Hi Zhanying.

**"suppose ,I will divided the 34 province into 5 region , then we will have 5  queues for each of the region, when we invoke the consequent task consumer, we will  explicitly identify the queue."**

**目前，我考虑将34个省划分为5个区域，然后我们将为每个区域制定一个queue， 当我们后续将任务分发给消费者的时候，我们需要明确的指定分发的queue。**

Yes that's kind of what I was thinking. You could give the information to the getter as to which task queue to use with **environment variables.** All the file getter workers would be based on the same docker image, but have a different set of environment variables depending on the region.

是的，这就是我考虑的一种实现方案。你可以将将相关的信息传递给getter，以便getter透过环境变量（**environment variables）来决定将任务分发**到哪一个任务queue。**这样所有的file getter** 处理单元将会基于相同的的docker 镜像来构建，然而不同的**file getter** 处理单元 将根据region参数来使用不同的环境变量参数集合。

**"It is a very complex work to prepare the route data to connect all the machines of the cluster to each of the FTP servers. That is why we have to de-coupling the file getter component and the file handling component, and distributed them in different  machines."**

如果将集群中的所有机器和所有的外部OMC FTP server建立全连接，这将是一个非常复杂的工作。 这也是为什么我们不得不将file getter 组件和file handling 组件进行解除耦合的原因，同时将这两个组件分布到不同的处理机器中。

So if I understand correctly, the file getters don't run in the same data center as the workers, and there are no routes to the FTP server from the cluster of workers, but there are from the file getters and the FTP and between the file getters and the cluster?

因此，如果我理解正确，组件file getter 和组件worker 是在不同的数据中心中运行， 同时在外部的FTP server 和组件worker 集群之间不存在路由连接， 然而组件file getter 和 FTP server组件，以及组件file getter 和组件worker 集群之间存在着路由。

If this is indeed the case, then it is really crucial that the nfs chosen as where to put the files is as local as possible with respect to the workers. That's why I think you should group machines per local area, and have each group assigned to a single queue, and the files listed in the queue are only files that are located on the nfs closest to the machine group.

假如真的是这样的情况，那么NFS服务器的设计就非常重要，我们要考虑将文件放置到离组件woker尽量接近的的位置。 这也是我考虑你应该根据34个省的区域划分将处理机划分为群组的原因，同时将每个处理群组指派一个单独的queue， 同时在这个queue中放置的文件列表，仅仅分发给了哪些和NFS所在的机器上最近的worker 集群上。

In the end I think the optimal architecture would be to have two levels of queue. One geographical, that is one queue per region. Then another level, with queues for each datacenter in a region. You can achieve these two levels of routing with rabbitmq.

另外，我考虑一个优化的架构师我们应该考虑采用一个两级的queue的结构。 1个是面向地理空间关系的queue，也就是1个queue1个 区域。 而另外一个级别是面向每个区域的数据中心中计算单元集群。你可以利用rabbitmq的路由功能来建立一个两个层级的queue。

This for me would be the most resilient because all of the regions and data centers would be independant so if one queue fails, the other can continue without interruptions. And it doesn't seem to difficult to implement. Good luck!

这对我来说是最有弹性的，因为所有的区域和数据中心都是独立的，所以如果一个队列失败，另一个队列可以继续而不会中断。 它似乎并不难实现。 祝好运！

Could you tell me more about your company and what you are trying to achieve in terms of business?

您能告诉我更多关于贵公司以及您在业务方面想要实现的目标吗？

Anis