



# Intel<sup>®</sup> Analytics Toolkit - Summary

The revolution in “big data” is transforming industries and research, while spawning new solutions to a range of societal challenges. Big data strategies usually begin by capturing the high volumes of varied data using Apache\* Hadoop\*-based platforms that have massive scalability, cost effectiveness, and a vibrant open-source ecosystem. But once captured, achieving anticipated insights remains elusive. High demand data science expertise is scarce, exacerbated by added skills to program across a myriad of open-source tools and working through workflows inefficient for iteration and collaboration. Finally, tools used are often geared to answering *known* questions, with limited workable methods to easily find hidden signals in data patterns and connections.

The Intel Analytics Toolkit addresses these barriers to achieving value from big data and enables data scientists to achieve greater insights, more quickly, and with reduced complexity. A simpler programming environment lets data scientists focus on analytics instead of mastering the details of programming to Hadoop and the myriad of open source tools. Data scientists can orchestrate and easily iterate through the end-to-end analytics workflow in a single program, using a familiar programming language that executes analytics using fully scalable algorithms. Out of the box, the platform unifies entity based machine learning with an end-to-end graph processing pipeline including powerful algorithms for uncovering relationships hidden in big data. And the modular framework enables users or developers to extend and integrate new analytics functionality and algorithms.

By bringing simpler analytics programming and the full range of graph processing capabilities to the Hadoop\* “data lake”, the Intel Analytics Toolkit is helping to democratize and accelerate big data powered solutions.

Capability	Benefit
Easier big data analytics programming using Python	<ul style="list-style-type: none"> <li>• Cluster scale programming as easy and intuitive as desktop analytics</li> <li>• One environment to use across multiple engines</li> <li>• Simpler iteration &amp; collaboration</li> </ul>
Fully scalable graph processing (graph database, distributed algorithms, graph ETL) for analyzing big data graphs.	<ul style="list-style-type: none"> <li>• Easier and faster analysis of connected data</li> <li>• Quickly query and traverse across relationships on any size of data</li> <li>• Reduce graph analytics effort and learning curve</li> </ul>
Unified graph and entity based analytics <sup>1</sup>	<ul style="list-style-type: none"> <li>• Enhance model performance with relationships</li> <li>• Learn and maintain fewer environments</li> <li>• Explore multiple analytic approaches for unconstrained data science creativity</li> </ul>
Scalable, distributed out-of-the-box algorithms	<ul style="list-style-type: none"> <li>• Easily build solutions to a range of big data challenges, such as link analysis, pattern recognition, segmentation, recommendations, topic modeling, and community detection</li> </ul>
Libraries to engineer features, transform data, and segment models <sup>2</sup>	<ul style="list-style-type: none"> <li>• Reduce the data preparation effort</li> <li>• Focus on analysis vs. data manipulation</li> <li>• Generate graphs from un- &amp; semi-structured data</li> </ul>
Extensibility of analytics functionality & programmability <sup>3</sup>	<ul style="list-style-type: none"> <li>• Easily incorporate custom algorithms</li> <li>• Adopt new analytics technology without burdening data scientist productivity</li> <li>• Standardize the data science workflow</li> </ul>

Please note the following for the .8 beta release scope

<sup>1</sup>Entity-based machine learning ships in a subsequent release

<sup>2</sup>Most .8 release transformations require user-defined Python functions

<sup>3</sup>Extensibility will be documented and exposed in a subsequent release