

**JupyterCon**  
August 22 -25, 2017  
New York City  
Ajay Asthana  
Partner, Cabot Partners



## **Expanding from Thousands to Millions – A Story of the Meteoric Growth of Jupyter Notebook**

JupyterCon 2017 was about Jupyter Notebooks, an open-source web application that enables creating and sharing documents with embedded live code, equations, visualizations as well as explanatory texts.

What is so unique about a platform that makes applications for notebooks? Well, what makes Jupyter appealing is its ability to provide a comprehensive, browser-based, interactive development environment that can organize all data analysis components. These include operations such as wrangling, exploration, visualization, analysis, and modeling, along with text and equation markdowns. Why should we be interested in such platforms? Jupyter has made one of the most significant advances in the scientific computing arena as it enables users to formulate and compose their thoughts with prose, enhanced by expressing these in mathematical equations and embellished with executable code. This synergy of thought, mathematics and code promote collaboration and open source development.

In August this year, I had the opportunity to attend JupyterCon 2017 at New York City and was amazed at the organization there. The venue was buzzing around with approximately 800 data scientists, educators, students, developers, and tool creators, participating in the insightful keynotes, training sessions, tutorials, vendor presentations, and platform explorations.

The conference was kicked off by [Fernando Perez](#), co-founder of Jupyter Notebooks and professor at the University of California, Berkeley, by giving us a historical perspective of the project, its dynamic acceleration, and ready acceptance among data scientists. Jupyter Notebooks started from humble beginnings in a university to a currently estimated six million to eight million copies, worldwide. The audience could feel the palpable energy, emotion, and efforts that went into the development and shaped this eventual success. Rachel Thomas of [fast.ai](#) explained how she taught profound practical learning using Jupyter notebooks which provided an encouraging environment for sharing best practices. Peter Wang of Anaconda, described the shared theme between Jupyter and Anaconda and indicated that both the technologies are at an inflection point to become mainstream.

The presentation from Wes McKinney's of Two Sigma investments dealt with the artificial intelligence arms race that is happening to get faster, more scalable, cost-effective machine learning. He made a case for a shared infrastructure and computation runtime for data scientists. Professor [Demba Ba](#) from Harvard University concluded the first-day keynotes by emphasizing that data science is a critical skill for the citizens of the modern world. He talked about how he brought Jupyter notebooks to the classroom by using it as a platform for teaching signal and data processing to students, democratizing and empowering them simultaneously.

The theme of the second day of the conference was on the adoption of Jupyter Notebooks across industries, collaboration required to continue the enhancement of the application, and acceptance of the Jupyter platform across both the academia and the industries. The second day's content complemented the previous day's talks which focused on history, education, growth, and issues of sustainability of increasing demand.

Jeremy Freeman of the Chan Zuckerberg Initiative indicated how the application is helping genomic and imaging research by eliminating the computation bottleneck, data sharing, interdisciplinary collaboration, modular analysis, and open knowledge dissemination. William Merchan from the DataScience.com, highlighted the three trends accelerating the adoption of Jupyter in the enterprise – the acceptance of data science across industries, tenfold growth in the use of Jupyter, and, Jupyter becoming the tool of choice for innovation.

Lorena Barba of George Washington University explored design for reproducibility and scientific advancement by using Jupyter to obtain better insights and use man and the computer collaboration in research. Andrew Odewahn from O'Reilly Media explained their long track record of spotting early innovation and how O'Reilly Media applied the Jupyter architecture to create the next generation of technical media. Brett Cannon of Microsoft gave very humorous and informative insights into the give and took of open source software, and presented guidelines to use, share, and provide constructive feedback. Nadia Ehnbal (GitHub), concluded the keynotes by highlighting the role of money in open source, by articulating the motivation of the contributors and spending available funds carefully.

In parallel to the talks, several presentations unfolding new frameworks, practical uses in industry and how universities are developing curriculum around it. I enjoyed listening to Subbu Rama of Bitfusion on deep learning and elastic GPUs session where he combined technology with economic value and how the platform could reduce the cost of processing. Raj Singh of IBM presented on mapping data in Jupyter notebooks with IBM's Pixi Dust to simplify spatial visualization and the magical self-healing servers. Pramit Choudhary of Datascience.com showed how to use their platform to explore better ways to evaluate and explain model learned decision policies, which is so important to refine the artificial intelligence algorithms continuously.

Finally, I would like to acknowledge the organizers, especially, Maureen Jennings, Media Relations Manager of O'Reilly for giving me an opportunity to hear these inspiring talks. I am excited for the future of Jupyter notebook as well as JupyterCon and am looking forward to more opportunities to hearing about them. Those of you who are interested in conference presentations and links to videos, please check <https://conferences.oreilly.com/jupyter/iup-ny>.