Season III, Challenge II

Project : Whatever clever guys are doing on Earth with cloud datacenters, do it on Mars and earn dough

Focus Area : The Cloudy Thing

Created by : Lubomir Zvolensky (lubomir@zvolensky.sk)

Contents

1. Executive summary

Human mankind, whatever is left of it, found an investor with deep pockets full of gold stripes. Hooray. As he doesn't understand technology at all, we need to prepare a project that will convince him to spend a buck or two to potentially replicate there what established guys are doing down here with clouds and datacenters. The idea is to have hybrid or public cloud infrastructure available on Mars hosting applications, providing benefits expected from such solution.

When checking out online calculators for several cloud offerings, one might easily get impression they are so f...reaking expen$$$$ive to be hard to believe they actually have customers. Got numerous calls from Columbia ganja guys, they noticed there is something more milky than what they do now, additionally requiring no private army "to protect the goods" [we've seen customers charged £4,015 per MONTH for 120GB RAM and 30GHz CPU, one can buy AND OWN server for that; not talking about £795 for 50Mbps connectivity per month, you are kidding, right ? and what about £665 production support PER MONTH, when you can have the same for much less for a year when you buy a box license ? ref.6 and they say some offers are 83% cheaper than competition, ref.7]

2. Cloud vendors

It is not possible to turn this pamphlet into "Extremely detailed cloud comparison PhD. thesis". Let's try to keep it short and concise because in recent Gartner's study, 205 criteria have been evaluated across compute, storage, networking, Security/access, service offerings, support levels, management and price/billing categories.

Numerous cloud providers and vendors compete on market with their offerings, for example – in alphabetical order – Akamai, Amazon, AT&T, CA, Citrix, Cloudera, Data Processing, Dell, Fujitsu, Google, Hewlett-Packard, IBM, Microsoft, Oracle, Rackspace, Red Hat, Soft Layer, T-Systems, VMware, and that is just short list not having an overview about providers in Asia as that market is totally unknown to me.

2.1 Brief details of existing major cloud infrastructures

It is next-to-impossible to create 100% detailed and thorough overview of functionality, features, advantages, shortcomings and pricing models of cloud solutions mentioned above. Out of the list, the biggest public/hybrid offerings are Amazon Elastic Compute Cloud (EC2) which is central part of Amazon Web Services platform, Google Cloud Platform, Microsoft Azure and VMware's vCloud Air. These are the most common choices for evaluation today.

All these vendors have some shortcomings. Microsoft Azure doesn't support FreeBSD and Red Hat operating systems at all (ref.1 + ref.2), some others don't support Microsoft Failover clustering or generally Microsoft Clusters at all due to deficiencies in storage infrastructure and virtualization platform (references not available, quite often you run into problems when talking to sales rep only).

Some infrastructures are not certified to run wide-spread critical business applications for example SAP HANA or Oracle stuff. Several do not integrate well or easily with customers' existing private datacenters what instantly disqualifies them from business perspective. Level of automation is also extremely different between vendors.

2.2 Choosing the best cloud solution

It is very daunting task to choose "the right" or "the best" hybrid cloud provider as these terms mean very different things for different customers. Some of reasons are :

- sheer amount of options and variants available with each vendor (ref.3)
- extremely problematic to DIRECTLY compare offers from various vendors
- bulks of CPU, RAM, storage, networking resources has to be bought and they are not aligned to competitors' offers, once again making "direct" comparison either utterly complicated, or impossible at all depending on scale of environment
- technical differences between competitors. Can you tell the EXACT difference between VMware's "SSD Accelerated Storage" and Amazon EC2's "Provisioned IOPS" or "General IOPS"

(no reference at all, I haven't found it on public materials) ? How much IOPS does VMware SSD Accelerated Storage provide for read operations, Ref.4 ? How much it does for write operations, Ref.4 ? With vCloud Air, are writes accelerated with SSD Accelerated storage or are they not ? From the look of it, this seems to be Flash Read Cache ONLY, so how much performance will my databases have ? Did this change with introduction of ESX 6.0 ? What is burstable IOPS and bandwidth in terms of MB/s and how long such burst is allowed to take place ? Will performance of writes drop after some time due to configuration limits, but also due to technology background ? We all know several SSD models can sustain their performance for longer (and better) than some others, also there is huge difference in latency evenness and you will never know what is exactly the storage used for your project in cloud.

