Navigating the Challenges of Building a Large Scale Distributed Application

A Case Study

Sridhar Komandur, PhD
Student Information Systems
What is MyPlan?

MyPlan is an academic planning tool for University of Washington Students.
MyPlan Usage

As of Jan 5, 2017 ...

Total Students who have a plan :  88,324
  UW Students               :  73,661  83%
  Non-UW Students          :  14,663  17%

  Freshmen adoption        :  5,903  81% (as of Sep 2016)
MyPlan Uptake

MyPlan Adoption Over Last 5 Years

Students

Year

MyPlan Uptake

Courses Planned in MyPlan Over Last 5 Years

![Bar chart showing the number of planned courses over the years 2011 to 2016. The number of courses planned increases significantly from 2014 onwards, peaking in 2016.]
# MyPlan: Ecosystem

<table>
<thead>
<tr>
<th>USERS</th>
<th>MYPLAN</th>
<th>PARTNERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advisors</td>
<td>Apps</td>
<td>First Year Programs (FYPs)</td>
</tr>
<tr>
<td>Faculty/Staff</td>
<td>Config</td>
<td>Registrar’s Offices</td>
</tr>
<tr>
<td>Students</td>
<td>Platform</td>
<td>Undergraduate Advising</td>
</tr>
<tr>
<td></td>
<td>Hosts</td>
<td>UW-IT Cross-functional Teams</td>
</tr>
<tr>
<td></td>
<td></td>
<td>… and more</td>
</tr>
</tbody>
</table>
MyPlan Tagline

We don't create a lot of data you see.

We make a lot of the data you see consumable!
# MyPlan: Data Aggregation

## Academic Year 2016-2017

Create your academic plan by finding and adding courses to prepare for registration, or choose "Build Your Schedule" scheduled course sections in the Quarter view in a calendar grid.

### Autumn 2016

<table>
<thead>
<tr>
<th>Course</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSE 332</td>
<td>4</td>
</tr>
<tr>
<td>BIOL 180</td>
<td>5</td>
</tr>
<tr>
<td>Backup</td>
<td></td>
</tr>
</tbody>
</table>

### Winter 2017

<table>
<thead>
<tr>
<th>Course</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGL 291</td>
<td>3</td>
</tr>
<tr>
<td>STAT 220</td>
<td>3</td>
</tr>
<tr>
<td>MATH 205</td>
<td>3</td>
</tr>
<tr>
<td>ECE 410</td>
<td>1-5</td>
</tr>
<tr>
<td>G H 403</td>
<td>1</td>
</tr>
<tr>
<td>No courses planned</td>
<td></td>
</tr>
</tbody>
</table>

### Spring 2017

<table>
<thead>
<tr>
<th>Course</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHEM 600</td>
<td>1-25</td>
</tr>
</tbody>
</table>

### Summer 2017

<table>
<thead>
<tr>
<th>Course</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backup</td>
<td></td>
</tr>
</tbody>
</table>

## Resources

- **Sys Academic Records**
- **SWB Enrollment Resource**
- **PWS**
- **SWB Term Resource**
- **Curriculum Management**
MyPlan: Data Aggregation
MyPlan: Data Sourcing Diagram
Scaling

Once upon a time (~2011) …

MyPlan was a *Monolith*: Put all its functionality into a single process
Scaling
The Monolith 2.0

The Monolith 1.0
MyPlan Scaling

The Monolith

Microservices
Put each element of functionality into a separate service
MyPlan Scaling
MyPlan Scaling: Data Sourcing Architecture

**APPLICATION TIER**
- KRAD APP CLUSTER
- REACT APPS CLUSTER

**CORE DATA SVCS TIER**
- SOLR MASTER
  - CORES
  - SECTION
  - STATUS
  - COURSE
  - PROGRAM
- POLLER (ETL)
- Amazon SQS

**REPLICATION**
MyPlan Scaling: DATA SOURCING ARCHITECTURE
MyPlan Scaling

- Currently we have 34 virtual machines *(and growing!)*

Challenges

How do we track and manage what’s on those VMs?

