

EXTREME COMPUTING GROUP

Defining the future.

Building Cloud Applications to Support Research:
the Microsoft Azure Research Engagement Project

Cloud Computing Research Engagement Initiative

Alexander Voss, Christian Geuer-Pollmann, Dennis Gannon, Fabrizio Gagliardi, Goetz Brasche, Jared Jackson, Nelson Araujo, Roger Barga, Wei Lu

- **International Engagements.**

- Offer cloud services to academic and research communities worldwide
- back up this offering with a technical engagements team.

- **Data.**


- Provide select reference data sets on Azure to enable communities of researchers.

- **Services for Research.**

- Provide applications and core services for research.

- **Ask the questions**, what does it take to catalyze a community of researchers, what are the core services, key products needed to support research.

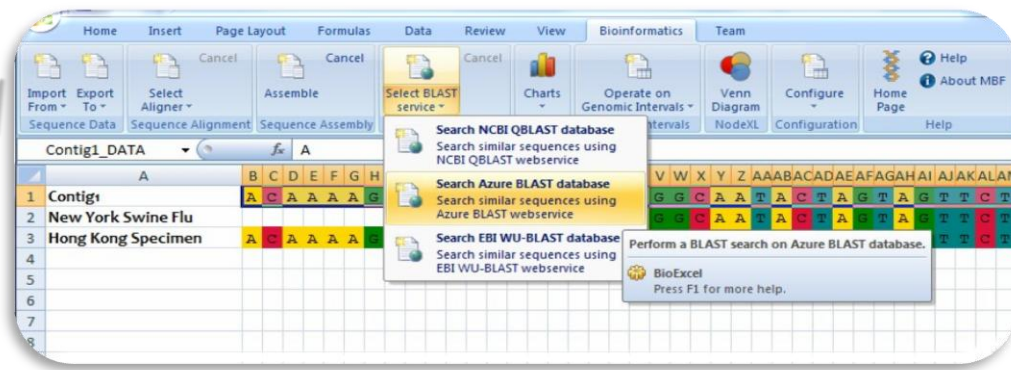
Microsoft's Goals for this Project

- Demonstrate that a client+cloud model can revolutionize research and learning
 - Illustrate that cloud computing is a cost-effective and easy-to-use way to outsource select components of research infrastructure
 - Provide feedback from research community to our product groups
 - Establish the Microsoft Cloud Computing platform as leader and trendsetter for basic research
- 

Some examples based on our ongoing
Cloud Research Engagements

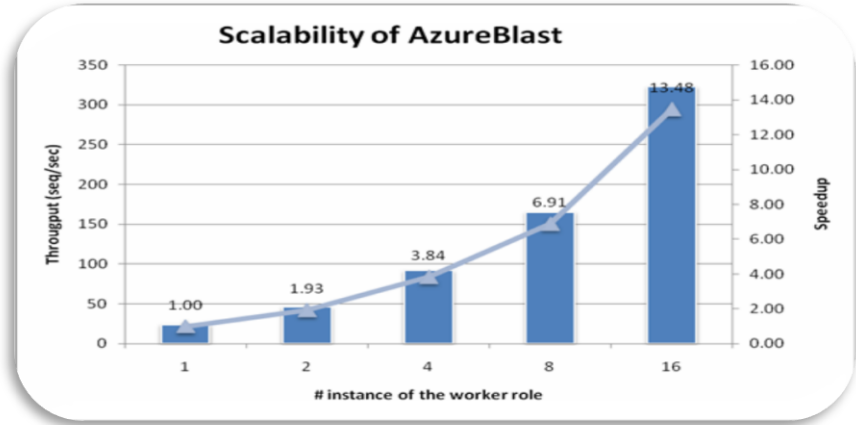
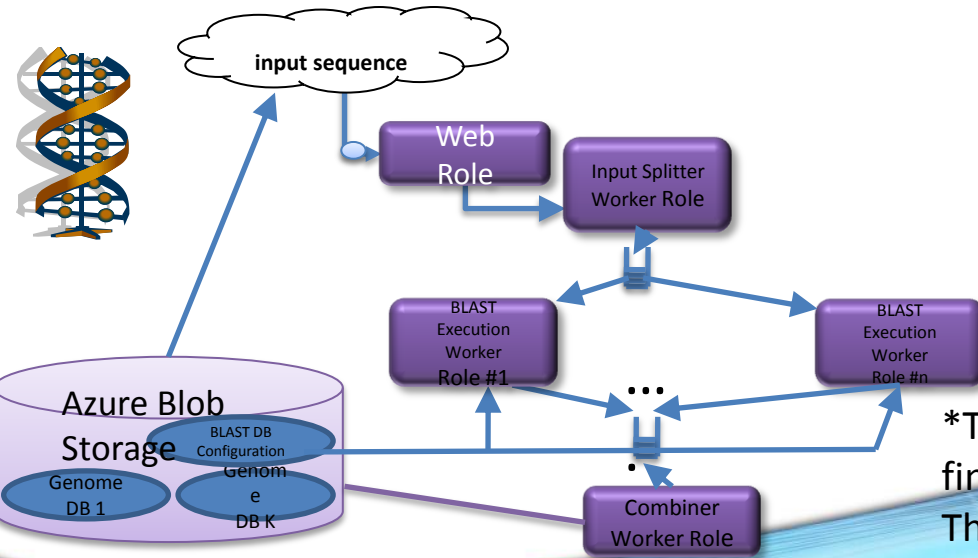


