|  |  |
| --- | --- |
| Design Statement | Cosmos Revolution |
| 4.1 | Expanding Cosmos: Cosmos++ |
|  | Bob Jenkins |

**Abstract**

The Cosmos project stores and processes big data for Bing within Microsoft. The existing Cosmos has scaling and usability issues that could be solved by a simple but fundamental infrastructure change: Cosmos++ will have four dimensions rather than three. This will be backwards compatible with most existing interfaces (Clientlib, C++, the speed of light), and will square our ultimate data capacity. However migration will require pandimensional replication.

**Revisions**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Summary** | | | **Author/Editor** | |
|  |  | | |  | |
| 2013/04/01 | | Created | Bob Jenkins | |
|  | |  |  | |

# Glossary

* **Alpha wave**: the speed of human thought, about 10 Hz, that is, 10 per second.
* **Atom**: Same as in the old Cosmos. Hydrogen has 1 proton, Helium has 2 protons and 2 neutrons, et cetera. Water is H20, and has an atomic mass of 18, in both Cosmos and Cosmos++.
* **Avogadro’s number**: 6.022e23 in both Cosmos and Cosmos++.
* **Breadth**: the second dimension, same as in the old Cosmos. Left and right. The direction from one hand to the other, one foot to the other, one eye to the other. Different from width.
* **Day**: one rotation of the earth (along the XY plane).
* **Depth**: the fourth dimension, the direction away from you. Same as in the old Cosmos. Things appear smaller at greater depth.
* **Gram:** water has an atomic mass of 18, and avogadro’s number of water molecules weighs 18 grams. The mass of a human is about 50 to 100 kilograms in both Cosmos and Cosmos++.
* **HDD:** Hard Disk Drive in Cosmos, Hard Drum Drive in Cosmos++. In Cosmos++ a platter is a cylinder rather than a disk, with a height equal to the diameter of the circle.
* **Height**: the first dimension, the direction of gravity. Up and down. Same as in the old Cosmos.
* **Lightspeed**: 3e8 meters per second, in both Cosmos and Cosmos++.
* **Meter**: roughly half the height of a human, in both Cosmos and Cosmos++. There are 200 Cosmos++ meters in a Cosmos meter.
* **Nanometer (nm)**: one billionth of a meter.
* **Night**: one rotation of the earth (along the WZ plane).
* **Second**: about ten alpha waves. There are 200 Cosmos++ seconds in a Cosmos second.
* **Time:** not one of the four dimensions in Cosmos++. You could call it the fifth dimension.
* **Width**: the third dimension, awid and away, widward and wayward. New to Cosmos++.

# Problems and a solution

The existing Cosmos has several issues and limitations.

* Capacity, throughput, and IO speeds must increase by several orders of magnitude
* Error messages are too big to fit on a screen
* Not enough hours in a day
* Not enough window offices
* Commutes are too long

These could all be fixed by a simple but fundamental change to our infrastructure. Building off of research from MSR and CERN, Cosmos++ will be backwards compatible with Cosmos, but it will have four dimensions rather than three: height, breadth, width, and depth.

# Constants and scaling factors

Certain things remain fixed, for backwards compatibility:

* the number of transistors in a processor (230)
* the number of atoms in a human (292)
* the clientlib interface: sdk\cosmos.h and sdk\cosmoserror.h
* the speed of light (3\*108 meters/second)

Other measures will be scaled to maintain the existing user interface:

* Humans are about 2 meters tall
* A second is perceived to be as long as before
* Gravitational pull at earth’s surface is enough to retain an atmosphere
* Clientlib message timeouts remain 50 seconds

Atoms compact better in four dimensions than in three, so there are 200 Cosmos++ meters per Cosmos meter. Our perception of time scales inversely with our size, so there are 200 Cosmos++ second per Cosmos second. This affects the measures of common things:

* The speed of light is 3\*108 meters/second in both Cosmos and Cosmos++
* Cosmos++ atoms are 20nm in diameter, while Cosmos atoms are .1nm in diameter
* Processors in Cosmos++ are equivalent in speed and power to those in Cosmos
* HDD’s are the same capacity and have the same seek times.

# Changes to hardware

Nanolithography is much coarser in Cosmos++ (5 micron resolution rather than 25nm resolution), but it has 3 dimensions to play with instead of 2. Intel has had 3d chip designs for ages. This is their opportunity to apply them in bulk. Heat dissipation and layering is no longer an issue for them, because every 3d point is on the surface of the chip. Intel can then start looking into 4d circuitry. Scaling of meters and seconds keeps the number of transistors per chip and chip speeds about the same in Cosmos and Cosmos++, so the experience of using computers remains about the same.

