Thank you for reaching out to the CPE Team to log your request. Please fill out the fields below. To ensure your request can be reviewed please ensure all fields are populated where you see a *. Thank you

Project details:

Name of Project:	DNF Counting
Submitted by:	Matthew Miller
Date:	Nov 2019
Dependency dates : (if known)	
Scope by Tech lead	Pierre- Yves Chibbon
Project timeline (approx high level est): see note below	1 month - 2 developers & 1 Smooge :)

Summary overview:

Request for a better DNF counting server-side work

There are some proposals for more complicated systems, but a quick thing we can do now to greatly improve what we have without a gigantic new infrastructure.

I don't need fancy graphs or anything; what I want is a database of digested (and therefore de-identified) count information. Ideally, this can be public and accessible from anywhere. The client side of this is already in place. And I can write something to process the results and create charts and reports. What I need is the "bridge" from the raw logs to aggregated weekly data.

Why?

Right now, we estimate installed Fedora systems by counting unique IP addresses which show up in our updates mirror statistics. We need better data than that.

Right now, we get IP-address-per-day counts in columns broken out by Fedora or EPEL release, architecture, and some other random stuff as has accumulated on an ad-hoc basis.This is problematic because we don't know how many systems are hidden by NAT or conversely overcounted due to rapid IP address changes. We can't see the difference between short lived CI instances and "real" installations. And the existing system can't answer questions like "what percentage of F30 is i686?" without asking for a new column.

How will this help?

This will help us better understand how Fedora's various offerings are used in the world, and give us better insight into the real-world lifecycle of our releases. In turn, that will help Fedora leadership make informed strategic decisions.

What happens to the old system?

It'd be nice to leave it running side by side so we can validate the new system and compare. It wouldn't need to get any further updates and could be eventually retired.

What platform does this project relate to: [Please insert ✓]*

✓	Fedora		CentOS

Is this idea... [Please insert ✓]*

	New	✓	Enhancement		Replacement

Is there a workaround in place? Y/N

If yes, please provide details:

What area does it relate to: [Please insert ✓]*

	Initiatives	✓	Infrastructure		Releng

Why is this important?	What is the benefit of doing this?	What happens if it doesn't happen?
What problem or opportunity are we addressing: This will help us better understand how Fedora's various offerings are used in the world, and give us better insight into the real-world lifecycle of our releases. In turn, that will help Fedora leadership make informed strategic decisions.	**Value to Fedora:** • Showcasing the number of downloads of fedora to chart its successes • Being able to merticise how fedora is used, and what aspects of the offering people are using • This ultimately helps the engineers who work on/in fedora to focus on what is generating the best value to the team	Very hard to find out how many people are using which version of Fedora easily & accurately.

Objectives/Goals *

Please insert as bullet points

How I want data in the new system

I want a database (sqlite, CSV, remote database access; whatever), with rows like this:

Datestamp	DNF countme	OS Id	OS Variant	OS Version	Architecture

DNF's countme value is defined here: https://bugzilla.redhat.com/show_bug.cgi?id=1672504#c15. It will be a value in the range 1 to 4; entries without a "countme" value or "countme" of zero should not appear in the log.

So, we might have entries like this:

2019-12-01	1	Fedora	Workstation	32	x86_64
2019-12-01	4	Fedora	Server	31	x86_64
2019-12-01	3	EPEL		8	aarch64
2019-12-01	3	Fedora	Cloud	32	aarch64
2019-12-01	3	Fedora	KDE	32	x86_64

Because having a single row per system would result in millions of rows every day, and because the DNF feature triggers once per week, I suggest we instead present this as **weekly aggregated values**, like this:

Week #	DNF countme	OS Id	OS Variant	OS Version	Architecture	Sum

where identical rows are collapsed into a single row counting the number of times that combination occurred that week in the "Sum" column.

Where?

I think this new data could be exposed to the public; the aggregation and relatively few data points means that this isn't sensitive. Note that there are no IP addresses or even timestamps in the proposed format. It could be dumped to alt.fedoraproject.org or some other server as sqlite or csv files. Or it could be in a "live" database open to the public — or at least to an OpenShift instance where I could run my processing and report generation.

How often?

Because of the weekly aggregation, having this data updated once per week is perfectly fine.

What does success look like to you? *

A database of digested (and therefore de-identified) count information. Ideally, this can be public and accessible from anywhere.

Note: You do not need to fill out fields below. Our PO will work with you and the assigned Tech Lead to scope these further

Please submit this request to cpe-requests@redhat.com & cc amoloney@redhat.com

Thank you, we will be in contact soon.

.....

Requirements: (Prioritized epics + deliverables)

Requirements

Dependencies (users, other teams & app's affected) *(If known)*

Internal	External

Risk *(If known)*

Risk title	Type of risk	Risk description	Level of risk	Actions to mitigate risk

Considerations:

[Pierre]The dev work is likely pretty light, a couple of people maybe be enough, the more important part will be the understand and knowledge of the current system and all the pitfalls it ran into in the past.
In other words, I think smooge needs to be involved in this one, at least at a consultant level

Skill Set/Resources required to deliver

People	Skillset	Length of time	Potential Team Member
2	Developers	1 month	
1	Sysadmin	1 month	Stephen Smoogen

Project timeline: 1 month

Any other information:

Any open questions, unknown's, other insights you would like to flag, add them here: