Tim Tully (1:34)

One very bright spot has been the work that you do. Entire companies, maybe yours, went remote overnight. Brick and mortar stores and restaurants embraced e-commerce. The use of digital tools by schools of medicine skipped ahead by years. Between navigating a pandemic, wildfires, homeschooling, and everything else, you made incredible things happen.

It's safe to say that none of this would be possible without the cloud. That's why the cloud has become an essential service, and you and so many of your teams across your organization have accelerated your cloud journey.

But that journey can be fraught with complexity that's amplified by your limited visibility. These challenges are exactly why we're accelerating Splunk innovation. As your cloud strategies and architectures change, you'll have new data platform requirements. You need to be able to stream data to collect, analyze, and publish data into Splunk or non-Splunk syncs, and you need automation and machine learning to detect and respond at machine speed. You need an index that's fantastically scalable for you to get more data in, and also federated search and orchestration all at once.

A few organizations will help you fill one or two of these new requirements. Only Splunk has industry-leading offerings in all five. This blows me away and should blow you away too. Again, only Splunk has the best offerings across each of these categories.

This is the way Splunk thinks about the single cloud platform you and all of your teams across IT, security, and observability and beyond will tackle these modern requirements. Splunk is the modern, flexible platform you need. Thousands of organizations around the world see Splunk as mandatory components to their data architectures and strategies while providing competitive advantages because of our unified approach and our continual innovation.

Tim Tully (03:19)

There's so much around what we're innovating on the platform, but today let's focus on three particular areas we're advancing data science. I can't wait to talk to you about what we're doing across stream processing, machine learning, and observability. Let's start with stream processing.
Today's IT and development environments are changing fast. More and more workloads are being moved to hybrid or full on cloud, and DevOps teams are now releasing code on a daily and hourly basis. Ephemeral and software-defined infrastructure spin up and down in a matter of minutes. Serverless functions last for seconds. Services are continually brought up and down on the fly.

We live in a dynamic software and network-driven world. In order to monitor all of this dynamic behavior, we need stream processing. Stream processing is a fundamental capability of data systems today. It allows you to see and process all of your data in real-time. It also provides a data delivery bridge, unifying your data between on-premise and the cloud.

Stream processing is built into many of our solutions already, including Mission Control, observability, and even things we haven't announced yet. We're the only company that can process structured and unstructured data with low latency and at any scale.

Tim Tully (04:33)
Last year, we introduced stream processing on our platform through our millisecond latency streaming engine, Splunk Data Stream Processor or DSP. This engine is central to our platform vision moving forward. It's essentially a consolidated data bus for the enterprise that helps you understand your data in the stream in real-time, transforms and aggregates your data from your formats to destination-specific formats such as Zipkin. Then based on that insight, intelligently routes the data to where it will have the most impact across your NOC, SOC, or APM observability. It doesn't matter if you're bringing in standard metrics, traces or log data, or even custom formats, Splunk analyzes all of that data in flight. We don't wait for it to rest or to settle.

What happens when you take the best of open-source software in the stream processing and analytics space, Apache Pulsar and Apache Flink, and add a lot of Splunk secret sauce engineering from the world's best data engineers? You get Splunk's stream processing engine. No one else can offer a streaming capability that processes petabytes of high cardinality data at millisecond level speed, with the ability to convert logs and traces to metrics on the stream.

Stream processing is how you gain better control of your data sprawl, enrich your data by looking at additional attributes to create more complete events, conduct the analytics on the stream, and route data to wherever you need it based on that dynamic decisioning.

Given that you can do a lot with the streaming technology, by collecting all data sources in one place, you can improve your data visibility across your multi and hybrid infrastructure sprawl. You can detect drift earlier on in your data life cycle and ensure downstream KPIs are accurate. You'll consolidate and filter based on what you want to query and add contextual details to
events. You'll identify and mask sensitive data early, reducing risk. Most importantly, you'll intelligently route data to where it should be going.

These are needs across IT, security, and observability. I personally know about these challenges because Splunk's IT and security orgs are part of our broader tech org here at Splunk. This helps ensure we build the best IT and security products because we are our own best customer. Splunk stream processing powers use cases across all three in a way no other vendor is doing. Let's talk out a security example.

