Fake Data for Real Apps

SimData as a new Simulated Data Generator

David J. Cavuto, CISSP | Principal Product Manager, Data Ecosystem
David Poncelow | Senior Software Engineer, Data Ecosystem

Date | Washington, DC
During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.
Cavuto Bio

► Bell Labs
  • Principal Engineer - Lucent VPN Firewall
► AT&T
  • Network security and analytics
► Narus
  • Product Manager – Narus Cyber Analytics
► Splunk
  • Sales Engineer, Security SME
  • Principal Product Manager – Splunk App for Stream
  • Principal Product Manager – Data Ecosystem Area
► David J. Cavuto
dcavuto@splunk.com
History of Data Simulation at Splunk
Help user envision how Splunk might address their specific needs
- Simulation must create scenarios that look familiar
- Simulation must create impairments that mimic users impairments

Help software testers QA apps and add-ons
- Performance
- Functional test

Help software developers build and test apps and add-ons
- Often, no access to original equipment
EventGen History
What EventGen Did / Does

- Sample-based
  - Replays “sample” events in original log file format
- Written in Python
- Written as a Modular Input
- Replace tokens based on rules
EventGen Limitations
Challenges users have with EventGen

- Realistic user cohorts was challenging
- Didn’t really scale (for testing)
- Multiple scenarios required hacks
- Data correlation required specific scenarios artificially inserted into data
Next Steps in Data Simulation: SimData

SimData Concepts and Execution
SimData Design Goals

- Create Entity/Event based simulations
- Allow for multiple data outputs
- Maintain internal state of entities
- Scale to support load testing on multiple indexers
- Allow external control of running simulations
- Provide backwards-compatibility for EventGen configurations
Persona Targeted By SimData

► Field Engineer
  • A Splunk Field Engineer will be using SimData to demonstrate Splunk Core and Splunk App

► Simulation Author
  • A Simulation Author will create Simulations (primarily for Field Engineers)

► Test/QA
  • Test/QA engineer will use SimData to exercise elements of their App for testing purposes
  • Test/QA engineer will use SimData to generate large amounts of load to test capacity and performance of distributed Splunk systems

► App Developer
  • A Splunk App Developer will use SimData to help populate dashboards in their App when a live data source is unavailable
Domain-Specific Language designed specifically for creating simulation elements

Development experience not required

Intended that you specify the elements, connections between them, and the messages they pass

Engine takes care of the rest
SimData Concepts
Basic concepts of a Data Simulation

► Scene – Specifies the connections and messages between entities
► Simulation – Specifies the entities and their behaviors
► Entity – The functional element of any simulation
  • Init
  • Every
  • On
► Message – How Entities communicate to each other
  • Send
  • Respond
► Log – How Entities send data to Splunk
  • Emit
Simulation Structure Diagram
How Entities Connect and Communicate

Entity 1 (Web Client) -> Entity 2 (Web Server)

Web Request
Web Response

Periodic Action (Every)

Entity 2 (Web Server) -> Log

Reactive Action (On)

Send Event to Splunk (Emit)

Splunk Indexer

Splunk
DSL Syntax (v0.8)
Yes, it's an eye chart

entity <entity-class-name> {
    init { assignment; [assignment]; };
    every (time) { statement };
    on <message-class-name> { statement };
    send <message-class-name> { parameters } [in (time)];
    respond { parameters } [in (time)];
    emit <event-class-name> { format-parameters };
    /
    /* this defined a function for invoking as
    self.method() */
    /* and can use 'send', 'emit'. In the right scope 'respond' too. */
    method <method-class-name>([parameters]) { statement };
};
message <class-name> { message-parameters };
event <class-name> { statement };
debug( format-parameters );

scene {
    connect {
        fromEntity: <entity-class-name>;
        messageType: <message-class-name>;
        toEntity: <entity-class-name>;
    } [, { conditional:
    fromEntity: <entity-class-name>;
    messageType: <message-class-name>;
    toEntity: <entity-class-name>;
    } ]
} if (<evaluation> <comparison> <evaluation>)
then { statement } [ else { statement } ]
assignments:
    <variable> = evaluation;

time: numeric-evaluation (milliseconds | seconds | minutes | hours)
evaluation: expression that evaluates to a type object
numeric-evaluation
expression that evaluates to a numeric scalar type
language builtins: (more TBD)
    random(<range>)
    now
    false
    true
    self
    if/else
    milliseconds
    seconds
    minutes
    hours
range: (<range>)
numeric-evaluation, numeric-evaluation
time, time
External Control Of Scenarios

