课程总结



徐 辰 cxu@dase.ecnu.edu.cn

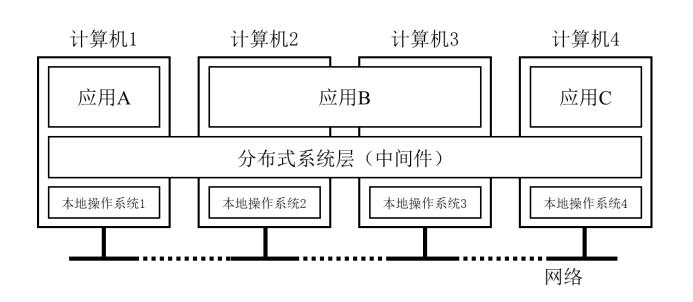
華東师絕大學

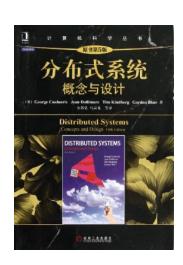


- □分布式计算
- □课程模块回顾
- □课程知识体系
- □发展趋势
- □课程考核

分布式系统的定义

□《分布式系统:概念与设计》认为:分布式系统:概念与设计》认为:分布式系统是若干独立计算机的集合,这些计算机对于用户来说就像是一个单机的系统









分布式计算系统的定义

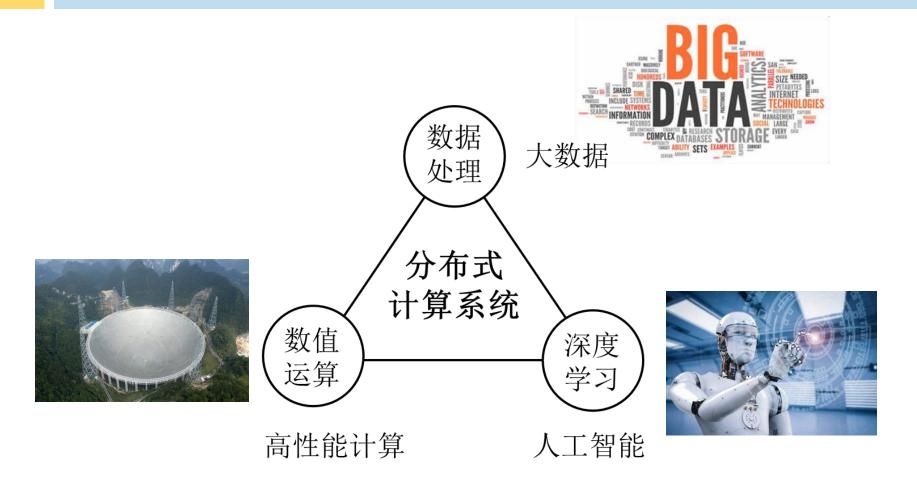
□ 分布式计算系统是作用于若干独立计算机 之上、使得这些计算机能够协同执行计算 完成某项应用的软件系统

