

Evolveum

midScale

First Mentoring Session

November 2020

Agenda



- Objectives
- Approach
- Progress
- Discussion

Project Objectives

- Increase midPoint scalability by an order of magnitude

Routinely manage: ~1 million identities

Possible deployments: ~10 million identities

- Identify scalability and performance obstacles

Performance testing environment

- Overcome scaling obstacles

Data store: data model, indexing, partitioning, etc.

Clustering / multithreading

Stability: thread safety, tooling

Visibility & usability: UX, GUI performance, diagnostics

Approach: Data Store

- Data are foundation of the system
- System cannot be improved if data structures are wrong
 - MidPoint high-level data model is (mostly) right.
 - MidPoint low-level database schema *is* wrong.
- Database layer is replaceable
 - Early design decision, going back to approx. 2011
 - Allows re-implementation of the database layers
- Focus: PostgreSQL

Approach: Horizontal Scalability

- Vertical scalability and single-node performance is not enough
- We need: multi-threading, clustering, multi-node tasks
- MidPoint supports clustering, but improvements are needed

Better and more reliable per-node multi-threading

Smarter multi-node tasks (load distribution)

Cluster auto-scaling

Misc multi-node improvements (e.g. thresholds)

Approach: Stability, Visibility & Usability

- **Stability in large deployments**

We are observing instability under high load, thread safety suspected

Focus: Prism – midPoint data representation layer

Building on Axiom, result of midPrivacy initiative

- **Visibility and usability**

Diagnostics: needed for both development and deployment

GUI performance

UX for administrators to handle massive data sets

Approach: Evaluation & Testing

- Continuous activity: from project start to finish
- Performance testing environment
 - Automation
 - Docker, Jenkins
- Diverse set of diagnostic tools
 - Stock tools (e.g. profiler)
 - Diagnostics integrated into midPoint

Approach: Methodology

- **Transparency**

 - Open source

 - Everything is public: code, documentation, meeting notes, ...

- **Iterative and Incremental**

 - Proceed in iterations (milestones)

 - Feed experience back to the project

 - Solution architecture evolves, updated for each milestone

Project Progress

- On track
- Works ahead of plan
- Results:
 - Design meeting notes
 - Baseline performance measurements (partial)
 - Progress on testing environment
- Added more people to the project, funded by Evolveum

Project Resources

- Project home page

<https://docs.evolveum.com/midpoint/midscale/>

- Solution Architecture (work in progress)

<https://docs.evolveum.com/midpoint/midscale/architecture/>

- MidPoint source code

<https://github.com/Evolveum/midpoint>

- Evolveum Blog

<https://evolveum.com/blog/>

Discussion

Questions & Answers

Thank you for your time

See other talks at <https://docs.evolveum.com/talks>

Also **follow us** on our social media for further information!



/Evolveum



/Evolveum



/Evolveum



@Evolveum



/Evolveum

Evolveum