AzureBLAST*



Seamless Experience

- Evaluate data and invoke computational models from Excel.
- Computationally heavy analysis done close to large database of curated data.
- Scalable for large, surge computationally heavy analysis.
- Test local, run on the cloud.



*The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases

Making Excel the user interface to the cloud

The image shows a screenshot of Microsoft Excel with a custom ribbon titled 'Bioinformatics'. The ribbon includes the following groups and options:

- Home**: Import From, Export To, Select Aligner, Assemble
- Cancel**: Cancel (multiple instances)
- Select BLAST service**: A dropdown menu with three options:
 - Search NCBI QBLAST database**: Search similar sequences using NCBI QBLAST webservice
 - Search Azure BLAST database**: Search similar sequences using Azure BLAST webservice
 - Search EBI WU-BLAST database**: Search similar sequences using EBI WU-BLAST webservice
- Charts**: Charts
- Operate on Genomic Intervals**: Operate on Genomic Intervals
- Venn Diagram**: Venn Diagram, NodeXL
- Configure**: Configuration
- Help**: Home Page, Help, About MBF

The spreadsheet below the ribbon shows a table with columns A through Z and rows 1 through 8. The data in row 1 is as follows:

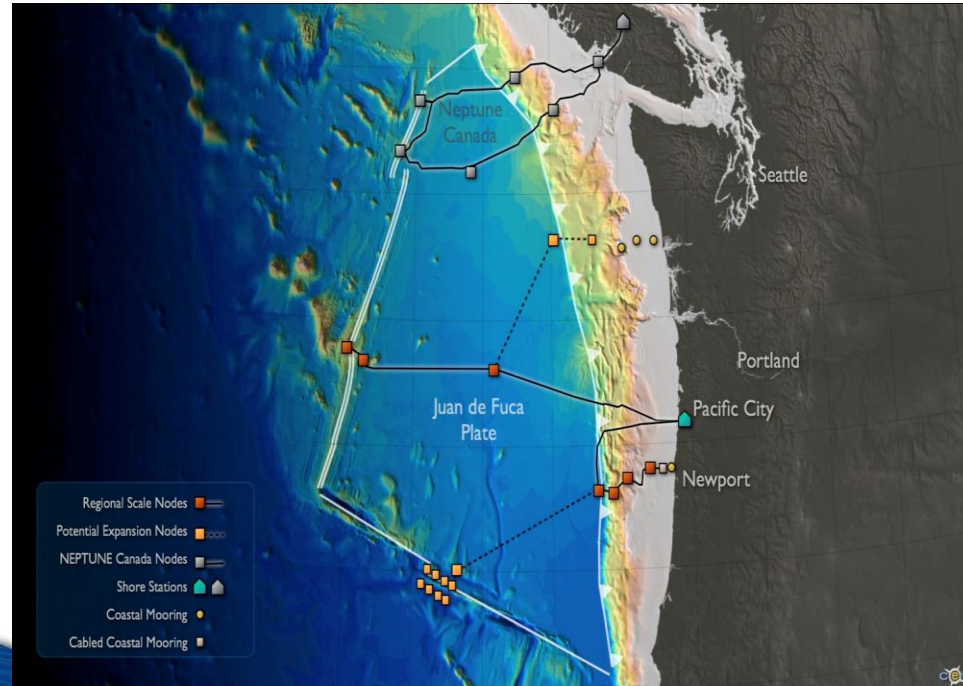
	A	B	C	D	E	F	G	H	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	Contig1	A	C	A	A	A	A	G	G	G	C	A	A	T	A	C	T	A	G	T	A	G	T	T	C	T
2	New York Swine Flu	A	C	A	A	A	A	G	G	G	C	A	A	T	A	C	T	A	G	T	A	G	T	T	C	T
3	Hong Kong Specimen	A	C	A	A	A	A	G	G	G	C	A	A	T	A	C	T	A	G	T	A	G	T	T	C	T
4																										
5																										
6																										
7																										
8																										

A BioExcel notification box is visible in the bottom right corner, stating: "BioExcel Press F1 for more help."

Supporting Smart Sensors and Data Fusion

The NSF Ocean Observing Initiative

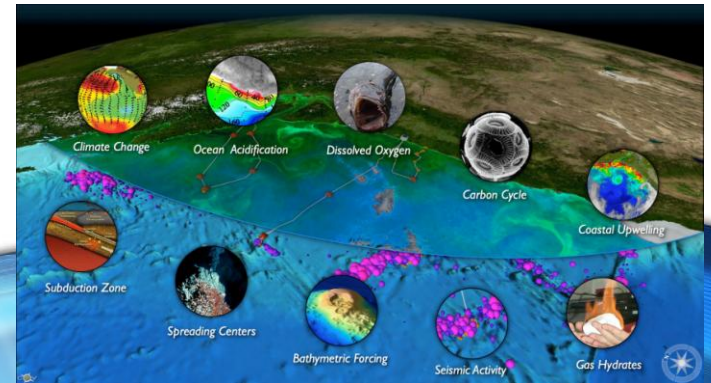
- Hundreds of cabled sensors and robots exploring the sea floor
- Data to be collected, curated, mined
- OOI Architecture plan of record, store this data in the cloud



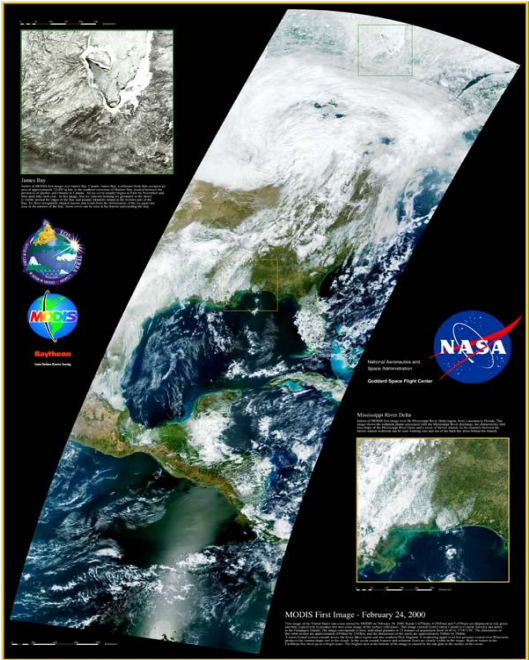
Data collected from:

- Ocean floor sensors, AUV tracks, ship-side cruises, computational models

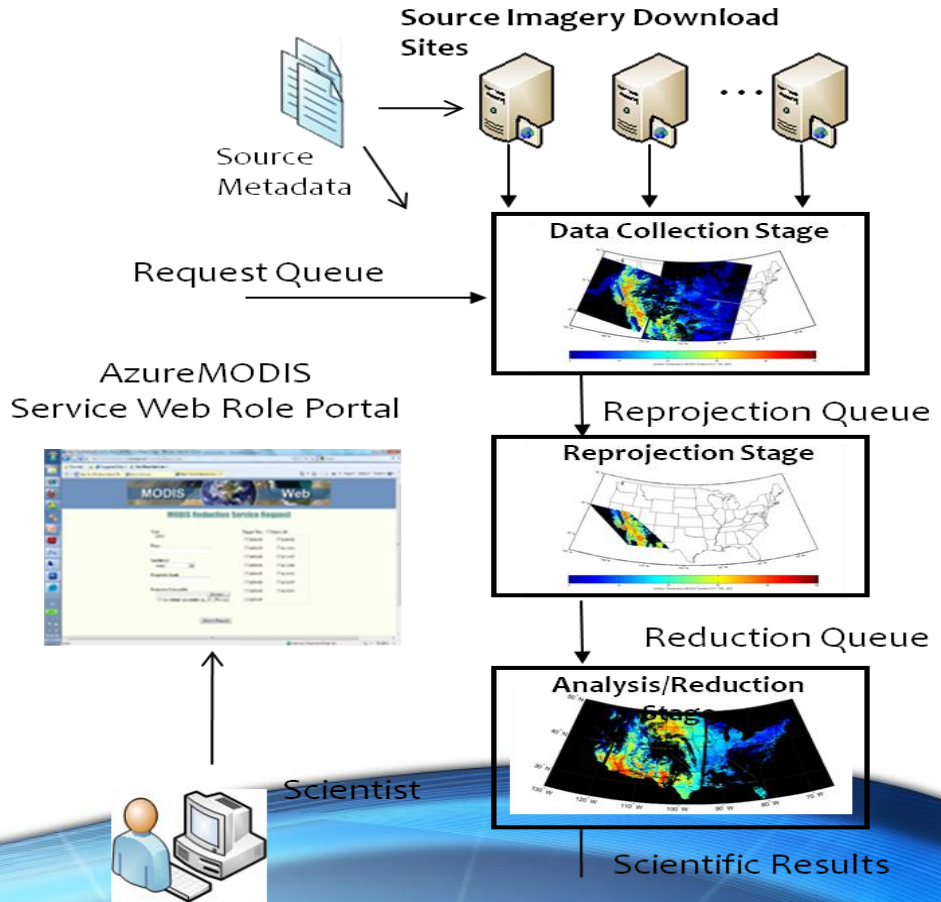
Data moves from **ocean** to shore side **data center** to the **Azure cloud** to your **computer**.



AzureMODIS – Azure Service for Remote Sensing Geoscience



- 5 TB (~600K files) upload of 9 different imagery products from 15 different locations (~6 days of download)
- 4 TB reprojected harmonized imagery ~35000 cpu hours
- 50 GB reduced science variable results ~18000 cpu hours (~14 hour download)
- 50 GB additional reduced science analysis results ~18000 cpu hours (~14 hour download)

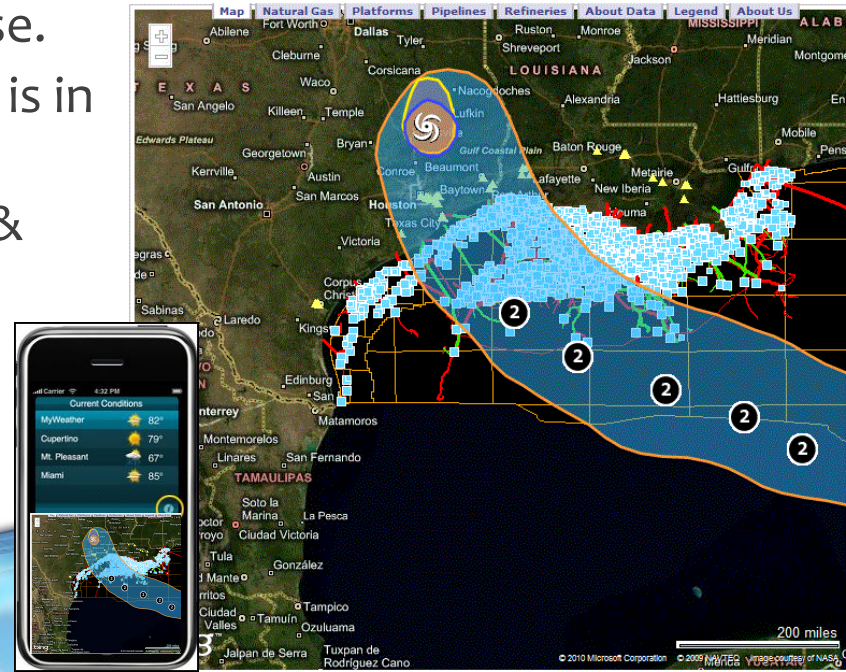


Potential Cloud Applications

- The cloud as a host to data collections
 - *Classical http data* - Web n-grams, photos, blogs, etc.
 - *Geographic & Semi-static* – Maps, satellite, geological records, weather radar, field reports/observations, geo-sensors (seismic, ocean)
 - *Geo-tagged Social record* - census, public government records, local newspapers, health and police records
 - *Scientific* – genome, pathology data, fMRI, astronomy, chemical
 - *Dynamic shared state* - multiplayer games, virtual worlds, social networks
 - *My context* - my devices, my application sessions, my agents.
- 

Geographic & semi-static data

- Maps, satellite, geological records, weather radar, field reports/observations, geo-sensors.
 - Mobile apps for emergency response.
 - Tell me when me or my team/family is in danger.
 - Integrate emergency data streams & field report tweets in the cloud and push to my client devices.