Tim Tully (06:41)

As I mentioned before, I know you're likely overwhelmed with the amount of data available, specifically data that is high volume and noisy, like Windows and Linux audit data along with flow data. Often these data sources are sent for longer-term storage, but they hold incredibly valuable information that needs to be leveraged to minimize your security vulnerabilities.

DSP allows you to build that data pipeline so you can centralize all data ingestion across your organization. Then it allows you to aggregate noisy events on the stream from your audit and flow data, pulling out relevant insights that you need now. Then you can enrich these events with anomaly scores.

You can route those events with high anomaly scores directly to your SOC team's index while maintaining full fidelity of events to the main index, helping you isolate critical data earlier. For larger organizations with multiple subsidiaries and cloud instances, you can help your security teams even further by consolidating and tagging any security-related data on the screen.

This means they don't have to look in multiple cloud instances for each business. Instead, this data is tagged then routed directly to the SOC for further investigation, improving your visibility and overall security monitoring across your data fabric. Only Splunk can do this at unprecedented scale without taxing your downstream systems.

Last year, our customers like you began sharing our vision for faster insight through stream processing, especially as they transition to the cloud. That's why today I'm excited to talk about what's new with DSP, Splunk Data Stream Processor 1.2.

New functionality within DSP will make it an incredible 400% faster. The Splunk Data Stream Processor will enable new access and support for multi-cloud environments, including Google Cloud Platform Pub/Sub and the Microsoft Azure Event Hub, enabling you to ingest and route data to other teams and systems while managing your cloud infrastructure sprawl.
DSP will also accelerate your insights through enrichment. Our new lookups provide in-stream data enrichment at scale, making your downstream searches more relevant and accurate. DSP also enables you to apply our machine learning functionality on data as it arrives in the stream. Let's talk more about machine learning, or ML as you might hear us call it.

Tim Tully (08:50)
With the explosive volume in a variety of data from your hybrid and multi-cloud environments, I know it's not practical to expect you to pour through log data looking for security or operational red flags, and yet offline machine learning and approaches have been building statistical modeling engines that aren't live or able to adapt to the evolution of data such as seasonality.

Drawing accurate insights from data requires online machine learning that can autonomously learn from the data they process to perform and improve the performance of specific tasks. None of the ML you've used from other vendors has allowed you to do that. But as we say, if you're not continuously learning, you're continuously behind.

As you learned from us at .conf19, Splunk rebuilt our machine learning algorithms in a completely novel way for ML to work in the stream. Splunk's machine learning unleashes your ability to learn, infer, and analyze data in the platform.

Unlike traditional batch ML systems where you have to train and validate models and constantly update them, our ML learns continuously, so our models are always up-to-date. Moreover, our machine learning scales seamlessly, independent of volume and cardinality.

Streaming data is an ideal application for machine learning, but high memory batch algorithms simply can't search across all of your valuable data without running into resource constraints. Even if these best jobs managed to complete, the results will be so delayed that the output is not actionable and the bad actors you identified have already gotten away.

We know streaming is only going to grow in importance, so we have focused the vast majority of our machine learning innovation to the stream, rewriting all algorithms from scratch, creating an online learning world where you're unrestricted in learning by volume, cardinality, and flexibility. We call this Streaming ML.

The ways customers use Streaming ML has set the Splunk platform apart from any other company or solution. Only Splunk has the guts, grit, and commitment to innovate in this way, which is why today I'm proud to announce that Splunk will be offering additional streaming machine learning algorithms in DSP 1.2.
We're now enabling data scientists to reach insights faster with new algorithms, such as online time series decomposition or online STL, which normalizes datasets by decoupling seasonality. This accelerates pattern detection and things like inventory management, which can lead to greater accuracy in purchasing, stocking, and staffing decisions.

Next, we enable categorical outlier detection, which is useful in security or IT domains when trying to detect outliers with non-numeric data values. Last, we offer approximate percentiles, helping you categorize data in specific groupings so you can answer questions like which machines are generating the top traffic or cell phones that are in the top 5% of bandwidth consumption.

