

LinkedIn Infrastructure and Operations Automation at WebScale

PROFILE



Industry

Online Professional Network

Corporate Headquarters

Mountain View, CA

Employees

More than 5000

Annual Revenue

Approximately \$2B

Website

<http://www.linkedin.com>

With CFEngine LinkedIn automates IT infrastructure and operations, reducing costs and boosting efficiency

“LinkedIn is the largest professional social network in the world and is currently the 10th largest website in the US by traffic. Our operations team will make around 5-10 production changes per day. CFEngine provides the automation framework and gives us the ability to continue to scale operationally.”

Mike Svoboda, Systems and Automation Engineer, LinkedIn

THE NUMBERS

- Automated 40K servers
- Team size – 6
- Build Datacenter in minutes
- Dozens of reliable changes across thousands of machines

IN BRIEF

Challenge

LinkedIn needed to accommodate rapid growth in business that translated to extreme infrastructure scaling, and agility in a dynamic environment.

Solution

A low touch, programmable approach to infrastructure automation and operations.

Results Increased efficiency with minimal risk while accommodating rapid growth in infrastructure. Increased visibility into infrastructure and operations Datacenter transformation leveraging CFEngine programmable infrastructure constructs

LinkedIn, one of the world's leading online professional networks, has enjoyed explosive growth, scaling from 300 machines in 2010 to over 40,000 servers growing very rapidly every month. As a result, the company turned to CFEngine to scale and automate its infrastructure throughout its entire lifecycle - from build and deploy to continuous operations

The Challenge

LinkedIn operates one of the world's premier online professional networks, allowing members to manage and share their professional identity online, find jobs, connect with other professionals, and locate business opportunities. LinkedIn has grown significantly since its inception and currently serves over 250 million users in over 200 countries. In March 2014, Quantcast reported that LinkedIn gets 237 million monthly unique U.S. visitors. Supporting this exponential business growth presented the IT infrastructure operations team at LinkedIn with a unique set of challenges and requirements:

Accommodating rapid growth: With a group of 40,000 servers growing over 5% per month, LinkedIn adds thousands of new machines every year. To keep pace, these machines needed to be configured and set up for production in 15 minutes or less, with minimal headcount addition.

Increasing developer efficiency: It took several weeks for new resources to be setup on the infrastructure using traditional methods, a process that was repeated many times every week. LinkedIn needed to automate this process so users could be removed or added to thousands of machines within minutes, freeing up IT operations to focus on higher value work.

Avoiding fear of change and increasing agility: With the site and traffic growing at a very fast clip, LinkedIn needed an automation solution that could introduce changes in a controlled fashion to minimize the risk of breaking the production environment.

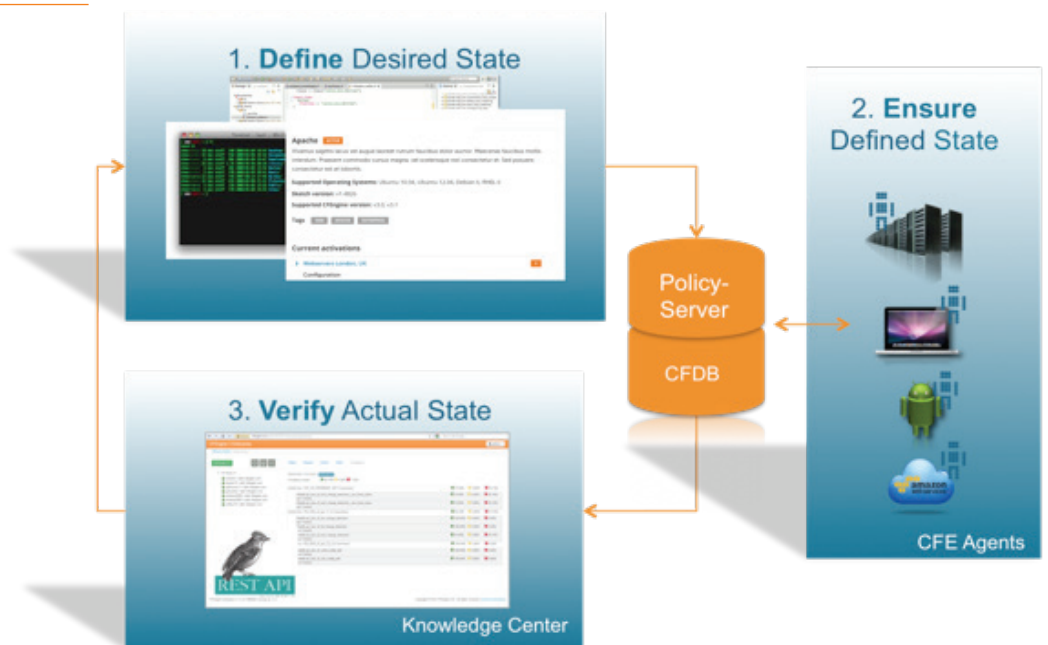
Establishing a culture of trust: LinkedIn needed to offer elevated access to production machines (sudo) to engineers while at the same time introducing self-healing capabilities to ensure that the systems could recover to their earlier state if an engineer were to cause production disruption.

Unlike other platforms based on Ruby, Python or Perl, CFEngine has provided LinkedIn with a lightweight solution based on 'C' with minimal dependencies that can easily scale to thousands of machines

Increasing visibility at scale: LinkedIn needed to be able to quickly diagnose and address issues such as hardware failure more quickly and at a greater scale than traditional monitoring and reporting solutions could offer.