- What is the cost of your downtime and how much of them will you have ? Nobody can exactly tell you these factors, but ref.5 http://blog.awe.sm/2012/12/18/aws-the-good-the-bad-and-the-ugly/#~pijAMzVRGudYi8

- Limited amount of resources. Some cloud infrastructures allow only 120GB of memory for VM, some limit projects to 6TB "SSD Accelerated storage" – damn, what if I need 50TB of flash storage ? Eh, can I really have PURE SSD storage at all or is it always some form of "pony trick" with the super expensive SSDs ? Even if I could, what the heck would be price of it ?

- Performance problems, a lot of existing BIG NAME infrastructures have very surprising and unexpected performance problems, mainly in terms of storage performance. **Storage seems to be the most limiting, the most troublesome technology today, ESPECIALLY IN BIG SCALE**, no matter how strange that seems (ref.5). Quote "I/O rates on virtualized hardware will necessarily suck relative to bare metal, but in our experience EBS has been significantly worse than local drives on the virtual host".

- Redundancy and business continuity: number of datacenters of particular provider, their geographical locality might be known, but "weak points", let's call them **real failure zones** , are invisible, unknown and never explained to customer as this is the top secret information. In order to protect workloads effectively, customers must accept additional unnecessary costs which vary between vendors and their BC/DR possibilities. It's extremely complicated to thoroughly understand cloud options and their shortcomings.

- Internal infrastructure "bindings" – for example several Amazon AWS services relying on EBS storage, for example Elastic Load Balancer (ELB), Relational Database Service (RDS), Elastic Beanstalk and some others were tied to EBS, so when EBS crashed, these services crashed too. Even when customers were not using EBS and paying huge extras for other storage options, they were affected. There is absolutely no visibility and no information available about these relations and infrastructure shortcomings.

- Several fuzzy specifications are provided, for example Amazon's EC2 used "ECU", evasive compute units, to specify CPU speed. According to some reference, ie. http://aws.typepad.com/aws/2013/11/a-generation-of-ec2-instances-for-compute-intensiveworkloads.html, Intel E5-2680 v2 CPU has ECU performance equivalent of 68. I can't imagine better reason to create such terrible "performance measurement units" than making customers' decisions as complicated as it possibly gets. There are no other words about it. How much CPU power needs my database/OLAP/OLTP ?   1000 ECUs or 10000 ECUs ? Will it be faster than quad-socket E5-2680 system ? What is it equivalent of ? Single socket 100core CPU ? Or 50sockets 2core ?

- several hidden costs, quite often not visible in official calculators, such as management costs.

There are huge number of other variations and details, efficiently spoiling comparisons. Rough estimates/calculations can and should be provided depending on exact sizing of particular project and expected bursts or growth. TCO is all what matters.

2.3 There's the problem with Microsoft Azure

Because we will be running mixture of operating systems on virtualized platform on Mars, Red Hat and FreeBSD inclusive, Microsoft's Azure cloud is disqualified from competition at the beginning. Similarly, VMware platforms are in very early adoption stage in Azure premises which doesn't sound like the right choice for life-critical systems we need to run on Mars.