□ 分布式计算系统归根结底是为了解决某些 类别的应用问题而设计的分布式系统





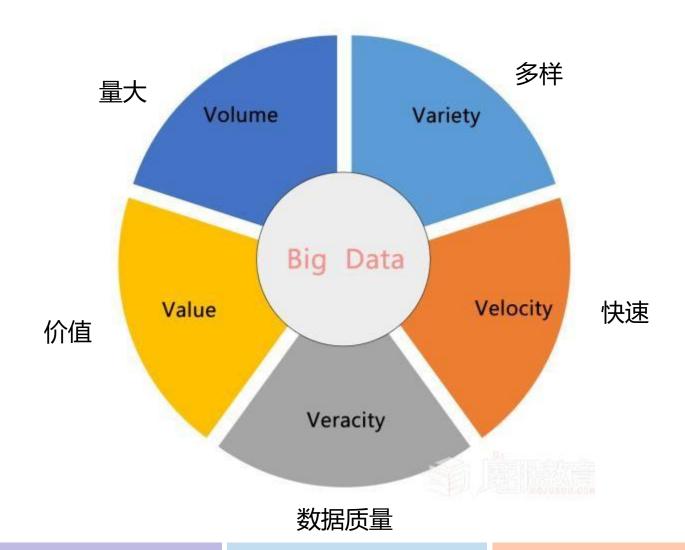
应用驱动分布式计算系统的发展创新







大数据5V





分布式系统遇到大数据

分布式系统





Big data landscape

BIG DATA & AI LANDSCAPE 2018 INFRASTRUCTURE APPLICATIONS - ENTERPRISE HADOOP ON-PREMISE STREAMING / IN-MEMORY -DATA ANALYST PLATFORMS — DATA SCIENCE PLATFORMS HADOOP IN THE CLOUD MARKETING - B2B CUSTOMER SERVICE cloudera Hortonwork aws sdatabricks striim Microsoft opentaho alteryx einstein CHORUS RADIUS Ann Annie MEDALLIA zendesk aws Microsoft Azure IBM AKNIME A data CLARABRIDGE pigital guavus AYASDI INSIDESALES.COM Lattice BlueYonder [PERSADO] @kahuna MAPR Pivotal confluent GridGain TOOMING mapidminer ATTIV/O Datameer Quid incorto. conversica ACTIONIQ SAILTHRU B BLUECORE dataArtisans | hazelcast MINTIGO ÓSENSE IBM InfoSphere TREASURE Dubole () altiscale CONTINUUM' ALGORITHMIA → tubular Datafox Refl QUANTIFIND ## mparticle Amplero. TERRACOTTA KX ST FASTDAT inter and ClearStory Origami clari A aviso tact.ai OUTCOMOT & frame at A bluedata jethro CAZENA CenturyLink fuse machines TROOPS #NGAGIO TO Samperity TROOPS & Simon Lytics ⊕msgai | INTERCOM CallDes WallarooLASS D. DATAWATCH SSAS **HUMAN CAPITAL** -LEGAL FINANCE -ENTERPRISE BACK OFFICE ¬ NoSQL DATABASES -MPP DBs ___ CLOUD EDW BI PLATFORMS VISUALIZATION MACHINE LEARNING PRODUCTIVITY How Vie entelo /inaplan Clustrix 📦 Google Cloud aws @neo4j aws @Sea StackPath 🔯 illumio 🔼 CODE42 🔮 CoherClou +ableau VERTICA hil GIGSTER Pivotal ORACLE Microsoft Azus △ Google Cloud H₂O zuora △ Google Cloud cetonis JUDICATA ORACLE IBM textic sestusses mongoDB MarkLogic SBREVIA MEMSQL () influxdata DataRobot ORACLE looker @november Oction Qlik @ Periscope gamalon ®Wode&Wendy Stella WOLTER VOLTER Pivotal. ATSCALE ELEMENT" Rognitio ⊚ clara talla□ ZEPL ZOMEATA PIOTIY SweTalon | Seconder feedzal Cyber (a) Custree (a) pymetrics @citusdata splice snowflake VISENZE ROSS Exasol butteral Kasisto MicroStrategy * birst mva uncommon CHARTIO PTOUCAN TOCO redislabs SCYLLA. APPLICATIONS - INDUSTRY COMPUTER VISION -SPEECH & NLP -ADVERTISING -**EDUCATION** -GOVERNMENT -△ Google Cloud (□) twillio MEM Watson Cortana, Escott BIT-20 TRANSFORMATION Microsoft Azure (OPENGOV REDFIN INVESTING ondeck Affirm convolesa narrative o O rubrik O APPDYNAMICS Sentient Voyager in worker criteol. PxAd Ontegral •talend pentaho Opendoo HAMPI CHANDILA snaplogic enigma @ podur Lemonade ORACLE OpenX and datorary VTS Kreditech AVANT Ordynatrace In V SoundHound Inc. **G**ieclara ○ 西湿 alooma xolehty ZALONI splunk> SignalFx Odruug A Numeria PETUUM SI kidaptive dstillery Livelntent Passport ₩ Moogsoft unravel # EVER Al @ deepomatic naralogics - CUROUSAL OSARO SENTIUM (WeLab We V Stitch import.io PANORAMA TAPAD datax A CHRUTA OKERA pagerduty ** Numerify knowre LOG ANALYTICS - SOCIAL ANALYTICS - WER / MORILE / -SEARCH . CLUSTER SVCS - APP DEV - CROWD -LIFE SCIENCES TRANSPORTATION -AGRICULTURE 1 COMMERCE - INDUSTRIAL COMMERCE ANALYTICS SOURCING splunk> Hootsuite spn Google TPU arm kinstica flatiron Clover KYRUUS HealthTop® Xassame Color Charges UBER TESLA Google Analytics TITCH FIX sumologic NETBASE CLEARPATH mixpanel A MPUTUS 5 Granular S synthesio Systems MYTHIC WUXINextCODE ZEPHYS Uowork *appen sumal Airtable 30Med 2 zebro Patria OVIA TACHYUS & g swiftype algolia Z Quint green RBLUERIVER floure eight scole nimblestorage RESCI ESIGOPT bitly predata A PILOTAI ANIO Core Corporation **PRI AZINGOR** OTHER -Brytlyt PG-Strom granify custora omni:us SINEQUA (logz.io **Similar**Web OPTIMUS ♥moovit FarmLogs nevar CITRINE CROSS-INFRASTRUCTURE/ANALYTICS twoRAR comma.ai **©** GAMAYA Atomake @ deep garamen N netrodyn Unbabel € Juleotock Ø Selond 0 aWS O Google Cloud # Microsoft IN Sevent Peace Sas 1010DATA VMWare TIBCO TERADATA ORACLE Netapp \$ syncson MAPR cloudera (X OWKN FRAMEWORK AI / MACHINE LEARNING / DEEP LEARNING LOGGING & SECURITY COORDINATION STREAMING SEARCH VISUALIZATION -COLLABORATION •talend python" TensorFlow theano Gerdana (anifi mongoDB Spark (A) Anache Ranger R K kibana KNOX Flink 3 beam Solr SLAMDATA DRILL Rodeo KSCIDE Chainer VELES WEKA A SENTRY | logstash 0 & kafka | druid 2 Nanty W OPENTAGE STICK DIMSUM SciPu neon" Spark DL4. MAHOUT Aerosolve HBASE (Charleson SCCUMULO Apache Airflow DSSTNE **DATA SOURCES & APIs** DATA RESOURCES FINANCIAL & ECONOMIC DATA LOCATION INTELLIGENCE DATA SERVICES **INCUBATORS & SCHOOLS** AIR / SPACE / SEA PEOPLE / ENTITIES Ontaital Insight Planet. Q Palantir 🔞 qualtries Bloombera (THOMSON REUTERS D | DOW JONES FOURSQUARE O mapbox acxiem. experian. OpenAI **OPERA** EPSILON SInsideView SAP CAPITAL IQ CBINSGHIS Xignite Quandle data secrid △ spire kespry @ @ me DATA SCIENCE DataCamp 24 DataElite PREMISE @estimize second second Place esri : factual enigma # fitbit GARMIN fractal kaggle VECTOR INSTITUTE Quantcast WINDWARD | telluslabs △ Mapillary CRUX BASIS ALLEN INSTITUTE SAFEGRAPH EXL innercxus