That's on top of what we already have with algorithms like adaptive thresholding, which is really just a fancy way of describing anomaly detection. But let's stop talking and show it to you. To give you a demo, please welcome Splunk's head of machine learning, Ram Sriharsha.

Ram Sriharsha (11:45)
Thanks, Tim. The scenario that I want to discuss today is a very common pain point that many of our customers in a NOC face. Imagine that you're monitoring hundreds of thousands of metrics that are being logged by diverse applications and hosts. How do you monitor them for outliers?

At this scale, you cannot possibly set static thresholds on a per entity or even a per KPI level, and setting one static threshold across all entities and KPIs simply doesn't work, as at least there's too much noise. You need outlier detection algorithms that can adaptively threshold based on the properties of the metrics you're monitoring and relieve you from the burden of having to manually configure and tune metrics at this scale.

It does seem like a daunting task, but we have you covered. These are the sort of problems that Streaming ML was designed to solve. In this Splunk dashboard, we can see all the data that's coming into our pipeline at the source, the rate at which it is being processed, and the outliers that are being detected in stream being processed and the outliers that are being detected in-stream. Let's see how this works. To do so, I'm going to open up the data stream processor and show you the pipeline that is streaming data to our Splunk Enterprise dashboard. You can get a preview of the volume and variety of data that's streaming through this pipeline through our preview UI. With lookups, you can add contextual information to events on the stream and quickly turn error codes into human-readable form.

By leveraging out-of-the-box functionality in DSP, we can easily perform a count of these events by error type and host to convert these logs into metrics. Now that we have our logs metricized,
we would like to go ahead and monitor these metrics for outliers. One such algorithm that's available in streaming ML for monitoring outliers is adaptive thresholding. Leveraging adaptive thresholding is as simple as dragging and dropping this operator into the pipeline. There is no need to pre-train models to figure out how to deploy these models or how often to retrain them. The algorithm simply learns on the fly as it sees data, what normal looks like, and uses that to detect outliers in the stream.

As you can see, the output column of this algorithm generates a Boolean label, which is true when the point is detected as an outlier. By activating this pipeline, we have an outlier detection algorithm that's running at scale on this metrics pipeline, routing the data to Splunk, SignalFx, and S3 for further observability and monitoring. That is all it takes for you to start leveraging out-of-the-box machine learning capabilities in your streaming pipelines. With that, I will turn back to Tim, and then see you again in a few minutes.

Tim Tully (14:20)
Ram, thanks for that demo. It's incredible to see these updates come to life after years of R&D. No one else in the industry is building ML into the stream and we know it's going to be transformative for our customers. In line with our goal to provide the world's best data platform, the next aspect we're focused on is data science. The amount of data in the Splunk platform makes it great for advanced analytics and data science use cases.

We know that a number of you have adjunct data scientist teams that collaborate with your everyday Splunk admins. We also know this collaboration at scale introduces a whole new set of challenges. Moving data from Splunk to a different platform at such volumes can be cumbersome at best, or come with such a high opportunity cost that it isn't worth it. We ask ourselves, how can we create a world where companies like yours can splunk data and do data science at the same time, but with familiar open-source tools? Our solution was a Splunk machine-learning environment.

Today, we are excited to announce a new beta program that we know will make you smile. Today, we are announcing the beta program of Splunk machine learning environment or smle for short. smle is Splunk's rethinking of the data scientist's workbench and is squarely in line with our data-platform view of the world. As your team's data scientists look to custom build your own algorithms, you'll no longer have to move data from Splunk. Here, we've created a world where SPL and data scientists come together on the same platform for optimal speed and insights together.

First, we folded in Jupyter notebooks, the fastest-growing open-source data science platform into Splunk, as an experimentation experience, with Jupyter speaking SPL under the hood. This
enables data scientists to switch between using R, Python, Scala, and SPL for greater user access. Your data scientist and Splunk admin can also collaborate organically without having to learn each other’s skills.

Second, in this experience, data scientists can embed batch or streaming machine learning seamlessly in their Splunk queries. This experience is not restricted to data scientists, but also those familiar with RSPL ecosystem. Splunk admins, for example.