- Change a running simulation to show how Splunk / App responds
- Create Sunny / Cloudy / Rainy-day scenarios
- Change parameters of systems
- Create impairments
  - Network outages or congestion
  - Enable / disable hosts
  - Create attack-based disruption
External Control Dashboard Example

Allows Real-time Control of Running Simulation

Welcome to the demo controller!

Basic Controls

Alter Database Status.

- Enable All Databases
- Disable All Databases
- /user/scene/Database/Sb
- /user/scene/Database/Sa

Alter WebServer Status.

Set maxPendingRequests for all WebServers:
- /user/scene/WebServer/Sb
  - Value: 0
- /user/scene/WebServer/Sa
  - Value: 0
Examples and Demonstration

Some syntax and a live simulation
message RequestPage {
    required: [requester, page, ip];
}

message PageResponse {
    required: page;
}

event GotPage {
    required: [name, ip];
    template: "{{_time}} {{name}} here- got my page. I'm at {{{ip}}}";
}

event WebRequest {
    required: [page, server_ip, client_ip];
    template: "{{_time}} {{{server_ip}}} Received request for ' {{{page}}}' from {{{client_ip}}}";
}

entity User {
    init {
        name = "joe";
        ip = "127.0.0.1";
    }
    on PageResponse {
        emit GotPage {
            name: name;
        }
    }
    every (3s) {
        send RequestPage {
            requester: self;
            page: "home";
            ip: ip;
        }
    }
}

entity Webserver {
    init {
        hostname = "webserver01";
        ip = "5.5.5.5";
    }
    on RequestPage {
        emit WebRequest {
            page: message.page;
            server_ip: ip;
            client_ip: message.ip;
        }
        send message.requester PageResponse {
            page: message.page;
        }
    }
}
Simulation Example 1 – Scene File
Entirely in JSON format

```json
{
    "update_interval": 1,
    "time_unit": "Seconds",
    "entities": [
        {
            "entity_name": "User",
            "initial_state": {
                "name": "Lucy",
                "ip": "185.19.32.1"
            },
            "count": 1
        },
        {
            "entity_name": "User",
            "initial_state": {
                "name": "Edward",
                "ip": "43.19.22.5"
            },
            "count": 1
        },
        {
            "entity_name": "User",
            "initial_state": {
                "name": "Susan",
                "ip": "39.182.16.4"
            },
            "count": 1
        },
        {
            "entity_name": "User",
            "initial_state": {
                "name": "Peter",
                "ip": "212.52.1.198"
            },
            "count": 1
        },
        {
            "entity_name": "User",
            "initial_state": {
                "name": "Mr. Tumnus"
            },
            "count": 1
        },
        {
            "entity_name": "Webserver",
            "initial_state": {
                "hostname": "webserver01"
            },
            "count": 1
        }
    ],
    "entity_wirings": [
        {
            "from": "User",
            "to": "Webserver",
            "message_type": "RequestPage",
            "wiring_type": "any",
            "filter": null
        }
    ],
    "default_transport": "Text"
}
```
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-09-07T15:35:01.6</td>
<td>userID=837 got response code=200 sessionID=c808e135-ea9c-4754-935b-148e4d251967</td>
</tr>
<tr>
<td>2017-09-07T15:35:01.2</td>
<td>Database Replied with status=ok connectionID=358d3b60-1792-4788-a95b-72ae619ad7f7</td>
</tr>
<tr>
<td>2017-09-07T15:35:00.8</td>
<td>Query Successful. connectionID=358d3b60-1792-4788-a95b-72ae619ad7f7 CPU_percent=0 disk_percent=0 duration=1</td>
</tr>
</tbody>
</table>
Sample Dashboards
Running Demo 1
Simple Splunk events
Running Demo 2
dashboards + investigation + scenario changes
Q&A?

Questions from the Audience
Thank You

Don't forget to rate this session in the .conf2017 mobile app