- □分布式计算
- □课程模块回顾
- □课程知识体系
- □发展趋势
- □课程考核



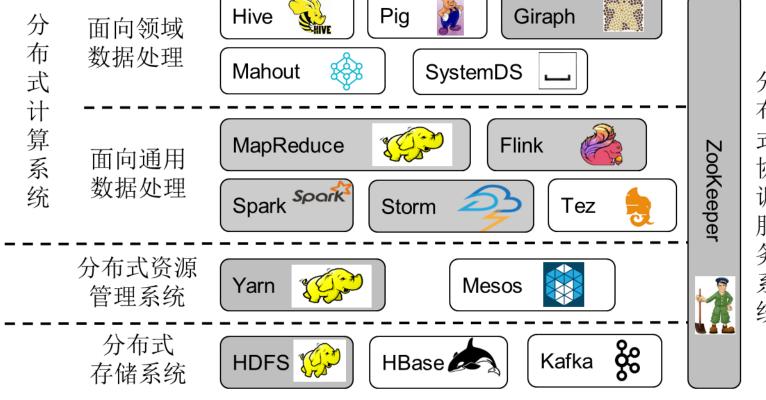
面向数据密集型应用的分布式计算系统

应用	类型	代表性系统		
通用数据处理	批处理系统	MapReduce, Spark, Tez		
	流计算系统	Storm、Spark Streaming		
	批流融合系统	Flink、Structured Streaming		
领域数据处理	图处理系统	Giraph、GraphX		
	机器学习系统	Mahout, SystemDS, Mllib		
	查询分析系统	Hive, Pig		





