---
1. Spark Core

Paper: Resilient Distributed Datasets: A Fault-Tolerant Abstraction forIn-Memory Cluster Computing

URL: https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final138.pdf

@INPROCEEDINGS{*SparkCore,*
***author** = "Matei Zaharia and Mosharaf Chowdhury and Tathagata Das and Ankur Dave and Justin Ma and Murphy McCauly and Michael J. Franklin and Scott Shenker and Ion Stoica",*
***title** = "Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing",*
***booktitle** = "9th Symposium on Networked Systems Design and Implementation (NSDI'12), 15--28",*
***publisher** = "USENIX",*
***year** = 2012,*
}

---
2. Spark SQL

Paper: Spark SQL: Relational Data Processing in Spark

URL: https://dl.acm.org/doi/pdf/10.1145/2723372.2742797

@INPROCEEDINGS{*SparkSQL,*
***author** = "Michael Armbrust and Reynold S. Xin and Cheng Lian and Yin Huai and Davies Liu and Joseph K. Bradley and Xiangrui Meng and Tomer Kaftan and Michael J. Franklin and Ali Ghodsi and Matei Zaharia",*
***title** = "Spark SQL: Relational Data Processing in Spark",*
***booktitle** = "15th International Conference on Management of Data (SIGMOD'15), 1383--1394",*
***publisher** = "ACM",*
***year** = 2015,*
}

---
3. Spark Streaming

Paper: Discretized Streams: Fault-Tolerant Streaming Computation at Scale

URL: https://dl.acm.org/doi/pdf/10.1145/2517349.2522737

@INPROCEEDINGS{*SparkStreaming,*
***author** = "Matei Zaharia and Tathagata Das and Haoyuan Li and Timothy Hunter and Scott Shenker and Ion Stoica",*
***title** = "Discretized streams: fault-tolerant streaming computation at scale",*
***booktitle** = "24th ACM Symposium on Operating Systems Principles (SOSP'13), 423--438",*
***publisher** = "ACM",*
***year** = 2013,*
}

---
4. Spark Structured Streaming

Paper: Structured Streaming: A Declarative API for Real-Time Applications in Apache Spark

URL: https://cs.stanford.edu/~matei/papers/2018/sigmod_structured_streaming.pdf

@INPROCEEDINGS{*SparkStructuredStreaming,*
***author** = "Michael Armbrust and Tathagata Das and Joseph Torres and Burak Yavuz and Shixiong Zhu and Reynold Xin and Ali Ghodsi and Ion Stoica and Matei Zahariaa",*
***title** = "Structured Streaming: A Declarative API for Real-Time Applications in Apache Spark",*
***booktitle** = "18th International Conference on Management of Data (SIGMOD'15), 601--613",*
***publisher** = "ACM",*
***year** = 2018,*
}

---
5. Spark GraphX

Paper: GraphX: Graph Processing in a Distributed Dataflow Framework

URL: https://amplab.cs.berkeley.edu/wp-content/uploads/2014/09/graphx.pdf

@INPROCEEDINGS{*SparkGraphX,*
***author** = "Joseph E. Gonzalez and Reynold S. Xin and Ankur Dave and Daniel Crankshaw and Michael J. Franklin and Ion Stoica",*
***title** = "GraphX: Graph Processing in a Distributed Dataflow Framework",*
***booktitle** = "11th Symposium on Operating Systems Design and Implementation (OSDI'14), 599--613",*
***publisher** = "USENIX",*
***year** = 2014,*
}

---
6. Spark GraphFrames

Paper: GraphFrames: An Integrated API for Mixing Graph andRelational Queries

URL: https://cs.stanford.edu/~matei/papers/2016/grades_graphframes.pdf

@INPROCEEDINGS{*SparkGraphFrames,*
***author** = "Ankur Dave and Alekh Jindal and Li Erran Li and Reynold Xin and Joseph Gonzalez and Matei Zaharia",*
***title** = "GraphFrames: An Integrated API for Mixing Graph andRelational Queries",*
***booktitle** = "4th International Workshop on Graph Data Management Experiences and Systems (GRADES'16), 2",*
***publisher** = "ACM",*
***year** = 2016,*
}