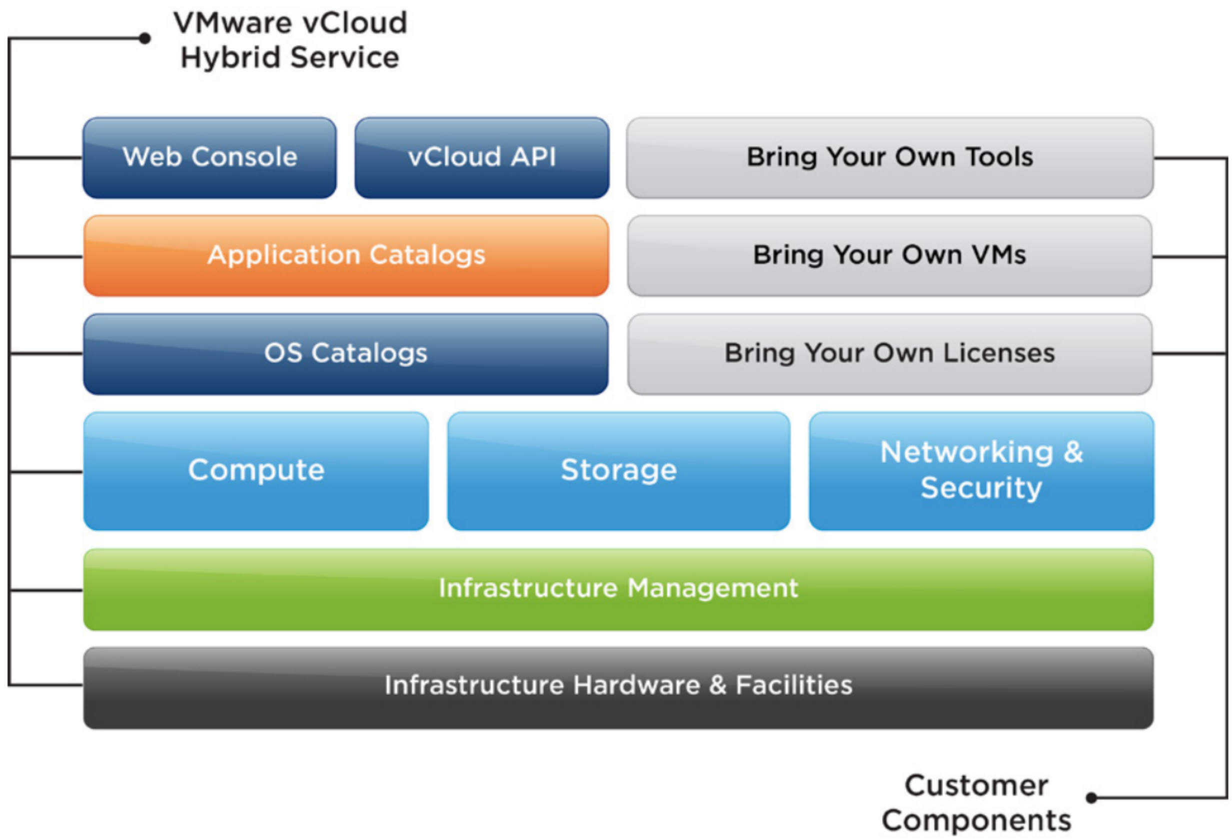
For these reasons, Microsoft Azure platform can't be considered for our purposes from technical perspective, without considering any other criteria.

2.4 VMware vCloud Air

One of the most investigated and native options for customers already running VMware platform is, for sure, VMware's vCloud Air. Physical datacenters are USA, United Kingdom, Germany, Japan, Australia. More than 5000 applications and 90 operating systems are certified to run on vCloud Air.

vCloud Air lets businesses move workloads between on-premise servers and the cloud products and services using the same VMware tools that are used in-house : eliminating additional costs, reconfiguration or new knowledge skills and learning curve. The service is comparable to Amazon Web Services, but also integrates with existing virtualized environments, their management and automation tools which protects investment customers already performed. The major advantage is no rewrites or recoding when workloads are moved from customers internal premises to vCloud Air, ESX v6.0 is targeted to utilize this functionality as much as possible with long-distance vMotion (up to 100ms latency) and across-ESX-logical-datacenters migrations.

Network virtualization in the form of VMware NSX product allows customers to configure firewalls and network to mirror on-site networks, including NAT rules and firewall rules, networks and public IPs to extend existing Layer 2 or Layer 3 networks from their datacenters to the vCloud Hybrid Service :



As a part of service, Data Protection is available. While not exactly cheap, it is self-service backup offering that gives granular control to the consumer. No workarounds need to be taken in order to PROPERLY protect workloads of customers – this is integrated part. List of features follows :

Feature Description	
Self-Service enablement for feature add-on	✓
Paid usage tier based on backup storage (FETB[2])	✓
Self-service vApp registration	✓
NEW Custom protection policy affinity settings per VDC or vApp	✓
NEW Custom retention policy settings[3]	✓
NEW Self-service (in-place) restores per vApp or VM	✓
NEW Self-service (out-of-place) restores per vApp or VM	✓
NEW Self-service backup scheduling	✓
NEW Ad-hoc backups	✓
NEW Full backup and restore activity reporting	✓
NEW At-a-glance admin dashboard views	✓
NEW Data-at-rest encryption	✓

For detailed options, consult http://vcloud.vmware.com/

With ESX v6.0, there is no better eco-system from any other vendor, providing tighter integration between private and public virtualized/cloud infrastructures. Microsoft with Windows 2012 R2 and Hyper-V offerings doesn't achieve what VMware easily provides in terms of functionality, manageability and integration.

Technical details, such as networking speed, number of virtual machines per vApp, maximum configurations of VMs, maximum number of virtual NICs per VM and maximum disk size are factors for choosing vCloud Air over AWS :

parameter	Google AWS	VMware vCloud Air
Networking speed (max)	1Gbps	10Gbps
Maximum RAM per VM	244GB	1024GB
Maximum CPU per VM	32	64
Number of vNics (max)	8	10
Maximum disk size	Approx. 44TB	62TB

We fully agree that extremely tight integration of most recent ESX platform, v6.0, and single original vendor support is by far the major decision point for going with vCloud Air for customers with existing local ESX infrastructures. Identical importance has been shown by Gartner study : despite severe shortcomings of Microsoft Azure platform, Gartner points out 64% of users' biggest reason for choosing Azure was their existing relationship with Microsoft.

For reasons discussed in paragraph 2.3 and 2.4, VMware vCloud Air has been chosen as the platform to build on.

3. Applications

Following applications will be running in hybrid or public premises:

a) life supporting critical systems : command and control centre for oxygen/water supply

b) time tracking web application for botanists in greenhouses plus reservation system for resources

c) email communication server

d) internet cache - list of favorite webpages mirrored nightly so Marsonauts can read their favorite content.

e) social networking stuff, collaboration platform

3.1 Application requirements

Application requirements are :

R01 : performance in terms of user experience and low hardware demands (HW power/cooling, space)

R02 : expandable capacity (online without downtime)

R03 : high availability, crucially important for life-supporting systems, RTO 20 minutes max.

R04 : compatibility : web-based apps must run in any browser, such as Internet Explorer, Chrome, Firefox, Opera and any platform (Windows OS, Apple, Android)

R06 : scalability (number of Mars citizens expected to grow exponentially)

R07 : low bandwidth usage

R08 : resistance to network disruptions and communication outages

R09 : effective storage usage.

R10 : if possible, use dockerable applications to save resources (RAM, storage)

3.2 Application infrastructure

3.2.A Application A : Life Supporting Critical Systems

This is by far the most critical technology we will every run on Mars as it is command and control centre for oxygen/water supply. This application only can tolerate 20 minutes recovery time objective, then building objects on Mars will run out of oxygen and people will die.

High-level application functionality :

- collect data from sensors
- transfer data from each object, each sensor to servers for processing
- store data on servers in protected database
- react on events (less oxygen)
- application-based redundancy, multiple technology controls can be bound to single instance
- master/slaves architecture (think active/multi-passive config)

All three datacenters available on Mars run this application simultaneously in replicated mode, providing active-active-active redundancy. Application runs concurrently in all three datacenters, behaving as active for closest object and passive (stand-by) for all distant objects. Technology available at each site can be controlled by ANY application, local or remote, this is used for redundancy purposes.

This application is NOT latency sensitive, as it is not that much important if data about oxygen composition in each object is delivered in 2ms or 89ms. At the same time, bandwidth required by this application is extremely low, too, because control and status messages only are 240 bytes large and they are sent every 30 seconds only.