分布式计算系统生态圈



分布式协调服务系统



大纲

- □分布式计算
- □课程模块回顾
- □课程知识体系
- □发展趋势
- □课程考核



知识体系

- □设计思想
- □ 体系架构
- □工作原理
- □ 容错机制
- □编程示例

内容组织

- □ 设计思想:
 - ▲该系统应对的应用场景如何?
 - →为什么需要构造这个系统?
 - ▲该系统所处理的数据模型是怎样的?
 - ▲该系统采用何种计算模型?
 - ♣如果系统支持迭代计算,那么采用何种迭代模型?
 - **4**



设计思想

□数据模型

系统	数据模型	
MapReduce	Key-value	
Spark	RDD	
Storm	Tuple	
Spark Streaming	Dstream	
Flink	DataStream	
Giraph	Graph	





设计思想

□计算模型

系统	计算模型	
MapReduce	DAG	
Spark	DAG	
Storm	Topology (DAG)	
Spark Streaming	DAG	
Flink	DAG	
Giraph	Vertex-centric	





设计思想

□ 迭代模型

系统	迭代模型	
MapReduce		
Spark		
Storm		
Spark Streaming		
Flink	$\sqrt{}$	
Giraph	$\sqrt{}$	





内容组织

□ 体系架构:

- ♣作为一个分布式系统,该系统包含哪些部件,即由哪些进程或线程构成?
- →各个部件都扮演着怎样的角色?
- ♣系统中的各个部件都以怎样的次序共同工作才能保障用户应用程序的顺利执行?
- **4**





- □ 架构图
- □应用程序执行流程





内容组织

□ 工作原理:

- ♣系统如何根据用户编写的应用程序生成相应的 执行计划,以及是否可以进行优化?
- ♣面对数据密集型应用,系统<mark>如何减少数据I/O</mark>的 代价?
 - ➤ 这里I/O的代价表现为节点本地读写数据的磁盘I/O 代价以及节点之间通过网络进行数据传输的代价, 后者在分布式场景下显得尤为突出
- 📥





工作原理

□数据输入、数据转换、数据输出

- □讨论主题
 - ♣执行引擎不同Task(运行某个或某些算子)之 间如何进行数据交换?
 - ♣Flink:如何基于执行引擎来实现Dataflow编程模型?



内容组织

□ 容错机制:

- 早期的并行计算系统部署在高性能服务器甚至 是超级计算机当中,通常不需要考虑容错问题
- ♣商用服务器集群中的节点发生故障是常见的现象,系统如何保证在某些节点有故障的情况下不影响用户应用程序正常执行,即故障对用户透明?
- 榫



容错机制

- □故障类型
- □ 检查点 (持久化)
- □故障恢复



内容组织

- □ 编程示例:
 - ♣使用这些系统的编程接口进行简单编程的一般 步骤是怎样的?
 - ♣使用这些系统进行编程是否有通用的框架?如何提高某些程序的性能?
 - ♣使用不同系统实现同一个应用所编写的程序有何差异?
 - 📥



