Real-World Education Data Management Performance Testing

Mike Reynolds, Co-Founder and President
Benjamin Silberglitt, Ph.D, Co-Founder and CSO
Why Performance Test?

• Demand for data continues to increase
• Data bottlenecks are rampant in education data integration projects – even with standards
• Is Unity THE future-proof standard?
The Demand for Data

- Real-time period attendance
- Grades pass-back
- Continued growth in number of software systems
- Data insights across systems (e.g. Medical & K-12 data)
- Example of what’s coming: AI and just-in-time feedback from online learning systems
Data insights across systems
Why the SIF specification?

Infrastructure
• Defines the “how”
• Transport and messaging
• Modern, RESTful API
• Key to ease of adoption and efficient data flow

Data Model
• Defines the “what”
• XML or JSON
• Set of data objects
• Aligned vocabulary with CEDS
• *SEPARATE* from infrastructure
Takeaways:

• ALL your data can be transported over a common infrastructure

• Many data models, one infrastructure

• Question: Can it handle the load?
SIF Infrastructure: Key Features

- Parallel Processing
- Endless Queue Enabling
- Data Bundling
- Scalability
- Data Model Agnostic
- Privacy Over-the-Wire
The Test

Three scenarios:
• SIF Infrastructure Specification 2.x
• Hybrid (SIF 2 Mock SIS + SIF Infrastructure 3.4 Data Digester)
• SIF Infrastructure Specification 3.4
Keeping it honest

- No auto-scaling
- Attendance was the use case, but same performance would be expected with any data
- Same level of server capacity and run time (1 hour) used in all three scenarios
## Results

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Total Records</th>
<th>Records /Second</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SIF 2 Infrastructure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock SIS</td>
<td>1,063,119</td>
<td>295</td>
</tr>
<tr>
<td>Digester</td>
<td>19,697</td>
<td>5</td>
</tr>
<tr>
<td><strong>Hybrid</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock SIS</td>
<td>1,063,119</td>
<td>295</td>
</tr>
<tr>
<td>Digester</td>
<td>1,311,408</td>
<td>364</td>
</tr>
<tr>
<td><strong>SIF Infrastructure 3.4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mock SIS</td>
<td>215,622,600</td>
<td>59,895</td>
</tr>
<tr>
<td>Digester</td>
<td>375,188,400</td>
<td>104,219</td>
</tr>
</tbody>
</table>
The Difference

- Parallel record processing
- Refined queues
- Batch events
What it means

- SIF Infrastructure Specification 3.4 takes better advantage of the hardware it runs on and can scale up or down to meet the needs of the moment.

- Efficiency gains mean users can do more work with less resources, saving money on implementations while providing a sizeable performance improvement.
With the hybrid scenario, any existing implementation can immediately start to take advantage of this powerful new architecture without having to embark on a huge multi-year implementation or rip up existing tooling and processes that are vital to operations.
SIF 3.4

• Built for the future
• Less expensive to operate
• Safe, proven migration path
Questions?

Full press release available at: https://home.a4l.org/news/