Local sensors are built with cache memory – if transmission fails, results are cached up to 34 because local cache memory is 8192 bytes. When connection is established again, all cache content is transferred. This gives opportunity for 1020 seconds outage = 17 minutes, because messages are sent in 30 second intervals and 34 of them can be cached (34x 30 = 1020). Each message contains incremental identifier, which serves as “order arbiter” in case of communication outage, think something like TCP/IP packet ordering mechanism.

Confirmation of message reception is sent by servers back to technology in each object ; in case when data are not received within five seconds of expected arrival (30 seconds timing), caching of data and retransmission occurs.

Operating system chosen for application : Red Hat version 6.6, recommended by Honeeyvell vendor

Clustered database : MySQL Cluster edition, requires no operating system clustering, NO SHARED STORAGE, can be geographically distributed, can be backed up ONLINE

Clustered web front-end : NGINX server is used to provide data to operators and administrators. If it fails in one datacenter, all remaining two can both control technology in each site and provide status messages.

Clustered data collection and distribution : after message is received by any single application instance (remember there are THREE running, each in separated datacenter, each with separated non-shared database), applications communicate together to verify if all remaining partners received the same message.

Configuration of 1 vCPU, 8GB RAM and 50GB storage is requested for each Oxygen VM :

CPU	1
RAM	8
HDD / tier / IOPS	50GB / fastest / 1000 IOPS guaranteed
Networking bandwidth / priority	1Mbit guaranteed, no max, highest priority

Total SIX Oxygen VMs will be running in Datacenters :

Datacenter 1	OxygenVM1
Datacenter 1	OxygenVM2
Datacenter 2	OxygenVM3
Datacenter 2	OxygenVM4
Datacenter 3	OxygenVM5
Datacenter 3	OxygenVM6

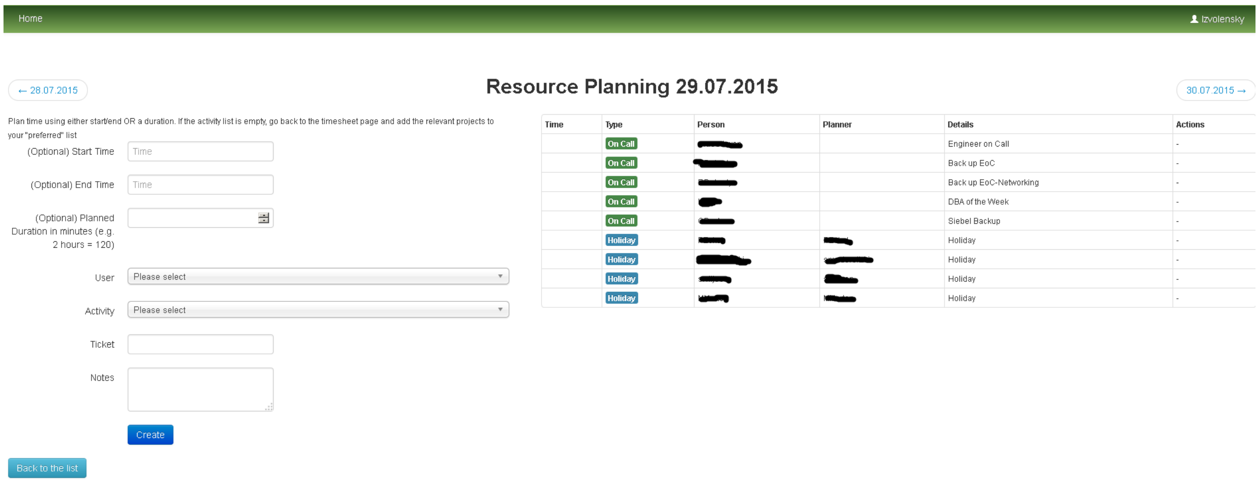
Easy backups and restores of configuration is mandatory. Applications can be restarted independently in each datacenter ; each datacenter also runs TWO separate copies (synchronized on application and database levels).