编程示例

- □词频统计
- PageRank
- □ K均值聚类



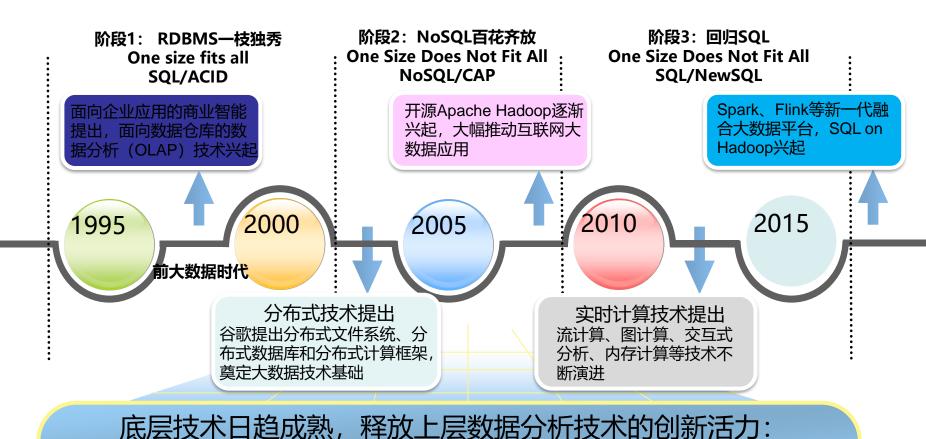
大纲

- □分布式计算
- □课程模块回顾
- □课程知识体系
- □发展趋势
- □课程考核





大数据管理技术的演进

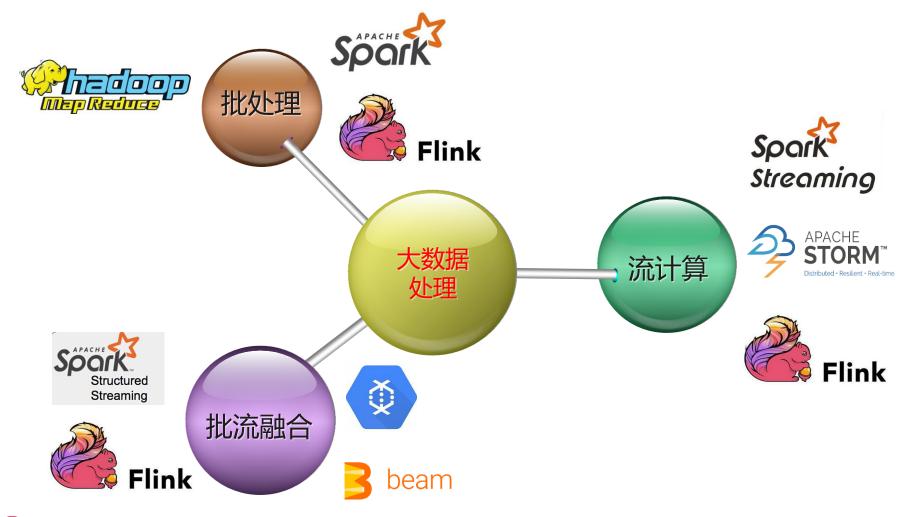


实时流分析、交互分析、机器学习





大数据处理系统







大数据处理系统的演进

Fourth Generation

In-memory + Out of Core Performance, Declarativity, Optimization of Iterative Algorithms, True Streaming e.g. Apache Flink

