



September 11, 2018

5 things I Learned at the JupyterCon 2018 Conference

Ajay Asthana, Cabot Partners

JupyterCon 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.

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 thinking, mathematics, and system promote collaboration and open source development.

Last week, I had the opportunity to attend the second **JupyterCon 2018 in New York City**. The addition of Education Track and Business Summit to supplement the Key Notes, Trainings and Tutorials was a pleasant surprise. The venue was buzzing around with more than 2,000 data scientists, educators, students, developers and tool creators, participating in the insightful explorations. I concentrated on keynotes and business summit, and there are the five things I learned at the conference:

1. Large organizations including many Fortune 500 companies are adopting Jupyter to address their domain-specific computations challenges
2. Respected educational institutions are using Jupyter to train data scientists and make them available to the industry
3. Prestigious research institutions are using Jupyter to collaborate on complex problems involving structures and semi-structured data
4. Industries with complex regulations are identifying ways to address business, legal and technical issues associated with using open source solutions.

5. Jupyter is addressing the “technical chasm” by bringing together business and technical personnel using open source software.

The next five paragraphs describe the sessions in detail associated with the five lessons learned based on the sessions I attended:

1. **Paco Nathan of Derven.ai** explained how Jupyter is helping large organizations to “future proof” the movement of software engineering practice from hardware to software to process and challenges associated with collaboration, data privacy, ethics, security, compliance, etc. **Ryan Abernathy of Columbia University** made a case for partnership between academia, government, and industry to move away from siloed and on-premise research to pooling all data centrally on a cloud to conduct experiments and share results. **Brian Granger from California Polytechnic San Luis** articulated the business case for the use of Jupyter in corporations outlining benefits of disruptive innovation – a rich ecosystem, risks of a slow movement, bug fixes, no licensing fees, and lack of enterprise support. **Julie Lane from NYU** presented on the need for public policy to ensure the safety of people, setting, projects, outputs, and data.

2. **Fernando Perez of University of California, Berkeley**, described the popularity of data science course where 40% of the students are enrolled in it; Using Jupyter their projects are extending data science into specific domains that cover virtually all disciplinary areas of the campus. **Carol Willing of Cal Poly San Luis Obispo** explained how Jupyter is revolutionizing education, research, and business by creating opportunities for exploration, storytelling with data, communication, and developing products that produce substantial margins. **Mark Hansen from Columbia School of Journalism** described how Jupyter is used as a computationally expressive “reporter’s notebook” by journalists for their investigations incorporating data, code, and algorithms both as tools to report with as well as tools to report on. **Tracy Teal of Data Carpentry** who explains how to bring people to data and empower them to address their questions, reach their potential, and solve issues that are important in science, scholarship, and society.

3. **William Far of Stony Brook University** defined how global scientific collaboration was facilitated by Jupyter in astronomy to detect first-ever gravitational waves from the merger of a pair of stellar-mass black holes about 1.5 billion light-years away. **Cristian Capdevila of Prognos** explained how they are combining machine learning techniques and clinical expertise to a large clinical database to predict the onset of diseases. Using Jupyter they are able to accelerate model development, training, and deployment. **George Williams and team of GSI Technology** described the dedicated extensions for data scientists working in cyber security that can be used and deployed via JupyterHub to detect threats quickly and address them. **Catherine**

Ordun of Booz Allen and Co. showed the value of transparency and how it helped in establishing a data science group at their customer by presenting the details abstract and detail

4. **Luciano Resende of IBM** described how IBM Watson products leveraged the Jupyter stack to address enterprise requirements and offer industry-leading and business-critical services to their clients. He also outlined a pattern for building deep learning for cost-effective full dataset training of deep learning models. **Dave Steward of Department of Defense's** presentation on using Jupyter at DoD to improve efficiency, collaboration, and approachability in an extremely complex organization and painted a positive picture on its adoption in highly regulated environments. **Shivraj Ramanan and team from Capital One** described the challenges the bank faced in using the Jupyter in a regulated environment, such as data sharing, metadata, data quality, decisions, steps, availability, no data loss, policy and market shifts and provided insights on how they overcame them. He concluded that the benefits of building models centrally and distributing them, showing results and computation together, using right data by right people compensate for all the challenges.

5. **Matt Greenfield of Two-Sigma** outlined his company's contribution to the open source community including - funding startups, paying for contractors, hiring engineers in-house, and building a consortium with committed stakeholders – at accelerate the development (currently there are 27 million downloads for Jupyter and only 10 full time personnel). **David Schaaf of Capital One** explained how data science and data engineering can work together in cross-functional teams—with Jupyter notebooks at the center of collaboration and the analytic workflow—to more effectively and more quickly deliver results to decision makers. **Gerald Rouselle of Teradata** reviewed trends in modern data and analytics ecosystems for large enterprises and shares some of the key challenges and opportunities for Jupyter adoption. **Michelle Ufford of Netflix** described how notebooks are the future and would be used by everyone in their organization to do their job and help collaboration between the business and the technical personnel.

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 for understanding about them. Those of you who are interested in conference presentations and links to videos, please check <https://conferences.oreilly.com/jupyter/jup-ny/public/schedule/proceedings>.