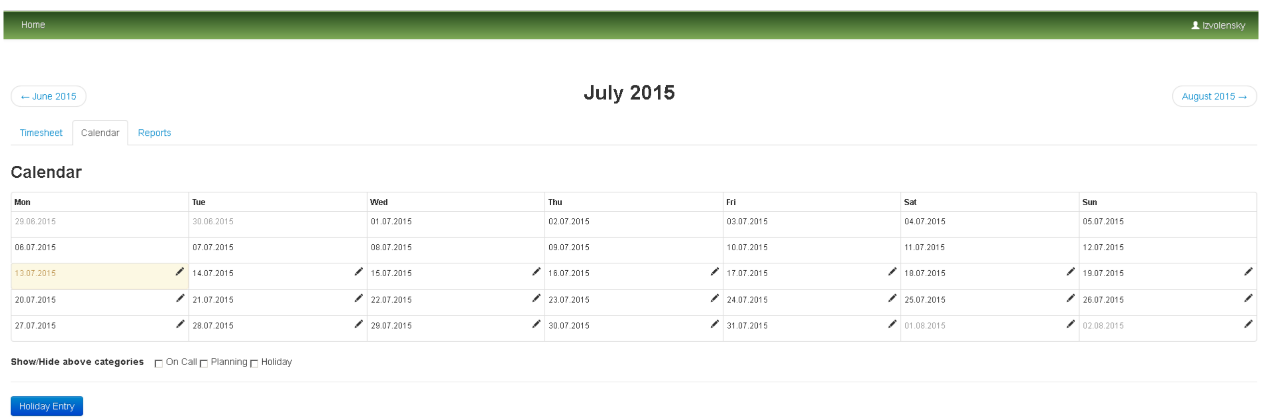
Remember we have perfect infrastructure running on Mars, based on my ~12% complete design from Challenge1 (with missing vCenter, network setup, redundancy, business continuity/disaster recovery and some other “unnecessary minor things”), so there is no problem with storage performance due to all-flash VSAN configuration, network throughput (multiple 10Gbit Ethernet interfaces, some 40GbE) or resources (huge 3TB RAM physical servers).

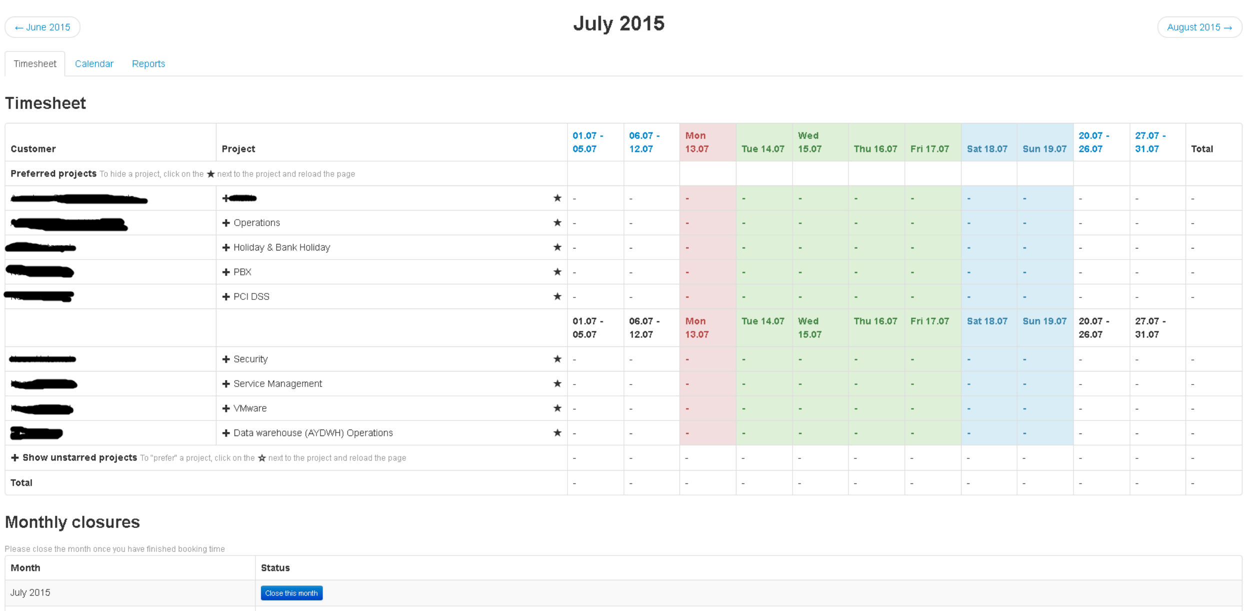
In terms of storage IO, maximum 50 IOPS is requested each 30 seconds for application as each transaction message can fit into single IO. This means very low demands despite extreme criticality.