Third, building on not just Jupyter, but also world-class capabilities like DSP in our cloud platform, we provide the first seamless experience, allowing data scientists to deploy, manage, and monitor models in production, all at scale, right from their Jupyter notebooks. Like DSP, this is groundbreaking science right here, offering machine learning as a service. All those streaming machine-learning algorithms we just talked about powered by DSP 1.2, are available in SMLE out of the box.

As you see, our Data-to-Everything Platform is state-of-the-art, in our innovation, in stream processing, and machine learning, equip you to address your biggest problems. In fact, our IT, observability, and security customers are using stream processing and machine learning for key capabilities no other engine can enable. To show you the power of SMLE, please welcome back our head of machine learning at Splunk, Ram Sriharsha.

Ram Sriharsha (17:03)
Thanks for inviting me back, Tim. In order to fully support end-to-end machine learning, we wanted to provide a way for data scientists to easily analyze data at scale, leveraging our machine learning capabilities together with those available in the rich Python ecosystem, to develop machine learning pipelines and easily deploy them to DSP at scale. With SMLE, we have built a Jupyter lab experience with Python R and deep-learning libraries baked in, all with one Splunky twist. We are extremely excited to announce that we have an SPL kernel that allows you to mix and match your coding language of choice while still leveraging the power of SPL.

We're going to use SMLE to fine-tune our outlier detection pipeline from earlier. To do so, we are going to read from the same S3 bucket that we saw in the last pipeline and start doing some interactive analysis. We will run adaptive thresholding on this data to see what it actually does. Incidentally, this is the exact same SPL that we ran in DSP. Except now, we are querying from the S3 sink that DSP is writing to. As you can see, adaptive thresholding adds a few columns to this input dataset.

One such column is the label column. We are interested in tracking the points where the label is
'true'. This would indicate the points that are detected as outliers. The interoperability between SPL 2 and Python allows us to leverage the rich visualization ecosystem in Python. We can easily put together a dashboard that overlays the original metric found by thresholding with default parameters.

As you can see, our pipeline does detect outliers, but it seems to have some noise. This noise seems to be coming from some seasonal data. The metric has seasonality that we haven't accounted for. A good example of seasonality, when you have say one day of the week that has less activity than other days. In that case, increased activity on that day can potentially be an outlier, even though it would be normal for other days. Similarly, on a day when there is usually more activity, incrementally more activity should not count as an outlier.

Ram Sriharsha (19:03)
For an algorithm to behave correctly in both scenarios, it has to be aware of seasonality. In streaming ML, we have a simple way to detect and correct for seasonality online. This algorithm, called online STL, is an algorithm for performing time series decomposition online. We can think of it as allowing us to decompose the original metric into its trend, seasonal, and residual parts. To correctly handle seasonality, therefore, we will use this decomposition and only monitor the residual for outliers. In the next cell, we will use this feature to modify the SPL to correctly handle seasonality in this specific metric.

As you can see, we have far fewer outliers now than before. Now, we are going to apply the same parameters to the entire dataset. This will allow us to train a model per host and error type on this dataset, without any further effort. Okay, this looks good. Now, let's make sure we achieve the expected accuracy in terms of picking relevant outliers, as well as the avoiding noise. For this, we'll again switch back to the Python mode to visualize the outliers. It looks like they're good to go.

Now, how do we productionize this? This is another area where smle really shines. Unlike other platforms where deploying machine learning pipelines requires re-engineering code and often rethinking algorithms at scale, in smle it is as simple as cutting and pasting the SPL into your DSP pipeline and reactivating the pipeline, and there you have it. We have our improved machine-learning pipeline tested and validated in SMLE, deployed at scale in DSP, without a single line of code re-engineered for operationalizing or scaling the pipeline. Thanks for having me today. I'll turn it back to Tim.

Tim Tully (20:47)
Thanks for the demos today, Ram. As my last topic today, you heard it from Doug in his talk. It's been a big year for Splunk and Observability. Last year, we jumped into observability, which at
its core is a real-time, high-cardinality, big-data problem. Perfect for Splunk to be incredible at.
Enterprise applications today fall somewhere on the distribution in this diagram. On the far left,
we have folks whose applications live in VMs in the data center. Typically monolithic apps, often
running in JVMs.