HDDs have the same seek time and capacity in Cosmos++, but they are more expensive and scans are faster. The platter is a disk in the XY dimensions extended to a drum in the Z dimension. It rotates in the XY plane, leaving W and Z dimensions fixed. The disk arm hovers above the platter in the W direction, reading one spot (X=?,Y=0) on the rotating XY disk, but an array of heads (which adds expense) reads the whole Z dimension in parallel. Disk scans are 10GB/sec instead of 150MB/sec (capacity^^2/3 rather than capacity^^1/2).

If disks read from just one zero-dimensional head, as in Cosmos, speeds would be about 1MB per second and HDDs would not be competitive.

This suggests HDD drive enhancements for old Cosmos. 4TB disks have 4 platters, and 8 heads (one on each side of each platter), but scan rates are no better than 1 platter with 1 head. This is because HDDs allow only 1 head to be active at once. The heads already move in sync. If data were striped and all heads were allowed to read/write at once, read/write speeds would be 1.2GB/sec rather than 150MB/sec. Having n heads packed at the tip of each arm would also allow scans to be n times faster.

Displays are better in Cosmos++.

* 3d monitors are the default, with headache-inducing 4d projections mostly confined to movie theaters.
* A typical textbox is 120\*120\*40 characters (576KB), not just 120\*40 (4KB).
* Error messages can be much larger without scrolling off the screen.
* Developers can view megabytes of code at once in Visual Studio, which allows gestalt understanding of larger problems.
* While 3d macho developers are limited to about 12 (3\*4) monitors, the 4d ultra geek can fit 48 (3\*4\*4) monitors on their desktop.

Cosmos++ data centers are better.

* Containers are 4d, so an extra level (rack) fits between pod and container. 45 machines per pod, 50 pods per rack, 40 racks per container, giving 90,000 machines per container.
* While the 3d Chicago datacenter could park up to 56 (7\*8) containers, the 4d Chicago datacenter can park up to 336 (6\*7\*8). This gives 4d Chicago a maximum capacity of a quarter billion machines, orders of magnitude greater than in old Cosmos.

The size of a data center limits the size of jobs that can avoid cross-data-center latencies, so this scales the ultimate size of Cosmos jobs by several orders of magnitude.

# Changes to software

Cosmos++ is backwards compatible with Cosmos. Although using four dimensions rather than three has clear implications for Intel and Seagate and Kinect, Cosmos code hardly changes at all. Assembly language doesn't change. Neither does C++, C#, CsIo, the clientlib interface, or Scope. Caches get bigger and faster, which require some tuning.

Existing tools for monitoring clusters and Scope jobs, which project data to displays which are now 3d rather than 2d, will need refactoring.

Images and movies in Cosmos++ are 3d rather than 2d. This means they take more storage and/or are grainier. A typical picture (or movie frame) will be 100x200x200 voxels in Cosmos++, rather than 1000x2000 pixels in Cosmos. Certain lossy image compression techniques – cartoonification, and fractal compression – will be bigger wins, because the idea that a picture represents typically has a fractal dimension of about 1.5. The mismatch between 1.5 and 2 dimensions isn’t as pressing as the mismatch between 1.5 and 3.

# Environmental changes

Offices are better in Cosmos++. For example, a Cosmos building with 5\*20 offices per floor would have 100 offices per floor, 54% of which have no windows. Growing a team beyond 100 people is problematic, because people avoid stairs. But a Cosmos++ building with 5\*5\*20 offices per floor has 500 offices per floor, with only 32% with no windows. All 500 people would be within 20 offices of each other without climbing stairs.

Shoelaces don't work, you can't make knots from 1d surfaces in 4d space. But velcro (hooks, plus hollow 3d spheres as loops) still work. The cosmos discussion list would have to deal with the frequent question, "why can't I tie my shoes?" Jeans and cotton T-shirts are out (again, anything held together by tying or weaving strings falls apart). But leather works, and clothes could be fashioned from woven 2d surfaces.

Tripods don't work. You need four legs to make something statically stable. You know how to place tripod legs in an equilateral triangle in 3d? In 4d, place the four legs at the four points of a tetrahedron on the ground. This makes the design of most vertebrates look more elegant. Birds and people remain bipedal. Spiders still have too many legs.