3.2.B Application B : time tracking web plus reservation system for resources

For time tracking and resource reservation system, an internal application has been created by one of Marsounauts. It provides phenomenal time tracking possibilities, easy to use… also reservation system is created on the same principles with definable resources :







← June 2015

July 2015

August 2015 →

Timesheet | Calendar | Reports

Timesheet

Customer	Project	01.07 - 05.07	06.07 - 12.07	Mon 13.07	Tue 14.07	Wed 15.07	Thu 16.07	Fri 17.07	Sat 18.07	Sun 19.07	20.07 - 26.07	27.07 - 31.07	Total
Preferred projects To hide a project, click on the ★ next to the project and reload the page													
[redacted]	+ [redacted] ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ Operations ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ Holiday & Bank Holiday ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ PBX ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ PCI DSS ★	-	-	-	-	-	-	-	-	-	-	-	-
		01.07 - 05.07	06.07 - 12.07	Mon 13.07	Tue 14.07	Wed 15.07	Thu 16.07	Fri 17.07	Sat 18.07	Sun 19.07	20.07 - 26.07	27.07 - 31.07	
[redacted]	+ Security ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ Service Management ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ VMware ★	-	-	-	-	-	-	-	-	-	-	-	-
[redacted]	+ Data warehouse (AYDWH) Operations ★	-	-	-	-	-	-	-	-	-	-	-	-
+ **Show unstarred projects** To "prefer" a project, click on the ☆ next to the project and reload the page		-	-	-	-	-	-	-	-	-	-	-	-
Total		-	-	-	-	-	-	-	-	-	-	-	-

Monthly closures

Please close the month once you have finished booking time

Month	Status
July 2015	Close this month

Each category has several sub-categories and/or projects as defined by approved users.

Lean, small, blatantly fast, new items can be defined by users based on their privileges (for example when new dock is built for rockets, approved users can enter it to this application so everybody can reserve it for their vehicles). Another examples are gas station - it takes long to fill the tanks of rockets so proper planning and queueing is necessary !!, and yeah we got new cinema EEMax style (because we only have 12 seats, bring your laptop, Red Camera, android tablet or anything with vga, dvi, hdmi, displayport connection and play it BIG BIG BIG!).

One day, we will be bribed errrr convinced we need 50 mil $ S.A.P. to track attendance for citizens and plan resources on Mars, but not now.

Configuration of 1 vCPU, 4GB RAM and 10GB storage is requested for Citadel VM :

CPU	1
RAM	4
HDD / tier / IOPS	10GB / slowest / no guarantee necessary
Networking bandwidth / priority	128kBit guaranteed, 1Mbit max, lowest priority

This thing is any linux distribution plus NGINX on top. A small directory with scripted software on top. As easy as that. No special requirements are necessary in terms of resources or performance, this is extremely, extremely lean software. We will never be able to overload the web server, even if there is million people on Mars : let's be realistic… http://g-wan.ch/benchmark/babel.html -> http://www.statisticbrain.com/google-searches/

No java, no flash, runs in any browser we can think of, even on Apple iEverything. Yikes.

3.2.C Application C : email communication server

3.2.D Application D : internet cache (proxy)

3.2.E Application E : social networking stuff, collaboration platform

Marsonauts like to read content available down on Earth on internet, they have their preferred web pages they would like to follow even being so terribly far away.

