

Couchio

CERN – CouchDB and Ease of Use

Situation

CERN, the European Organization for Nuclear Research, is the home of the Large Hadron Collider (LHC), the world's most powerful particle collider. There are four main experiments that run on the LHC and the Compact Muon Solenoid (CMS) experiment, one of two general purpose detectors, uses CouchDB extensively in its Data Management and Workflow Management (DMWM) project. DMWM is responsible for providing all the offline processing infrastructure to CMS. This includes cataloguing data, transferring it between their ~100 computing centers, running processing tasks across the distributed computing infrastructure at all stages of the data's lifetime (from when it first gets to CERN from the detector to end-user analysis), and creating simulated data so scientists can match experiment to theory. CMS will collect roughly 10PB per year when the Large Hadron Collider is running at design luminosity (a measure of how many proton-proton collisions happen) with at least as much simulated data being created.

The DMWM team has issues that don't fit well into standard relational databases or files in a filesystem. Being able to easily access and consolidate data from distributed locations with minimal latency is required routinely. Typically, external access to a site is limited, so incoming connections to a database aren't possible. The team often doesn't have clear requirements to address, which means metadata is either not collected or not effectively used. Generally, the team members must prototype tools quickly and be able to demonstrate that they are ready to go into production.

Solution

All the features of CouchDB are called upon in various combinations across their applications. The speed of development on CouchDB is of huge value to the DMWM team. For example, a summer intern, who had never seen CouchDB before, was able to create a complete application built on top of CouchDB to serve the project. The core of an application used to manage the recall of data from petabyte tape stores was written in an afternoon. This could not have been accomplished as quickly with other solutions.

Results

The DMWM team in CMS found CouchDB to be a natural fit for their environment. Using CouchDB means they don't have to write and maintain large pieces of code. This becomes critically important when the application is needed urgently or is a small component of a wider system — which is the norm for the DMWM team's workload.

About CERN

CERN, the European Organization for Nuclear Research, is one of the world's largest and most respected centers for scientific research. Its business is fundamental physics, finding out what the Universe is made of and how it works. At CERN, the world's largest and most complex scientific instruments are used to study the basic constituents of matter — the fundamental particles. By studying what happens when these particles collide, physicists learn about the laws of Nature.

Founded in 1954, the CERN Laboratory sits astride the Franco-Swiss border near Geneva. It was one of Europe's first joint ventures and now has 20 Member States.

www.cern.ch

Quote

Simon Metson, Convener of the DMWM group, says “CouchDB really fits the way we work, from the distributed nature of our collaboration to the way we exchange data to our distributed systems. We have found CouchDB to be really easy and enjoyable to develop against. It's nice using something ‘built of the Web’ at the lab where the ‘web was born’, too.”

Who we are



About CouchDB

Apache CouchDB bridges the cloud and the local network by making it simple to write web apps that can scale up to the datacenter and scale down for local use on a smartphone. The synchronization engine (inspired by Lotus Notes) allows sharing of data and applications across ad-hoc groups. Offline replication gives mobile devices better performance (lower latency) while saving battery life by avoiding radio usage. With offline replication solved in a simple way, application developers are free to focus on the code that matters

CouchDB is popular with developers because its document model is a closer fit for many domains, leading to less overall complexity. The JSON HTTP API allows developers to write dynamic database-backed applications without a middle tier: 2-tier CouchApps require only a web-browser and a CouchDB, and can be written by anyone with Ajax / jQuery experience. We firmly believe that empowering people to share data and custom applications at the edge of the network will lead to a more humane web.

Apache CouchDB is currently deployed by the BBC, Meebo, Assay Depot, Engine Yard, (among others) and is an integral part of the Ubuntu operating system.

Couchio

About Couchio

Couchio <http://couch.io> (founded by 3 CouchDB core committers) offers support, hosting, and professional training for CouchDB.

Contact CouchDB

couchdb.apache.org

Contact Couchio

www.couch.io

hello@couch.io