Commutes are better in Cosmos++. Roads remain 1d (lines connecting here to there), but the surface of the earth changes from 2d to 3d (plus up/down). So there would be no intersections, no bridges, just on and off ramps. Lines must intersect on a 2d surface, but not on a 3d surface. Also, commutes are at least 6x shorter on average, due to 3d packings of houses being denser than 2d packings.

Proportionately less real estate needs to be spent on roads. More can be spent on parks and backyards.

In Cosmos it takes a city to have enough density of people to support a shop dedicated to umbrellas. In Cosmos++, most places will have enough people nearby to support such a shop. Crime is also a function of how many people you can reach in a given amount of time, so crime will pay better, so there will be more public surveillance to counter that.

# Changes to the world

Cosmos++ would change our world.

As a consequence of Gauss's law, electric and gravitional fields will drop off with the inverse cube of distance.

The earth's radius would be 200x bigger. Earths' gravity has to be the same as today in order to hold an atmosphere. Mass grows with fourth power of radius but gravity drops with the third power, so that much gravity requires earth's radius to be the same number of atoms deep as today. But atoms are 200x wider in Cosmos++. The earth's would appear to be twice as big as today's sun.

The speed-of-light delay from one side of the earth to the other would be 8 seconds, so local data centers become vital.

There would be no moon. There’s no way it could keep a stable orbit around the earth.

Assuming a 20\*30\*30 meter house lot in Cosmos++ feels as big as a 20\*30 meter house lot in Cosmos, the Cosmos++ earth would have 2.6 trillion times the living space as Cosmos's earth. More potential Bing customers!

Despite having such a big world, and despite momentum feeling the same, gravity would feel 200x less. It's meters per second per second, which multiplies by 200 once but divides by 200 twice, so it feels like 0.049 meters/second/second. Everyone would be superman, able to leap 40-story buildings in a single bound.

Human-powered flight will be easy. You can get wings at Costco.

The Cosmos++ sun would be a line across the whole sky, which would almost never entirely set.

A day in Cosmos++ would be 200 Cosmos days. More hours in a day!

Earth would have two independent rotation rates in the XY and WZ planes. One would be called day, and the other, night. Nights are longer than days.

In Cosmos++, earth would have no poles (poles only appear in odd-dimensional spaces). But the earth would have two special non-intersecting circumferences, a quarter circumference apart, which rotate in circles in only two of the four dimensions. One would have a period of a day, the other a period of a night.

# Cosmic changes

Cosmos++ would be a better Cosmos.

* Two spheres can only form stable gravitational orbits in 3d. Orbits about a line, though, are stable in 4d. Globular clusters are fairly stable in 4d as well.
* The Cosmos++ sun would not be a ball. It would be a ring one light-year in circumference, rotating slowly in the XY plane. High density spots would prompt fusion, which explodes, which causes low density, preventing the ring from collapsing into a chain of balls.
* The Cosmos++ earth (which would be a ball) would orbit a fixed XY spot on the sun, circling in the WZ plane. (This should be extensively unit tested, since unexpected instabilities in earth’s orbit could easily cause live site issues, leading to the DRI being called at odd hours.)
* Cosmos++ galaxies would be globular clusters or spirals, same as today. Spiral galaxies happen when two globular clusters have a close pass and pull out bits of each other.
* Cosmos can only hold an amount of data linearly dependent on its radius, otherwise it collapses on itself as a black hole. That's why today's universe is so big and empty. Cosmos++ could store an amount of data proportional to the square of its radius. Cosmos++ squares Cosmos's data capacity.

# Migration

In-place migration will not be supported. Customers will experience this as a large shock, not an incremental change. Migrating from Cosmos to Cosmos++ will require pandimensional data replication.

MSR recommends utilizing a spinning black hole. However CERN advises thorough integration testing, because spinning black holes are prone to causality violations. CERN also suggests that we may have to replicate in two hops, passing through 26-dimensional space in the middle.

All vital data on the 3d Cosmos earth could be replicated to an identical 3d region on the surface of the 4d Cosmos++ earth. This would preserving affinity, while leaving an unoccupied but usable center that legacy applications can expand into.

# Summary

Cosmos needs to scale by many orders of magnitude. A four-dimensional Cosmos++ can deliver those gains. I for one am super-excited about this opportunity for Cosmos, Bing, and Microsoft as a whole.