The Solution

In the face of these mounting challenges, LinkedIn chose CFEngine to automate their infrastructure configuration and lifecycle management by enforcing intended system state for compliance. Unlike other platforms based on Ruby, Python or Perl, CFEngine has provided LinkedIn with a lightweight solution based on 'C' with minimal dependencies that can easily scale to thousands of machines. CFEngine controls virtually everything in production except for the application deployment, including setting up new servers from bare metal, all OS configurations, software updates, and Java-based lifecycle management.



The Results

Rapid and Dynamic Server Management

CFEngine provides LinkedIn with a fully-automated provisioning process that can put a new machine into production within minutes, so they can easily add hundreds of new machines a month into production. This includes bare-metal provisioning, operating system configuration, account administration (assigning ssh access and sudo account elevation to machines), auditing against desired state, hardware failure detection, and system monitoring. CFEngine automation means that LinkedIn operations staff no longer needs to log into individual machines to maintain them – they instead make changes into CFEngine as a policy and the changes are propagated to all the relevant servers.

“With CFEngine, I can define a new Software Defined Datacenter and offer IAAS and PAAS to my customers within 10 minutes.”

Mike Svoboda, Lead Automation Engineer, LinkedIn

Increased Engineer Efficiency

Prior to CFEngine, LinkedIn needed to ssh to every machine to add accounts and it would 2-3 weeks to get new hires set up in all of the infrastructure. Now, using CFEngine, they have the ability to install and remove hundreds of users from thousands of machines in minutes. They have also taken advantage of this capability to minimize other repetitive tasks, so when the operations team solves an issue once and commits it to CFEngine, it is automatically replicated on other machines. As a result, wasted time is minimized and freed up to pursue higher value work.

Minimized risk with insightful phased deployment rollouts

LinkedIn can now make large-scale changes across thousands of machines in a controlled manner that minimizes the risk of breaking the production environment. To implement this, they use CFEngine to assign each policy to a range class of specific machines affected by the policy. When the operations staff commits a change to production (such as sudo rules, account access, or software installs), they assign it to its related range classes at 0%, - no systems initially affected. Next they increase the range class to contain 10% of machines and CFEngine’s monitoring capabilities immediately notify them if anything breaks as a result of the change, in which case they can immediately roll it back to its previous state. If not, they gradually increase to 100% while continually monitoring for issues. Phased rollout is one of the most important features CFEngine has enabled LinkedIn to implement because it gives them the confidence to automate operations change management aggressively with the knowledge that it will not break the production environment. As a result, it is now common for LinkedIn to push up to 15 CFEngine-related changes per day.

A culture of trust that leads to agility

CFEngine enables LinkedIn to minimize the possibility of engineers inadvertently causing “configuration drift” and disrupting the production environment. Rather than adding significant delay by carefully evaluating every elevated access request from engineers before granting it, LinkedIn now grants root access to engineers by leveraging the powerful flexibility of CFEngine classes to determine who gets elevated access and where. By doing this, an engineer can’t disrupt a machine because in the case of any issues, CFEngine will immediately restore the system to its desired system state using its policy engine. Thanks to CFEngine, LinkedIn believes that their account privilege escalation infrastructure is probably one of the most advanced and flexible solutions in the industry.

Granular insight of actual states in seconds regardless of scale

LinkedIn uses CFEngine to publish data about the status of any machine to a custom-built monitoring solution which can answer environment questions almost immediately. CFEngine is used as a closed-loop system enabling LinkedIn to make configuration changes while visualizing the impact of those changes immediately to ensure they have taken place as intended.

The fact that LinkedIn can manage the entire infrastructure lifecycle of their 40,000 servers, growing by about 800 new servers every month with only a headcount of 6 in their infrastructure operations group is a testimony to the power of CFEngine and programmable automation at WebScale.

The Bottom Line

LinkedIn Challenges	LinkedIn Results with CFEngine
Accommodate rapid growth Drastically reduce machine provisioning times (from days)	Fully Automated Infrastructure Provision and bring online in minutes reliably, consistently
Increase Developer Efficiency Reduce setup time for new accounts on all infrastructure (2 to 3 weeks)	Automated User-Acct. Mgmt. Add remove accounts from thousands of machines in minutes
Increase Change Frequency Operations needed to make multiple changes to production, daily	Phased Rollout with Reporting Make dozens of changes across thousands of machines reliably
Minimize Repetitive Tasks Operations performing repetitive tasks such as configuration changes	Make Change Only Once Operations commits changes to CFEngine policy, propagated across
Establish a Culture of Trust Grant elevated access to production machines to engineers without risk of environment compromise	Safely Grant Access to Team Elevated access grants, reporting, ability to remediate and restore to operational configuration
Increase Visibility At Scale Issues such as hardware failures need to be detectable in seconds	Scalable In-Memory Reporting Publish data about any machine to reporting system and provide answers in seconds
Inefficient Hardware Purchase Lack insights into failure rates, manufacturer reliability	Failure Detection and Repair Automatically detect and replace failed hardware, report and analyze manufacturer reliability
Datacenter Transformation Would have taken a long time to move entire infrastructure from co-location to own datacenter	Programmable Infrastructure Able to leverage programmable infrastructure to move to self-owned facility in 5 months



About CFEngine

CFEngine is the industry leader of IT automation at WebScale. CFEngine enables some of the largest IT organizations to provision resources and deploy new applications orders of magnitude faster, while ensuring high availability, security and compliance in highly complex environments. CFEngine has users in more than 100 countries, including many of the world's largest financial organizations, websites and Fortune 1000 companies.