Geo-tagged social record

- Census, public government records, local newspapers, health and police records
 - Build cloud tools to understand societies problems
 - » Links between history of disease, poverty and pollution
 - » Correlate crime rates and bad weather
 - Discover the lost heritage of a people or a place



Dynamic shared state

- multiplayer games, virtual worlds, social networks
 - New modalities of collaboration



Reaching Out: Azure Research Engagement project

In the U.S.

Memorandum of Understanding with the National Science Foundation

- Provide a substantial Azure resource as a donation to NSF
- NSF will provide funding to researchers to use this resource

In Europe

- We interested in direct engagement with the thought leaders in the U.K., France and Germany
- EC engagements where possible

In both we provide our engagement team

- We provide workshops, tutorials, best practices and shared services, learn from this community, shape policy...

In Asia

- We wish to explore possibilities.
- 

VENUS-C




Virtual multidisciplinary EnvironMents USing Cloud infrastructures

*Funding Scheme: Combination of Collaborative Project
and Coordination and Support Action: Integrated
Infrastructure Initiative (I3)*

*Program Topic: INFRA-2010-2 1.2.1. Distributed
Computing Infrastructures*

Goals

1. Create a platform that enables user applications to leverage cloud computing principles and benefits.
 2. Leverage the state of the art to on-board early adopters quickly, incrementally enable interop with existing DCI and push the state of the art where needed to satisfy on-boarding and interop
 3. Create a sustainable infrastructure that enables the cloud computing paradigms for the user communities inside the project, the one from the call for applications, as well as others.
- 

Supporting multiple basic research disciplines

Biomedicine: Integrating widely used tools for Bioinformatics (UPV), System Biology (CosBI) and Drug Discovery (NCL) into the VENUS-C infrastructure

Civil Protection and Emergency: Early fire risk detection (AEG), through an application that will run models on the VENUS-C infrastructure, based on multiple data sources.

Civil Engineering: Support complex computing tasks on Building Information Management for green constructions (provided by COLB) and dynamic building structure analysis (provided by UPV).

D4Science: Integrating computing through VENUS-C on data repositories (CNR). In particular focus will be on Marine Biodiversity through Aquamaps.

Open Call for 20 e-Science Applications

20K€ funding each (in addition to Azure Compute ,
Storage and Network Resources)