As they start to move all the way to the right side, where they're full-on microservice-oriented on
the orchestration platforms like Kubernetes, they typically wind up somewhere in the middle. My
observation is that it winds up being a normal distribution. In order to begin to monitor your
applications as you shift right, you need a new way to monitor your applications. This is where
Observability comes into play. The most sophisticated, real-time way to pull together logs,
metrics, and traces from your microservice-based apps into a single pane of glass.

You're going to hear more later from Sendur and others talk more about why Splunk is perfectly
suited to take you on this journey, but this problem we’re solving for is directly why we’re
focused on observability. We want to be there for you as you evolve your applications to
become more cloud-native, or as you shift right, or right-shift for the more technically inclined.
We seeded the tech for this software, with the acquisitions of SignalFx and Omnition.

We married metrics and traces with Splunk's logs expertise, to help you achieve better
outcomes through comprehensive observability across all of your data, which brings me to our
brand new Observability suite. As Doug mentioned, today we're very proud to announce the
launch of this suite, which combines our best-in-class metrics, tracing and logging capabilities,
into the most comprehensive, entirely integrated suite of products. All in one UI.

Tim Tully (22:30)

As you guessed, we leveraged architectural components based on what I've already talked
about today, stream processing, built in ML, and much more. Even more importantly, we ingest
all of your data. Unlike other approaches that sample or drop data on the floor, we're collecting
and analyzing all of your data, every single metric, transaction, log line, and dimension, every
single day. We call this our no-sample full-fidelity ingestion, and we're the only ones in the
industry that can do it cost-effectively and at scale.

Full fidelity is exactly what the name implies. It allows our customers to get a complete picture of
what's happening at all times. Because you have so much data, we're applying machine
learning analytics to your data in-flight to automatically detect patterns, catch all anomalies, and
surface recommendations to assist users in answering any question.

As part of our new approach to observability and APM, we're also launching two new products
within the Splunk Observability suite, real user monitoring or RUM, and a purpose-built log
observer for dev ops teams, both in beta.

Tim Tully (23:33)
We build RUM with the same innovations behind APM. We're the only solution that can provide real-time, end-to-end and full fidelity visibility into every single customer and user interaction within an application, so dev ops teams can get a full perspective of the user experience. Make sure you check out our demo of real user monitoring on our observability suite. It's an especially powerful and critical for dev ops teams, now that the world has shifted to digital channels.

Second, here you see our purpose build log observer for DevOps. Log observer puts the power of Splunk logging into the hands of SRS and developers, and rounds out our three pillar approach to observability.

It's cloud hosted, so you can get started in minutes, and offers out of the box integrations for fast time to value. We've tuned this experience for SRS and developers that uses logs for debugging, error reporting and exception handling.

We also leverage our streaming architecture to provide a live tail, which helps you keep an eye on things in production when you deploy code. And for those of you that need quick access to logs, but don't want to use SPL, we also provide a point and click search experience for rapid log exploration. Over the past year, our teams have been busy pulling together logs, metrics and traces into a brand new re imagined experience that allows you to seamlessly navigate across all the aspects of observability. Our suite delivers a beautiful, modern and unified user experience. Check out your screen for me to show you.

This new UX delivers a seamless workflow across entire problem detection to incident resolution workflow. In this particular example, we quickly see that an incorrect token is being used for a new version of code that was pushed to canaries.

As the checkout process is happening, it hits the canaries and fails, but retries and gets to a node that hasn't been swapped to the new code, and succeeds. We can see all of this happening in real time, starting with real-time and accurate problem detection in Splunk APM, verifying which end users have seen the problem and experiencing high latencies with Splunk RUM.

Then, looking at the end to end trace, but full fidelity and APM and seeing that there's a problem with the payment service, then using Splunk infrastructure monitoring to verify that there are no problems with the underlying Kubernetes cluster.
Finally, using Splunk log observer to see a live tail of events and do root cause analysis. Everything that you just saw and all of our capabilities around mealtime streaming, no sample full fidelity ingestion, and built-in ML are central to not just our post jobs observability, but also application performance monitoring.