Moreover, challenge 1 specifies that some form of social collaboration is preferred for Mars citizens. Of course, everybody needs email today and that's not going to disappear anytime soon.

In order to save resources, which are very scarce on Mars, we decided to integrate three "applications" into one virtual machine running Citadel software.

Following is feature list of Citadel platform :

- Email, calendaring, address books, bulletin boards, instant messaging
- Wiki and blog engines built in. Citadel is a collaboration server **and** a content management system
- Web browser, telnet/SSH, local client software accessible
- Standards-compliant e-mail built in: IMAP, POP3, ESMTP
- Group calendaring and scheduling (WebDAV, GroupDAV, and Kolab-1 compatible)
- Built-in listserv (mailing list server)
- Built-in RSS Feed Aggregation
- Support for push e-mail and mobile devices
- Database-driven, single-instance message store
- Authenticated SMTP for remote email submission
- Multiple domain support

- Built-in integration with perimiter email filtering technologies such as Realtime Blackhole Lists (RBL's), SpamAssassin, and ClamAV antivirus

- Server-to-server replication. Users in any number of domains can be spread out across any number of Citadel servers, allowing you to put data where you need it, and enabling infinite horizontal scalability.

- Web-based access to email, calendars, and everything else through a powerful AJAX-style front end

- Very strong support for “public folders” and message forums.

- Built-in instant messenger service

- SSL/TLS encryption for all protocols

Configuration of 1 vCPU, 8GB RAM and 50GB storage is requested for Citadel VM :

CPU	1
RAM	8
HDD / tier / IOPS	50GB / slowest / no guarantee necessary
Networking bandwidth / priority	128kBit guaranteed, 16Mbit max, lowest priority

Due to low number of citizens on Mars and their low usage of “Internet”, one virtual CPU and 8GB RAM are more than enough for Citadel software. Capacity of storage depends only on amount of data to store, we chose 14GB for proxy cache in beginnings as not more than 0.5GB can be transferred during night [!!] and no more than 28 days worth of “caching” is considered necessary which equals to 14GB consumption. Other capacity will be consumed by local operating system (RedHat) and database. There is no need to separate database from frond-end web services, because one can’t exist without the other and there is no significant additional risk putting all eggs into one basket. This isn’t a critical application at all.

Please note web content and social media are TYPICALLY compressible extremely well, so compressed filesystem will be used on RedHat which will allow to hold much more than projected 28 days of data cached from internet. If disk space gets to critically low levels, oldest data (three days) will be simply thrown away automatically.

All resources can be expanded online if necessary. In terms of storage IOPS, 500 is expected normal usage maximum but because we have phenomenal all-flash VSAN in place, we don’t need to limit this on VM level.

Network traffic demands are pretty low, with no huge bursts expected at all. 16Mbit maximum should be more than adequate for “intra-LAN” Mars network – this is absolutely enough for displaying web pages, social activities, forums, wiki, blogs and similar activities. Lowest priority was chosen due to non-critical nature of these applications.

Because we initially have enough compute resources and this is pretty lean VM, disaster recovery counts with simply restarting this VM in surviving datacenters in case of catastrophy. No special protection is necessary.

During time, when number of citizens on Mars expands and configuration of this VM will not be sufficient anymore, it can be easily expanded in terms of adding CPU, RAM or storage space.

References

1. https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-linux-endorsed-distributions/

2. https://support.microsoft.com/en-us/kb/2941892

3. http://calculator.s3.amazonaws.com/index.html

4. http://pubs.vmware.com/vca/index.jsp?topic=%2Fcom.vmware.vca.od.ug.doc%2FGUID-57C2D483-6BF6-46A8-8A2C-D591F29D4FB9.html

5. http://blog.awe.sm/2012/12/18/aws-the-good-the-bad-and-the-ugly/#~pijAMzVRGudYi8

6. http://vcloud.vmware.com/uk/service-offering/pricing-guide

7. http://blogs.vmware.com/vcloud/2014/08/vcloud-hybrid-service-35-cheaper-azure-83-cheaper-aws.html