo4j.com/contact-us**

---