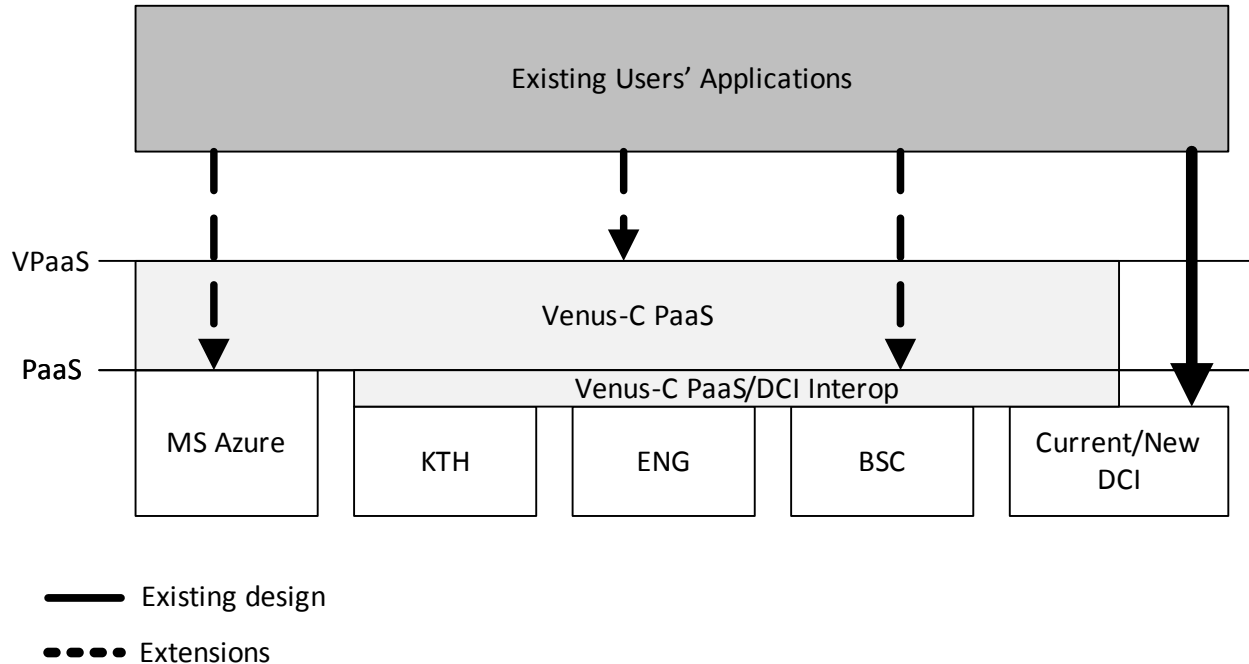
-> porting applications to the cloud

-> education and training

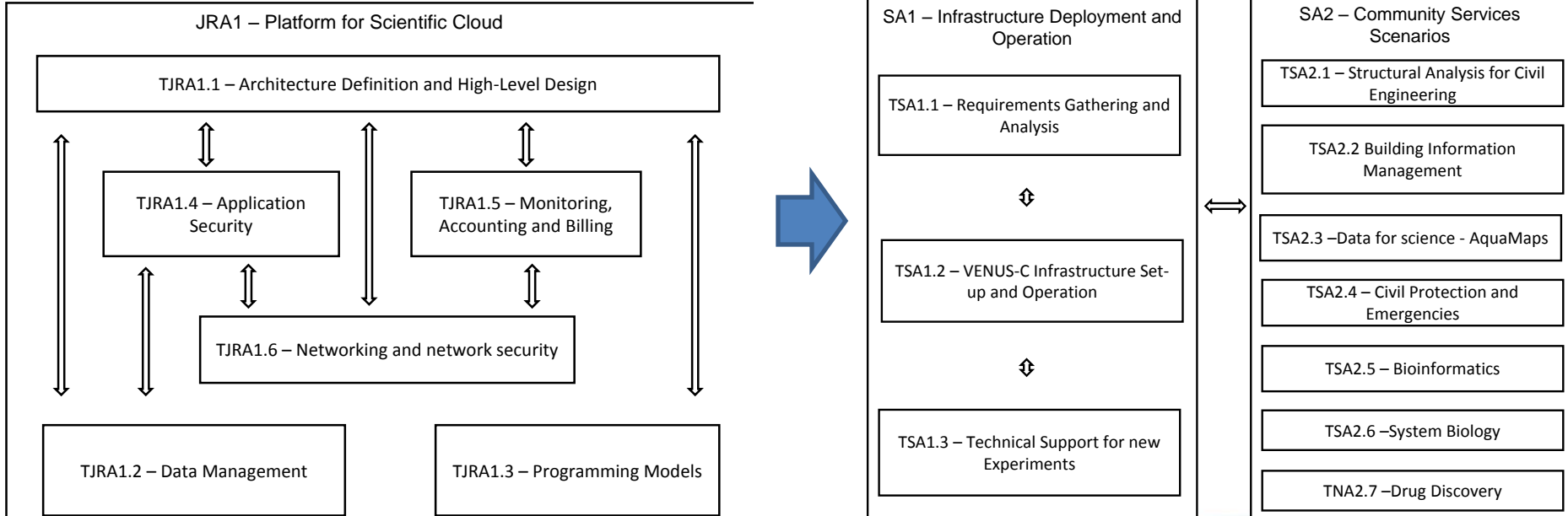
-> scalability tests



Software + Services to facilitate e-Science Applications in the Cloud



Strategy/ Approach



Consortium

1 (co)	Engineering Ingegneria Informatica S.p.a.	ENG	IT
2	European Microsoft Innovation Centre	EMIC	DE
3	European Charter of Open Grid Forum	OGF.eeig	UK
4	Barcelona Supercomputing Center – Centro Nacional de Supercomputación	BSC-CNS	ES
5	Universidad Politecnica de Valencia	UPV	ES
6	Kungliga Tekniska Hoegskolan	KTH	SE
7	University of the Aegean	AEG	GR
8	Technion	TECH	IL
9	Centre for Computational and Systems Biology	CoSBi	IT
10	University of Newcastle	NCL	UK
11	Consiglio Nazionale delle Ricerche	CNR	IT
12	Collaboratorio	COLB	IT

Questions?