Third Generation

In-memory Performance and Improved Programming Model, e.g. Apache Spark

Second Generation

Scale-out, Map/Reduce, UDFs, e.g., Apache Hadoop

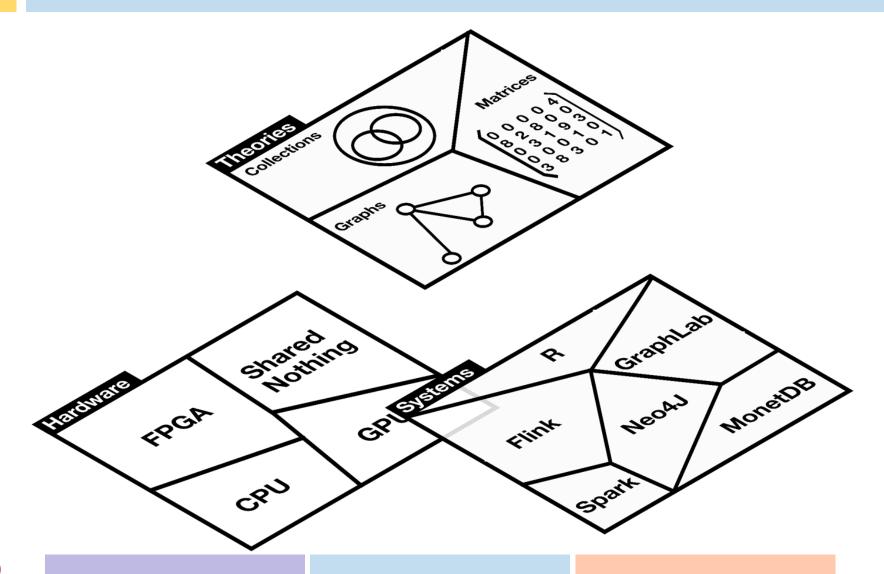
First Generation

Data Warehouses, e.g, relational DBMS





Vision: Mosaics





未来趋势

□ 算法: 与系统的深度结合

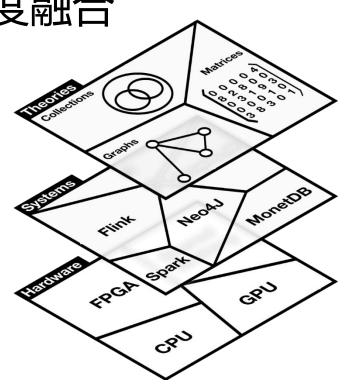
□ 系统:与AI、DB系统的深度融合

□ 硬件:新体系架构

▲运算: GPU

♣存储: NVM

₩网络: RDMA







大纲

- □分布式计算
- □课程模块回顾
- □课程知识体系
- □发展趋势
- □课程考核





期末考试

题型	每题分值	数量	单项总分
选择题	2	5	10
判断题	1	10	10
填空题	2	10	20
简答题	5-6	5-6	30
分析论述题	10	3	30





题型示例

一、单项选择题

- 1. 以下哪一个系统不适合用于流计算(A)
- A. MapReduce B. Spark C. Storm D. Flink
- 2. 分布式系统中常用的"选主"功能通常使用哪一个系统实现(C)
- A. MapReduce B. Yarn C. ZooKeeper D. Mesos
- 3. 在批流融合系统中,以下哪个系统认为批处理被包含在流计算当中(B)
- A. Spark B. Flink C. Beam D. 以上答案都不对

二、判断题

- 1. 类似于 Spark Streaming, Hadoop Streaming 提供了一个工具库使得 MapReduce 程序能够支持流计算(×)
- 2. 远程过程调用在分布式数据处理引擎中广泛使用(√)

三、 填空题

- 1. 在 MapReduce 系统中,Map 和 Reduce 二者之间数据传输的阶段称为___slnuffle__。
- 2. 在 Storm 系统的架构中,作为主节点的部件的名称是__nimbus___。

四、 简答题

1. 为什么 Storm 无法满足 exactly-once 的容错语义?

参考 storm 容错部分内容

五、 论述题

1. 画出 Lambda 架构图,简述各部件的功能并指出该架构的不足之处。

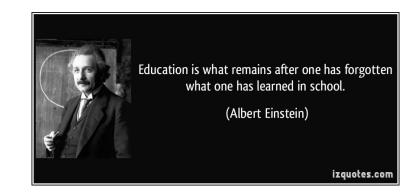
参考批流融合概述部分内容



课程成绩评定

- □ 平时成绩: 50%
 - →考勤: 10%
 - ▲课堂讨论: 15%
 - →编程作业 (2+1次) : 30%
 - ♣上机实验(10个实验共3组): 45%

- □期末成绩:50%
 - ▲期末考试: 100%







On Education

I want to oppose the idea that the school has to teach directly that special knowledge and those accomplishments which one has to use later directly in life. The demands of life are much too manifold to let such a specialized training in school appear possible. Apart from that, it seems to me, moreover, objectionable to treat the individual like a dead tool. The school should always have as its aim that the young man leave it as a harmonious personality, not as a specialist. This in my opinion is true in a certain sense even for technical schools, whose students will devote themselves to a quite definite profession. The development of general ability for independent thinking and judgement should always be placed foremost, not the acquisition of special knowledge. If a person masters the fundamentals of his subject and has learned to think and work independently, he will surely find his way and besides will better be able to adapt himself to progress and changes than the person whose training principally consists in the acquiring the detailed knowledge. (On Education: Einstein, 1931)





论教育

我想反对这样的想法——学校必须直接教那些专业知识 和那些在今后的人生中能够直接使用的技能。生活的需 求是如此多样,以至于要在学校进行这样的专业训练似 乎是不可能的。除此之外,在我看来,令人反感的是, (这样的教育) 像一个死工具一样对待人。学校应该始 终把学生培养为具备和谐人格的人作为目标,而不是仅 培养专家。在我看来,即使对于那些为某个特定职业培 养学生的技术学校, 我的这一点观点在某种程度上也是 正确的。应该始终把独立思考和判断的一般性能力的培 养放在首位,而不是获取专业知识。如果一个人掌握了 某个学科的基础,并且学会了独立思考和工作,他就一 定能够找到属于自己的方式; 另外, 相对那些主要的训 练都集中在获得某种具体知识的人们,他也能让自己更 好地适应各种发展和变化。



谢谢! Q&A