They power highly detailed, but intuitive services maps, and enable unlimited cardinality exploration of our data, helping developers better understand and optimize their applications.

Modern applications in Cloud environments need a fundamentally new approach to monitoring, and that's Splunk observability.

Tim Tully (26:11)
If you like what you're hearing, take a look at what I just announced in my Splunk blog post. Splunk is announcing its acquisition of Plumber and its intent to acquire Rigor. Two companies with deep expertise in intellectual property that extends Splunk's product and market leadership in observability.

On Splunk's March toward building the most complete and robust observability solution possible, we had a strong desire to offer you DEM, or digital experience monitoring functionality.

To that end. We're excited to announce our acquisition of Plumbr, helping to bring the power of Java profiling and BCI, or bytecode instrumentation, which is critical for you to have deep visibility into traditional Java and .net enterprise applications.

The Plumbr team brings particularly rare expertise in the world of APM and BCI, and their team is going to help us accelerate Splunk's customers BCI, and existing RUM and database monitoring capabilities.

We plan to donate and incorporate all of the applicable technology to the open telemetry initiative, and I'll talk in a few minutes about why this is so important to Splunk and our customers like you.

We're also excited to announce our intent to acquire Rigor, which will bring advanced synthetic monitoring or synthetics to our observability suite.

Synthetics enables you to simulate requesting applications and services to verify what's working and what's not before your real users are impacted. The capability ultimately helps you deliver
better digital experiences.

The Rigor team is made up of leaders in digital experience monitoring with products that combine the power of synthetic monitoring with an intelligent optimization engine to help you find, fix and prevent the speed and web performance issues impacting user experience.

Together, the acquisitions of Plumbr and Rigor help accelerate our vision of delivering a comprehensive observability suite with best-in-class APM and DevOps for all applications, existing and new. You should have no doubt that Splunk has redefined application monitoring and troubleshooting with our next generation enterprise grade approach to APM. We offer the industry’s only real time, full fidelity and ML driven APM solution that helps it and dev ops teams quickly, confidently and reliably modernize existing applications and build new Cloud native applications.

We look forward to welcoming the Plumbr and Rigor teams to the Splunk family to help us continue the position to position Splunk as a top leader in the observability and APM markets.

Tim Tully (28:32)
The last thing I want to talk about, as promised, is open telemetry. Our approach to observability and data ingestion for observability starts with our commitment to this open source project.

Open telemetry enables our customers to use a standards based approach to collect all metrics, traces and log data from their applications and environments with a single, open and flexible set of APIs in instrumentation.

Open telemetry is the most active CNCF project, after only Kubernetes. As Co-founders of the project, we’re a hundred percent behind it. In fact, we’ve become so passionate about otel, that we’re now the number one contributor to OTel overall, ahead of Google and Microsoft to name a few. And as of this month, as part of our team, we also have the top contributor from across all projects in the CNCF. Congrats to the Bogdan.

Open telemetry is all about democratizing access to data. As you move forward and embrace your inner developer, it’s no longer about how you collect the data, it’s about what you do with it. You need to be able to innovate quickly and respond to market conditions.

Splunk believes heavily in open source software and open standards, and you’re going to see us continue to fight for them and push the rest of the industry to do the same.
You know Splunk's history with helping security and IT teams to remove the barriers between data and action. Now you'll use our innovations to go even further using Splunk as a single platform.

As you've seen today, our innovations addressing and combining extreme processing, machine learning and observability equip you to tackle your organization's biggest challenges in a way no other platform can enable.

Look, this year has been like no other year, and what the data industry has done is stunning. As Doug talked about, no other industry ever has scaled to meet a challenge like ours has in 2020. We know the future runs on data, and we at Splunk are focused on arming you to deliver data to everything, to reach that future.

This is the year the Cloud became business critical. This is the year IT professionals became essential workers. This is the year the digital business became the business. Thank you for being the data heroes the world needs.

###