- Apps
- Platform
- Configuration (includes app config, certs etc)
- Provisioning
- Orchestration
MyPlan Scaling

Automation using Ansible!

- Deployment
- Provisioning
- Orchestration
Monitoring & Troubleshooting

Short term and long term concerns

- Active monitoring: PageWatch -> UW Connect tickets
- During Registration (predictable peaks)
- Short term/immediate troubleshooting
- Long term issue trend analysis (creep up over time)
- Which Apps are deployed and where? (features, testing, troubleshooting)
# Monitoring & Troubleshooting

## MyPlan

### Section Status Poller

<table>
<thead>
<tr>
<th>App</th>
<th>Host</th>
<th>Build Number</th>
<th>Build Time</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>audit</td>
<td>stratoseval01</td>
<td>34</td>
<td>2017-02-24T10:56:48.318-08:00</td>
<td>1.4</td>
</tr>
<tr>
<td>audit</td>
<td>stratoseval02</td>
<td>34</td>
<td>2017-02-24T10:56:48.318-08:00</td>
<td>1.4</td>
</tr>
<tr>
<td>console</td>
<td>stratoseval01</td>
<td>141</td>
<td>2017-03-17T08:08:823-07:00</td>
<td>2.2</td>
</tr>
<tr>
<td>console</td>
<td>stratoseval02</td>
<td>141</td>
<td>2017-03-17T08:08:823-07:00</td>
<td>2.2</td>
</tr>
<tr>
<td>core</td>
<td>troposeval01</td>
<td>443</td>
<td>2017-03-15T14:41:54.610-07:00</td>
<td>1.8-SNAPSHOT</td>
</tr>
<tr>
<td>core</td>
<td>troposeval02</td>
<td>443</td>
<td>2017-03-15T14:41:54.610-07:00</td>
<td>1.8-SNAPSHOT</td>
</tr>
<tr>
<td>course</td>
<td>stratoseval01</td>
<td>23</td>
<td>2017-03-21T13:18:28.941-07:00</td>
<td>2.0</td>
</tr>
<tr>
<td>course</td>
<td>stratoseval02</td>
<td>23</td>
<td>2017-03-21T13:18:28.941-07:00</td>
<td>2.0</td>
</tr>
<tr>
<td>program</td>
<td>stratoseval01</td>
<td>1048</td>
<td>2017-03-17T11:27:29.136-07:00</td>
<td>3.0-SNAPSHOT</td>
</tr>
<tr>
<td>program</td>
<td>stratoseval02</td>
<td>1048</td>
<td>2017-03-17T11:27:29.136-07:00</td>
<td>3.0-SNAPSHOT</td>
</tr>
<tr>
<td>student</td>
<td>uwkisew01</td>
<td>91</td>
<td>2017-03-21T20:27:20.27 UTC</td>
<td>3.6</td>
</tr>
<tr>
<td>student</td>
<td>uwkisew02</td>
<td>91</td>
<td>2017-03-21T20:27:20.27 UTC</td>
<td>3.6</td>
</tr>
<tr>
<td>student</td>
<td>troposeval01</td>
<td>91</td>
<td>2017-03-21T20:27:20.27 UTC</td>
<td>3.6</td>
</tr>
<tr>
<td>student</td>
<td>troposeval02</td>
<td>91</td>
<td>2017-03-21T20:27:20.27 UTC</td>
<td>3.6</td>
</tr>
</tbody>
</table>
Operational Data Aggregation

We use Splunk for operational data aggregation

- Collects logs from all our hosts (troubleshooting, trend analysis)
  - Potential for tracing activity end-to-end through the entire system (e.g., instrumenting uuid)
- Instrumented Apps to collect user experience data
  - Potential for User Behavior Analytics
- Potential for Splunk alerts to trigger actions
Future Work

● Change management
● User Behavior Analytics
● Predictive Analytics/Demand Forecasting
Reference

- Microservices, Martin Fowler
  (https://martinfowler.com/articles/microservices.html)